

Nanoscience

P. Boisseau • P. Houdy • M. Lahmani
(Eds.)

Nanoscience

Nanobiotechnology and Nanobiology

With 628 Figures and 40 Tables



EUROPEAN MATERIALS
RESEARCH SOCIETY



Springer

Editors

Patrick Boisseau
CEA
LETI-MINATEC
17 rue des martyrs
38054 Grenoble CX 9
France
patrick.boisseau@cea.fr

Prof. Philippe Houdy
Université d'Evry
bd. F. Mitterrand
91025 Evry CX
France
philippe.houdy@univ-evry.fr

Marcel Lahmani
Université d'Evry
Dépt. Sciences des Matériaux
rue du père Jarlan
91025 Evry CX
France
marcel.lahmani@univ-evry.fr

Translation from the French language edition of
“Les nanosciences – 3. Nanobiotechnologies et nanobiologie”
© 2007 Editions Belin, France

ISBN 978-3-540-88632-7 e-ISBN 978-3-540-88633-4
DOI 10.1007/978-3-540-88633-4
Springer Heidelberg Dordrecht London New York

Library of Congress Control Number: 2009926250

© Springer-Verlag Berlin Heidelberg 2010

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilm or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer. Violations are liable to prosecution under the German Copyright Law.

The use of general descriptive names, registered names, trademarks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

Cover design: WMX Design GmbH

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Foreword to the French Edition

The Living Nanoworld

The slogan ‘small is beautiful’ is perfectly suited to the field of biology. The human body is composed of about five thousand billion cells, each of which functions by virtue of a whole range of nanoscale phenomena and nanomachines. The perfect harmony of the information systems and molecular devices at work in our cells is today a fertile source of inspiration for scientists engaged in the development of nanotechnology and nanomaterials. Within a volume of a few picoliters, DNA stores absolutely all the genetic information required to program embryonic development, cell differentiation, and the functioning of living organisms. It also contains all the machinery and molecular systems needed to replicate this information and distribute it to daughter cells during cell division. Exchanges between the interior of a cell and its environment, but also between the different compartments of a given cell, occur via a whole system of complex valves with finely tuned aperture, the ion channels. ATP, the molecular form of chemical energy storage, is synthesised by micromotors in which the rotor is turned by a flow of protons. Convoys purvey the constituents of cells in every direction through a dense network of microtubules and microfilaments. Bacteria, protozoans, and the cells of metazoans move themselves around by means of a range of different types of motors, flagella, or microhelices.

Naturally, these biological nanomachines constitute a wonderful source of inspiration for research scientists keen to reproduce their achievements for scientific or industrial purposes. One aim is to perfect methods for exploring the ultimate structure of biological objects: DNA, protein, or cell chips, nanoparticles revolutionising the field of microscopy; optical tweezers for micromanipulating nanoscale entities or measuring the forces acting on them. Several types of bioassay can today be brought together in a nanolaboratory, or lab-on-a-chip, with astonishing properties.

One of the challenges in this kind of exploration of the cellular nanoworld, when using some of its solutions to achieve the same level of performance, is

to build molecular nanorobots capable of carrying out investigative tasks or medical treatment in the inner workings of the human body, if not within the cells themselves.

In the third volume of this series devoted to the nanosciences, readers will realise with amazement and wonder just how ingenious are the molecular systems retained by natural selection. No doubt they will then share the hopes and aspirations of the present authors to devise ways for humankind to design *de novo* nanomachines imitating biological systems, with a view to understanding those systems better and where necessary finding solutions when they malfunction; in a word, with a view to fighting illness.

Institut Cochin
IFR Alfred Jost
August 2007

Axel Kahn

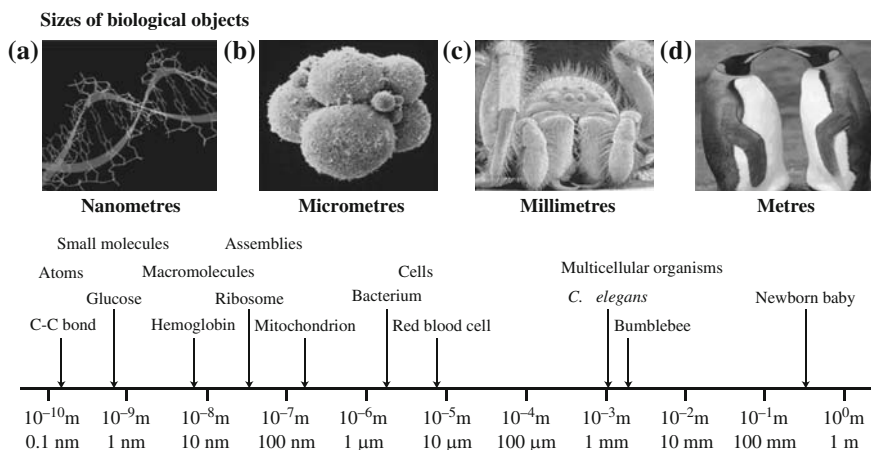
Preface to the French Edition

The Size of Biological Entities

From ancient times until the invention of the optical microscope by Van Leeuwenhoek in the seventeenth century, humankind could only study those biological objects that were visible to the naked eye, which means to say, bigger than one millimetre across. Optical microscopy took observation down to the micron, an improvement by a factor of a thousand. Then, in the last century, the invention of the electron microscope opened up the world on a scale of ten nanometers. Today, the nanometer and even the angstrom unit (0.1 nm) mark the limits of resolution in a whole range of modern molecular imaging techniques. Indeed, since the middle of the twentieth century, crystallographic studies of DNA and proteins have gradually revealed the structure of biological entities with a resolution slightly greater than 0.1 nm, which is the size of a hydrogen atom. Extremely complex protein molecules, built up from several smaller proteins, have now been successfully characterised and their three-dimensional coordinates can be obtained from data bases.

Molecular imaging has made considerable progress since the heroic work carried out by crystallographers at the Cavendish Laboratory in Cambridge. It is possible today to obtain the complete structure of a single molecule, whatever its size. The most difficult thing when working with a single molecule, if it is complex, is to be able to visualise its motion within a protein or, an even more delicate task, within a multiprotein structure. Electron microscopy and crystallographic techniques can usually only be applied to preparations in which there is no molecular motion, although over the past twenty years or so, NMR techniques have become available to follow the internal motions of small proteins with a size of a few nanometers.

What is really new about the novel techniques described in the present book is that it is now possible to study the functioning of a living cell with nanometric resolution, i.e., on a scale a million times smaller than what can be observed with the naked eye. This is a genuine technological revolution, not restricted to observation of the living world, but which can be used to build



biological objects. One can now contemplate the reconstruction of a cell from its molecular components. Such a construction will exploit the self-assembly properties of biological constituents, using the fact that proteins recognise each other and can self-assemble. But it will also be possible to guide this assembly process using recently developed physical techniques.

The Convergence of Nanoscience and the Life Sciences

Physics and biology have long been in contact. The beginning of the twentieth century saw the fruitful collaboration between chemistry and biology, which made it possible to devise medicines, not from the plant extracts provided directly by nature, but by chemical synthesis from scratch, leading to a significant increase in the number of effective medicines available to treat an ever larger number of illnesses. From the middle of the twentieth century, the methods of physics were successfully applied to the living world, leading to the new field of molecular biology. One of the first conquests of this new field was the determination of the three-dimensional structure of nucleic acids and other macromolecules using the methods of X-ray diffraction, at which point crystallographic studies were effectively extended to biology.

Today one can speak of a new convergence, this time between the various branches of the nanosciences and the life sciences. One can already begin to imagine how this encounter will revolutionise our approach to the life sciences, just as microtechnology and nanotechnology have completely recast the fields of data processing and communications.

Everyday experience shows us that the capabilities of our personal computers double roughly every 18 months, while their cost remains approximately constant or even decreases. This has been made possible by the extreme miniaturisation of electronic devices and the reduction in size of the transistor. In

1950, a transistor would have measured several centimeters across (10^{-2} m), while today, it occupies a space of just a few tens of nanometers (10^{-8} m), a reduction by a factor of about a million in the linear dimensions. A genuine revolution has been occurring under our very eyes over the past 25 years. We can no longer fully appreciate the essential role played by electronic components in our everyday environment, from the portable telephone to the car.

This technology requires engineers to be able to observe and manipulate matter with a resolution varying between the micron (10^{-6} m or one millionth of a meter) and the nanometer (10^{-9} m or one billionth of a meter). Over the last few years, nanoscientists have also developed investigative methods appropriate to these new structures. The atomic force microscope (AFM) serves as an example, used to displace atoms one by one and place them according to specific arrangements.

For the purposes of comparison, the components of living organisms are also of micron or submicron dimensions. Our blood vessels and bronchial tubes are capillaries of diameter a few microns transporting fluids or gases. The cells, functional units of living beings, are tiny globules measuring a few microns across. Within the cells, chemical reactions occur within compartments of a few attoliters (10^{-18} L). Compounds such as medicines, chemical mediators, metabolites, and so on, enter the cells via pores measuring a few nanometers in aperture. Enemies of the cell such as bacteria or viruses are also micrometric or even nanometric entities. For example, the envelope of the influenza virus is built up from several protein macromolecules and has a diameter of a few hundred nanometers.

Operating on this kind of length scale, and often subject to the same laws, it was only natural that nanotechnology would eventually come into contact with the biological sciences to form the joint venture we now call nanobiotechnology.

There is no point trying to give an exhaustive definition of nanobiotechnology here. Indeed, there is still no definition that would obtain a general consensus. Some authors see this as an inevitable state of affairs in a newly born discipline, or one that is just coming into being. One may note in passing that some definitions are perhaps unnecessarily restrictive. For example, the National Nanotechnology Initiative, created under the auspices of the United States government, defines nanotechnology as anything involving structures with dimensions less than 100 nm. The problem with this kind of definition is that it runs the risk of leaving out devices that currently manipulate objects or fluids rather on the micrometric scale, not to mention truly macroscopic devices, but which merely make use of nanometric objects or structures.

In order to glimpse the way nanobiotechnology may develop in the future, it is tempting once again to draw a parallel with the rise of nanotechnology in the field of data processing and communications. Indeed, one can already discern two lines of attack, still quite distinct, which approach the question from completely opposite directions. The so-called top-down approach consists in miniaturising the investigative or analytic tools we possess in order to

move from subjects of centimetric or millimetric dimensions to ones with the same function but much smaller. In a sense it is as though one were climbing down the length scale. The opposite approach aims to climb up this same scale by arranging nanoscopic elements like atoms or clusters of atoms into novel structures, assembled in some appropriate way. It is clear that today most nanobiotechnological activity subscribes to the first approach, for while the second, bottom-up approach remains extremely attractive, it still faces many difficulties. One of these is that we are unable to predict the properties of elements conceived in this way, working solely from the knowledge of the individual properties of their components. In association with these two approaches, new imaging and measurement techniques are being developed to observe phenomena that have now become accessible to us.

There can be no doubt that within a few years these new tools will completely revolutionise our understanding of the most intimate mechanisms occurring within the cell, on a molecular scale and in real time, opening the way to novel and extraordinary therapeutic prospects.

The present book provides a cross-section of current understanding in several areas of nanobioscience. It is divided into three main parts:

- *Biological Nano-Objects*. This part describes the basic building blocks, i.e., DNA, typical biological structures, etc., used as example and support in nanobiotechnology.
- *Methods of Nanobiotechnology*. Chapters here present the physical, chemical and electrical tools and methods used to investigate biological nano-objects.
- *Applications of Nanobiotechnology*. The last part deals with the most common of current applications, pointing the way from nanobiotechnology to nanomedicine.

Neurobiology Department,
ESPCI, Paris

Jean Rossier
Vincent Studer

Acknowledgements

We would like to thank all members of the French nanoscience community (CNRS, CEA, universities, Grandes Ecoles, industry) who gave a very favourable welcome to the writing of these pedagogical introductions to nanotechnology and nanophysics, nanomaterials and nanochemistry, and nanobio-technology and nanobiology, and without whom they would have been impossible. Special thanks go, of course, to all those who contributed to these books.

We are particularly grateful to the late Hubert Curien of the Academy of Sciences (Paris), Jean-Marie Lehn (Nobel Prize for Chemistry), and Axel Kahn (Director of the Institut Cochin and IFR Alfred Jost) for contributing the forewords to volumes I, II and III of this series, respectively, and also to Patrice Hesto who gave invaluable advice when the project first began.

We warmly acknowledge the material and financial support of the French Ministry of Research and the French atomic energy authority (CEA), orchestrated by Jean-Louis Robert of the Department of Physics, Chemistry, and Engineering Sciences, and Jean Therme, Director of the CEA, Grenoble.

Likewise, our warmest thanks go to Claude Puech, President of the Club NanoMicroTechnologie, everyone at the LMN (Laboratoire d'étude des Milieux Nanométriques at the University of Evry, France) and the GIFO (Groupement des Industries Françaises de l'Optique) for their administrative and logistical support.

Finally, we would like to thank Alain Brisson and Pierre Schaaf for their continued scientific support, especially during copy-editing sessions, and Paul Siffert of the European Materials Research Society for supporting the English edition of the book.

Marcel Lahmani, Patrick Boisseau and Philippe Houdy

Contents

Part I Biological Nano-Objects

1 Structural and Functional Regulation of DNA: Geometry, Topology and Methylation

C. Auclair

1.1	Geometry of the DNA Double Helix	4
1.2	The Z Conformation of DNA	7
1.3	Supercoiled DNA	12
1.4	Methylation of DNA	17
1.4.1	Methylation of Cytosine	19
1.4.2	CpG Sequences	21
1.4.3	Structure of Methylated CpG Dinucleotides	22
1.4.4	Specific Recognition of Symmetric Methylation by Proteins	23
1.5	Conclusion	25
	References	25

2 Protein–Lipid Assembly and Biomimetic Nanostructures

A. Girard-Egrot, L. Blum, R. Richter, A. Brisson

2.1	Introduction: Biological Membranes	29
2.2	Lipid Membranes: Structure and Properties	31
2.2.1	The Main Classes of Lipid Membranes	31
2.2.2	Self-Assembly	36
2.2.3	Lipid Polymorphism	39
2.2.4	Lipid Shapes	44
2.3	Models and Methods for Characterising Membranes	46
2.3.1	Liposomes	46
2.3.2	Langmuir Monolayers	50
2.3.3	Supported Membranes	57
2.3.4	Suspended Membranes	72
2.3.5	Bilayer Lipid Membranes (BLM)	75

2.4	Protein–Lipid Assembly	77
2.4.1	Functionalising Langmuir–Blodgett Films	78
2.4.2	Two-Dimensional Organisation of Proteins on Lipid Surfaces	85
2.4.3	Reconstitution of Membrane Proteins in Supported Lipid Bilayers	91
2.5	Applications of Biomimetic Membranes in Nanobiotechnology	92
2.5.1	Bio-Optoelectronic Micro- and Nanosensors	92
2.5.2	Composite Assemblies	95
	References	95

3 Supramolecular Complexes of DNA

G. Zuber, D. Scherman

3.1	Introduction	101
3.2	Different Stages of Gene Transfer	105
3.2.1	Presentation	105
3.2.2	Condensation and Protection of DNA	105
3.2.3	Circulation in a Multicellular Organism	106
3.2.4	Cell Adhesion and Crossing of the Plasma Membrane	107
3.2.5	Intracellular Circulation and Entry into the Nucleus	109
3.2.6	State of the Art	110
3.3	Polymolecular DNA Assemblies: Synthesis, Characterisation and Properties	110
3.3.1	Polyplexes	110
3.3.2	Lipoplexes	115
3.3.3	Modification of Polyplexes and Lipoplexes for in Vivo Gene Transfer	120
3.4	Monomolecular DNA Assemblies (Nanoplexes): Synthesis, Characterisation, and Properties	122
3.4.1	Monomolecular Condensation of DNA	122
3.4.2	Chemical Synthesis	124
3.4.3	Synthesis and Characterisation of Nanoplexes	124
3.4.4	Nanoplex Modification for in Vivo Gene Transfer	126
3.5	Conclusion and Prospects	127
	References	127

4 Functionalised Inorganic Nanoparticles for Biomedical Applications

E. Duquet, M. Treguer-Delapierre, M.-H. Delville

4.1	Synthesis and Chemical Surface Modification of Inorganic Nanoparticles	130
4.1.1	The Main Strategies	130
4.1.2	Iron Oxide Nanoparticles	133

4.1.3	Semiconductor CdSe Colloids	141
4.1.4	Noble Metal Nanoparticles: Gold and Silver	145
4.2	Biological Tagging in Vitro and in Animals	146
4.2.1	Biological Tagging by Semiconductor Colloids	147
4.2.2	Biological Tagging by Metal Colloids	151
4.3	In Vivo Applications	154
4.3.1	Fate of Particles in the Blood Compartment	154
4.3.2	Tools for Medical Diagnosis: MRI Contrast Agents	159
4.3.3	Therapeutic Tools	164
4.4	Conclusion	167
	References	168

5 Living Nanomachines

M.-F. Carrier, E. Helfer, R. Wade, F. Haraux

5.1	Introduction	171
5.2	Force and Motion by Directed Assembly of Actin Filaments	174
5.2.1	General Considerations	174
5.2.2	Assembly Dynamics of Actin in Vitro. Intrinsic Properties	177
5.2.3	Regulation of Actin Filament Assembly in Cell Motility	179
5.2.4	Biomimetic Motility Assay	182
5.2.5	Measuring the Force Produced by Directional Actin Polymerisation	183
5.2.6	Theoretical Models for Force Production by Actin Polymerisation	187
5.2.7	Prospects	192
5.3	Molecular Motors: Myosins and Kinesins	193
5.3.1	Introduction	193
5.3.2	Actin Filaments and Microtubules	194
5.3.3	Motor Proteins	196
5.3.4	Motion and Forces	198
5.3.5	Motion and Structural Conformation	202
5.4	ATP Synthase: The Smallest Known Rotary Molecular Motor	206
5.4.1	Basics of ATP Synthase	206
5.4.2	How ATP Synthase Was Recognised as a Molecular Motor: A Story of Two Conceptual Leaps	208
5.4.3	Rotation Mechanism: Current Understanding	212
5.4.4	Thermodynamics, Kinetics, and Nanomechanics	215
5.4.5	Conclusion	219
	References	220

6 Aptamer Selection by Darwinian Evolution*F. Chauveau, C. Pestourie, F. Ducongé, B. Tavitian*

6.1	Some Theoretical Aspects of Molecular Evolution	224
6.1.1	Darwin and the Theory of Evolution	224
6.1.2	Molecular Evolution and Properties of Nucleic Acids	225
6.2	Structural Features of Nucleic Acids	227
6.2.1	General Considerations: The Double Helix	227
6.2.2	Intrahelical Interaction Sites	228
6.2.3	From Secondary to Tertiary Structure: Supercoiling	229
6.2.4	Role of Cations and Water Molecules	230
6.2.5	Binding of an Aptamer to Its Target: Examples of Resolved Structures	230
6.3	SELEX	231
6.3.1	History	231
6.3.2	General Selection Principle	232
6.3.3	Chemical Modifications	235
6.4	Applications	236
6.4.1	Aptamers as Research Tools	236
6.4.2	Aptamers as Purification Tools	237
6.4.3	Aptamers as Detection Tools	238
6.4.4	Aptamers as Regulatory Tools	240
6.4.5	Aptamers as Therapeutic Tools	242
6.5	Conclusion	244
	References	246

Part II Methods of Nanobiotechnology

7 Optical Tools

*E. Roncali, B. Tavitian, I.e Texier, P. Peltié, F. Perraut,
J. Boutet, L. Cognet, B. Lounis, D. Marguet, O. Thoumine,
M. Tramier*

7.1	Introduction to Fluorescence Microscopy	253
7.1.1	Conventional Fluorescence Microscopy	253
7.1.2	Examples of Biological Applications	256
7.1.3	Confocal Microscopy	259
7.1.4	Two-Photon and Multiphoton Microscopy	261
7.1.5	Conclusions and Prospects	262
7.2	Labels	262
7.2.1	Introduction	262
7.2.2	Exogenous Probes	263
7.2.3	Endogenous Probes: Reporter Genes	281
7.2.4	Conclusion	291
7.3	In Vivo Detection Systems	292
7.3.1	Introduction to in Vivo Optical Imaging	292
7.3.2	Basic Principles of in Vivo Optical Imaging	293

7.3.3	Experimental Setups for Fluorescence and Bioluminescence Imaging (Continuous Irradiation)	298
7.3.4	Applications of Fluorescence and Bioluminescence Imaging	300
7.3.5	Time-Resolved Fluorescence Imaging	303
7.4	In Vitro Detection Systems	307
7.4.1	Introduction to Biochips and Microarrays	307
7.4.2	Conventional Read Instruments	313
7.4.3	Detection by Surface Plasmon Resonance (SPR)	320
7.4.4	Fluorescence Enhancement	324
7.4.5	Current Trends in Biological Instrumentation	326
7.5	Other Detection Systems.	
	Dynamics of Molecular Interactions	333
7.5.1	Fluorescence Recovery after Photobleaching (FRAP) and Associated Techniques	334
7.5.2	Fluorescence Correlation Spectroscopy (FCS)	341
7.5.3	Tracking Single Molecules and Particles	351
7.5.4	Fluorescence Resonance Energy Transfer (FRET)	355
	References	361

8 Nanoforce and Imaging

C. Le Grimellec, P.-E. Milhiet, E. Perez, F. Pincet, J.-P. Aimé, V. Emiliani, O. Thoumine, T. Lionnet, V. Croquette, J.-F. Allemand, and D. Bensimon

8.1	Molecular and Cellular Imaging Using AFM	375
8.1.1	Introduction	375
8.1.2	Atomic Force Microscopy	375
8.1.3	Imaging Soluble Molecules	378
8.1.4	Membrane Imaging	380
8.1.5	AFM and Cells: Cell Imaging, Mechanical Properties, and Adhesion	384
8.1.6	Current Limits and Future Developments	387
8.1.7	Developments in Nanobiotechnology and Medecine	389
8.2	Surface Force Apparatus and Micromanipulation	390
8.2.1	Surface Force Apparatus (SFA)	390
8.2.2	Micromanipulation	397
8.3	Atomic Force Microscopy in Contact and Tapping Modes	402
8.3.1	Introduction	402
8.3.2	Force Measurements in Contact (Static) Mode	405
8.3.3	AFM Oscillating Modes: Introduction and Definitions	413
8.3.4	Oscillations in a Liquid Medium	422
8.3.5	Force Measurements and Height Images. DNA Measurements	427

8.4	Optical Tweezers	434
8.4.1	Basic Principles and Main Parameters	434
8.4.2	Estimating the Stiffness Constant of the Trap	436
8.4.3	Different Types of Optical Tweezers	437
8.4.4	Experimental Setup.....	441
8.4.5	Biological Applications of Optical Tweezers	442
8.5	Magnetic Tweezers	447
8.5.1	General Idea	447
8.5.2	A Mechanical Model for a Force Sensor: A Bead Attached to a Spring	449
8.5.3	Measuring the Bead Position with Nanometric Resolution	451
8.5.4	Calibrating the Force Measurement by Brownian Motion	454
8.5.5	Magnets Used for Magnetic Tweezers	456
8.5.6	Advantages of Magnetic Tweezers	458
8.5.7	Examples of Studies Using Magnetic Tweezers	461
8.5.8	Manipulating an Object with Magnetic Tweezers	466
	References	467

9 Surface Methods

D. Altschuh, S. Ricard-Blum, V. Ball, M. Gaillet,

P. Schaaf, B. Senger, B. Desbat, P. Lavalley, J.-F. Legrand

9.1	Biosensors Based on Surface Plasmon Resonance: Interpreting the Data.....	477
9.1.1	Introduction	477
9.1.2	Evaluating the SPR Data	479
9.1.3	Measurements Under Mass Transport or Kinetic Conditions	484
9.1.4	Other Experimental Adaptations	489
9.1.5	Applications	495
9.2	Ellipsometry	500
9.2.1	Introduction	500
9.2.2	Theory of Light and Polarisation	501
9.2.3	Basic Principles and Possibilities of Ellipsometry	508
9.2.4	Instrumentation	510
9.2.5	Ellipsometric Data and Its Use	515
9.2.6	Applications	520
9.2.7	Conclusion	523
9.3	Optical Spectroscopy Using Waveguides.....	525
9.3.1	General Features of Optical Biosensors	525
9.3.2	Optical Spectroscopy of Normal Modes Coupled in a Waveguide.....	527
9.3.3	Applications of Optical Waveguide Lightmode Spectroscopy.....	534
9.3.4	Conclusions.....	537

9.4	Vibrational Spectroscopy	540
9.4.1	General Features	540
9.4.2	Infrared Spectroscopy	541
9.4.3	Raman Spectroscopy	551
9.4.4	Prospects for Vibrational Spectroscopy in the Study of Nano-Objects	555
9.5	Brewster Angle Microscopy	556
9.6	Quartz Crystal Microbalance with Dissipation Monitoring (QCM-D)	561
9.6.1	Introduction	561
9.6.2	Vibration of a Damped Harmonic Oscillator Subject to Forces	563
9.6.3	Crystal in Vacuum	563
9.6.4	Crystal in Contact with a Viscous Medium	564
9.6.5	Crystal Covered with a Stratified Viscoelastic Medium in Contact with a Viscous Medium	566
9.6.6	Numerical Simulation of the QCM Response	572
9.6.7	Analysis of a Specific Experiment: Construction of a Polyelectrolyte Multilayer Film	575
9.7	Grazing Incidence Neutron and X-Ray Reflectometry	578
9.7.1	Reflection of X-Rays by a Plane Interface. Critical Angle and Fresnel Law	578
9.7.2	Interference Produced by a Homogeneous Film of Nanometric Thickness	580
9.7.3	Determining the Density Profile of a Stratified Layer. Resolution	582
9.7.4	Neutron Reflectometry: Contrast Variation	584
	References	586

10 Mass Spectrometry

D. Pflieger, E. Forest, J. Vinh

10.1	Principles and Definitions	595
10.1.1	What Is Mass Spectrometry?	596
10.1.2	The Mass Spectrometer	596
10.1.3	Terminology	596
10.2	Ionisation Sources for Biomolecules	598
10.2.1	Applications in Biology and Biochemistry	598
10.2.2	Electrospray Ionisation (ESI)	601
10.2.3	MALDI	606
10.2.4	NanoSIMS and Ion Microscopy	609
10.3	Analysers	611
10.3.1	General Considerations	611
10.3.2	Time-of-Flight Analyser	612

10.3.3	Quadrupole Analyser	616
10.3.4	Ion Trap	619
10.3.5	Fourier Transfer Ion Cyclotron Resonance (FT-ICR) Analyser	620
10.4	Combined Liquid Phase Separation and Mass Spectrometry	622
10.4.1	Chromatographic Techniques	622
10.4.2	Electrophoretic Techniques	624
10.5	Which Mass Spectrometer Should Be Coupled with Separation Techniques: ESI or MALDI?	625
10.5.1	Combinations with HPLC	625
10.5.2	Coupling with Electrophoretic Techniques	627
10.6	Nanotechnology for the MS Interface	628
10.6.1	Microfluidic Chip Associating Chromatography and Nanospray Tip	629
10.6.2	Nanospray Tip Array Chip	629
	References	630

11 Electrical Characterisation and Dynamics of Transport

*N. Picollet-D'Hahan, C. Amatore, S. Arbault, L. Thouin,
A.-L. Biance, G. Oukhaled, L. Auvray, J. Weber, N. Minc,
J.-L. Viovy*

11.1	Ion Channels and the Patch-Clamp Technique	639
11.1.1	What Is an Ion Channel?	639
11.1.2	Physiological Role of Ion Channels	644
11.1.3	Pharmacological Dysfunction	645
11.1.4	Direct Ways of Studying Ion Channels	649
11.1.5	Conclusion: Prospects for the Patch-Clamp Technique and the High-Throughput Revolution in Electrophysiology	666
11.2	Amperometry	667
11.2.1	Basics of Faradaic Electrochemistry	668
11.2.2	Concentration Profiles	678
11.2.3	Conclusion Regarding Faradaic Electrochemical Detection	682
11.2.4	Artificial Synapses: Biological Applications to Single Cells	684
11.3	Macromolecular Transport Through Natural and Artificial Nanopores. Electrical Detection	695
11.3.1	Introduction	695
11.3.2	Electrical Detection of Particle Transport in a Pore	698
11.3.3	Polymers Confined in Pores. Statics and Dynamics	703
11.3.4	Some Natural and Artificial Systems	713
11.3.5	Conclusion and Prospects	718

11.4	Electrophoretic Techniques	719
11.4.1	Introduction	719
11.4.2	Migration of a Charged Species in Solution	720
11.4.3	Use of Polymer Matrices	721
11.4.4	Microfluidic Systems for Separation of Long DNA Fragments	726
11.4.5	Conclusion	731
References		731

12 Microfluidics: Concepts and Applications to the Life Sciences

A. Buguin, Y. Chen, P. Silberzan

12.1	Introduction	743
12.2	Physics of Microfluidic Flows	745
12.2.1	Fluid Mechanics on Microscopic Scales	745
12.2.2	Setting the Fluid in Motion	748
12.3	Fabrication, Materials, Functions	753
12.3.1	Lithography	754
12.3.2	Different Technologies	755
12.3.3	Silicone Elastomers	758
12.3.4	Elementary Components: Pumping, Mixing, and Separating in Microvolumes	760
12.4	Applications	761
12.4.1	Crystallisation of Proteins	761
12.4.2	Separation of DNA Molecules	764
12.4.3	Cell Sorting	768
12.5	Conclusion	771
References		771

13 Data Processing

P. Grangeat

13.1	Nanobiotechnology and Data Systems	775
13.1.1	Nanobiotechnology	775
13.1.2	Data Systems	776
13.1.3	Three Examples	778
13.1.4	Technological Bottlenecks	781
13.1.5	Automated Measurements	783
13.1.6	Layout of this Chapter	783
13.2	Representing Data	784
13.2.1	Data Structures	784
13.2.2	Sampling and Quantification	785
13.2.3	Measurement Noise	785
13.2.4	Direct or Indirect Measurement	786
13.3	Correcting for Sensor Defects and Improving the Data	786
13.3.1	Linearity and Calibration	787
13.3.2	Independence and Normalisation	788

13.3.3	Noise and Filtering	789
13.3.4	Outliers	789
13.3.5	Distortion and Geometric Corrections	790
13.4	Data Extraction	790
13.4.1	Extracting Physical Quantities	790
13.4.2	The Systems Approach	791
13.4.3	Inverse Problems	793
13.4.4	Regularised Solutions	794
13.5	Data Analysis	795
13.5.1	Selecting the Relevant Measurements	795
13.5.2	Statistical Analysis	796
13.5.3	Geometrical Analysis	796
13.5.4	Classification Methods	796
	References	798

14 Molecular Dynamics. Observing Matter in Motion

C. Chipot

14.1	Introduction.....	803
14.1.1	Relating the Microscopic to the Meso- and Macroscopic	803
14.1.2	Legitimacy of Molecular Dynamics Simulations.....	805
14.2	Basic Principles of Molecular Dynamics	806
14.2.1	Validity of Molecular Dynamics Simulations	806
14.2.2	Multistep Integration of the Equations of Motion	807
14.3	Potential Energy Function	808
14.3.1	Meaning of Different Terms in the Force Field.....	809
14.3.2	Parametrisation of Unbound Atom Terms	810
14.3.3	Beyond the Usual Force Fields	812
14.4	Integrating the Equations of Motion	814
14.4.1	Molecular Dynamics Integrators	814
14.4.2	Integration with Constraints	817
14.4.3	Molecular Dynamics at Constant Temperature	818
14.4.4	Molecular Dynamics at Constant Pressure	821
14.5	Rigorous Treatment of Electrostatic Interactions	823
14.6	Some Properties Accessible to Simulation	826
14.6.1	Structural Properties from Simulations	826
14.6.2	Dynamical Properties from Simulations	827
14.6.3	Molecular Dynamics and Free Energy	829
14.7	Molecular Dynamics and Parallelisation.....	831
14.8	Conclusion	834
	References	835

Part III Applications of Nanobiotechnology

15 Real-Time PCR*A. Evrard, N. Boulle and G.s Lutfalla*

15.1	Real-Time PCR	841
15.1.1	Polymerase Chain Reaction	841
15.1.2	Equipment Used for Quantitative Real-Time PCR	846
15.1.3	Fluorescence Formats	847
15.2	Implementing Quantitative Real-Time PCR	853
15.2.1	Denaturation and Amplification Curves	854
15.2.2	Optimising the Annealing Temperature: Specificity	857
15.2.3	Determining the Amplification Efficiency	857
15.2.4	Relative Quantification	860
15.2.5	Multiplex PCR	861
15.3	Applications of Real-Time PCR	862
15.3.1	Real-Time PCR for the Quantification of Viral Genomes	862
15.3.2	Real-Time PCR in Pharmacogenetics	865
	References	869

16 Biosensors. From the Glucose Electrode to the Biochip*L. Blum and C. Marquette*

16.1	Bioreceptors	872
16.1.1	Natural Bioreceptors	873
16.1.2	Artificial Bioreceptors	874
16.1.3	Using Ligand–Receptor Systems	876
16.2	Immobilisation Methods	877
16.2.1	Adsorption	877
16.2.2	Inclusion	878
16.2.3	Confinement	878
16.2.4	Crosslinking	878
16.2.5	Covalent Bonding on an Activated Substrate	879
16.3	Biosensors with Electrochemical Detection	880
16.3.1	Enzyme Electrodes	880
16.3.2	ENFET or Enzyme ISFET	885
16.4	Mass Transducer Biosensors	887
16.5	Enzyme Thermistors	889
16.6	Fibre Optic Biosensors	891
16.6.1	Fibre Optic Chemical Sensors	892
16.6.2	Setups for Fibre Optic Biosensors	893
16.6.3	Enzyme Fibre Optic Biosensors	894

16.6.4	Affinity Biosensors.....	895
16.6.5	Biosensors Based on Chemiluminescent or Bioluminescent Detection.....	897
16.7	Biochips	900
16.7.1	DNA Microarrays	903
16.7.2	Protein and Other Microarrays	905
16.8	Conclusion	906
	References	908

17 DNA Microarrays

C. Nguyen and X. Gidrol

17.1	Introduction.....	911
17.2	Analysing the Transcriptome	912
17.2.1	Basic Idea	912
17.2.2	Different Types of DNA Microarray for Transcriptome Analysis	913
17.2.3	Some Applications of DNA Microarrays	919
17.2.4	Some Remarks Concerning Transcriptome Data Analysis.....	919
17.2.5	Transcriptome Applications.....	920
17.3	Beyond the Transcriptome	923
17.3.1	CGH Microarrays	923
17.3.2	ChIP on Chip.....	925
17.3.3	When DNA Microarrays Become Cell Microarrays	928
17.3.4	Prospects and Conclusion	930
	References	930

18 Protein Microarrays

S. Ricard-Blum

18.1	Overview of Proteins	937
18.2	Fabricating a Protein Array on a Flat Support.....	939
18.2.1	Preparation of Purified Proteins	939
18.2.2	Substrates for Protein Microarrays.....	942
18.2.3	Immobilising Proteins on the Array	943
18.2.4	Spotting Proteins.....	944
18.2.5	Detection Systems	947
18.3	Other Formats for Protein Microarrays	953
18.4	Applications of Protein Microarrays	953
18.4.1	Analytical Microarrays	953
18.4.2	Functional Microarrays	956
18.5	Conclusion	958
	References	959

19 Cell Biochips*B. Le Pioufle, N. Picollet-D'Hahan*

19.1	Biochips for Analysing and Processing Living Cells	965
19.1.1	From Single Cells to Reconstituted Tissue	965
19.1.2	Cell Micromanipulation Methods	967
19.1.3	Methods for Characterising Microcultured Cells on Chip	969
19.2	Patch-Clamp Microarrays	972
19.2.1	Motivations	972
19.2.2	Emergence of New Patch-Clamp Platforms	974
19.2.3	A Cultural Revolution? Prospects	987
	References	992

20 Lab on a Chip*P. Puget*

20.1	The General Idea	999
20.2	Implanted Functions	1001
20.2.1	Sample Preparation	1002
20.2.2	Transduction	1007
20.3	Technological Aspects	1009
20.4	Conclusion	1011
	References	1013

21 Polyelectrolyte Multilayers*P. Schaaf, J.-C. Voegel*

21.1	The Idea	1017
21.1.1	Construction and Properties	1017
21.1.2	Physical Origin of Interactions Between Polyanions and Polycations	1019
21.2	Linear Growth and Exponential Growth of Polyelectrolyte Films	1021
21.2.1	Linear Growth	1021
21.2.2	Exponential Growth	1023
21.2.3	Fabrication of Polyelectrolyte Multilayers	1026
21.3	Biological Functionalisation	1027
21.3.1	Biologically Inert Films	1027
21.3.2	Functionalisation by Protein Insertion	1029
21.3.3	Functionalisation by Peptides	1033
21.3.4	Functionalisation by Drugs	1035
21.3.5	Development of Nanoreactors	1035
21.4	Making Hollow Particles from Multilayers	1037
21.5	The Route to More Complex Architectures	1038
21.6	Prospects	1040
	References	1041

22 Biointegrating Materials

*J. Amédée, L. Bordenave, M.-C. Durrieu, J.-C. Fricain,
L. Pothuaud*

22.1	Cell and Tissue Engineering	1043
22.2	Modifying Material Surfaces	1045
22.2.1	Using Nanoparticles to Deliver Active Ingredients	1045
22.2.2	Macroscale Functionalisation of Biomaterial Surfaces	1048
22.2.3	The Relevance of Controlled Nanotopochemistry and Nanodomains	1052
22.3	Applications of Biointegrated Biomaterials	1054
22.3.1	Applications to Bone Tissue	1054
22.3.2	Applications to the Vascular System	1056
22.4	In Vivo Assessment of Tissue Engineering Products	1058
22.4.1	Animal Models	1058
22.4.2	Which Animal Model for Which Application?	1059
22.4.3	Standard Methods for in Vivo Evaluation of Tissue Engineering Products	1060
22.5	Investigative Methods Associated with Tissue Engineering	1061
References		1063

23 Viral Vectors for in Vivo Gene Transfer

E. Thévenot, N. Dufour, N. Déglon

23.1	Introduction	1069
23.1.1	In Vivo Gene Transfer	1069
23.1.2	Viral Vectors	1071
23.2	Main Types of Viral Vector	1071
23.2.1	Retroviral and Lentiviral Vectors	1072
23.2.2	Adenoviral Vectors	1077
23.2.3	Adeno-Associated Vectors	1078
23.3	Biomedical Applications of the Viral Platform	1079
23.3.1	Gene Therapy	1080
23.3.2	Animal Models of Human Pathologies	1085
23.4	Controlling and Visualising Transgene Expression	1089
23.4.1	Controlling Transgene Expression	1089
23.4.2	Imaging Transgene Expression	1092
23.5	Prospects	1092
References		1093

24 Pharmaceutical Applications of Nanoparticle Carriers

B. Heurtault, F. Schuber, B. Frisch

24.1	Introduction to Drug Delivery in Pharmaceutics	1097
24.2	Nanoparticle Carriers	1098
24.2.1	The Main Nanoparticle Carriers	1098
24.2.2	Carrier Characteristics	1101

24.3	Development of Carriers for Pharmaceutical Applications	1105
24.3.1	Thermosensitive and pH-Sensitive (Fusogenic) Liposomes	1106
24.3.2	Modifying the Carrier Surface	1106
24.4	Applications of Carriers	1111
24.4.1	Medical Mycology and Parasitology	1111
24.4.2	Ophthalmology	1112
24.4.3	Infectious Diseases	1113
24.4.4	Cancerology	1113
24.5	Conclusion	1114
	References	1115

25 Activatable Nanoparticles for Cancer Treatment.

Nanobiotix

V. Simon, A. Ceccaldi, L. Lévy

25.1	Introduction	1121
25.2	NanoTherapeutics	1123
25.3	Different Families of Nanoparticles	1125
25.4	NanoTherapeutic Action Mechanisms	1126
25.4.1	NanoMag	1126
25.4.2	NanoPDT	1126
25.4.3	NanoXRay	1127
25.4.4	Nano(U)Sonic	1128
25.5	Synthesising NanoMag Particles	1128
25.5.1	Coating the Fe ₂ O ₃ Particles with SiO ₂	1129
25.5.2	Adding the Spacer	1130
25.5.3	Adding the Ligand	1130
25.6	Advantages of the NanoTherapeutic Families	1130
25.6.1	NanoMag	1130
25.6.2	NanoPDT	1131
25.6.3	NanoXRay	1133
25.6.4	Nano(U)Sonic	1133
25.7	Results (NanoMag)	1133
25.7.1	In Vitro Experiments	1133
25.7.2	In Vivo Experiments	1138
25.8	Conclusion	1141
	References	1141

26 The Medical, Social, and Economic Stakes of Nanobiotechnology

J. Hache, F. Berger

26.1	From Current to Future Applications	1143
26.1.1	Diagnosis and Therapy	1143
26.1.2	Cosmetics	1149

XXVIII Contents

26.1.3	Product Quality and Traceability	1150
26.1.4	Environment and Risk Prevention	1151
26.2	From Individual Players to Clusters	1152
26.2.1	Different Players Around the World and the Position of France	1152
26.2.2	Clusters and Other Poles of Competitiveness	1152
26.3	From Funding to Industrialisation	1152
26.3.1	Patents	1152
26.3.2	Funding Nanobiotechnological Activity	1153
26.3.3	The Markets: Between Fantasy and Reality	1154
26.4	From Risks to Precautions	1155
26.4.1	New Risks and Ethical Considerations	1155
26.4.2	Science Fiction or Future Reality?	1156
26.4.3	Image and Communication	1157
26.4.4	Convergence of Nanoscience and the Life Sciences	1157
26.5	The Advent of Nanomedicine	1158
	References	1161
	Index	1163

List of Contributors

Jean-Pierre Aimé

Centre de Physique Moléculaire
Optique et Hertzienne
Université Bordeaux 1 and CNRS
351 cours de la libération
33405 Talence Cedex France
`j.p.aime@cpmoh.u-bordeaux1.fr`

Jean-François Allemand

Laboratoire de Physique Statistique
de l'École Normale Supérieure
24 rue Lhomond, 75005 Paris
`allemand@lps.ens.fr`

Danièle Altschuh

UMR 7100
Ecole supérieure de biotechnologie
de Strasbourg
Boulevard Sébastien Brant
BP 10413
67412 Illkirch Cedex, France
`daniele.altschuh@esbs.
u-strasbg.fr`

Christian Amatore

Ecole Normale Supérieure
Département de Chimie
UMR CNRS 8640
24 rue Lhomond
75231 Paris cedex 05
`christian.amatore@ens.fr`

Joëlle Amédée

Biomatériaux et réparation
tissulaire
Université Victor Segalen
Bordeaux 2
Zone Nord, Bâtiment 4A
Case postale 45
146 rue Léo-Saignat
33076 Bordeaux Cedex, France
`joelle.amedee@bordeaux.
inserm.fr`

Stéphane Arbault

Ecole Normale Supérieure
Département de Chimie
UMR CNRS 8640
24 rue Lhomond
75231 Paris cedex 05
`stephane.arbault@ens.fr`

Christian Auclair

Laboratoire de biotechnologies
et de pharmacologie génétique
appliquée
Ecole Normale Supérieure de Cachan
61 avenue du Président Wilson
94235 Cachan, France
`auclair@lbpa.ens-cachan.fr`

Loïc Auvray

Laboratoire de Matériaux
Polymères aux Interfaces
Université d'Evry-Val d'Essonne
Bâtiment Maupertuis
Bd F. Mitterrand
91025 Evry Cedex
loic.auvray@univ-evry.fr

Vincent Ball

UFR d'Odontologie
Université Louis Pasteur
4 rue Kirschleger
67000 Strasbourg
vincent.ball@medecine.
u-strasbg.fr

David Bensimon

Laboratoire de Physique Statistique
de l'École Normale Supérieure
24 rue Lhomond
75005 Paris
david.bensimon@lps.ens.fr

François Berger

INSERM U318
Université Joseph Fourier
BP 53
38041 Grenoble Cedex 9
francois.berger@ujf-
grenoble.fr

Anne-Laure Biance

Laboratoire de Photonique
et de Nanostructures
Université de Marne-la-Vallée
Route de Nozay
91460 Marcoussis, France
anne-laure.biance@univ-mlv.fr

Loïc Blum

Institut de Chimie et Biochimie
Moléculaires et Supramoléculaires
UMR 5246 Université Lyon 1
CNRS-INSA Lyon-CPE Lyon

Laboratoire de Génie
Enzymatique et Biomoléculaire
Bât. CPE
Université Claude Bernard Lyon 1
43 bd. du 11 novembre 1918
69622 Villeurbanne cedex
loic.blum@univ-lyon1.fr

Patrick Boisseau

CEA-LETI
17 rue des Martyrs
38054 Grenoble Cedex
patrick.boisseau@cea.fr

Laurence Bordenave

INSERM 577
Laboratoire de Biophysique
Université Victor Segalen
Bordeaux 2
Bâtiment 4A zone Nord
146 rue Léo Saignat
33076 Bordeaux cedex
laurence.bordenave@biophys.
u-bordeaux2.fr

Nathalie Boulle

Laboratoire de biologie cellulaire
et hormonale
Hôpital Arnaud de Villeneuve
34295 Montpellier
n-boulle@chu-montpellier.fr

Jérôme Boutet

CEA-LETI
17 rue des Martyrs
38054 Grenoble Cedex
jerome.boutet@cea.fr

Alain Brisson

Laboratoire d'Imagerie Moléculaire
et NanoBioTechnologie
IECB/UMR 5258 CNRS
Université Bordeaux 1
2, rue Robert Escarpit
33607 Pessac cedex
a.brisson@iecb.u-bordeaux.fr

Axel Buguin

Laboratoire de Physicochimie
CNRS UMR 168
Institut Curie
26 rue d'Ulm
75005 Paris, France
axel.buguin@curie.fr

Marie-France Carlier

Laboratoire d'Enzymologie et
Biochimie Structurales
CNRS UPR 3082
Bâtiment 34
Avenue de la Terrasse
91198 Gif-sur-Yvette Cedex
carlier@lebs.cnrs-gif.fr

Alexandre Ceccaldi

Nanobiotix
60 rue de Wattignies
Bâtiment B, 3ème étage
75012 Paris
alexandre.ceccaldi@
nanobiotix.fr

Fabien Chauveau

Laboratoire d'imagerie moléculaire
expérimentale
4 place du Général Leclerc
91401 Orsay Cedex, France
fabien.chauveau@cea.fr

Yong Chen

Ecole Normale Supérieure
24 rue Lhomond
75231 Paris Cedex 05
yong.chen@ens.fr

Christophe Chipot

Université Henri Poincaré
Campus Victor Grignard BP 239
54506 Vandoeuvre-les-Nancy Cedex
France
christophe.chipot@edam.
uhp-nancy.fr

Laurent Cognet

Centre de Physique Moléculaire
Optique et Hertzienne
Université Bordeaux 1 and CNRS
351 cours de la libération
33405 Talence cedex France
l.cognet@cpmoh.u-bordeaux1.fr

Vincent Croquette

Laboratoire de Physique Statistique
de l'École Normale Supérieure
24 rue Lhomond
75005 Paris
vincent.croquette@lps.ens.fr

Nicole Déglon

MIRcen
CEA Fontenay-aux-Roses
18 route du Panorama
92265 Fontenay-aux-Roses Cedex
nicole.deglon@cea.fr

Marie-Hélène Delville

ICMCB-CNRS
Université Bordeaux 1
87, Ave. du Dr A. Schweitzer
33608 Pessac Cedex, France
delville@icmcb-bordeaux.
cnrs.fr

Bernard Desbat

Institut des Sciences Moléculaires
Université Bordeaux 1
CNRS UMR 5255
Bâtiment A12
351 cours de la Libération
33405 Talence Cedex
b.desbat@ism.u-bordeaux1.fr

Frédéric Ducongé

Laboratoire d'imagerie moléculaire
expérimentale
4 place du Général Leclerc
91401 Orsay Cedex, France
frederic.duconge@cea.fr

Noëlle Dufour

MIRCen
CEA Fontenay-aux-Roses
18 route du Panorama
92265 Fontenay-aux-Roses Cedex
noelle.dufour@cea.fr

Étienne Duguet

ICMCB-CNRS
Université Bordeaux 1
87, Ave. du Dr A. Schweitzer
33608 Pessac CEDEX, France
duguet@icmcb-bordeaux.cnrs.fr

Marie-Christine Durrieu

Biomatériaux et réparation
tissulaire
Université Victor Segalen
Bordeaux 2
Zone Nord, Bâtiment 4A
Case postale 45
146 rue Léo-Saignat
33076 Bordeaux Cedex, France
marie-christine.durrieu@
bordeaux.inserm.fr

Valentina Emiliani

Laboratoire de Neurophysiologie
et Nouvelles Microscopies
INSERM U603 CNRS UMR 8154
3rd Floor
45 rue des Saints Pères
75006 Paris
valentina.emiliani@univ
-paris5.fr

Alexandre Évrard

Laboratoire de Toxicologie du
Médicament
Faculté de Pharmacie
Université Montpellier 1
15 avenue Charles Flahault
34060 Montpellier
aevrard@univ-montp1.fr

Éric Forest

CEA
17 rue des Martyrs
38054 Grenoble Cedex
eric.forest@ibs.fr

Jean-Christophe Fricain

Biomatériaux et réparation
tissulaire
Université Victor Segalen
Bordeaux 2
Zone Nord, Bâtiment 4A
Case postale 45
146 rue Léo-Saignat
33076 Bordeaux Cedex, France
jean-christophe.fricain@
u-bordeaux2.fr

Benoît Frisch

Institut Gilbert Laustriat
UMR 7175 LC1
Université Louis Pasteur
Faculté de Pharmacie
74 route du Rhin BP 60024
67401 Illkirch, France
frisch@bioorga.u-strasbg.fr

Mélanie Gaillet

Horiba Jobin Yvon SAS
ZI de la Vigne aux Loups
5 avenue Arago
91380 Chilly-Mazarin
France
melanie.gaillet@jobinyvon.fr

Xavier Gidrol

Laboratoire d'Exploration
Fonctionnelle des Génomes
CEA d'Evry
2 rue Gaston Crémieux
CP 5722
91057 Evry Cedex, France
xavier.gidrol@cea.fr

Agnès Girard-Egrot

Institut de Chimie et Biochimie
Moléculaires et Supramoléculaires
UMR 5246 Université Lyon 1
CNRS-INSA Lyon-CPE Lyon
Laboratoire de Génie
Enzymatique et Biomoléculaire
Bât. CPE
Université Claude Bernard Lyon 1
43 bd. du 11 novembre 1918
69622 Villeurbanne cedex
agnes.egrot@univ-lyon1.fr

Pierre Grangeat

CEA-LETI, MINATEC
17 rue des Martyrs
38054 Grenoble Cedex
pierre.grangeat@cea.fr

Jean Hache

ESIEE Management (CCIP)
Cité Descartes
2 bd Blaise Pascal
BP 99
93162 Noisy le Grand cedex

Université d'Évry-Val d'Essonne
Bld F. Mitterrand
91025 EVRY Cedex, France
hachej@istm.fr

Francis Haraux

Laboratoire des protéines
membranaires, CEA Saclay
91191 Gif-sur-Yvette Cedex
France
francis.haraux@cea.fr

Emmanuèle Helfer

Laboratoire d'Enzymologie et
Biochimie Structurales
CNRS UPR 3082
Bâtiment 34
Avenue de la Terrasse
91198 Gif-sur-Yvette Cedex
emmanuele.helfer@lebs.
cnrs-gif.fr

Béatrice Heurtault

Institut Gilbert Laustriat
UMR 7175 LC1
Université Louis Pasteur
Faculté de Pharmacie
74 route du Rhin BP 60024
67401 Illkirch, France
heurtault@bioorga.u-strasbg.fr

Philippe Houdy

Laboratoire d'étude
des milieux nanométriques
Université d'Evry
Bld F. Mitterrand
91025 EVRY Cedex, France
philippe.houdy@univ-evry.fr

Marcel Lahmani

Laboratoire d'étude
des milieux nanométriques
Université d'Evry
Bld F. Mitterrand
91025 EVRY Cedex, France
lahmani@bp.univ-evry.fr

Philippe Laval

INSERM UMR 595
Université Louis Pasteur
11, rue Humann
67085 Strasbourg Cedex
philippe.lavalle@medecine.
u-strasbg.fr

Jean-François Legrand

Institut Charles Sadron
Université Louis Pasteur
Campus CNRS de Cronembourg
23 rue du Loess
BP 84047
67034 Strasbourg Cedex 2
legrand@ics.u-strasbg.fr

Christian Le Grimmelc

Centre de Biochimie Structurale
29 rue de Navacelles
34090 Montpellier Cedex
clg@cbs.cnrs.fr

Bruno Le Pioufle

Ecole Normale Supérieure de Cachan
61 avenue du Président Wilson
94235 Cachan, France
pioufle@bretagne.ens-cachan.fr

Laurent Lévy

Nanobiotix
60 rue de Wattignies
Bâtiment B, 3ème étage
75012 Paris
laurent.levy@nanobiotix.fr

Thimothée Lionnet

Albert Einstein College of Medicine
Yeshiva University, Bronx, USA
lionnet@lps.ens.fr

Brahim Lounis

Centre de Physique Moléculaire
Optique et Hertzienne
Université Bordeaux 1 and CNRS
351 cours de la libération
33405 Talence cedex France
b.lounis@cpmoh.u-bordeaux1.fr

Georges Lutfalla

CNRS UMR 5235
Dynamique des Interactions
Membranaires Normales
et Pathologiques (DIMNP)
Université Montpellier 2
CC086 Place Eugène Bataillon
34095 Montpellier Cedex 2
lutfalla@infobiogen.fr

Didier Marguet

Centre d'Immunologie de
Marseille Luminy
Parc Scientifique de Luminy
Case 906
13288 Marseille cedex 09
marguet@ciml.univ-mrs.fr

Christophe Marquette

Institut de Chimie et Biochimie
Moléculaires et Supramoléculaires
UMR 5246 Université Lyon 1
CNRS-INSA Lyon-CPE Lyon
Laboratoire de Génie
Enzymatique et Biomoléculaire
Bât. CPE
Université Claude Bernard Lyon 1
43 bd. du 11 novembre 1918
69622 Villeurbanne cedex
christophe.marquette@univ-lyon1.fr

Pierre-Emmanuel Milhiet

Centre de Biochimie Structurale
29 rue de Navacelles
34090 Montpellier Cedex
pem@cbs.cnrs.fr

Nicolas Minc

Laboratoire de Physicochimie
CNRS UMR 168
Institut Curie
26 rue d'Ulm
75005 Paris, France
nicolas.minc@curie.fr

Catherine Nguyen

Laboratoire de Technologies
Avancées pour le Génome
et la Clinique
ERM INSERM 206
Parc Scientifique de Luminy
163 avenue de Luminy
Case 928, 13288 Marseille Cedex 09
nguyen@tage.univ-mrs.fr

Ghani Oukhaled

Laboratoire de Matériaux
Polymères aux Interfaces
Université d'Evry-Val d'Essonne
Bâtiment Maupertuis
Bd F. Mitterrand
91025 Evry Cedex
abdel-ghani.oukhaled@univ-evry.fr

Philippe Peltié

CEA-LETI
17 rue des Martyrs
38054 Grenoble Cedex
philippe.peltie@cea.fr

Éric Perez

Laboratoire de Physique Statistique
de l'École Normale Supérieure
24 rue Lhomond
75005 Paris
eric.perez@lps.ens.fr

François Perraut

CEA-LETI
17 rue des Martyrs
38054 Grenoble Cedex
francois.perraut@cea.fr

Carine Pestourie

Laboratoire d'imagerie moléculaire
expérimentale
4 place du Général Leclerc
91401 Orsay Cedex, France
carine.pestourie@cea.fr

Delphine Pflieger

Laboratoire analyse et modélisation
pour la biologie et l'environnement
Université d'Evry-val-d'Essonne
Boulevard F. Mitterrand
91025 Evry Cedex
delphine.pflieger@espci.fr

Nathalie Picollet-D'Hahan

Commissariat à l'Energie Atomique
DSV/iRTSV/Biopuces
Bâtiment 4020
17 rue des Martyrs
38054 Grenoble Cedex 09
nathalie.picollet-dhahan@cea.fr

Frédéric Pincet

Laboratoire de Physique Statistique
de l'École Normale Supérieure
24 rue Lhomond
75005 Paris
frederic.pincet@lps.ens.fr

Laurent Pothuaud

MED-IMAPS
Avenue du Haut Leveque
33600 Pessac, France
lpothuaud@medimaps.fr

Pierre Puget

Département Technologies
pour la Biologie et la Santé
CEA-LETI, MINATEC
17 rue des Martyrs
38054 Grenoble Cedex 09
pierre.puget@cea.fr

Sylvie Ricard-Blum

Institut de Biologie et Chimie
des Protéines
UMR 5086 CNRS
Université Lyon I
7 passage du Vercors
69367 Lyon Cedex 07
s.ricard-blum@ibcp.fr

Ralf Richter

CIC BiomaGUNE
Parque Tecnológico
de San Sebastian
Po Miramón 182
20009 San Sebastian, Spain
rrichter@cicbiomagune.es

Émilie Roncali

Laboratoire d'imagerie moléculaire
expérimentale
4 place du Général Leclerc
91401 Orsay Cedex, France
roncali@shfj.cea.fr

Jean Rossier

Ecole supérieure de physique
et de chimie industrielles
10 rue Vauquelin
75231 Paris Cedex 05
jean.rossier@espci.fr

Pierre Schaaf

Institut Charles Sadron
Université Louis Pasteur
Campus CNRS de Cronembourg
23 rue du Loess
BP 84047
67034 Strasbourg Cedex 2
schaaf@ics.u-strasbg.fr

Daniel Scherman

Laboratoire de Pharmacologie
Chimique et Génétique
U 640 INSERM UMR 8151 CNRS
Université Paris Descartes
Faculté de Pharmacie
4 avenue de l'Observatoire
75006 Paris
daniel.scherman@univ-paris5.fr

Francis Schuber

Institut Gilbert Laustriat
UMR 7175 LC1
Université Louis Pasteur
Faculté de Pharmacie
74 route du Rhin BP 60024
67401 Illkirch, France
schuber@bioorga.u-strasbg.fr

Bernard Senger

INSERM UMR 595
Université Louis Pasteur
11, rue Humann
67085 Strasbourg Cedex
bernard.senger@medecine.
u-strasbg.fr

Pascal Silberzan

Laboratoire de Physicochimie
CNRS UMR 168
Institut Curie
26 rue d'Ulm
75005 Paris, France
pascal.silberzan@curie.fr

Virginie Simon

Nanobiotix
60 rue de Wattignies
Bâtiment B, 3ème étage
75012 Paris
virginie.simon@nanobiotix.fr

Vincent Studer

Ecole supérieure de physique
et de chimie industrielles
10 rue Vauquelin
75231 Paris Cedex 05
vincent.studer@espci.fr

Bertrand Tavitian

Laboratoire d'Imagerie Moléculaire
Expérimentale (LIME), CEA
INSERM 803
Imagerie de l'Expression des Gènes
4 place Leclerc
91400 Orsay, France
bertrand.tavitian@cea.fr

Isabelle Texier

CEA-LETI
17 rue des Martyrs
38054 Grenoble Cedex
isabelle.texier-nogues@cea.fr

Étienne Thévenot

Laboratoire des Processus
Stochastiques et Spectres
Commissariat à l'Energie Atomique
LIST, 91191 Gif-sur-Yvette
etienne.thevenot@cea.fr

Laurent Thouin

Ecole Normale Supérieure
Département de Chimie
UMR CNRS 8640
24 rue Lhomond
75231 Paris cedex 05
laurent.thouin@ens.fr

Olivier Thoumine

UMR CNRS 5091
Université Bordeaux 2
Institut Magendie
146 rue Léo Saignat
33077 Bordeaux
olivier.thoumine@pcs.
u-bordeaux2.fr

Marc Tramier

Institut Jacques Monod
2 place Jussieu
Tour 43
75251 Paris Cedex 05
tramier@ijm.jussieu.fr

Mona Treguer-Delapierre

ICMCB-CNRS
Université Bordeaux 1
87 Ave. du Dr A. Schweitzer
33608 Pessac Cedex, France
treguer@icmcb-bordeaux.cnrs.fr

Joëlle Vinh

Ecole supérieure de physique
et de chimie industrielles
10 rue Vauquelin
75231 Paris Cedex 05
joelle.vinh@espci.fr

Jean-Louis Viovy

Laboratoire de Physicochimie
CNRS UMR 168
Institut Curie
26 rue d'Ulm
75005 Paris, France
jean-louis.viovy@curie.fr

Jean-Claude Voegel

INSERM UMR 595
Université Louis Pasteur
11, rue Humann
67085 Strasbourg Cedex
jean-claude.voegel@medecine.
u-strasbg.fr

Richard Wade

Institut de Biologie Structurale
41 rue Jules Horowitz
38027 Grenoble Cedex 1
richard.wade@ibs.fr

Jérémy Weber

Société Fluigent
4 avenue de l'Observatoire
75006 Paris
jeremie.weber@fluigent.com

Guy Zuber

Guy Zuber
Université Louis Pasteur
Faculté de Pharmacie
74 route du Rhin
67400 Illkirch
zuber@bioorga.u-strasbg.fr

Structural and Functional Regulation of DNA: Geometry, Topology and Methylation

C. Auclair

The work of Rosalind Franklin, then Watson and Crick [1], established the architecture of deoxyribose nucleic acid (DNA), carrier of all genetic information. The idea that DNA was structurally organised in the form of a double helix comprising two antiparallel and complementary polymer chains was one of the great scientific discoveries of the twentieth century. It revealed not only the way in which genetic information is stored, but also the mechanism by which the genetic code is read, and the way this code can be faultlessly copied from one cell to another during cell division.

The structural organisation of genomic DNA varies significantly from one organism to another, or from one cell to another, depending as it does on the physiological constraints specific to each organism or tissue. This complexity can be observed in particular in the diversity of genomic sequences, the size of the human genome being something like 3 gigabases for about 30,000 genes, whereas yeast, a lower eukaryotic organism, only possesses 6,200 genes for a size of 13 megabases (see Table 1.1).

The fraction of protein-coding sequences is also highly variable (1.4% for the human genome, 68% for the yeast genome), and so too is the size of the genes. Particularly interesting is the variation in the content of G+C bases, which determines the overall stability of the DNA helices. Sequences rich in G+C bases are involved in the key processes regulating gene expression and probably in a dominant way in dynamical processes. An important point is the possibility of methylating cytosines, especially the CpG sequences, a crucial process in the control of gene expression. The presence of alternating sequences of GC base pairs, associated with the methylation of the cytosines in these sequences, favours in particular the transition from the B to the Z conformation (see below). Within a given genome, the G+C content can vary significantly, reaching 80% in some regions of mammal genomes, and there seems to be a correlation between the GC base content (especially GCs3) and the gene density in the relevant region.

The complexity of DNA depends directly on the kinds of sequences, but is also characterised by the broad range of micro- and macrostructures resulting

Table 1.1. Genomic characteristics of some eukaryotic organisms

	Yeast (<i>S. cerevisiae</i>)	Nematode (<i>C. elegans</i>)	Drosophila (<i>D. melanogaster</i>)	Human
Size (Mb)	13	100	180	3,000
[G+C] content	38%	36%	43%	41%
Number of genes	6,200	19,100	13,600	~30,000
Coding fraction	68%	27%	13%	1.4%
Number of exons per gene	1.04	5.5	4.6	8.7
Size of genes (kb)	1.4	2.7	3	28

from the physicochemical constraints imposed by sequencing and pairing of bases. The wide range of possible conformations of DNA, not to mention molecular arrangements such as cruciform, triple helix, and base tetrad structures, among others, plays a key role in the way genomes work, especially through their specific recognition by proteins carrying out important genetic functions.

One could describe DNA and its protein environment as a nanoworld of the most complex kind. Naturally, this complexity reflects the functions the system has to fulfill. However, one can nevertheless identify certain representative elements in the workings of the genetic machinery. In this context, the aim of the present chapter will be to provide, with the help of some examples, a succinct review of certain features arising from the geometric and topological flexibility of DNA, and to describe rather briefly the structural modifications related to methylation of cytosines and the functional modifications that result from them.

1.1 Geometry of the DNA Double Helix

The conformation originally described by Watson and Crick was a double helix known as the B conformation (see Fig. 1.1), which is the one most often observed in the natural state.

However, the double helix can occur in three geometric forms denoted A, B and Z, characterised in particular by different degrees of hydration. These various conformations of the DNA are made possible by an extraordinary level of geometric freedom allowed between the constituents of DNA. The main point is the existence of conformers: orientation of the sugar constituent (C2'-endo or C3'-endo) and orientation of the base with respect to the sugar (syn or anti) (see Fig. 1.2).

The DNA helix is characterised by the C2'-endo/anti conformation in the case of the B conformation, and the C3'-endo/anti conformation for the A conformation. The situation is a little more involved in the case of the Z

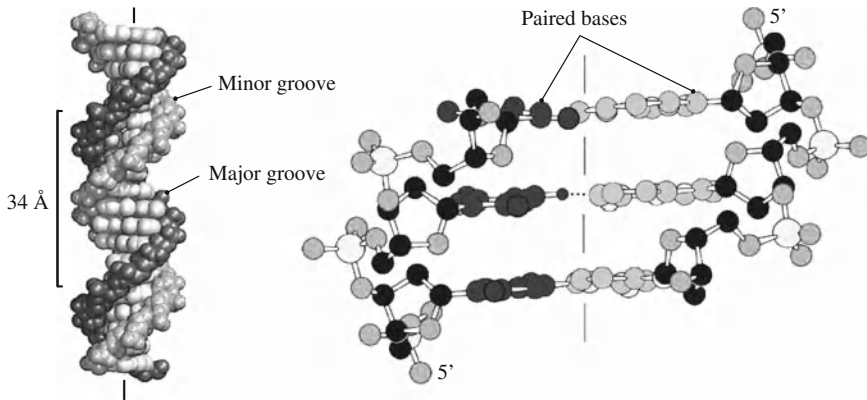


Fig. 1.1. *Left:* Molecular model of a DNA double helix in the B conformation. *Right:* Detail of base pairing

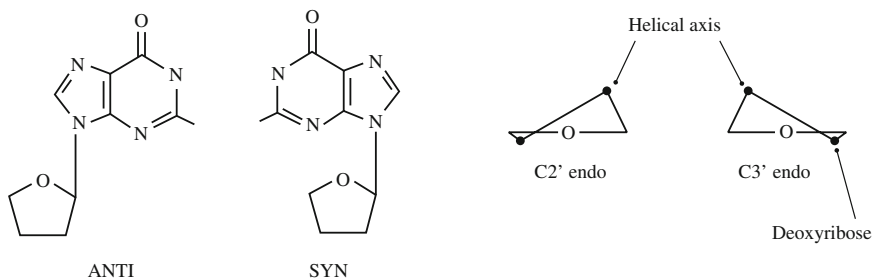


Fig. 1.2. Orientation of the base (guanine) with respect to the deoxyribose (anti and syn) and orientation of the deoxyribose (C2' endo and C3' endo)

conformation, where the purine bases adopt the C3'endo/syn conformation while the pyrimidine bases adopt the C2'endo/anti conformation. In fact, the structural characteristics and the evolution of the DNA double helix toward one of the possible conformations are conditioned by the parameters known as twist, roll, and tilt specifying the helix and the set of six torsion angles α , β , γ , δ , ε , and ζ of the phosphate-sugar backbone.

Figure 1.3 shows the main geometric arrangements that the stacks of bases can adopt in relation to one another. The twist is the angle of rotation of two adjacent stacks of base pairs about the helical axis of symmetry. The roll corresponds to the angle of rotation of two adjacent stacks of base pairs about the third axis of the helix. The tilt is the angle of rotation of two adjacent stacks of base pairs about the pseudosymmetry axis of the helix. It should be noted that the position of the plane of each paired base can also vary relative to the other base in the pair (opening, propeller, and buckle), thus increasing the flexibility of the whole construction.

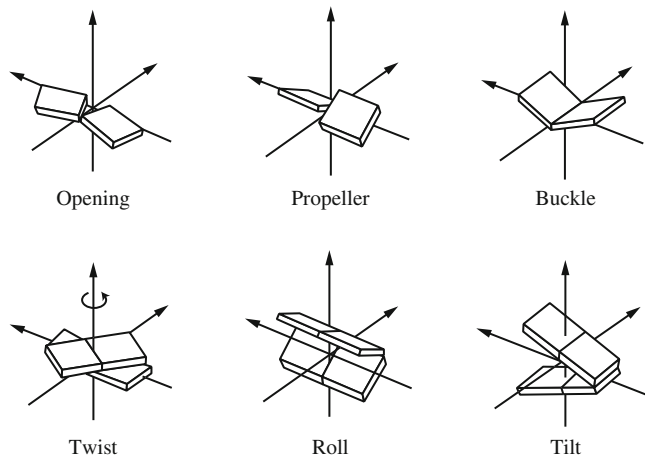


Fig. 1.3. Helical parameters specifying the geometry of the DNA helix. *Arrows* indicate the axes of symmetry of the helix

Table 1.2. Estimated values of the parameters specifying a B-DNA helix for different sequences. Angles are given in degrees and correspond to the angles between a stack of base pairs and its nearest neighbour. The distance between stacks of base pairs is set to 3.4 Å. The method here is due to Bolshoy [3]

Sequence	Twist	Roll	Tilt
A	—	—	—
A	35.67	−6.5	3.2
A	35.67	−6.5	3.2
T	—	—	—
A	36	0.9	0
T	31.2	2.6	0
G	—	—	—
G	33.67	1.2	−1.8
G	33.67	1.2	−1.8
C	—	—	—
G	29.8	6.7	0
C	40.1	−5	0

There are rules [2,3] for assessing the static structure of the DNA molecule as a function of its sequence. The problem is to estimate the above helical parameters. Using the computation program devised by the Georgia Institute of Technology [4], it is easy to demonstrate the dependence of the sequence on the these parameters. As an illustration, Table 1.2 shows the changes in the twist, roll, and tilt for two non-alternating sequences, viz., $-AAAA \dots$ and $-GGGG \dots$, and for two alternating sequences, $-TATA \dots$ and $-CGCG \dots$.

Among the remarkable points regarding the data in the table, note the high twist and roll engendered by poly A sequences. This largely explains the curvature of the helix observed in regions with this type of sequence. Indeed, such non-alternating sequences can induce significant deviations from helical symmetry. Modelling and molecular dynamics simulations have shown that this kind of planar curvature results from a tendency to stretch the sugar-phosphate backbone, causing compression with a modification of the torsion and the alignment of the bases, which in turn leads to curvature. This kind of curvature can play an important part in interactions between DNA and its ligands, especially ligands of the minor groove and multimer proteins.

Furthermore, high-resolution structural analysis of the torsion angles of the B helix (angles ϵ C3'–O3' and ζ O3'–P) has revealed the existence of BI and BII subconformations [5]. Note also that the ionic strength of the environment of the helix and its level of hydration play predominant roles in determining the conformation that is finally adopted.

Apart from its ability to store and transmit genetic information, it has now been shown that the DNA molecule can itself carry out some degree of regulation, controlling among other things the reading of the code, i.e., the level of gene expression depending on the specific needs of each cell. This regulatory ability arises partly from the presence of regulating sequences located upstream of gene reading frames, and partly from the many different structural and conformational modifications to which the DNA molecule may be subjected. Structural patterns arising from the kind of sequences and/or the architectural organisation of DNA are recognised by a wide range of different proteins, which then act as effectors for genetic functions. This system is effectively based on a kind of molecular recognition, suggesting a great diversity and high level of flexibility in the regulating structural patterns. In the light of more recent work, the phenotype of a cell, i.e., its functional characteristics and morphology, can also be considered to depend just as much on the kinds of genes as on the architectural organisation of the genome. From this standpoint, the available conformational and topological variants of DNA look more and more like key regulatory parameters.

1.2 The Z Conformation of DNA

Since the initial description of the structural parameters specifying the DNA double helix in its most common form (the B conformation depicted in Fig. 1.1), a great deal of further work has shown that this architectural organisation exhibits an exceptional level of flexibility, able to generate an enormous number of variants, each of which would appear to contribute in a crucial way to the functional activities of DNA, including regulating the expression of coded genes.

A representative example of the conformational flexibility of DNA is provided by the Z conformation (see Fig. 1.4). Indeed, it is a striking fact that,

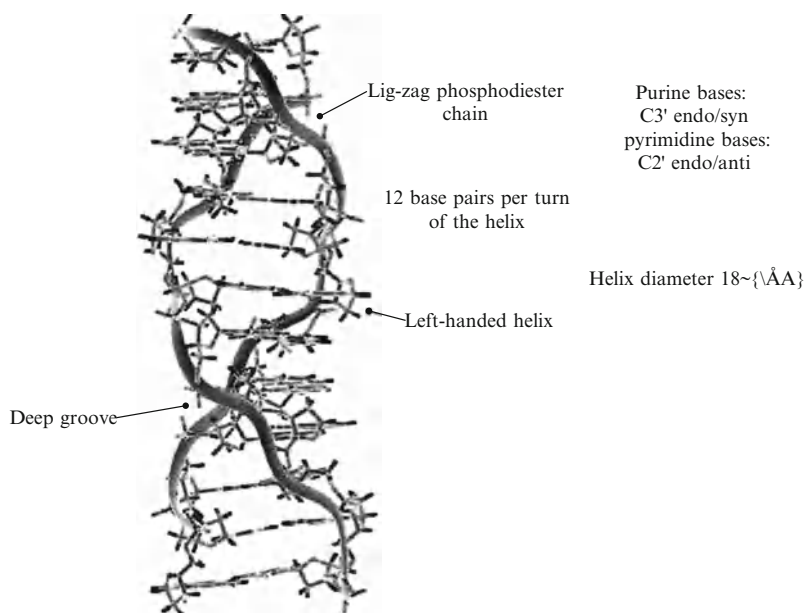


Fig. 1.4. Molecular model of a DNA double helix in the Z conformation, the left-handed helix generally favoured in regions rich in G-C base pairs

following a relatively small energy input, the DNA double helix can change from its B conformation (right-handed helix) to a Z conformation, a left-handed helix with very different structural characteristics (see Table 1.3). In fact, the activation energy required for transition from the B conformation to the Z conformation is about 22 kcal mol^{-1} , roughly equivalent to the energy needed for base pair breaking, whether the DNA is in the B or the Z form [6]. This indicates the facility with which the transition can occur from an energetic point of view and also the identity of the molecular dynamics of the two conformations.

In its B conformation, the angle of rotation between consecutive bases is about 36° , so that there are 10 pairs of bases for each turn of the helix. One turn occupies about $34 \text{ \AA}$, implying a distance of $3.4 \text{ \AA}$ between two consecutive base pairs. In its Z conformation, the DNA helix is characterised by a high value of the distance occupied by one turn of the helix and the presence of 12 base pairs per turn. One also finds a reduction in the diameter of the helix and only one rather deeply indented groove. The Z helix is also characterised by the coexistence of glycosidic bonds in the syn form for purine bases and in the anti form for pyrimidine bases. One consequence of this bond alternation is that the sugar-phosphate backbone of the helix adopts a zig-zag shape rather than a regular spiral as it does in B-type DNA. Another consequence that is probably important from a functional point of view is the non-uniform

Table 1.3. Comparative structural parameters for the B and Z forms of the DNA double helix

Structural parameters	B	Z
Orientation of helix	Right-handed	Left-handed
Repetition	1 bp	2 bp
Rotation/bp	35.9°	−30°
Average number of bp/turn	10.0	12
Angle between base and axis	−1.2°	−9°
Distance between bp along axis	3.4 Å	3.7 Å
Distance for one turn of the helix	34 Å	45 Å
Average torsion	+16°	0°
Glycosidic bond	anti	C: anti, G: syn
Sugar conformation	C2'-endo	C: C2'-endo, G: C2'-exo
Diameter of helix	20 Å	18 Å

distribution of negative charges along the helix, which can significantly modify interactions with the various ligands, especially protein ligands.

The transition from the B conformation to the Z conformation is favoured by a high ionic strength, indicating a strong electrostatic component in the stability of the two conformers. The compacted form of the Z helix is to a large extent stabilised by a high salt concentration and the particular hydration network solvating the charges and polar groups [7].

There is a connection between the nucleotide sequence in the DNA and the ability of the double helix to adopt a Z conformation. This is particularly so for alternating sequences of type GCGCGC. The determining factor explaining this phenomenon is in fact the presence of an amine in position 2 of the guanine, which stabilises the Z helix. Replacing the guanine by inosine (which is a guanine but without the amine in position 2) considerably destabilises the Z conformation [8]. In a poly d(G-C) alternating sequence, the difference of free energy between the B form and the Z form is slight, being about $0.33 \text{ kcal mol}^{-1}$ per base pair [9], and this favours the great stability of the Z conformation.

The tendency of pyrimidine–purine dinucleotides to form sequences of Z-type DNA is, in decreasing order:

$$\text{m5CG} > \text{CG} > \text{TG} = \text{CA} > \text{TA} \quad (\text{m5C:5-methylated cytosine}).$$

This brings us to the interesting role played by methylation of cytosines, which also tends to stabilise the Z conformation. This stabilising effect probably comes from the fact that the methyl group at position 5 on cytosine prevents the setting up of a hydration network which stabilises the B form. In fact, alternating sequences of G-C type can occur in the B form or in the Z form and are subject to changes in equilibrium with the environment of the helix, energy constraints applied to the helix, and the presence of chemical or protein

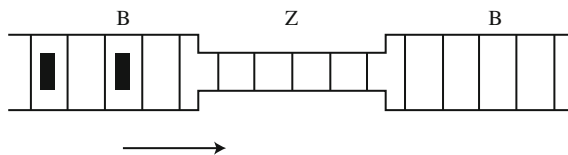


Fig. 1.5. Schematic view of a DNA sequence in the Z conformation inserted into a B helix. Intercalating a plane molecule in the B sequences adjacent to the Z-type sequence induces a Z-to-B transition at a distance

ligands. Intercalating agents bound to sequences adjacent to Z-conformation regions (see Fig. 1.5) can easily induce the Z-to-B transition, whereas groove ligands have almost no effect [10].

This way of inducing the transition results from the significant unwinding of the B helix, which induces stresses from a distance. On the other hand, torsion stresses induced by the transcription process clearly favour the transition from B to Z. This is consistent with the fact that the Z conformation is favoured by negative superhelicity in the DNA helix (see Sect. 1.3). From the biological point of view, note that regions close to transcription initiation sites are rich in sequences favouring the B-to-Z transition. Furthermore, as the RNA polymerase moves along the transcribed DNA strand, regions of increasingly negative superhelicity form upstream. Globally, the formation of Z helices near promoter sequences seems to stimulate transcription. It seems likely that the possibility of conformational transitions in regions of high topological stress provide a way of minimising the energy required for the process to go ahead, or even a way of temporarily stabilising favourable architectural arrangements.

From a structural point of view, the coexistence of B and Z forms raises many questions, and the architectural arrangement of the helix at the B–Z junction always seemed somewhat mysterious, until teams led by A. Rich and K.K. Kim succeeded in resolving this junction at 2.6 Å using X-ray diffraction [11].

The structure of the B–Z junction is characterised by pair breaking in an A–T base pair, while the unpaired bases are extruded outside the helix as shown in Fig. 1.6. This pair breaking corresponds to an energy relaxation which allows the system to maintain a regular stacking of the bases. This is consistent with the free energy estimate of 5 kcal mol^{−1} for the B–Z junction as put forward by Peck and Wang.

There can be no doubt that the special structure of this junction characterised by the two extruded bases could constitute a motif for recognition by certain proteins.

Several proteins have been identified that look likely to bind preferentially to Z-type DNA. This is the case for the protein E3L of the vaccinia virus, the protein AF2008 of *Archaeoglobus fulgidus*, the protein DLM-1, and the RNA-editing enzyme ADAR-1. These proteins have a Z-DNA binding region

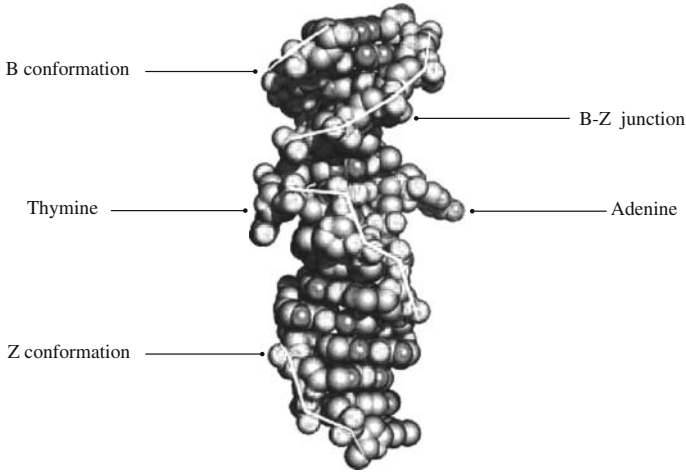


Fig. 1.6. Structural organisation of the B–Z junction [11]

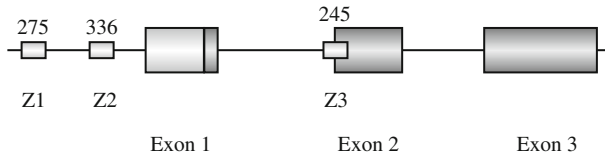


Fig. 1.7. Region of the c-myc gene in which there are three sequences Z1, Z2, and Z3 assuming the Z conformation during transcription [14]

(Za) located in the N-terminal region. The 3D structure of the Za region of ADAR-1 shows that there is a helix–turn–helix motif and a single bond with a guanine in the syn conformation, characteristic of Z-DNA [12].

The 3D structure of the protein AF2008 from *Archaeoglobus fulgidus* resolved at 1.55 Å [13] reveals the dimeric organisation of the protein, in which the Z-DNA binding region of each monomer is separated by 45 Å, a distance corresponding to one turn of the DNA helix in the Z conformation.

Many observations have now been made suggesting that the Z conformation of DNA plays a determining role in the control of gene expression.

Using permeabilised nuclei of U937 cells, it has been shown that, during transcription of the c-myc gene, three Alu I restriction sequences located in the vicinity of the promoter sequences adopted a Z conformation (see Fig. 1.7) [14].

At the end of transcription, the Z conformations rapidly disappear, this being related to relaxation of negative supercoiling by topoisomerase I. Many other examples can be found in the literature. One recent observation has confirmed the crucial biological role of Z-DNA. Indeed, it has been shown that integrity of the Z-DNA binding region of the protein EL3 is crucial for the virulence of the vaccinia virus [15]. Deleting the 83 amino acids of the N-terminal region of EL3 totally deactivates the virus. However, replacing these

83 amino acids by the Z-DNA binding regions of DLM1 or ADAR1 does not affect the virulence of this virus.

The early detection of Z helices *in vivo* by virtue of the fact that it is easy to obtain anti Z antibodies long remained a mystery. However, it is clear today that the B-to-Z transition has an important regulatory effect with regard to genetic functions. Stresses induced when the negative superhelicity increases, which is almost the rule during replication and transcription of DNA, systematically lead to the formation of Z helices in poly-purine/poly-pyrimidine sequences. This is also the case in nucleosomal DNA (see Chap. 2), especially in the close-packed regions of chromatin. Proteins binding on the Z helices can thereby act to stabilise the structure and possibly also to activate or repress transcription.

The biological significance of the Z form of DNA has been indirectly confirmed by the presence of a large number of sequences for which the probability of undergoing a B-to-Z transition is high, and this for a moderately high level of superhelicity. As an example, on the human chromosome 22, there are 7,580 regions exhibiting these characteristics [16].

1.3 Supercoiled DNA

A second example demonstrating the molecular diversity of DNA helices is provided by the topological variations of DNA so well characterised in circular DNA. In bacteria, some viruses, and mitochondria, DNA helices do indeed adopt a closed circular form as shown in Fig. 1.8.

Circular DNA helices can arrange themselves in space to form positive or negative supercoils, leading to what are known as topoisomers. The

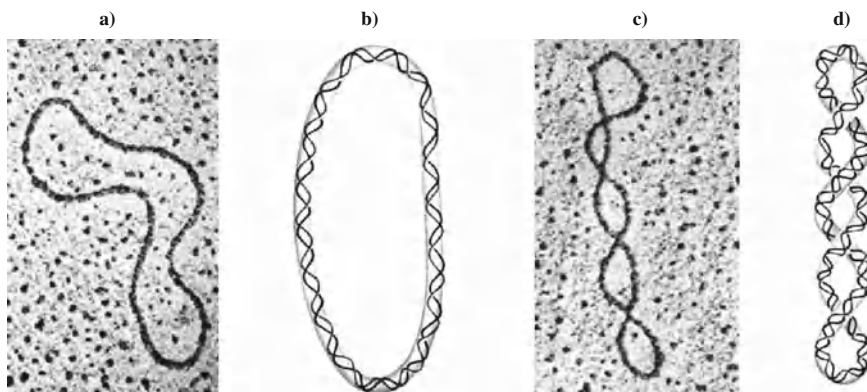


Fig. 1.8. (a) and (b): Relaxed circular DNA. (c) and (d): Supertwisted DNA with negative supercoiling. (a) and (c) are electron microscope images, while (b) and (d) are diagrammatic

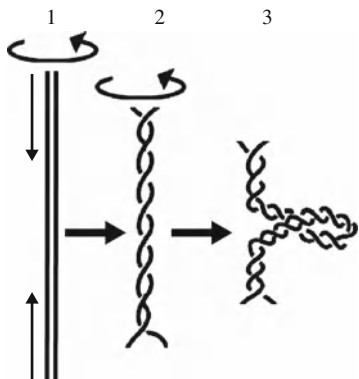


Fig. 1.9. Writhe of a straight, double-stranded helix leading to supercoiling. (1) $T = 0$, $W = 0$. (2) $T = 8$, $W = 0$, $L = 8$. (3) $W = 1$, $\Delta L = 1$

supercoiling of DNA is characterised by three parameters: the linking number L , the twist T , and the writhe W . These are related by the simple formula

$$L = T + W .$$

In a relaxed circular DNA, the writhe corresponding to the degree of supercoiling is zero. In supercoiled DNA (see Fig. 1.8c), $W = -4$, and the linking number is less than the twist. Figure 1.9 illustrates schematically the supercoiling phenomenon and the relation between twist and writhe.

The difference in supercoiling energy between two topoisomers depends on the square of the change in linking number:

$$\Delta G_{sc} = \frac{1}{N} KRT(\Delta L)^2$$

In the above example, $(\Delta L)^2 = 16$.

It is a striking thing that DNA always has negative supercoiling in bacteria, but positive supercoiling in archaeobacteria, and both geometries in eukaryotes.

The degree of supercoiling and its orientation depend on the angle of rotation between consecutive bases, itself determined by the distance between base pairs. As an example, intercalating a planar molecule such as ethidium bromide, or antitumor molecules such as the ellipticines and acridines, between base pairs of a circular DNA induces a rotation of adjacent base pairs of the order of -20° to -26° at the intercalation site. Depending on the increasing number of intercalating molecules, this rotation induces a relaxation of negatively supercoiled DNA first to circular DNA, then to positively supercoiled DNA. However, ligands binding in the minor groove of DNA, such as netropsin, induce a twist of a few degrees, favouring the formation of a negative superhelicity and eventually a close-packing of the DNA. Non-closed DNA can also undergo negative or positive supercoiling. This happens, in particular, for nucleosomal DNA in chromatin structures (see Fig. 1.10).

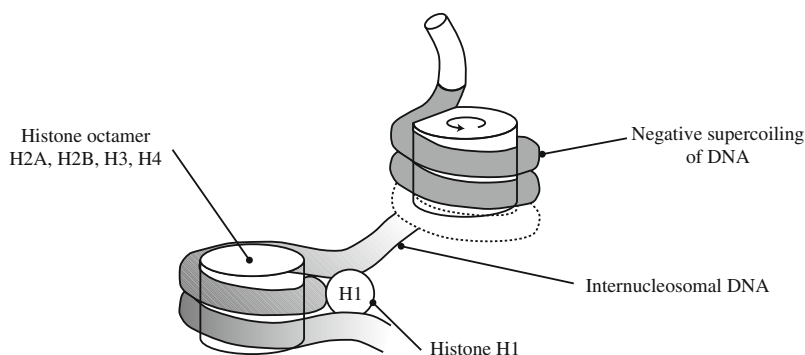


Fig. 1.10. Schematic view of nucleosomal DNA. The DNA molecule is wrapped around a histone octamer stabilised by histone H1

In nucleosomes, the wrapping of DNA around histones corresponds to negative superhelicity. Note that, excepting the nucleosome case, DNA can adopt a positive superhelicity. The situation is in fact relatively complex insofar as the nucleosome can fluctuate between three conformational states. Two of these conformations are characterised by a wrapping of about 1.7 turns around the octamer and a negative or positive crossing over of the leading and lagging DNA strands. The third conformation corresponds to a more open architecture, characterised by a partial wrapping of about 1.45 turns without crossing over of the leading and lagging DNA strands, following an unwrapping of the DNA on either edge of the nucleosome. The possibility of attaining these conformations is regulated by the wrapped DNA sequence. This affects the local twist of the helix, which in its turn acts on the organisation of the leading and lagging DNA strands. It is clear that the nucleosomal architecture must have a very significant level of conformational flexibility. This is illustrated by the fact that histone (H3–H4)₂ tetramer can associate equally well with either a positively or a negatively supercoiled DNA minicircle [17].

The extent of DNA supercoiling is also directly regulated by enzymes called topoisomerases which play a determining role in many genetic functions such as replication, transcription, and so on. These enzymes act by cutting then religating DNA strands, one strand for type I topoisomerases and two for type II topoisomerases. In prokaryotic organisms, topoisomerase I reduces negative supercoiling, while gyrase, acting as a type II topoisomerase, preferentially reduces positive supercoiling. Changes in structure and close-packing introduced in this way play a determining role in regulating DNA functions. Moreover, the enzymes known as helicases can unwind the DNA helix in order for replication to take place and can in this way induce topological changes in the vicinity of the replication fork. The same is true during transcription, where the RNA polymerase itself has a helicase action. As an example, the RNA polymerase in *E. coli* unwinds the DNA helix by 140° [18]. The key observations in this field were made by Liu and Wang [19], who showed that, during

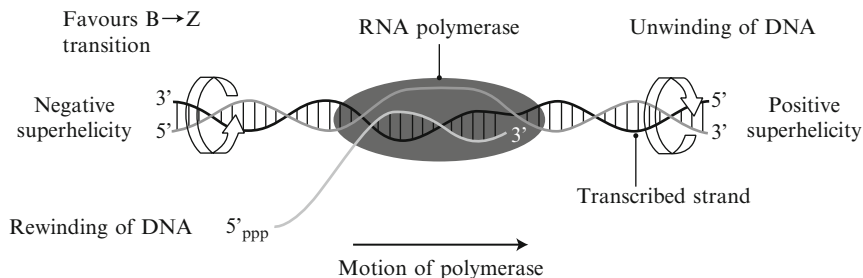


Fig. 1.11. Winding and unwinding of DNA during transcription

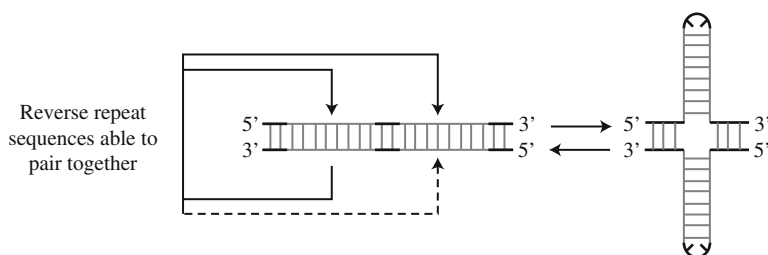


Fig. 1.12. Equilibrium between linear and cruciform architectures at the site of a palindromic sequence in double-stranded DNA

transcription, RNA polymerase appears to generate on the DNA array a negative superhelicity upstream of the polymerase and a positive superhelicity downstream (see Fig. 1.11).

As a secondary effect, the unwinding of the DNA downstream of the polymerase induces a positive superhelicity, while the winding upstream produces a negative superhelicity. As mentioned in the last section, the negative superhelicity generated upstream of the polymerase favours the transition from a B conformation to a Z conformation in the presence of the right sequences, e.g., poly d(G-C), and this probably helps to minimise structural stresses in the transcription machinery. Note that a negative superhelicity favours transcriptional activity. Indeed, the RNA polymerase binds very well on circular DNA with a negative superhelicity but only weakly on relaxed circular DNA. This is one reason why transcription occurs near nucleosomes.

One of the last points to consider is the fact that stresses induced on the helix by a positive or negative superhelicity can generate major structural changes in the helix as a secondary effect. This is the case with regard to the formation of cruciform structures when there are palindromic sequences (see Fig. 1.12), structures revealed in supercoiled circular DNA [20].

For palindromic sequences, the linear form is in equilibrium with the cruciform structure. Note that this equilibrium transiently generates single-strand structures sensitive to SI nuclease. The shift in equilibrium depends on a certain number of parameters such as the presence of magnesium and the degree

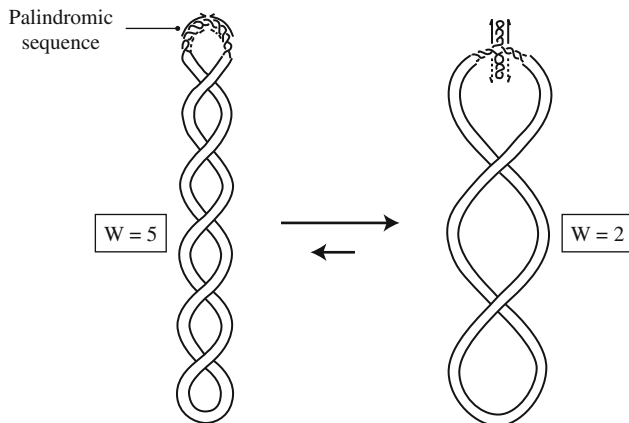


Fig. 1.13. Relation between the degree of supercoiling of DNA and generation of a cruciform structure from a palindromic sequence

of superhelicity. High superhelicity induces formation of the cruciform structure, whereas relaxation favours the linear form. In actual fact, formation of the cruciform structure induces as a consequence a reduction in superhelicity (see Fig. 1.13), and in the end reduces the free energy of the system. In the example of Fig. 1.13, if the twist T is considered to remain constant, then formation of the cruciform structure reduces the linking number L : $\Delta L = 3$.

It should be stressed that the role of these cruciform structures remains rather mysterious. Although it is easy to identify cruciform structures in vitro in circular DNA with a high level of superhelicity, it is extremely difficult to identify such structures in the genomes of eukaryotic cells under physiological conditions. However, the use of monoclonal antibodies directed against these structures and also photoinduced crosslinking experiments suggest that cruciform structures do indeed exist in living cells [21].

For other reasons, their presence as transient structures would appear highly probable:

- Type I and II topoisomerases, enzymes involved in the relaxation of superhelicity, preferentially recognise cruciform structures and cut DNA in the vicinity of these structures. In the light of this observation, cruciform structures can be considered as topological stress markers.
- Palindromic sequences are often found in sequences regulating gene expression (promoter sequences), the cruciform structures generated by these palindromes being specifically recognised by proteins regulating gene expression (transcriptional activators or repressors). This last point suggests that cruciform structures in relation with topological DNA variants might play an important part in regulating genetic expression.

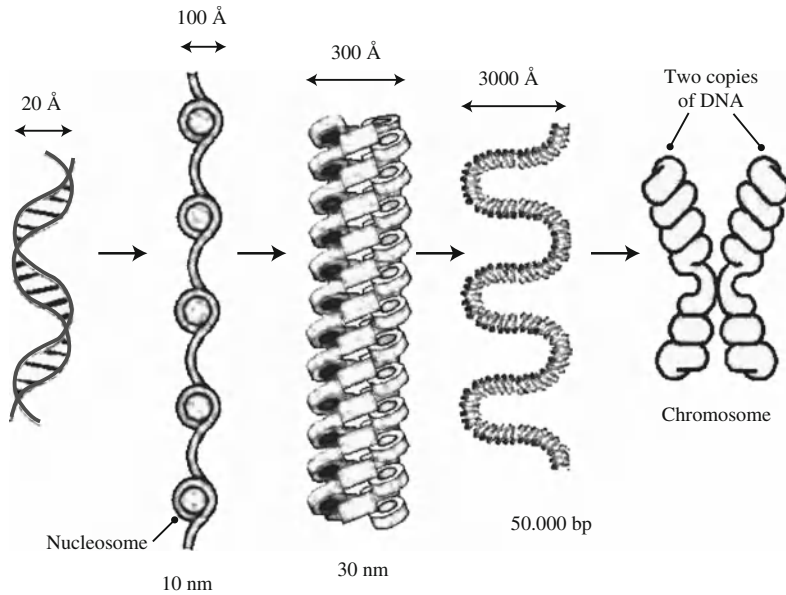


Fig. 1.14. Organisation and packing of DNA in chromatin and chromosome

1.4 Methylation of DNA

In association with geometric and topological modification of DNA, another essential ingredient in the control of genetic expression is chemical modification of the bases themselves, and in particular the methylation of cytosines.

Before discussing this point in more detail, it will be useful to review briefly the way DNA is organised in the nuclei of eukaryotic cells. The architectural organisation of DNA in the nucleus, and the eventual formation of chromosomes, are largely dictated by the need for a closely packed structure. Indeed, the human genome comprises some 3 billion base pairs in the form of double helices. In linear form, this would have a length of about two meters. The problem here is that these two meters of helix must somehow be packed into the nucleus, which measures only about one micron in diameter. Figure 1.4 summarises the organisation of the DNA and the different levels of packing.

The first level of organisation is the association of DNA fragments, containing some 200 base pairs each, with a globular protein structure comprising four histone dimers H2A, H2B, H3 and H4 (a histone octamer) to form the nucleosome, a structure mentioned in the last section (see Figs. 1.8 and 1.14). Note that the wrapping of DNA around the histone octamer is stabilised by H1 histone, which fixes the structure into place (see Fig. 1.8). The chain of nucleosomes thereby formed constitutes a 'fibre' of diameter about 10 nm (see Fig. 1.14). Under biological conditions, this fibre folds up to form the chromatin fibre of diameter 30 nm. The latter is then compacted in the form of

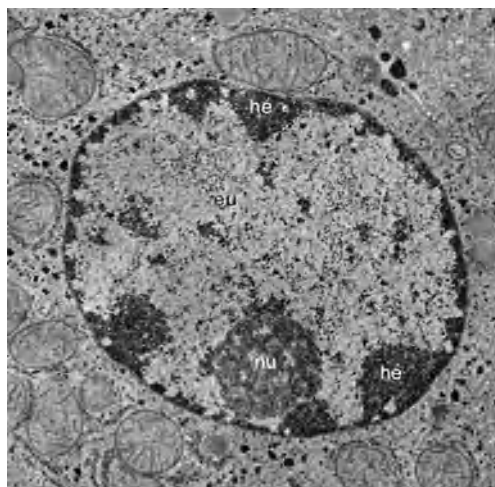


Fig. 1.15. Electron microscope image of the nucleus of an epithelial cell showing the non-uniformity of the chromatin. He: heterochromatin regions where DNA is more densely packed and hence more opaque to electrons. Eu: euchromatin regions where the DNA is less densely packed. Nu: nucleolus

a chromosome before cell division. To a large extent, DNA compaction is effected by electrostatic forces resulting from interactions between negatively charged phosphates on the DNA and basic amino acids on the positively charged histones.

In fact, within the nucleus of a non-multiplying cell, the degree of close-packing of the DNA is variable from one point to another. There are regions where it is less densely packed called euchromatin and others where it is more densely packed called heterochromatin (see Fig. 1.15). It was shown very early on that the heterochromatin with its densely packed DNA and low genetic expression is characterised among other things by a high level of methylation of the DNA, whereas the euchromatin, a region of strong genetic expression, is usually not significantly methylated and not very densely packed. Despite this kind of observation, DNA methylation was originally viewed by the scientific community as a secondary phenomenon of little biological interest.

However, major progress over the last few years has shown that methylation does in fact play an essential role in several biological processes associated with development, such as deactivation of the X chromosome in female mammals, genomic imprinting, and the expression of genes specific to different tissues. It is now accepted that the methylation of cytosines, in the context of the architectural organisation of chromatin, is one of the key mechanisms for regulating gene expression. This chemical modification of DNA is said to be epigenetic, because it can be transmitted from one cell to another and modulates the activity of a gene without directly affecting the sequence. The methylation of cytosines is preferentially located in specific genomic regions

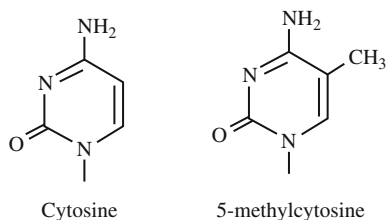


Fig. 1.16. Chemical structure of cytosine and 5-methylcytosine

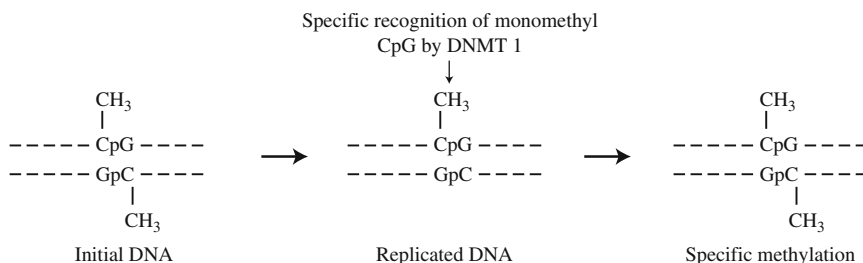


Fig. 1.17. Conservation of methylation in CpG sequences during DNA replication

within CpG dinucleotides. Indeed, globally speaking, 2–7% of cytosines in DNA are methylated, while some 70–90% are methylated in CpG sequences.

Finally, it seems more and more obvious that an exaggerated level of methylation in DNA plays a fundamental role in cancerogenesis. Such hypermethylation could deactivate genes suppressing tumour formation, thereby leading to tumorigenesis comparable to that induced by genetic mutation [22].

1.4.1 Methylation of Cytosine

In eukaryotes, the methylation of cytosine at position 5 is catalysed by a methyl transferase (see Fig. 1.16). The catalysed reaction involves transfer of a methyl group from S-adenosyl-methionine to a cytosine built into the DNA double helix. There are in fact three families of methyl transferases: DNMT 1, 2, together with 3a and 3b. DNMT 1 seems to be more specialised in maintaining overall levels of DNA methylation during cell multiplication. DNMT 3a is the methylation enzyme for gene regulating sequences and DNMT 3b is more specifically involved in methylation of centromeric sequences. These two enzymes can catalyse methylation of CpG sequences *de novo*, whereas DNMT 1 catalyses the methylation of semi-methylated CpG sequences.

An important point here is that the double methylation observed in CpG sequences is a structural characteristic that can be transmitted during cell multiplication. This is related to the semi-conservative nature of DNA replication and the specific recognition of monomethylated CpG sequences by methyl transferases, in particular DNMT 1 (see Fig. 1.17).

The enzyme DNMT 1 involved in maintaining the methylation of cytosines in CpG sequences is a very large protein, containing 1,618 amino acids. This protein is characterised by the presence near its N-terminal of three recognition regions denoted RTR1, RTR2, and RTR3, where RTR stands for replication target region, which are DNA replication regions. The C-terminal region contains the catalytic region and the recognition region of the semi-methylated DNA.

It is worth stressing that the level of expression of DNMT 1 varies during the cell cycle with a maximum in the G1S and S phases (S = DNA synthesis). This corresponds to the need to methylate cytosines on newly synthesised DNA strands. Consequently, and rather unexpectedly, one may consider that a high potential for methylation is not only compatible with, but even favours active cell proliferation. In confirmation of this idea, high levels of DNMT 1 expression have been observed in acute and chronic myelogenous leukaemia [23]. Moreover it would seem that many tumour cell lines can be characterised by hypermethylation of tumour-suppressing genes, and this favours the growth of these tumours. Such hypermethylation is mainly observed in CpG islands (see the next section) preferentially located in promoter regions. The kind of methylation related to the tumour phenotype can in fact be characterised by an overall hypermethylation of the genome, associated with hypermethylation of CpG islands controlling the expression of tumour-suppressing genes and, more generally, genes involved in the negative regulation of cell proliferation.

It should be stressed that the epigenetic mechanism of gene regulation and expression is a rather complex system, involving several concomitant processes. As an example, the methylation of gene regulating sequences (promoters), in a context of inhibition of expression, is accompanied by methylation of histones, in particular H3, on lysine 9, this leading to a form of cooperation between two methylation processes. We shall see in Sect. 1.4.4 that bimethylated CpG sequences are specifically recognised by proteins such as MeCP2, which binds onto these sequences. The complex formed in this way subsequently recruits histone deacetylases (HDAC), which catalyse the elimination of acetyl groups present on the histones. Acetylation of the histones facilitates the action of chromatin remodelling factors, leading to architectural changes allowing the opening and activation of promoters. The regulation machinery is then completed by the binding of other regulatory proteins such as HP1 (heterochromatin protein 1). This machinery can be summarised as follows:

- **Active Region for Gene Expression:**
 - Euchromatin: lightly packed chromatin.
 - Little methylation of CpG sequences.
 - Little methylation of H3.
 - Absence of HP1 proteins.
 - High level of acetylation of histones (action of histone acetylases HAT).
- **Inactive Region for Gene Expression:**
 - Heterochromatin: densely packed chromatin.

- High level of methylation of CpG sequences.
- High level of methylation of H3.
- Presence of HP1 proteins.
- Little acetylation of histones (action of histone deacetylases HDAC).

1.4.2 CpG Sequences

As we have seen, cytosines in the CpG dinucleotide sequence are favoured targets for methylation. Interestingly, the CpG sequence is much less well represented in vertebrate genomes [24]. This statistical anomaly resulting from selection pressure is probably due to the fact that cytosine is very easily methylated to give 5-methylcytosine (see Fig. 1.16), which is then easily deaminated to give thymine. This leads to a guanine–thymine mismatch that is not recognised by repair systems. This under-representation of CpG sequences is also observed in mitochondrial DNA, exemplifying the adaptation of prokaryotic DNA to a eukaryotic environment and functionality.

Despite their comparative overall scarcity in the genomes of higher eukaryotes, CpG sequences play a major part in the control of gene expression, with methyl-CpG sequences displaying a significant repressive potential. One remarkable feature relating to the regulatory function of CpG sequences is the presence of high CpG concentrations (CpG islands) [26] in the vicinity of promoter sequences for genes essential to the functioning of the cell and usually constantly expressed. These islands, comprising more than 200 nucleotides, are characterised by a high density of GC bases and a low level of methylation of the cytosines in the CpG sequences (in contrast to what is observed in isolated CpG sequences). Such hypomethylation is a prerequisite for strong gene expression, since methylation usually corresponds to inhibition of gene expression. Note that over-representation of CpG islands, associated with hypermethylation, can lead to anomalous under-expression of genes downstream of these islands. This happens, in particular, in the mental retardation syndrome associated with the fragile X site, characterised by deactivation of the FRM1 (fragile X mental retardation 1) gene. Deactivation of FRM1 can be imputed to an increase in the number of CGG triplets downstream of the promoter for this gene. In the population as a whole, the number of repeats is somewhere between 5 and 59 CGG, whereas patients affected by this syndrome have more than 200 repeats [27].

In order to understand the role played by CpG sequences in the functional regulation of the genome, it is essential to study the structural characteristics of these sequences. Note to begin with that the G-C pairing is stabilised by three hydrogen bonds, in contrast to the pairing of A-T nucleotides which only involves two hydrogen bonds (see Fig. 1.18). The first gas-phase measurements of the binding energy [28], later confirmed by other techniques, gave values of 21 kcal/mol for the G-C pairing, compared with 13 kcal/mol for the A-T pairing.

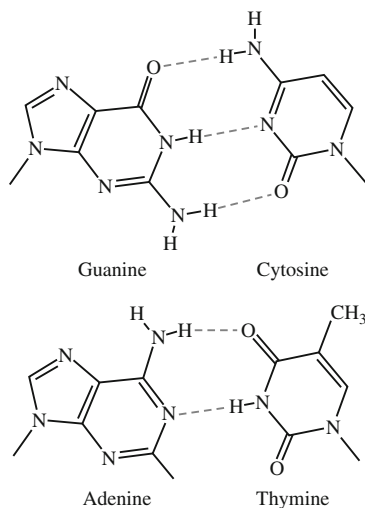


Fig. 1.18. Comparative pairing of G-C and A-T bases in a double-strand DNA molecule

The high binding energy of the G-C pairing makes the CpG dinucleotide sequence an extremely stable entity, which determines several of its characteristic properties. Apart from its stability, the CpG sequence displays an extraordinary flexibility. As an example, in B helices it has been observed that elongation and unwinding are more energetically favourable in $d(\text{CpG})_2$ (9.8 kcal) than in $d(\text{GpC})_2$ (27.8 kcal). Note also the low twist angle (about 30° , compared with 40° for GpC sequences) characterising the geometry of stacks of base pairs in CpG sequences and the high positive roll which leads to an opening of the stacks of bases towards the minor groove of the helix. The geometric parameters of CpG sequences facilitate specific molecular recognition, in particular by methylases catalysing *de novo* methylation of CpG sequences.

1.4.3 Structure of Methylated CpG Dinucleotides

In the double helix, the two methyl groups symmetrically positioned on the cytosines of the CpG sequence are located in the major groove of the helix (see Fig. 1.19) and thus form a highly distinctive motif which can be specifically recognised, in particular by (effector) proteins carrying out genetic functions.

A lot of work has been done to investigate the effects of methylation on the local structure of the CpG sequence and adjacent sequences. To a first approximation, the main consequences of the presence of two methyl groups situated close to one another and protruding into the major groove are as follows:

- increased hydrophobicity of the groove,
- establishment of hydrogen bonds between protons of the methyl groups and amino acids of the protein ligands (see the next section),

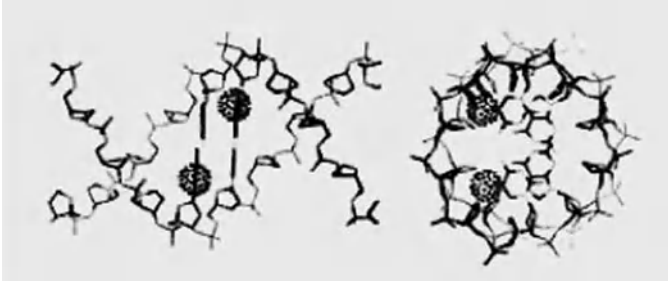


Fig. 1.19. Positioning of methyl groups in a CpG sequence symmetrically methylated on the cytosines. *Left:* View of methyl groups on the major groove side. *Right:* View in a plane perpendicular to the axis

- modified accessibility of the deep part of the groove, given the steric hindrance of the two methyl groups.

Regarding the DNA geometry, the general consensus is that methylation of a cytosine has little effect on the local structure of the helix. The main impact concerns the stability of the helix and molecular dynamics. Indeed, crystallographic studies [29] carried out on a d(ACCGCCGGCGCC) dodecamer have shown that the geometry of a duplex methylated on the central cytosine led to two new hydrogen bonds being set up between protons of the methyl group and oxygen atoms of the phosphates, with consequent stabilisation of the double helix. Moreover, NMR studies combined with molecular dynamics simulations [30] have shown that methylation of the CRE sequence (cAMP responsive element) d(GAGATGAmCGTCATCTC)₂ leads to the adoption of a BII conformation with reduced flexibility of the helix, notably in 5' adjacent sequences, inducing steric hindrance due to the methylated cytosine. These relatively modest structural and dynamic changes are enough to cause a drastic change in the interactions between the DNA and regulating proteins such as transcription factors.

1.4.4 Specific Recognition of Symmetric Methylation by Proteins

In vertebrates, there is a family of proteins that specifically recognise the symmetrically methylated CpG sequence. This family includes the protein MeCP2, already mentioned above, and also the proteins MBD1, MBD2, MBD3 and MBD4. These proteins share a methyl-CpG binding domain. This domain, located near the N-terminal region, comprises 70 amino acids [31].

From a biological standpoint, it is interesting to note that MeCP2 [32], after binding onto a methylated CpG sequence, is then able to recruit histone deacetylases. Deacetylation of the histone lysines releases positive charges and this favours electrostatic interactions between histone and DNA, thereby increasing the compaction of the chromatin and rendering it mute from a genetic point of view. When it binds to methylated CpG sequences, MeCP2

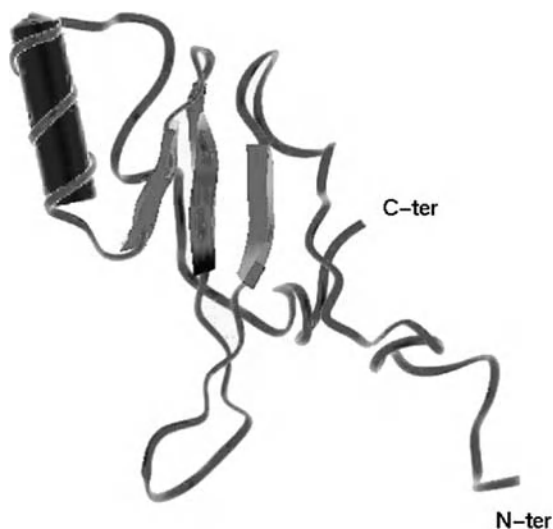


Fig. 1.20. Structure of the DNA binding domain of the protein MeCP2

acts as a key element in transcriptional repression. MeCP2 is assumed to bind onto a single methyl-CpG motif via essentially hydrophobic interactions. In fact the molecular mechanism for recognition of methyl-CpG sequences by the MeCP2 protein is not yet fully understood. However, the three-dimensional structure of the DNA binding domain is now known (see Fig. 1.20).

This domain is characterised in particular by a low level of structuring and a consequently high flexibility. This flexibility allows the protein to adapt itself to the rigidity of methylated CpG sites and to the steric hindrance caused by the presence of the two methyl groups. Mutagenesis and NMR studies [33] have shown that arginine-111, which interacts with aspartate-121, is one of the key amino acids controlling the specific binding of the protein to the methylated CpG sequence. Moreover, it would seem that the presence of sequences adjacent to the CpG site that are rich in AT base pairs favours a high-affinity binding of MeCP2 onto the CpG sequence.

In any case, through its binding on methyl-CpG sites, MeCP2 plays an essential role in the control of genetic expression. This is confirmed by the fact that mutations perturbing the binding of the protein onto methylated CpG sequences lead to the appearance of a pathology which mainly affects girls (Rett's syndrome), characterised by anomalous development of the central nervous system. This pathology, now considered to be a genetic disease affecting the X chromosome (the MeCP2 gene is carried by the X chromosome), is transmitted as a dominant character. The molecular etiology of this pathology clearly demonstrates the importance of epigenetic transcriptional repression processes.

1.5 Conclusion

The examples described in this chapter show the diversity and complexity of processes involved in regulating the expression of the genetic code. The most striking point is that DNA serves both as the physical support for genetic information and as the major regulator for reading this information. Changes in geometry, changes in topology, and changes in chemical structure all contribute in a concerted way, on the molecular scale, to the precise mechanisms regulating gene expression. Indeed, these modifications are signals triggering the mobilisation of effectors (mainly proteins) present in the environment of the DNA. These effectors are both regulatory elements (activators or repressors) and also synthesising elements carrying out the transformation of DNA code into RNA code (transcription), and then the transformation of RNA code into protein (translation).

Research over the last few years has revealed the fundamental role of epigenetic regulation of gene expression, i.e., regulation that is not directly linked to the gene sequence, and in this context, the equally fundamental role of repression processes, particularly those linked to DNA methylation and the compaction of chromatin structures (DNA plus proteins). Such processes repressing gene expression are probably the key features of development and cell differentiation. One emerging feature is that anomalies relating to these repression mechanisms may lie at the origin of, or at least large contribute to the occurrence of many major pathologies. This is the case for tumour transformation, which seems to result from an anomalous repression of genes whose function is to inhibit cell proliferation and maintain cell differentiation. Some of these genes have been clearly identified as tumour-suppressing genes.

From this point of view, it is striking to observe that the level of methylation of the genome increases steadily with the age of the individual. This observation may throw new light on the relationship between the incidence of cancer and aging. There can be no doubt that one of the great scientific challenges in biology will be to clarify the epigenetic mechanisms regulating gene expression, and beyond this, in the case of anomalies, to find ways to act and restore these regulation processes, by pharmacological means if need be.

References

1. Watson, J.D., Crick, F.H.C.: Molecular structure of nucleic acids, *Nature* **171**, 737 (1953)
2. De Santis, P., Palleschi, A., Savino, M., and Scipioni, A.: *Biochemistry* **29**, 9269 (1990)
3. Bolshoy, A., McNamara, P., Harrington, R.E., Trifonov E.N.: *Proc. Natl. Acad. Sci. USA* **88**, 2312 (1991)
4. <http://rumour.biology.gatech.edu>
5. Djuranovic, D., Hartmann, B.: Conformational characteristics and correlations in crystal structure of nucleic acid oligonucleotides: Evidence of sub-states, *J. Biomol. Struct. Dyn.* **20** (6), 1 (2003)

6. Mirau, P.A., Kearns, D.R.: Unusual proton exchange properties of Z-form poly[d(G-C)], *Proc. Natl. Acad. Sci. USA* **82**, 1594 (1985)
7. Misra, V.K., Honig, B.: The electrostatic contribution to the B to Z transition of DNA, *Biochemistry* **35**, 1115 (1996)
8. Kawaga, T.F., Howell, M.L., Tseng, K., Ho, P.S.: Effects of base substituents on the hydration of B- and Z-DNA: Correlations to the B- to Z-DNA transition, *Nucleic Acids Research* **21**, 255978 (1993)
9. Peck, L.J., Wang, J.C.: Energetics of B-to-Z transition in DNA, *Proc. Natl. Acad. Sci. USA* **80**, 6206 (1983)
10. Le Ber, P., Schwaller, M.A., Auclair, C.: Effect of intercalative binding compared to external binding on Z/B equilibrium of poly-d(Gme5C) using fluorescent oxazolopyridocarbazoles as probes, *J. Mol. Recognit.* **2** (4), 152–157 (1989)
11. Ha, S.C., Lowenhaupt, K., Rich, A., Kim, Y.G., Kim, K.K.: Crystal structure of a junction between B-DNA and Z-DNA reveals two extruded bases, *Nature* **437**, 1183 (2005)
12. Schwartz, T., Rould, M.A., Lowenhaupt, K., Herbert, A., Rich, A.: Crystal structure of the Za domain of the human editing enzyme ADAR1 bound to left-handed Z-DNA, *Science* **284**, 1841–1845 (1999)
13. Osipiuk, J., Skarina, T., Edwards, A., Savchenko, A., Joachimiak, A.: 1.55 Å crystal structure of putative Z-DNA binding protein AF2008 from *Archaeoglobus fulgidus*, ACA05 W0243
14. Witting, B., Wolff, S., Dorbic, T., Vahrson, W., Rich, A.: Transcription of human *C-MYC* in permeabilized nuclei is associated with formation of Z-DNA in three discrete regions of the gene, *EMBO J.* **11**, 4653 (1992)
15. Kwon, J.A., Rich, A.: Biological function of the vaccinia virus Z-DNA-binding protein E3L: Gene transactivation and antiapoptotic activity in HeLa cells, *Proc. Natl. Acad. Sci. USA* **102**, 12759 (2005)
16. Champ, P.C., Maurice, S., Vargason, J.M., Camp, T., Ho, P.S.: Distributions of Z-DNA and nuclear factor I in human chromosome 22: A model for coupled transcriptional regulation, *Nucleic Acids Research* **32**, 6501 (2004)
17. Hamiche, A., Carot, V., Alilat, M., De Lucia, F., O'Donohue, M.F., Revet, B., Prunell, A.: Interaction of the histone (H3-H4)₂ tetramer of the nucleosome with positively supercoiled DNA minicircles. Potential flipping of the protein from a left- to a right-handed superhelical form, *Proc. Natl. Acad. Sci. USA* **93**, 7588 (1996)
18. Wang, J.C., Jacobsen, J.H., Saucier, J.-M.: Physiochemical studies on interactions between DNA and RNA polymerase. Unwinding of the DNA helix by *Escherichia coli* RNA polymerase, *Nucleic Acids Res.* **4**, 1225 (1977)
19. Liu, L.F., Wang, J.C.: Supercoiling of the DNA template during transcription, *Proc. Natl. Acad. Sci. USA* **84**, 7024–7027 (1987)
20. Lilley, D.M.: The inverted repeat as a recognizable structural feature in supercoiled DNA molecules, *Proc. Natl. Acad. Sci. USA* **77**, 6468–6472 (1980)
21. Ward, G.K., McKenzie, R., Zannis-Hadjopoulos, M., Price, G.B.: The dynamic distribution and quantification of DNA cruciforms in eukaryotic nuclei, *Exper. Cell Res.* **188**, 235 (1990)
22. Baylin S.B., Herman, J.G.: DNA hypermethylation in tumorigenesis: Epigenetics joins genetics, *Trends Genet.* **16**, 168 (2000)

23. Mizuno, S., Chijiwa, T., Okamura, T., Akashi, K., Fukumaki, Y., Niho, Y., Sasaki, H.: Expression of DNA methyltransferases *DNMT1*, *3A*, and *3B* in normal hematopoiesis and in acute and chronic myelogenous leukaemia, *Blood* **97**, 1172 (2001)
24. Bird, A.P.: CpG-rich islands and the function of DNA methylation, *Nature* **321** (6067), 209 (1986)
25. Pollack, Y., Kasir, J., Shemer, R., Metzger, S., Szyf, M.: Methylation pattern of mouse mitochondrial DNA, *Nucleic Acids Res.* **12** (12), 4811 (1984)
26. Gardiner-Garden, M., Frommer, M.: CpG islands in vertebrate genomes, *J. Mol. Biol.* **196** (2), 261 (1987)
27. Fu, Y.H., Kuhl, D.P., Pizzuti, A., Pieretti, M., Sutcliffe, J.S., Richards, S., et al.: Variation of the CGG repeat at the fragile X site results in genetic instability: Resolution of the Sherman paradox, *Cell* **67**, 1047 (1991)
28. Sukhodub, L.F., Yanson, I.K.: Mass-spectrometric studies of binding energies for nitrogen bases of nucleic acids in vacuo, *Nature* **264** (5583), 245 (1976)
29. Mayer-Jung, C., Moras, D., Timsit, Y.: Effect of cytosine methylation on DNA–DNA recognition at CpG steps, *J. Mol. Biol.* **270** (3), 328 (1997)
30. Derreumaux, S., Chaoui, M., Tevanian, G., Fermandjian, S.: Impact of CpG methylation on structure, dynamics and solvation of cAMP DNA responsive element, *Nucleic Acids Research* **29** (11), 2314 (2001)
31. Nan, X., Meehan, R.R., Bird, A.: Dissection of the methyl-CpG binding domain from the chromosomal protein MeCP2, *Nucleic Acids Res.* **21**, 4886 (1993)
32. Lewis, J.D., Meehan, R.R., Henzel, W.J., Maurer-Fogy, I., Jeppesen, P., Klein, F., Bird, A.: Purification, sequence, and cellular localization of a novel chromosomal protein that binds to methylated DNA, *Cell* **69**, 905 (1992)
33. Free, A., Wakefield, R.I., Smith, B.O., Dryden, D.T., Barlow, P.N., Bird, A.P.: DNA recognition by the methyl-CpG binding domain of MeCP2, *J. Biol. Chem.* **276**, 3353 (2001)

Protein–Lipid Assembly and Biomimetic Nanostructures

A. Girard-Egrot, L. Blum, R. Richter, and A. Brisson

Molecular and supramolecular understanding of the architecture of biological systems, and in particular membranes, can provide an extraordinary source of inspiration for making ‘intelligent’ nanostructures, based on the self-assembly properties of biological molecules. Bioelectronic interfacing between living and inert matter constitutes one of the most promising areas of development in nanobiotechnology.

The present chapter is particularly concerned with the description of biological membranes and the self-association properties of the molecules making them up. These provide the basis for a natural and spontaneous formation of structure, which can be used to develop biomimetic membranes and a wide range of nanostructured protein–lipid structures with undeniable scope for application in the field of nanobiotechnology, e.g., lipid nanoparticles, encapsulation and delivery of medicines, targeting, molecular sorting, surface functionalisation, nanobiosensors, etc.

2.1 Introduction: Biological Membranes

Membranes play a central role in the life of a cell. Artists’ impressions of cells showing the current state of our understanding as a result of an enormous body of investigation reveal the presence of many compartments within any given cell, while the cell itself forms a compartment in its own right. One may consider compartmentalisation as a strategy adopted by the cell to share out its various tasks into well-defined regions, each of which disposes of specific means and elements. Cell compartments, or organelles, have specific functions: the nucleus is the place where the DNA is stored and rRNA, tRNA, and mRNA are synthesised; the endoplasmic reticulum is where proteins and lipids are synthesised; the mitochondria provide the energy supply for the cell, and so on.

All cell compartments share the fact of being surrounded by a common structure, the biological membrane (see Fig. 2.1). This membrane is a highly

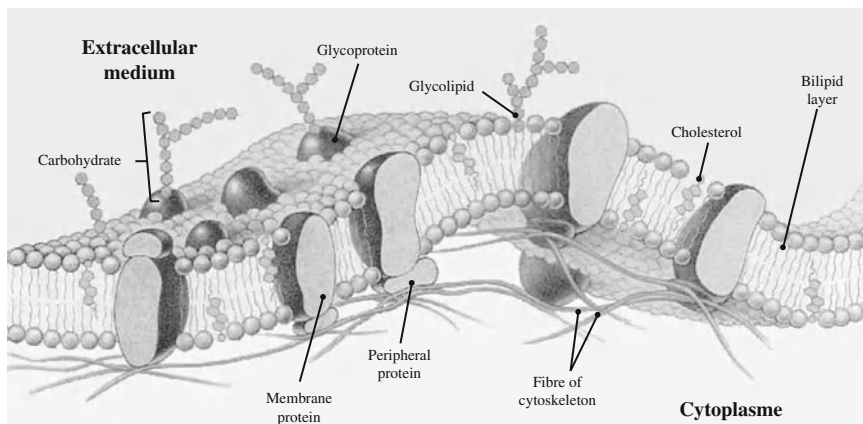


Fig. 2.1. General view of a biological membrane. This is the plasma membrane separating the inside from the outside of the cell. Biological membranes have a functional protein–lipid architecture comprising a lipid bilayer together with integrated or associated proteins. While it is the lipid bilayer which gives the membrane its structural properties, the proteins are essential to the way the membrane carries out its functions in the cell. In particular, they deal with the transfer of molecules across the lipid bilayer, and they also catalyse reactions occurring specifically at the cell surface. Furthermore, they ensure the transduction of signals beyond the membrane. Apart from its role of compartmentalising the cell, the biological membrane is thus the scene of many recognition and transduction phenomena, not to mention the exchange of energy, matter, and information between the interior and the exterior of the cell. Taken from [1]

complex supramolecular structure, mainly composed of lipids arranged together to form a lipid bilayer, transmembrane proteins which may contain glycosidic residues, and peripheral membrane proteins associated with the membrane either directly or through some kind of anchoring system.

Biological membranes constitute excellent natural models for structuring and organisation on the molecular level. As we shall see, this organisation builds upon the self-association properties of biological macromolecules. Using such properties, biomimetic membranes, i.e., self-assembled entities corresponding to organised supramolecular arrangements, can be reconstituted *in vitro*.

There are many potential applications of biomimetic membranes to nanobiotechnology. Examples are the encapsulation and controlled release of active ingredients, drug targeting, gene therapy, composite structures, fabrication of lipid nanoparticles for detection applications (quantum dots, magnetic nanoparticles used in medical imaging), reconstitution of membrane proteins, functionalisation and structuring of surfaces, and as a substrate for the 2D

Table 2.1. Main classes of lipid membranes

Class	Name
Glycerophospholipids (or phosphoglycerides)	1,2 diacylphosphoglycerides Plasmalogens
Glyceroglycolipids	1,2 diacylglycoglycerides
Sphingolipids	Phosphosphingolipids Glycosphingolipids
Sterols	Cholesterol (animal kingdom) Stigmasterol (plant kingdom) Ergosterol (fungi)

crystallisation of proteins. They are also used in nanobiosensors, processors (chips), and nanocarriers for delivering medicines.

2.2 Lipid Membranes: Structure and Properties

Lipids play a dual role in biological membranes, being both structural and functional. They have the property of self-assembling in an aqueous medium to form double layers, or bilayers, which constitute the basic structure of the membrane. Apart from this, some lipid membranes have important functional properties. In particular, they are involved in adhesion mechanisms and cell growth control, as well as platelet activation. They can play the role of toxin receptors or they may be precursors of intracellular second messengers. Lipid membranes display a wide range of different structures. They are classified into four many categories, each one corresponding to a family of compounds.

2.2.1 The Main Classes of Lipid Membranes

The main classes of lipid membranes are (see Table 2.1) [2]:

- glycerophospholipids,
- glyceroglycolipids,
- sphingolipids,
- sterols.

Glycerophospholipids

The glycerophospholipids, also called phosphoglycerides, include the 1,2 diacyl phosphoglycerides and the plasmalogens.

The 1,2 diacyl phosphoglycerides are the most abundant phospholipids in most cell membranes. They all derive from phosphatidic acid, obtained by

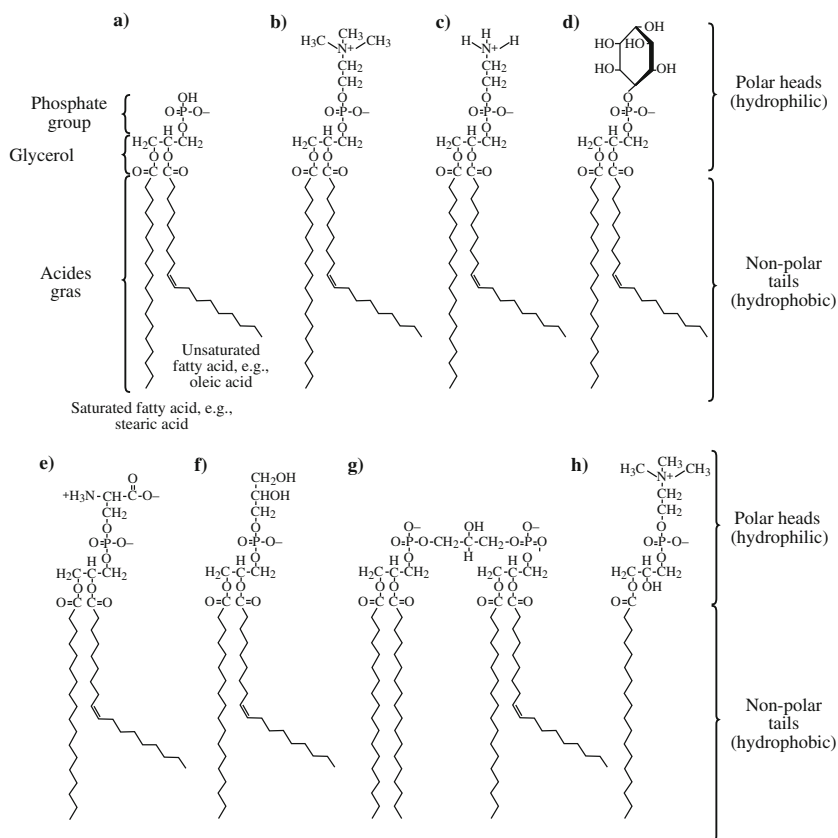


Fig. 2.2. Structure of the main diacyl phosphoglycerides (phospholipids). (a) Phosphatidic acid. (b) Phosphatidylcholine. (c) Phosphatidylethanolamine. (d) Phosphatidylinositol. (e) Phosphatidylserine. (f) Phosphatidylglycerol. (g) Diphosphatidylglycerol. (h) Lysophosphatidylcholine. Structures have been drawn in such a way as to bring out the amphipathic nature of the molecules. The polar end carrying the hydrophilic groups, often called the polar head, is at the top of the structure, while the non-polar end carrying the hydrophobic hydrocarbon chains, often called the hydrophobic tail, points downwards

esterification of glycerol, in position *sn*-1 and *sn*-2 by a long-chain fatty acid, and in position *sn*-3 by phosphoric acid. Phosphatidic acid is the simplest phosphoglyceride (see Fig. 2.2a). It is not strictly speaking a membrane lipid, but rather an intermediate in the biosynthesis of other phospholipids.

The different classes of phospholipids are defined in terms of the substituent bound to the phosphate group of the phosphatidic acid. This can be choline (Fig. 2.2b), ethanolamine (Fig. 2.2c), inositol (Fig. 2.2d), serine

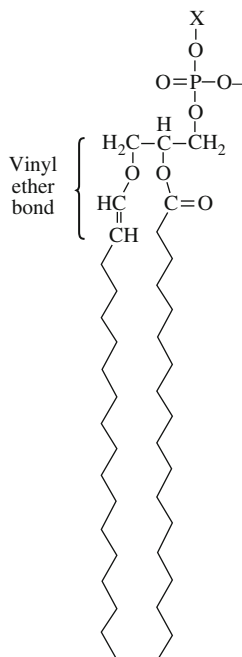


Fig. 2.3. Structure of a plasmalogen. The substituent X is usually ethanolamine, but sometimes choline

(Fig. 2.2e), or glycerol (Fig. 2.2f), bound in position C3 of the glycerol by a phosphodiester bond. These are thus called phosphatidylcholines, phosphatidylethanolamines, phosphatidylinositols, phosphatidylserines, and phosphatidylglycerols, respectively. The cardiolipids, or diphosphatidylglycerols (Fig. 2.2g), are particular phospholipids, comprising two molecules of phosphatidic acid joined together by a glycerol molecule. The phosphate group carries a neutral pH (pH = 7.0) negative charge. The overall charge of the phospholipid thus depends on the charge carried by its substituent. At physiological pH, some phospholipids are anionic (negatively charged), such as the phosphatidylinositols, phosphatidylserines, phosphatidylglycerols, or zwitterions (electrically neutral), such as the phosphatidylcholines or phosphatidylethanolamines. The constitutive fatty acids of the glycerophospholipids can have a wide range of different structures. This is why each class of phospholipids constitutes a family of compounds. In general, the glycerophospholipids contain a saturated fatty chain at position C1 and an unsaturated fatty chain at position C2. The length of the hydrocarbon chain can be anything between 14 and 24 carbon atoms. The commonest lengths are 16, 18 and 20 carbon atoms, but there are many exceptions.

The plasmalogens are phosphoglycerides in which one of the hydrocarbon chains is associated with glycerol by a vinyl ether bond. These are glyceride ethers. The ether bond is usually located in position C1 (see Fig. 2.3). The plasmalogens are common in the cardiac tissues of vertebrates, the peripheral nervous system, and muscle.

Glyceroglycolipids

The 1,2 diacyl glycolipids, like the diacyl phosphoglycerides, derive from glycerol by esterification by a fatty acid at positions *sn*-1 and *sn*-2. In contrast, however, position *sn*-3 of the glycerol is bound by a glycosidic bond to glucidic structures such as galactose (see Fig. 2.4). These glycolipids are very abundant in photosynthetic membranes of algae and plants, but rarely encountered in the animal kingdom.

Sphingolipids

The sphingolipids make up the second large family of membrane lipids. They contain sphingosine, a long-chain aliphatic amino alcohol. In sphingolipids, sphingosine plays the role of glycerol in glycerophospholipids. Amidification of the amine function $-\text{NH}_2$ of sphingosine by a long-chain fatty acid with variable degree of unsaturation produces a ceramide, a basic component of all the sphingolipids (see Fig. 2.5). The latter are obtained by addition of a substituent on the primary alcohol function of the ceramide.

The sphingolipid class includes the phosphosphingolipids and the glycosphingolipids. The phosphosphingolipids have the same types of polar groups as the glycerophospholipids. As an example, sphingomyelin, one of the main constituents of plasma membranes in most animal cells and in the myelin sheath enclosing axons and some neurons, is a phosphosphingolipid with a phosphocholine group substituted on the primary alcohol of the ceramide (see Fig. 2.5a). In the sphingomyelins, the predominant fatty acid is the fatty acid with a hydrocarbon chain of 24 carbon atoms, saturated or otherwise. There are other sphingolipids with a phosphoethanolamine, phosphoinositol or phosphoglycerol group substituted on the ceramide. The phosphosphingolipids are classified as phospholipids along with the glycerophospholipids.

The glycosphingolipids are derivatives of ceramide associated with sugars via a glycosidic bond on the primary alcohol of the ceramide. They contain no phosphate and are thus grouped together in the glycolipid class. The cerebrosides are neutral glycolipids comprising a single glucidic residue such as galactose or glucose (see Fig. 2.5b). They are characteristically found in plasma membranes of nerve tissue. The gangliosides, the most complex of the sphingolipids, are anionic glycolipids involving one or more sialic acid molecules (*N*-acetylneuraminic acid) carrying a negative charge at the physiological pH, branching off from the oligosaccharide chain. They constitute 6% of membrane lipids in the grey matter of the human brain and are present in smaller amounts in the membranes of animal neuronal tissues.

The glycosphingolipids are generally minor components of membranes. They are all located in the outer leaflet of the plasma membrane bilayer. The hydrocarbon chains of their ceramide group are buried within the hydrophobic core of the membrane, while their glucidic part projects outside the cell. For this reason they are involved in many recognition functions. In addition, some viruses use their glucidic part as an anchor before infection.

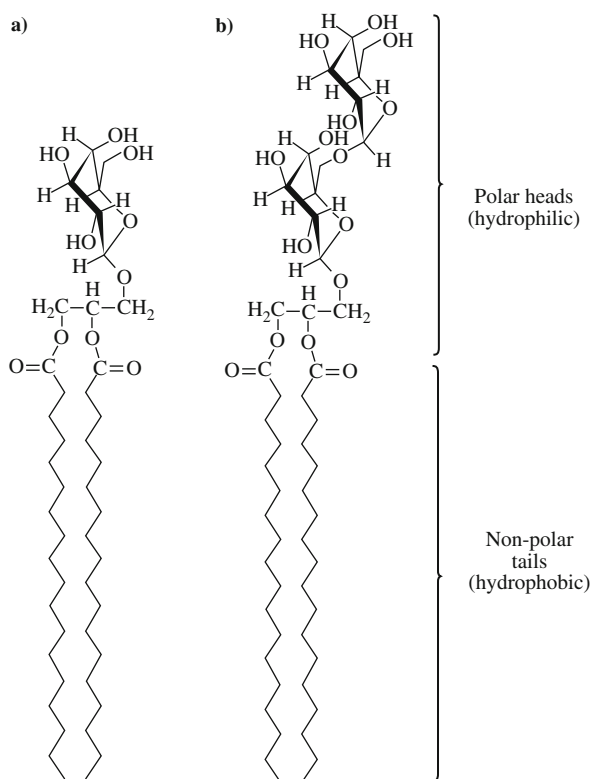


Fig. 2.4. Structure of several diacyl glycolipids (glycolipids). (a) Monogalactosyldiacylglycerol (MGDG). (b) Digalactosyldiacylglycerol (DGDG)

Sterols

Sterols are lipids that are well represented in biological membranes. Their characteristic structure is the steroid nucleus consisting of four carbon rings A, B, C, and D, fused together in a relatively rigid and almost planar configuration as shown in Fig. 2.6. Cholesterol constitutes some 30% of all lipids found in plasma membranes of animal cells (see Fig. 2.6a). It is also found in the membranes of some organelles in animal cells. The only polar group on this molecule is the hydroxyl group -OH carried by carbon 3 of the steroid nucleus. Stigmasterol is found in the plant kingdom (see Fig. 2.6b) and ergosterol is a constituent of fungal cell membranes (see Fig. 2.6c). With a few very rare exceptions, bacteria do not have sterols.

Cholesterol integrates into the bilayer with its OH group at the interface as shown in Fig. 2.6d. Due to its small size, it only occupies one leaflet of the bilayer.

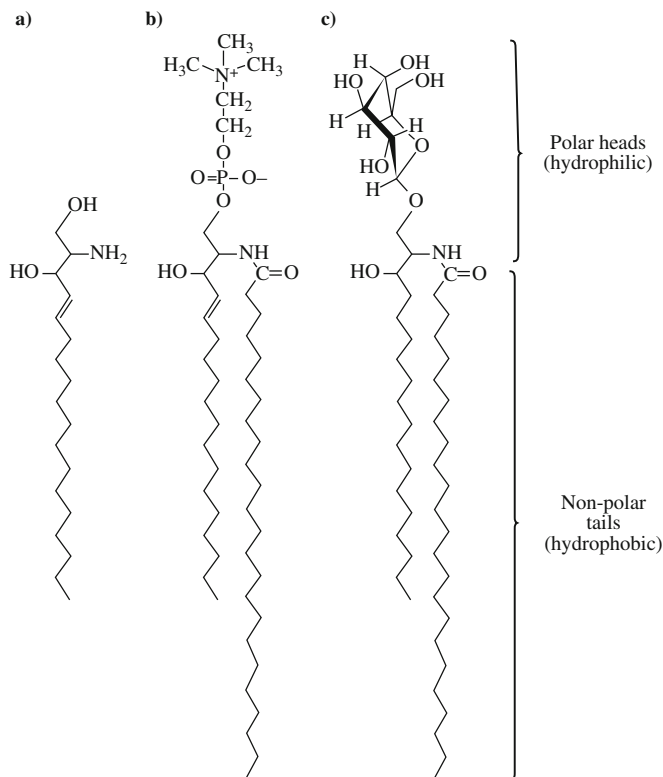


Fig. 2.5. Structure of several sphingolipids. (a) Sphingosine. (b) Sphingomyelin (phosphosphingolipid). (c) Galactocerebroside (glycosphingolipid)

Minor Components

Free fatty acids and lysophospholipids, phospholipids formed from a single hydrocarbon chain (see Fig. 2.2h) are present in biological membranes, but in extremely small amounts.

2.2.2 Self-Assembly

The characteristic feature of all membrane lipids (phosphoglycerides, glyco-glycerides, sphingolipids) is their amphipathic character. They are amphiphilic molecules with a hydrophilic polar end and a hydrophobic non-polar end located in two different regions of the molecular space. The polar part consists of the substituted phosphoglycerol or the glycosylated glycerol in the case of the glycerides, or the phosphorylated or glycosylated group substituting for the sphingosine in the case of the sphingolipids. The long hydrocarbon chains represent the apolar part. Amphipathic lipids are often represented schematically in the form of a molecule with a polar head and one or two hydrophobic

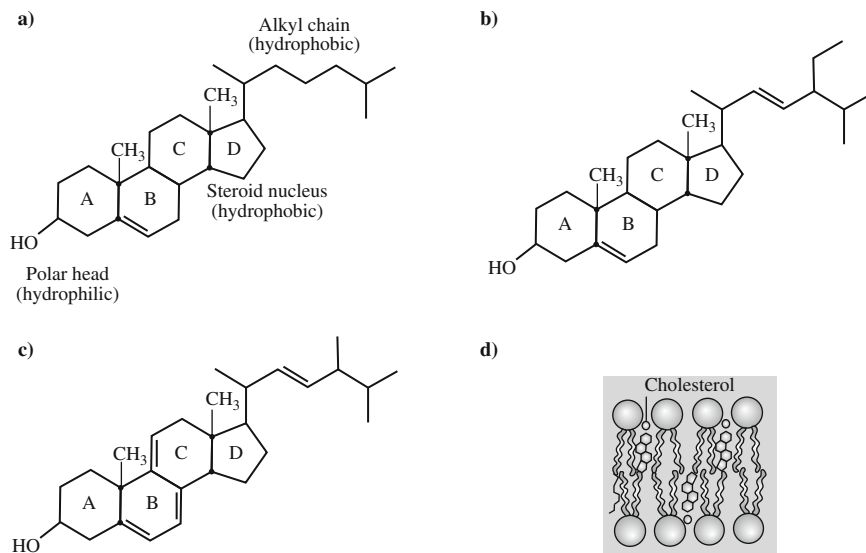


Fig. 2.6. (a)–(c) Structure of some sterols. The sterols are derived from the steroid nucleus, which comprises four fused carbon rings A, B, C, and D. (a) Cholesterol. (b) Stigmasterol, a common sterol in plant cell membranes. (c) Ergosterol, a common sterol in fungal cell membranes. (d) Relative sizes of a phospholipid and a sterol inserted in a membrane

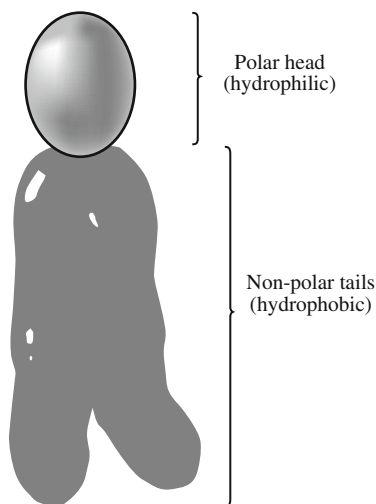


Fig. 2.7. Profile of an amphipathic lipid molecule. The molecule sketched here has two non-polar tails

tails (see Fig. 2.7). In the case of the glycolipids and sphingoglycolipids, the bigger the glucidic fragment, the more polar the lipid.

The apolar part of the lipids severely limits their solubility in water in the form of monomers. Membrane lipids are amphipathic lipids that are

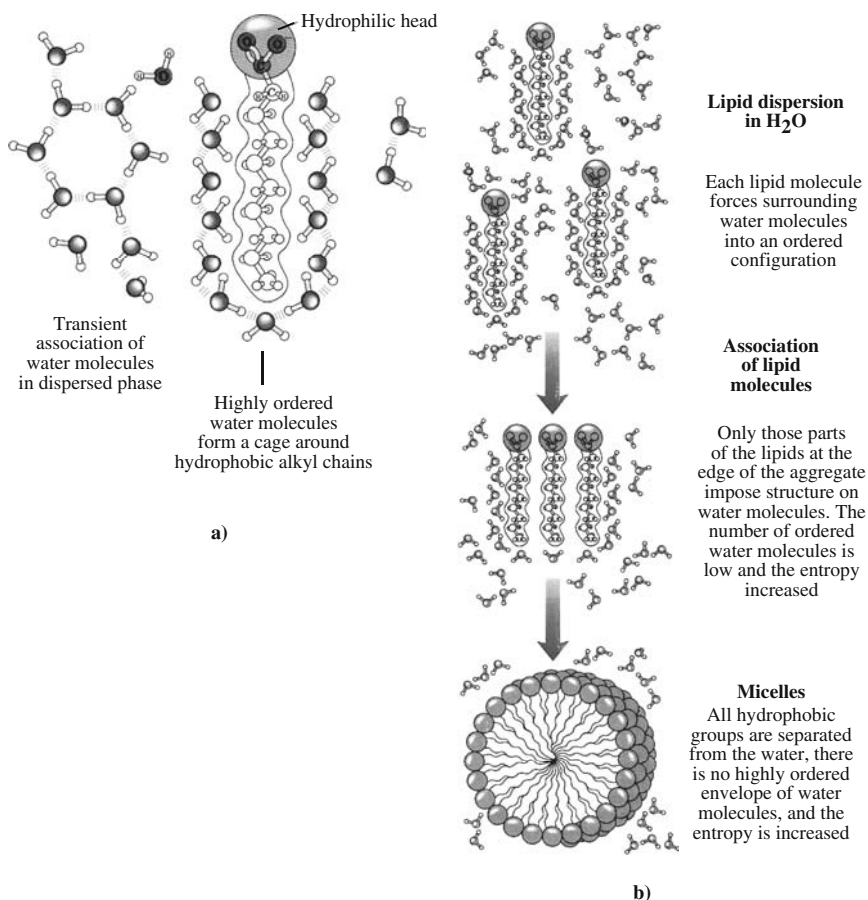


Fig. 2.8. Hydrophobic effect explaining the aggregation of lipid molecules. The example shown concerns a long-chain fatty acid. From Lehninger et al. [3]. Hydrophobic interactions set up between the lipid molecules supply the thermodynamic force required to form and maintain the aggregate

practically insoluble in water. When they are mixed with water, the two ends of the molecule tend to behave in opposite ways: the hydrophilic polar head interacts favourably with the aqueous medium and tends to dissolve, while the hydrocarbon chains avoid contact with the water as far as they can. For this reason, amphipathic lipids form aggregates in a separate phase from their aqueous environment. The shape of these structures depends on the relative sizes of the hydrophilic and hydrophobic parts (the so-called hydrophilic/hydrophobic balance), while their cohesion will be governed by the hydrophobic effect, the key to understanding the self-association properties of amphiphilic molecules in water.

The hydrophobic effect corresponds to the propensity of hydrophobic groups to clump together owing to their rejection by the aqueous solvent. It can be explained in thermodynamic terms (see Fig. 2.8). Indeed, hydrocarbon chains are non-wettable compounds, incapable of creating interactions with water molecules. When inserted into an array of free water molecules, those molecules in the immediate vicinity, enclosing the non-polar chain and bound together by hydrogen bonds, are highly restricted in the orientations they can adopt, and this leads to an enveloping structure of highly ordered water molecules around each hydrocarbon chain. This structure is called a cage. When the non-polar chains come together in order to present the smallest possible hydrophobic surface to the solvent, the dismemberment of the cage structure is accompanied by an increase in the entropy of the water molecules which leave this ordered envelope to return to the bulk of the solvent. The increased translational entropy of the water molecules greatly exceeds the drop in entropy associated with the gathering of the hydrocarbon chains and the formation of the aggregate.

When membrane lipids aggregate, the hydrophobic tails come together in such a way as to minimise their area of contact with the water, while the polar headgroups adopt positions which allow them to increase their interaction with the aqueous solvent to a maximum. Within the aggregate, the hydrocarbon chains interact together via van der Waals forces. The whole set of forces holding the non-polar regions of the molecules together are called hydrophobic interactions. It should be mentioned that the force arising from these interactions is not in any sense due to a mutual attraction between the apolar chains. Rather it results from the fact that the arrangement achieves optimal thermodynamic stability, minimising the reduction in entropy due to the arrangement of the water molecules around the hydrophobic regions of the amphiphilic molecules. These hydrophobic interactions are in fact the cohesive forces binding lipid aggregates. Although these interactions involve little energy, it is the fact that they are so numerous within the aggregate that provides the cohesion of the different structures adopted by membrane lipids.

The message here is thus that amphiphilic molecules first associate together, then self-order, under the effect of attractive/repulsive forces affecting their hydrophilic/hydrophobic parts. In membrane lipids, the structure or shape of the consequent supramolecular arrangements depends on the amphiphilic balance of the molecule.

2.2.3 Lipid Polymorphism

Lipid polymorphism results from the propensity of membrane lipids, once isolated from their biological context and suspended in water, to form supramolecular complexes with very varied structural configurations. These ordered arrangements will of course depend on the nature of the lipid – more precisely, on the relative steric properties of the hydrophilic and hydrophobic parts – and for a given lipid, experimental conditions such as solvent temperature, lipid

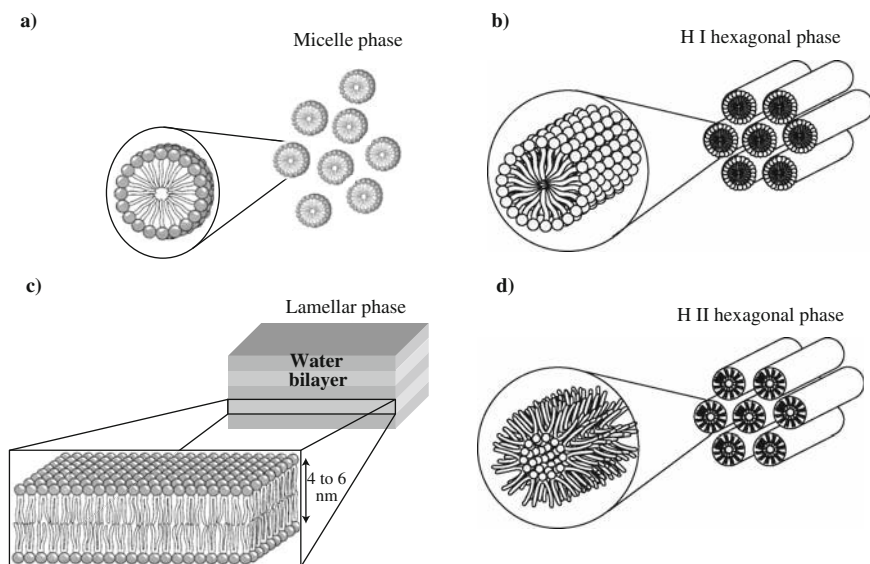


Fig. 2.9. Main phases and structural arrangements adopted by lipids suspended in water. (a) In a spherical micelle, the structure adopted by the salts of fatty acids (soaps), the hydrophobic chains are trapped in the core of the sphere. There is almost no water in the hydrophobic center of the micelle. All the water is forced to remain outside. The micelles themselves are randomly distributed through the aqueous phase. (b) In a cylindrical micelle, the structure adopted by the lysophospholipids, the polar part of the lipids coat the surface of the cylinder and the hydrocarbon chains are buried at its center. The cylinders are arranged in a hexagonal lattice in the aqueous phase. This corresponds to the H_I hexagonal phase. (a) and (b) are referred to as normal micelles. (c) In a bilayer, all the hydrophobic side chains apart from those at the edge of the leaflet are protected from interaction with the water. In a lamellar phase, the planar lipid bilayers are separated by layers of water with thicknesses depending on the level of hydration. (d) In the reverse micelle, the polar parts of the lipids coat the inner surface of a hollow cylinder surrounding the aqueous phase. The hydrocarbon chains project out of the cylinder. In the type H_{II} hexagonal phase, the cylinders also arrange themselves into a hexagonal lattice

concentration, pH, presence or absence of dissolved salts (ionic strength), and pressure. Depending on the nature of the lipid and the chosen experimental conditions, one encounters three types of lipid aggregate:

1. *Micelles*. These relatively small, closed structures can be either spherical or cylindrical. They involve between ten and a few hundred molecules in the case of spherical micelles, and a few thousand molecules in the case of cylindrical micelles, these being arranged in such a way that the hydrophobic regions come together inside the structure, thereby excluding the water, and in such a way that the hydrophilic headgroups are located

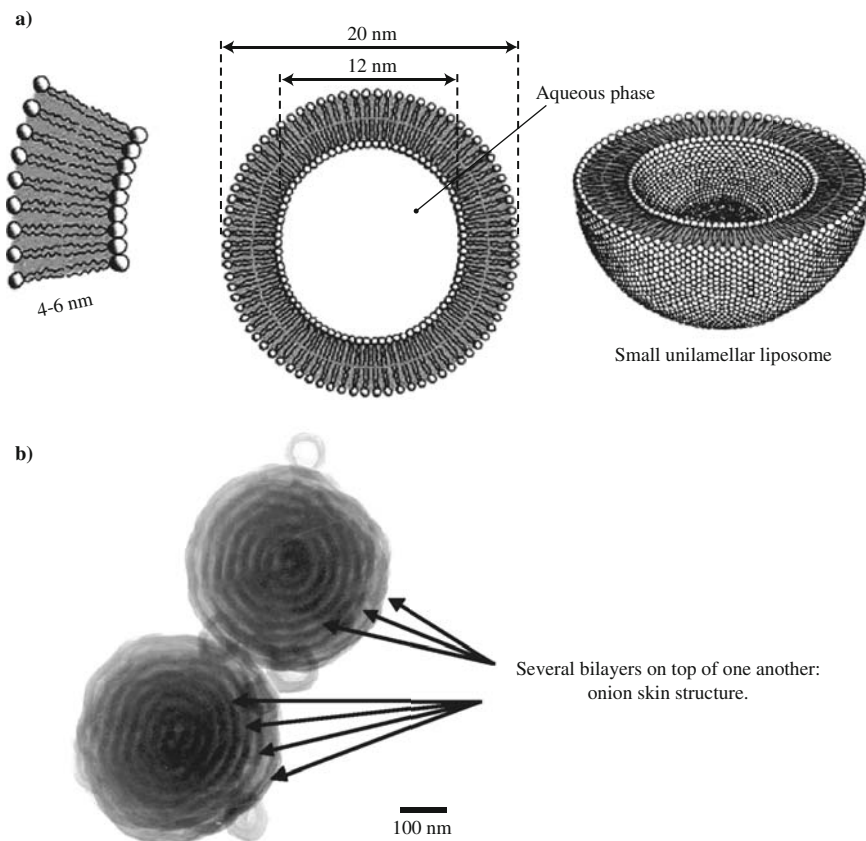


Fig. 2.10. Structure of a lipid vesicle, or liposome. When an extended 2D bilayer closes up on itself, it forms a 3D vesicle, enclosing an aqueous cavity. The phospholipids can form (a) unilamellar vesicles, comprising a single bilayer, or (b) multilamellar vesicles. (a) Schematic view. (b) ‘Onion skin’ structure observed by transmission electron microscope. Taken from [4]. Reproduced with the kind permission of Wiley-VCH Verlag (copyright)

at the surface, in contact with the water (see Figs. 2.9a and b). These two types of micelle, spherical and cylindrical, are formed by molecules with a single hydrophobic tail, such as fatty acids (or their salts, i.e., soaps), the lysophospholipids (phospholipids with a single hydrocarbon chain), or detergents.

2. *Lipid Bilayers.* This is the main supramolecular complex formed by the membrane phospholipids and glycolipids (see Fig. 2.9c). Owing to the presence of two hydrocarbon chains, these lipids cannot easily aggregate into the micelle structure, but prefer to arrange themselves into bilayers, in which two lipid layers join together to form a 2D structure. Lipid bilayers

constitute the basic architecture of all biological membranes, including the plasma membrane and the intracellular membranes of the organelles in eukaryotic cells. The amphipathic lipid molecules making up these bilayers have a quite remarkable orientation: the hydrophobic tails are directed toward the interior of the bilayer, while the hydrophilic headgroups maintain contact with the water on either face of the bilayer. The positive and negative charges of the bilayer components, to which the polar heads of the phospholipids (phosphoglycerides and phosphosphingolipids) contribute, give each leaflet of the membrane an ionised surface. Unlike micelles, which are always very small, lipid bilayers can cover large areas (10^8 nm^2 and more). Since it is energetically unfavourable for the edges of a bilayer to be exposed to the aqueous solution, extended bilayers tend to close themselves up to form hollow spheres, called vesicles or liposomes (see Fig. 2.10). When it forms such a vesicle, each layer disposes of its hydrophobic peripheral region, and the structure thereby achieves maximal stability in the aqueous surroundings. Phospholipids can form unilamellar vesicles, comprising a single bilayer, or multilamellar vesicles. The latter are reminiscent of the superposed layers found in the common onion. The nature and stability of such structures depends to a large extent on their lipid composition. In the lab, synthetic unilamellar vesicles of well-defined and variable sizes can be synthesised with a good yield. There are many applications. Section 2.3.1 is devoted to the preparation and study of these biomimetic entities.

3. *Reverse Micelles*. This is the third type of aggregate that can be formed by some membrane lipids, in particular the phosphatidylethanolamines. In this molecular arrangement, the polar heads of the lipids line a hollow cylinder filled with the aqueous phase, the hydrocarbon chains being forced outside the structure as shown in Fig. 2.9d.

The various phases of lipid polymorphism are distinguished by the different structural associations obtained, i.e., bilayers, normal micelles, or reverse micelles, but also by the way these structures arrange themselves in the aqueous phase. For example, lipids forming bilayers give rise to lamellar phases. In the lamellar arrangement, the lipid bilayers are separated by layers of water of well-defined thicknesses that depend on the amount of water in the system (see Fig. 2.9c). This type of phase is usually obtained when phospholipids such as phosphatidylcholines or sphingomyelins are put in suspensions at high concentrations, i.e., with lipid/water ratios of the order of 50/50. Under such conditions, the vesicles formed by the lipid bilayers (such as those shown in Fig. 2.10) are flattened by the reduced amount of water in the mixture, and the lipids, while still maintaining their bilayer organisation, are located within lamellae which can be treated as flat and of infinite area compared with their thickness. In type H_I or H_{II} hexagonal phases such as are formed by the ordering of normal or reverse cylindrical micelles, respectively, the lipid cylinders

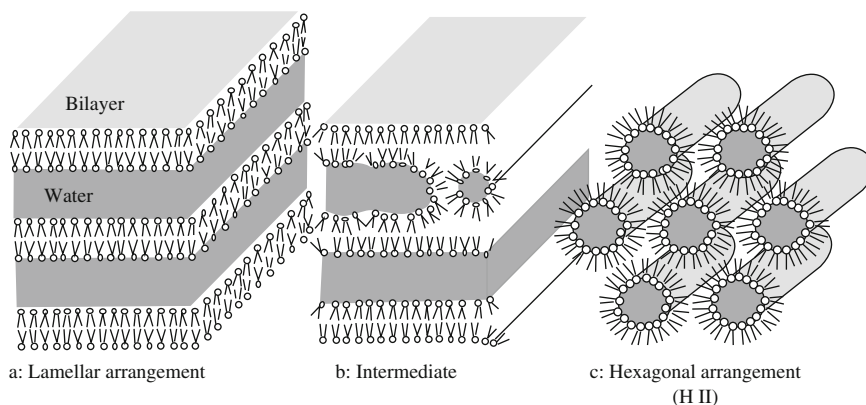


Fig. 2.11. Phase transition from the lamellar to the H_{II} hexagonal phase. This transition is favoured by decreasing the amount of water (increasing the lipid concentration) or raising the temperature of the system. When the lipid concentration increases, the thickness of the water layer decreases, the bilayers come closer together, and there is closer contact between bilayers. The outer monolayers in contact with two distinct bilayers sometimes roll up around one another, trapping the water that separates them. One then obtains a mixed intermediate phase. If the lipid concentration increases further, one obtains a hexagonal lipid arrangement of type H_{II} . In the same way, when the bilayers come into sufficiently close contact, the transition can be induced by raising the temperature. From Shechter [2]

arrange themselves in a regular hexagonal lattice as shown in Figs. 2.9b and 2.9d.

The way in which the molecules of a given lipid arrange themselves in the lipid/water mixture depends on the experimental conditions, in particular the temperature and the lipid concentration. A change of conditions can induce a phase transition. For some lipids, the transition from the lamellar phase to the H_{II} hexagonal phase can be favoured by reducing the water concentration (or increasing the lipid concentration) or raising the temperature (see Fig. 2.11). This transition involves overcoming an energy barrier, and the energy can be supplied by thermal excitation. The temperature ranges and water concentrations at which such phase transitions can take place depend on the kind of lipid. The greater the propensity of the lipid to adopt an H_{II} arrangement, the lower will be the transition temperature and the higher can be the water concentration at which the transition can take place.

Even if the structural arrangements obtained when membrane lipids are put in suspension are not always physiological, the *in vitro* study of the different phases adopted by membrane lipids allows one to draw conclusions concerning their propensities to adopt arrangements other than lamellar structures in biological membranes. An arrangement related to the one found in the H_{II} hexagonal phase may be relevant in the mechanisms occurring during membrane fusion.

Table 2.2. Structural organisation of different membrane lipids at 37°C. L lamellar phase, H_{II} corresponding hexagonal phase (reverse micelle), H_I corresponding hexagonal phase (normal cylindrical micelle). Although the difference between lysophosphatidylcholine (arrangement H_I) and phosphatidylcholine (lamellar arrangement) is clear, because just one hydrocarbon chain is lacking in the first, it is not immediately obvious in the case of phosphatidylcholine (lamellar arrangement) and phosphatidylethanolamine (H_{II} arrangement), which are both phospholipids. The difference in shape is explained by the different degrees of hydration of their polar heads. Indeed, phosphatidylcholine is hydrated by about twenty water molecules. The lipid has a cylindrical shape, considering the size of the hydrated polar head. But the polar region of phosphatidylethanolamine is only hydrated by 5 water molecules and this lipid therefore has a conical shape. From Shechter [2]

Lipid	Phase
Phosphatidylcholine	L
Sphingomyelin	L
Phosphatidylethanolamine	H _{II}
Phosphatidylserine	L
Phosphatidylglycerol	L
Cardiolipid	L
Monogalactosyldiglyceride	H _{II}
Lysophosphatidylcholine	H _I
Decreased unsaturation	L → H _{II}
Dehydration	L → H _{II}
Increased T	L → H _{II}
+ cholesterol	L → H _{II}

2.2.4 Lipid Shapes

A given lipid can form different phases under different physicochemical conditions, but under physiological conditions, viz., 37°C, pH 7, 150 mM NaCl, membrane lipids adopt a specific organisation. Table 2.2 shows the preferred arrangement adopted by different membrane lipids at the physiological temperature. Note in particular that the monogalactosyldiglycerides, common lipids in photosynthetic membranes, and the phosphatidylethanolamines, common in many biological membranes, adopt an H_{II} arrangement.

One idea put forward to try to explain the different types of lipid organisation refers to the shape of the lipid molecule, which depends on the relative spatial hindrance of its polar and hydrophobic parts (see Fig. 2.12). Hence, the formation of a normal micelle is favoured when the cross-section of the polar head is broader than that of the hydrocarbon side chains, as happens for free fatty acids or lysophospholipids. Under these conditions, the shape of the lipid fits into a cone (an inverted cone), and after aggregation, these lipids form a micelle phase (spherical micelles) or a type H_I hexagonal phase (cylindrical micelles) as shown in Fig. 2.12a. Bilayers form very easily when

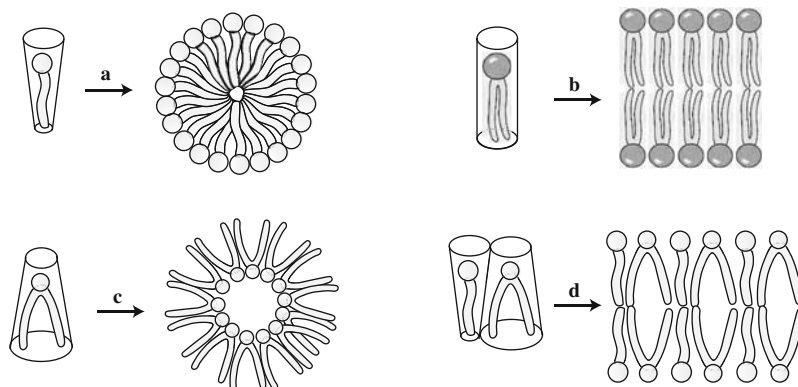


Fig. 2.12. Relation between structural shape of a lipid molecule and the way it arranges itself in a lipid/water phase (shape theory). If the lipid fits into an inverted cone, a micelle arrangement is produced (a). If the lipid fits into a cylinder, the arrangement is lamellar (b). If the lipid fits into a cone, the arrangement is H_{II} hexagonal (c). By a compromise between the different shapes, a lipid mixture may adopt a lamellar arrangement (d)

the cross-sections of the polar headgroup and the hydrophobic hydrocarbon chain of the lipid are the same. This happens for the glycerophospholipids and sphingolipids. Under these conditions, the shape of the lipid fits into a cylinder and, after aggregation, these lipids form a lamellar phase, as shown in Fig. 2.12b. Finally, the formation of a reverse micelle is favoured when the cross-section of the polar headgroup is less than the volume of the hydrocarbon chains, as happens in certain phospholipids or glycolipids (see Table 2.2). Under these conditions, the shape of the lipid fits into a cone and, after aggregation, these lipids form an H_{II} hexagonal phase as shown in Fig. 2.12c. In a mixture of different lipids with different phases when in the pure state, a compromise is found between the different shapes which may lead to a totally different structural arrangement in the end (see Fig. 2.12d).

The shape theory can explain the changes in structural organisation adopted by a given lipid when experimental conditions, such as temperature, concentration, ionic strength, degree of unsaturation of the hydrocarbon chains, and the presence of cholesterol, are varied. Hence, an increase in temperature, by increasing the motion of the chains, or dehydration, by reducing the volume of the polar part, favours an H_{II} arrangement (see Table 2.2). Likewise, this arrangement is favoured by unsaturation of the chains or by the presence of cholesterol, which induce an increase in the disorder of the chains, and hence in increase in the hydrophobic volume of the lipid. When many different lipid species with different structures are mixed together in biological membranes, this leads, by compensation between the different shapes, to a bilayer lamellar organisation.

2.3 Models and Methods for Characterising Membranes

Biological membranes provide excellent natural models of molecular structure and ordering, based on the self-association properties of amphipathic biological macromolecules. One of the main themes in nanobiotechnology consists in exploiting these properties to develop biomimetic membranes and ordered protein–lipid assemblies.

There are various membrane models with applications in the field of nanobiotechnology. These are:

- *Liposomes*. Closed vesicles bounded by a lipid bilayer enclosing a small aqueous compartment.
- *Langmuir Monolayers*. Ordered monomolecular films on a water surface or the surface of a neutralised solution (membrane leaflet).
- *Supported Bilayers*. Lipid bilayers formed on a solid substrate and obtained by depositing liposomes (supported lipid bilayer SLB) or by transferring monolayers occurring at the air/water interface (Langmuir–Blodgett films).
- *Suspended Membranes*. Lipid bilayers held at the surface of a solid substrate by spacers.
- *Bilayer Lipid Membranes (BLM)*. Flat lipid bilayers held across an aperture separating two aqueous compartments.

The aim in this section is to describe the various ways of obtaining and characterising these membrane models. Each system has its own particular advantages and points of interest for applications in different areas of nanobiotechnology.

2.3.1 Liposomes

Liposomes are vesicles comprising a lipid bilayer that separates two aqueous regions, the interior and exterior of the vesicle (see Fig. 2.10). Depending on the number of bilayers or lipid lamellae forming the wall, one can have unilamellar or multilamellar liposomes. The size of these liposomes is extremely variable, ranging from a few hundred μm for the largest to about 25 nm for the smallest. Liposomes are usually divided into the following categories:

- *Giant Unilamellar Vesicles (GUV)*. These have diameters greater than 5 μm .
- *Large Unilamellar Vesicle (LUV)*. These are large liposomes with sizes in the range from 50 nm to 1 μm .
- *Multilamellar Vesicles (MLV)*. These onion-like structures are made up of several concentric bilayers and have sizes greater than 100 nm.
- *Small or Sonicated Unilamellar Vesicles (SUV)*. These liposomes have diameters of the order of 25 nm.

Since they were introduced by Bangham around the middle of the 1960s [5], research on liposomes has increased steadily and much literature is regularly

devoted to them [6]. Liposomes constitute natural model systems for biological membranes. In this respect, they are widely used for biophysical studies of lipid phases and also structural and functional studies of membranes. Moreover, following the pioneering work of Gregoriadis [7], there has been a constant growth in applications of liposomes to areas such as drug targeting or cosmetology [8–10].

Methods for Synthesising Liposomes

Lipids are insoluble in aqueous solution but soluble in organic solvents such as chloroform, chloroform/methanol mixtures, and ether. Most ways of preparing liposomes use lipids dissolved in an organic solvent as starting point.

Multilamellar Vesicles (MLV)

Lipids dissolved in an organic solvent are placed in a glass container and the solvent is evaporated in vacuum. By spinning the container constantly during evaporation, the lipids are distributed uniformly over the walls. The lipids are then hydrated by simply adding an aqueous solution to the lipid film and the result is vortexed to produce multilamellar liposomes (see Fig. 2.13D). The sizes of these MLV and the number of bilayers composing them are variables that depend mainly on the conditions of synthesis and the kind of lipids used.

Cryo-TEM (cryo-transmission electron microscopy) is an almost ideal method for studying the structure of liposomes and complexes formed between proteins and membranes on the scale of the single liposome [12]. Problems due to collapse and structural reorganisation encountered when samples dry out during negative staining are avoided by maintaining an aqueous environment. Despite the absence of heavy atoms normally used as contrast agents in TEM, the intrinsic contrast of cryo-TEM images is sufficient to be able to resolve the two leaflets of the lipid bilayers, which are only about 4 nm apart. This intrinsic contrast comes about due to the presence of phosphorus atoms in the polar headgroups. These phosphorus atoms are indeed heavier than the other atoms and their cross-section with respect to electron scattering is consequently greater.

Large Unilamellar Vesicles (LUV)

Large unilamellar liposomes, i.e., larger than the SUV, can be obtained by several means: solubilisation of MLV in the presence of detergents, followed by elimination of the detergent; extrusion of MLV through polycarbonate filters; or phase inversion.

Detergents such as β -octylglucoside, dodecylmaltoside, Triton-X100, or CHAPS are amphiphilic molecules characterised by their critical micelle concentration (c.m.c.), the maximal monomer concentration beyond which the monomers are in equilibrium with micelle assemblies. When detergent molecules are incorporated into the lipid phase of the MLVs, the latter are

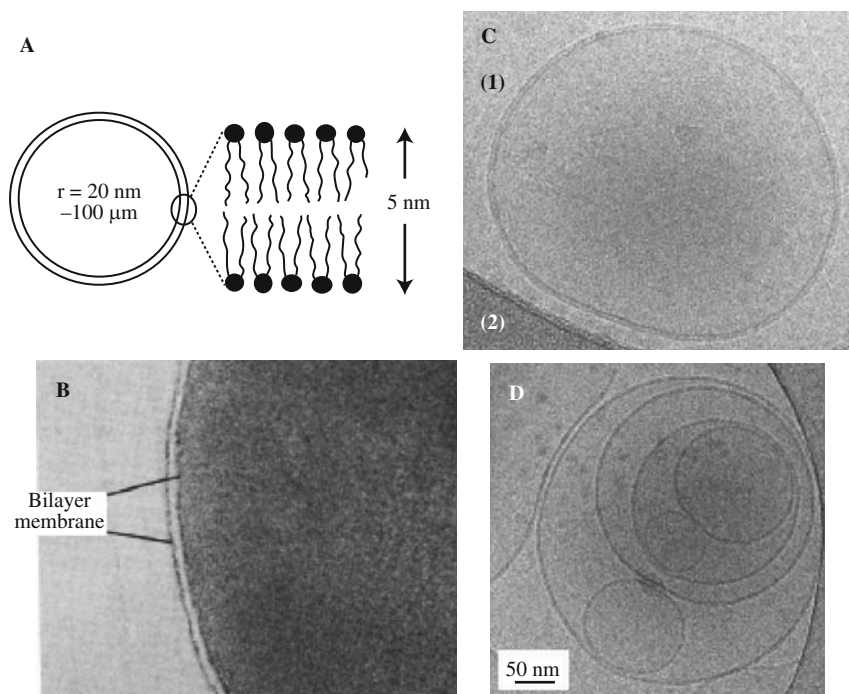


Fig. 2.13. (A) Schematic view of a liposome. (B) Membrane of a red blood cell, imaged by standard ultramicrotomy. Adapted from Robertson [11]. (C) LUV and (D) MLV, imaged by cryo-transmission electron microscopy (cryo-TEM): (1) amorphous ice film, (2) carbon film

solubilised beyond a critical concentration which depends among other things on the c.m.c. and the total amount of lipids. The MLVs are dispersed in the form of mixed micelles containing a mixture of detergent and lipid molecules. When the detergent is eliminated, and this can be done by dialysis, by adsorption onto polystyrene beads, or by dilution, the MLV solubilisation processes is reversed, in the sense that LUV-type lipid assemblies are reconstituted [13]. The size of the LUVs formed by detergent dialysis is of the order of 100 nm (see Fig. 2.13C). The term LUV is often used loosely. For example, the LUVs obtained by adsorbing the detergent on polystyrene beads can have several bilayers depending on the experimental conditions.

Extruded Unilamellar Vesicles (EUV)

Unilamellar liposomes can also be obtained by mechanically dispersing MLVs. The mechanical process known as extrusion consists in forcing the MLVs several times through polycarbonate filters with pores of definite size, e.g., with diameters in the range 50–400 nm. This process yields unilamellar liposomes with calibrated diameters called EUVs. However, it should be noted that

EUVs are often bilamellar or plurilamellar, and that a considerable amount of material can be lost when the lipids are forced through these membranes. It is therefore recommended to reportion the lipids after this process [14].

Small Unilamellar Vesicles (SUV)

Small unilamellar vesicles (SUV) are obtained when solutions of MLVs (or indeed LUVs or EUVs) are subjected to ultrasonic vibration. The MLVs are then transformed into SUVs with minimal dimensions of around 30 nm.

Preparing LUVs by Phase Inversion

Lipids are deposited on the walls of a glass container as described above, after which they are solubilised by first adding ether and then an aqueous solution. An ether/water emulsion is obtained by sonication, droplets of the aqueous solution being coated by a monolayer of lipids with their hydrocarbon chains pointing into the ether phase. Slow evaporation of the ether then leads to the formation of unilamellar liposomes [15]. The liposomes can subsequently be sorted according to size by filtering through a polycarbonate membrane. Large quantities of liposomes can be prepared by this method, with high concentrations, greater than 15 mg/mL.

Giant Unilamellar Vesicles (GUV)

Starting with a dry lipid film deposited on a substrate and hydrating without agitation, the lipid phase will swell up and very large liposomes are formed, with sizes up to a few hundred μm . One widely used technique called electroformation produces GUVs in an alternating electric field [16]. Owing to their large sizes, close to the size of a cell, giant vesicles have become very popular systems for studying the biophysics of complex membrane processes, especially using the many and varied techniques of optics.

Properties of Liposomes

To a first approximation, liposome properties reflect the properties of the lipids and the way they self-assemble. Since they are so easy to prepare and manipulate, liposomes constitute a choice model system for studying lipid assemblies and membranes. A large part of the field of membrane biophysics is taken up in studying the physical and physicochemical properties of these systems using the methods of physics. For example, there has been investigation of the relation between lipid composition and the type of lipid phase, membrane dynamics, permeability properties, and so on. There is a great deal of literature on this subject and it will not be discussed further here.

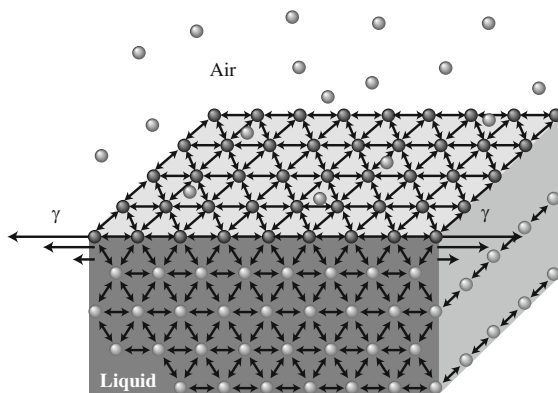


Fig. 2.14. Binding forces within a liquid and surface tension γ . In water, the molecules are arranged in such a way that the binding forces between them are shared equally over neighbouring molecules, so that the resultant force on any given molecule is zero. At the air/water interface, however, this molecular arrangement is no longer possible and the resultant force on any given molecule is nonzero. The interaction forces between molecules at the surface are strengthened, thereby creating an excess of free energy. The surface tension of water is 72.8 mN/m at 20°C

2.3.2 Langmuir Monolayers

Langmuir monolayers are ordered monomolecular films, made up of oriented amphiphilic molecules at an air/water interface. These are membrane models for a single leaflet of a biological membrane.

Forming an Insoluble Langmuir Monolayer

The formation of an (insoluble) Langmuir monolayer depends on the amphiphilic nature of the molecules making up the film and the specific thermodynamic properties of the air/water interface. At the surface of a liquid, there is an excess of free energy produced by the different environment of molecules in the surface and molecules in the bulk of the solution. In water, hydrogen bonds form and this produces a well-defined lattice, in which molecular binding forces are equal in all directions (see Fig. 2.14). At the interface, however, this equilibrium is broken. The attractive force which tends to pull molecules toward the interior of the liquid and reduce its free surface area is not counterbalanced. As a consequence, the lateral binding forces between molecules at the water surface are strengthened, thereby creating an excess free energy [17,18]. Making an interface thus costs a certain amount of surface energy, proportional to the area of the interface. This is the surface tension, denoted by γ .

When amphiphilic molecules, dissolved in a solvent such as chloroform that is immiscible with respect to water, are deposited on the surface of water, the dispersion forces due to the surface tension quickly cause the solution to

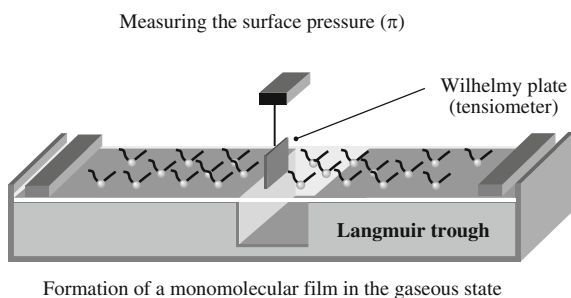
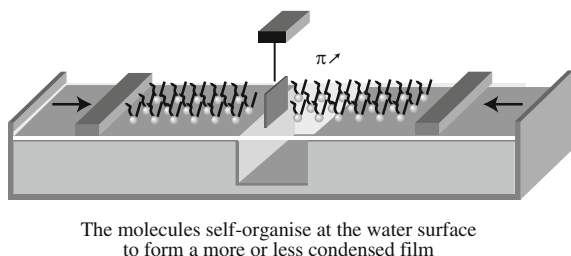
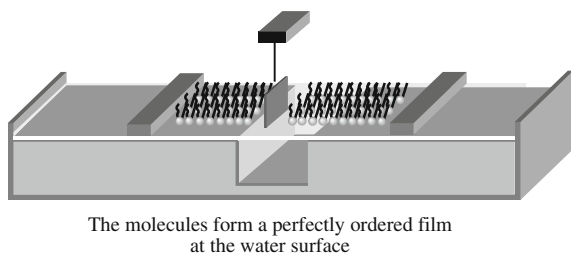
a) Spreading of amphiphilic molecules at the air/water interface**b) Compressing the interfacial film by means of the movable barriers****c) After compression: monomolecular film in the condensed state**

Fig. 2.15. Formation of a Langmuir monolayer. Monolayers are usually formed in a Langmuir trough (or balance). The trough is equipped with two movable barriers, one at each end, with which to compress the monolayer, and a Wilhelmy plate (or tensiometer), with which to measure the surface tension. For more details concerning the techniques and associated equipment, useful references are the books by Roberts [19] and Ulman [20] on Langmuir and Langmuir–Blodgett films

spread over the whole available surface. The surface tension can be likened to a negative pressure, related to the lateral cohesive forces between the surface molecules, which is reduced by the presence of certain molecules (surfactants) at the interface. When the solvent is evaporated, a monomolecular film called a monolayer is formed on the water surface, owing to the amphipathic nature of the molecules spread on it. The molecules orient themselves at the interface, with the polar head immersed in the aqueous phase (also called the subphase),

and the hydrocarbon chains remaining in the air (see Fig. 2.15). Langmuir monolayers are thus interfacial monomolecular films composed of insoluble amphipathic molecules. They correspond to an extreme case of interfacial adsorption, since all the molecules spread on the interface are concentrated into one interfacial monomolecular film.

Isotherms of a Langmuir Monolayer

Generally, the amount of molecules spread on the water surface is small enough to ensure that, initially, intermolecular distances are large and the presence of the monolayer at the interface only slightly affects the surface tension of the liquid. Under these conditions, interactions between molecules are weak and the monolayer can be treated as a 2D gas (see Fig. 2.15a). However, if the area occupied by the monolayer is reduced by means of two movable barriers (compression of the monolayer as shown in Fig. 2.15b), intermolecular distances are also reduced and the surface tension decreases (interactions between water molecules at the surface and the polar headgroups of the lipids). The molecules, which begin to interact via their hydrocarbon chains, tend to repel one another. The force exerted by the film is analogous to a 2D pressure field. It is defined as the surface pressure and denoted by π . It corresponds to a reduction in the surface tension of the liquid due to the presence of the film at the surface:

$$\pi = \gamma_0 - \gamma,$$

where γ_0 is the surface tension of the pure liquid and γ is the surface tension in the presence of the film. It is expressed in mN/m.

During compression, the amphiphilic molecules self-order and the monolayer undergoes different phase transitions, passing successively from the gaseous state to the liquid state, then from the liquid state to the solid state. As in three dimensions, the various states of the monolayer correspond to the different degrees of freedom and organisation of the molecules. At the end of the compression, a monolayer forms a perfectly ordered monomolecular film at the water surface (see Fig. 2.15c). During the compression process, the amphipathic nature of the molecules ensures that they remain aligned at the interface. Since the molecules making up the film are insoluble, the total number of molecules remains unchanged during compression. One can thus calculate at any time the average area occupied by a molecule at the water surface. This area, the molecular area A , is obtained by dividing the total area occupied by the film by the total number of molecules deposited there. It is usually expressed in nm²/molecule. By monitoring the surface pressure as the molecular area is varied, one can produce the isotherm π - A of the monolayer, commonly known as the Langmuir isotherm (see Fig. 2.16).

Quite generally, the Langmuir isotherm provides information about the interfacial properties of the film. It reveals the different states or phases that

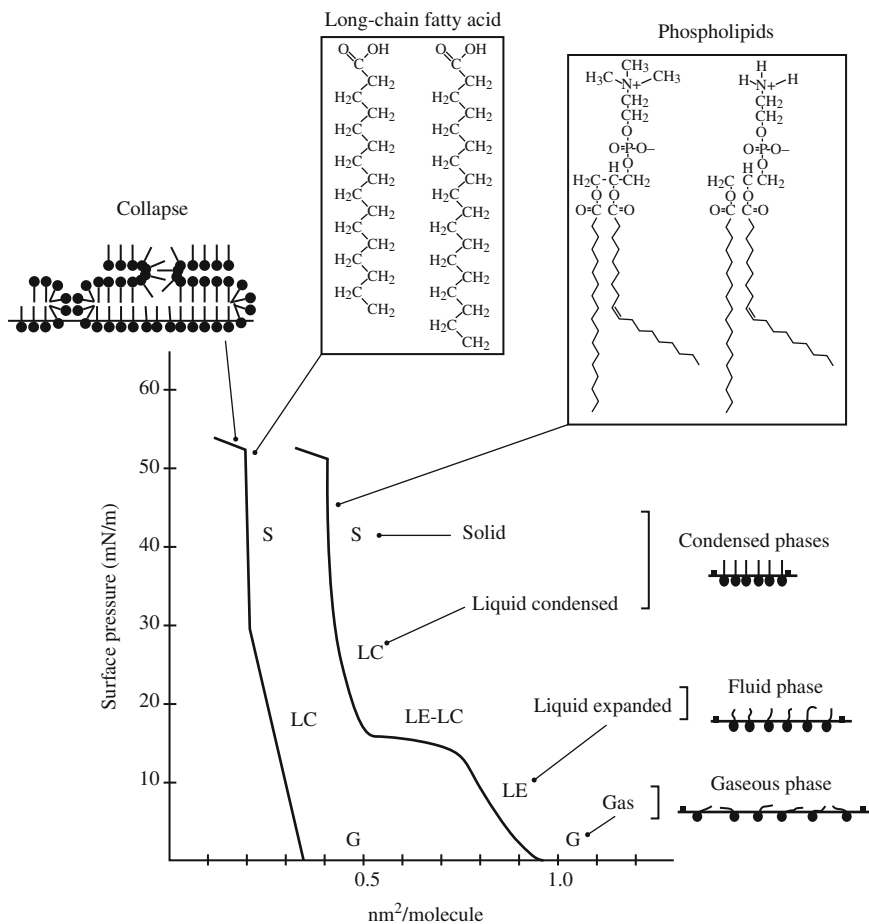


Fig. 2.16. Schematic isotherms of a Langmuir monolayer formed from phospholipids or long-chain fatty acids. The overall profile of the isotherm depends on various parameters including the subphase temperature, the degree of unsaturation and the length of the hydrocarbon chain(s), the spatial hindrance of the polar headgroup, and the presence of salts in the aqueous phase. Quite generally, a reduction in temperature or an increase in the length of the hydrocarbon chains strengthens the interactions between molecules and favours the formation of a condensed monolayer. Conversely, an increase in temperature or a decrease in the length or unsaturation of the hydrocarbon chains favours disorder (*gauche* conformation) which leads to the formation of a fluid phase monolayer. For a detailed discussion of these π -A isotherms, the reader is referred to the literature [17, 19–22]

may exist in a lipid monolayer (under given physicochemical conditions), and also the different phase transitions which may occur during compression. It

informs us not only about reorientation and conformational changes of the molecules in a 2D system, but also about the stability and purity of the monolayer [20].

The various phases obtained in these monolayers correspond to different levels of ordering of the molecules at the water surface, resulting from the molecular interaction forces arising in the film, and between the film and the aqueous phase. They can be identified by discontinuities in the isotherm. In the gaseous state (G), obtained for very large molecular areas, the molecular interaction forces are weak. The hydrocarbon chains of the molecules lie in the plane of the water surface and benefit from a high degree of freedom. There is no lateral cohesion in the film. In the liquid expanded (LE) phase, the monolayer becomes coherent, but the molecules still retain a certain amount of freedom. The LE phase corresponds to the fluid phase of the monolayer. The hydrocarbon chains have gauche conformations and random orientations. As the system is compressed, the molecular interactions grow stronger and the chains begin to stand up above the water surface. The monolayer then reaches its condensed state.

The condensed states of a monolayer correspond to the liquid condensed (LC) phase and the solid (S) phase. In the condensed states, the monolayer displays strong lateral cohesion. Interactions between hydrophobic tails are maximised. The chains are perfectly ordered at the water surface, in the trans-zigzag conformation. The condensed state forms by virtue of van der Waals forces between the chains. Indeed these constitute the cohesive forces in the monolayer. The two states of the condensed monolayer (LC or S) are crystalline states. They differ in the angle between the hydrocarbon chains and the plane of the water surface. In the solid phase of fatty acid monolayers, the hydrophobic chains are vertical. Compressing the film beyond the actual size allowed by the molecules leads to collapse of the monolayer, characterised by the formation of multilayers at the water surface and destruction of the 2D character of the monomolecular film.

The interfacial behaviour of a monolayer depends on the properties of the molecules making it up, i.e., length and degree of unsaturation of the hydrocarbon chains and steric hindrance of the polar head, and also the experimental conditions, i.e., temperature, subphase composition, pH, and ionic strength. Depending on the nature of the amphiphilic molecule and the physicochemical conditions, not all phases will necessarily occur. A direct transition from the gas phase to the liquid condensed phase is often observed in the isotherms of long-chain fatty acid monolayers. Likewise, the phase transition between the liquid expanded (LE) phase and the liquid condensed (LC) phase, denoted LE-LC, usually occurs in the isotherms of phospholipid monolayers. At the LE-LC transition, domains of the condensed (LC) phase appear in the fluid state (LE) monolayer. This coexistence of two phases can be seen directly in the monolayer using Brewster angle microscopy (see Sect. 9.5) or fluorescence microscopy, after incorporating a fluorescent lipid probe in the film (the

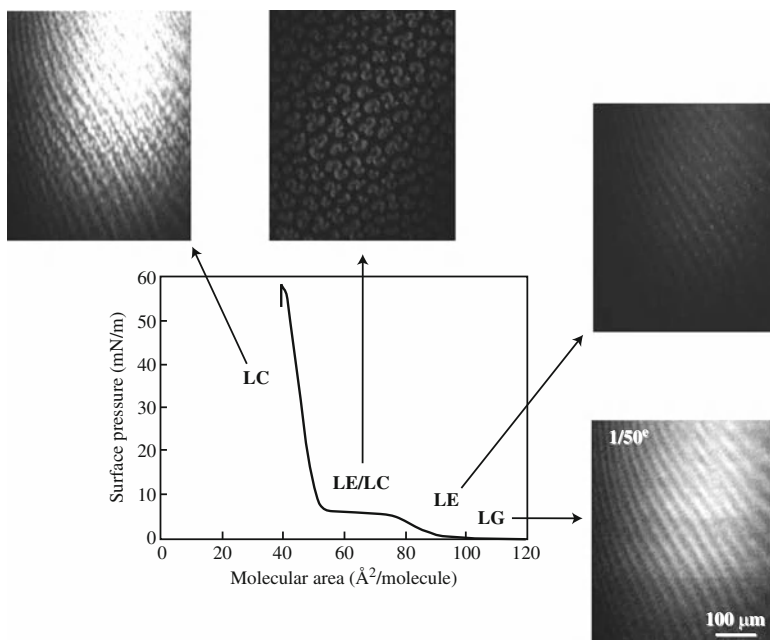


Fig. 2.17. Formation of condensed domains (*bright regions*) during compression of a dipalmitoylphosphatidylcholine (DPPC) monolayer. Observations made using Brewster angle microscopy [23,24]. Due to their abundance in biological membranes, the phosphatidylcholines are often taken as a model for studying biological membranes. Temperature 20°C

distribution of the probe molecules between the fluid and condensed phases allowing one to visualise the lipid domains).

As an example, Fig. 2.17 shows the aggregation of molecules and appearance of condensed domains in a model monolayer of dipalmitoylphosphatidylcholine (DPPC) when it is compressed. With the advent of these techniques for characterising monolayers directly at the air/water interface, it has become possible to study the aggregation and morphology of condensed domains as a function of the physicochemical conditions (temperature, ionic strength, chirality of the molecules, and so on) or during the interaction of a molecule with the monolayer (see below). These techniques are currently under rapid development, as a means of studying the formation or the reconstitution of microdomains in biological membranes.

Uses of Langmuir Monolayers

Langmuir monolayers are membrane models exploiting the self-association properties of amphipathic lipid molecules at the air/water interface. The main advantage with them is the possibility of obtaining a perfectly ordered state at the water surface and then being able to control this aggregated state by

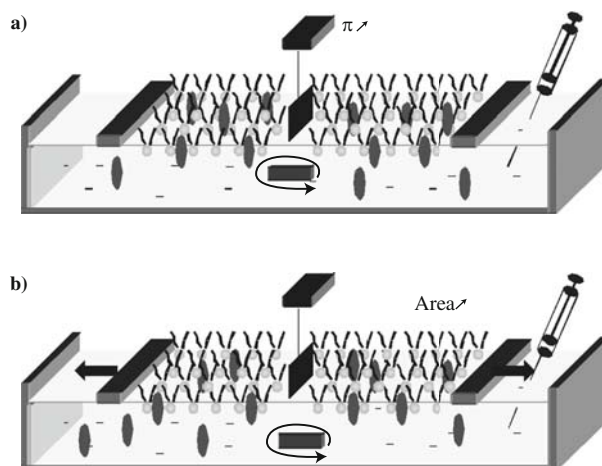


Fig. 2.18. Investigating the interaction of a protein or peptide using phospholipid monolayers. If a protein or peptide is inserted into the monolayer, there is either (a) an increase in the surface pressure (constant area) or (b) an increase in the area (constant pressure). By monitoring the increase in the surface pressure for different initial pressures, one can determine the interfacial exclusion pressure for the relevant protein (the maximum pressure above which the protein can no longer enter the monolayer [29]). By monitoring the increase in area after injection at different initial pressures, one can estimate the apparent molecular area of the molecule once it has entered the monolayer (if and only if the variation obtained is proportional to the amount injected, which can happen for measurements at low surface pressures [26]). Measuring the exclusion pressure is the simplest way of characterising the insertion of a protein in a lipid membrane and determining its specificity with regard to the kind of lipid in the monolayer. By comparing the exclusion pressure with the pressure generally assumed to correspond to the internal lateral pressure in biological membranes (30–35 mN/m depending on their composition [30]), one can assess the capacity of the protein to insert itself within the membrane

varying the imposed surface pressure. In nanobiotechnology, the interest in developing Langmuir monolayers is double-edged. On the one hand, they can be used to form supported lipid bilayers by transferring the monolayer onto a solid substrate (see Sect. 2.3.3), while on the other hand they are well-suited to the study of lipid/protein interactions or of macromolecules in general. Indeed, a molecule such as a protein or peptide can be injected into the aqueous phase under the monolayer. If it inserts itself into the interfacial film, thus indicating an interaction, the surface pressure increases, provided that the area is held constant (see Fig. 2.18a), or else the area increases, provided that the surface pressure is held constant (see Fig. 2.18b) [25–29]. This means that one can simulate, under realistic biological conditions, what happens when a water-soluble molecule (peptide, cytoplasmic protein, hormone, etc.) in the extracellular or intracellular medium interacts at the surface of the target cell (or organelle) membrane.

Compared with other model membrane systems, such as vesicles (see Sect. 2.3.1) or bilayer lipid membranes (lipid bilayers held across an aperture separating two aqueous compartments, see Sect. 2.3.5), all physicochemical parameters of this system, such as the kind of phospholipid (not all phospholipids can form vesicles), its state of aggregation, the composition of the subphase (ionic strength, pH) or the temperature, can be strictly controlled and easily modified [31]. The plane geometry of these monolayers means that all molecules in the film have a well-defined and specific orientation, while the presence of curvature would impose stresses on the polar headgroup. So by means of relatively simple measurements, this perfectly well-defined 2D system can be used to study the interactions of extrinsic soluble peptides, proteins, or other macromolecules with a monomolecular lipid layer corresponding to a membrane leaflet. This system can also be used to study the interaction of membrane or membrane-anchored proteins, but in this context, the model has limitations, because it is not strictly compatible with the biological membrane, in the sense that the bilayer structure of these membranes may be essential for the insertion of intrinsic proteins.

Although the main interest in studying lipid–protein interactions lies mainly in the study of the relation between structure and function in biological membranes, this kind of approach can lead to the formation of mixed monolayers, inserting macromolecules by self-association with the lipid leaflet. Once protein–lipid monolayers have been obtained in the condensed state, they can be transferred to a solid substrate in order to carry out supramolecular arrangements (see Sect. 2.4.1) and develop supported biomimetic membranes. In the context of nanobiotechnology, these could be used to functionalise surfaces or to develop biomimetic sensors (see Sect. 2.5.1).

2.3.3 Supported Membranes

Supported Lipid Bilayers

Lipid bilayers formed on a solid substrate, called supported lipid bilayers or supported bilayers for short, currently provide a very popular membrane model. One reason for the considerable interest in these systems as subjects of fundamental research is the constant development of physicochemical methods for studying interface phenomena. In addition, such systems have many potential applications in biotechnology. The most commonly used methods for preparing supported bilayers are liposome deposition, Langmuir–Blodgett transfer, a combination of these methods, or surface modification following liposome rupture. Langmuir–Blodgett films are discussed on p. 62. Supported lipid bilayers obtained by liposome deposition are commonly called SPBs (supported phospholipid bilayers) or SLBs (supported lipid bilayers). We shall use the more general term of SLB for the present discussion.

SLBs were first introduced as a model system for biological membranes in the mid-1980s, when it was observed that such structures form spontaneously

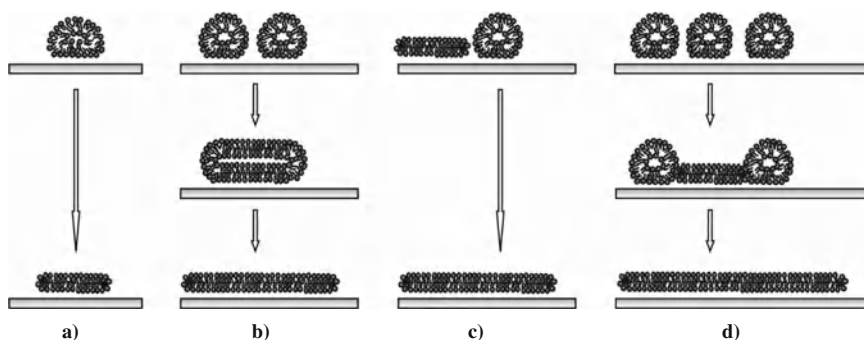


Fig. 2.19. Liposome rupture mechanisms on a solid substrate. (a) Spontaneous rupture of individual vesicles induced by deformation. (b) Fusion of vesicles followed by rupture. (c) The active edge of a bilayer domain induces the rupture of a neighbouring vesicle. (d) Cooperative action of several vesicles leading to rupture, at a critical local coverage, of a first vesicle, leading to the rupture in series of neighbouring vesicles [35]

when suspensions of liposomes or biological membranes are deposited on glass slides [32]. A considerable amount of work has since been devoted to studying these systems [33–35].

Our understanding of the processes involved in the deposition of lipid vesicles and SLB formation (see Fig. 2.19) has significantly improved recently, mainly thanks to the arrival of a series of methods capable of characterising molecular surface processes in a very detailed way. These methods can be divided into two types:

- Global physicochemical methods able to monitor adsorption of molecules onto a substrate in real time, to determine a range of thermodynamic and kinetic quantities, and in some cases to reveal conformational changes associated with the transition between vesicles and bilayer. Among these methods, the most widely used are the quartz crystal microbalance with dissipation monitoring (QCM-D), ellipsometry, and surface plasmon resonance (SPR).
- Imaging methods, and in particular atomic force microscopy (AFM) and the many fluorescence techniques, such as fluorescence recovery after photobleaching (FRAP), total internal reflection fluorescence (TIRF), fluorescence correlation spectroscopy (FCS), etc., provide an extremely effective way of obtaining local structural information, right down to the scale of the individual molecule or vesicle, concerning adsorption processes, transitions between vesicles and bilayers, or SLB dynamics.

The adsorption of liposomes, followed by their deformation, fusion, and rupture into bilayer domains, and the coalescence of these domains into an

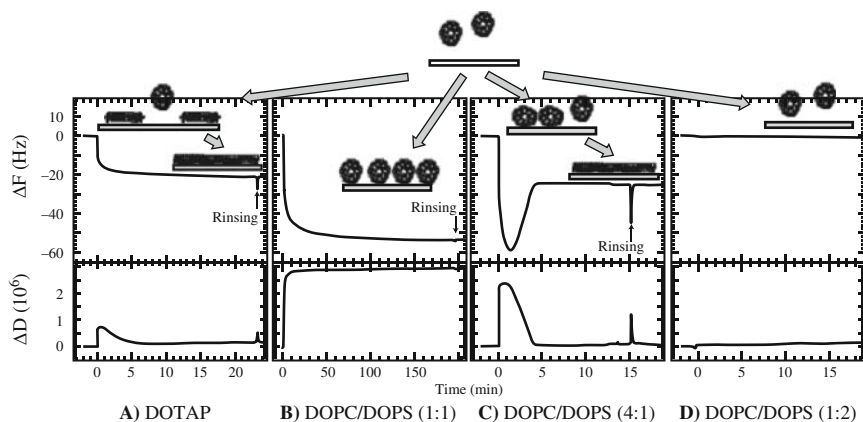


Fig. 2.21. Deposition of lipid vesicles on a silica substrate, observed by QCM-D. Adapted from Richter et al. [38]. QCM-D can measure in real time the adsorption of compounds on the surface of a quartz sensor. Variations ΔF in the resonance frequency of the quartz sensor provide a determination of the adsorbate mass, while variations ΔD in the dissipation characterise the viscoelastic nature of the coupling between substrate and adsorbate. The sensitivity of the measurement is about 10 ng/cm^2 and the time resolution of the order of one second. The SUVs used have the following composition and net charge (at neutral pH): (A) DOTAP positive net charge; (C), (B), (D) DOPC:DOPS (4:1), (1:1), (1:2) increasingly negative net charges. (A) DOTAP SUVs adsorb and decompose instantaneously to form an SLB. (C) DOPC:DOPS (4:1, w/w) SUVs adsorb up to a certain critical density beyond which they transform into bilayer domains which coalesce to form an SLB. (B) DOPC:DOPS (1:1) SUVs adsorb and remain intact, forming a vesicle monolayer. (D) DOPC:DOPS (1:2) SUVs do not adsorb onto the silica substrate. DOTAP: dioleoyltrimethylammonium propane. DOPC: dioleoylphosphatidylcholine, DOPS: dioleoylphosphatidylserine

vesicles is obtained (Fig. 2.20C). For moderate attraction, adsorption of the vesicles is only followed by their cooperative rupture when the density of adsorbed vesicles reaches a critical value, after which a continuous supported bilayer without defects covers the whole surface (Fig. 2.20B). This is generally what happens for silica, mica, titanium oxide, and glass, which are all negatively charged hydrophilic substrates [38, 43, 44].

These results were obtained for the main part using joint QCM-D and AFM studies. Such methods can characterise processes of adsorption and conformational change at previously unattainable resolutions [37, 38, 43, 45–47]. QCM-D [48] (see Sect. 9.6) is able to monitor the transition between adsorbed vesicle and bilayer domain, and is therefore particularly well-suited to investigating this phenomenon (see Fig. 2.21). For its part, AFM provides a way of tracking in real time the adsorption and rupture of the vesicles, not to mention the dynamics and in particular the coalescence of the lipid domains, and the formation of defects (see Fig. 2.22).

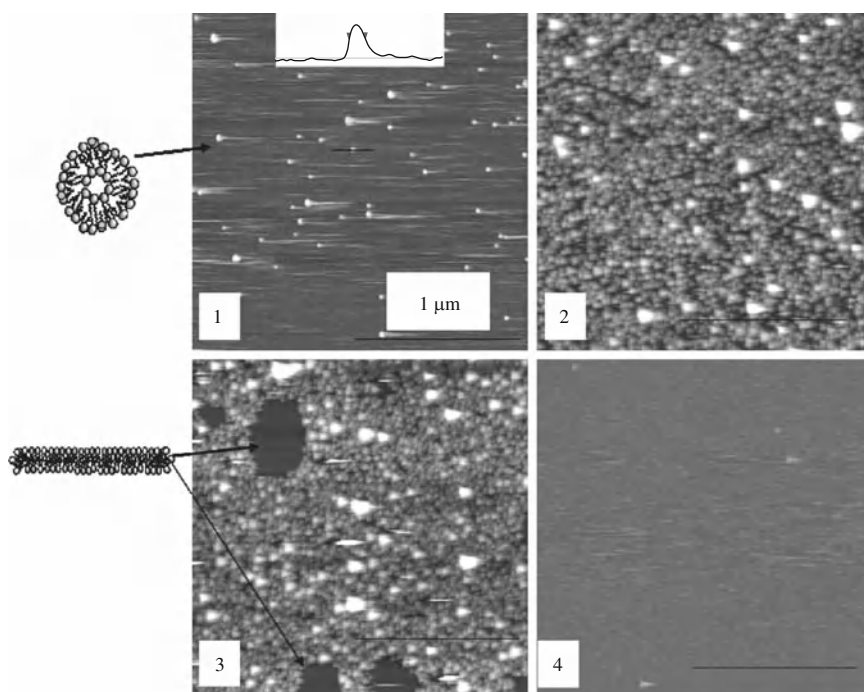


Fig. 2.22. AFM images of intermediate states observed during formation of a supported bilayer. Adapted from Richter et al. [38], with kind permission of the Biophysical Journal. The silica substrate is exposed to increasing amounts of SUV composed of DOPC:DOPS (4:1). (1) Individual adsorbed vesicles. *Inset:* Histogram of a vesicle of height 12 nm and width 25 nm. (2) High density of vesicles. (3) Several bilayer domains have formed and coexist with intact vesicles. (4) Supported bilayer entirely covering the substrate. Image size 2 μm. Height scale 50 nm

The combination of several investigative methods to study one process, as here for the formation of supported bilayers, is a common approach in the field of nanotechnology. However, it should be noted that these characterisation methods are often only applicable for certain types of substrate. For example, SPR requires gold substrates, whereas AFM works with planar substrates such as mica or silicon. For this reason, it is important not to extrapolate the conclusions obtained for a given substrate to other substrates: the characteristics of the substrate, such as chemical composition, charge, roughness, hydrophilicity, and so on, play a determining role for many of the processes under investigation.

The influence of chemical and physicochemical factors like vesicle composition, physical state of the lipids, temperature, liposome dimensions, ionic strength, pH, and so on, on the formation of supported bilayers has been studied in great detail [37,46,47,49]. The deformation and rupture of vesicles

and the formation of supported bilayers are processes dominated by electrostatic interactions, mainly those between the vesicles and the solid substrate. Several studies have revealed the key part played by calcium ions and negatively charged lipids in SLB formation [36–38, 44]. Hence, although it was long thought that titanium oxide was a substrate for which vesicles would not transform into a bilayer, experiments carried out recently with neutral lipids such as DOPC have clearly demonstrated the possibility of forming a supported bilayer on titanium oxide using liposomes containing DOPS, a negatively charged lipid, in the presence of calcium ions [35, 44].

It is commonly assumed that a water layer a few nanometers thick separates the substrate from the lipid bilayer [50]. Although the nature of the interactions between substrate and lipid bilayers is still not fully understood, it has been shown recently that the nature of the substrate can influence the distribution of lipids in the two leaflets of a bilayer [51]. Whereas DOPS distributes itself symmetrically over the two leaflets of an SLB on a silica substrate, it does so in a highly asymmetrical way on mica or titanium oxide substrates [52].

AFM studies of the stability of adsorbed lipid vesicles have identified different situations, depending on the nature of the substrate and lipids. The vesicles can either burst instantaneously, e.g., DOTAP SUV on silica (see Fig. 2.20A), or they can remain stable for periods of several days, e.g., DOPC:DOPS (1:1) SUV on silica (see Fig. 2.20C). But they may also rupture on a timescale of the order of one hour, e.g., DOPC:DOPS (4:1) SUV on mica [52]. The latter case is certainly due to the anisotropy in the lipid distribution, suggesting that some kind of flip-flop mechanism is responsible for this asymmetry [35, 52].

To sum up, we are just beginning to understand the processes involved in the formation of lipid bilayers on a solid substrate. Future studies will no doubt make it possible to specify the exact nature of the forces acting on the vesicles, leading to their rupture and the formation of supported bilayers.

Surface Nanopatterning by Supported Lipid Bilayers

The development of miniaturised systems such as chips – DNA, protein, or cell chips – provides the motivation for a great deal of research, owing to the biotechnological applications of these systems. Various surface patterning methods, including photolithography, μ -contact printing, nanodroplet deposition, and a whole range of chemical methods have been developed recently.

Surfaces have been patterned using supported lipid bilayers and exploiting the different behaviour of certain substrates with regard to the adhesion of vesicles and the formation of the supported bilayer [44, 53] (see Fig. 2.23).

Langmuir–Blodgett (LB) Films

The Langmuir–Blodgett technique was invented in 1935 by I. Langmuir and K. Blodgett [54, 55]. The idea was to pick up monomolecular films formed on a

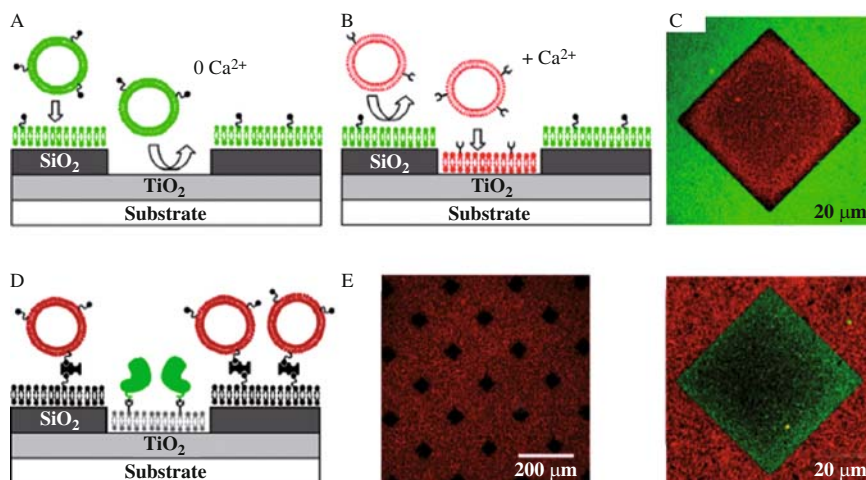


Fig. 2.23. Nanopatterning substrates by means of supported bilayers. Reproduced with the kind permission of the American Chemical Society (copyright 2005) [44]. (A) When vesicles containing a mixture of DOPC and DOPS are added in the absence of Ca^{2+} to substrates in which silica and titanium oxide are juxtaposed, a bilayer forms on the silica but the vesicles do not adsorb onto the titanium oxide. (B) In the presence of Ca^{2+} , vesicles containing DOPS form a bilayer on the titanium oxide zones. (C) Fluorescence microscopy image showing the presence of two types of bilayer: labelling by TRITC (red) on the titanium and NBD (green) on the silica. (D) Diagram showing the addition of different functional groups – biotin, Ni-NTA – in the bilayers. (E) Fluorescence microscopy images using a histidine tag to show the bond between the GFP protein (green) and a bilayer containing Ni-NTA lipids on the titanium oxide zones. Red fluorescent regions correspond to vesicles, containing a fluorescent lipid, fixed by means of streptavidin to a biotin-tagged bilayer

water surface and transfer them to solid surfaces. Since then this technique has been considerably developed to become one of the most widely used methods of supramolecular engineering, in which a thin film can be deposited under conditions of controlled orientation and thickness. The Langmuir–Blodgett (LB) technique is particularly useful for preparing lipid layers, with good control over thickness (of nanometric order) and molecular organisation.

Basic Idea and Advantages of the Langmuir–Blodgett Technique

This technique uses the property of insoluble amphiphilic molecules such as lipids, phospholipids, and glycolipids of forming an ordered monomolecular film at the air/water interface (see Sect. 2.3.2). When this film is compressed and the molecules are sufficiently cohesive (generally in the liquid condensed phase), the monolayer can be lifted and transferred like a carpet onto a solid substrate cutting the interface in the vertical direction (see Fig. 2.24). Depending on whether the substrate is hydrophilic or hydrophobic, the first layer will

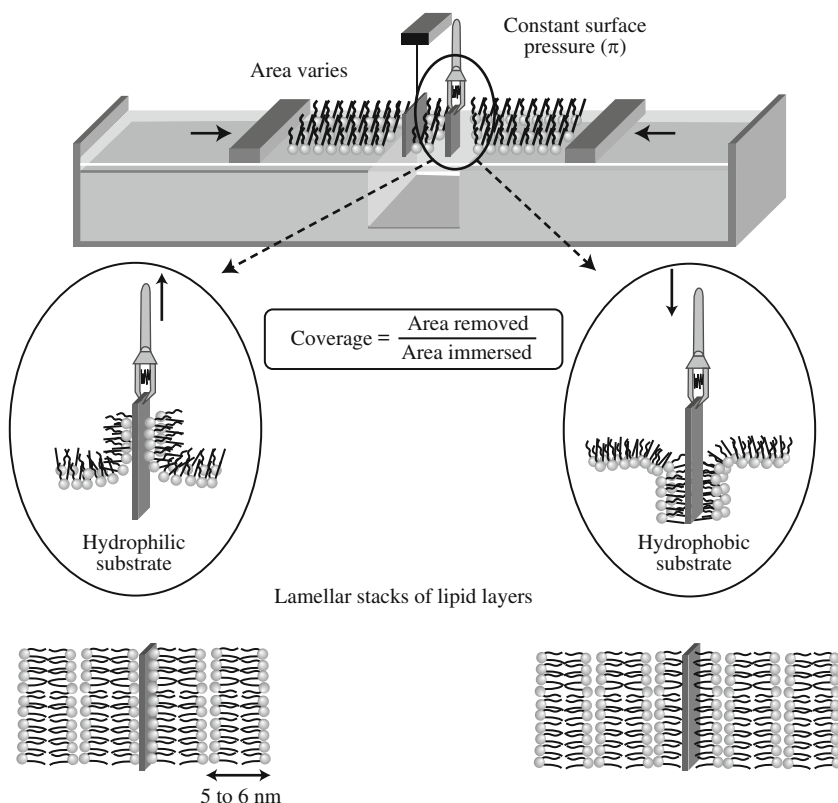


Fig. 2.24. Transfer of monomolecular layers by the Langmuir–Blodgett technique. During transfer, the monolayer is held at constant pressure. The change in area due to removal of the monolayer, which induces a drop in surface pressure in the film, is then compensated by adjusting the movable barriers. This change can then be used to calculate the coverage of the substrate, defined as the ratio of the area of the monolayer lifted from the water surface during the transfer to the theoretical area of the immersed substrate. The coverage gives an indication of the quality of the lift. Transfer is totally efficient when the transfer rate is equal to unity

be transferred during emersion (upstroke) or immersion (downstroke) of the substrate, respectively. By passing the substrate through the monolayer held at constant pressure, supported lipid stacks called Langmuir–Blodgett films are produced, monomolecular layer by monomolecular layer. Their thickness and the hydrophilicity or hydrophobicity of their surfaces depend on the number of layers deposited.

This transfer of lipid molecules generally produces a stack of bilayers, with one monolayer being deposited each time the substrate goes through the interface. This type of deposition, said to be of type Y, is representative of the natural organisation of biological membranes (see Fig. 2.25b). However, depending on the amphiphilic balance of the molecules, the kind of substrate,

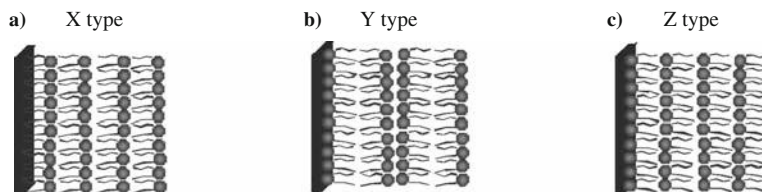


Fig. 2.25. Different types of Langmuir–Blodgett films. Y type is the most stable. Given the amphiphilicity of the molecules, reordering may occur in X (or Z) type structures in such a way that the molecules finally adopt a Y type structure [56]

and transfer conditions (transfer pressure, deposition rate, subphase pH and composition), monolayer removal may only occur during immersion (X type) as shown in Fig. 2.25a, or during emersion of the substrate (Z type) as shown in Fig. 2.25c.

The Langmuir–Blodgett technique allows a good level of control in each stage of the preparation of supported films, viz., controlled formation of the interfacial film, preparation of the substrate, and transfer parameters. The main advantage with membranes obtained by building up lipid layers, and Langmuir–Blodgett films in general, lies in the highly ordered molecular arrangement that can be achieved on the water surface, and which can be conserved during transfer onto the substrate when all transfer parameters (surface pressure, rate of immersion of the substrate, temperature, composition of the aqueous phase) have been optimised. Strictly optimised experimental conditions, almost for each type of molecule, are required to obtain repeatable results. Subtle changes in these conditions may modify the behaviour of the monolayer during transfer. The reader is referred to the literature on Langmuir and Langmuir–Blodgett films [19–21] for a detailed description of the experimental conditions required to guarantee high-quality monolayer formation and transfer to the solid substrate, viz., purity of the lipid material, choice of solvent for spreading the molecules, purity of the water and the compounds in the aqueous phase, control of subphase temperature, trough environment (antivibration table, controlled atmosphere), substrate preparation, and so on.

Phospholipid LB Films

In the context of biomimetic membranes, Langmuir–Blodgett films made from phospholipids and/or glycolipids are obviously the most representative. However, these molecules are complex lipids (see Sect. 2.2.1). They comprise two hydrocarbon chains per molecule (not necessarily of the same length), and their polar headgroup, which may be electrically charged, is generally highly hydrated. The primary hydration shell of the polar headgroup of a phosphatidylethanolamine or phosphatidylcholine molecule contains 5 or 20 water molecules, respectively [57]. Hydration of the polar headgroups of glycolipids depends on the number of glucidic structures they contain. Owing to this complex structure, it is not always easy to transfer the membrane lipids, and

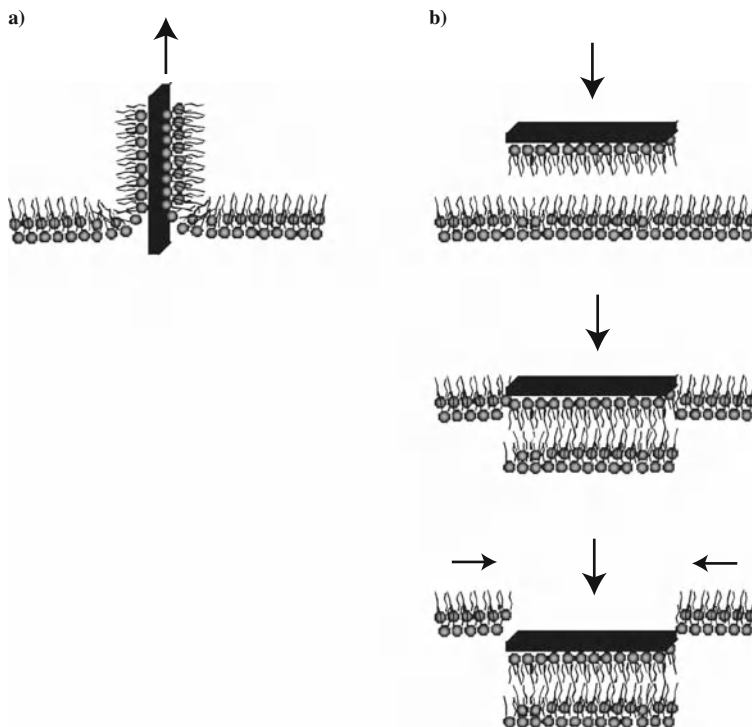


Fig. 2.26. Hybrid method for depositing a phospholipid bilayer on a solid substrate. (a) The first layer is picked up by vertical Langmuir–Blodgett transfer. (b) The second layer is deposited by horizontal transfer using the Langmuir–Schaefer method. After transfer, the bilayer is immersed in the aqueous phase

the structures produced on hydrophilic substrates are not always perfectly Y type. Owing to the strong affinity of the polar headgroups for the aqueous medium, it is not unusual for the first layer transferred upon extraction of the substrate to redisperse itself on the water surface when the substrate is pushed back through the interface upon immersion. Likewise, once several layers have been deposited, the efficiency of transfer can drop and the interaction forces set up between the last layer transferred and the interfacial film are not always strong enough to tear the monolayer from the water surface. However, phospholipids with a small polar headgroup, such as phosphatidic acid (see Fig. 2.2a), generally form bilayer stacks with good coverage [58, 59]. The transfer efficiency of phospholipid films is generally improved by adding divalent cations to the aqueous phase or mixing phospholipids together or with fatty acids. In the same way, a surface treatment such as oxidation, deposition of a thin metal film or an LB lipid film comprising fatty acids, can strengthen the interaction of the first transferred layer with the substrate and avoid any subsequent redispersion on the surface. The strong affinity of the polar

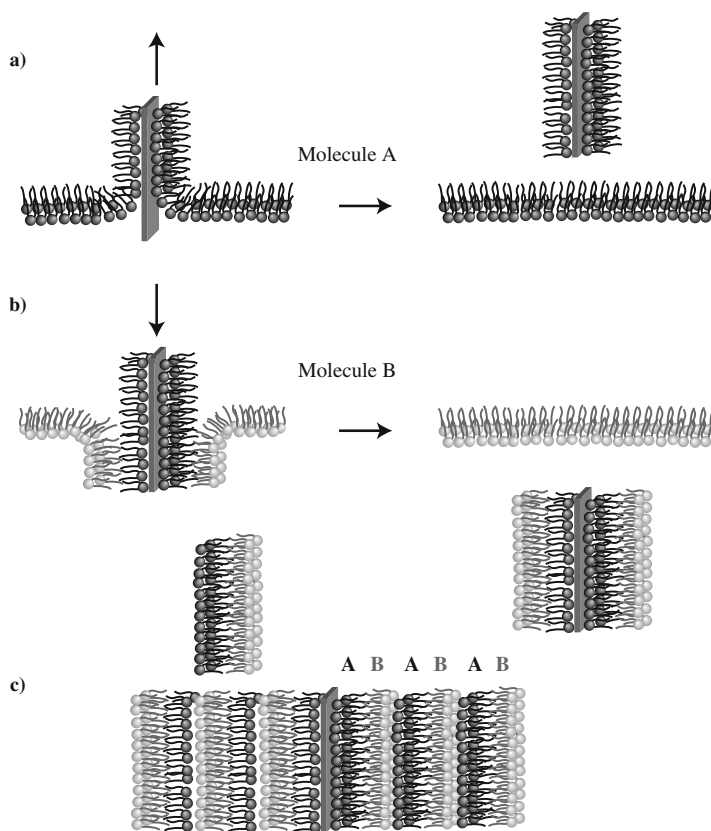


Fig. 2.27. Alternating Langmuir–Blodgett layers. These are produced by (a) pulling the substrate out through a monolayer made from a compound A, then (b) reimmersing the substrate through a monolayer made from a compound B, and so on. (c) Final arrangement of alternating layers made up of several bilayers

headgroups for their aqueous substrate can generally be counterbalanced by using a highly hydrophilic substrate, thus favouring the adhesion of the first layer.

In this respect, silicon is a suitable substrate for forming high quality phospholipid bilayers [60]. The surface properties of this type of material tend under condensation to strengthen the adhesion of the molecules and favour an orientation (relative to the substrate surface) such that interactions with the following layer can be favourably set up to facilitate the transfer. On this type of substrate, up to 21 layers of dipalmitoylphosphatidic acid can be transferred, and up to 5 layers if it is mixed with dipalmitoylphosphatidylcholine [60]. To avoid back-transfer of the first layer, in 1985 Tamm and McConnell [61] suggested a combined approach for elaborating lipid bilayers on a hydrophilic substrate (see Fig. 2.26). In this method, the first layer

is deposited by vertical Langmuir–Blodgett transfer. The substrate is then rotated through 90° and brought horizontally to the monolayer surface. The second layer is deposited by pushing the substrate through the interface (horizontal Langmuir–Schaefer transfer). After the transfer, the bilayer finds itself immersed in the aqueous phase.

Asymmetric Bilayers

The Langmuir–Blodgett technique can be used to prepare asymmetric bilayers called alternating layers. In these alternating layers, the composition of each lipid leaflet is different. These layers are obtained by successively passing the substrate through monolayers with different compositions (see Fig. 2.27). To some extent, this type of layer can reproduce the asymmetric composition of biological membranes. Indeed, the lipid composition of the inner leaflet of a biological membrane is usually different from that of the outer leaflet. As an example, consider the composition of the membrane of the human erythrocyte (red blood cell). The inner leaflet is rich in phosphatidylethanolamines (see Fig. 2.2c), phosphatidylinositols (see Fig. 2.2d), and phosphatidylserines (see Fig. 2.2e), while the outer leaflet contains more phosphatidylcholines (see Fig. 2.2b) and sphingolipids (see Fig. 2.5b) [62]. Asymmetric membranes can also be produced by the combined Langmuir–Blodgett and Langmuir–Schaefer method [63].

The asymmetric membranes formed by Langmuir–Blodgett deposition are currently being used to study phase separation and the formation of membrane lipid microdomains in lamellar structures [64–68]. Indeed, biological membranes are no longer considered as totally fluid systems, as had been proposed in 1972 in the fluid mosaic model due to Singer and Nicolson [69]. More and more experimental evidence suggests that there are condensed membrane microdomains called lipid rafts within this structure, rich in cholesterol and sphingolipids [70]. These microdomains are thought to play a crucial role in many cell functions, such as endocytosis [70], biological signal transduction [71–74], or molecular targeting (specific sorting and transport of lipids and proteins to well-defined locations in the plasma membrane) [75]. By studying the aggregation of lipids and the proteins making up these microdomains, in monolayers at the air/water interface or in supported asymmetric bilayers, we now have a better understanding of the way in which these microdomains form and organise themselves.

Finally, the transfer method for alternating LB layers can be used to make bilayers in which the first leaflet is composed of zwitterionic phospholipids and the second of anionic phospholipids. These structures can serve as membrane supports for incorporating proteins, since many of them can associate with membranes through interactions with the anionic phospholipids [76].

Characterising Langmuir–Blodgett Films

A wide range of techniques exists for characterising LB films. These techniques are summarised in Table 2.3 together with the information that can be deduced from them.

The Langmuir–Blodgett method provides a good way of forming perfectly ordered and structured films. However, with the advent of high-resolution techniques for characterising surfaces, a certain number of defects have been identified. These include differences in the tilts of the hydrocarbon chains [77], the formation of inhomogeneous crystal domains [78], nanoscale holes [76, 79–82], local collapse [83], empty regions that have not been coated [64], and lateral and between-bilayer heterogeneities [84]. Today, some defects can be identified as being directly due to the properties of the molecules making up the film, such as the occurrence of crystal domains with different tilt angles or lipid segregation causing heterogeneities in the bilayers. Others, i.e., empty zones, can be eliminated by optimising transfer techniques coupled with high-resolution characterisation methods. Note, however, that the presence of some defects such as molecular scale holes, which can create electrical defects, put a limit on applications of LB films to electrode–molecule–electrode structures. Conversely, some applications such as surface patterning can exploit discontinuities generated by instabilities in the meniscus during transfer of the monolayer to create nanoscale channels in a bilayer [85–88]. However, the formation of biomimetic membranes with applications in the field of nanobiotechnology can only be envisaged if the lipid films form homogeneous structures.

We have already considered the difficulties arising in the transfer of phospholipid monolayers. Indeed, it is not unusual for the coverage of the second layer deposited under immersion, when the hydrocarbon chains have to interact with those of the first layer transferred to the hydrophilic substrate, to take a value in the range 0.5–1.0, which must theoretically be interpreted as incomplete transfer [59, 76, 80, 88–91]. By a combined analysis of the coverage values and AFM observations of the structures obtained, Bassereau and Pincet put forward an explanation for this phenomenon in 1997 [80]. It may be that, during transfer of the second layer, some molecules from the first layer already transferred desorb from the substrate to join the interfacial film. Under these conditions, the fact that the coverage is less than unity does not imply a poor quality transfer, i.e., molecular overlap, but reflects rather the balance between molecules of the interfacial film in the process of being transferred and those leaving the substrate. Desorption of lipid molecules during transfer of the second layer is what causes the holes seen by AFM on the nanometric scale in different phospholipid bilayers [64, 76, 80, 88]. The number, shape and size of these holes (between 30 nm, typical size of the AFM tip, and 500 nm [80]) depend on the pressure and the transfer rate of the film, the number of defects in the lower layer, the type of lipids used, and the condensation state of the lipids at the water surface [76, 80]. Quite generally, their size and density decrease with an increase in the transfer rate or pressure [80]. Since the desorption

Table 2.3. Overview of the most widely used techniques for characterising the structure and organisation of LB films. For more detail on the various techniques, the reader is referred to the textbooks by Roberts [19], Ulman [20], Petty [21], Tredgold [77], and references Peng et al. [174] for GIXD, Chazalet et al. [175] and Heywang et al. [176] for SERS, Leverette and Dluhy [177] for SERS and IRRAS, Vandevyver and Barraud [178] for microscopy, Nomarski, Dufrène and Lee [81], Rinia and de Kruijff [84], Sanchez and Badia [179] and Chap. 7 of this book for AFM, Cordero et al. [83], Hollars and Dunn [180] for FL-NSOM. Readers will also find the basic principles of STM and AFM in [181]

Technique	Information obtained
Ellipsometry	Determination of thickness (to within 2 Å) and refractive index
Neutron diffraction, neutron reflection	Structural information, overall thickness of LB film, average thickness of a molecular layer
X-ray diffraction	Thickness of multilayers, thickness of a monolayer, periodicity between planes, electron density profile normal to the plane of the substrate (molecular arrangement: head-to-head or tail-to-tail), crystal structure of LB films (vertical or tilted orientation of hydrocarbon chains)
Grazing incidence X-ray diffraction (GIXD)	Information concerning parameters of the molecular lattice in the plane of the structure (parallel to the leaflet) and tilt of hydrocarbon chains
Electron diffraction	
Transmission electron diffraction (TED)	• In-plane intermolecular distances, crystal structure, and defects
Reflection high energy electron diffraction (RHEED)	• In-plane structural information, direction and tilt of aliphatic chains
Low energy electron diffraction (LEED)	• Distances between chains
Infrared spectroscopy	
Polarised IR spectroscopy, infrared reflection adsorption spectroscopy (IRRAS), attenuated total reflection-Fourier transform infrared (ATR-FTIR)	• Conformation of hydrocarbon chains, degree of ionisation of the polar head, hydrogen bonds and structural changes • Organisation and tilt of hydrocarbon chains, molecular orientation
Surface-enhanced Raman spectroscopy (SERS)	Conformation of alkyl chains and polar head, interaction of biomolecules with LB films
X-ray photoelectron spectroscopy (XPS)	Quantitative analysis of chemical composition
Electron spectroscopy for chemical analysis (ESCA)	Quantitative analysis of chemical composition
Optical microscopy	
Nomarski differential interference contrast microscopy	• Visualisation of birefringent crystal defects, path difference in the optical thickness of the LB film (difference of thickness between two monolayers) (to within 30 Å, with lateral resolution 1 µm)
Fluorescence near-field scanning optical microscopy (FL-NSOM)	• Orientation of molecules, morphology of lipid domains, defects at domain boundaries, microcollapse (lateral resolution 0.1 µm)
Electron microscopy	
Scanning electron microscopy (SEM), transmission electron microscopy (TEM) (with shadowing, replication, silver decoration and charge decoration)	Surface morphology, structure of lipid domains, patterns, holes and defects (inhomogeneous crystal domains, microcollapse, etc.). Resolution 5 nm (SEM), 2 nm (TEM)
Scanning tunnelling microscopy (STM), atomic force microscopy (AFM)	Surface topography, morphology of lipid domains, defect visualisation (grains, holes, lateral heterogeneity, differences in tilt, etc.), molecular organisation (lateral resolution 0.2 nm)

phenomenon depends on the balance between the energy of adsorption of the molecules onto the substrate and that of the molecules at the air/water interface, the reduction in surface tension due to compression of the interfacial film reduces the affinity of the polar headgroups for the air/water interface and favours adsorption of molecules onto the substrate [80]. The condensation state of the monolayer at the interface will thus determine the number of molecules of the first layer that will leave the substrate during transfer of the second, and in general fewer defects are present in the bilayer if the second leaflet is transferred at higher surface pressures, i.e., when the monolayer is in the liquid condensed (LC) phase [68, 76, 82, 92]. In addition, the formation of asymmetric bilayers generally produces more uniform bilayers. This is exemplified by bilayers made up of a first leaflet of phosphatidylethanolamines and a second leaflet of phosphatidylcholines [67, 76, 80], which reflects the natural asymmetry of biological membranes. In this case, the vertical orientation of the phosphatidylethanolamine hydrocarbon chains favours transfer of the phosphatidylcholine monolayer, which has tilted chains, by avoiding back-transfer.

The presence of holes through the bilayer in LB phospholipid films may seem surprising since the phospholipids, when suspended in water, self-associate and self-organise into vesicles made from perfectly sealed bilayers (see Sect. 2.2.2). As reported by Bassereau and Pincet [80], the desorption phenomenon is probably shared by all supported bilayer systems. Furthermore, as pointed out by Benz et al. in 2004 [82], hole formation in supported membranes can be related to the temporary existence (over a few picoseconds) of aqueous pores in sufficiently thin, free membranes, formed from phospholipids with medium-length chains (C12, C14) [93]. Finally, the LB transfer of a phospholipid bilayer onto a hydrophilic substrate is by far the most relevant approach for producing biomimetic membranes. Note, however, that the whole of the bilayer or part of the outer leaflet can detach itself from the substrate when the bilayer/substrate system crosses a pure water interface vertically. This detachment phenomenon is once again related to the energy balance of the interactions between the polar headgroup of the lipid on the substrate and the air/water interface, for which it has such an affinity, and the strength of the hydrophobic interactions between hydrocarbon chains of the two consecutive layers. In addition, a lipid bilayer deposited on a hydrophilic surface must be manipulated with great care to avoid any risk of desorption [80] and to maintain the stability of its outer leaflet [82].

To preserve its integrity, the supported bilayer must ideally be kept in an aqueous medium. After transferring the last layer by immersing the substrate, the supported bilayer can be recovered in a small container already placed in the aqueous phase before formation of the monolayer at the interface and transfer of the bilayer [64, 65, 67, 68, 76, 80, 82, 94]. This precaution also makes it possible to work with perfectly hydrated bilayers, and this is the most representative state for biological membranes immersed in the intra- and extracellular media. Clearly, the success of applications of biomimetic

membranes in nanobiotechnology will require a very good understanding of the fundamental self-ordering properties of the membrane itself.

To sum up, in comparison with other membrane models, the Langmuir–Blodgett technique can be used to produce patterned and ordered lipid bilayers on different types of substrate, e.g., silica, mica, silicon, platinum, etc. Homogeneous bilayers cannot be formed on all types of substrate by liposome adsorption. By virtue of their lamellar morphology (Y type film), which is reproducible when all transfer parameters have been optimised, Langmuir–Blodgett films can be used to study biological membranes after inserting a biological element (see Sect. 2.4.1). They are indeed good substrates for directly fabricating ordered protein–lipid molecular assemblies at the surface of a transducer. In addition, the possibility of producing asymmetric membranes, and biological membranes are of this kind, provides new opportunities for developing biomimetic systems integrating complex membrane recognition systems. In the field of nanobiotechnology, the elaboration of functionalised lipid bilayers on a silicon substrate looks like a promising way of miniaturising hybrid bioelectronic systems.

2.3.4 Suspended Membranes

Supported lipid bilayers (SLB), formed by adsorbing liposomes onto a hydrophilic substrate, are generally held on the substrate by means of a thin film of water, between 1 and 2 nm thick. This aqueous film allows the membrane to maintain its fluidity, the lipids and proteins associated with the lipids conserving their lateral mobility within the two leaflets of the bilayer (see p. 57). These membrane models have been successfully used to study the adsorption, interaction, function and self-association of peripheral membrane proteins (see Sect. 2.4) or membrane-anchored proteins when reconstituting membrane microdomains [95]. However, this film separating the membrane from the substrate is not thick enough if one hopes to reincorporate integral membrane proteins in this type of membrane. The hydrophilic parts of the intrinsic proteins, sticking out from the bilayer, can interact with the hydrophilic surface of the substrate, which fixes the protein and restricts its lateral mobility. The membrane then loses its fluid dynamical properties.

Suspended membranes are supported lipid bilayers held at the surface of a solid substrate by means of spacers, usually made from a specific ligand/receptor recognition system. These spacers are used to hold the bilayer away from the substrate, which favours the reinsertion of integral membrane proteins without their losing their lateral mobility.

Two systems stand out in the development of suspended membranes. The first exploits the chelation properties of the polar headgroup of a lipid functionalised by a N-nitrilotriacetic acid (NTA) group which can form a metal–chelate complex with a metal ion such as nickel. This Ni-NTA complex is then able to bind two imidazole nuclei of a histidine tag, i.e., a polyhistidine sequence, usually comprising a six-residue peptide. In the system illustrated

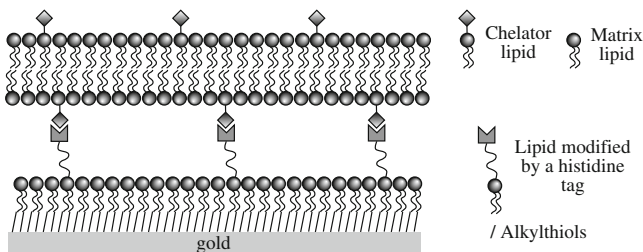


Fig. 2.28. Schematic view of the formation of a suspended membrane using a Ni-NTA/histidine complex. One monolayer of phospholipids inserting a lipid modified by a six-histidine tag is formed at the surface of an octadecylthiol layer self-assembled on a gold substrate. The suspended membrane is formed by adsorbing liposomes of phospholipids incorporating a lipid with a head functionalised by an N-nitrilotriacetic acid (NTA) chelator able to form a metal–chelate complex with a metal ion, viz., nickel. The membrane is then maintained by the Ni-NTA complex binding two imidazole nuclei of the polyhistidine sequence. From Rädler et al. [96]

in Fig. 2.28, a lipid modified by a tag of six histidines (His) on its polar part is inserted into a phosphatidylcholine monolayer formed on an octadecylthiol prelayer self-assembled on a gold substrate. Phospholipid vesicles inserting the phospholipid functionalised by an NTA chelator charged with nickel ions are adsorbed at the surface of lipids carrying the histidine tag. The bilayer thereby formed is thus maintained at the surface of the substrate functionalised by means of the Ni-NTA/His complex [96]. The same type of membrane can be obtained by functionalising the gold substrate by the NTA chelator and inserting the lipid modified by the histidine tag on its polar region into phospholipid vesicles [97].

The second exploits the specific recognition capability of the avidin/biotin system, which has a very high affinity constant (10^{15} M^{-1}). Phospholipid vesicles inserting biotinylated lipids are adsorbed onto a plane functionalised surface by a layer of streptavidin (see Fig. 2.29).

Addition of polyethylene glycol (PEG), a fusing agent for lipid vesicles, leads to the formation of a continuous bilayer, whose integrity has been checked by AFM [98]. This type of membrane formed inside a microporous electrode with honeycomb structure [99] has been used to study lateral diffusion and electrochemical properties of ubiquinone (coenzyme Q_{10}), a major protein involved in electron transport chains in mitochondrion and chloroplast membranes [100].

Very recently, Ataka et al. [101] have devised a method for oriented incorporation of a membrane protein by reconstituting the membrane directly at the surface of a gold electrode after immobilising the protein. In this system, a recombinant membrane protein (cytochrome c oxidase), solubilised by means of a detergent and carrying a histidine tag at its C-terminus, is immobilised on a gold surface functionalised by an Ni-NTA chelator (see Fig. 2.30a). The membrane is then reconstituted by adding mixed micelles of

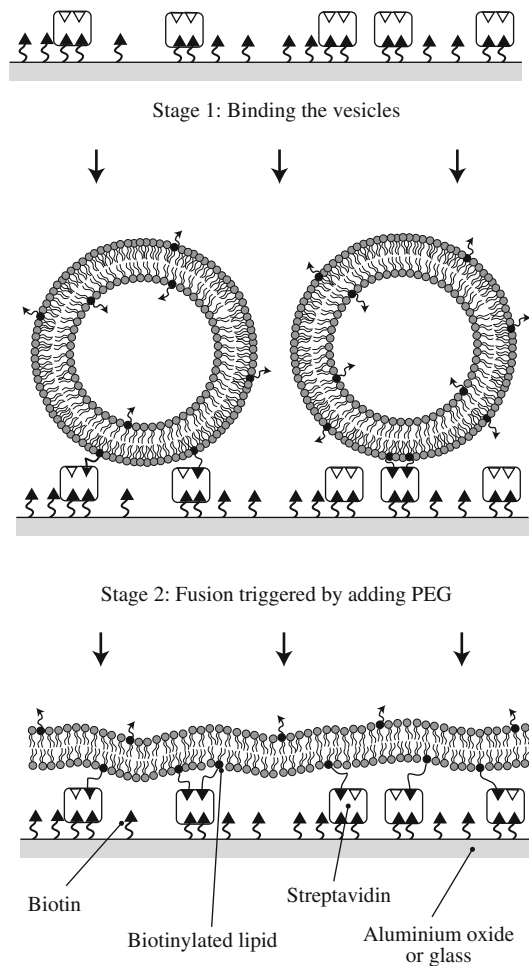


Fig. 2.29. Schematic view of the formation of a suspended membrane using an avidin/biotin system. (1) Immobilisation of phospholipid vesicles incorporating a biotinylated lipid on a streptavidin film previously formed on an aluminium oxide or glass surface functionalised by biotin. (2) Addition of polyethylene glycol (PEG) induces rupture and fusion of the membranes of the immobilised vesicles to form a homogeneous and contiguous bilayer. From Berquand et al. [98]

dimirystoylphosphatidylcholine (DMPC) and detergent in the presence of Bio-Beads. The latter are porous microbeads able to bind detergents with a strong affinity. By eliminating the detergent molecules in the mixed micelles, the phospholipid molecules are compelled to form a bilayer around the immobilised protein (see Fig. 2.30b).

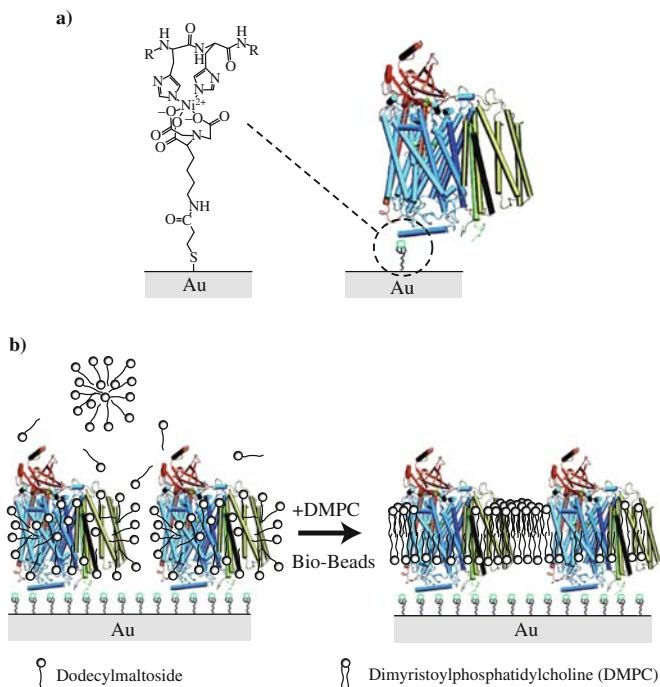


Fig. 2.30. Schematic view of the reconstitution of a suspended membrane incorporating a membrane protein, cytochrome c oxidase, immobilised by means of a Ni-NTA/His complex. **(a)** The recombinant cytochrome c oxidase with a histidine tag on its C-terminus is immobilised at the surface of a gold electrode modified by the chelator N-nitrilotriacetic acid (NTA), forming a metal–chelate complex with nickel (Ni-NTA) that can bind two imidazole nuclei of the histidine residues. **(b)** A lipid bilayer is formed by adding mixed micelles of dimyristoylphosphatidylcholine (DMPC) and dodecylmaltoside (detergent) in the presence of Bio-Beads. Elimination of the detergent by the Bio-Beads favours formation of the bilayer surrounding the protein. From Ataka et al. [101]

2.3.5 Bilayer Lipid Membranes (BLM)

Bilayer lipid membranes (BLM), originally called black lipid membranes and now sometimes called planar lipid bilayers, are obtained using a circular aperture of diameter 0.5 to 1 nm separating two aqueous compartments. These bilayers were first described by Mueller and coworkers in 1962 [102]. There are two techniques for making these membranes (see Fig. 2.31):

- In the first approach shown in Fig. 2.31a, a brush is immersed in a solution containing lipids dissolved in an organic solvent that does not mix with water, e.g., decane. A stroke of the paintbrush across an aperture in a Teflon wall separating two aqueous compartments causes a lipid film to form across the aperture. The film thickness decreases spontaneously with

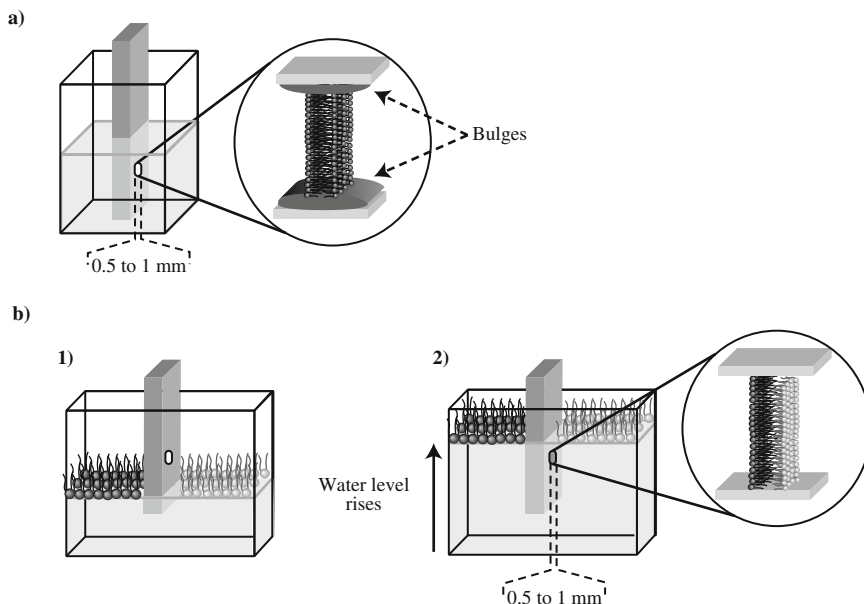


Fig. 2.31. Formation of planar lipid bilayers. These freely held bilayers are formed across an aperture in a wall separating two aqueous compartments. The diameter of the aperture is between 0.5 and 1 mm. (a) Paintbrush method. (b) Water level method

time and the excess lipid forms a bulge around the edge of the aperture. The film continues to thin down until a lipid layer has formed. The polar parts are in contact with the aqueous medium and the hydrocarbon chains are located inside the bilayer. However, the bilayer thereby formed still contains the hydrophobic solvent, giving rise to slight defects in the membrane structure.

- The second approach consists in forming a planar lipid bilayer from two monolayers made by spreading the molecules out and evaporating the solvent (see Fig. 2.31b). The monolayers are first formed on the surfaces of two solutions separated by the vertical wall, with the aperture above the surface level of the solutions (1). The level of the solutions is then raised above the aperture (2). This technique can also be used to obtain asymmetric bilayers, if the lipid composition of the two original monolayers is different. These films, which have two interfaces, are less than 10 nm thick. For this reason, they do not absorb visible light and appear black, which is why they were originally called black lipid membranes.

Planar lipid bilayers behave like biological membranes in which membrane proteins can be reinserted, in particular, intrinsic or integral membrane proteins which contain hydrophobic segments. This reconstitution is obtained

either by mixing the protein, solubilised by detergents, with the lipid solution, before formation of the bilayer, or by fusing protein-containing liposomes to a previously formed bilayer. In addition, by its very design, this model system provides two easily accessible aqueous compartments, in which electrodes can be immersed. As a consequence, it is perfectly suited to studying the ionic permeability of the bilayer. If the activity of the inserted protein induces temporary electrical changes, like a channel protein which causes a sudden variation in the impedance of the membrane by inducing a transmembrane ionic flux, changes in the current and/or potential differences on either side of the bilayer can then be recorded.

Planar lipid bilayers are widely used to investigate the molecular mechanisms underlying a great many membrane functions, such as membrane permeability and transport, ionic selectivity, signal transduction, electrical signal transmission, or photosynthesis, after reinserting different proteins, in particular ion channels, bacterial porins, β -adrenergic receptors, bacterial rhodopsins (intrinsic proteins in photosynthetic membranes), and cytochrome *c* (intrinsic protein in the respiratory chain of mitochondrion membranes) [103]. However, the big drawback with BLMs is their fragile structure, limiting their mechanical stability to around 8 h or less. In addition, they are very hard to manipulate and this limits technological applications. Current developments are thus aimed at forming bilayers on Teflon tubes with diameters of 0.5 to 1 mm, containing conducting metal wires (sBLM) or filled with hydrated polymers like those used in electrochemistry (sb-BLM) [104]. Under these conditions, the membrane is formed by rapidly immersing the end of the tube in a lipid solution. The tube, connected to a working electrode, is placed in an electrolyte solution in which a reference electrode has been immersed. It has been possible to improve the stability of planar bilayers, while preserving their electrical properties, by forming BLMs on this type of substrate [105]. In addition, this topography makes it easier to manipulate these thin films, which now have some applications in the design of biosensors and bioelectronic systems based on electrical detection methods.

For more information concerning planar lipid bilayers and their applications, the reader is referred to the literature, in particular, the reviews by Nikolelis et al. in 1999 [106] or Tien and Ottova in 2001 [103].

2.4 Protein–Lipid Assembly

Lipid membranes are self-assembled entities that can be used as substrates for the incorporation, immobilisation, self-association, or crystallisation of macromolecules which may or may not have specific biological activities. The functionalisation of biomimetic membranes to develop protein–lipid assemblies is a crucial step in many nanobiotechnological applications. This section aims specifically to discuss the association of proteins with lipid membrane systems of the kind described above.

2.4.1 Functionalising Langmuir–Blodgett Films

Langmuir–Blodgett lipid films are functionalised by association (or immobilisation) of proteins with specific recognition properties, such as enzymes, antibodies, or receptors, in order to develop supported, ordered protein–lipid molecular assemblies. These biomimetic nanostructures, corresponding to supramolecular arrangements, can be used to functionalise surfaces, upon which the protein confers its biospecificity.

Over the past twenty years, a lot of research has been carried out on the association of proteins, and in particular enzymes, with Langmuir–Blodgett films. The bioactive films obtained in this way have been studied for their potential application in the design of biosensors, with the protein–lipid LB layers integrated into these systems as ultrathin sensitive films [107]. Since they can be transferred to many different types of substrate, in particular, micronic substrates, these films exhibit many advantages for the development of novel micro- or nanobiosensors, inspired by biological models. As for other systems imitating natural membranes, their (highly ordered) structural organisation and their ultrathin dimensions (a few nanometers) are the main characteristics for designing micronic sensors operating on the molecular scale and displaying rapid response times (of the order of one second), fundamental criteria in the design of ‘intelligent’ sensors or biochips. However, Langmuir–Blodgett films are not only useful as a result of these structural features. The following possibilities are worth mentioning:

- The bioactive sensitive layer and its association with the transducer can be fabricated in a single stage.
- Only a very small amount of protein is required to prepare the membrane.
- They can be used at room temperature and pressure, hence avoiding the kind of thermal treatments required in the design of electronic systems, which would damage biological components.
- The performance of the sensor in terms of detection range and limits, and sensitivity, can be modulated by depositing different numbers of protein–lipid layers [108–113].

From a more general standpoint, protein–lipid membranes associated with sensors (or transducers) for biological signals can be used to study their functional properties. This association corresponds to a biomimetic simulation as close as one could hope to get to one of the main functions of biological membranes, namely, the recognition and transduction of biological signals. The direct contact between biological element and transducer allows a detailed study of its recognition properties and the resulting physicochemical modifications, providing information about the structure–function relations of biological membranes. If the protein is an enzyme, one can investigate its catalytic properties in a heterogeneous medium in a biomimetic lipid environment at the nanoscale. In particular, this type of study is relevant to the field of nanobioscience.

The crucial stage in the fabrication of these biomimetic membranes and the supported nanostructures they are used for is still the incorporation of the biological element in LB films, without alteration or loss of activity. Several methods have so far been developed to produce self-associated protein–lipid assemblies in bilayers or multilayers. The basic idea and the advantages and disadvantages of these specific methods based on the Langmuir–Blodgett technique are discussed in the following sections. The association of proteins, and especially enzymes, with Langmuir–Blodgett films, using the various techniques presented below, has recently been reviewed in [107], which discusses in particular the main points of interest of this type of membrane and its applications in nanobioscience.

Inserting Proteins in the Interfacial Monolayer Before Transfer to a Solid Substrate

The insertion of proteins in the monolayer at the air/water (or air/buffer) interface is one of the most commonly used methods. It derives from the procedures developed to study protein/lipid interactions with a Langmuir monolayer (see p. 55). Proteins present in the subphase are adsorbed onto the interfacial monomolecular film, before it is compressed and transferred to a solid substrate (see Fig. 2.32). It depends on the ability of the protein to associate with and integrate into a lipid leaflet. This method is particularly well suited to extrinsic peripheral proteins capable of associating with biological membranes or to anchoring proteins inserting themselves into one leaflet of a bilayer.

This insertion method does have some drawbacks, however. It often requires a relatively large amount of proteins since the volume of even the smallest LB trough may be as much as a few tens of millilitres, and the protein concentration must be relatively high in order to accelerate the penetration kinetics and avoid mechanical instability problems of the kind manifested by compressed films. Furthermore, if the monolayer is prepared on an aqueous phase which already contains the dissolved protein, a layer of inactive denatured protein, which will eventually be associated with the film, may form at the air/buffer interface due to the extremely high surface tension at the surface of a pure liquid (see Fig. 2.14 and associated discussion) [114]. The denaturation problem is also encountered when proteins penetrate weakly compressed films (generally less than 5 mN/m). Likewise, if the protein is injected directly under the monolayer, it can penetrate close to the point of injection and may not be uniformly distributed throughout the film. Gradual injection of the protein under the protein or beyond the compression barriers, as shown in Fig. 2.32, may provide a solution to the problem of the heterogeneity of the protein–lipid film at the buffer surface.

To avoid denaturing the protein at the interface, one possibility is to spread the protein or protein/detergent mixtures directly on a previously formed lipid film, which leads to the formation of a homogeneous, mixed interfacial

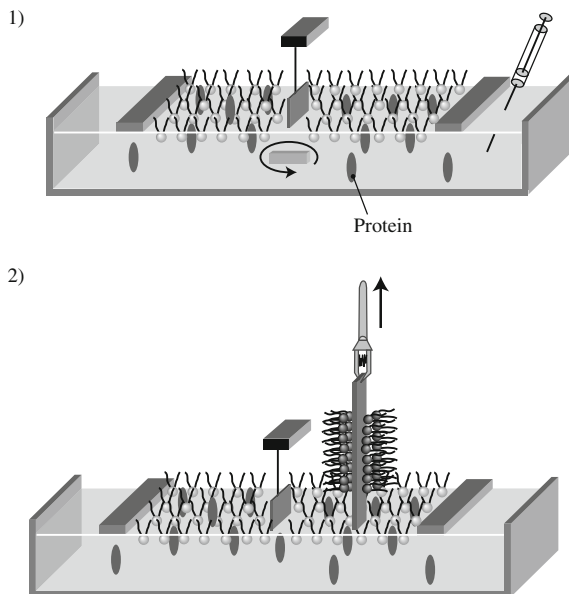


Fig. 2.32. Inserting proteins in the interfacial monolayer before Langmuir–Blodgett transfer

monolayer [108, 111, 114]. The formation of complexes between the protein and dialkylated synthetic amphiphilic molecules, soluble in solvents that are immiscible with water (such as benzene or chloroform), has also been exploited to form the mixed monolayer before transfer [115, 116]. Another approach recently reported is to mix the protein with the lipids before spreading the two together [117–120].

Finally, the presence of the protein in the interfacial film may affect its aptitude for transfer [27, 112, 121]. The minimum surface pressure imposed when lifting the monolayer may not be well suited to the conditions for insertion of the protein. High surface pressures usually lead to the protein being expelled from the film. In addition, the adhesion of the film on the substrate may be reduced by presence of the protein, and this can prevent the formation of bilayers by the first layer spreading back over the interface on the downstroke (see Sect. 2.3.3).

Protein Association on Previously Formed LB Lipid Films

Another immobilisation approach consists in adsorbing the protein on the surface of an already formed LB film. The main advantage with this procedure lies in the possibility of associating the protein with a hydrophilic lipid surface (polar head at the surface) or a hydrophobic lipid surface (hydrocarbon chains at the surface), depending on the number of layers deposited on the substrate (see Sect. 2.3.3), which allows one to control the lipid environment of the

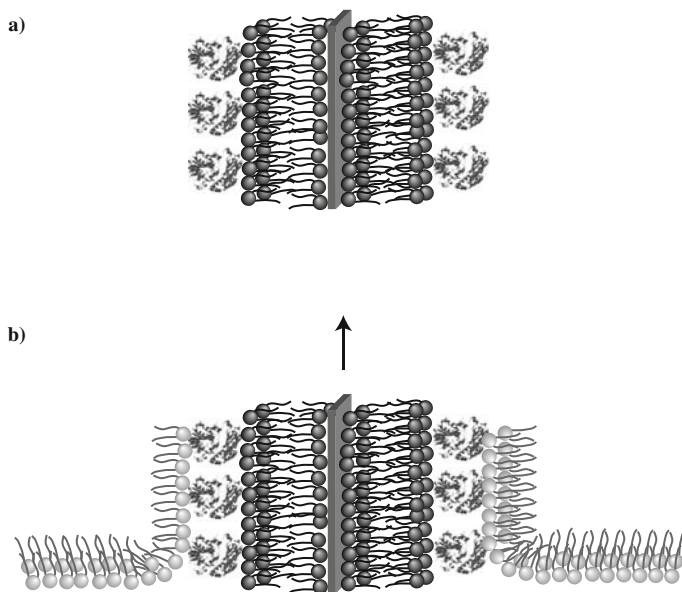


Fig. 2.33. Inclusion of proteins in Langmuir–Blodgett films. After adsorbing the protein on previously formed films (a), the protein molecules are coated with a further lipid layer (b). The protective leaflet may have a different lipid composition to the supporting layers

protein. It also means that one can control the thickness and homogeneity of the lipid substrate receiving the biological macromolecules.

The interactions involved in this type of association are sometimes too weak to prevent the protein from being released again, and this is a major cause of lack of stability and/or lack of repeatability when preparing this type of membrane, or when carrying out subsequent procedures with these assemblies. In order to minimise the desorption of protein molecules, some authors have suggested immobilising the protein adsorbed on the surface of LB films in a covalent manner by means of cross-linking agents [122, 123]. However, covalent immobilisation destroys the self-associating nature of the biomimetic membrane. Another idea is to stabilise, after formation of the protein–lipid LB films, by cross-linking in the presence of vapours of an agent like glutaraldehyde [121, 124]. The fact remains, however, that covalent association of the protein with the lipid structure can induce conformational changes which may cause it to lose its biological activity.

Another alternative for limiting desorption and avoiding covalent immobilisation of the protein consists in covering the protein molecules by transferring a further lipid layer onto the surface of the adsorbed molecules (see Fig. 2.33). This procedure, also called inclusion, serves to hold the protein in a totally hydrophilic or totally hydrophobic environment, while preserving the

homogeneity of the supported layers [125]. It is also possible to modify the lipid composition of the protective leaflet [126] and, to some extent, to reproduce the membrane asymmetry which can favour the physical retention of the protein and preserve its biological activity. In addition, the possibility of independently controlling the transfer and adsorption stages may be an advantage, particularly for overcoming the difficulties sometimes encountered during the transfer of mixed protein–lipid interfacial films (see the last section).

Finally, an appropriate choice of composition for the lipid film on which the protein is to associate may significantly reduce its tendency to detach itself. Monolayer studies of the affinity of the protein for a given type (or mixture) of lipids (see the caption of Fig. 2.18) may help to determine the best suited lipid matrix for immobilising the protein. Furthermore, an apposite choice of composition for the ‘acceptor’ leaflet may also limit denaturation problems of the kind sometimes encountered when proteins are adsorbed onto hydrophilic surfaces in which the surface tension is too strong (choice of polar headgroup, charged or otherwise).

Oriented Insertion of Proteins in LB Films

In the last two sections, we discussed the functionalisation of Langmuir–Blodgett films by association of proteins before or after transferring the lipid membrane. Now in both cases, these methods bind the associated protein randomly to the membrane. One of the great challenges in the development of ordered protein–lipid assemblies and functionalised biomimetic membranes is to control the orientation of the associated protein, just as it is in biological membranes where the binding of the protein on (or in) the bilayer determines its own orientation for optimal functionality.

In order to control the orientation of the protein associated with lipid membranes in general, and Langmuir–Blodgett films in particular, several strategies have been developed independently. One of these is the covalent coupling of the antigen binding fragment (or Fab) of an antibody, i.e., the fragment endowed with specific antigen recognition properties, via a disulfide bridge on the polar headgroup of a linker lipid inserted into a lipid monolayer [127–129]. This monolayer is directly transferred onto the solid substrate by horizontal Langmuir–Schaefer transfer (see Fig. 2.26). The supported membrane thereby obtained is then composed of a membrane half-leaflet.

Another approach for achieving oriented immobilisation of a protein, viz., myoglobin, involves binding its histidine residues, naturally present at the protein surface, onto a monolayer of a lipid with its polar headgroup functionalised by an iminodiacetate (IDA) group with chelating properties. In this way, several orientations, defined by the spatial distribution of the histidine residues, have been obtained [130]. The possibility of immobilising an anchoring protein in the interfacial monolayer before transferring it to the solid substrate can solve the problem of multiple orientation. The unique orientation of the protein is then guaranteed by inserting its anchor into the

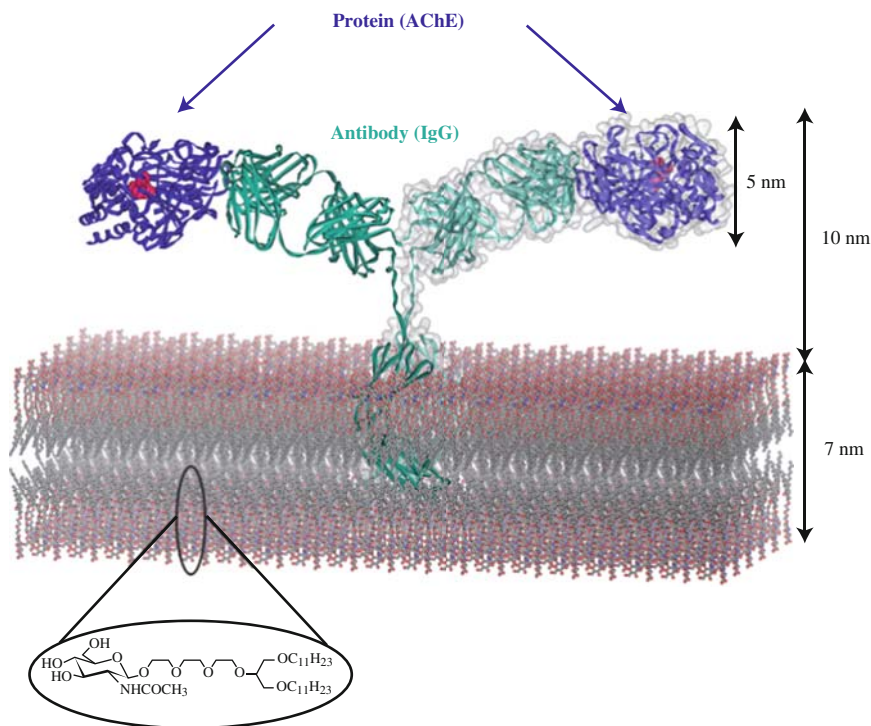


Fig. 2.34. Structural model of a functionalised biomimetic membrane for the oriented immobilisation of proteins. This membrane was obtained by combining two techniques exploiting the self-assembly properties of biological molecules (protein lipid vesicles spread at the air/buffer interface and the Langmuir–Blodgett technique), resulting in preferential orientation of the antibody (or immunoglobulin G, IgG) in the transferred bilayer (see text). The lipid bilayer is made from a synthetic neoglycolipid with fluid hydrocarbon chains at room temperature (gauche conformation). The antibody is held in the bilayer by what are assumed to be sugar–sugar-type interactions between the glycosylated residues of the protein and the polar headgroups of the glycolipid, and hydrophobic interactions between the Fc fragment of the antibody (a region rich in aliphatic residues) and the hydrocarbon chains of the glycolipid. In this model, the protein (a monomer of acetylcholinesterase, AChE) is associated with the functionalised lipid bilayer after LB transfer to a solid substrate, by specific recognition of the non-inhibitor monoclonal antibody. This structurally stable biomimetic membrane can maintain the protein activity over several months. From Godoy et al. [135]

lipid monolayer [120]. However, although this very attractive method is as biomimetic as one could hope, it only works for a well-defined class of proteins.

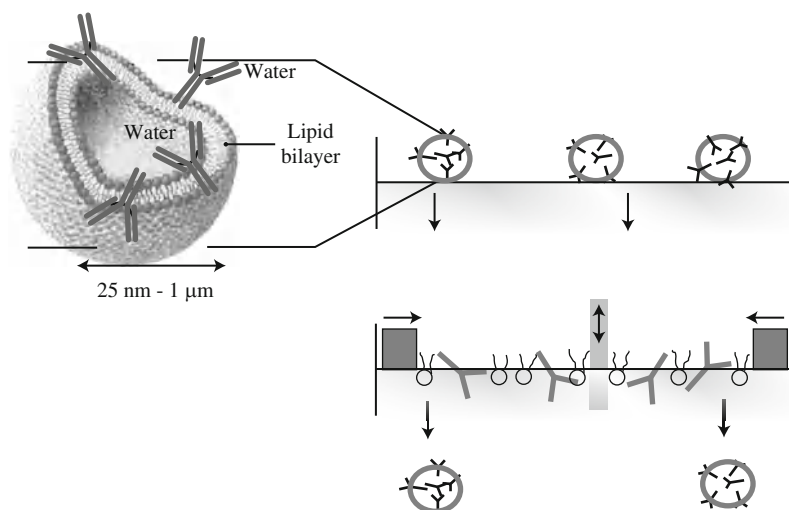


Fig. 2.35. Formation of a mixed interfacial monolayer by bursting protein–lipid vesicles at the air/buffer interface. This method exploits the instability of the protein–lipid structures, which open at the interface under the effect of surface tension forces to form a stable monomolecular film. The monolayer is compressed to maintain the lateral ordering of the molecules. The compressed monolayer can be transferred vertically to a solid substrate by the Langmuir–Blodgett technique

With the aim of developing functionalised biomimetic membranes with unique recognition sites for oriented binding of the protein, another strategy has been proposed recently. The idea is to insert a monoclonal antibody that does not inhibit biological activity in the Langmuir–Blodgett lipid bilayers. The antibody serves as an anchor to tether the protein in an oriented position at the membrane surface (see Fig. 2.34) [131]. The nanostructures obtained are polyvalent and the nature of the protein that is retained is defined by the specificity of the inserted antibody.

In this approach, the functional insertion of the antibody in the lipid was achieved by targeted transfer of immunoglobulin (soluble amphiphilic protein) in the interfacial lipid film before deposition on the solid substrate, using a suitable combination of two techniques: spreading protein–lipid vesicles at the air/buffer interface and then applying the Langmuir–Blodgett technique. This procedure thus exploits the possibility of forming a monomolecular mixed film at the surface of a buffer solution using surface tension forces able to destructure the membranes of a weakly stable protein–lipid vesicle [132, 133]. After compression, the mixed film is transferred by vertical Langmuir–Blodgett (LB) transfer (see Fig. 2.35) [134].

The point about forming protein–lipid vesicles before forming the monolayer is that interactions can then be set up by self-association between the lipid molecules and the antibodies in the vesicle membranes to improve

insertion of the antibody in the interfacial film and hence transfer the film without ejection of the protein. The vesicles are thus used as vectors for carrying the antibody directly to the air/buffer interface in a lipid environment. By combining these two techniques, i.e., liposomes and the LB technique, exploiting the self-association properties of biological molecules, the orientation of the antibody in the liposome membrane can be predetermined, and this orientation will be preserved when the liposomes open at the interface. The organisation of the protein–lipid film is then maintained by lateral compression of the monolayer.

From a structural point of view, these nanostructures are stable and remain functional for several months. After immuno-association of a model enzymatic protein (acetylcholinesterase, involved in the neurotransmission of the nerve influx), they have been used to study the kinetic behaviour of the enzyme in a lipid environment on a membrane with nanometric dimensions. The results clearly demonstrated a catalytic behaviour of the enzyme characteristic of immobilised enzymes, with the marked effects of diffusion constraints (for high substrate concentrations) in the microenvironment of the enzyme [135]. Hence these functional nanostructures, which allow an oriented binding of the enzyme at the surface of the lipid bilayer by offering a similar topography to the one found in biological membranes, represent a great potential for applications in the field of enzymology in structured media. Indeed, the behaviour of catalytic enzymes immobilised by a lipid bilayer is fully representative of enzyme biocatalysis of the kind observed in cells; most cell enzymes are held in membranes and are found *in vivo* in a phospholipid environment that is absolutely necessary for them to function correctly. Likewise, some proteins, called peripheral proteins, can associate with a membrane in a temporary manner, e.g., via electrostatic interactions with the polar parts of the lipids (or membrane proteins), and this association is essential for them to function in an optimal way.

Finally, these supramolecular assemblies can also be used to functionalise surfaces and to develop biomimetic sensors exploiting sensitive layers structured on the nanoscale (see Sect. 2.5.1).

2.4.2 Two-Dimensional Organisation of Proteins on Lipid Surfaces

Biological membranes provide a natural 2D space in which proteins – both extrinsic and intrinsic – can diffuse, interact and organise themselves in a highly ordered way. Several biological systems involve 2D ordered assemblies of proteins. One example is bacterial rhodopsin, a proton pump induced by light and prototype of seven-helix α -transmembrane proteins, which form almost perfect 2D crystals in the membranes of halophilic bacteria [136]. Other examples are provided by the S-layers which cover the surfaces of many species of bacteria in the form of 2D protein lattices and whose exact function is still being debated [137], or again the regular assemblies of cadherin molecules

forming adherens junctions and desmosomes [138]. The ability of certain proteins to form 2D crystals generally depends on the presence on their surface of groups able to establish intermolecular interactions, and also geometric features related to shape complementarity.

Historically, the use of lipid membranes as substrates for 2D crystallisation of proteins arose from structural studies carried out by transmission electron microscopy (TEM), originally on bacterial rhodopsin [139]. Indeed, by analysing TEM images of 2D protein crystals, the structure of the proteins can be determined at medium (~ 1 nm) or high (~ 0.4 nm) resolution [140]. This approach, called 2D electron crystallography by analogy with X-ray crystallography of 3D crystals formed from macromolecules, is a field of structural biochemistry in its own right [141, 142].

The study of molecular organisation in two dimensions, or in 2D multilayers, is of great interest on the fundamental level, because restricting to 2D can lead to quite different behaviour from 3D systems [143, 144]. Furthermore, there are many potential applications for controlled 2D organisation, from nanoelectronics to μ -array systems.

There are three strategies for 2D crystallisation of proteins using lipid surfaces, depending on whether they are membrane proteins reconstituted in lipid bilayers or soluble proteins crystallised in lipid monolayers or bilayers (see Fig. 2.36). The reconstitution of membrane proteins in lipid bilayers by elimination of a detergent, following an analogous process to the one used for LUV formation, described on p. 47 ff, does lead to the formation of 2D crystals in many cases. Several detailed reviews have been devoted to this question [140, 145]. The understanding acquired from these developments is now applied to the design of biosensors containing membrane proteins reconstituted in suspended lipid bilayers [101, 146], a very active area of nanobiotechnology (see Sects. 2.3.4 and 2.4.3).

Two-Dimensional Crystallisation of Soluble Proteins in Lipid Monolayers

The 2D crystallisation of soluble proteins in lipid monolayers at the air/water interface is a general method that has been widely used in structural biology [147, 148]. The method works as follows (see Fig. 2.36):

1. A lipid monolayer, containing lipids with a ligand group of the relevant protein (macromolecule) in their polar headgroup, is formed at the air/water interface.
2. The proteins, circulating in the aqueous phase, bind to the lipid ligands by molecular recognition.
3. The protein-lipid complexes diffuse and concentrate in the plane of the monolayer.
4. Depending on the nature of the interactions between the proteins, the latter arrange themselves at the interface in the form of disordered or crystalline 2D assemblies.

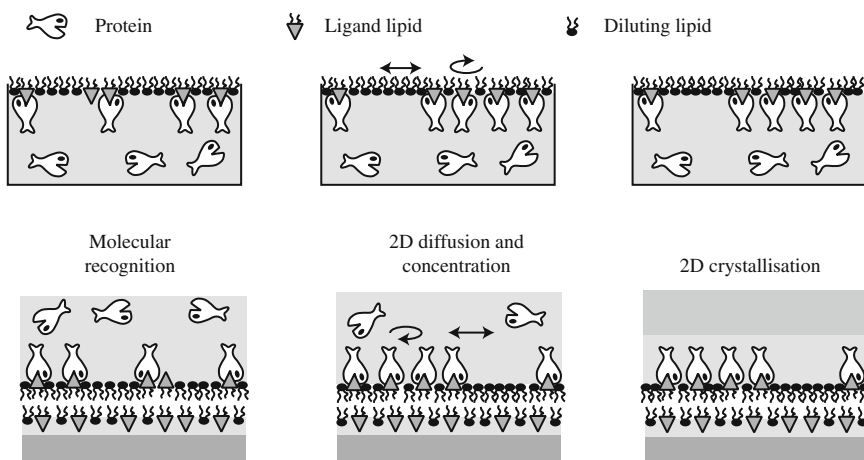


Fig. 2.36. Two-dimensional crystallisation of biological macromolecules by affinity in lipid monolayers at the air/liquid interface (*upper*) and in supported lipid bilayers at the solid/liquid interface (*lower*)

The resulting systems can be classified in terms of the lipid ligands used for the 2D crystallisation of the proteins [148, 149]:

- Natural lipids, e.g., ganglioside G_{M1} , a physiological receptor of the cholera toxin [150], or phosphatidylserine, ligand of the annexin A5 [151].
- Synthetic lipids obtained by grafting a soluble ligand of a protein onto a lipid [152]. The classic example is provided by lipids coupled with biotin, a natural ligand of streptavidin [153]. The system in which lipids carry an Ni-NTA (nitrilotriacetic acid) group, a ligand with affinity for polyhistidine tags [154, 155], is worth a special mention owing to the widespread use of histidine tags for purifying recombinant proteins.
- Positively or negatively charged. RNA polymerases [156] or 50S subunits of ribosomes [157] have been crystallised in this way in monolayers containing positively charged fatty acids and negatively charged phosphatidylserine, respectively.

Determining factors in the 2D crystallisation of proteins in lipid monolayers are the 2D limitation of protein diffusion, and the orientation and concentration of the proteins in a plane. Hence, in the case of annexin A5, the formation of trimers and 2D trimer lattices results directly from the oriented binding of the molecule on a plane surface, allowing the formation of salt bridges between residues of opposite charges which find themselves opposite one another. These same interactions must exist between the molecules in solution, but they are not strong enough to allow the formation of oligomers and stable 2D assemblies.

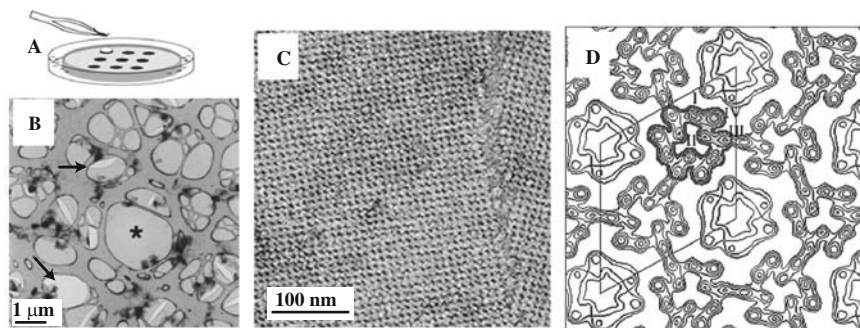


Fig. 2.37. Two-dimensional crystallisation of soluble proteins on a lipid monolayer. Adapted from Brisson et al. [158]. Copyright Elsevier (1999). (A) Crystallisation well in a Teflon substrate. A drop of protein solution ($\sim 20 \mu\text{L}$) is deposited in each well and covered with a lipid droplet ($\sim 1 \mu\text{L}$). After a certain incubation time (a few minutes to hours), the interfacial film is lifted by touching its surface with the grid of the electron microscope (EM) coated with a holey carbon film. (B) EM grid coated with a protein-lipid film after transfer. The homogeneity of the film signals the high quality of the transfer (*). Some squares of the grid mesh are broken (arrows). (C) 2D streptavidin crystals formed on a monolayer of biotinylated lipids. (D) 2D projection of an annexin A5 crystal calculated by EM image analysis [159]

The sequence of operations leading from 2D crystal formation to the analysis of TEM images is illustrated in Fig. 2.37. The interfacial films comprising a lipid monolayer firmly associated with a monolayer of proteins assembled into a crystalline mosaic constitutes a particular type of Langmuir film. The potential applications of these films have hardly been investigated as yet.

Several variants have been developed in the 2D crystallisation of soluble proteins in lipid monolayers. One of these uses lipids forming assemblies with tubular morphology as a matrix for the helical crystallisation of proteins. Some lipids, natural or synthetic, have the property of forming tubular assemblies [160]. This property, which depends among other things on the chirality of the polar groups, is also found in functionalised lipids. Lipid tubes containing biotinylated lipids have thus been used as a matrix support for helical crystallisation of streptavidin [161]. An analogous approach consists in incorporating lipids functionalised by Ni-NTA groups within lipids that naturally form into tube structures. The helical self-assembly of proteins carrying polyhistidine tags has been reported in several protein systems [162].

A second variant involves the 2D crystallisation of membrane proteins in lipid monolayers [163]. Membrane proteins carrying polyhistidine tags bind to lipid monolayers containing complementary Ni-NTA groups, in detergent concentrations that pose no threat to the integrity of the monolayer. Elimination of the detergent in the presence of lipids leads in certain cases to the reconstitution of lipid bilayers within which the proteins have 2D order. These

potentially very useful approaches deserve further development in the context of membrane protein biosensors.

The transfer of protein–lipid films from the air/water interface to EM grids is a delicate operation and often difficult to control, which can damage the molecular arrangement of samples [158]. However, the main difficulties arising during transfer can be countered using holey (perforated) carbon films.

The main reason why 2D crystal formation and crystallography are not more widely used is that there is no fast screening method. The approach employed is TEM observation of samples with negative staining. Although this operation is simple enough, it is very time-consuming compared with assays for 3D macromolecular crystal formation where thousands of conditions can be tested robotically and where the screening method is even simpler, being based on observation by optical microscopy. The development of a simple optical method for assessing 2D crystal formation would represent a significant step forward.

Two-Dimensional Organisation of Proteins in Supported Lipid Bilayers

The advent of effective physicochemical and structural methods for analysing molecular processes on surfaces has stimulated the development of methods for preparing biological samples on solid substrates. With regard to 2D protein–lipid assemblies, there are two types of sample, depending on whether the assemblies already exist in solution and are subsequently deposited on a substrate or whether they are assembled *in situ* on the substrate. Two-dimensional crystals of membrane proteins belong to the first category. Over the last decade, an impressive number of structural studies of 2D crystals of membrane protein has been carried out, mainly using AFM [164,165]. The possibility of imaging membrane proteins forming disordered assemblies within native membranes has recently been demonstrated on retinal disks, and also on photosynthetic membranes [166,167]. These results open up a huge field of investigation for functional studies of membrane processes.

The second system concerns molecular structures formed *in situ*, in the context of dynamic and functional studies of 2D molecular assembly mechanisms. The possibility of AFM monitoring the dynamics of complex molecular processes in real time, with a resolution of a few seconds and a quasi-physiological buffer, has been widely exploited to investigate assemblies with a whole range of different 2D organisational states, from disorder to crystalline order.

Figure 2.38 presents a study of the protein distribution within model membranes, for the example of lipid rafts [95].

The above method of 2D crystallisation for soluble proteins in lipid monolayers at the air/liquid interface has been extended to lipid bilayers at the solid/liquid interface (see Fig. 2.39).

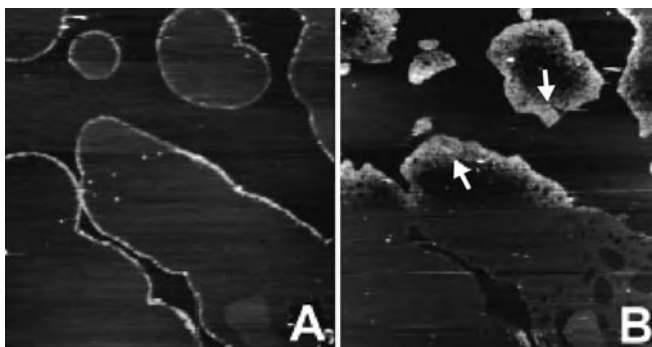


Fig. 2.38. Lateral distribution of alkaline phosphatase in supported bilayers. (A) SLBs containing a lipid mixture of lecithin and sphingomyelin exhibit a phase separation with sphingomyelin-rich regions of micrometric dimensions, a gel phase, slightly higher than the lecithin-rich regions. The molecules of alkaline phosphatase, a protein with a GPI (glycosylphosphatidylinositol) lipid anchor known to target raft domains, visible in the form of white dots, are mainly located around the edge of sphingomyelin-rich regions. (B) Incorporating cholesterol in the SLB induces a redistribution of the proteins which migrate toward the interior of the sphingomyelin- and cholesterol-rich regions. From Milhiet et al. [95]

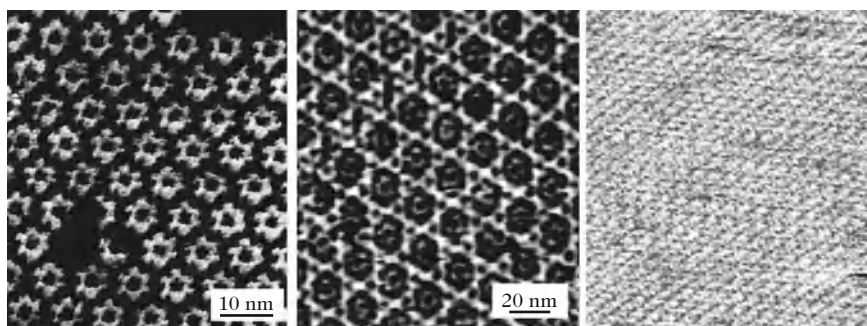


Fig. 2.39. AFM images of 2D crystals of proteins on an SLB. *Left:* α -hemolysin on a DOPC bilayer [92]. Copyright Elsevier 1995. *Center:* Annexin A5 on a DOPC:DOPS (4:1) bilayer [168]. Copyright Elsevier 1998. *Right:* Streptavidin on a lipid bilayer containing lipids carrying a biotin group [169]. Reproduced with kind permission of the American Chemical Society. Copyright 2001

A detailed analysis of the adsorption process and 2D organisation of proteins in supported bilayers has been carried out for annexin A5, establishing a quantitative relationship between these processes [52]. Combined quantitative data on protein adsorption from QCM-D and ellipsometry, and structural data on the various 2D organisational states from AFM, have provided a highly accurate description of binding, nucleation and crystal growth (see Fig. 2.40). The lessons to be drawn from this model system can certainly be generalised to other systems.

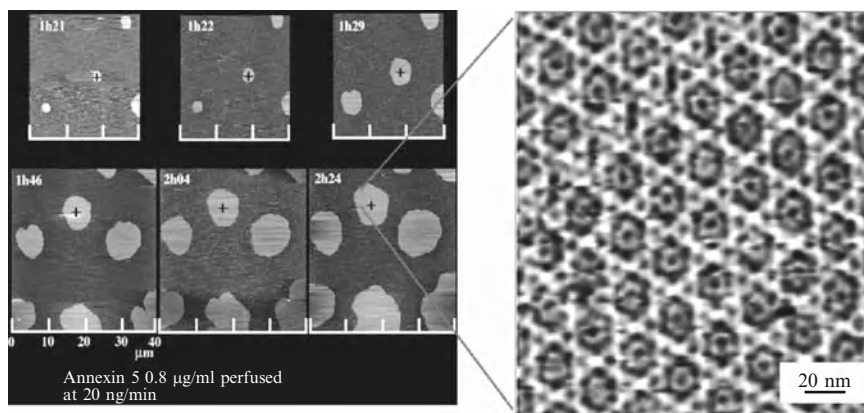


Fig. 2.40. Binding and 2D organisation of annexin A5, a soluble protein in the annexin family, on negatively charged lipid surfaces in the presence of calcium ions. The annexin A5 and streptavidin systems are the macromolecular systems for which the 2D ordering process has been most accurately characterised at the current time. From I. Reviakine and A. Brisson

The role played by the substrate in the 2D ordering of proteins has been revealed in a most striking way. For example, annexin A5 forms either 2D crystals or disordered close-packed assemblies, depending on whether the lipid bilayers are formed on a mica or a silica substrate. The reason for this different behaviour remains to be identified, another example of our incomplete understanding of the way the substrate affects the properties of molecules or molecular layers.

There are many potential nanobiotechnological applications for these systems, in the design of biosensors, different types of chip, or nanocarriers for drug targeting.

2.4.3 Reconstitution of Membrane Proteins in Supported Lipid Bilayers

Membrane proteins carry out many key functions in the life of a cell, and failure of the associated functional mechanisms can explain a good number of diseases. This is why so much effort has gone into identifying and synthesising new pharmacological agents which target membrane proteins. In this context, one of the main aims in biotechnology is to integrate membrane proteins into biosensors.

Most of the approaches that have been developed are based on the fabrication of lipid membranes suspended over a solid substrate, but decoupled and separated from it in such a way as to avoid direct interaction of the proteins with the substrate. Several recent studies report very encouraging results regarding the reconstitution of membrane proteins in suspended

membranes [101, 146] (see Sect. 2.3.4). Another approach involves controlling the orientation of membrane proteins in inorganic substrates by modifying the chemical properties of the surface [171].

The way now seems open for the development of biosensors and biochips incorporating membrane proteins.

2.5 Applications of Biomimetic Membranes in Nanobiotechnology

Biomimetic membranes and protein–lipid assemblies in general are the basic structures in many nanobiotechnological applications. The main examples are the fabrication of lipid nanoparticles, encapsulation and release of medicines, drug targeting, gene therapy, reconstitution of membrane proteins, composite assemblies, functionalisation of surfaces, and the development of nanobiosensors, but also as substrates in the 2D crystallisation of proteins.

The aim in this section is to present several recent examples of biomimetic structures and membranes in the field of nanobiotechnology.

2.5.1 Bio-Optoelectronic Micro- and Nanosensors

The first example application of biomimetic membranes concerns the design of bio-optoelectronic micro/nanosensors. Biochemical sensors, or biosensors for short, are high-performance analytical tools, combining the specific recognition capacity of a sensitive biological element, the bioreceptor, with the sensitivity of the (electro)chemical, physical, or optical sensor, the transducer. The latter detects physicochemical changes generated by the bioreceptor upon contact with the target substance and translates them into a measurable and interpretable electrical signal (see Chap. 16). The performance of a biosensor is closely linked to the properties of the sensitive layer and the quality of its association with the transducer. Current developments follow the marked trend toward miniaturisation of recognition systems, to allow analysis of a single molecule. Molecular scale patterning of the sensitive layer is therefore a crucial step in the miniaturisation of these biosensors. The development of new biospecific substrates like ordered thin films, directly deposited on a transducer, constitutes a key stage in the evolution of biosensors toward integrated micro- or nanosensors and toward the design of biochips, which apply the idea of miniaturised biosensor arrays.

In this context, self-assembled supramolecular assemblies of nanometric dimensions can be developed by the controlled synthesis of ordered molecular systems using biomolecules as the elementary structure. Ordered biomimetic membranes associating the relevant biological element in an oriented way can be used as an interface with a miniaturised micro- or optoelectronic system. Surfaces can then be functionalised by integrating biochemical functions. The development of new bioelectronic hybrids based on biological models opens a

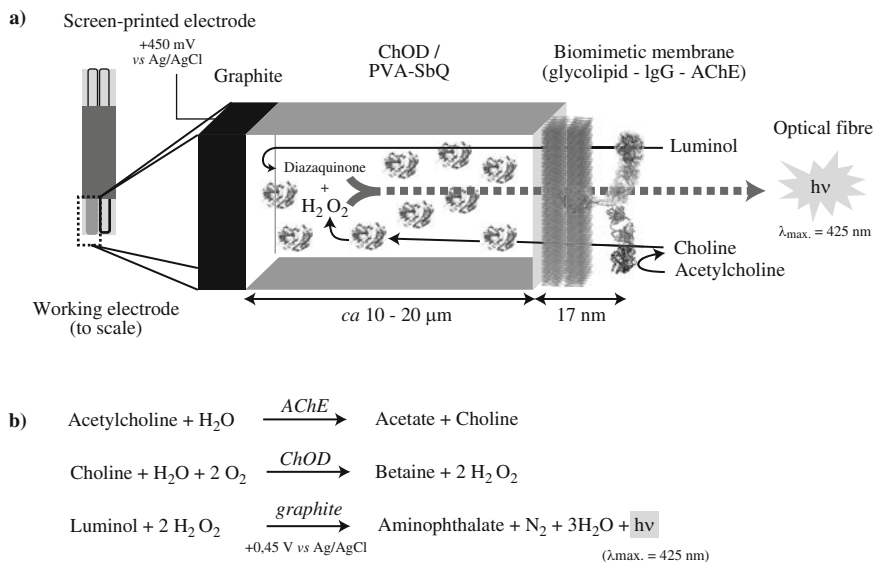


Fig. 2.41. (a) Bio-optoelectronic microsensor. This sensor has been obtained by direct transfer of a biomimetic Langmuir–Blodgett membrane orienting acetylcholinesterase at the surface of a screen-printed electrode. (b) Reaction sequence for detecting acetylcholine in the reaction medium. Adapted from Godoy et al. [172]

vast field of investigation for the development of new miniaturised biomimetic sensors and analytical tools operating on the molecular level.

With the aim of designing a new miniaturisable bio-optoelectronic sensor, a lipid membrane obtained by the Langmuir–Blodgett technique and associating acetylcholinesterase (AChE) in an oriented way (as described on p. 82 and shown in Fig. 2.34) has been combined with a high-performance optical sensor. The latter exploits the electrochemiluminescence reaction of luminol (see Fig. 2.41) [172]. (See also Chap. 16 on biosensors for more about the chemiluminescence reactions of luminol and their application to high-performance biochemical sensors.) Very briefly, acetylcholinesterase catalyses choline formation from acetylcholine present in the reaction medium. The choline is then oxidised by choline oxidase immobilised in a photopolymer of poly(vinyl alcohol) (PVA-SbQ) at the surface of a screen-printed electrode. This produces hydrogen peroxide (H_2O_2). In the presence of electro-oxidised luminol, the hydrogen peroxide induces light emission. The light signal is then detected via an optical fibre connected to the photomultiplier tube of a light meter.

This new type of sensor combines the advantages of using a biomimetic membrane as sensitive layer with those of using screen-printed electrodes for the electrochemiluminescence reaction of luminol:

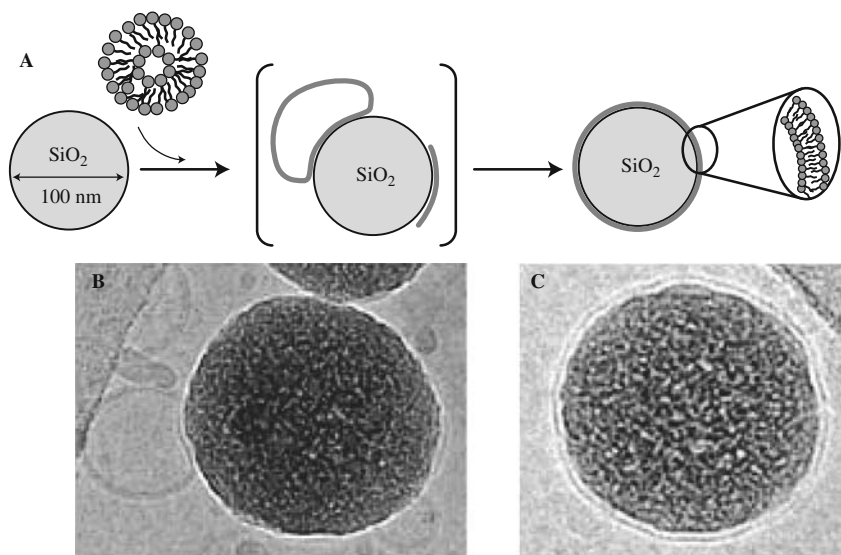


Fig. 2.42. Deposition of lipid vesicles around silica nanoparticles. From Mornet et al. [173]. (A) Formation of the lipid bilayer around silica nanoparticles. (B) Cryo-EM image of a vesicle adsorbed on a silica nanoparticle (diameter 100 nm). The vesicle is deformed and faithfully adapts to the asperities of the particle surface. (C) Silica nanoparticle completely coated with a supported bilayer

1. The membrane is ordered at the molecular level, so there is hope for miniaturising the analysed region.
2. One can have a functional orientation of the enzyme at the membrane surface, something that is not always possible with the usual methods for immobilising enzymes (see Chap. 16).
3. The membrane, exploiting the specific recognition properties of a non-inhibitor monoclonal antibody, is multipurpose, whence different enzymes could be immobilised there by changing the antibody.

Since the detection system is triggered by hydrogen peroxide, it can be applied to many oxidases able to detect a range of different metabolites of medical, industrial, or pharmaceutical interest. Environmental applications are also envisaged. Final, the intimate contact between the different enzyme layers means that one can directly detect products generated during the enzyme reaction sequence, without loss in the reaction medium (the product generated by the first enzyme concentrates in the vicinity of the second, i.e., in its microenvironment). This favours the flow of metabolites toward the detection device, avoiding back-scattering into the reaction medium and increasing sensitivity. It is the ultrathin dimensions of the biomimetic membrane that leads to the high performance of this sensor, especially in terms of response time.

The association of Langmuir–Blodgett biomimetic membranes with high-performance optical sensors illustrates the way in which ‘natural’ patterning

of the sensitive layer by self-association of biomolecules can be combined with surface functionalisation in the design of miniaturised bio-optoelectronic sensors.

2.5.2 Composite Assemblies

The development of functionalised nanoparticles is a central concern in nanobiotechnology, both in detection applications (quantum dots, magnetic nanoparticles in medical imaging) and in drug delivery. The functionalisation of inorganic nanoparticles has the double role of stabilising the nanoparticles against their natural tendency to flocculate and rendering them biocompatible. All approaches so far developed and optimised for large substrates ($\sim 1\text{ cm}^2$) can in principle be applied to nanoparticles.

For example, the formation of supported lipid bilayers around silica nanoparticles obeys the general rules established for silica substrates (see Fig. 2.42) [173]. This study also showed that cryo-EM, the method used up to now to study biological assemblies and polymer aggregates, is particularly well-suited to the characterisation of functionalised nanoparticles. Cryo-EM can reveal the nanoscale structure of supramolecular complexes in aqueous solution and it should therefore play a key part in the field of nanobiotechnology.

References

1. omega.dawsoncollege.qc.ca/ray/cellmemb/membrane.jpg
2. Shechter, E.: *Biochimie et biophysique des membranes. Aspects structuraux et fonctionnels*, Masson, Paris (1993)
3. Lehninger, A.L., Nelson, D.L., Cox, M.M.: *Principes de biochimie*, 2nd edn., Médecine-Sciences, Flammarion (1994) p. 87, Chap. 4
4. bravo436.chem.mcgill.ca/groups/eisenberg/morphologies/
5. Bangham, A.D., Horne, R.W.: *J. Mol. Biol.* **8**, 660 (1964)
6. Düzüğnes, N. (Ed.): Liposomes, parts A–E, *Methods in Enzymology*, Vols. 367, 372, 373, 387, 391, Elsevier Academic Press (2004)
7. Gregoriadis, G.N.: *Engl. J. Med.* **295**, 704; 765 (1976)
8. Düzüğnes, N. (Ed.): Liposomes, part D, *Methods in Enzymology*, Vol. 387, Elsevier Academic Press (2004)
9. Düzüğnes, N. (Ed.): Liposomes, part E, *Methods in Enzymology*, Vol. 391, Elsevier Academic Press (2005)
10. Lasic, D.D.: *Liposomes in Gene Delivery*, CRC Press (1997)
11. Robertson, J.D.: *J. Biophys. Biochem. Cytol.* **3**, 1043 (1957)
12. Dubochet, J., Adrian, M., Chang, J.-J., Homo, J.-C., Lepault, J., McDowell, A.W., Schultz, P.: *Q. Rev. Biophys.* **21**, 129 (1998)
13. Rigaud, J.-L., Pitard, B., Levy, D.: *Biochim. Biophys. Acta* **123**, 223 (1995)
14. Rouser, G.S., Fleischer, R., Yamamoto, A.: *Lipids* **5**, 494 (1970)
15. Szoka, F., Papahadjopoulos, D.: *Proc. Natl. Acad. Sci. USA* **75**, 4194 (1978)
16. Dimitrov, D.S., Angelova, M.I.: *Bioelectrochem. Bioenerg.* **19**, 323 (1988)

17. Gaines, G.L. Jr.: *Insoluble Monolayers at Liquid-Gas Interfaces*, Wiley Interscience, New York (1966)
18. Birdi, K.S.: *Lipid and Biopolymer Monolayers at Liquid Interfaces*, Plenum Press, New York (1989)
19. Robert, G.G.: *Langmuir-Blodgett Films*, Plenum Press, New York (1990)
20. Ulman, A.: Langmuir-Blodgett films. In: *An Introduction to Ultrathin Organic Films from Langmuir-Blodgett to Self Assembly*, Academic Press, New York (1991)
21. Petty, M.C.: *Langmuir-Blodgett Films. An Introduction*, Cambridge University press, Cambridge (1996)
22. Dynarowicz-Latka, P., Dhanabalan, A., Osvaldo, N.O. Jr.: Adv. Colloid Interface Sci. **91**, 221 (2001)
23. Vollhardt, D.: Adv. Colloid Interface Sci. **64**, 143 (1996)
24. Vollhardt, D.: Mater. Sci. Eng. C **22**, 121 (2002)
25. Fromherz, P.: Biochim. Biophys. Acta **225**, 382 (1971)
26. Verger, R., Pattus, F.: Chem. Phys. Lipids **30**, 189 (1982)
27. Zhu, D.C., Petty, M.C., Ancelin, H., Yarwood, J.: Thin Solid Films **176**, 151 (1989)
28. Demel, R.A.: Subcell. Biochemistry **23**, 83 (1994)
29. Brockman, H.: Curr. Opin. Struct. Biol. **9**, 438 (1999)
30. Marsh, D.: Biochim. Biophys. Acta **1286**, 183 (1996)
31. Maget-Dana, R.: Biochim. Biophys. Acta **1462**, 109 (1999)
32. Brian, A.A., McConnell, H.M.: Proc. Natl. Acad. Sci. USA **81**, 6195 (1984)
33. Sackmann, E.: Science **271**, 43 (1996)
34. Boxer, S.G.: Curr. Opin. Chem. Biol. **4**, 704 (2000)
35. Richter, R.P., Bérat, R., Brisson, A.R.: Langmuir **22**, 3497 (2006)
36. Nollert, P., Kiefer, H., Jahnig, F.: Biophys. J. **69**, 1447 (1995)
37. Reviakine, I., Brisson, A.R.: Langmuir **16**, 1803 (2000)
38. Richter, R.P., Mukhopadhyay, A., Brisson, A.R.: Biophys. J. **85**, 3035 (2003)
39. Jass, J., Tjarnhage, T., Puu, G.: Biophys. J. **79**, 3153 (2000)
40. Granéli, A., Rydström, J., Kasemo, B., Höök, F.: Langmuir **19**, 842 (2003)
41. Raedler, J., Strey, H., Sackmann, E.: Langmuir **11**, 4539 (1995)
42. Cremer, P.S., Boxer, S.G.: J. Phys. Chem. **103**, 1554 (1999)
43. Richter, R.P., Brisson, A.R.: Biophys. J. **88**, 3422 (2005)
44. Rossetti, F.F., Bally, M., Michel, R., Textor, M., Reviakine, I.: Langmuir **21**, 6443 (2005)
45. Keller, C.A., Kasemo, B.: Biophys. J. **75**, 1397 (1998)
46. Reimhult, E., Höök, F., Kasemo, B.: J. Chem. Phys. **117**, 7401 (2002)
47. Reimhult, E., Höök, F., Kasemo, B.: Langmuir **19**, 1681 (2003)
48. Rodahl, M., Höök, F., Krozer, A., Brzezinski, P., Kasemo, B.: Rev. Sci. Instrum. **66**, 3924 (1995)
49. Seantier, B., Breffa, C., Felix, O., Decher, G.: J. Phys. Chem. B **109**, 21755 (2005)
50. Bayerl, T.M., Bloom, M.: Biophys. J. **58**, 357 (1990)
51. Richter, R.P., Maury, N., Brisson, A.R.: Langmuir **21**, 299 (2005)
52. Richter, R.P., Lai Kee Him, J., Tessier, B., Tessier, C., Brisson, A.R.: Biophys. J. **89**, 3372 (2005)
53. Dahlin, A., Zäch, M., Rindzevicius, T., Käll, M., Sutherland, D.S., Höök, F.: J. Am. Chem. Soc. **127**, 5043 (2005)

54. Blodgett, K.B.: J. Am. Chem. Soc. **57**, 1007 (1935)
55. Blodgett, K.B., Langmuir, I.: Phys. Rev. **51**, 964 (1937)
56. Grundy, M.J., Musgrove, R.J., Richardson, R.M., Roser, S.J.: Langmuir **6**, 519 (1990)
57. Gennis, R.B.: *Biomembranes. Molecular Structure and Function*, Springer, New York (1989)
58. Hasmonay, H., Caillaud, M., Dupeyrat, M.: Biochim. Biophys. Res. Commun. **89**, 338 (1979)
59. Girard-Egrot, A.P., Morélis, R.M., Coulet, P.R.: Langmuir **12**, 778 (1996)
60. Girard-Egrot, A.P., Morélis, R.M., Coulet, P.R.: Langmuir **9**, 3107 (1993)
61. Tamm, L.K., McConnell, H.M.: Biophys. J. **47**, 105 (1985)
62. Devaux, P.F.: Biochemistry **30**, 1163 (1991)
63. Crane, J.M., Kiessling, V., Tamm, L.K.: Langmuir **21**, 1377 (2005)
64. Schneider, J., Dufrêne, Y.F., Barger Jr., W.R., Lee, G.U.: Biophys. J. **79**, 1107 (2000)
65. Schneider, J., Barger Jr., W.R., Lee, G.U.: Langmuir **19**, 1899 (2003)
66. Yuan, C., Johnston, L.J.: Biophys. J. **79**, 2768 (2000)
67. Yuan, C., Johnston, L.J.: Biophys. J. **81**, 1059 (2001)
68. Yuan, C., Furlong, J., Burgos, P., Johnston, L.J.: Biophys. J. **82**, 2526 (2002)
69. Singer, S.J., Nicolson, G.L.: Science **175**, 720 (1972)
70. Simons, K., Ikonen, E.: Nature **386**, 569 (1997)
71. Anderson, R.G.W.: Annu. Rev. Biochem. **67**, 199 (1998)
72. Sheets, E.D., Holowka, D., Baird, D.: Curr. Opin. Cell Biol. **3**, 95 (1999)
73. Okamoto, T., Schlegel, A., Scherer, P.E., Lisanti, M.P.: J. Biol. Chem. **273**, 5419 (1998)
74. Viola, A., Schroeder, S., Sakakibara, Y., Lanzavecchia, A.: Science **283**, 680 (1999)
75. Anderson, R.G.W., Jacobson, K.: Science **296**, 1821 (2002)
76. Rinia, H.A., Demel, R.A., van der Eerden Jan, P.J.M., de Kruijff, B.: Biophys. J. **77**, 1683 (1999)
77. Tredgold, R.H.: *Order in Thin Organic Films*, Cambridge University Press, Cambridge (1994)
78. Okamura, E., Umemura, J., Iriyama, K., Araki, T.: Chem. Phys. Lipids **66**, 219 (1993)
79. Lesieur, P., Barraud, A., Vandevyver, M.: Thin Solid Films **152**, 155 (1987)
80. Bassereau, P., Pincet, F.: Langmuir **13**, 7003 (1997)
81. Dufrêne, Y.F., Lee, G.U.: Biochim. Biophys. Acta **1509**, 14 (2000)
82. Benz, M., Gutschmann, T., Chen, N., Tadmor, R., Israelachvili, J.: Biophys. J. **86**, 870 (2004)
83. Cordero, S.R., Weston, K.D., Buratto, S.K.: Thin Solid Films **360**, 139 (2000)
84. Rinia, H.A., de Kruijff, B.: FEBS Lett. **504**, 194 (2001)
85. Gleiche, M., Chi, L.F., Fuchs, H.: Nature **403**, 173 (2000)
86. Lu, N., Gleiche, M., Zheng, J., Lenhert, S., Xu, B., Chi, L., Fuchs, H.: Adv. Mater. **14**, 1812 (2002)
87. Moraille, P., Badia, A.: Langmuir **18**, 4414 (2002)
88. Moraille, P., Badia, A.: Langmuir **19**, 8041 (2003)
89. Matuoka, S., Asami, H., Hattai, I.: Thin Solid Films **180**, 123 (1989)
90. Charitat, T., Bellet-Amalric, E., Fragneto, G., Graner, F.: Eur. Phys. J. B **8**, 583 (1999)

91. Hughes, A.V., Goldar, A., Gerstenberg, M.C., Roser, S.J., Bradshaw, J.: *Phys. Chem. Chem. Phys.* **4**, 2371 (2002)
92. Mou, J., Yang, J., Shao, Z.: *J. Mol. Biol.* **248**, 507 (1995)
93. Zahn, D., Brickmann, J.: *Chem. Phys. Lett.* **352**, 441 (2002)
94. Yuan, C., Johnston, L.J.: *J. Microscopy* **205**, 136 (2002)
95. Milhiet, P.E., Giocondi, M.-C., Baghdadi, O., Ronzon, F., Roux, B., Le Grimellec, C.: *EMBO Reports* **3**, 485 (2002)
96. Radler, U., Mack, J., Persike, N., Jung, G., Tampé, R.: *Biophys. J.* **79**, 3144 (2000)
97. Stora, T., Dienes, Z., Vogel, H., Duschl, C.: *Langmuir* **16**, 5471 (2000)
98. Berquand, A., Mazeran, P.-E., Pantigny, J., Proux-Delrouyre, V., Laval J.-M., Bourdillon, C.: *Langmuir* **19**, 1700 (2003)
99. Proux-Delrouyre, V., Laval, J.-M., Bourdillon, C.: *J. Am. Chem. Soc.* **123**, 9176 (2001)
100. Proux-Delrouyre, V., Elie, C., Laval, J.-M., Moiroux, J., Bourdillon, C.: *Langmuir* **18**, 3263 (2002)
101. Ataka, K., Giess, F., Knoll, W., Naumann, R., Haber-Pollmeier, S., Richter, B., Heberle, J.: *J. Am. Chem. Soc.* **126**, 16199 (2004)
102. Mueller, P., Rudin, D.O., Tien, H.T., Wescott, W.C.: *Nature* **194**, 979 (1962)
103. Tien, H.T., Ottova, A.L.: *J. Membr. Sci.* **189**, 83 (2001)
104. Tien, H.T., Wurster, S.H., Ottova, A.L.: *Bioelectrochem. Bioenerg.* **42**, 77 (1997)
105. Tien, H.T., Ottova, A.L.: *Colloids Surf. A: Physicochem. Eng. Aspects* **149**, 217 (1999)
106. Nikolelis, D., Hianik, T., Krull, U.J.: *Electroanalysis* **11**, 7 (1999)
107. Girard-Egrot, A.P., Godoy, S., Blum, L.J.: *Adv. Colloid Interface Sci.* **116**, 205 (2005)
108. Li, J.R., Cai, M., Chen, T.F., Jiang, L.: *Thin Solid Films* **180**, 205 (1989)
109. Moriizumi, T.: *Thin Solid Films* **160**, 413 (1988)
110. Arisawa, S., Arise, T., Yamamoto, R.: *Thin Solid Films* **209**, 259 (1992)
111. Pal, P., Nandi, D., Misra, T.N.: *Thin Solid Films* **239**, 138 (1994)
112. Choi, J.W., Min, J., Jung, J.W., Rhee, H.W., Lee, W.H.: *Thin Solid Films* **327–329**, 676 (1998)
113. Yasuzawa, M., Hashimoto, M., Fujii, S., Kunugi, A., Nakaya, T.: *Sens. Actuators B* **65**, 241 (2000)
114. Heckl, W.M., Zaba, B.N., Möhwald, H.: *Biochim. Biophys. Acta* **903**, 166 (1987)
115. Okahata, Y., Tsuruta, T., Ijio, K., Ariga, K.: *Langmuir* **4**, 1373 (1988)
116. Okahata, Y., Tsuruta, T., Ijio, K., Ariga, K.: *Thin Solid Films* **180**, 65 (1989)
117. Singhal, R., Gambhir, A., Pandey, M.K., Annapoorni, S., Malhotra, B.D.: *Biosens. Bioelectron.* **17**, 697 (2002)
118. Singhal, R., Takashima, W., Kaneto, K., Samanta, S.B., Annapoorni, S., Malhotra, B.D.: *Sens. Actuators B* **86**, 42 (2002)
119. Sharma, S.K., Singhal, R., Malhotra, B.D., Sehgal, N., Kumar, A.: *Electrochim. Acta* **49**, 2479 (2004)
120. Caseli, L., Furriel, R.P.M., de Andrade, J.F., Leone, F.A., Zaniquelli, M.E.D.: *J. Colloid Interface Sci.* **275**, 123 (2004)
121. Wan, K., Chovelon, J.-M., Jaffrezic-Renault, N.: *Talanta* **52**, 663 (2000)
122. Tsuzuki, H., Watanabe, T., Okawa, Y., Yoshida, S., Yano, S., Koumoto, K., Komiya, M., Nihei, Y.: *Chem. Lett.* 1265 (1988)

123. Tatsuma, T., Tsuzuki, H., Okawa, Y., Yoshida, S., Watanabe, T.: *Thin Solid Films* **202**, 145 (1991)
124. Zhang, A., Hou, Y., Jaffrezic-Renault, N., Wan, J., Soldatkin, A., Chovelon, J.-M.: *Bioelectrochemistry* **56**, 157 (2002)
125. Girard-Egrot, A.P., Morélis, R.M., Coulet, P.R.: *Langmuir* **13**, 6540 (1997)
126. Berzina T.S., Piras, L., Troitsky, V.I.: *Thin Solid Films* **327–329**, 621 (1998)
127. Vikholm, I., Albers, W.M.: *Langmuir* **14**, 3865 (1998)
128. Vikholm, I., Viitala, T., Albers, W.M., Peltonen, J.: *Biochim. Biophys. Acta* **1421**, 39 (1999)
129. Ihalainen, P., Peltonen, J.: *Langmuir* **18**, 4953 (2002)
130. Kent, M.S., Yim H., Sasaki, D.Y., Majewski, J., Smith, G.S., Shin, K., Satija, S., Ocko, B.M.: *Langmuir* **18**, 3754 (2002)
131. Godoy, S., Chauvet, J.-P., Boullanger, P., Blum, L.J., Girard-Egrot, A.P.: *Langmuir* **19**, 5448 (2003)
132. Girard-Egrot, A.P., Chauvet, J.-P., Boullanger, P., Coulet, P.R.: *Langmuir* **17**, 1200 (2001)
133. Girard-Egrot, A.P., Godoy, S., Chauvet, J.-P., Boullanger, P., Coulet, P.R.: *Biochim. Biophys. Acta* **1617**, 39 (2003)
134. Girard-Egrot, A.P., Chauvet, J.-P., Boullanger, P., Coulet, P.R.: *Colloids Surf. B: Biointerfaces* **23**, 319 (2002)
135. Godoy, S., Violot, S., Boullanger, P., Bouchu, M.-N., Leca-Bouvier, B.D., Blum, L.J., Girard-Egrot, A.P.: *ChemBioChem* **6**, 395 (2005)
136. Henderson, R., Baldwin, J.M., Downing, K.H., Lepault, J., Zemlin, F.: *Ultra-microscopy* **19**, 147 (1986)
137. Sara, M., Sleytr, U.B.: *J. Bacteriol.* **182**, 859 (2000)
138. Gumbiner, B.M.: *J. Cell. Biol.* **148**, 399 (1988)
139. Henderson, R., Baldwin, J.M., Ceska, T.A., Zemlin, F., Beckmann, E., Downing, K.H.: *J. Mol. Biol.* **213**, 899 (1990)
140. Walz, T., Grigorieff, N.: *J. Struct. Biol.* **121**, 142 (1998)
141. Nogales, E., Wolf, S.G., Downing, K.H.: *Nature* **391**, 199 (1998)
142. Unwin, N.: *J. Mol. Biol.* **346**, 967 (2005)
143. Strandburg, K.J.: *Rev. Mod. Phys.* **60**, 161 (1988)
144. Chernov, A.A.: *J. Struct. Biol.* **142**, 3 (2003)
145. Jap, B.K., Zulauf, M., Scheybani, T., Hefti, A., Baumeister, W., Aepli, A., Engel, A.: *Ultramicroscopy* **46**, 45 (1992)
146. Bieri, C., Ernst, O.P., Heyse, S., Hofmann, K.P., Vogel, H.: *Nat. Biotechnol.* **17**, 1105 (1999)
147. Uzgiris, E.E., Kornberg, R.D.: *Nature* **301**, 125 (1983)
148. Brisson, A., Lambert, O., Bergsma-Schutter, W.: Two-dimensional crystallization of soluble proteins on planar lipid films. In: *Crystallization of Nucleic Acids and Proteins, a Practical Approach*, ed. by A. Ducruix and R. Giegé, Oxford University Press, Oxford (1999) pp. 341–63
149. Schultz, P., Bischler, N., Lebeau, L.: *Methods Mol. Biol.* **148**, 557 (2001)
150. Ludwig, D.S., Ribí, H.O., Schoolnik, G.K., Kornberg, R.D.: *Proc. Natl. Acad. Sci. USA* **83**, 8585 (1986)
151. Mosser, G., Brisson, A.: *J. Struct. Biol.* **106**, 191 (1991)
152. Lebeau, L., Olland, S., Oudet, P., Moiskowski, C.: *Chem. Phys. Lipids* **62**, 93 (1992)
153. Blankenburg, R., Meller, P., Ringsdorf, H., Salesse, C.: *Biochemistry* **28**, 8214 (1989)

154. Kubalek, E.W., Stuart, F., Le Grice, J., Brown, P.O.: *J. Struct. Biol.* **113**, 117 (1994)
155. Dietrich, C., Schmitt, L. Tampé, R.: *Proc. Natl. Acad. Sci. USA* **92**, 9014 (1995)
156. Darst, S., Edwards, A.M., Kubalek, E.W., Kornberg, R.D.: *Cell* **66**, 121 (1991)
157. Avila-Sakar, A.J., Chiu, W.: *Biophys. J.* **70**, 57 (1996)
158. Brisson, A., Bergsma-Schutter, W., Oling, F., Lambert, O., Reviakine, I.: *J. Crystal Growth* **196**, 456 (1999)
159. Oling, F., Bergsma-Schutter, W., Brisson, A.: *J. Struct. Biol.* **133**, 55 (2001)
160. Selinger, J.V., Schnur, J.M.: *Phys. Rev. Lett.* **71**, 4091 (1993)
161. Ringler, P., Mueller, W., Ringsdorf, H., Brisson, A.: *Chem. Eur. J.* **3**, 620 (1997)
162. Wilson-Kubalek, E.M., Brown, R.E., Celia, H., Milligan, R.A.: *Proc. Natl. Acad. Sci. USA* **95**, 8040 (1998)
163. Lévy, D., Mosser, G., Lambert, O., Moeck, G.S., Bald, D., Rigaud, J.-L.: *J. Struct. Biol.* **127**, 44 (1999)
164. Muller, D.J., Schoenenberger, C.A., Schabert, F., Engel, A.: *J. Struct. Biol.* **119**, 149 (1997)
165. Muller, D.J., Amrein, M., Engel, A.: *J. Struct. Biol.* **119**, 172 (1997)
166. Fotiadis, D., Liang, Y., Filipek, S., Saperstein, D.A., Enegl, A., Palczewski, K.: *Nature* **421**, 127 (2003)
167. Scheuring, S., Sturgis, J.N., Prima, V., Bernadac, A., Levy, D., Rigaud, J.-L.: *Proc. Natl. Acad. Sci. USA* **101**, 11293 (2004)
168. Reviakine, I., Bergsma-Schutter, W., Brisson, A.: *J. Struct. Biol.* **121**, 356 (1998)
169. Reviakine, I., Brisson, A.: *Langmuir* **17**, 8293 (2001)
170. Ratanabanangkoon, P., Gast, A.P.: *Langmuir* **19**, 1794 (2003)
171. Trépout, S., Mornet, S., Benabdelhak, H., Ducruix, A., Brisson, A., Lambert, O.: *Langmuir* **23**, 2647 (2007)
172. Godoy, S., Leca-Bouvier, B., Boullanger, P., Blum, L.J., Girard-Egrot, A.P.: *Sens. Actuators B* **107**, 82 (2005)
173. Mornet, S., Lambert, O., Duguet, E., Brisson, A.: *NanoLetters* **5**, 281 (2005)
174. Peng, J.B., Barnes, G.T., Gentle, I.R., Foran, G.J.: *J. Phys. Chem. B* **104**, 5553 (2000)
175. Saint-Pierre Chazalet, M., Masson, M., Bousquet, C., Bolbach, G., Ridente, Y., Bolard, J.: *Thin Solid Films* **244**, 852 (1994)
176. Heywang, C., Saint-Pierre Chazalet, M., Masson, M., Bolard, J.: *Biophys. J.* **75**, 2368 (1998)
177. Leverette, C.L., Dluhy, R.A.: *Colloids Surf. A: Physicochem. Eng. Aspects* **243**, 157 (2004)
178. Vandevyver, M., Barraud, A.: *J. Mol. Electron.* **4**, 207 (1988)
179. Sanchez, J., Badia, A.: *Thin Solid Films* **440**, 223 (2003)
180. Hollars, C.W., Dunn, R.C.: *J. Chem. Phys.* **112**, 7822 (2000)
181. Lahmani, M., Dupas, C., Houdy, P. (Eds.): *Nanoscience I: Nanotechnologies and Nanophysics*, Springer, Berlin Heidelberg New York (2007)

Supramolecular Complexes of DNA

G. Zuber and D. Scherman

3.1 Introduction

Deoxyribose nucleic acid or DNA is a linear polymer in the form of a double strand, synthesised by sequential polymerisation of a large number of units chosen from among the nucleic bases called purines (adenosine A and guanosine G) and pyrimidines (cytosine C and thymidine T). DNA contains all the genetic information required for life. It exists in the form of a limited number (a few dozen) of very big molecules, called chromosomes. This genetic information is first of all transcribed. In this process, a restricted fragment of the DNA called a gene is copied in the form of ribonucleic acid, or RNA. This RNA is itself a polymer, but with a single strand in which the sequence of nucleic acids is schematically analogous to the sequence on one of the two strands of the transcribed DNA. Finally, this RNA is translated into a protein, yet another linear polymer. The proteins make up the main part of the active constituents ensuring the survival of the cell.

Any loss of information, either by mutation or by deletion of the DNA, will cause an imbalance in the cell's metabolism that may in turn lead to incurable pathologies. Several strategies have been developed to reduce the consequences of such genetic deficiencies or, more generally, to act, by amplifying or suppressing them, on the mechanisms leading from the reading of the genetic information to the production of proteins:

- Strategies aiming to introduce synthetic DNA or RNA, which selectively block the expression of certain genes, are now being studied by an increasing number of research scientists and pharmacologists. They use antisense oligodeoxyribonucleotides or interfering oligoribonucleotides and they already have clinical applications. This kind of therapy is often called gene pharmacology.
- Other, more ambitious strategies aim to repair in situ mutated or incomplete DNA within the chromosomes themselves, by introducing short

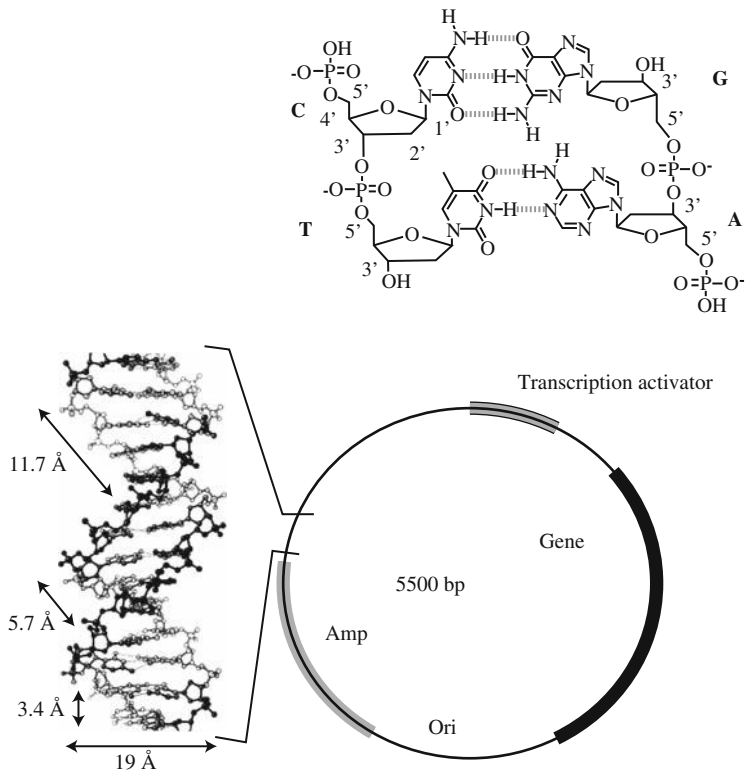


Fig. 3.1. Different DNA structures. One strand of DNA is a linear polymer made up of only four different nucleotides. The specific formation of base pairs by hydrogen bonds leads to the formation of the DNA double helix. A gene is generally produced in bacterial systems by inclusion in a plasmid

sequences of DNA or RNA which recognise and take the place of mutations. This is the underlying principle of genetic correction.

- Yet other strategies aim to reintroduce the deficient DNA fragments into the cells in the form of genes. Indeed, in certain diseases, the only solution is to bring genetic information back into the cells by transferring exogenous DNA into the cell nucleus. This approach goes by the name of gene therapy.

There are two fundamental problems that must be faced in order to implement these procedures:

- How can one produce therapeutic nucleic acids in sufficient quantities and with sufficient uniformity?
- How can one administer these nucleic acids to the cell nuclei, the process known by specialists as transfection?

The subject of the present chapter is the second question rather than the first. In fact, some synthetic molecules spontaneously form macromolecular complexes with DNA or RNA. Under certain conditions, these systems have proven to be efficient vehicles for transporting genetic material into cells. For this reason, these molecules, sometimes called vectors, are used on a daily basis in the laboratory by almost all biologists. And as we have just seen, they also provide a potential means of devising new medicines.

It is worth saying a little more here about the idea of gene therapy. Up until the beginning of the 1970s, DNA was the most difficult biological compound to analyse owing to its length and the repetition, over sequences of several billion units, of its four modular constituents (see Fig. 3.1): the nucleic acids A and G (the purines) and T and C (the pyrimidines). To give an example, a non-sexual human cell, i.e., containing two copies of all the chromosomes, except X in the case of males, includes 10^9 pairs of bases of a linear double helix with only four different nucleotide units. This genome would have a diameter of 2 nm and a total length of 1 m if the DNA molecules of all 23 chromosomes was unwound and placed end to end.

Today, DNA has become one of the easiest biological macromolecules to produce, in the form of plasmid in bacteria, and also one of the easiest to manipulate with the help of enzymes. In addition, processes for analysing DNA and RNA have been automated and it is easy today to process very large amounts of data. This has made it possible to sequence whole genomes, including the human genome. This kind of progress, combined with a better understanding of cell processes on the molecular level and of the relevance of genetics to many diseases makes it possible today to envisage using DNA as a medicine. In a simple case, an exogenous gene will take over from a deficient one causing the disease. But going beyond this, we know that genetic information can be used to combat a range of diseases, including cancer and blood circulation problems (ischemia), where the expression of a protein or even an RNA molecule can bring therapeutic benefits.

The interest of gene therapy lies mainly in the amplification of genetic information and the highly selective action of the products (proteins). Having delivered the gene to the cell nucleus, the transcription of the DNA into several messenger RNA molecules, which will in turn be translated into a set of therapeutic proteins, leads to a significant amplification of the information (see Fig. 3.2). Furthermore, this signal can be very tightly controlled via sequences called promoters, which adjust the transcription with respect to many different parameters such as cell type, cell activity, or hormone signalling. This combination of amplification, specificity, and control has many advantages.

Although nucleic acids are undoubtedly powerful active ingredients, there are major pharmacological problems in exploiting them as medicines. Indeed, like any active ingredient, nucleic acids must be delivered to the right site in order to exercise their therapeutic effects. However, the chemical nature, size, and (hydro)dynamic behaviour of DNA polymers considerably hinders their diffusion in a multicellular organism, crossing of the plasma membrane of cells

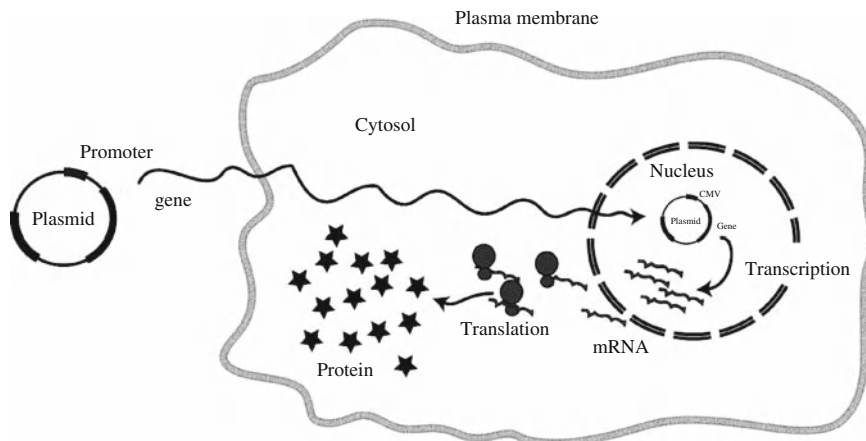


Fig. 3.2. The cell amplifies the genetic information by transcribing a gene into many messenger RNA molecules, then by translating the messenger RNAs into a protein. Amplification occurs in each step

(hydrophobic lipid bilayer), diffusion through the intracytoplasmic sieve made up of actin filaments and microtubules, and even passage through nuclear pores. In addition to this, DNA is a biodegradable nanomaterial: outside the nucleus, where enzymes are able to repair it, DNA degrades in a few minutes!

Despite the difficulties mentioned here, certain nanoscale biological entities, viz., viruses, are able to deliver their own genetic material right into the cell nucleus, where amplification of the information leads to the production and ejection of viral replicons. There is a tremendous diversity in the world of viruses. They generally measure some 100 nm in diameter, and another property they all share is that they exploit certain cell signalling and internalisation mechanisms in an extremely effective way. Accumulated knowledge of viruses has been put to use to develop biological systems for producing nanoparticles called defective recombinant viruses, incorporating almost exclusively therapeutic exogenous genetic material (see Chap. 23). In an alternative and complementary way, chemists have been developing new molecules called synthetic vectors, which associate with DNA and improve its pharmacological attributes (see Chap. 24).

In the next section, we shall give a schematic overview of the general problem of gene transfer. We then discuss the elaboration and characterisation of polymolecular DNA assemblies (polyplexes and lipoplexes), which have outstanding gene transfer properties on the cellular level. Finally, we describe how to prepare monomolecular DNA nanoparticles (nanoplexes).

3.2 Different Stages of Gene Transfer

3.2.1 Presentation

One can consider five main functions for synthetic vectors and their assemblies with DNA:

- To pack and protect DNA. Plasmids, which have sizes in the range from 100 to 500 nm in aqueous solution, depending on the number of base pairs and the supercoiled state, are packed and protected by the synthetic vector from various sources of degradation, such as nucleases.
- To facilitate good circulation through tissue barriers in an animal (in vivo administration), right down to cell membranes.
- To help fix the DNA in the cell. DNA is a polyanion with no spontaneous ability to adhere to the polyanionic plasma membrane in cells.
- To help the DNA to cross the plasma or endosomal membrane.
- To promote intracellular circulation and penetration of the nucleus. The latter is a major challenge when trying to transfect quiescent cells.

3.2.2 Condensation and Protection of DNA

At the nanoscale, double-strand DNA of the size of a gene is a semi-rigid, anionic linear polymer. In solution it adopts a wormlike hydrodynamic behaviour, taking up a large volume, and diffusion is limited in consequence. It has been estimated that a DNA double helix with more than 2,000 base pairs (which corresponds to the size of a gene coding for a 60 kDa protein) has a low diffusion coefficient in aqueous solution (about 10^{-8} cm²/s) and is immobile in cytosol (see Figs. 3.3A and B). This implies a consequent increase in the time the biological polymer spends in contact with nucleases in the serum or cytoplasm, and this has a significant impact on its stability.

The condensation of filiform DNA into a compact structure improves not only its dynamical behaviour thanks to the smaller volume it occupies, but also its chemical stability. The typical macroscopic example of a semi-rigid polymer is a very long hose pipe which gets twisted up and soon becomes unusable unless it is carefully rolled up for storage.

The transformation of a polymer from an unfolded and disorganised state into a compact structure depends on the intrinsic flexibility of the polymer and the possibilities for stabilising the system by interactions between segments of the polymer or with external molecules present in the solvent. The condensation of native DNA is opposed by electrostatic repulsion due to the negative charges placed at intervals of about 7 Å along the deoxyribose–phosphate backbone of the DNA. It is only by neutralising these charges by cations (complexation) that one can annul these repulsive forces between the anionic segments of the polymer and favour a transition into a condensed and protected form. Of course, the size, structure and properties of DNA:cation

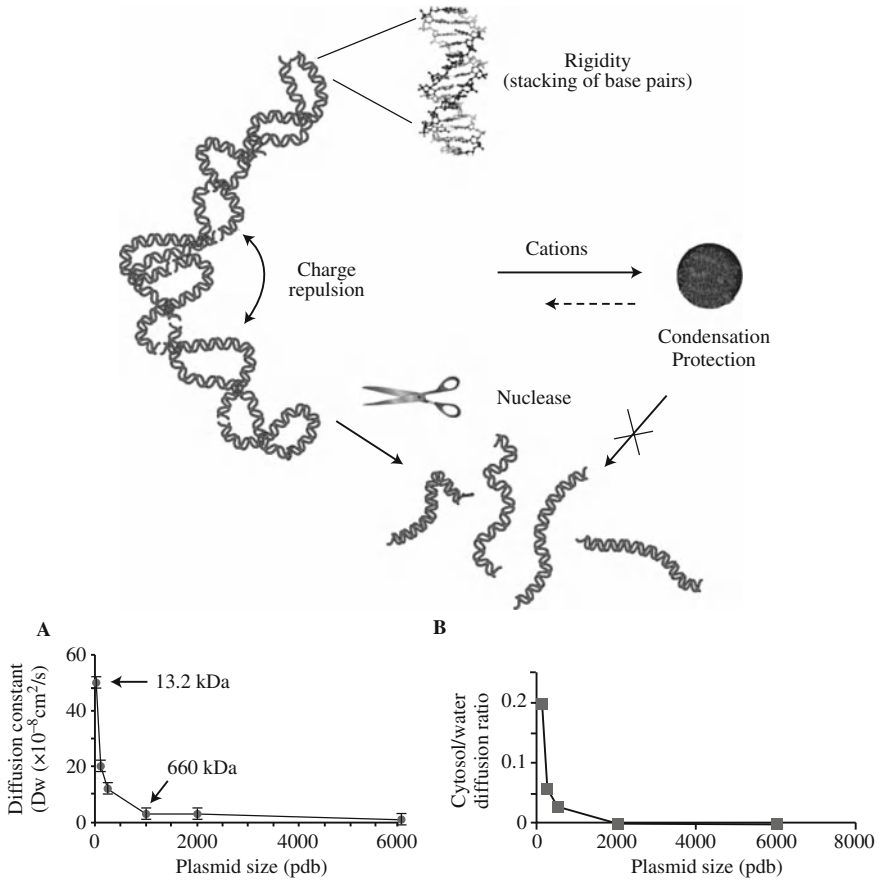


Fig. 3.3. Gene-sized DNA is a polymer taking up a large volume, hence slow to diffuse through water and immobile in cytosol. Close-packing by cations improves its chemical stability

complexes depend on the kind of cationic agent, its association constant, and also the concentrations of the partners and their stoichiometry and charge ratio. Naturally, for biological applications, the DNA:cation complexes must have sufficient stability with respect to the macromolecules circulating in the organism, while still allowing the DNA to be released in the cell.

3.2.3 Circulation in a Multicellular Organism

A mammal is composed of several organs which are themselves made up of differentiated cells and extracellular polymers. This type of organisation constitutes a very effective barrier to DNA or small particles. Indeed,

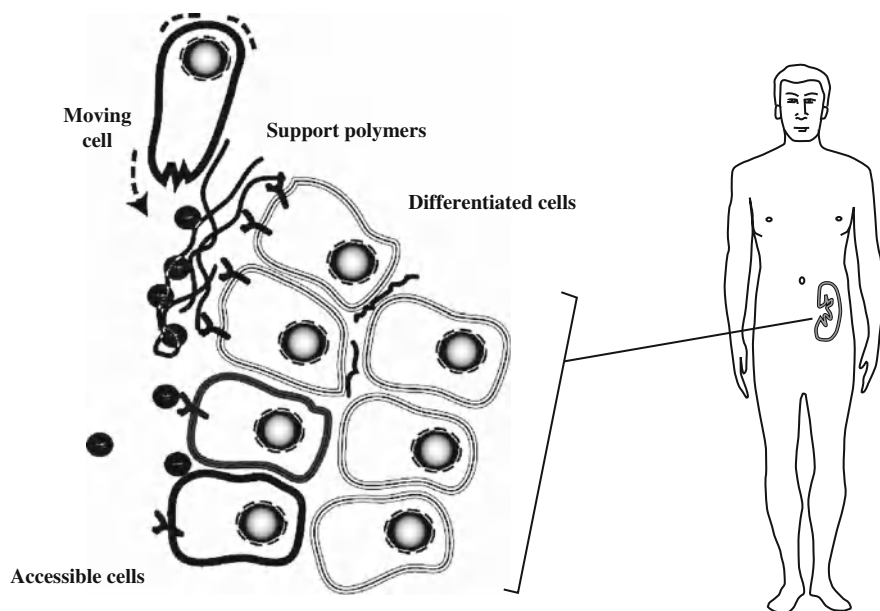


Fig. 3.4. In an organism with a tissue organisation based on differentiated cells, it is difficult for nanomaterials to reach the surface of many cells. Most of the material is caught by support polymers or adheres to the first cell layer. In some cases, reptating cells, e.g., lymphocytes, can carry the trapped nanomaterial

intercellular contacts and extracellular polymers physically impede any diffusion of molecules with dimensions greater than 8 nm through the organism (see Fig. 3.4). DNA delivery systems thus have only a very restricted access to tissues where a large cell area is physically accessible by surgery, local administration, or blood injection. Systemic injection is particularly interesting because it provides access to remote and deeply imbedded tissues. The composition of the blood and the presence of specific blood filtering and purifying organs means that these assemblies must be stable, as small as possible, and if possible stealthy (see Chap. 24).

3.2.4 Cell Adhesion and Crossing of the Plasma Membrane

The plasma membrane is a major obstacle to DNA (and other hydrophilic macromolecules) trying to enter a cell. If the lipid bilayer is directly rendered permeable, this destroys the integrity of the cell and seriously threatens its viability. However, it can be done over a short time scale, e.g., when the cells are subjected to suitable electric fields. One solution is to make use of cellular mechanisms for membrane invagination, internalising material in vesicles known as endosomes. This sub-compartmentalisation has many advantages, because it is generally associated with changes in the medium which can be

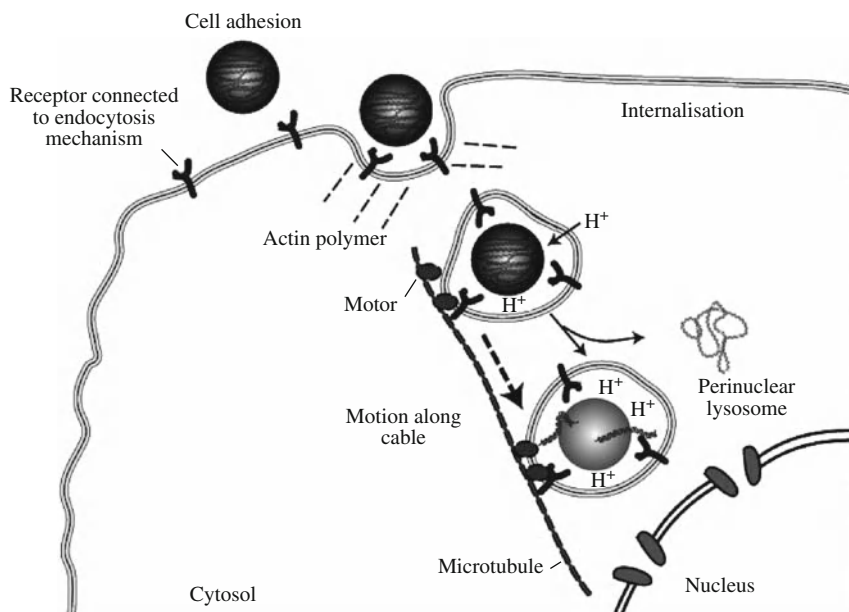


Fig. 3.5. Endocytosis of particles in cells. Particles adhere to specific membrane domains and are internalised in endosomes, of which there exist several types. These vesicles then undergo transformations and are actively guided into cell recycling compartments

put to use to induce a lysis of the membrane bilayer. In addition, the rupture of a limited amount of endosomal membrane has little incidence on cell viability.

This type of mechanism requires particles to be bound to membrane receptors which are themselves connected to endocytosis mechanisms (see Fig. 3.5). Endocytosis results from internalisation of specific domains of the plasma membrane in intracellular compartments. These compartments are then actively steered by actin filaments or by protein 'motors' along microtubules into recycling compartments which undergo chemical modifications (reduction of the pH) or enzymatic modifications (for complete degradation).

The choice of the way the DNA/vector particles bind onto the plasma membrane has important repercussions. Electrostatic interactions with anionic molecules present on all cells are particularly interesting *in vitro* for a very general application to a large number of different cell monocultures (lines). Alternatively, by binding to specific protein receptors such as integrins, or those binding transferrin, folic acid, or asialo glucoprotein, one can selectively target cells or tissues which overexpress them, or even target a specific mode of internalisation.

Once in the endosomes, the assemblies must induce a rupture in the membrane surrounding them in order to be released in the cytosol. However, this

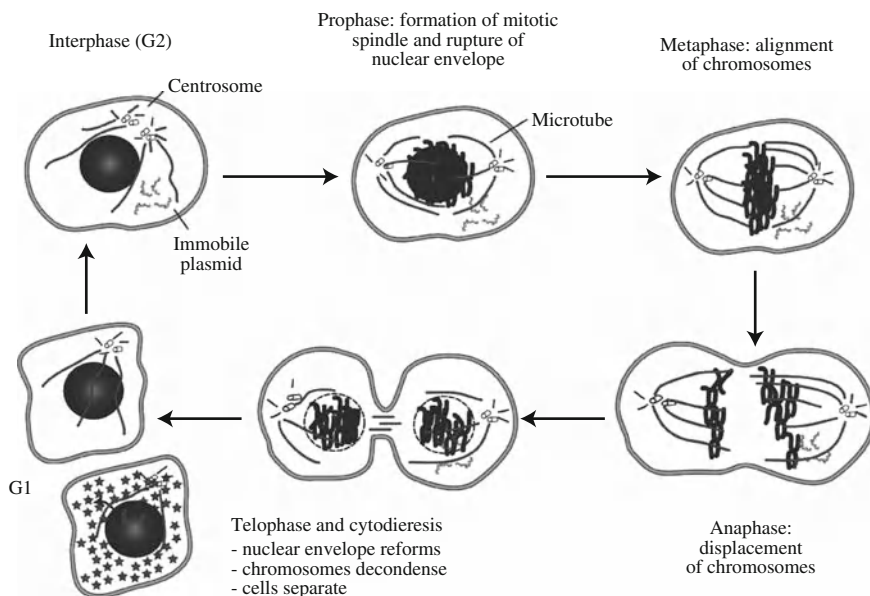


Fig. 3.6. Model for internalisation of particles in the nuclei of dividing cells. The immobile plasmid, in free or complexed form, is taken by moving chromosomes and sequestered in the nucleus after reconstitution of the nuclear envelope. Stars represent the protein or gene product

kind of rupture is not easy to engineer. One solution is to incorporate molecules with a fusogenic function (e.g., dioleoylphosphatidylethanolamine or DOPE, some peptides derived from viral sequences or designed *ab initio*, etc.), i.e., able to destabilise the membrane in such a way as to facilitate intracellular release of the plasmids. Another idea is to use the active acidification of the endosomes. This differentiated cell mechanism allows one to exploit the properties of cationic polymers that can act as a ‘proton sponge’, such as polyethylene imine (PEI).

3.2.5 Intracellular Circulation and Entry into the Nucleus

Studies of intracellular traffic show that DNA and its complexes transit via the cytosol before entering the nucleus. Physically speaking, a DNA molecule has little chance of getting through the nuclear envelope of a quiescent cell owing to its immobility in the cytoplasm. However, there are transport mechanisms for macromolecules to get from the cytoplasm into the nucleus via nuclear pores. While macromolecules smaller than 40 kDa diffuse passively through these pores, larger macromolecules or nanometric particles of diameter 39 nm can get through with the help of specific nuclear import mechanisms, using specific signals known as nuclear localisation signals (NLS) and ‘cargo’ proteins. When

compacted, a single plasmid collapses to a particle with diameter about 25 nm, hence compatible with assisted entry via the nuclear pores.

While the nuclear pore provides the only means of entry into the nucleus in the case of quiescent cells, cell division offers another possibility for incorporating an immobilised exogenous gene in the cytosol. During cell division, the nuclear envelope disappears into a continuum with the endoplasmic reticulum and no longer constitutes an obstacle between the cellular genetic material and the cytoplasm. During mitosis, chromosomes are actively transported along microtubules and, with a bit of luck (a statistical or probabilistic factor), they can carry the exogenous gene and sequester it in the nuclei of the daughter cells (see Fig. 3.6). A cell division thus leads to a tenfold increase in the efficiency of transfection mediated by synthetic vectors.

3.2.6 State of the Art

These functions have now been achieved, at least partially, by synthesising supramolecular structures containing DNA and associative molecules. The association of polymers and cationic lipids with DNA produces assemblies known as polyplexes and lipoplexes, respectively, capable of efficiently transfecting cell lines cultivated in boxes. The best gene vectors would seem to have found some therapeutic applications *in vivo* for treating easily accessible cancerous cells. Finally, solutions based on supramolecular chemistry are beginning to emerge, providing precise control over the way DNA is compacted in stable assemblies of minimal dimensions known as nanoplexes.

3.3 Polymolecular DNA Assemblies: Synthesis, Characterisation and Properties

3.3.1 Polyplexes

Introduction and Structure

Condensation of DNA into stable assemblies is achieved by a simple cooperation effect using polycationic materials (see Fig. 3.7). The most obvious method is to use cationic polymers to condense and protect the DNA, and many synthetic polymers are effective in compacting DNA. Furthermore, in eukaryotic cells, chromosomal DNA is compacted in this way using proteins rich in the amino acids lysine and arginine: the histones. Likewise, in spermatozooids, it is another protein rich in arginine, *viz.*, protamine, which leads to maximally compacted genetic material.

Polycationic polymers such as poly(L-lysine), protamine, polyethylene imine (PEI), and also cationic dendrimers, associate with DNA by means of multiple electrostatic interactions producing a cooperation process that generates the particles called polyplexes. Poly(L-lysine) was originally the most

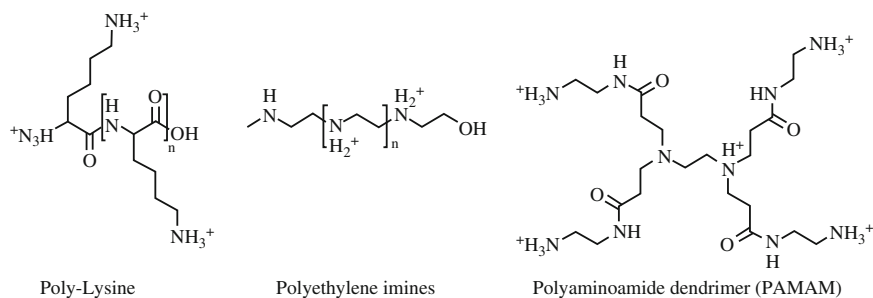


Fig. 3.7. Chemical structures of polycations used for DNA transfection. In general, the degree of polymerisation for polylysine and PEI is around 500. Fourth generation dendrimers are beginning to have gene transfer applications *in vitro*

widely used cationic polymer, but lacks endosomolytic properties (i.e., it is unable to rupture endosomes). Since then, it has been shown that polyethylene imine (PEI) and intact or fragmented polyamidoamine dendrimers can function as efficient promoters of gene transfer. These polymers seem to owe their efficiency as gene transfer agents to their chemical structure and their electron density, which considerably increases the acidity of ammoniums. At physiological pH, a large fraction of the amines is not protonated and would function in endosomes as a proton sponge, with endosomolytic properties. The relation between the proton sponge property and endosomolytic activity is discussed below. So to preserve the coherence of the discussion, we shall only describe polyplexes of linear PEI.

Synthesis of Polyplexes

A precise knowledge of the stoichiometry or charge ratio between the DNA polymer and the polycation is essential, because it determines the extent to which the DNA is compacted and also the stability of the result. It is easy to calculate the number of negative charges of the DNA because, at physiological pH (around 7.4), the phosphates (pK_a around 1) of the DNA are always negative. The titration of negative charge or phosphate (P) is thus simply obtained by measuring the nucleic base concentration by spectrophotometry. For a plasmid, the specific absorbance of a nucleic base (and hence a phosphate) at 260 nm is $6,600 \text{ M}^{-1} \text{ cm}^{-1}$.

The cationic charge number of a synthetic polymer is more difficult to determine. To begin with, polymerisation techniques are not as precise as biological methods and produce polymer populations that must be described statistically, e.g., with reference to average mass, polydispersity, etc. A second point is that cationic polymers are generally polyammoniums. Although the nitrogen number of these polymers is easily established by colorimetric reaction, the number of cationic charges is not the same as the nitrogen number, because environmental constraints induced by the polymer considerably

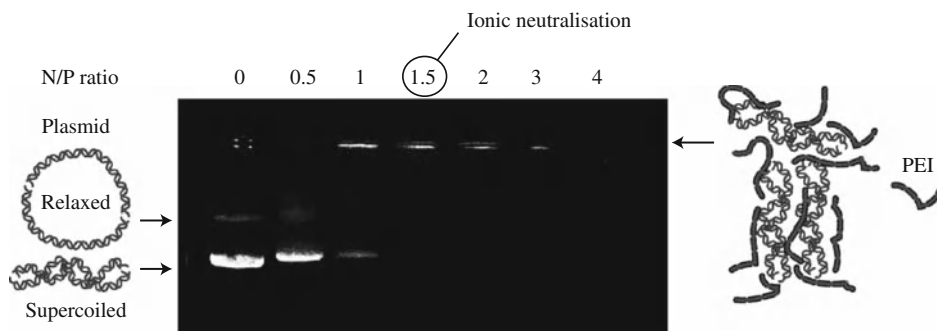


Fig. 3.8. Determining the ionic stoichiometry of a DNA/PEI polyplex by retardation in an electrophoresis gel

reduce the pKa of the ammoniums. Hence, the overall charge of the polymer may be distinctly smaller than the nitrogen number, but vary also with the pH of the medium. An indirect but interesting method here to approximate the stoichiometry of these polyplexes uses a technique based on electrophoresis in agarose gel (see Fig. 3.8).

When an electric field is applied, molecules and complexes migrate through the pores of the gel at a rate proportional to their electrical charge, but inversely proportional to their apparent volume. The different topological shapes of a given plasmid can thus be separated in terms of their respective apparent volumes (see Fig. 3.8, line 1). In the experiment of Fig. 3.8, a plasmid is condensed with increasing amounts of PEI and the resulting complexes analysed by electrophoresis on agarose gel. The plasmid is then detected by intercalation of ethidium bromide. The irreversibility and speed of condensation of the DNA by PEI leads to the formation of complexes that remain completely trapped in the wells. When the number of positive charges of the polymer is greater than or equal to the number of negative charges of the DNA, the bands corresponding to the plasmid disappear completely. As can be seen from Fig. 3.8, the nitrogen number of the PEI required to condense the DNA is at least 1.5 times the phosphate number of the DNA contained in the sample, showing that for this polymer one nitrogen in three does not carry a positive charge at physiological pH.

Stability of Polyplexes

The rapidity of electrostatic interactions incites several plasmids to condense and intertwine. When there is an excess of the polycation, several plasmid molecules are compacted and protected in particles surrounded by a matrix of cationic polymers. Residual positive charges at the surface contribute to cell internalisation properties and also explain the stability of complexes in suspension in the aqueous phase (see Fig. 3.9).

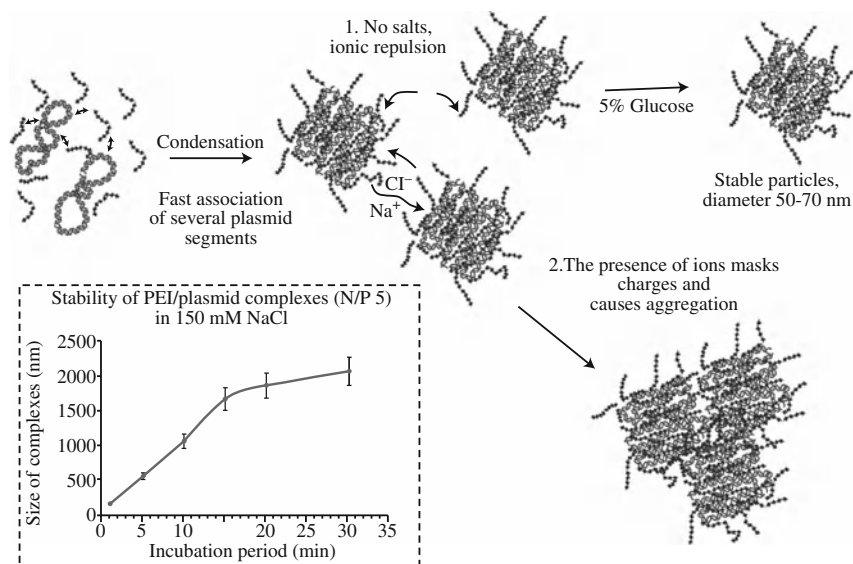


Fig. 3.9. The stability of polyplexes formed with excess cationic charge depends on the ionic strength of the reaction medium

Hence, when PEI is added to a plasmid solution, it very quickly associates with several plasmid segments. There then follows a rapid condensation of several plasmids in a complex which becomes positive when an excess of PEI is used. When this PEI/DNA condensation is carried out in the absence of salts, these complexes, positively charged at the surface, are particles measuring between 50 and 70 nm in diameter. Ionic repulsion then ensures their stability in solution. However, these same particles aggregate very quickly when their ionic surface is screened by other ions at physiological concentrations, e.g., 150 mM NaCl. It has thus been demonstrated that aggregates of diameter 500–1,000 nm are obtained after a 10 min incubation in 150 mM of NaCl.

Using DNA/PEI Complexes for in Vitro Gene Transfer

Gene transfer using PEI in adherent and rapidly dividing cell lines is easy to implement and efficient. The efficiency of gene vectors is quantified using reporter genes, which code for easily detected proteins and which remain localised in the cell. The most common are the green fluorescence protein (GFP) which, as its name suggests, is detectable by its fluorescence, and luciferase, the quantity of which is determined through its enzyme activity.

Classically, a plasmid (2 μ g for 100,000 cells/mL) coding for an easily detected protein (a reporter gene) is mixed with an excess of PEI in 150 mM NaCl. An N/P ratio of 5 is used to form the positive particles, while nevertheless limiting the amount of free PEI. These conditions lead to the formation of

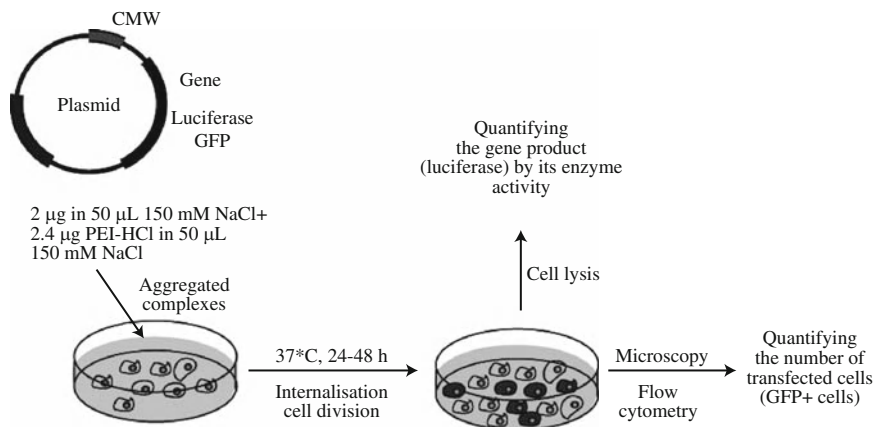


Fig. 3.10. Using PEI as transfection agent in cell lines. The lines are at about 70% confluence when the polyplexes are added and adhere to the bottom of the box

aggregates with positive surfaces and dimensions in the range 200–1,000 nm. These complexes are then diluted in the culture medium. Dilution (usually 1/10) inhibits the aggregation kinetics. The polyplexes then sediment out at the bottom of the box and adhere by electrostatic interaction to the outer surface of the cells (see Fig. 3.10).

Mechanical and dynamical studies show that heparan sulfate proteoglycans (HSPG) are the membrane receptors for cationic particles. These highly negatively charged proteins are involved in adhesion of the cell to a substrate, and some of them are connected to membrane invagination mechanisms. This leads to cellular internalisation of a large number of particles in vacuoles which are normally acidified by a proton pump. When protons are trapped inside the endosome by the protonatable amines of the PEI, this induces overactivity of the proton pump, together with a concomitant and large influx of water molecules and chloride ions (see Fig. 3.11). The vacuoles then swell up and may rupture the lipid membrane by hyperosmotic shock.

After release in the cytosol, the DNA must still get into the nucleus to be transcribed. The mechanisms underlying this stage of the process are still poorly understood. However, it has been shown that cell division contributes significantly to transfection efficiency. During cell division, the nuclear envelope disappears in a continuum with the endoplasmic reticulum and no longer constitutes a barrier between the genetic material and the cytoplasm. During mitosis, the chromosomes are actively transported along microtubules and with a little luck can carry the exogenous gene and sequester it in the nuclei of the daughter cells. Of course, this entry mechanism also works in the other direction. A plasmid that does not integrate into a chromosome of the host cell may well be eliminated from the nucleus during subsequent divisions by simple dilution. Indeed, experiment corroborates this hypothesis. With synthetic

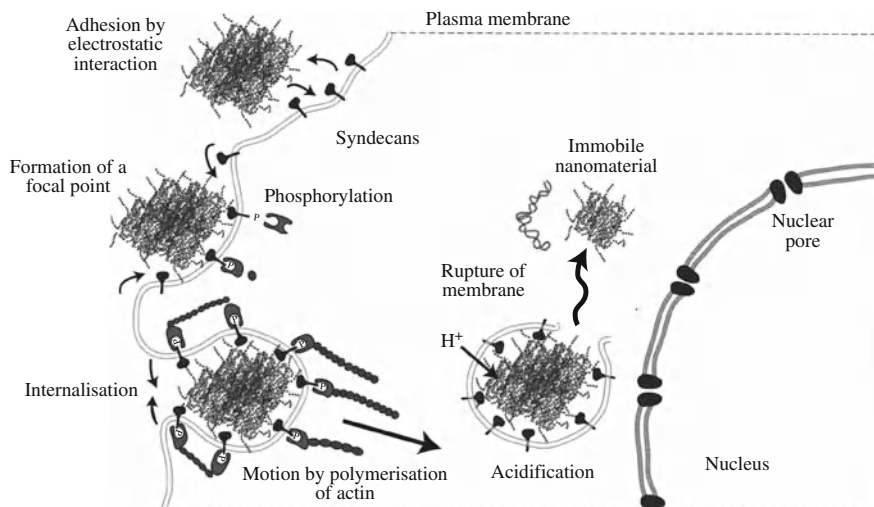


Fig. 3.11. Internalisation of cationic particles in cells. The polyplex particles bind to adhesion receptors of the cell and use the cell motion to internalise themselves. After internalisation, active transport of protons (but also chloride ions and water) leads to rupture of the membrane by hyperosmotic shock when the PEI prevents acidification through its buffer effect. Note that lipoplexes enter in the same way. It seems that nucleic acids are released by dilution of cationic lipids in anionic endosomal membranes

(polymer or lipid) DNA vectors, transfection is transient. Expression is optimal after 24–48 h, then drops off, all the more quickly as the cells divide rapidly.

3.3.2 Lipoplexes

Introduction and Structure

There are three parts to a cationic lipid: a cationic head that can bind to the DNA, one or more hydrophobic chains, and a spacer separating these two elements (see Fig. 3.12). In an aqueous solution, the hydrophobic chains join up to form particles (micelles or liposomes), leaving the cationic part in the aqueous phase. Electrostatic interactions between these self-assembled structures and the DNA polymer lead to the formation of stable lipoplexes.

A whole range of different cationic lipids have been synthesised and tested for the transfer of genes, either with or without the presence of serum, or in vivo, with a view to reducing the cytotoxicity of the formulations and improving the bioavailability of the DNA–lipid particles. Some cationic lipids are now commercially available, e.g., Lipofectine, Lipofectamine, Transfectam, TransfectACE, etc., and are routinely used in fundamental research. These are

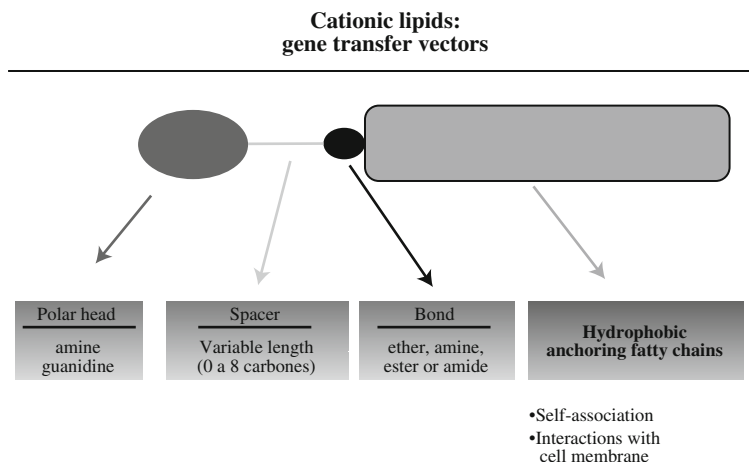


Fig. 3.12. Structure of the lipid used for gene transfer

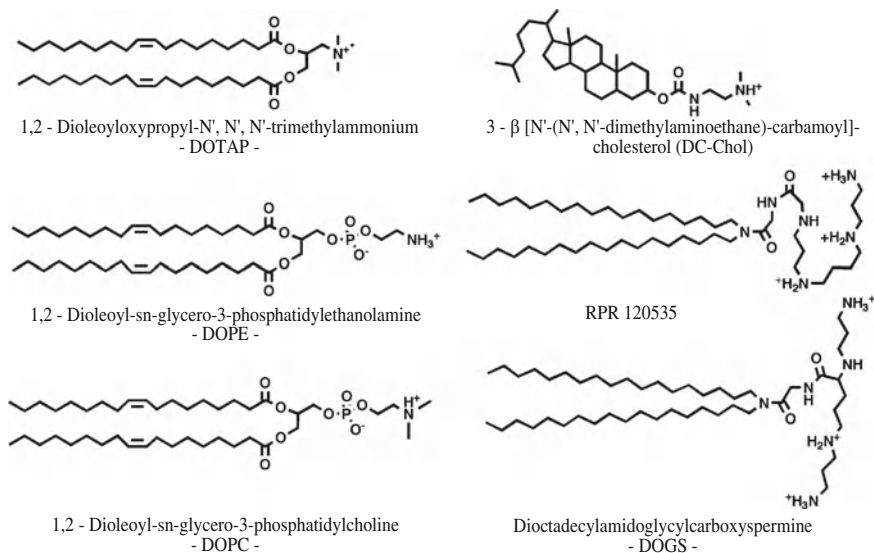


Fig. 3.13. Chemical structure of different cationic lipids

analogues of the structure of DOTAP, cationic lipids containing cholesterol, or lipopolyamines whose head is usually derived from spermine (see Fig. 3.13).

Synthesis of Lipoplexes

As has just been explained, the interest in cationic lipids stems from the fact that, in aqueous solution, they form stable superstructures with positively

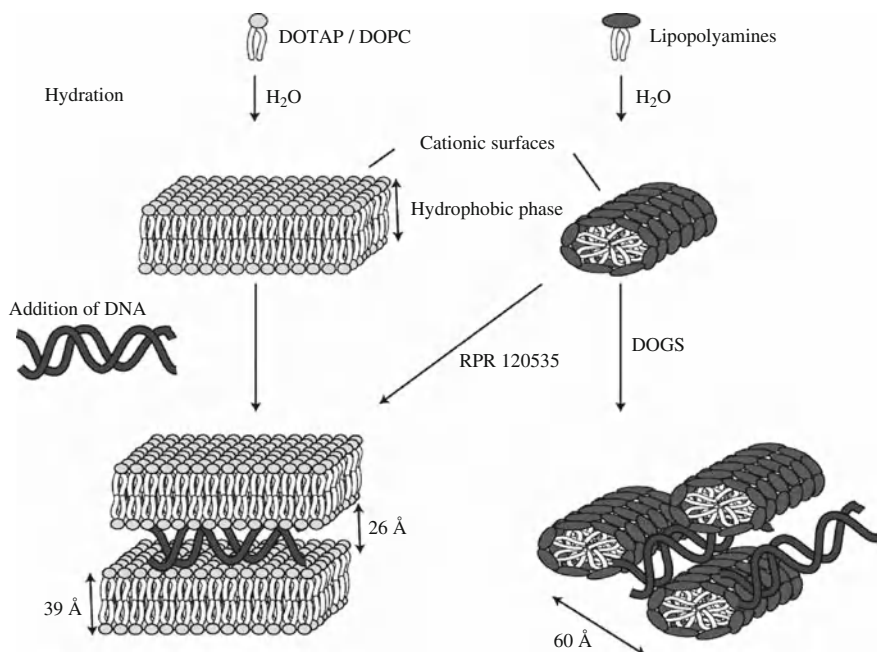


Fig. 3.14. In aqueous solution, cationic lipids form lamellar or micellar phases with positive surfaces which interact with DNA to form structures ordered into lamellar phases (DOTAP, RPR120535) or hexagonal phases (DOGS)

charged surfaces by self-association of their hydrophobic chains. There are many different ways of suspending cationic lipids in an aqueous solution. Those that form micelles (usually lipospermines) are simply solubilised in ethanol, added to an aqueous solution and homogenised by agitation. Alternatively, the cationic lipid can be formulated as a liposome, associating it with cholesterol and its derivatives or with other lipids. Liposomes are unilamellar membrane vesicles (small unilamellar vesicles SUV, or large unilamellar vesicles LUV) or multilamellar membrane vesicles, containing a volume of solvent. Standard techniques for synthesising liposomes are then preferred: mixing the cationic lipid and the colipid in a solvent such as chloroform–ethanol, evaporating the solvent in a vacuum flask, then creating a suspension by adding a buffer and supplying energy by sonication or heating. These vesicles are then put in contact with the DNA and used as in the case of DNA polyplexes.

Structure and Characterisation of Lipoplexes

Structural studies use electron microscopy or X-ray diffraction. Those studies that have been carried out so far show that the electrostatic interactions of DNA with lipids self-assembled in lipid bilayers cause the complexes to

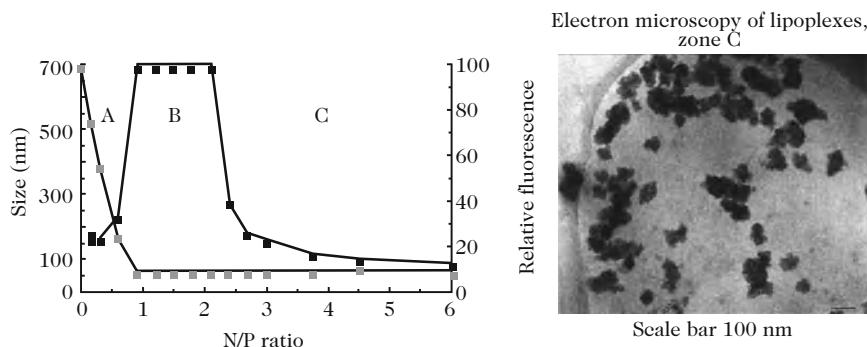


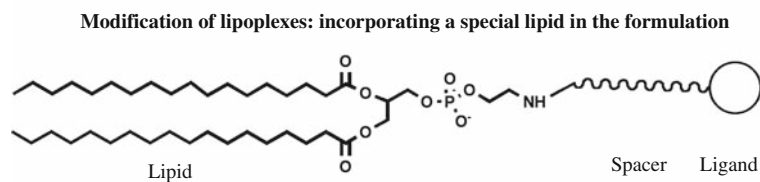
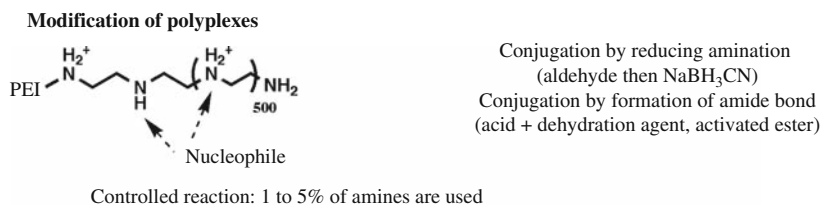
Fig. 3.15. Colloidal stability of lipoplexes as a function of the charge ratio (lipid amine to DNA phosphates). Electron microscope view of complexes. The size of the complexes (*black squares*) was determined by light scattering. DNA close-packing is measured indirectly by reduction in the fluorescence of intercalating ethidium bromide (*grey squares*)

organise themselves into lamellar phases, with the DNA sandwiched between the lipid layers. Dioctadecylamidoglycylcarboxyspermine (DOGS) molecules assemble into tubes which interact with DNA to form direct hexagonal phases that can be observed by electron microscope (see Fig. 3.14). Note that one lipospermine (RPR120535), which forms micelles in the aqueous phase, seems to reorganise itself into a lamellar phase in the presence of DNA (see Pitard et al. [3]). Finally, the presence of a sufficient amount of dioleoyl-*sn*-glycero-phosphatidylethanolamine (DOPE), a lipid known to assemble into a reverse hexagonal phase, may lead the DNA lipoplexes to adopt this same structure.

Although the precise mechanisms involved in lipoplex formation are still poorly understood, they are essentially governed by electrostatic interactions. The experiment illustrated in Fig. 3.15 shows three zones of colloidal stability, depending on the lipopolyamine/DNA (+/−) charge ratio, in which the resulting particles carry an overall negative, neutral, or positive charge, respectively. In an experiment of the same type, it has been observed that the stability and physicochemical characteristics of DNA/cationic vector lipoplexes do not depend on the size of the compacted DNA polymer.

DNA Lipoplexes for Gene Transfer

Cationic lipids are used almost in the same way as polymers for transfection of culture cells. It suffices to prepare the cationic lipoplexes. There are a few differences, however. For one thing, it is preferable to remove the serum temporarily from the culture medium during the cell adhesion stage in order to avoid harmful interactions between the serum constituents and the lipid aggregates. Another point is that it would seem that these lipids cause endosome rupture by perturbing the membrane. It seems that the lipoplexes



Example ligands

D-Mannose	Mannose receptor (dendritic cells)
Peptide with RGD motif	Integrins (endothelial cells)
Transferrin	Transferrin receptor (dividing cells)
Multi-antennary galactosides	Asialo glycoprotein receptor

Espaceur

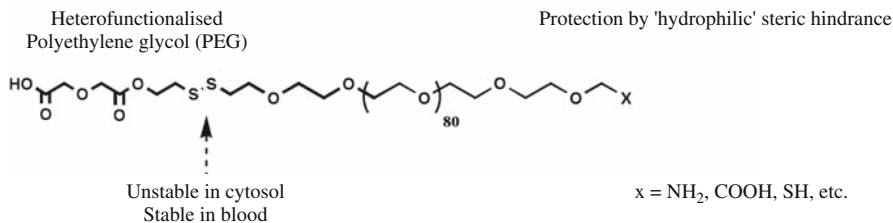


Fig. 3.16. Chemical modification of gene vectors, showing examples of ligands used for cell targeting, e.g., galactosides to target the asialo glycoprotein receptor, present on the major cells of the liver, the hepatocytes. Folic acid is a ligand for which there is a high-affinity receptor on tumour cells, and which has also received much attention. See Hofland et al. [7]

are able to disassemble by diluting the cationic lipid in the endosomal lipid bilayer charged by anionic lipids, with release of the nuclear material in the cytoplasm. This property is currently exploited for delivery of RNA, e.g., interfering RNA, either with cytosolic targets, or for which they exist natural nuclear import mechanisms.

Additives for Improving Lipoplex Properties

Many specific functions are easily incorporated into lipoplexes by anchoring in the hydrophobic phase. Incorporating 20–50% DOPE in the formulation of the cationic lipid gives an overall improvement in the transfer efficiency of nucleic acids in cells. Alternatively, molecules with a fusogenic function, i.e., able to destabilise membranes in such a way as to favour the intracellular release of plasmids, can be coupled with lipid anchors. In another approach, additives such as dodecyl-2-(1'-imidazolyl) propionate (DIP) are added to the lipid formulation, to give cationic liposomes a pH-sensitive character that will promote the release of lipoplexes from endosomes.

3.3.3 Modification of Polyplexes and Lipoplexes for *in Vivo* Gene Transfer

Cationic lipids and polymers are excellent vectors for delivering genes to a large number of cell lines in rapid division. This property depends largely on electrostatic interactions. Not only do the latter protect several genes in cationic particles; they also provide anchoring to the cell and bring about endocytosis of a large number of particles. Classic studies, carried out either by empirical methods, or by relating structure and function, have led to the development of cationic complexes with DNA that also possess endosomolytic functions. This allows the release of DNA or complexes back into the cytosol and transfection of cells in division. As they are easy to use, these vectors were soon adopted by the scientific community for a wide range of applications involving cell cultures (see Fig. 3.16).

Most cells in an animal are totally differentiated and no longer divide. These cells are not therefore easily transformed genetically by synthetic cationic vectors. This limitation can be turned to selective advantage for the transfection of cancer cells, whose surfaces are easily accessible to polyplexes. A therapeutic suicide gene formulated with PEI has thus proved to be effective in reducing bladder cancers, where cells are easily accessible by topical (local) application.

Other cancer cells can be infiltrated by particle systems injected into the blood. This particularly concerns metastases, these being irrigated by blood vessels that are not completely impervious because the cells in the wall of the vessel do not form perfectly sealed connections. In order for this to work, the DNA complexes must obviously remain stable in the blood, with sizes smaller than 100 nm, avoid elimination by blood filtration organs (kidneys) and purification organs (liver), and attach themselves selectively to cell surfaces.

Polyplexes (and also lipoplexes), in their active (cationic) form, interact quickly with blood constituents and aggregate in particular with erythrocytes. They are rapidly eliminated by hepatic degradation or remain blocked in blood capillaries, especially in the lungs. In order to improve the stability of these

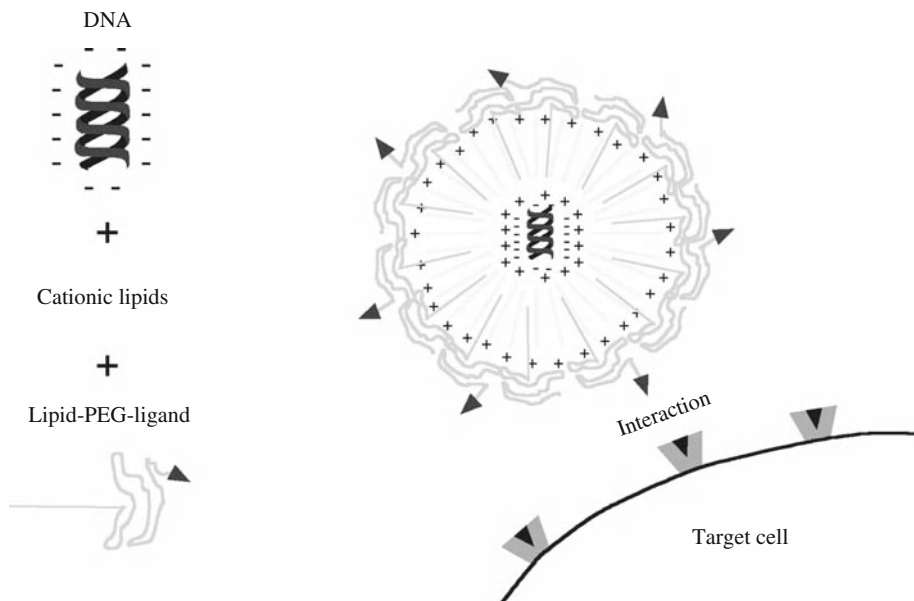


Fig. 3.17. Lipoplex delivery to extracellular receptors

assemblies in the blood, one approach is to mask the surface charges by grafting on hydrophilic polymers. The best known example of such a polymer is polyethylene glycol (PEG). On the other hand, there is a price to pay for this stealth, because it interferes with cell adherence, whence the need to incorporate cell targeting molecules. For example, the addition of triantennary galactosides allows preferential binding of nucleolipid particles *in vitro* onto the receptors of the asialo glycoproteins of human HepG2 hepatoma cells. Other ligands such as mannosylated residues, transferrin, folic acid, or antibodies also provide a way of targeting specific cellular receptors, leading to particle endocytosis mechanisms. The targeting capability of lipoplexes (see Fig. 3.17) is generally obtained by conjugating these targeting ligands with lipid anchors. For polyplexes, these targeting ligands are easily coupled with cationic polymers such as PEI or polylysine by chemical reaction with amines, which are nucleophilic. Without going into the details, this level of sophistication leads to a certain number of difficulties and requires very careful formulation, because the new elements can perturb condensation of the DNA. Likewise, some receptors may be present in very limited amounts. Consequently, an excess of polycations modified by ligands may lead to competition with the DNA complexes.

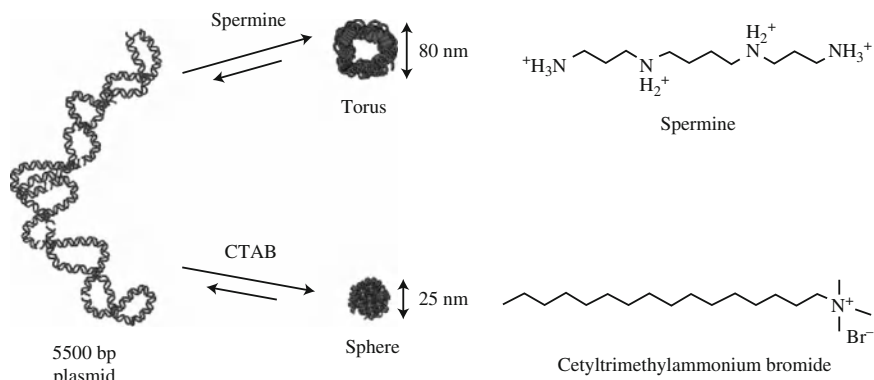


Fig. 3.18. DNA condensation with low valence ions is reversible

3.4 Monomolecular DNA Assemblies (Nanoplexes): Synthesis, Characterisation, and Properties

3.4.1 Monomolecular Condensation of DNA

For *in vivo* application, the size of DNA complexes is a crucial parameter. Owing to the high mass of the gene, the ideal solution would be monomolecular condensation. Reagents with a low affinity for DNA, mainly low valence cations like Na^+ , Mg^{2+} and spermine, interact with the DNA phosphates in a reversible way (see Fig. 3.18). Here the final system is essentially controlled by thermodynamics and one obtains varied superstructures, including tori of constant diameter with an approximately neutral surface potential.

The use of cationic detergents such as cetyltrimethylammonium bromide (CTAB) greatly improve the synthesis of DNA complexes with unit structures, because these detergents cause a single plasmid to collapse on itself in a reproducible way. Detergents are molecules with a high critical micelle concentration (CMC), usually with a single alkyl chain. This means they are only weakly hydrophobic and have a high CMC. The formation of these complexes is driven by electrostatic interactions between the cationic polar head of the unit cationic detergent and a phosphate group of the DNA polymer. Although adhesion is initially bimolecular, it is immediately followed by the aggregation of CTAB molecules into micelles, in a highly cooperative way, leading to the collapse of a single plasmid onto itself. However, rapid exchange of a detergent molecule between its monomer and micelle forms can quickly destabilise cationic surfactant/DNA complexes after dilution or addition to cell membranes.

One solution to this problem is to combine the advantages of single-chain cationic detergents in the monomolecular condensation of DNA with the stability of aggregates formed with double-chain lipids. The selective chemical transformation, after DNA condensation, of the single-chain detergent into

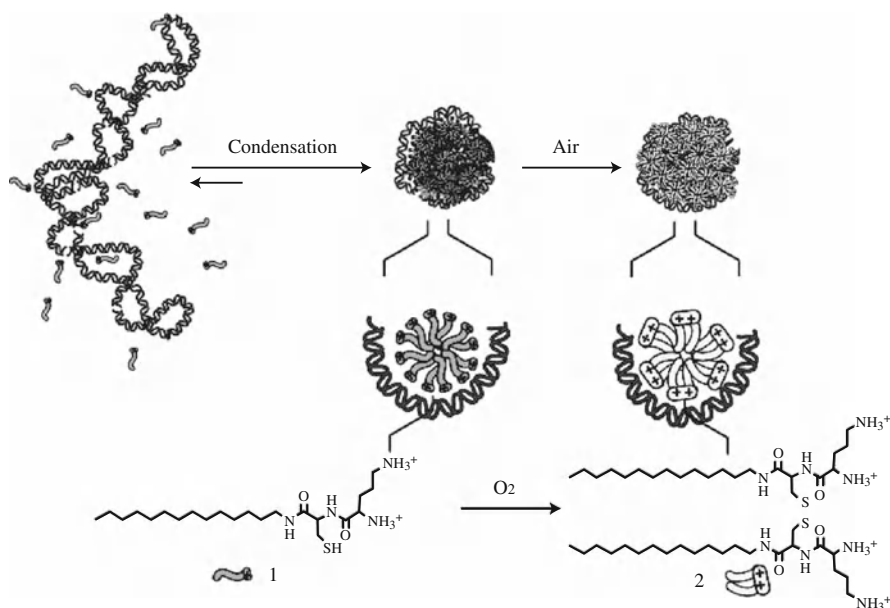


Fig. 3.19. Condensation of plasmids in stable nanometric particles. There are two stages: (1) Electrostatic association of a detergent **1** with the DNA leads to formation of micelle domains and collapse of a single plasmid onto itself. (2) In situ conversion of the detergent into a lipid **2** by formation of disulfide bridges stabilises micelle domains in interaction with the DNA

a two-chain lipid with lower solubility will stabilise the DNA particles. The chemical transformation reaction used must be inert with respect to the compacted DNA polymer in order to maintain the integrity of the genetic information. Gentle oxidation of thiol groups, in the presence of oxygen dissolved in the water, to form disulfide bridges provides a suitable solution. To begin with, a plasmid is condensed with cationic detergent containing a free thiol. In order to prevent the formation of polycationic micelles and the condensation of several plasmids, the concentration of detergent in the medium is held below its critical micelle concentration. Under these conditions, individual neutralisation of the plasmid phosphates favours hydrophobic associations of detergent molecules in DNA/micelle complexes. An increase in the thiol concentration near the DNA then favours the formation of disulfide bridges by increasing the reaction rate, which leads to a population of nanoplexes that is stable in the presence of ions (150 mM), and has uniform size distribution (see Fig. 3.19).

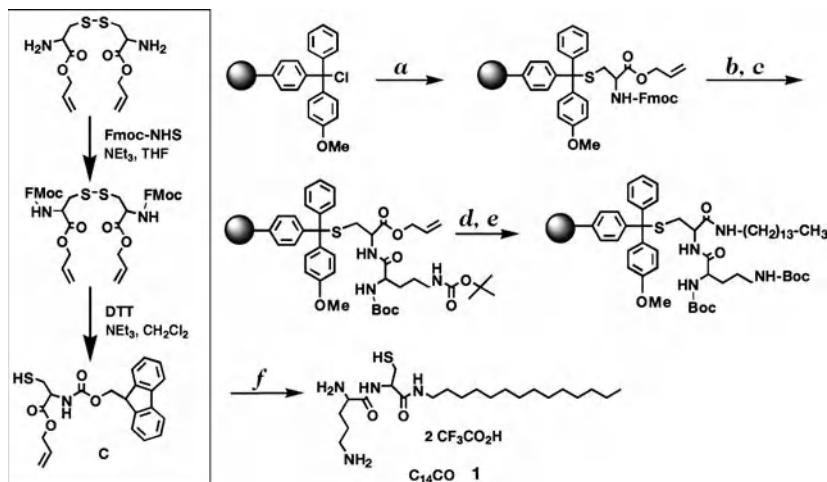


Fig. 3.20. Detergent synthesis on a solid substrate. *Inset:* Preparation of protected cysteine. (a) c, DIEA, CH_2Cl_2 . (b) 20% piperidine, DMF. (c) Bis-boc-ornithine, PyBOP, DIEA, DMF. (d) $\text{Pd}(\text{PPh}_3)_4$, $\text{CHCl}_3/\text{AcOH}/\text{N-methylmorpholine}$ (37:2:1). (e) Tetradecylamine, PyBOP, DIEA, DMF. (f) TFA

3.4.2 Chemical Synthesis

For biological applications, the properties of the detergent must be optimised. Its critical micelle concentration must exceed the phosphate concentration of the DNA in order to induce monomolecular condensation of the plasmid. In addition, after oxidation, the symmetric double-chain lipid obtained in this way (gemini surfactant) must have negligible solubility in order to maintain the cohesion of the particles under biological conditions. Dimerisable detergents with these properties have been investigated, keeping cysteine as reagent but modifying the polar head or the length of the carbon chain. Synthesis on a solid substrate is well-suited to these investigations (see Fig. 3.20) and a wide variety of compounds can be made rather quickly, including ornityl-cysteine-tetradecylamine **1**, which has been the most exhaustively investigated.

3.4.3 Synthesis and Characterisation of Nanoplexes

Monomolecular condensation of DNA is achieved as follows (see Fig. 3.21). A circular plasmid (cDNA) of 5,500 base pairs with a phosphate concentration of $60\text{ }\mu\text{M}$ (negative charge) is mixed with increasing concentrations of bicationic detergent **1** (between 24 and $40\text{ }\mu\text{M}$) in a slightly alkaline solution (pH 7.4). The concentrations of the reagents are selected to maintain detergent concentrations below its critical micelle concentration ($45\text{ }\mu\text{M}$ for **1**). The [cDNA/**1**] complexes are then stabilised by oxidising the thiols to form disulfide bridges.

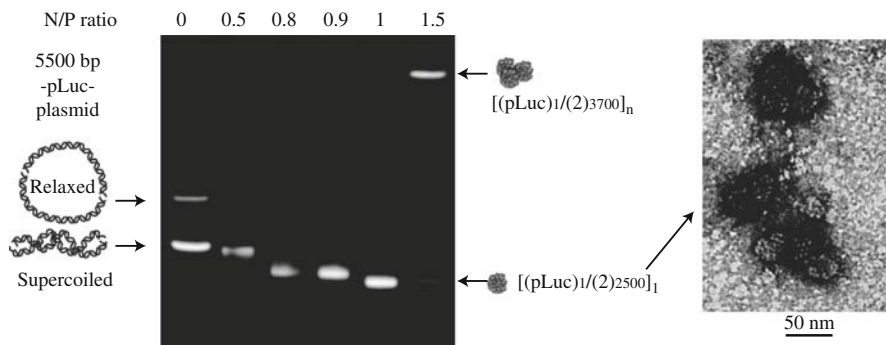


Fig. 3.21. Electrophoresis analysis of DNA nanoplexes

The slightly alkaline conditions and the oxygen dissolved in the aqueous phase ensure a measured oxidation of the thiols without unnecessary degradation of the genetic material.

After total transformation of the detergents **1** into lipids **2** (16 h at 20°C), the [cDNA/**2**] complexes were once again analysed by agarose gel retardation. An excess of detergent **1** unbound to DNA, which then transforms into lipid **2**, leads to a standard formation of lipoplexes which remain trapped in wells (line 6 of Fig. 3.21). In contrast, a gradual increase in the charge ratio (+/–) between the detergent and the DNA up to values 0.9 and 1.0 produces a complex that migrates in the gel as a single band relative to the plasmid. This property is explained by the collapse of the unfolded plasmid into a spherical structure allowing better circulation in the pores of the gel compared with reptation of the lone plasmid. This result clearly demonstrates that the [cDNA/(**2**)] nanoplexes, formed for charge ratios (+/–) in the range 0.9–1 (lines 4 and 5) are anionic, monomolecular in cDNA, and remain stable in an electric field where unstable complexes are quickly dissociated, while polymolecular cDNA complexes remain trapped in wells.

The formation of anionic particles at charge equivalence may seem surprising! However, it must be remembered that these particles were originally formed by reversible electrostatic interactions and obey the condensation theory due to Mannings. According to this model, the reversible condensation of DNA by counterions comes into play when 90% of the polymer charges are neutralised by a fast and highly cooperative process. Neutralisation of the remaining charges then becomes more and more difficult and follows an anti-cooperative process. Part of the plasmid is thus at the interface between the core of the particle and the aqueous phase, which explains why the nanoplexes have a negative surface potential (measured between –50 and –20 mV). It can be estimated that, between 2,475 (90% neutralised charge) and 2,700 (100% neutralised charge) molecules of **2** are bound to a plasmid of 5 500 bp in each [cDNA/**2**] nanoplex.

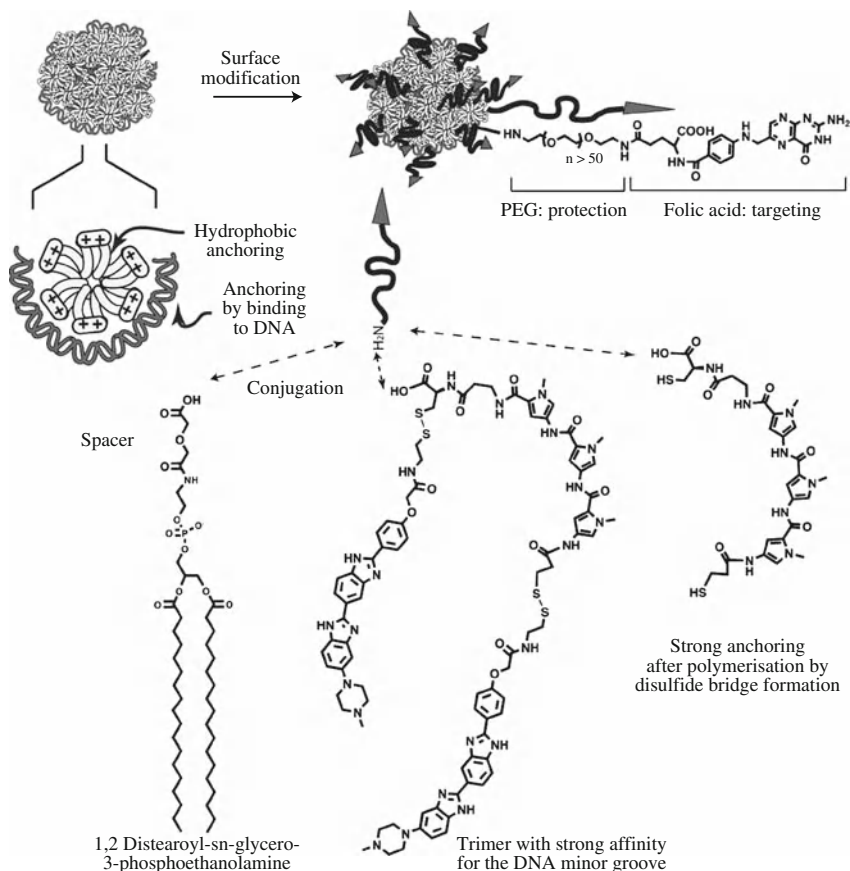


Fig. 3.22. Strategies for modifying nanoplex surfaces

Indeed, transmission electron microscope observations show that these complexes form a uniform population of spherical particles with diameters of about 25 nm. These particles also exhibit an irregular but repetitive texture. This texture is reminiscent of the appearance of DNA/DOGS complexes (see Fig. 3.13) and seems to indicate that double-chain lipids **2** arrange themselves into tubular micellar domains when associated with a plasmid.

3.4.4 Nanoplex Modification for in Vivo Gene Transfer

In vivo studies using a murine model show that nanoplexes synthesised as above have improved pharmacokinetic properties compared with naked DNA, or with polyplexes and lipoplexes, after injection into the blood system. Although the size and stability of these nanoplexes seems particularly well-suited for in vivo applications using intravenous administration, they appear to have no

advantages when it comes to binding to cancerous cells or helping genetic material to reach a cell nucleus.

Given the physical characteristics of nanoplexes, there are several strategies for modifying their surfaces (see Fig. 3.22). To begin with, a lipid can anchor itself by hydrophobic interactions in the micelle phase. Alternatively, the DNA at the nanoplex surface can also be used:

- directly via a DNA binding domain with very high affinity for the DNA minor groove,
- indirectly via DNA-assisted oligomerisation of a monomer into an oligomer with higher affinity.

All these strategies have proved themselves equivalent as regards enveloping nanoplexes with a PEG–folate film. Their biological properties turn out to be compatible with the protective role of the PEG and endocytosis of the nanoplexes after anchoring to specific folic acid receptors.

3.5 Conclusion and Prospects

Chemists have improved the efficiency of vectors by structural modification using standard pharmacological methods relating each structure to a given function. At the present time, the best synthetic gene vectors, using polyplexes or lipoplexes, are able to transfect rapidly dividing cells whilst remaining easy to apply and efficient enough to have a large number of in vitro applications. Even if the colloidal properties of these complexes considerably limit their therapeutic applications (incompatibility in vivo), some invasive dividing cells (tumours) are perfectly accessible by local administration and offer some interesting prospects.

Methods are now emerging for fabricating homogeneous synthetic DNA assemblies with ever better control over the stoichiometry of the various partners in the complex. Hence, it is now possible to make DNA assemblies with new functionalities, and even nanoplexes containing a single genetic copy and enveloped by functional elements. The next step in this field will be to integrate all the functions required for gene delivery in a cooperative manner, from the syringe to the cell nucleus.

References

1. Saenger, W.: *Principles of Nucleic Acid Structure*, Cantor C.A., Springer, Berlin Heidelberg New York (1984)
2. Scherman, D., Bessodes, M., Cameron, B., Herscovici, J., Hofland, H., Pitard, B., Soubrier, F., Wils, P., Crouzet, J.: Application of lipid and plasmid design for gene delivery to mammalian cells, *Curr. Opin. Biotechnol.* **9** (5), 480–485 (1998)

3. Pitard, B., Aguerre, O., Airiau, M., Lachages, A-M., Bouknikachvillit, T., Byk, G., Dubertret, C., Daniel, J-C., Herviou, C., Scherman, D., Mayaux, J.-F., Crouzet, J.: Virus-sized self-assembling lamellar complexes between plasmid DNA and cationic micelles promote gene transfer, *Proc. Natl. Acad. Sci. USA* **94**, 14412–14417 (1997)
4. Boussif, O., Lezoualc'h, F., Zanta, M.A., Mergny, M., Scherman, D., Demeneix, B., Behr, J.P.: A versatile vector for gene and oligonucleotide transfer into cells in culture and in vivo: Polyethylenimine, *Proc. Natl. Acad. Sci. USA* **92**, 7297–7301 (1995)
5. Verkman, A.S.: Solute and macromolecule diffusion in cellular aqueous compartments, *Trends Biochem. Sci.* **27**, 27–33 (2002)
6. Bloomfield, V.A.: DNA condensation, *Curr. Opin. Struct. Biol.* **6**, 334–41 (1996)
7. Hofland, H.E.J., Masson, C., Iginla, S., Osetinsky, I., Reddy, J.A., Leamon, C.P., Scherman, D., Bessodes, M., Wils, P.: Folate-targeted gene transfer in vivo, *Molecular Therapy* **5** (6), 739–744 (2002)
8. Dauty, E., Remy, J.S., Zuber, G., Behr, J.P.: Intracellular delivery of nanometric DNA particles via the folate receptor, *Bioconjugate Chem.* **13**, 831–839 (2002)
9. Nicolazzi, C., Garinot, M., Mignet, N., Scherman, D., Bessodes, M.: Cationic lipids for transfection, *Current Med. Chem.* **10** (14), 1263–1277 (2003)
10. Zuber, G., Zammuto-Italiano, L., Dauty, E., Behr, J.P.: Targeted gene delivery to cancer cells: Directed assembly of nanometric DNA particles coated with folic acid, *Angew. Chem. Int. Ed.* **42**, 2666–9 (2003)
11. Chittimalla, C., Italiano, L., Zuber, G., Behr, J.P.: Monomolecular DNA nanoparticles for intravenous delivery of genes, *J. Am. Chem. Soc.* **127**, 11436–114 (2005)
12. Demeneix, B., Hassani, Z., Behr, J.P.: Towards multifunctional synthetic vectors, *Curr. Gene Ther.* **4**, 445–55 (2004)

Functionalised Inorganic Nanoparticles for Biomedical Applications

E. Duguet, M. Treguer-Delapierre, and M.-H. Delville

The recent development of effective and reproducible techniques has made it possible to synthesise stable aqueous dispersions of individual particles with sizes that can be accurately adjusted from a few nanometers to a few tens of nanometers. These objects are thus small enough to circulate within the human body without causing a risk of embolus, because the narrowest capillaries (those in the lungs) have a minimal diameter of 5 μm . Such particles can also escape from the blood compartment through windows of diameter around 100 nm in certain epithelia with increased permeability, such as those located in tumours and centres of infection, thus favouring their accumulation in precisely these tissues. Finally, the smallest particles can enter cells and their different compartments. Research scientists and doctors thus have new tools at their disposal for understanding biological processes, improving medical diagnoses, and even developing new therapeutic strategies.

Liposomes and particles made from polymers were discussed in some detail in volume II of this series [2], especially with regard to drug targeting. In the present chapter, we shall be concerned with inorganic nanoparticles, such as metal chalcogenides and oxides, and noble metals, whose intrinsic magnetic or optical properties are complementary to the properties of polymers. They are soon expected to play a key role in biological tagging, enhancing contrast in magnetic resonance imaging (MRI), and the hyperthermal treatment of many pathologies, such as cancers.

While the properties of the particle core motivate the choice of a specific type of nanoparticle, surface properties turn out to be equally fundamental. Indeed, it is the surface along with whatever molecular adaptations can be created on it that provides control over the interactions between the particles (single or clustered objects) and the interactions with biological molecules, macromolecules, and cells. Surface functionalisation plays an essential part here, and effective applications of these new diagnostic and/or therapeutic tools will largely be due to progress in this field.

4.1 Synthesis and Chemical Surface Modification of Inorganic Nanoparticles

4.1.1 The Main Strategies

When nanoparticles are produced, the size of the inorganic core must be correctly adjusted to control its intrinsic properties. Synthesis is generally carried out in such a way as to favour particle nucleation rather than particle growth. However, the hydrodynamic volume of the object is just as important, including the inorganic core, molecules grafted onto the surface, and the accompanying water molecules that structure themselves at the periphery, for this predetermines the plasma pathway and hence the final biodistribution of the particles. Hence, once the inorganic core has been synthesised, all subsequent surface modifications must be carried out under optimal dispersion conditions to affect the surfaces of single particles and not those of clusters.

Whatever biomedical application one considers, it must be possible to use inorganic nanoparticles in physiological conditions, i.e., at 37°C in an aqueous medium where the pH is 7.4 (extracellular medium) and the ionic strength is equivalent to 150 mM NaCl. Here, the notion of chemical stability is quite relative: if the nanoparticles cannot be eliminated by natural channels, it is better if they can be biodegraded in the medium term and that the byproducts of this degradation should not be toxic. However, it is essential that, once administered, they should not begin to aggregate, but maintain their colloidal stability (see below).

Colloidal Stabilisation of Nanoparticles in Water

The main attractive forces between nanoparticles dispersed in a liquid are van der Waals forces, i.e., dipole–dipole interactions. In the case of two spheres of radius R separated by a distance d that is much smaller than R , the interaction potential can be written

$$V_{\text{vaw}} = -\frac{AR}{12d},$$

where A is the effective Hamaker constant of the system, depending on the intrinsic Hamaker constants of the relevant solvent and solid. As a general rule, there are two possible strategies for avoiding aggregation of the nanoparticles in water (see Fig. 4.1).

Electrostatic repulsion arises when electric charge is introduced at the surface of the nanoparticles. These charges may come from the adsorption of charged ions at the surface, or they may be due to ionisation of groups such as hydroxyl groups at the surface of oxides. The structure of the solution in the immediate vicinity of the particle surface is in fact relatively complex. It is often described by a double-layer model: the inner layer, called the Stern layer, where the water molecules are highly structured by solvation of surface charges, and the diffuse layer which involves counterions, the latter being subjected simultaneously to electrostatic surface interactions and thermal agitation [3]. This kind of stabilisation works effectively provided that

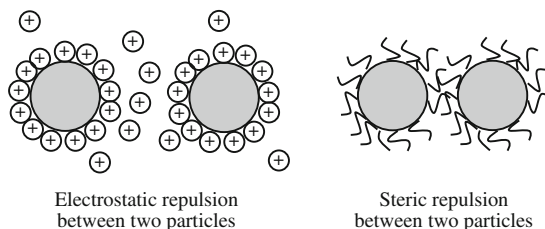


Fig. 4.1. Electrostatic and steric repulsion between two particles

Table 4.1. Possible coupling (bioconjugation) between the surface of an inorganic nanoparticle and the functions available on the biomolecule to be grafted. NHS: *N*-hydroxysuccinimide, SMCC: succinimidyl 4-(*N*-maleimidomethyl)cyclohexane-1-carboxylate, CDI: carbonyl diimidazole

Organic functions commonly grafted at the surface of inorganic nanoparticles to couple with biomolecules	Functional groups of biomolecules through which coupling can occur without affecting their activity	Examples of coupling agents
–COOH	–NH ₂	Sulfo-NHS
–NH ₂	–COOH	Sulfo-NHS
–NH ₂	–NH ₂	Glutaraldehyde
–NH ₂	–SH	Sulfo-SMCC
–SH	–NH ₂	Sulfo-SMCC
–OH	–NH ₂	CDI

the ionic strength of the solution is not too high. When it is high, electrostatic repulsion is significantly reduced, causing irreversible aggregation of the nanoparticles. The pH can also affect this form of stabilisation depending on the kind of ions adsorbed onto the particle surface. The charge of certain ionic ligands depends on the pH of the solution. For example, molecules functionalised by carboxyl groups are negatively charged for pH values greater than 5–6 (COO[–]) and neutral at lower pH values (COOH). Hence, particles stabilised by ligands carrying such a function are only correctly dispersed at neutral or alkaline pH values. The use of electrostatic forces to counterbalance the van der Waals attraction between nanoparticles in water is nevertheless limited to the case where the particles are able to sustain a surface charge, e.g., metal oxides.

Steric repulsion provides a more general method, exploiting the fact that the strength of van der Waals forces falls off very quickly with the distance d . This alternative consists in coating the surface with neutral and hydrophilic molecules that are long enough to maintain a certain distance between the nanoparticles, thereby rendering the van der Waals attraction negligible. A variety of different molecules or macromolecules can be used, provided that they interact strongly with the particle surface and that they cannot form ‘bridges’ between different particles. In a physiological medium, extremely hydrophilic macromolecules like poly(ethylene glycol) (PEG) or dextran are often used to stabilise colloidal particles. The advantage

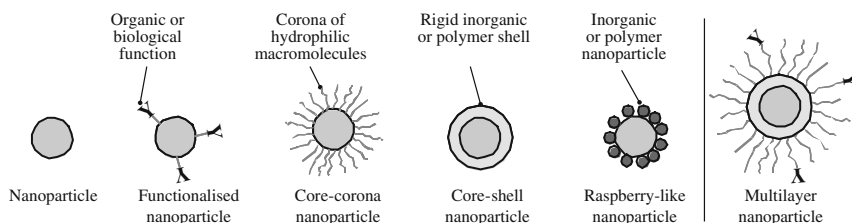


Fig. 4.2. Cross-sectional view of the main morphologies for functionalised nanoparticles used in water in biology and/or medicine. A commonly made distinction is between coronas and shells. A corona is made up of macromolecules solvated in water, which are thus mobile and must be partially bound to the surface by strong interactions, or cross-linked by covalent bonds. In the latter case, the particle is mechanically trapped within the corona. On the other hand, a shell is insoluble in water. If it is made up of macromolecules, these are precipitated onto each other and physically tangled up. All combinations are possible and, depending on the application and associated requirements, multilayer morphologies are often necessary

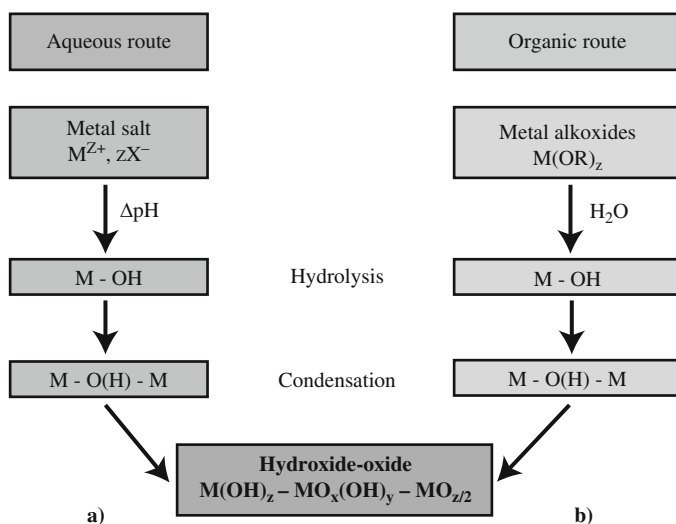


Fig. 4.3. Flow chart for synthesis of metal oxide nanoparticles in solution. (a) Aqueous medium. (b) Organic medium

with this type of stabilisation is that it is a priori insensitive to pH and ionic strength conditions.

By chemical modification of a particle, one understands any treatment leading to the adsorption of a reagent at its surface. Chemisorption occurs when the surface and reagent are bound by covalent bonds (sharing of electron doublets), while physisorption refers to all other cases. Functionalisation can be either direct or intermediary. For example, organic functions such as $-\text{NH}_2$, $-\text{SH}$, $-\text{COOH}$, etc., can only be used for temporary stabilisation of

particles in their medium or as anchor points for biological macromolecules like enzymes, antibodies, and so on, which themselves play a role when the particles are put to use (see Table 4.1).

Figure 4.2 shows the different kinds of morphology that can be obtained by chemical modification of a nanoparticle and which are currently considered to present a potential for biomedical applications. Clearly, in the case of the more complex morphologies, the size of the various components must be controlled in order to regulate the overall hydrodynamic volume, which is generally predefined. The particle size distribution is generally required to be as narrow as possible and, if the method of synthesis does not incorporate any way of controlling it, separation techniques must be brought in.

In the following sections, we shall describe in more detail the synthesis and chemical modification of inorganic nanoparticles currently used or under development in biology and/or medicine.

4.1.2 Iron Oxide Nanoparticles

Core Synthesis and Description of the Surface

Metal oxides M_xO_y are generally prepared by solid-state reaction. Precursor salts are finely milled, mixed, and heated to a high temperature under pressure in an oxygen atmosphere [3]. This procedure must be repeated several times until a homogeneous product is obtained. However, when synthesising nanoparticles, this procedure does not work very well and a chemical approach in solution has been developed. The organometallic method uses metal alkoxides $M(OR)_z$ in an organic solvent (see below) and a more inorganic approach in aqueous solution requires the use of metal salts, e.g., chloride, oxychloride, nitrate, etc. (see Fig. 4.3). These methods are referred to as soft chemistry, because they are generally carried out at room temperature (or a few hundred degrees in the case of hydrothermal synthesis). Metal salts are less expensive and less sensitive to humidity than metal alkoxides, but reactions are more difficult to control. On the other hand, alkoxides do not exist for all transition metals.

Hydrolysis/Condensation of Metal Alkoxides

The most frequently used organometallic compounds are alkoxide precursors, viz., $M(OR)_z$ in which the metal atom $M = \text{Si, Ti, Zr, Al, Sn, Ce, etc.}$, is surrounded by hydrolysable alkoxy groups OR. The coupling between two metal alkoxide molecules occurs in two steps: hydrolysis which forms $M\text{--OH}$ groups and condensation which sets up a hydroxo (olation) or oxo (oxolation) bridge by eliminating a water or alcohol molecule (Fig. 4.4).

When metal alkoxides are used alone in solution, these reactions lead to the formation of oligomers with varying degrees of branching which remain stable in the solution (sol) from the colloidal point of view. If their concentration is too low, they eventually couple together to form a gel and then an oxide after drying and

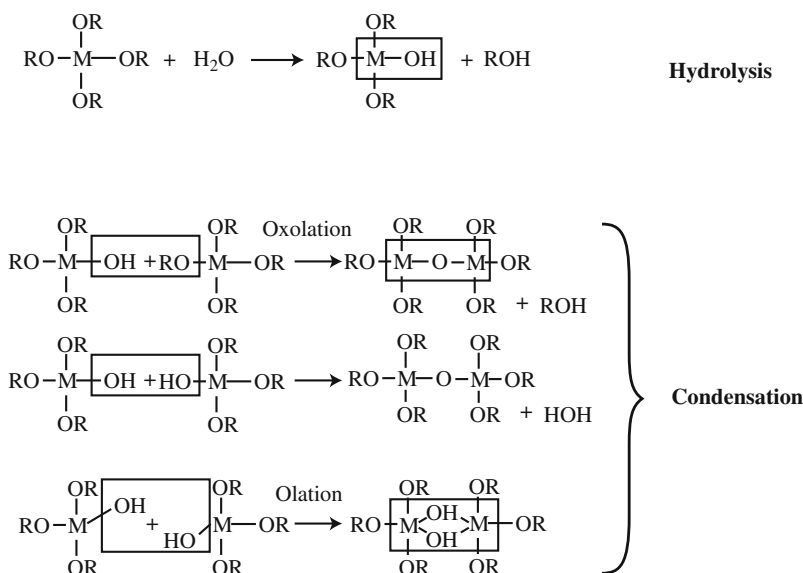


Fig. 4.4. Eliminating a water or alcohol molecule from the most frequently used organometallic compounds

heat treatment. Under certain pH and concentration conditions, they can also yield precipitates or nanoparticles.

This is what happens with iron oxides such as magnetite Fe_3O_4 or maghemite $\gamma\text{-Fe}_2\text{O}_3$, which are exclusively obtained in aqueous solution, in particular by alkaline coprecipitation in water from precursor salts of Fe^{2+} and Fe^{3+} [4]. The size, shape, and chemical composition of the particles is controlled via the type of salts (chlorides, nitrates, sulfates, perchlorates), the $\text{Fe}^{2+}/\text{Fe}^{3+}$ ratio, the pH, and the ionic strength of the medium [5]. To ensure dispersion in water, the nanoparticles are stabilised by means of electrostatic repulsion in a peptisation stage in which either acids or bases selectively attack the surface to create positive or negative charges there, respectively (see Fig. 4.5). Since the particles are magnetic and can move in a magnetic field gradient, this type of dispersion is said to constitute a ferrofluid. Indeed, if the particle concentration is high enough, it is the whole liquid that moves under the effect of an external magnet.

The water/metal oxide interface is a region in which surface cations, in order to satisfy their coordination, carry free hydroxyl groups which terminate the crystal lattice (see Fig. 4.6) [3]. These functions, sometimes referred to as chemisorbed water, interact strongly with water molecules more classically adsorbed by hydrogen bonds (physisorbed water) and structured into several layers. The strength of the interactions, and hence the quality of this structuring, fades out with the number of the layer. The total thickness of

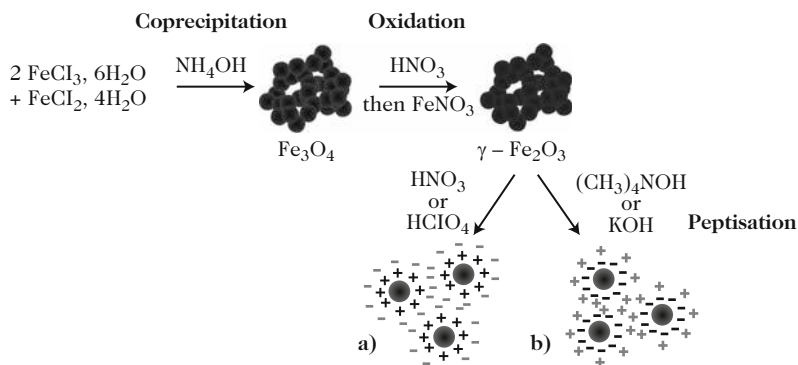


Fig. 4.5. Reaction diagram for the Massart process for preparing ionic ferrofluids of maghemite. (a) Acid cationic ferrofluid and (b) alkaline anionic ferrofluid [4]. The peptisation stage for stabilising the dispersion by electrostatic repulsion is all the more important for iron oxides in that magnetic interactions add to the van der Waals attractive forces in this case. The nanoparticles are thus dispersed in the form of single nanoparticles under pH conditions below 4–5 or above 9–10. Indeed, for iron oxides, it is at pH 7 (the isoelectric point) that the charge density is zero at the surface. Under such conditions, the nanoparticles flocculate

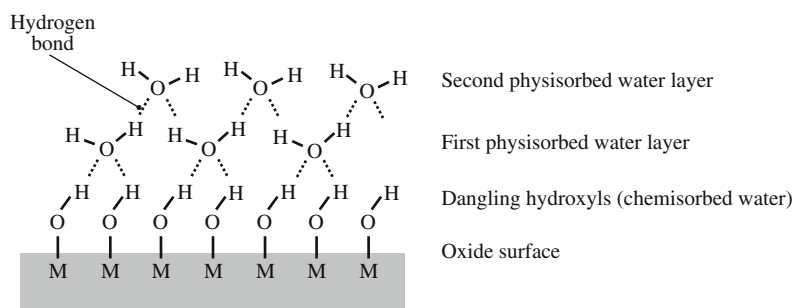


Fig. 4.6. Schematic cross-sectional view of hydroxyls at the water/metal oxide interface. Physisorbed water layers constitute the Stern structured water layer

the hydration layer is directly related to the polarity and to the charge density of the surface. The free hydroxyls explain the reactivity of oxide surfaces. If reagents can reach them, they provide a way of creating covalent bonds, polarised to varying degrees, with the particle surface.

From this point of view, not all oxides exhibit the same surface reactivity, because the kind of metal cation, its degree of oxidation, its coordination number, and its environment all have to be taken into account. While the surface chemistry of silica SiO_2 or alumina Al_2O_3 is a rich area that has been extensively discussed in the literature, that of transition metal oxides such as iron has received much less attention and is not nearly so widely used.

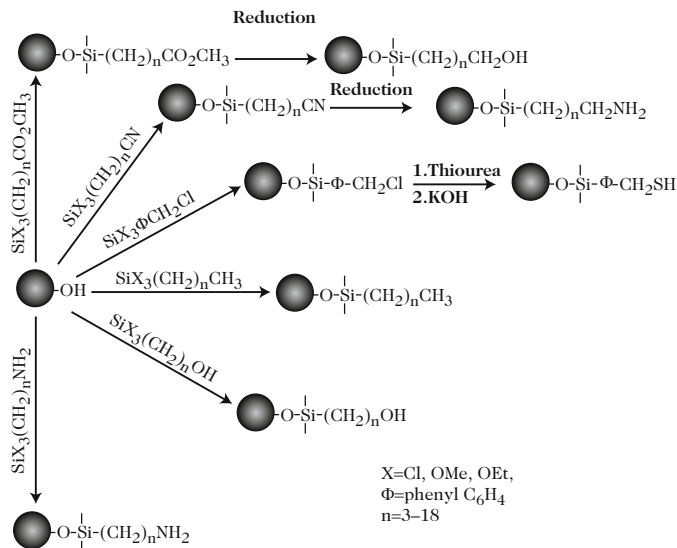


Fig. 4.7. Simplified reaction diagrams showing the diversity of chemical functions accessible at oxide surfaces using the chemistry of alkoxy silanes RSiX_3 . Adapted from [7]

In this chapter, we describe the chemical modification of oxide surfaces by organic and organometallic reagents, and the encapsulation of nanoparticles by hydrophilic macromolecules (core–corona morphology) and silica shells (core–shell morphology).

Chemical Modification by Organic and Organometallic Molecules

Silanisation is undoubtedly the most widespread method for modifying oxide surfaces [6]. In industry, it is used for composite materials to improve interactions between reinforcement and polymer matrix, but also for stationary phases in chromatography, substrates for catalysts, and so on.

Organosilanes are used here. A wide variety of these are available commercially. They have the general formula $\text{R}_n\text{SiX}_{4-n}$, where X is a hydrolysable substituent, e.g., halogeno, alkoxy $-\text{OR}'$, etc. During hydrolysis, a reactive silanol is formed and condenses to make an oxo bridge $\text{M}-\text{O}-\text{Si}$ with free hydroxyls at the surface of metal oxides (see above). The organic group R is not hydrolysable. If it carries a reactive function Y that can react with a third compound, one then speaks of a coupling agent (see Fig. 4.7). When Y carries a positive or negative charge, ionic interactions with the surface can lead to spontaneous ‘upside-down’ adsorption. In order to get the molecule the right way up and force it to graft onto the surface with an oxo bridge, a suitable heat treatment is often required (see Fig. 4.8). In general, the particles are transferred to a heavy liquid like glycerol and the dispersion is outgassed

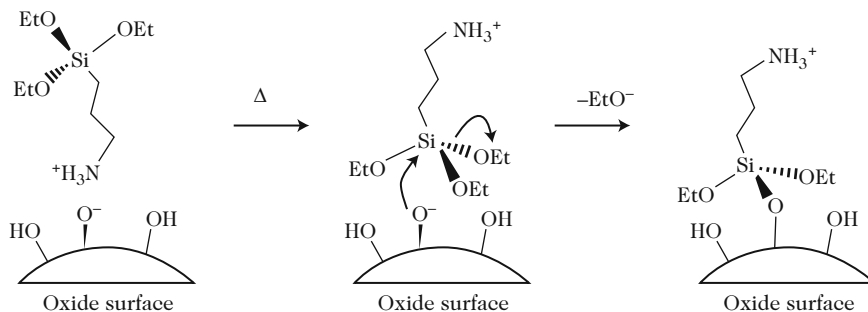


Fig. 4.8. Forced condensation reactions between a molecule of γ -aminopropyltriethoxysilane (APS) and the surface of a metal oxide

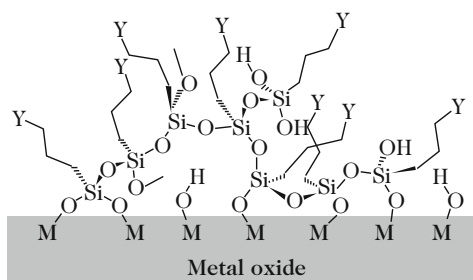


Fig. 4.9. Schematic cross-section of the surface of a metal oxide after chemical modification by an organosilane with formula $(\text{CH}_3\text{O})_3\text{SiCH}_2\text{CH}_2\text{CH}_2\text{Y}$. In the case of iron oxide, the presence of Fe-O-Si bonds remains hypothetical and the success of the grafting process is more readily attributed to the network of Si-O-Si bonds that covers the surface and emprisons the particle

by heating to favour dehydration and elimination of alcohol molecules. Once oriented toward the liquid, the charged Y functions are often very effective in redispersing the modified particles in water and preserving their colloidal stability by electrostatic repulsion.

Organosilanes with $n = 1$ are the most widely used. At oxide surfaces, they also set up oxo bridges with their neighbours and hence graft in several layers, leading to the formation of a highly cross-linked polysiloxane film which covers the whole surface of the oxide. In the case of iron oxide, for which the formation of the Fe-O-Si bond has never been clearly demonstrated, it is through this network of siloxane bonds that one explains the durability of the chemical treatment by organosilanes (see Fig. 4.9).

When grafting reactions are carried out in water, self-condensation of the organosilanes in the solution is often quicker than condensation on the surface. It is thus the condensates that graft onto the surface, and whose polycondensation is finalised by ripening at higher temperature (see Fig. 4.10).

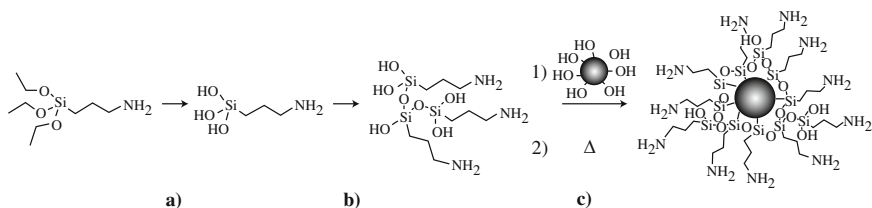


Fig. 4.10. Reaction diagram showing the chemical modification of oxide particles by reaction with γ -aminopropyltriethoxysilane (APS) in an aqueous medium: (a) Hydrolysis in solution. (b) Self-condensation in solution. (c) Grafting at the surface. Adapted from [8]

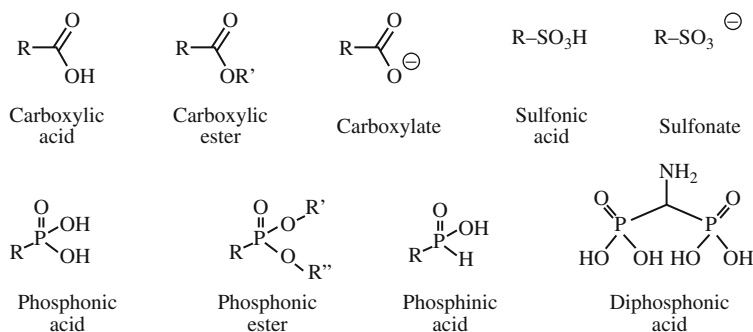


Fig. 4.11. Main coupling agents apart from organosilanes

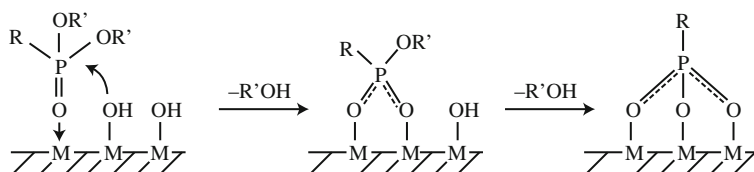


Fig. 4.12. Proposed mechanism for grafting a phosphonic ester to the surface of a metal oxide. Adapted from [9]

Other reagents that can efficiently modify iron oxide surfaces are carboxylic, sulfonic, and phosphonic acids (see Fig. 4.11) [9]. The particular interest of phosphonic acid derivatives lies in the fact that they cannot self-condense, even in an aqueous medium, to form P–O–P bridges. They therefore adsorb to form monolayers according to a coordination mechanism (see Fig. 4.12). In the case of iron oxides, grafting is apparently not tridentate, but bidentate [10]. Furthermore, the derivatives of diphosphonic acid, which should in principle provide more efficient anchoring, are currently under investigation [11].

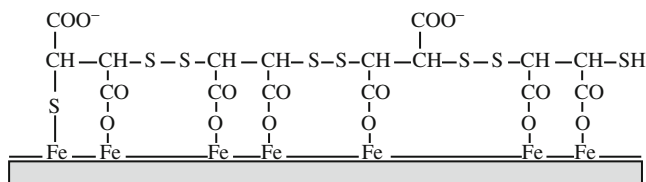


Fig. 4.13. Schematic cross-section of the surface of an iron oxide after chemical modification by dimercaptosuccinic acid. The carboxylic groups contribute both to anchoring the molecule and to stabilising the particles electrostatically. The thiol functions also undergo partial oxidation by the surface Fe^{3+} ions, leading to self-condensation of the molecules and formation of a polydisulfide film. Residual thiol groups can later serve to attach biomolecules

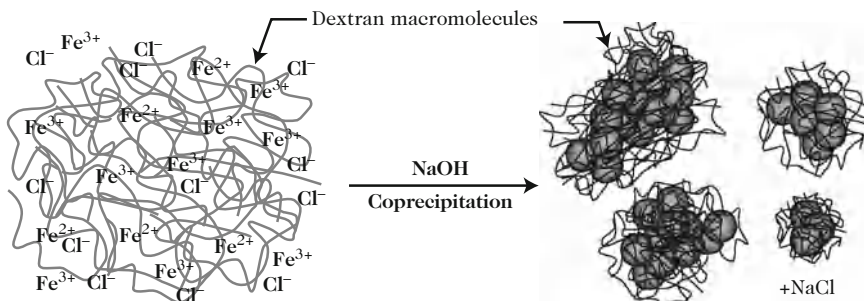


Fig. 4.14. Simplified reaction diagram for direct synthesis of magnetite (Fe_3O_4) nanoparticles stabilised by dextran via the Molday process [13]

Carboxylic acids are often used in the form of multifunctional derivatives, such as dimercaptosuccinic acid $\text{HOOC}-\text{CH}(\text{SH})-\text{CH}(\text{SH})-\text{COOH}$ (see Fig. 4.13) [12].

Encapsulation by a Corona of Hydrophilic Macromolecules

A corona of hydrophilic macromolecules serves both to stabilise the particles in water via steric repulsion effects (see p. 130) and also to control the biodistribution of the nanoparticles in vivo (see Sect. 4.3.1). These macromolecules can be of natural origin, such as dextran, a polysaccharide obtained by bacterial fermentation of sucrose, followed by hydrolysis and fractionation. Synthetic macromolecules are also widely used, in particular poly(ethylene glycol) or PEG with formula $\text{HO}-(\text{CH}_2-\text{CH}_2-\text{O})_n-\text{H}$.

The simplest method here involves carrying out synthesis of the iron oxide nanoparticles in an aqueous solution in which the macromolecules have already been dissolved (see Fig. 4.14). They are then naturally adsorbed onto the surface of the magnetic cores. They also play a structuring role by limiting particle growth. This phenomenon is explained by the increased viscosity of

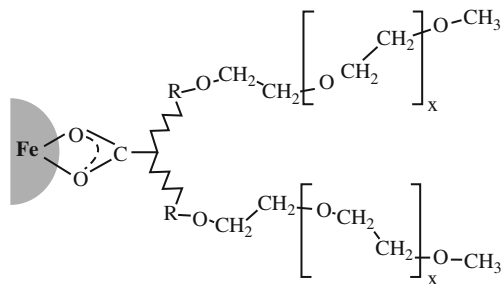


Fig. 4.15. Schematic view of the surface of an iron oxide nanoparticle after reacting with a triblock copolymer comprising a polyurethane central segment carrying carboxylate functions and PEG end blocks. Adapted from [14]

the medium, which slows down the diffusion of metal cations, and the adsorption of macromolecules onto nuclei, thereby hindering their growth. The drawbacks with this method are:

- partial aggregation of nanoparticles depending in particular on the molar mass of the macromolecules,
- polydispersity in the hydrodynamic volume of the synthesised particles, requiring subsequent column separation,
- weak interactions between particles and macromolecules, sometimes leading to progressive desorption as time goes by and/or in the presence of other compounds.

To avoid the last effect, macromolecules can also be cross-linked by creating chemical bonds between them, which amounts to mechanically trapping the particle within the corona.

With a view to better controlling aggregation effects and strengthening the interactions between particles and macromolecules, multi-stage methods of synthesis have been devised. The first stage consists in synthesising the nanoparticles and stabilising them by electrostatic repulsion. The next step is either to have them react with macromolecules carrying functions able to interact strongly with the surface (see Fig. 4.15), or else to use silane or other coupling agents, that will react with the macromolecules in a third stage. These methods are very efficient and do not require subsequent separation if the transition from electrostatic stabilisation to steric stabilisation is successful.

Encapsulation by a Silica Shell

Because the surface chemistry of iron oxide is less well known than that of silica, some effort has been invested in attempts to encapsulate these magnetic cores in a silica shell. It is then the surface of the encapsulated particle that is functionalised. Silica also has the advantage that its isoelectric point occurs at pH 2. Hence, the presence of a silica layer at the surface means that

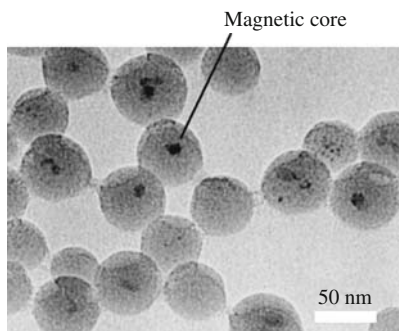


Fig. 4.16. Transmission electron microscope image of maghemite nanoparticles coated with a layer of silica (core-shell morphology). Adapted from [16]. Reproduced with the kind permission of ecmjournal.org

subsequent chemical modifications can be carried out at neutral pH, because under these conditions the silica particles are stabilised by negative surface charges, whereas the naked iron oxide cores would flocculate (see Fig. 4.15).

There are several ways of encapsulating iron oxides in silica. One of these begins by letting a thin film of sodium silicate adsorb onto the maghemite surface to sensitise it [15]. The nanoparticles are then transferred to an ethanol solution containing ammonia and a tetraalkoxysilane $\text{Si}(\text{OR})_4$, which hydrolyses and polycondenses at the surface according to a similar mechanism to the one described on p. 133. Another method uses the same chemical reactions, but in order to confine the reaction to the neighbourhood of the nanoparticles, the synthesis is carried out in a reverse microemulsion [16]. The ammoniated aqueous dispersion of nanoparticles is emulsified in heptane in the presence of a surfactant. The magnetic cores are isolated from one another in an aqueous envelope containing ammonia. The alkoxysilane is then introduced into the heptane phase, where it is gradually hydrolysed by contact with the water surrounding the particles. Once it has become hydrophilic in this way, it enters this region to polycondense around the nanoparticles (see Fig. 4.16).

4.1.3 Semiconductor CdSe Colloids

Over the last few years, many different ways of fabricating semiconductor nanoparticles, often called quantum dots, have been devised. In this chapter, we shall be concerned mainly with type II–VI semiconductors, i.e., involving elements from columns II and VI of the periodic table, whose synthesis in nanoparticle form has been much more developed than that of type III–V semiconductors, for example. Techniques for producing CdSe nanocrystals have received particular attention, because they emit light in visible wavelengths. Unlike oxides, for which electrostatic particle stabilisation can at least be considered by virtue of the acid–base properties of the surface hydroxyls, metal selenides are not naturally susceptible to this manner of stabilisation. For this reason, metal selenide nanoparticles must be stabilised by means of steric repulsion effects.

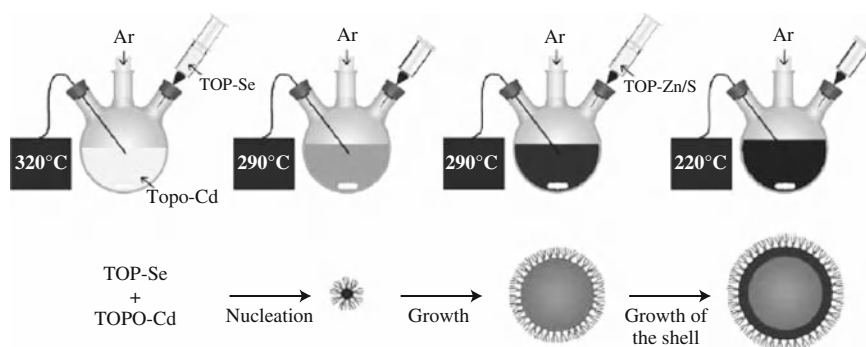


Fig. 4.17. Experimental setup and schematic view of quantum dot synthesis. CdSe core in a ZnS shell. From [18]. Copyright Elsevier (2004)

Fabricating Semiconductor Cores

To take advantage of their remarkable optical properties, it is essential to synthesise perfectly uniform particles, in terms of shape, size, and surface state. To obtain a size dispersion less than 5%, the best conditions for synthesis are a temperature range 250–300°C in a complexing organic medium of molecules such as trioctylphosphine oxide (TOPO), with formula $\text{O}=\text{P}[(\text{CH}_2)_7\text{CH}_3]_3$, or trioctylphosphine (TOP) (see Fig. 4.17) [17]. The stabilising mixture can also include alkylamines and/or carboxylic or phosphonic acids. The structuring role played by these molecules allows one to control the light-emitting core: growth stops when the particle reaches the size for which the envelope of organic molecules is at its most stable. If the chalcogenide/ligand ratio is high, larger particles are obtained, but they are less stable because there are fewer complexing molecules. Conversely, a low ratio favours stabilisation of very small particles (less than 3 nm).

When the synthesis is complete, the organic molecules (TOPO, TOP) remain adsorbed, the polar head being bound to the particle surface and the hydrocarbon chain being directed outward, thereby ensuring colloidal stability in organic media by steric repulsion. However, they cannot saturate all the dangling bonds (Cd^{2+} and Se^{2-}). The latter add to the crystal defects, which trap charge carriers and contribute to limiting the fluorescence quantum yield (see Sect. 4.2). In general, the latter does not exceed 10%.

Improving Light Emission by Surface Passivation

Two approaches have been devised to control the confinement of the exciton within the nanocrystal and thereby increase the fluorescence quantum yield:

- encapsulation in a shell made from a material with greater bandgap energy than the core (see p. 147) [19],

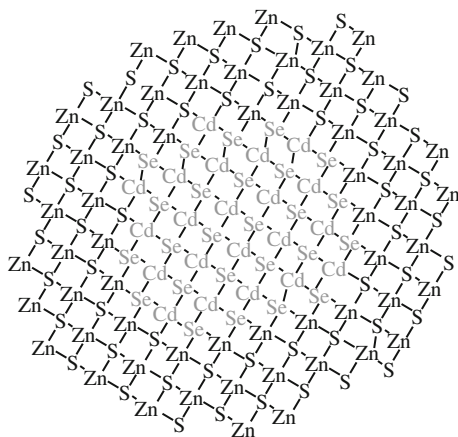


Fig. 4.18. Schematic cross-section of a core-shell nanoparticle made from CdSe and ZnS. Since ZnS does not have exactly the same crystal lattice as CdSe, adding too many layers will lead to the formation of defects in the ZnS shell and a gradual drop in luminescence yield

- doping of particles by a light-emitting element to which excited carriers transfer their energy.

The first approach is the most often used. It involves passivating the surface by saturating all the dangling bonds, those of both the metal and the chalcogenide. The lattice parameter of the shell must be close to that of the core to facilitate epitaxial growth (see Fig. 4.18).

The shell is generally grown in an organic solution by precipitation, using the core particles as heterogeneous nucleation sites. This reaction does not affect molecules adsorbed onto the surface, which continue to fulfill their stabilising role. The optimal thickness of the shell is controlled by the difference in lattice parameters (the lattice mismatch) between the constituents of the core and shell. Above a certain thickness, the materials can no longer absorb the elastic strains resulting from the lattice mismatch and the structure relaxes by creating defects, e.g., dislocations, which open up new, non-radiative channels and tend to reduce the quantum yield of the particle. For example, the passivation CdSe nanocrystals proves to be optimal for a ZnS thickness between one and two atomic layers. The quantum yield can then be pushed up to 85% [20].

Transferring Nanoparticles to an Aqueous Medium

For biological purposes, semiconductor nanocrystals require subsequent functionalisation so that they can be dispersed in the physiological aqueous medium. This passivation should not modify the size distribution, upon which

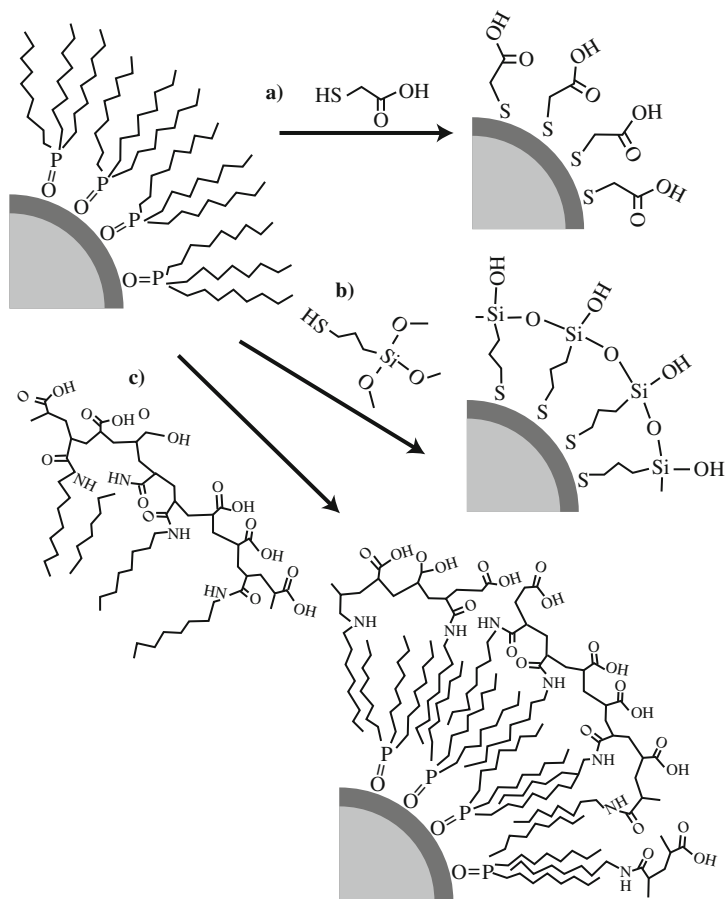


Fig. 4.19. Different strategies for dispersing semiconductors in water when they have originally been prepared in an organic medium and stabilised by the TOPO ligand. **(a)** Swapping TOPO for mercaptoacetic acid. **(b)** Swapping TOPO for γ -mercaptopropyltrimethoxysilane. **(c)** Hydrophobic interactions with polyacrylic acid partially modified by octylamine. From [18]

the optical properties depend. Several strategies can be considered to make these initially hydrophobic particles hydrophilic (see Fig. 4.19).

The simplest approach consists in replacing the organic molecules that coat the particle surface by hydrophilic molecules. For example, a molecule carrying thiol functions, such as mercaptoacetic acid, is known to replace TOPO molecules and adsorb specifically onto zinc atoms in the outer layer of CdSe/ZnS nanocrystals. The latter are then surrounded by molecules with carboxyl groups at the end, which ensure particle dispersion through the repulsive interactions of the carboxylate ions, but also conjugate with biomolecules (see Fig. 4.19a). However, dispersion is only achieved for weakly acidic, even

alkaline pH, and for a small number of bonds in order to preserve a maximum of carboxylate ions. This method for functionalising the surface is easy to implement, but the particles formed in this way turn out to be rather unstable. The ZnS–thiol bonds are gradually broken by hydrolysis or photo-oxidation of the thiol group and the particles end up clustering together after a few days [21]. Strategies currently under development exploit the idea of attaching hydrophilic stabilising ligands able to establish several anchoring points on the particle surface [22].

The second, more sophisticated approach is also based on the idea of replacing the stabilising ligands: the organic molecules are replaced by tri-alkoxysilanes $(\text{RO})_3\text{Si}-(\text{CH}_2)_3-\text{Y}$, where Y is typically an $-\text{SH}$, $-\text{NH}_2$, $-\text{PO}(\text{O}^-)\text{CH}_3$, or $-\text{COOH}$ group, able to interact with the ZnS surface. The alkoxy functions can then hydrolyse and trigger polycondensation of these molecules, with the formation of a silica ‘monolayer’, which is functionalised by other alkoxy silanes to allow subsequent bioconjugation (see Fig. 4.19b). This multi-stage approach is more difficult to implement, but the anchoring of the hydrophilic layer is in principle more stable.

The third strategy consists in modifying, rather than replacing, the hydrophobic molecules around the nanoparticle surface. The idea is to encapsulate the particles with a mixture of amphiphilic macromolecules, which associate with the TOPO molecules by hydrophobic interactions. These macromolecules can be polyacrylic acid partially modified by octylamine, for example (see Fig. 4.19c) [23]. The main advantage of this approach is that it avoids the rather delicate step in which hydrophobic and hydrophilic ligands are exchanged. Furthermore, it provides a way of stabilising the nanocrystals for several months, even in a physiological medium. However, it involves a series of steps in which functionalised molecules are grafted to the surface and this leads in the end to much bigger particles than those initially created in the organic phase. Starting with particles having core diameters of 3–5 nm, the final object has a hydrodynamic diameter of 15 nm.

4.1.4 Noble Metal Nanoparticles: Gold and Silver

The fabrication of metal colloids has been known since Faraday (1857) and has been widely developed and described over the last few decades by a large scientific community. Biomedical research uses mainly gold or silver colloids. The idea is to reduce a precursor metal salt (sulfate, nitrate, perchlorate, etc.), in a single or double phase medium, using chemical agents or other reduction techniques (photochemical, electrochemical, or radiolytic) in the presence of structuring and stabilising molecules (see Fig. 4.20). Ligands can have differing degrees of complexity, like those of organic or organometallic dendrimers, for example. Particles can be spherical, or have a more novel morphology, e.g., rods, triangles, wires, and there are even particles with four arms [24].

In contrast to semiconductor nanocrystals, stable, monodispersed metal nanoparticles can be fabricated just as easily in an aqueous medium (preferred

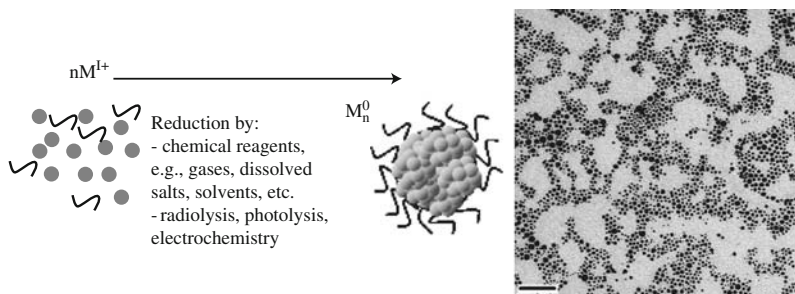


Fig. 4.20. Synthesis of metal nanoparticles in a liquid phase. Transmission electron microscope image of gold nanoparticles stabilised by mercaptoundecanoic acid. *Scale bar:* 50 nm. From [25]. Copyright Wiley-VCH Verlag GmbH & Co. KGaA. Reproduced with kind permission of the publisher

for the applications we are concerned with here) as in an organic medium. However, there are two advantages when synthesising in an organic medium:

- solutions can be made with a high concentration of particles,
- synthesis can be carried out at temperatures above 100°C (boiling point of water), so the range of precursor metal salts and stabilising ligands can be extended, and the number and types of crystal defects in the particles can be reduced (see Fig. 4.20).

After synthesis, the ways of functionalising the metal nanoparticles and transferring them from the organic to an aqueous medium are similar to those described above for semiconductor nanocrystals (see p. 143) [26]. The formation of a silica shell around gold and silver nanoparticles is widely used, applying a strategy similar to the one discussed on p. 140 [27]. The thickness of the silica layers can vary from a few nanometers to a few tens of nanometers depending on the particle and alkoxide precursor concentrations.

4.2 Biological Tagging in Vitro and in Animals

Nanomaterials made from semiconductors and metals are commonly used in biology and medicine as specific tags for various entities such as proteins, cells, nucleic acid fragments, viruses, and so on (see Fig. 4.21). For effective tagging, these objects must fulfill several conditions: they must produce a strong and long-lasting signal, they must be dispersible in an aqueous medium, and it must be possible to bind them specifically to the relevant biomolecule without perturbing its normal activity. Progress in chemistry has provided ways of fabricating and functionalising these nano-objects.

Over the past few years, semiconductor and metal nanocrystals have been used to improve tools for in vitro use. With recent progress in detection techniques, they are now even used in vivo in certain animal models to gain a better understanding of the evolution mechanisms underlying certain pathologies.

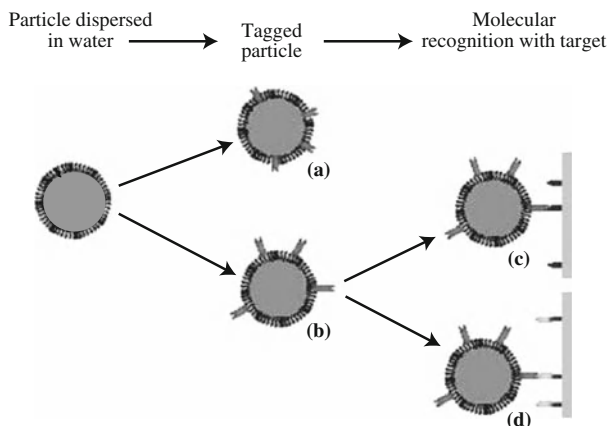


Fig. 4.21. Biological tagging using nanoparticles. If the tag is to be specific, the relevant biomolecules, e.g., antibodies, must be bound to the nanoparticles, either directly to the inorganic core (a), or to the hydrophilic stabilising agent (b). The tagged particle then conjugates with the specific receptor of the target, either by direct fusion with the receptor (c) or via a specific ligand previously bound to the target (d). Adapted from [28]

However, owing to the high level of toxicity of metal chalcogenides, semiconductor nanocrystals cannot yet be used for human beings.

4.2.1 Biological Tagging by Semiconductor Colloids

The most commonly used tagging and imaging technique used in biology is fluorescence microscopy, for which the usual markers are molecular fluorophores or fluorescent proteins (see Fig. 4.22). However, the rapid photodegradation (or photobleaching) of these tags means that long observation sequences of biomolecules are not possible. Semiconductor nanoparticles provide an alternative to organic fluorophores that are unstable with regard to photobleaching. Their properties come from their nanoscopic dimensions. Their spectral characteristics can be adjusted simply by modifying these dimensions (see below).

Light Emission and Optical Characteristics of Semiconductor Nanocrystals

The physical properties of a semiconductor crystal are radically changed when its dimensions are reduced to a few nanometers. In macroscopic semiconductors, electrons do not occupy discrete levels, but rather two broad energy bands called the valence and conduction bands. They absorb light over a broad range of colours, provided that the photons have enough energy to cross the gap separating the two bands. On the other hand, they only emit light at a specific wavelength corresponding to the gap energy.

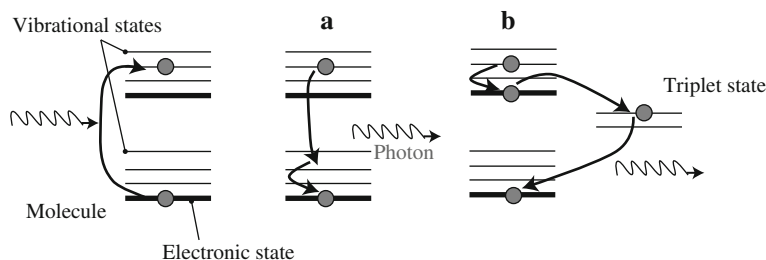


Fig. 4.22. Fluorescence by molecular fluorophores. Fluorescence is a form of light emission which occurs when a body is illuminated. Unlike phosphorescence, fluorescence stops suddenly when the light stimulation is switched off. Many molecules, like dyes for instance, are fluorescent. In order to stimulate fluorescence, they are placed under a laser beam at the appropriate frequency. When the energy of a laser photon is absorbed, one of the outer electrons of the molecule goes from the ground state to one of the higher levels. When this electron drops back down to its ground state, it restores the energy in two ways: in a non-radiative form via vibrations of the molecule, and in the form of a photon (a). Hence, some of the light energy absorbed by the molecule is dissipated in the form of vibrations. The fluorescence photon thus has lower energy than the laser photon, i.e., it is shifted toward the red. In most molecules, another mechanism reduces fluorescence. Instead of dropping back directly into the ground state, the electron goes through an intermediate state, the triplet state (b). This state is metastable, meaning that the electron only remains there for a limited time, during which the fluorescence emission is interrupted (photobleaching)

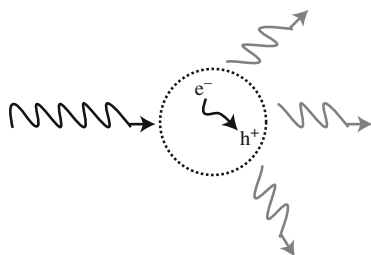


Fig. 4.23. Light emission by a semiconductor nanoparticle

The properties of semiconductor crystals of nanometric size are governed by the laws of quantum mechanics. They are characterised by discrete energy levels. Under the action of light, an electron-hole pair is created (the hole being denoted by h^+) which interacts to form an exciton (see Fig. 4.23). The exciton energy involves the kinetic energy, the confinement potential, and a Coulomb interaction term between the hole and electron. When semiconductor crystals have sizes less than the Bohr diameter (11.2 nm for CdSe), the exciton is in the strong confinement regime. The energy levels of the electron and hole are then those of a free particle enclosed inside a spherical quantum 'box'. The exciton then has discrete energy levels and can be likened to a pseudo-atom, whence the name quantum dot. In this approximation, the energy difference between the ground state and excited states increases in inverse proportion to the particle size [29].

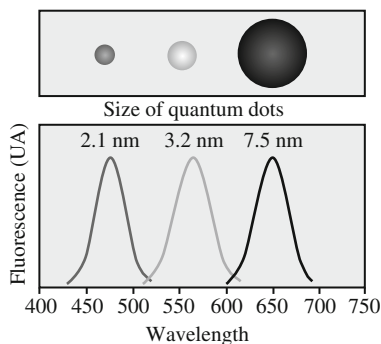


Fig. 4.24. Size effect on the fluorescence spectrum of CdSe nanocrystals. The colour of colloidal suspensions goes from dark red to orange, then yellow, for particles with diameters between 2.5 nm and 7.5 nm. From [18]

At room temperature, a photon with energy greater than the bandgap energy of the semiconductor creates an exciton which is thus carried to a higher excited level. From there, it relaxes non-radiatively by molecular coupling into the first excited state. The electron and hole then recombine, thereby generating a photon (light emission). However, the quantum efficiency of a nanocrystal never exceeds 10%. As explained in Sect. 4.1.3, this low yield is essentially due to the poorly controlled surface state. In the excited state, carriers are delocalised as far as the particle surface. Surface defects can then trap the carriers, which recombine radiatively far from the gap or in a non-radiative way. This trapping effect effectively reduces the quantum efficiency. In order to confine carriers within the crystal, the nanocrystals must be coated with a material with a higher bandgap than the core. This surface passivation reduces the number of traps and makes them less accessible to the carriers, thereby increasing the quantum efficiency.

Like semiconductors with larger dimensions, semiconductor nanocrystals absorb all photons whose energy is greater than their bandgap. However, the wavelength of emitted photons depends on the size of the nanocrystals. Hence, a single semiconductor material can generate a whole family of markers with different colours. One only has to make nanocrystals of different sizes (see Fig. 4.24).

Semiconductor nanocrystals can withstand many excitation and light emission cycles. Typically, the number of photons detected before photodestruction is of the order of 10^6 for a dye molecule, whereas it can be as high as 10^8 for a quantum dot [30]. This stability means that biologists can monitor what is happening in cells and tissues over longer periods of time (see Fig. 4.25). Their quantum efficiencies for light emission are comparatively high. The other advantage is that their absorption spectrum is very broad, in contrast with their emission spectrum, which remains the same whatever the excitation wavelength within the absorption band. Quantum dots thus provide a means for multicolour and simultaneous tagging of different species contained in cells with a single excitation wavelength. As biological systems are extremely complex, biologists often seek to observe several components at the same time. Up until now, this kind of monitoring was difficult to arrange, because each organic dye had to be excited at a specific wavelength that differed from the one required by the others. By tagging biological molecules with semiconductor

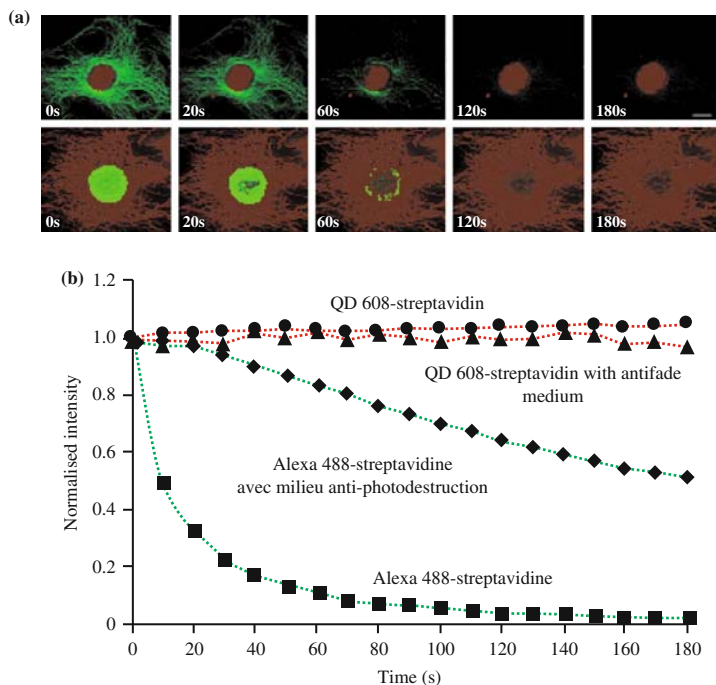


Fig. 4.25. Comparative study of the photostability of semiconductor nanocrystals (*red*) and the dye Alexa 488 (*green*). (a) Fluorescence images of 3T3 cells taken at different times over a total period of 3 min (excitation at 505 nm). (b) Time dependence of the intensity. From [35]. Copyright Elsevier

nanocrystals of different sizes, a single light source is sufficient to be able to monitor all the molecules simultaneously.

In parallel, progress made in methods for biofunctionalising nanocrystals mean that these objects are now commercially available. Today it is thus possible to buy semiconductor nanocrystals coated with macromolecules and coupled with antibodies or streptavidins which carry specific functional groups for targeting membrane proteins (Quantum Dot Corp., Hayward, California). By virtue of all these developments, a wide variety of biological systems, such as enzymes, proteins, nucleic acids, DNA, and so on, have already been tracked, and different types of investigation, such as the detection of particular gene sequences and protein conformations, and monitoring of enzyme reactions, have been carried out *in vitro* [31], and more recently *in vivo* in animals [32–34]. The latter studies are quite remarkable, because it is often difficult to visualise living systems owing to the optical noise caused by intrinsic fluorescence of the cell medium.

Semiconductor nanocrystals are currently the most effective tags we possess from the point of view of resistance to photobleaching. However, there are a small number of disadvantages. Their size of around fifteen nanometers, once functionalised, cannot be neglected in some applications. For example, nanocrystals used as tags for receptors of glutamate diffusing on neurons in a synaptic environment have some difficulty in passing through the synaptic gap, which varies between 20 and 30 nm. Furthermore, semiconductor nanocrystals tend to blink on and off in a way which complicates the tracking of biomolecules, especially when trying to reconstruct their trajectories. It also complicates use of the emitted signal to monitor time-dependent processes. Although still poorly understood, this blinking is attributed to ionisation of the nanocrystal, probably when an electron (or hole) transits by tunnel effect into non-radiative traps situated at the surface of the nanocrystal [36]. Once ionised, the nanocrystal is in a dark state in which it no longer fluoresces. Electron-hole pairs created by the excitation energy recombine with the remaining charge by interaction processes involving several particles and give up their recombination energy without photon emission. Finally, since semiconductor nanocrystals are made up of toxic elements such as Cd and Se, their use for observing *in vivo* phenomena in humans raises a certain number of toxicological questions. Several toxicological tests have been carried out with CdSe and CdSe/ZnS nanoparticles. They show that prolonged exposure of these nanocrystals to both UV radiation and water can lead to surface oxidation and release of Cd²⁺ ions [35]. The level of toxicity depends sensitively on the way the tags are administered in cells and also on the nature of the hydrophilic shell encapsulating the nanocrystals. An alternative solution consists in enclosing the particles by a stabilising ZnS shell. However, some work remains to be done before obtaining a completely biocompatible marker. At the present time, although semiconductor nanocrystals cannot be said to represent a universal biological marker, they nevertheless represent a promising candidate for use *in vivo* in the near future.

4.2.2 Biological Tagging by Metal Colloids

Metal nanoparticles, mainly gold, and more recently silver, constitute another generation of biological labels. In comparison with organic structures and semiconductor nanocrystals, they are non-toxic, extremely resistant to corrosion, and have an optical response without fluctuation in time. Furthermore, they exhibit excellent stability and many possibilities for chemical surface functionalisation. Like semiconductor nanocrystals, they are mainly used for tagging biomolecules. Initially followed by electron microscope, they were then detected optically, whence they could be used to study living systems.

As early as the 1970s, gold nanoparticles were already used in immunochemistry for *in situ* localisation of tagged macromolecules [37,38]. Provided with outer groups for conjugation with the chosen biomolecules, gold nanoparticles (0.8 or 1.4 nm in diameter, containing 11 or 67 atoms, respectively) were

used and detected by electron microscope. Given their small size, these objects are able to penetrate to a depth of 40 μm in biological tissues. This technique thus gave access for the first time to a considerable amount of information with nanometric resolution. It provided a way, and still does so, to see more clearly and hence understand better the behaviour of biomolecules such as proteins, cells, DNA, etc. However, this technique only furnishes a static view of biological systems. Since it can only be used with electron microscopy, it is incompatible with any study on living systems.

With *in vivo* studies in mind, a significant effort was devoted to finding more flexible detection methods, in particular all-optical methods. The advantage with optical detection of biomolecules is to avoid mechanical contact in 'remote' studies. Such techniques were finally developed in the 1980s. They exploit the rather special optical properties of metal nanoparticles. These specific properties stem from the existence of a resonance in their absorption spectrum: the surface plasmon resonance. The latter corresponds to a coherent oscillation of the conduction electrons generated by their interaction with an external electromagnetic field [1, Chap. 7]. Excited near their plasmon resonance, metal nanoparticles exhibit strong Rayleigh scattering. They are thus highly sensitive probes for analysing biomolecules [39]. It has been possible to considerably reduce the optical detection threshold for biological structures and hence to study living systems.

Detection techniques have progressed so far that it is now possible to observe isolated proteins in cells and follow their diffusion over a lapse of time. The advantage in detecting isolated objects is that one can ascertain individual heterogeneous behaviour that would be concealed in an average measurement over an ensemble, e.g., specific fluctuations due to interaction of a molecule with its immediate environment. This last approach really does provide a new way of looking at biological systems, because one can follow the time development of objects on the scale of single molecule. However, there is one major limitation due to the weak signal of the single object under investigation, which must be isolated from all background signals due to the surroundings. Given that the intensity of the scattered light drops off rapidly with the radius R of the particle (as R^{-6}), it is currently impossible to detect gold particles with diameters less than 40 nm. Detection is made even more difficult in a scattering medium like a cell. This means that the use of optical imaging *in vivo* is limited to exploring tissues accessible at the surface of a small animal.

A more recent and rather promising breakthrough for visualising living systems is detection by the photothermal effect [40]. This technique detects local heating induced by light absorption by a metal nanoparticle. Indeed, metal nanoparticles exhibit relatively high absorption cross-sections. They emit little or no light. The absorbed energy excites conduction electrons in the metal. The electrons then thermalise via electron–electron and electron–phonon interactions. Phonons in the metallic crystal lattice interact with phonons in the surrounding medium, giving rise to heat transfer. Hence, the metal particles

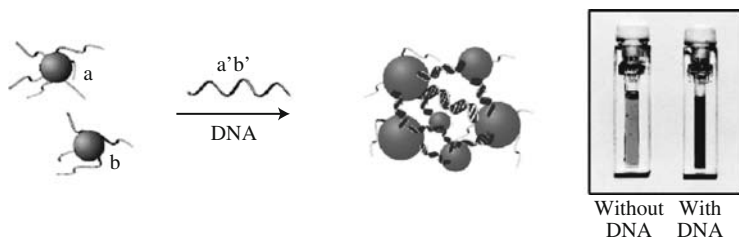


Fig. 4.26. Using gold nanoparticles to detect DNA sequences. From [35]. Copyright 2004, with the kind permission of Elsevier

become heat sources, inducing a local temperature rise in their immediate environment, and a consequent modification of the refractive index of the medium. The interference method is the most sensitive for detecting such a local variation. At the present time, this technique has been used *in vitro* to visualise biological systems comprising fixed cells tagged by gold nanoparticles of diameter 5 nm. The temperature rise is estimated at a few kelvin for this size of particle, a value that is undoubtedly a little too high to be acceptable, since there is a risk of damaging biological media. However, by improving the detection sensitivity, it should become possible to reduce this level. It is one of the only available methods able to image metal nanoparticles in scattering media. It should thus soon be possible to develop *in vivo* imaging and tagging of cells.

Apart from their use as tags, metal nanoparticles can also be used to determine the composition of biological samples. A technique has recently been devised for testing the presence of a specific genetic sequence in a solution, e.g., a pathogenic bacterium [41]. It uses gold nanoparticles of diameter 13 nm on which are attached DNA strands. The trick is to use two types of gold particle. Particles in the first group carry half of the target sequence while those in the second group carry the other half (see Fig. 4.26). The two particles associate with the target. Each particle carries several strands and therefore binds to target sequences, which thus form a kind of a bridge between particles. One by one, the particles thus begin to stick together to form a lattice. The optical properties of the nanoparticles then change. When the particles are dispersed randomly, they produce a blue colour. The gold particles reveal at a glance the presence or otherwise of a given sequence. This technique looks extremely useful for quick, portable DNA testing.

With this new technique, scientists hope above all to achieve progress in medicine, with faster diagnoses so that disease can be identified in its early stages, e.g., in the case of cancer, heart disease, or viral infections, but also the development of new treatments involving biomolecules.

Another application of gold colloids can be found in Chap. 5, which discusses molecular motors.

4.3 In Vivo Applications

Intravascular administration (by injection or perfusion) is in principle the most general method for reaching any given organ or tissue, because all cells are either directly or indirectly supplied by the blood flow (see below) [42]. In order to be administered by this channel, nanoparticles must not only be sure not to stick together in the blood or undergo premature decay of any kind, but they must also be able to get around the complex mechanisms set up by the body's immune system, for which they immediately become the public enemy number one! It is thus essential to begin by examining the plasma elimination mechanisms to be faced by these nanoparticles, along with factors affecting their biodistribution, and solutions devised to provide them with the necessary stealth with regard to the immune system. We then discuss ways of attaining given targets as specifically as possible. Finally, we will be able to describe current and projected medical applications for nanoparticles administered intravenously, with regard to both diagnosis and therapy.

The Blood Compartment

The blood circulatory system comprises two subsystems: the cardiovascular system and the lymph system. The lymphatic vessels collect the lymph, which surrounds organs and recovers some of the waste products from cells. It is very rich in lymphocytes and is filtered in the lymphatic ganglions before being rejected into the blood compartment just before the heart. The lymphatic system plays an important role in the immune system because, during any kind of infection, the infectious agent very quickly ends up in the lymph.

Blood is a tissue made up of cells distributed throughout a liquid called the plasma. The whole ensemble is transported through the circulatory system in a one-way flow guided by the contractions of the heart. The plasma is made up of water, inorganic salts, and organic molecules such as sugars, lipids and amino acids. The main blood cells are listed in Table 4.2.

The inner walls of the blood vessels are composed of vascular endothelium, a specialised tissue layer. The cells there are joined together into a single structure by intercellular junctions. The endothelium can withstand significant mechanical strains (particularly important near the heart) and remain permeable to water, substances with low molar mass, and white blood cells. At the blood–brain barrier (BBB), tight junctions between the endothelial cells appear to prevent any intercellular transfer of molecules, forcing them to cross the cells which thus seem to play the part of highly selective filters.

4.3.1 Fate of Particles in the Blood Compartment

Mononuclear Phagocyte System and Hepatosplenic (Passive) Targeting

It is the mononuclear phagocyte system (MPS), also called the reticulo-endothelial system, which actively extravasates any foreign body of an infectious kind, or not part of the blood system, thereby constituting the front

Table 4.2. Main cell types in the blood compartment

Blood cells	Size [μm]	Concentration [mm^3]	Lifetime [day]	Features and main functions
Red blood cells or erythrocytes	7–8 diameter 1–2 thick	4–6 million	120	• Biconcave disk without nucleus • Deformability to pass through vessels of different diameters • Contains hemoglobin (red pigment) which combines with oxygen in lungs to form oxyhemoglobin • Transports oxygen to all tissues
White blood cells or leucocytes		4,000–10,000		• Nucleus with several lobes, wrongly suggesting that they had several nuclei, whence the name polynuclear cells
Neutrophils	10–12	1,800–7,000	1–4	• 45–70% • Involved in defense mechanisms, especially against bacteria
Eosinophils and basophils	12–18	60–350		• 1–3% of leucocytes • Role poorly understood
Lymphocytes	6–15	1,500–4,000		• 20–40% of leucocytes • Can also circulate in lymphatic system • Coordinate immune response
Monocytes	15–18	100–700		• 3–7% of leucocytes • Circulatory form of mononuclear phagocyte system
Platelets or thrombocytes	1.5–2	150,000–300,000	8–12	• No nucleus • Involved in blood clotting

line of the immune defense system. It comprises an army of macrophage cells arranged in a lattice and positioned at strategic points in the organism. They are mainly found in the bone marrow where they are produced, the blood in which they circulate in the form of monocytes, the alveoli in the lungs, the spleen, and especially the liver where they are known as Kupffer cells.

Their role is to recognise and eliminate from the blood compartment any senescent cells, micro-organisms, and particles. In particular, in the latter

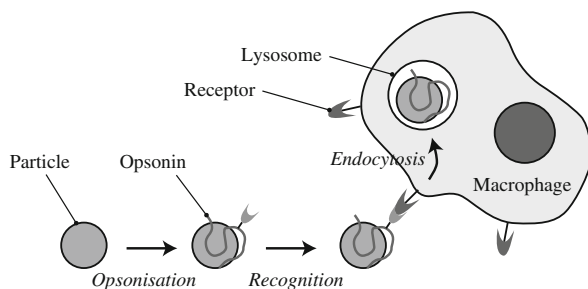


Fig. 4.27. Mechanisms used by the MPS for recognition and clearance of nanoparticles arriving accidentally in the blood compartment

case, the blood clearance process is often triggered by a preliminary step called opsonisation, in which circulating proteins (various subclasses of immunoglobulins, elements of the blood complement, fibronectin, etc.) adsorb onto the particle surface (see Fig. 4.27). Labelled in this way, the particles are then recognised by macrophages via specific receptors. They are subsequently internalised by endocytosis and accumulate in lysosomes where they are eventually degraded by lysosomal enzymes. Hence, within a few minutes, the particles are eliminated from the blood compartment and find themselves fixed in the liver (up to 90%), the spleen, and to a lesser degree in the bone marrow. One can speak of the hepatosplenic biodistribution.

The unpromising fate of the nanoparticles can in fact be turned to advantage when specifically targeting these organs to diagnose or treat pathologies that concern them. One then speaks of hepatosplenic or passive targeting, because the MPS spontaneously takes control. However, if one has in mind other organs than the liver and/or the spleen, there is no option but to inhibit or at least drastically slow down the opsonisation process. The problem here is to make particles stealthy with respect to the MPS and thereby increase their probability of reaching the target.

Designing Particles with Prolonged Intravascular Lifetime

To extend the plasma half-life of a particle, it has been clearly established today that one must hinder or even prevent the adsorption of opsonins by adjusting its size and surface properties [43]. For example, it has been shown that, the smaller the particle radius, the longer the particles can circulate within the blood compartment, because the value of the radius of curvature affects the nature and/or quantity of adsorbed opsonins. In addition, the lower the surface charge density, the longer the particles can remain in place. Finally, hydrophobic particles are so quickly eliminated from the blood compartment that it would seem that the preliminary opsonisation stage is not even necessary for their clearance. In contrast, particles with a rather hydrophilic surface have every chance of circulating for longer, just like the red blood

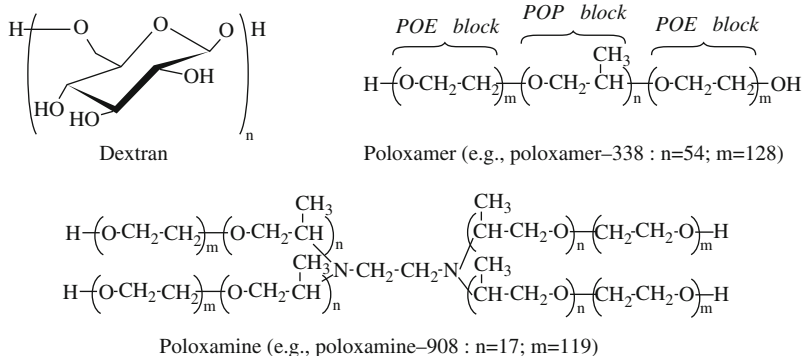


Fig. 4.28. Hydrophilic macromolecules in a corona formation around particles injected into the blood compartment extend the plasma half-life of the particles

cells themselves, whose lifetime of 120 days is explained by the presence of a hydrophilic barrier of oligosaccharide groups on their surface. In the case of synthetic particles, this physicochemical idea of steric repulsion of proteins can be reproduced by coating the particle surface by flexible and hydrophilic macromolecules. Arranged rather perpendicularly to the surface, they form a kind of brush which serves as an effective steric barrier.

The macromolecules used are generally polysaccharides like dextran. It is produced by bacterial fermentation of sucrose, followed by hydrolysis and separation to give batches of different molar mass (see Fig. 4.28). Other macromolecules of biological origin have been used, such as polysialic acid or heparin, but their development has been hindered by the fact that they can be expensive and there is sometimes a risk of immunological consequences. As a consequence, synthetic macromolecules have received much more attention. For example, poly(ethylene glycol) PEG, which has the chemical formula $\text{HO}-(\text{CH}_2-\text{CH}_2-\text{O})_n-\text{H}$, is widely used in galenic pharmacy, where it is commonly conjugated with active ingredients (small molecules or peptides, proteins, antibodies, oligonucleotides) to reduce their immunogenicity, increase their plasma half-life (by reducing the rate of renal clearance) and hence also their bioavailability [44]. This approach is so commonplace that it has become known as 'PEGylation'. PEG can be anchored at the surface when the particles are synthesised, grafted on later by one of its ends, or simply physisorbed. In the last case, it is then used in the form of block copolymers, the main families being the poloxamers and the poloxamines (see Fig. 4.28). It is the most hydrophobic poly(oxypropylene) block (POP) that interacts with the particle surface, leaving the poly(oxyethylene) blocks (POE) to orient themselves spontaneously toward the outside. For maximal efficiency with regard to opsonisation, the molar mass of PEG/POE chains must lie in the range 2,000–5,000 g/mol.

It is thereby possible to extend the plasma lifetime of these so-called stealth particles by a few hours. If they are made small enough, they can also find their

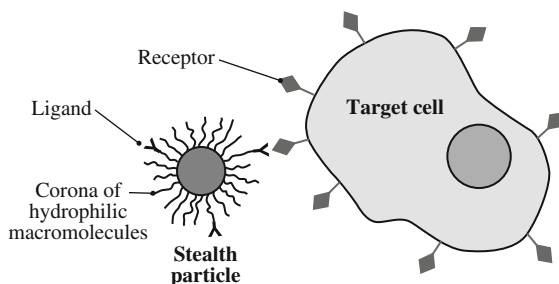


Fig. 4.29. Using molecular recognition to get a nanoparticle to a target cell. The ligand grafted onto the particle surface must be specific to receptors at the surface of the target cell

way out of the blood compartment in regions where the vascular epithelium has discontinuities. Indeed, the integrity of the endothelial barrier is often perturbed near centres of infection and in the vicinity of some types of tumour. Stealth particles can thus passively target these tissues and accumulate there.

In addition, it has been shown that stealth particles can penetrate the brain tissue of health animals even though the blood–brain barrier (BBB) turns out to be impenetrable for the majority of therapeutic molecules. The mechanism for crossing this barrier has not been completely understood yet, but these results open the way to promising diagnostic and therapeutic applications.

Active Targeting Via Molecular Recognition Ligands

In order to target some particular population of cells, e.g., tumour cells, a specific strategy must be worked out in each case. This is known as active targeting. It consists in attaching a molecule at the particle surface that is capable of binding in a specific way with the surface receptors of the target cells, using a molecular recognition mechanism such as the antigen–antibody interaction (see Fig. 4.29). It must be possible to graft these ligands at the particle surface in such a way that they do not lose their targeting function after grafting, while preserving the stealth of the whole entity. This is why antibodies are still rarely used, being bulky (around 20 nm), costly to synthesise, and potentially risky in terms of immunogenicity. They no doubt also have the disadvantage that they are too specific for antigenic epitopes which change in time. This is why peptides, sugars, or small molecules like folic acid are preferred at the present time. Folic acid is a group B vitamin essential to the cell division mechanism. Folic acid receptors are thus overexpressed at the surface of cells that need them most, such as tumour cells [45]. Hence, nanoparticles decorated with folic acid could be used specifically to target tumours.

4.3.2 Tools for Medical Diagnosis: MRI Contrast Agents

Magnetic resonance imaging (MRI) is certainly the most effective non-invasive radiological technique available to medicine at the current time [42, 46]. It is particularly useful for assessing diseases of the bone, articulations, and central nervous system. The spectacular development of MRI over the past thirty years is due to the discovery of materials with low resistivity for the design of superconducting magnets and also the improvement of computer facilities for data storage and processing.

Magnetic Resonance Imaging

MRI exploits the nuclear magnetic resonance (NMR) properties of components of the human body and especially protons in water contained within tissues, lipid membranes, proteins, and so on [47]. The underlying principle is the same as for NMR spectroscopy in chemical analysis, combining a strong static magnetic field B_0 (up to 2 T in standard hospital equipment) and a perpendicular radiofrequency field (5–100 MHz). After the radiofrequency pulse, the spins of the protons seek to realign with B_0 . This relaxation phenomenon can be decomposed into two independent mechanisms: longitudinal relaxation, corresponding to a gradual increase in the longitudinal component of the magnetisation, and transverse relaxation, which is a gradual decay of the transverse component. They are characterised by relaxation times T_1 (the time required for 63% of the longitudinal component to reestablish) and T_2 (the time required for 37% of the transverse component to disappear), respectively.

In order to reconstruct a 3D image, the NMR signals are collected in each volume element (voxel) of the sample and correlated with their original coordinates. To do this, B_0 or the rf field must vary in space in such a way that to each voxel there corresponds a specific resonance frequency. The relaxation values are processed by a 2D Fourier transform. By adjusting the various parameters, in particular the repeat and echo times, the operator can obtain images weighted in T_1 or in T_2 . Since the tissues, fats, fluids, and so on, do not have the same relaxation times, they can then be differentiated. For example, since fluids have very long T_2 , images weighted in T_2 are used to detect certain pathologies such as internal bleeding and cancerous lesions.

With the development of the MRI technique, it became apparent that exogenous contrast agents could be used to obtain a better view of tissue boundaries, and hence enhance the efficiency of the diagnosis. These act indirectly, because they are substances with magnetic properties that increase relaxation rates of nearby protons. Hence, in MRI images, it is not the contrast agents themselves that are visualised, but rather a consequence of their presence. This property is called relaxivity and is defined by $R_1 = 1/T_1$ and $R_2 = 1/T_2$. The different contrast agents available are described in the following sections. They are now used in 40% of MRI scans. Current effort aims

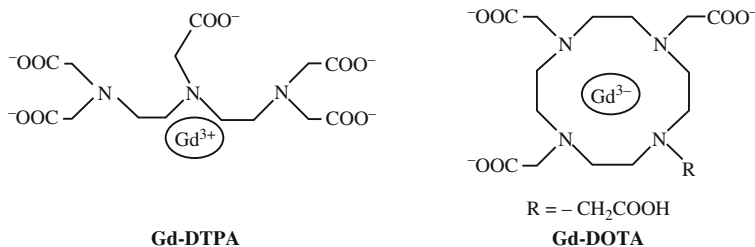


Fig. 4.30. Chemical structure of the gadolinium chelates most commonly used as T_1 contrast agents in MRI

to create contrast agents carrying ligands, which will thus be able to target specific cell populations or tissues and thereby reduce the doses to be administered and enhance the signal from the target. This is the burgeoning field of molecular imaging, concerned with biological phenomena on the scale of the cell or the molecule, and including also *in vivo* techniques like positron emission tomography (PET), single-photon emission computed tomography (SPECT), near-infrared optical imaging and scintigraphy.

Paramagnetic Contrast Agents (T_1 Agents)

The first generation comprises strongly paramagnetic ions such as Gd^{3+} (7 unpaired electrons). They are stabilised in the form of molecular chelates to reduce their intrinsic toxicity (see Fig. 4.30). They accelerate T_1 relaxation of protons with which they are in direct interaction. Hence, in T_1 weighted images, they contribute to signal enhancement (positive contrast). They have been used clinically since the end of the 1980s to label intravascular and extracellular spaces (kidney functioning, BBB integrity, etc.). They are administered at an average concentration of 0.1 mmol/kg of patient. The main problems are a rather short plasma half-life (70–100 min), due to their low molar mass, and a relatively low contrast due to the fact that there is only one paramagnetic ion per molecule.

Paramagnetic Metal Chelates Trapped or Grafted onto Particles

To extend the intravascular lifetime of gadolinium chelates and simultaneously increase their local concentration, the main strategies involve grafting them along macromolecules such as dextran [48] or onto liposome [49] or metal oxide [50] surfaces, or confining them in polymer particles [51]. In the last case, they are core-shell particles in which the polyacrylic acid core is able to complex Gd^{3+} ions, while the more hydrophobic shell is porous and controls access of water molecules to the Gd^{3+} ions (see Fig. 4.31). The size of the nanoparticles, less than 120 nm, allows them to pass easily through the vascular system. The observed reduction in T_1 is proportional to the level of Gd^{3+} ions in the particle.

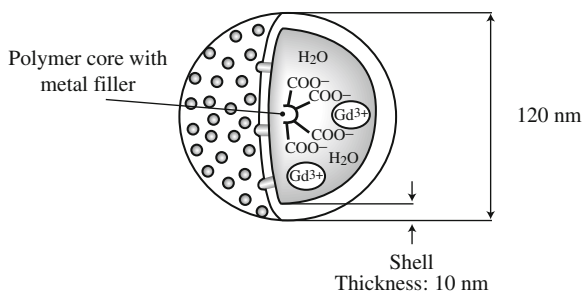


Fig. 4.31. Core-shell polymer nanoparticle containing Gd³⁺ ions and used as a T_1 contrast agent for MRI. Taken from [51]

Magnetic Susceptibility Contrast Agents (T_2 Agents)

This is a new generation of contrast agents, now used in hospitals. They are in fact iron oxide nanoparticles, i.e., magnetite Fe₃O₄ or maghemite Fe₂O₃- γ , with diameters in the range 3–10 nm, stuck together to varying degrees and encapsulated in a hydrophilic dextran corona [13, 52]. Since they are smaller than a magnetic domain, they lose their magnetisation as soon as the magnetic field is switched off. Their magnetic moment is nevertheless much higher than that of paramagnetic compounds, which is why they are called superparamagnetic compounds.

They are commonly referred to as (ultrasmall) superparamagnetic iron oxides, or (U)SPIOs for short. They are very simply obtained by a single-stage process involving alkaline coprecipitation of Fe(II) and Fe(III) precursors in an aqueous solution of dextran [13] (see Fig. 4.14). The role of the macromolecules is

- to limit particle growth,
- to stabilise particles sterically,
- and later, in vivo, to hinder opsonisation effects.

There are administered by perfusion at an average concentration of 1 mg of iron per kg of patient. After endocytosis by macrophages, they end up being metabolised in the lysosomes. Hence, after solubilisation, the metal ions join the iron pool of the organism, estimated at 3,500 mg per person.

Relaxivity measurements show that, unlike gadolinium chelates, (U)SPIOs have high and different R_1 and R_2 values (see Table 4.3). In most cases, they are used for their efficiency in reducing the signal in T_2 weighted images (negative contrast). Expressed simply, their effect is explained by non-uniformities they create in the field around them and the consequences this has on water molecules which diffuse through these regions (phase shifts which tend to shorten T_2). Unlike paramagnetic agents, they exert their effects at a distance without the need for any direct contact with the water molecules.

Table 4.3. Relaxivities measured at 37°C and clinical dose (per kg of patient) of MRI contrast agents

Contrast agent	Commercial example	B_0 [T]	R_1 [mM ⁻¹ s ⁻¹]	R_2 [mM ⁻¹ s ⁻¹]	Dose
Gd-DTPA	Magnevist ^a	1.0	3.4	3.8	0.1–0.3 mM/kg
Gd-DOTA	Dotarem ^b	1.0	3.4	4.8	0.1 mM/kg
SPIO	Endorem ^b	0.47	40	160	15 μM _{Fe} /kg
USPIO	Sinerem ^b	0.47	23	49	2.6 mg _{Fe} /kg

^a Schering, Berlin; ^b Guerbet, Roissy

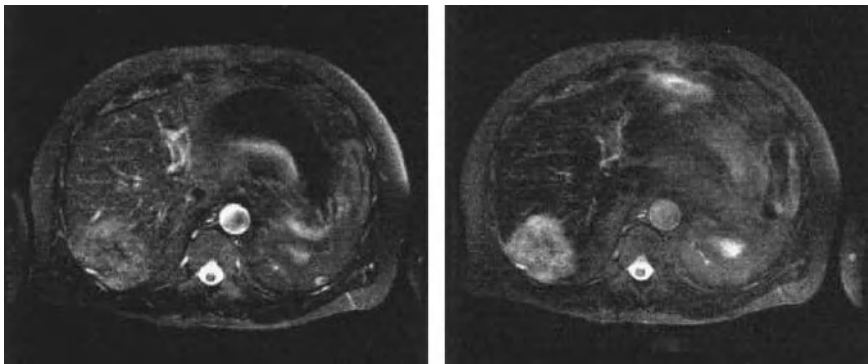


Fig. 4.32. T_2 weighted MR images of a liver affected by metastases, showing the effect of magnetic susceptibility agents on the contrast. *Left:* Natural contrast reveals the metastases in the form of a positive signal. *Right:* The presence of SPIO (Endorem) effectively removes the signal from the healthy parts of the liver and makes diagnosis easier. From [54]

Using several separation stages, the size polydispersity of (U)SPIO can be reduced and a median value chosen. SPIO has hydrodynamic volume greater than 40 nm and, despite the presence of dextran, accumulates rapidly in the organs of the MPS, with a plasma half-life of less than 10 min. They can be used to image the liver, where malignant tumours and metastases, which are typically stripped of Kupffer cells, show up in T_2 weighted images in the form of hyperintense lesions (Fig. 4.32 left). These show up more clearly when the signal from the healthy part of the liver is reduced using SPIO (Fig. 4.32 right). USPIO, also called monocrystalline iron oxide nanocompounds or MION, has lower hydrodynamic volume. Its plasma half-life is longer than 2 h and allows imaging of blood vessels (angiography). The smallest particles can escape from the blood compartment via the interstitium and enter the lymph system. After drainage or capture by macrophages, they accumulate in the lymph nodes where they then allow imaging (lymphography) [53]. Signal collapse indicates

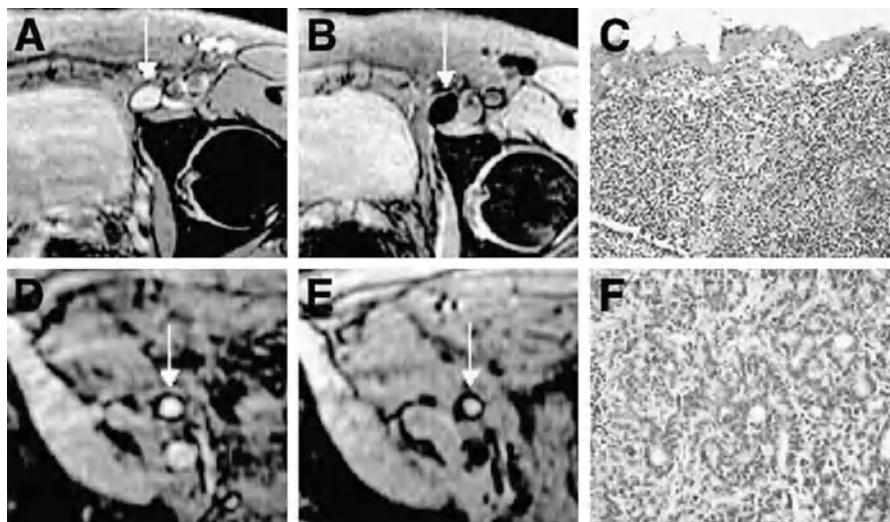


Fig. 4.33. T_2 weighted MR images of iliac lymph nodes (*arrow*) in a patient suffering from a prostate cancer, showing the effect of magnetic susceptibility agents on the contrast. In the series (A)–(C), the ganglion is healthy. Natural contrast (A) shows the lymph node in the form of a positive signal. 24 h after intravenous administration of the USPIO Sinerem (B), the signal decreases in a uniform way, demonstrating the accumulation of nanoparticles and hence the integrity of the lymph node functioning. Image (C) confirms by a histological cross-section ($\times 125$, hematoxylin and eosin staining). In comparison, in the series (D)–(F), the lymph node has been affected by the tumour and its permeability has been increased in consequence. Administration of USPIO has no effect on the image, because the nanoparticles cannot accumulate there [image (F), $\times 200$]. From [55]. Copyright 2003 Massachusetts Medical Society

normal functioning of the ganglions, while a positive signal attests to capillary permeability characteristic of a tumour (see Fig. 4.33).

Because the interactions between the dextran and the magnetic cores are weak, essentially of van der Waals type or hydrogen bridges, chemical modification of the dextran is a delicate matter that often leads to a depletion effect (macromolecular desorption). To deal with this problem, which represents a serious obstacle to efficient and reproducible grafting of ligands, the research effort has been aiming for better cohesion in the dextran/iron oxide system. A first approach consists in cross-linking the dextran macromolecules by means of epichlorohydrin, mechanically trapping the magnetic cores and producing cross-linked iron oxide or CLIO (see Fig. 4.34). Another approach uses silane coupling agents of the form $(\text{RO})_3\text{SiCH}_2\text{CH}_2\text{CH}_2\text{NH}_2$ which can polycondense at the oxide surface and also form reducible Schiff bases with previously oxidised dextran macromolecules. This new generation is called versatile USPIO or VUSPIO (see Fig. 4.35). The first attempts to graft ligands (IgG antibodies, folic acid, etc.) showed that, in vitro (cell cultures)

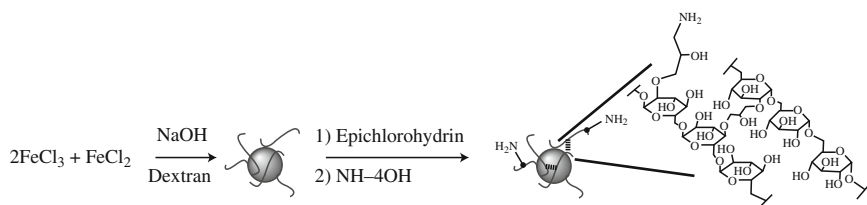


Fig. 4.34. Single-stage synthesis of iron oxide nanoparticles encapsulated in a cross-linked dextran corona and functionalised by amine groups. Adapted from [56]

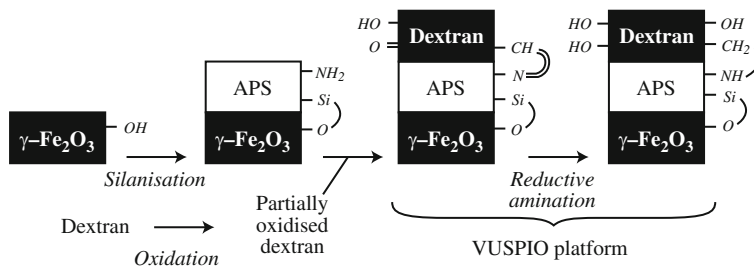


Fig. 4.35. Multi-stage synthesis in which dextran can be covalently grafted onto maghemite nanoparticles. The coupling agent used here is γ -aminopropyltriethoxysilane (APS). Adapted from [57]

and/or in vivo (animal models), contrast agents can be targeted. However, the antibody doses required to obtain a good enough contrast are dissuasive from a commercial point of view.

4.3.3 Therapeutic Tools

It has been known for over 5,000 years that heat can be used to treat a great many illnesses, and in particular some cancers [42, 46]. In modern oncology, hyperthermia is one of the four main therapeutic solutions, along with surgical removal, radiotherapy, and chemotherapy, which are often combined. The most recent hyperthermic techniques fall into three categories:

- Contact from outside with a hot liquid.
- Heating without contact using devices that transmit energy from a distance, e.g., ultrasound, microwaves, radio frequencies or infrared radiation.
- Implantation of optical fibres, antennas, probes or mediators in the human body which are able in vivo to transport or convert an energy supply controlled from the outside into heat.

These techniques generally require a considerable effort to implement them, and surgery is required for some of them. In addition, they are far from being completely and universally effective, i.e., they cannot be applied to all types of cancer, but depend on its localisation, extent, and state of advancement.

In this context, the idea of using small enough particles to reach all tissues via the blood compartment, equipped with ligands specific to the target cells and responsive to remote heating, has gradually moved forward over the past few decades. Magnetic nanoparticles were a natural choice for research interest, because the human body does not contain magnetic materials and, under certain amplitude and frequency conditions, they can release heat when excited by an alternating magnetic field. More recently, gold nanoparticles have been developed for the purposes of photothermal treatment [40].

Magnetic Hyperthermia

In order to create a magnetic field, the magnetic component of the radiation must be favoured over the electric component, by using a solenoid. From a physiological point of view, acceptable frequencies must be higher than 50 kHz to avoid neuromuscular electrostimulation phenomena, and less than 10–100 MHz to enter sufficiently deeply into the body. In addition, the product $H\nu$ of the magnitude H of the magnetic field and the frequency ν must be kept below 4.85×10^8 A/m s to allow 1 h sessions without discomfort to the patient. Indeed, it is impossible to completely eliminate the electric component of the radiation, and this can cause general heating of body tissues and fluids by eddy currents.

There are three mechanisms for heating the magnetic particles, depending on the field parameters H and ν , the particles themselves (intrinsic magnetic properties and surface chemistry), and the medium in which the particles are dispersed. Above about ten nanometers in diameter, the particles are ferro- or ferrimagnetic and dissipate heat by hysteresis loss. The amount of heat released is then proportional to the frequency and the area of the hysteresis cycle. When the particle diameter is less than 10 nm, the area of the hysteresis cycle is zero and it is Néel relaxation that controls heating. Finally, whatever the size of the particles, they can also choose to oscillate under the effect of the alternating field, and heating then results from friction with the surrounding liquid (Brownian relaxation). The latter phenomenon is then highly sensitive to the viscosity of the medium, the presence of molecules or macromolecules at the particle surface, and the question of whether the particle is fixed to its target or not. Whatever mechanism is involved, there is, for each homogeneous batch of nanoparticles, an optimal frequency for which the specific loss power (SLP) is maximal. The efficiency of nanoparticles in this task can be compared via the SLP (Table 4.4). It can be measured quite simply by recording the temperature rise of an aqueous colloidal dispersion in a calorimeter as time goes by. At a given temperature, it is then proportional to the slope of the curve and the average heat capacity C of the medium:

$$\text{SLP} = C \frac{\Delta T}{\Delta t}.$$

Today, we are still a long way from understanding and controlling these phenomena, because existing theories are rarely confirmed by experiment. Current

Table 4.4. Specific loss power (SLP) of different dispersions of magnetic nanoparticles. Results from different authors are difficult to compare because the experimental conditions are not the same [42]

Magnetic nanoparticle			Experimental conditions			
Magnetic compound	Core diameter [nm]	Corona	H [kA/m]	ν [kHz]	Dispersion medium	SLP [W/g _{metal}]
Fe ₃ O ₄	100–150	–	7.2	880	Phys. serum	45 ± 3
γ-Fe ₂ O ₃	100–150	–	7.2	880	Phys. serum	42 ± 3
Fe ₃ O ₄ (Endorem)	6	Dextran	6.5	300	Water	< 0.1
γ-Fe ₂ O ₃	< 10	–	8.0	1,000	Water	115
γ-Fe ₂ O ₃	< 10	Dextran 9,000 g/mol	8.0	1,000	Water	170
γ-Fe ₂ O ₃	< 10	Dextran > 70,000 g/mol	8.0	1,000	Water	400
γ-Fe ₂ O ₃	3	Dextran	12.5	500	Water	106
γ-Fe ₂ O ₃	5	Dextran	12.5	500	Water	524
γ-Fe ₂ O ₃	7	Dextran	12.5	500	Water	626

effort is as much concerned with the synthesis of nanoparticle batches with controlled granulometry, surface chemistry, and colloidal stability as with the development of specific frequency scanning devices able to determine the optimal frequencies.

With regard to clinical testing, the first human tests were carried out in Berlin in 2003. Since the economic stakes are no doubt considerable, the exact chemical nature of the nanoparticles used was not revealed. It is probably magnetite (Fe₃O₄) with a surface treatment for presenting amine functions. Under physiological pH conditions, they are therefore protonated. The dispersion is stable in a physiological medium, but could not be used for intravenous administration, because these are not stealth particles. They are therefore injected directly into the tumour, where they are captured in large numbers by the tumour cells. The therapy consists in alternating radiotherapy and hyperthermia sessions in a custom-built human-sized machine ($H \sim 10$ kA/m and $\nu \sim 100$ kHz). The results are very encouraging and must be considered as a first step.

Future developments must find ways of synthesising particles that can be administered intravenously, i.e., with a surface treatment that increases their plasma half-life and allows them to target tumour cells. Optimal field and frequency conditions must also be found to maximise the SLP and hence minimise the required doses. Finally, the in vivo temperature must be controlled

in some way to avoid the risks of overheating and causing necrosis in healthy tissues in the vicinity. A novel solution would be to use ferro- or ferrimagnetic nanoparticles whose Curie temperature can be adjusted to a value just above the temperature not to be exceeded *in vivo*. In this way, if this temperature were ever attained, each nanoparticle mediator would blow its own fuse, so to speak, by losing its magnetic properties and hence its heating potential [42]. Studies are currently under way.

Photothermal Treatment

This idea is more recent and seeks to use the optical properties of silica nanoparticles with diameters of a few tens of nanometers, coated with a thin layer of gold (core-shell morphology), or gold nanoparticles with a raspberry morphology [58]. Depending on the geometry of the metal, they absorb almost all wavelengths, but those absorbing energy in the near infrared are the most interesting, because these wavelengths can penetrate the human body to depths of at least 10 cm. They then convert this infrared radiation energy and heat their environment locally. Cancer treatment also constitutes a possible application by virtue of passive targeting (retention by increased permeability of the vascular epithelium, see p. 156). *In vivo* studies on mice show that the heating generated by infrared absorption in the nanoparticles under the effect of a diode laser (808 nm, 4 W/cm², 3 min) is sufficient to destroy cancer cells, while at the same time preserving healthy neighbouring tissues [59]. In this study, remission was total, extending the lifespan of treated animals by at least three months.

Other work has tried to associate gold colloids with a thermosensitive hydrogel polymer such as poly(N-isopropylacrylamide-co-acrylamide), which has the particular feature of suddenly contracting above a certain temperature called the lowest critical solution temperature or LCST [60]. This mechanical withdrawal phenomenon releases molecules originally imprisoned between the macromolecules. Such a device could be used to operate the controlled release of drugs by optical illumination of the relevant region. However, much remains to be done, especially with regard to making these thermosensitive macromolecules biocompatible.

4.4 Conclusion

While (bio)organic chemists have been involved in the search for active ingredients for decades and polymer scientists have been designing capsules to deliver them for twenty years, it is now time for inorganic chemists to structure the resulting particles on the nanometric scale, to combine their unique magnetic or optical properties with a view to biomedical applications. A complete set of such tools should greatly accelerate our understanding of biological mechanisms, allow earlier and better diagnosis, and instigate new therapeutic

strategies. This aspect of the development of nanotechnology in the interests of medicine is without doubt one of the finest examples of concerted effort by chemists, physicists, biologists, pharmacists, and physicians to achieve a common purpose. At a time when there is growing concern, and quite justifiably so, over the possible consequences of nanotechnology for human beings and our environment, it should not be forgotten that a non-negligible part of research on nanoparticles aims only to favour the survival and comfort of humankind.

Acknowledgements

The authors would like to thank Stéphane Mornet, Sébastien Vasseur, Frédéric Rocco, Jean-Marie Devoisselle, Catherine Dubernet, Jean-Michel Franconi, Vincent Dousset for helpful discussion, and more generally, all members of the nanohybrid group GDR CNRS no. 2486.

References

1. Dupas, C., Houdy, P., Lahmani, M. (Eds.): *Nanoscience: Nanotechnologies and Nanophysics*, Springer, Berlin Heidelberg New York (2007)
2. Lahmani, M., Bréchnignac, C., Houdy, P. (Eds.): *Nanoscience: Nanomaterials and Nanochemistry*, Springer, Berlin Heidelberg New York (2008)
3. Jolivet, J.P.: In: *De la solution à l'oxyde: condensation des cations en solution aqueuse, chimie de surface des oxydes*, InterEditions/CNRS Editions, Paris (1994)
4. Massart, R.: IEEE Trans. Magn. MAG-17, 1247 (1981)
5. Gupta, A.K., Gupta, M.: Biomaterials **26**, 3995 (2005)
6. Plueddemann, E.P.: In: *Silane Coupling Agents*, 2nd edn., Plenum Press, New York, 1991; Plueddemann, E.P.: In: *Silane and Other Coupling Agents*, ed. by K.L. Mittal, VSP, 3 (1992)
7. Margel, S., Gura, S., Bamnolker, H., Nitzan, B., Tennenbaum, T., Bar-Toov, B., Hinz, M., Seliger, H.: In: *Scientific and Clinical Applications of Magnetic Carriers*, Plenum Press, New York (1997)
8. Del Campo, A., Sen, T., Lellouche, J.P., Bruce, I.J.: J. Magn. Magn. Mater. **293**, 33 (2005)
9. Mutin, P.B., Guerrero, G., Vioux, A.: C.R. Chimie **6**, 1153 (2003)
10. Yee, C., Kataby, G., Ulman, A., Prozorov, T., White, H., King, A., Rafailovich, M., Sokolov, J., Gedanken, A.: Langmuir **15**, 7111 (1999)
11. Portet, D., Denizot, B., Rump, E., Lejeune, J.J., Jallet, P.: J. Colloid Interface Sci. **238**, 37 (2001)
12. Fauconnier, N., Pons, J.N., Roger, J., Bee, A.: J. Colloid Interface Sci. **194**, 427 (1997)
13. Molday, R.S., Mackenzie, D.: J. Immunol. Methods **52**, 353 (1982); Molday, R.S.: US Patent no. 4 452 773 (1984)
14. Harris, L.A., Goff, J.D., Carmichael, A.Y., Riffle, J.S., Harburn, J.J., St. Pierre, T.G., Saunders, M.: Chem. Mater. **15**, 1367 (2003)

15. Liu, Q., Xu, Z., Finch, J.A., Egerton, R.: *Chem. Mater.* **10**, 3936 (1998)
16. Mornet, S.: PhD Thesis, University of Bordeaux 1 (2002); Mornet, S., Grasset, F., Portier, J., Duguet, E.: *Eur. Cells Mater.* **3**, 110 (2002)
17. Murray, C.B., Norris, D.J., Bawendi, M.G.: *J. Am. Chem. Soc.* **115**, 8706 (1993)
18. Bailey, R.E., Smith, A.M., Nie, S.: *Physica E* **25**, 1 (2004)
19. Hines, M.A., Guyot-Sionnest, P.: *J. Phys. Chem.* **100**, 468 (1996)
20. Peng, X., Schlempp, M.C., Kadavanich, A.V., Alivisatos, A.P.: *J. Am. Chem. Soc.* **109**, 7019 (1997)
21. Chan, W.C.W., Nie, S.: *Science* **281**, 2016 (1998); Bruchez, M., Moronne, M., Gin, P., Weiss, S., Alivisatos, A.P.: *Science* **281**, 2013 (1998); Aldana, J., Wang, Y.A., Peng, X.: *J. Am. Chem. Soc.* **123**, 8844 (2001)
22. Kim, S., Bawendi, M.G.: *J. Am. Chem. Soc.* **125**, 14652 (2003)
23. Tribet, C., Audebert, R., Popot, J.L.: *Proc. Natl. Acad. Sci. USA* **93**, 15047 (1996); Wu, X.Y., Liu, H.J., Liu, J.Q., Haley, K.N., Treadway, J.A., Larson, J.P., Ge, N.F., Pearle, F., Bruchez, M.P.: *Nature Biotechnol.* **21**, 452 (2003); Petruska, M.A., Bartko, A.P., Klimov, V.I.: *J. Am. Chem. Soc.* **126**, 714 (2004); Pelligrino, T., Manna, L., Kudera, S., Liedl, T., Koktysh, D., Rogach, A.L., Keller, S., Radler, J., Natile, G., Parak, W.J.: *Nano. Lett.* **4**, 703 (2004); Parak, W.J., Pellegrino, T., Plank, C.: *Nanotechnology* **16**, R9 (2005)
24. Liz-Marzan, L.M.: *Materials Today* **7**, 26 (2004)
25. Gittins, D., Caruso, F.: *Chem. Phys. Chem.* **1**, 110 (2002)
26. Brust, M., Walker, M., Bethell, D., Schiffrin, D.J., Whiman, R.: *J. Chem. Soc. Chem. Commun.* 801 (1994)
27. Kobayashi, Y., Horie, M., Konno, M., Rodrigiez-Gonzalez, B., Liz-Marzan, L.M.: *J. Phys. Chem.* **107**, 7420 (2003)
28. Gittins, D., Caruso, F.: *Chem. Phys. Chem.* **1**, 110 (2002)
29. Brus, L.E.: *J. Phys. Chem.* **80**, 4404 (1984)
30. Peterman, E.J.G., Brasselet, S., Moerner, W.E.: *J. Phys. Chem. A* **103**, 10553 (1999)
31. Sutherland, A.J.: *Curr. Opin. Solid State Mater. Sci.* **7**, 365 (2002); Gao, X., Nie, S.: *Trends Biotechnol.* **21**, 371 (2003)
32. Clapp, A.R., Medintz, I.L., Mauro, J.M., Fischer, B.R., Bawendi, M.G.: *J. Am. Chem. Soc.* **124**, 2293 (2002)
33. Dubertret, J., Skouride, P., Norris, D.J., Noireaux: *Science* **298**, 1759 (2002)
34. Dahan, M., Levi, S., Luccardini, C., Rostaing, P., Riveau, B., Triller, A.: *Science* **302**, 442 (2003)
35. Ozkan, M.: *Drug Discovery Today* **9**, 24, 1065 (2004)
36. Brokmann, X., Hermier, J.P., Desbiolles, P., Dahan, M.: *Images de la Physique* 126 (2005) and www.spm.cnrs-dir.fr/actions/publications/idp/IdP2005/20Brockmann.pdf
37. Faulk, W.P., Taylor, G.M.: *Immunochemistry* **8**, 1081 (1971)
38. Hainfeld, J.F., Powell, R.D.: *J. Histochem. Cytochem.* **48**, 471 (2000)
39. Schultz, S., Smith, D.R., Mock, J.J., Schultz, D.: *Proc. Natl. Acad. Sci.* **97**, 996 (2000)
40. Boyer, D., Tamarat, P., Maali, A., Lounis, B., Orrit, M.: *Science* **297**, 1160 (2002)
41. Mirkin, C.A.: *Nature* **382**, 607 (1996)
42. Mornet, S., Grasset, F., Vasseur, S., Duguet, E.: *J. Mater. Chem.* **14**, 2161 (2004)

43. Moghimi, S.M., Hunter, A.C., Murray, J.C.: *Pharmacol. Rev.* **53**, 283 (2001)
44. Monfardini, C., Veroneses, F.M.: *Bioconjugate Chem.* **9**, 418 (1998)
45. Stella, B., Arpicco, S., Peracchia, M.T., Desmaëlle, D., Hoebeke, J., Renoir, M., D'Angelo, J., Cattel, L., Couvreur, P.: *J. Pharm. Sci.* **89**, 1452 (2000)
46. Andrä, W., Nowak, H. (Eds.): *Magnetism in Medicine*, Wiley-VCH (1998)
47. Pankhurst, Q.A., Connolly, J., Jones, S.K., Dobson, J.: *J. Phys. D: Appl. Phys.* **36**, R167 (2003)
48. Brasch, R.C.: *Magn. Reson. Med.* **22**, 282 (1991)
49. Unger, E., Tilcock, C., Ahkong, Q.F., Fritz, T.A.: *Invest. Radiol.* **25**, S65 (1990)
50. Voisin, P., Ribot, E.J., Miraux, S., Bouzier-Sore, A.-K., Lahitte, J.-F., Bouchaud, V., Mornet, S., Thiaudiere, E., Franconi, J.-M., Raison, L., Labrugere, C., Delville, M.-H.: *Bioconjugate Chemistry* **18** (4), 1053–1063 (2007)
51. Reynolds, C.H., Annan, N., Beshah, K., Huber, J.H., Shaber, S.H., Lenkinski, R.E., Wortman, J.A.: *J. Am. Chem. Soc.* **122**, 8940 (2000)
52. Bonnemain, B.: *J. Drug Targeting* **6**, 167 (1998)
53. Bulte, J.W.M., Douglas, T., Mann, S., Vymazal, J., Laughlin, P.G., Frank, J.A.: *Acad. Radiol.* **2**, 871 (1995)
54. Wallis, F., Gilbert, F.J.: *J.R. Coll. Surg. Edinb.* **44**, 117 (1999)
55. Harisinghani, M.G., Barentsz, J., Hahn, P.F., Deserno, W.M., Tabatabaei, S., Hulsbergen van de Kaa, C., de la Rosette, J., Weissleder, R., *New Engl. J. Med.* **348**, 2491 (2003)
56. Shen, T., Weissleder, R., Papisov, M., Bogdanov Jr., A., Brady, T.J.: *Magn. Reson. Med.* **29**, 599 (1993)
57. Mornet, S., Portier, J., Duguet, E.: *J. Magn. Magn. Mater.* **293**, 127 (2005)
58. Hirsch, L.R., Stafford, R.J., Bankson, J.A., Sershen, S., Price, R.E., Hazle, J.D., Halas, N.J., West, J.L.: *Proc. Natl. Acad. Sci. USA* **100**, 13549 (2003)
59. O'Neal, P., Hirsch, L.R., Halas, N.J., Payne, J.D., West, J.L.: *Cancer Lett.* **209**, 171 (2004)
60. Sershen, S.R., Westcott, S.L., Halas, N.J., West, J.L.: *J. Biomed. Mater. Res.* **51**, 293 (2000)

Living Nanomachines

M.-F. Carlier, E. Helfer, R. Wade, and F. Haraux

5.1 Introduction

The living cell is a kind of factory on the microscopic scale, in which an assembly of modular machines carries out, in a spatially and temporally coordinated way, a whole range of activities internal to the cell, including the synthesis of substances essential to its survival, intracellular traffic, waste disposal, and cell division, but also activities related to intercellular communication and exchanges with the outside world, i.e., the ability of the cell to change shape, to move within a tissue, or to organise its own defence against attack by pathogens, injury, and so on.

These nanomachines are made up of macromolecular assemblies with varying degrees of complexity, forged by evolution, within which work is done as a result of changes in interactions between proteins, or between proteins and nucleic acids, or between proteins and membrane components. All these cell components measure a few nanometers across, so the mechanical activity of these nanomachines all happens on the nanometric scale. The directional nature of the work carried out by biological nanomachines is associated with a dissipation of energy. As examples of protein assemblies, one could mention the proteasome, which is responsible for the degradation of proteins, and linear molecular motors such as actomyosin, responsible for muscle contraction, the dynein–microtubule system, responsible for flagellar motility, and the kinesin–microtubule system, responsible for transport of vesicles, which transform chemical energy into motion. Nucleic acid–protein assemblies include the ribosome, responsible for synthesising proteins, polymerases, helicases, elongation factors, and the machinery of DNA replication and repair; the mitotic spindle is an integrated system involving several of these activities which drive chromosome segregation. The machinery coupling membranes and proteins includes systems involved in the energy metabolism, such as the ATP synthase rotary motor, signalling cascades, endocytosis and phagocytosis complexes, and also dynamic membrane–cytoskeleton complexes which generate protrusion forces involved in cell adhesion and migration. The ideas of molecular

recognition and controlled interfaces between biological components provide the underlying mechanisms for biological machinery and networks [1]. Many proteins illustrate this principle by their modular organisation into domains. The juxtaposition of catalytic domains of known function and domains of interaction with different partners leads to the emergence of new biological functions. It can also create threshold mechanisms, or biological switches, by triggering the activity of a given domain only when several partners interact with the regulatory domains. Many of these interaction domains are well understood. They exist inside different proteins, in particular, in cell signalling networks, and could potentially be used as building blocks in the construction of new proteins.

Living nanomachines are characterised by their high efficiency and reliability. Even in 1959, Richard Feynman [2] marvelled at the ability of biological systems to anabolise and catabolise the biomass, store and transfer information, self-assemble and reproduce, and move around. He realised that biological systems are remarkable examples of miniaturised tools, and that the concepts underpinning the functioning of such systems could be exploited to create controllable, biomimetic chemical nanomachines with a huge range of applications, from molecular electronics to nanomedicine.

Conversely, the tools of nanophysics, used to observe and manipulate single molecules like molecular motors have made an enormous contribution to our understanding of biological mechanisms for producing force and motion, showing us how to relate mechanical properties to chemical properties in living matter.

In addition, modular biological systems lend themselves well to the idea of reconstituting a functional system and its regulation, starting out from a minimal number of cell components [3]. This type of biomimetic approach aims to understand the relevant mechanisms by testing proposed operating concepts for the system. Furthermore, there is an obvious benefit in the direct use of molecules from living matter, whose functions have been optimised over thousands of years of evolution, as spare parts in the construction of hybrid nanostructures, associating biological substances with biocompatible materials for applications in bioengineering and medicine.

An exhaustive description of all the biological machines mentioned above would take us beyond the scope of this book. We have thus selected a number of examples of mechanisms for producing force and motion in a living cell:

- Motility processes resulting from spatially directed self-assembly of actin filaments against a membrane (M.-F. Carlier and E. Helfer).
- ATPase linear molecular motors (myosins, kinesins, dyneins) moving along actin filaments and microtubules, major polymers of the cytoskeleton (R. Wade).
- A rotary motor: ATP synthase (F. Haraux).

We begin with a brief introduction to the polymers of the cytoskeleton as a prerequisite for the discussion in Sects. 5.2 and 5.3.

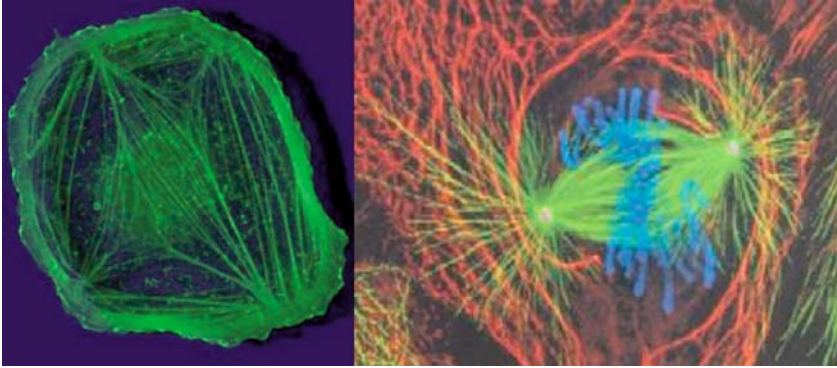


Fig. 5.1. Organisation of intracellular space by the cytoskeleton. *Left:* Network of actin filaments in a cell. Note the presence of stress fibres in rigid cables, a peripheral cortical network of filaments, and a filament-rich region, forming the motile undulating boundary of the cell. *Right:* Cell undergoing mitosis. The mitotic spindle (*green*) is made of microtubules (anti-tubulin immunofluorescence), chromosomes are stained *blue* with DAPI, and the intermediate filament network is stained *red* (anti-vimentin)

In eukaryotes, the intracellular space is organised by a network of fibrous polymers called the cytoskeleton (see Fig. 5.1). These fibres are a few nanometers in diameter and several micrometers long. There are three different substructures: actin filaments (diameter 7–8 nm), the so-called intermediate filaments (diameter 10 nm), and microtubules (diameter 25 nm). The actin filaments and microtubules play a major role in the life of the cell. They are the mainstay of the cellular architecture, but they also control all cell motility in response to signals from the outside world. The key properties explaining these functions are their dissipative, dynamic assembly and structural polarity.

Actin filaments and microtubules form by self-assembly of a single globular protein, i.e., actin and tubulin, respectively. ATP (adenosine triphosphate) binds to the actin monomer, and GTP (guanosine triphosphate) to the tubulin ($\alpha\beta$ heterodimer); then hydrolysis of the ATP into ADP and the GTP into GDP leads to the assembly of actin filaments and microtubules, respectively. Actin and tubulin are common in all cells, and highly conserved in the eukaryotic kingdom, with counterparts or remote ancestors in the prokaryotic kingdom. The assembly of these polymers is tightly controlled *in vivo* by a wide range of associated proteins. These polymers serve as tracks to guide the movement of motor ATPases such as myosins (which move along the actin filaments) and dyneins or kinesins (which move along microtubules), thus accomplishing muscle contraction or intracellular transport. Further, variations in the assembly dynamics of these polymers and their significant spatial reorganisation underlie changes in cell shape and motility or adhesion, e.g., at the beginning of mitosis, the two networks depolymerise and reassemble into structures constituting the mitotic spindle and the contractile ring of

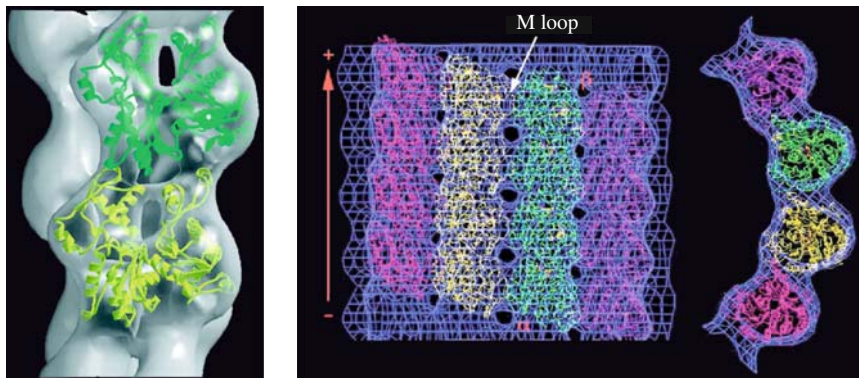


Fig. 5.2. Actin and tubulin structures. *Left:* Crystallographic structures of actin monomers (green and yellow) fitted to the electron microscope reconstruction of an actin filament (grey) [4]. Copyright 2002, National Academy of Sciences, USA. *Right:* Atomic model of a microtubule as deduced from electron microscope diffraction images and the structure of tubulin at resolution 4 Å, derived from the electron diffraction of tubulin-Zn polymers [5]. With the kind permission of Springer Science and Business Media

cytokinesis. For this reason, the dynamics of the two networks are intimately related to ensure a harmonious motile response from the cell.

The structures of actin and tubulin are known at the atomic scale (see Fig. 5.2). Atomic models of the actin filament and the microtubule can be constructed. The self-assembly properties of actin and tubulin have been widely investigated *in vitro*, together with many features of their regulation by associated proteins. However, much progress is still being made in this area.

5.2 Force and Motion by Directed Assembly of Actin Filaments

5.2.1 General Considerations

Living cells have an internal architecture or cytoskeleton comprising three protein polymer networks: microtubules, actin filaments, and intermediate filaments. This architecture is not frozen, but is extremely dynamic. It orchestrates several vital functions in the cell, playing the role of sensor with regard to the outside world, and its remodelling of the cell in response to extracellular signals is what allows the cell to change shape, to polarise itself, and to migrate in a directional manner, while controlling its interactions with neighbouring cells and the extracellular matrix [6]. The microtubules and actin networks react in a coordinated way to establish and maintain cell polarity during motion.

The basic changes in the shape of the cell are the formation of membrane protrusions of lamellar type (lamellipodia) or digitiform (filopodia).

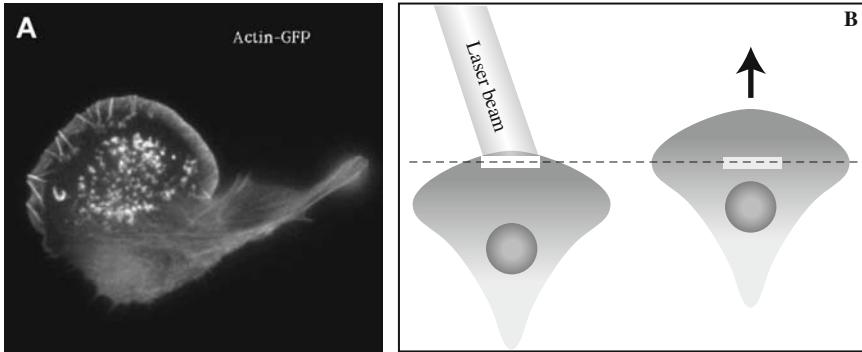


Fig. 5.3. Polymerisation of actin in the lamellipodium of a moving cell. **(A)** Melanocyte transfected by actin-GFP [9]. The lamellipodium contains a dense polarised network of actin filaments. **(B)** Dynamics of actin during motion. Schematic view of a photobleaching experiment, showing that the actin network remains stationary relative to the substrate as the cell moves forward, i.e., the actin filaments continuously polymerise at the leading edge of the cell

These motile processes are generated by spatially directed polymerisation of the actin filaments against the plasma membrane. They are involved in many complex physiological processes, such as the migration of embryonic cells and formation of the neural tube, the chemotactic locomotion of neutrophils or amoebas, the activation of T lymphocytes in the immune response, cytokinesis, angiogenesis, repair of damaged tissue, extension of axons during development of the nervous system, and the synaptic plasticity that presides over learning and memory processes.

In a living organism, a cell generally moves in the three dimensions of space within a tissue. Little is yet known about motion in three dimensions. Laboratory studies have been limited to analysing the 2D migratory motion of culture cells on a substrate (extracellular matrix).

Reptating cells have a polarised morphology: the body of the cell including the nucleus hangs behind, while the cell extends a rather thin, fan-shaped lamellar protrusion (thickness 200 nm, width a few micrometers) in the direction of motion (see Fig. 5.3A). Cell migration occurs at any speed between 0.5 and 20 $\mu\text{m}/\text{min}$ and involves a repeated cycle of four consecutive steps: protrusion (extension of the lamellipodium), adhesion to the substrate, contraction at the back of the cell which projects the cell body forward, and breaking of adhesive contacts at the rear [7]. These steps can take place in a highly concerted way. One then observes a continuous reptating motion, during which the cell maintains a steady-state morphology. An acceptable motion, conserving the integrity of the cell, requires the reactions of protrusion and adhesion/deadhesion to be concerted. The mechanisms guaranteeing this level of coordination are not yet understood. However, the protrusion movement is produced by the directional polymerisation of actin filaments against the

membrane. The molecular mechanism here is reasonably well understood and can be reconstituted from purified cell components.

A simple experiment provides a visual demonstration of the autonomy of the lamellipodium as a motile structure. When detached mechanically from the rest of the cell by microsurgery using a fine needle, the lamellipodium continues its reptating motion for as long as it is allowed by the supply of chemical energy (ATP) it contains [8]. Hence, the lamellipodium represents a site of modular motile activity within the cell.

Actin filaments can be observed 'live' in the moving cell by the technique of fluorescence microscopy. Cellular actin is tagged by green fluorescent protein (GFP). The lamellipodium then appears as an actin-rich region of the cell, sometimes referred to as an 'actin factory', where actin filaments form a polarised fan-shaped network, with one end of the filaments directed perpendicularly to the leading edge of the migrating cell (see Fig. 5.3A). The migratory motion can be arrested by drugs that inhibit polymerisation of the actin. The density of actin in the network exhibits a gradient, higher at the leading edge, and lower toward the rear of the lamellipodium. The actin density gradient and also the width of the network are held constant during motion. By analysing the recovery of fluorescence after photobleaching the actin network in a region near the leading edge, it can be shown that the dark spot remains stationary relative to the substrate while the cell moves forward, with new fluorescent filaments forming all the time at the leading edge (see Fig. 5.3B). All these results point to the idea that actin filaments continuously polymerise, in an insertional way, at one end, at the leading edge of the membrane, and depolymerise at the other end, at the rear of the lamellipodium, in a motion known as treadmilling [9]. More detailed analyses of this kinetics carried out recently using the technique known as speckle fluorescence microscopy have confirmed this interpretation and determined the turnover rate of individual filaments in the extending lamellipodium [10].

These observations and their analysis suggest that the polymerisation of actin filaments against a membrane produces a protrusion force and underlies the motion of the leading edge of cells. Owing to its molecular and macroscopic features, this phenomenon of cell biology has attracted much interest from biophysicists and biochemists, and raised questions regarding the mechanism of force production by directional polymerisation, the relation between molecular reactions and mechanical properties, and the possibility of exploiting a biomimetic approach to motion.

A model system for studying lamellipodia has turned up somewhat providentially in the form of the intracellular pathogens *Listeria monocytogenes*, *Shigella flexneri*, and the vaccinia virus. These pathogens express at their surface a protein that mimics the cell machinery of the host cell. They induce at their surface the polymerisation of actin in filaments organised in a 'comet', using the propulsion force produced in this way to move around inside the infected cell, at speeds comparable with the extending lamellipodium. These pathogens were used in the 1990s as biochemically and genetically manipulable

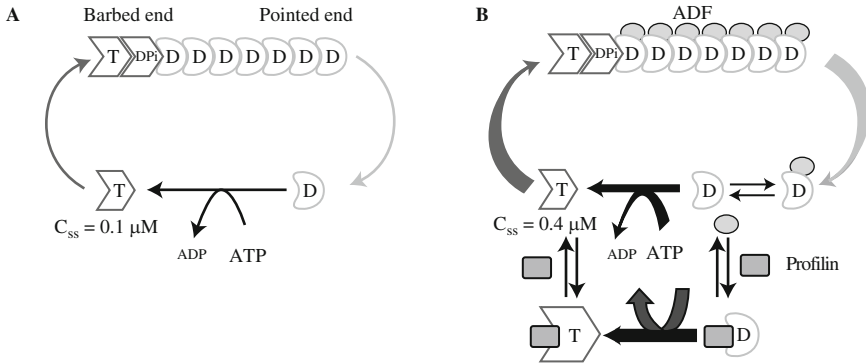


Fig. 5.4. Treadmilling of an actin filament. (A) Intrinsic treadmilling of pure actin. Dissociation from the pointed end is the limiting stage of the ATPase cycle. (B) Actin depolymerising factor (ADF) accelerates treadmilling by enhancing the dissociation rate of the actin at the pointed end. Profilin favours the association reaction at the barbed end

models of the migration front of motile cells. By observing their movement inside infected cells, it was for the first time possible to appreciate the significant force produced by actin polymerisation. These pathogens are not blocked either by intracellular obstacles (organelles), or by the plasma membrane which they easily invaginate, hence passing from one cell to another of the epithelial tissue of the intestine, and thereby favouring their dissemination.

The dynamics of actin in lamellipodia must be understood in the context of the assembly properties of actin filaments and their regulation by associated proteins.

5.2.2 Assembly Dynamics of Actin in Vitro. Intrinsic Properties

Actin is one of the best conserved proteins throughout evolution. It is abundant in all eukaryotic cells (20–500 μM , or 5–10% of the protein mass in most cells, 20% of the protein mass in muscle). The polymerisation of G-actin (monomeric, globular actin) in helical filaments called F-actin (filamentous actin) follows a nucleation–growth process. As required by the helical structure of the filament, the nucleus comprises three actin subunits, sufficient to initiate the genetic helix of the filament. Indeed, the subunits within the filament are connected by a rotation through -167° (screw angle) accompanied by an axial rise of 2.7 nm [11]. Under physiological ionic conditions, pure actin polymerises above a critical concentration C_c , which can vary between 0.1 and 0.5 μM .

A key property of actin, directly responsible for its role in cell motility, is that it is an ATPase. The monomeric actin binds an ATP molecule (G-ATP), the ATP bound to the actin is hydrolysed into ADP during polymerisation,

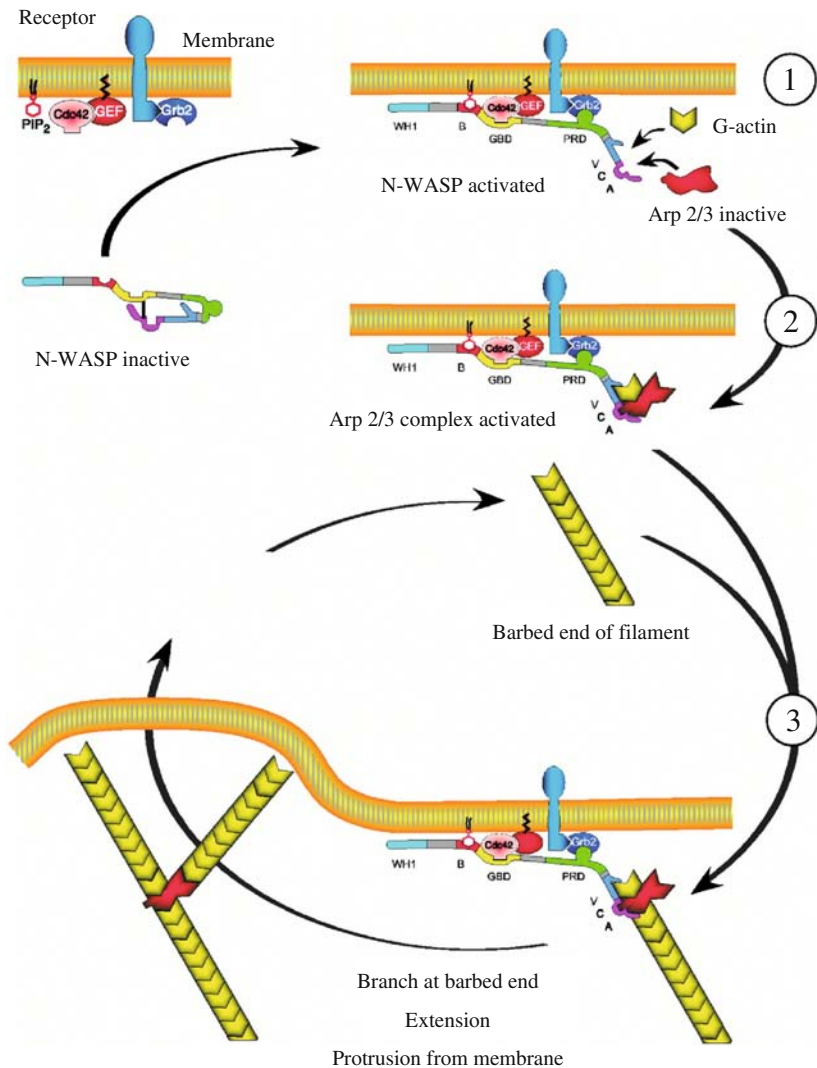


Fig. 5.5. Molecular mechanism for generation and rapid turnover of branched filaments in lamellipodia. N-WASP is recruited and activated at the membrane by signalling molecules (1). G-actin and the Arp2/3 complex bind to the C-terminal domain of N-WASP to form a branching complex (2). This branching complex associates with a pre-existing filament to generate a branch (3): the two branches grow at the same rate. N-WASP catalyses several consecutive branching cycles and increases the number of filaments growing against the membrane

the inorganic phosphate is released in solution, and the actin filaments are thus essentially made up of F actin-ADP. Since this reaction is irreversible, polymerisation of actin is a dissipative process. It cannot be described by

the laws of reversible polymerisation, according to which, at equilibrium, the polymer coexists with the monomer at the critical concentration, defined as the monomer–polymer equilibrium dissociation constant. The F-actin filaments are in a steady state defined by a thermodynamic disequilibrium of the monomer–polymer exchanges at the two ends. A slight excess of actin–ADP dissociation reactions at the pointed end is balanced by an excess of actin–ATP association reactions at the barbed end, thereby creating a net flow of subunits along the filament, in a process called treadmilling (see Fig. 5.4A). This actin ATPase cycle is responsible for the turnover of the filaments, which thus occurs in a polarised manner. The kinetically limiting stage in this cycle is the dissociation of actin at the pointed end (0.5 s^{-1}). This reaction imposes a very slow turnover of pure actin filaments *in vitro* (about $4\text{ }\mu\text{m/hr}$). In living cells, actin is polymerised in a steady state and the filament polymerisation process must therefore be understood in the treadmilling framework. In other words, the barbed ends of the filaments are continuously elongating, unless they are blocked by capping proteins, while the pointed ends are continuously depolymerising. The rate of this process is tightly regulated by specific proteins and spatially controlled by signalling pathways as described below.

5.2.3 Regulation of Actin Filament Assembly in Cell Motility

In a cell, the morphology of the actin filament network accounts for the motile state of the cell. There are two main networks. Firstly, there are cables of filaments arranged in antiparallel bundles, called stress fibres, which stretch from one point of the cell to another to form a rigid architecture, kept under tension by myosin. The anchor points of the stress fibres are focal adhesion complexes, i.e., protein assemblies which set up a bond between the extracellular matrix, the plasma membrane, and the filaments. The turnover of the filaments within these structures is very low, indicating that the barbed ends are capped at the anchor points. Secondly, a peripheral cortical region is formed by a more diffuse network of filaments (see Fig. 5.1). A non-motile cell adheres firmly to the substrate and contains a large number of stress fibres. Under the effect of extracellular stimuli such as hormones, growth factors, chemotactic agents, etc., a cell can become motile by reorganising the actin cytoskeleton. The stress fibres collapse, the cell polarises and the actin polymerises into a new network at the leading edge, thus generating the lamellipodium in which the filament turnover by treadmilling is very fast ($0.5\text{--}20\text{ }\mu\text{m/min}$, or two orders of magnitude faster than the treadmilling of filaments observed *in vitro*). Observations in cell biology can be transcribed in terms of biochemical reactions:

- **Actin Filament Turnover.** This is controlled with respect to rate and efficiency by two regulatory proteins (see Fig. 5.4B): cofilin or actin depolymerizing factor (ADF), which enhances depolymerisation of the filaments at the pointed end, the limiting stage of the ATPase cycle; and profilin, which favours association reactions at the barbed end and thereby

enhances the processivity of the ATPase cycle. This effect can be reconstituted *in vitro*, and one observes an enhancement of treadmilling by a factor of 150 when ADF and profilin are added, corresponding to the filament turnover measured in lamellipodia.

- **Continuous Generation of Filaments.** Filaments are generated all the time at the membrane at the leading edge of the cell, by protein machinery associated with the membrane, and directly controlled by signalling. This machinery comprises an enzyme (proteins in the WASP family) which catalyses the branching reaction of the filaments at the barbed end, thereby doubling the number of growing filaments in each catalytic cycle (see Fig. 5.5). Each barbed end can exist in two states: a state attached to the membrane via N-WASP during the branching reaction and a detached state during the growth reaction. Free barbed ends directed against the membrane undergo rapid growth supplied by treadmilling, while the pointed ends depolymerise at the rear of the lamellipodium. A dendritic polarised network of fractal type is thus constructed. Insertional polymerisation of filaments produces the protrusion force of the lamellipodium. In addition, the lamellipodium adheres to the extracellular matrix at focal points, at which bundles of non-branching filaments are initiated. Assembly of such filaments is catalysed by different machinery, a processive motor called formin.
- **Filament Growth.** The growth of filaments is arrested by capping proteins at the barbed ends. The concentration of capping proteins determines the lifetime and hence the average length of the filaments in the branching network. In the steady-state regime, the network morphology, the number of growing filaments, and the protrusion rate are all constant, because the creation of filaments by branching is balanced by the disappearance of filaments due to the action of the capping proteins. It is the ratio of the surface density of N-WASP and the concentration of capping proteins that determines the branching density.

To sum up, for a biochemist, the mechanism for motion by directed polymerisation of actin against an obstacle is based on an autocatalytic local stimulation of polymerisation, counterbalanced by inhibition of the growth of filaments by a diffusive soluble factor. The constant speed and directionality of the motion are related to the (dissipative) steady-state character of the polymerisation: filament treadmilling serves as fuel for the motion [12]. The rate of migration of the lamellipodium must depend on the concentration of proteins controlling the treadmilling rate. This molecular view of the protrusion mechanism indicates the nature and function of the cell components required for the motion, but it is not adequate to provide a mechanical description of the way the force is produced. This novel self-organising system for producing a force has attracted the attention of soft-matter physicists, interested in the properties of the stimulatable gels represented by actin filaments in motile processes. It has also proved interesting to mathematicians who try to model complex

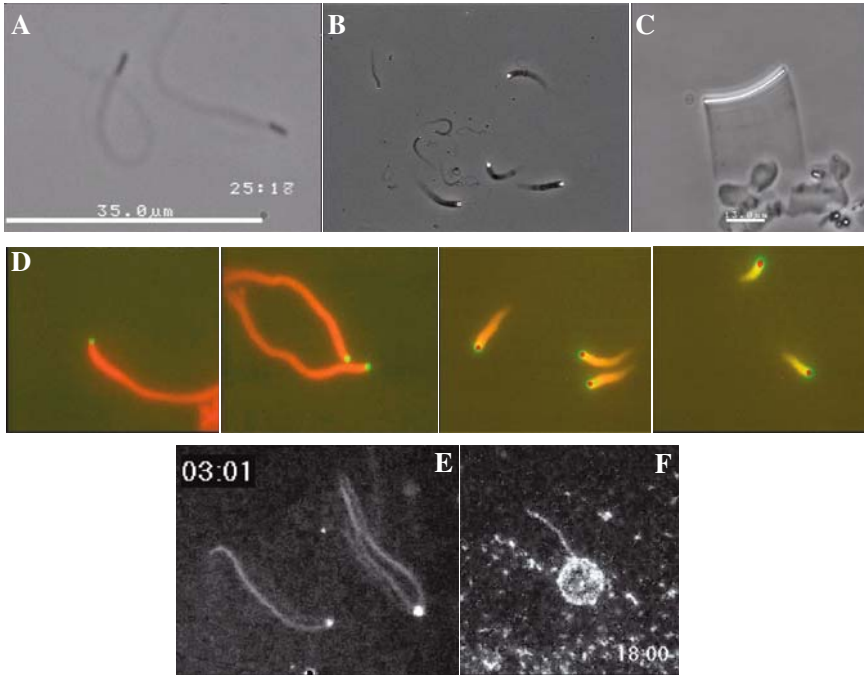


Fig. 5.6. Motility reconstitution by directional polymerisation of actin filaments: from the *Escherichia coli* model to biomimetic systems. (A)–(C). Motion in the reconstituted motility medium containing actin and four pure proteins, observed by phase-contrast microscopy: *Escherichia coli* (A) and beads of different diameters (B) form actin comets. A glass rod (C) generates a lamellar filament network. (D) Monitoring the dependence of the filament branching density on one of the regulatory proteins using two fluorophores. The motility medium contains actin–rhodamine and Arp2/3–Alexa 488. (E)–(F) Biomimetic motility assay with microspheres functionalised by formin, observed by fluorescence microscopy (actin labelled with rhodamine). The medium contains actin and profilin, needed for the formin processive motor. (E) The formation of a large number of unbranched filaments by processive assembly catalysed by immobilised formin leads to propulsion of the bead. (F) Catalysis of processive assembly of a single filament by a formin molecule immobilised on a bead. The filament only detaches after 22 min

biological systems via a computational approach. The progress made in these two fields has been made possible through the elaboration of a biomimetic system in which the motion of a functionalised particle is reconstituted in vitro using a minimal number of purified cell components.

5.2.4 Biomimetic Motility Assay

Motility Generated by Formation of a Branching Filament Network

The aim of this assay is to mimic the advancing motion of the leading edge of a cell in a chemically controlled way, in order to measure the mechanical parameters (speed, force) produced by actin polymerisation, and then to relate them to the controlled physicochemical parameters. A particle of micrometric dimensions (a glass rod or microsphere), of variable size and shape, is functionalised by simple adsorption of the N-WASP protein machinery. The average surface density of the protein covering the surface can be controlled and measured. This particle is placed in an actin solution containing ATP as energy source and regulatory proteins for the treadmilling process (ADF/cofilin, capping protein profilin), together with the Arp2/3 complex, substrate for N-WASP in the filament branching reaction. It has been shown that this reconstituted motility medium is necessary and sufficient for the constitutive propulsion of the particle, in terms of the elements it contains. The robustness of the medium is manifested by the wide range of concentrations of the regulatory proteins in which motion is observed [13]. Insertional polymerisation at the functionalised solid surface leads to formation of a branched filament network with morphology dictated by the geometry of the particle itself (see Fig. 5.6A–C): the network resembles a comet if the particle is spherical, but adopts a lamellar morphology, as in a lamellipodium, if the particle is a rod, the network organising itself spontaneously from a region centered around the cylinder axis. The motion and the formation of the actin network are observed by phase-contrast optical microscopy, or fluorescence microscopy if the proteins used in the assay (actin, Arp2/3) are labelled with a fluorophore. The motility medium functions as a chemostat which maintains a high steady-state concentration of G-actin–ATP, so that the rate of propulsion of the particle generated by the steady-state polymerisation of actin remains constant. This speed is of the order of a few $\mu\text{m}/\text{min}$, as observed *in vivo*.

The *in vitro* reconstitution of this motion corroborates the molecular explanation of the motion presented above. The data show that the propulsive motion of a functionalised solid particle has the same characteristics (speed, identity of reagents, morphology of actin network) as the protrusion motion of the plasma membrane, suggesting that the molecular mechanisms producing the force are in fact the same in the two systems. A recent study using antisense RNA technology has shown that, *in vivo*, the essential proteins for lamellipodium protrusion are the same as those identified on the basis of their biochemical properties.

The biomimetic motility assay has many advantages for analysing the force production mechanism through directional actin polymerisation and testing the predictions of different physical models for the motion. Indeed, the surface density of the enzyme branching the filaments and the concentration of

the various regulatory proteins are controllable parameters. The particles can be manipulated or subjected to external forces. It is thus possible to measure the force/speed relation. In addition, the effect of external forces on the structure of the network can be observed by microscopy using the double fluorescence of actin and Arp2/3. Since Arp2/3 is incorporated at each branching point of the filaments, the ratio of Arp2/3 to actin in the network provides a measure of the branching density (see Fig. 5.6D).

Motility Generated by Processive Assembly of Unbranched Filaments

The reconstituted motility assay has been extended to the processive assembly of filaments by formin molecules immobilised on a particle [14]. In this system, the polymerisation of single unbranched filaments is catalysed by formin which, in association with profilin and actin, plays the role of a genuine processive assembly motor. The formin remains associated at the growing barbed end for periods of up to 4,000 consecutive assembly cycles, which corresponds to generation of a filament several micrometers long (see Figs. 5.6E–F). As the formins are rather close to one another at the particle surface, the filaments synthesised by insertional polymerisation associate together in bundles to form a non-branching network which also produces propulsion of the particle.

In motile systems based on polymerisation into either branched filaments or unbranched filaments, it should be noted that the propulsion speed of the particle is closely related to the polymerisation rate, regardless of the macroscopic organisation and mechanical or rheological properties of the actin network.

5.2.5 Measuring the Force Produced by Directional Actin Polymerisation

The motion of *Listeria* bacteria or functionalised polystyrene microspheres has been studied both in cell extracts with non-controlled composition and in the chemically controlled reconstituted motility medium. Trajectories and propulsion speeds can be analysed. The force is assessed by applying an external force to slow down the motion.

Micromanipulation Using Optical Tweezers

The first force measurements were carried out by Gerbal et al. to estimate the interaction between *Listeria* and its actin comet [15]. Using optical tweezers to attempt to detach the comet from the bacterium, it was shown that a force of 10 pN is not enough to separate them. This experiment shows that the actin comet and the bacterium are firmly connected, thereby contradicting the first physical models put forward to explain the force production mechanism

(see Sect. 5.2.6). The second experiment consists in opposing the propulsion with an external force, by trying to trap the bacterium with the optical tweezers. It is observed that the bacterium can be stopped temporarily, but that the comet continues to extend, causing it to buckle: the thrust due to growth of the comet increases until the bacteria is expelled from the optical trap. After release, the comet relaxes into its initial configuration, proving that the deformation is elastic. This conclusion led to the proposal of a model for propulsion by an elastic gel, which will be discussed in Sect. 5.2.6.

Effect of Viscosity on the Propulsion of Functionalised Particles

Two other experimental studies provide an estimate of the propulsion force. In both cases, a variable external force is applied by tuning the viscosity of the medium via a polymer [16, 17]. Stokes' law is then used to establish the force-speed relation. The results of the two studies differ. McGrath et al. observed a two-stage decrease in the speed, with the bacteria significantly decelerated for forces up to 50 pN, but less so beyond that. On the other hand, Wiesner et al. measured a much lesser effect (10% reduction in the propulsion speed) over the force range investigated, up to around 100 pN. Several hypotheses can explain these contradictory measurements, in particular the fact that the two groups used totally different techniques to measure the viscosity induced by methylcellulose (local viscosity 10^4 Pa s for McGrath et al. and nominal viscosity a few times 10 Pa s for Wiesner et al.). Finally, it may also be that this difference arises only from the biochemical system used.

Mechanical Measurement of the Deformation of Functionalised Membranes

More recently, several groups have developed a new experimental approach, which provides a closer approximation for the membrane protrusion. It consists in studying deformable functionalised objects like vesicles or oil droplets in cell extracts. The growth of an actin network can deform them during propulsion. Liposomes, whose elastic properties have now been well established [18], deform significantly under the influence of an external force. The forces acting on them can be deduced by observing their change in shape. It was thus a natural choice to use them as model systems for reproducing the deformation of the cell membrane. By studying the shape of these objects, the local force distribution can be determined along their contour. Vesicles functionalised with ActA [19, 20] or oil droplets coated with VCA [21] move by generating an actin comet at the rear and are deformed into a pear shape. In the case of vesicles, the authors argue that the pressure is uniform inside and conclude that the vesicle is deformed under the action of compressive lateral forces and retractile forces at the rear, estimated to be of the order of a few nN (see Fig. 5.7). In the case of the oil droplets, the authors took into

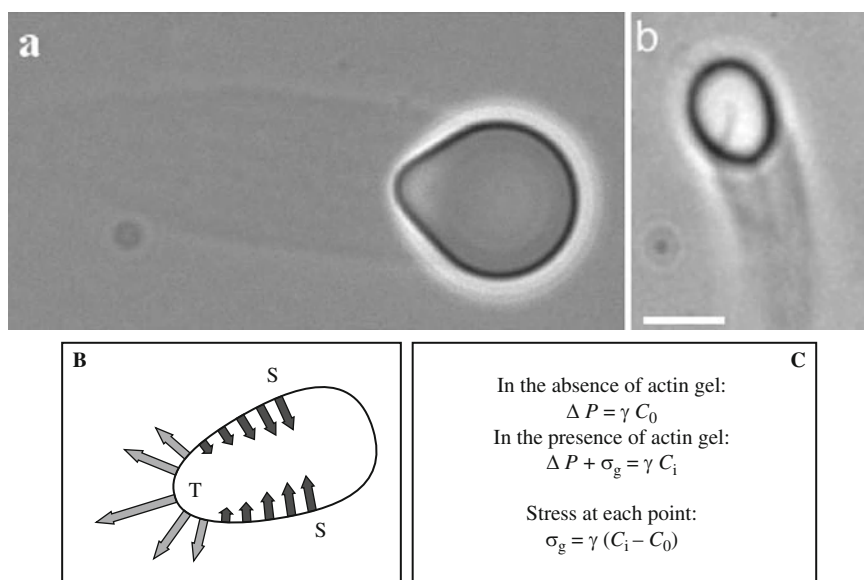


Fig. 5.7. Propulsion of deformable objects. Taken from [19,21]. **(A)** Photographs of an oil droplet **(a)** and a vesicle **(b)** deformed by actin comets. **(B)** Surface force distribution. The vesicle is subjected to lateral compression forces (*black*) and traction forces at the rear (*grey*). **(C)** Local force measurement. At the front, the object undergoes no local stresses since there is no actin gel. At the rear, the stress due to the presence of the gel is added to the force due to osmotic pressure on the surface. The local stress σ_g due to the actin gel is given directly as a function of the surface tension γ of the membrane by measuring the local deformation (or curvature C_i)

account variations in the surface tension and internal flow within the droplet to calculate a propulsion force of the order of 100 pN.

As in the case of rigid beads, these systems should subsequently be used in a purified protein medium, in order to establish the dependence of this phenomenon on controllable physical and biochemical parameters. Furthermore, the potential segregation of activators at the membrane surface could be monitored by using fluorophore-tagged proteins.

Micromanipulation by Micropipette

The forces estimated in the experiments described above vary over several orders of magnitude, from the fN to a few nN, and the results are sometimes contradictory. For this reason, the force produced by actin polymerisation must be measured in some direct way over a wide range of forces.

In 2004, Marcy et al. carried out a novel micromanipulation experiment in order to make a direct measurement of the force produced by polymerisation of the actin comet [22]. The actin comet generated by a bead fixed to a flexible glass fibre (see Fig. 5.8) and placed in the reconstituted motility medium is manipulated by means of a micropipette. With this setup, external

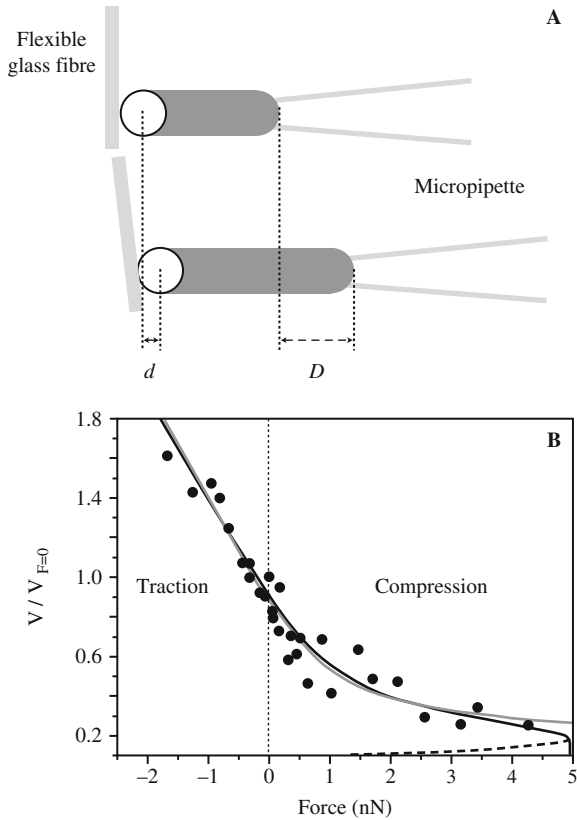


Fig. 5.8. Micromanipulation measurement of the force generated by actin polymerisation. Taken from [22]. **(A)** Experimental setup. The shift d of the bead results from the displacement D imposed by the micropipette and the extension of the comet. **(B)** Force–speed curve. Negative forces correspond to experiments in which the comet undergoes traction, while positive forces correspond to compression experiments

compression and traction forces up to a few nN can be applied. Two types of measurement are carried out: a fast traction measurement at constant speed to estimate the elastic modulus of the actin gel and the force required to detach the comet, and a speed measurement at constant force, either a (negative) traction force or a (positive) compressive force. The first measurements show that a force of a few nN is required to break the biochemical connections between the surface of the bead and the actin gel. The second series of measurements gives a force–speed curve which reveals two regimes: a linear regime for negative forces, i.e., traction applied to the comet, and a regime of slower decrease for positive forces. This result is very different from the one obtained by McGrath et al. [17]. However, the measurements made by

Wiesner et al. [16] fit well with this curve. Extrapolation of the curve shows that a force of 7 nN is required to arrest comet growth.

The experiment carried out by Marcy et al. remains the most accomplished for the moment. However, this experiment does not give a definitive answer to the question of the molecular mechanism producing the force. What is required is a systematic study of the dependence of the force–speed relation on the parameters controlling the number of filaments. Indeed, the measurements shown in Fig. 5.8 were obtained with beads of given diameter and functionalisation density. Because the system is self-organising, it responds to external forces by creating new filaments. The number of filaments thus varies right along the curve. It is not known whether this curve can be normalised by the number of filaments to obtain a kind of master curve representing the behaviour of n filaments. Making a rough estimate of the N-WASP density at the surface, based on the work by Wiesner et al. [16], the authors deduce a force per filament of about 10^{-2} pN, which is clearly too low compared with the values established by Upadhyaya and van Oudenaarden [23]. The amount of activator has been overestimated in these experiments and requires further checking. If the activator density at the surface could be varied in a controlled way, it would be possible to observe the behaviour of the force–speed curve as a function of the number of filaments initiated at the surface. Recent progress in nanolithography should make it possible to functionalise a surface by depositing N-WASP molecules one by one at regular intervals, controlled to the nearest nanometer [24]. Another important quantity is the branching density in the comet. The evolution of the degree of cross-linking can be monitored using fluorescence-tagged proteins (see Fig. 5.6 and Sect. 5.2.4). The actin comet is in fact a self-regulating system that modifies its kinetic parameters in response to an external force. Applying a compression or traction force should lead to changes in the structure of the comet.

AFM Force Measurements

Another possible experiment for measuring the force due to an actin gel would be to use atomic force microscopy (AFM). This technique can explore a wide range of forces, from about 100 pN to a few μ N. The AFM tip is functionalised with the actin polymerisation activator, causing a gel to grow out from the tip when it is placed in the reconstituted motility medium, at a very small distance from the glass slide. When it reaches the slide surface, the actin gel is expected to exert a force, thereby deflecting the AFM tip. The force–speed relation could be deduced from this kind of measurement.

5.2.6 Theoretical Models for Force Production by Actin Polymerisation

The experiments described above clearly demonstrate that it is indeed the polymerisation of the actin that produces the force deforming the cell membrane or propelling objects like the beads. But the mechanism producing this

force has not yet been clearly established. Indeed, although the thermodynamic basis for the motion seems to be well understood, we are still a long way from understanding the relation between the biochemical reactions (controlled directional polymerisation at a surface) and the resulting mechanical properties (elementary displacements and forces at the particle surface). Several models have thus been put forward to explain this process, some of which remain to be validated.

The theoretical models proposed today fall into two categories: microscopic models based on the molecular biochemical reactions and individual properties of filaments, and mesoscopic models based on the properties of the filament network treated as a continuous actin gel.

Microscopic Models of Motion Generated by Actin Polymerisation

In the first category are the Brownian ratchet models, which have been evolving steadily since the first of their kind due to Peskin in 1993 [25–28]. In the first version of the Brownian ratchet, Peskin assumed that fluctuations in the position of the bacterium (or lamellipodium membrane) allowed insertion of actin monomers at the end of the filaments, in the space produced when the bacterium moved away; extension of the filaments then prevented the bacterium from moving back again [25]. Actin filament growth thus resulted in a diffusive forward motion of the object, whose speed would depend on the dimensions of the object. The last point was refuted by the fact that *Listeria* and *Shigella* move at the same speed despite their different sizes [26]. Mogilner and Oster therefore adapted this model to an elastic Brownian ratchet. This assumes that the monomer insertion space is freed by thermal fluctuations of the actin filament, which can move away from the wall [27].

These first two Brownian ratchet models are based on the assumption that the actin filaments are not permanently attached to the membrane and that monomers insert themselves whenever a big enough space is made available by thermal fluctuations between the wall and the barbed end of the filaments. However, micromanipulation experiments by Gerbal et al. using optical tweezers clearly showed that the actin comet is firmly tethered to the bacterium: a force of 10 pN is not enough to detach the actin comet from *Listeria* [15]. The model was recently updated to give the tethered Brownian ratchet model [28]. In this version, the authors make the distinction between two filament populations which exert opposing forces on the bacterium: tethered filaments which hold the bacterium back and free filaments which can extend and push the bacterium as long as they are not capped (Fig. 5.9). The propulsion speed predicted by this model does not depend on the number of existing filaments, but on the ratio between the two types of filaments, tethered and free. It is independent of the density of functionalisation of the beads and the rate of nucleation of filaments. The authors also predict a two-stage behaviour in response to an external force: first a rapid drop in speed for forces of a few

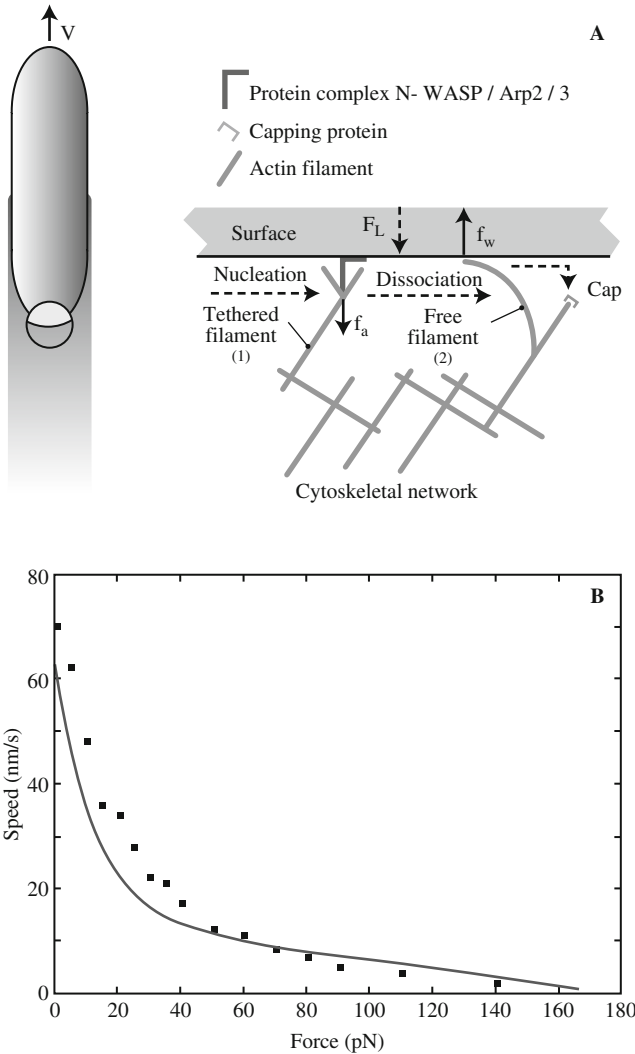


Fig. 5.9. Tethered Brownian ratchet model. Taken from [28]. **(A)** Molecular reactions on the scale of actin filaments: (1) Tethered filaments can nucleate new filaments or detach themselves from the surface. (2) Free filaments can polymerise or be capped by capping proteins. The thrust generated by extending filaments is counterbalanced by the retaining force due to tethered filaments and the external force. **(B)** Prediction for the propulsion speed as a function of the external force

tens of pN, followed by a slower decrease beyond that. This theory agrees with the experimental results due to MacGrath et al. [17], who observe a spectacular viscosity effect, i.e., due to the external viscous force, on bacterial motion. However, measurements made by Wiesner et al. [16], in which the

beads exhibit little or no slowing down when the viscosity of the medium is increased, cannot be explained by this model.

Carlsson has put forward a different model for the growth of actin filaments against an object, which compares two possible scenarios for filament generation at the surface [29, 30]: an autocatalytic branching process and a dendritic nucleation process (Fig. 5.10). In the first case, the rate of formation of new branches is proportional to the number of filaments or to the amount of polymerised actin near the surface. In the second case, the rate is independent of the number of filaments already in existence. However, the propulsion force exerted by the comet is proportional to the number of filaments in both cases. In 2001, Carlsson carried out a stochastic simulation of filament growth near a surface, but only for the case of autocatalytic generation [29]. The growth rate was controlled by 5 regulatory phenomena (polymerisation at the barbed end, depolymerisation at the pointed end, capping, branching, and detachment of branches). The author then produced a model agreeing with the initial simulations and including the nucleation process [30]. The main assumption of the model is that new filaments can only form in a region close to the surface, the filaments being oriented randomly relative to the surface. The results concern the structure of the comet as a function of the kinetic parameters for actin assembly. The nucleation model gives a two-stage force–speed relation very similar to the one obtained by Mogilner and Oster [28] and in agreement with the experimental results due to McGrath et al. The autocatalytic model for its part is in better agreement with the observations made by Wiesner et al.: the branch spacing decreases with increasing amounts of capping proteins; the density of functionalisation of the beads affects the density of actin comets but not the propulsion speed; and last but not least, it predicts a speed independent of the force over the range 0–100 pN.

The models due to Mogilner and Oster reveal significant differences to those developed by Carlsson. The first does not take into account filament orientation, and the formation rate does not depend on the amount of existing filaments, but it does treat the traction exerted by filaments tethered to the surface. On the other hand, Carlsson does not consider the attachment of filaments to the surface, but explicitly includes the orientation of filaments in his model. Unfortunately, he does not predict the force production mechanism.

Mesoscopic Models for Force Production

The second class of models applies to a quite different length scale, treating the actin network as an elastic continuum. This type of analysis does not take into account the action of individual filaments, but rather their cooperative action. Gerbal and Prost thus put forward a stress model [31, 32]. In the simplest case of an object with spherical geometry, the particle is surrounded by an actin gel which grows out from the surface and gradually cross-links. When a new layer of actin is inserted between the surface and the gel by filament polymerisation, the pre-existing gel is deformed, storing elastic energy which it then relaxes

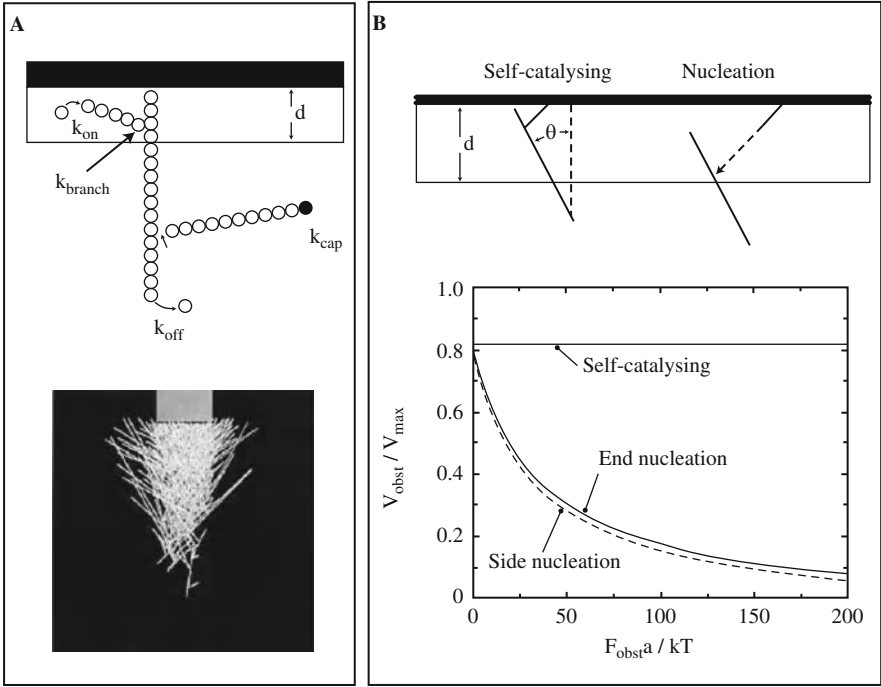


Fig. 5.10. Growth of branched actin networks. Taken from [29, 30]. Reproduced with the kind permission of the Biophysical Society. (A) Simulation of the growth of an actin network against an obstacle. *Top*: Simulation parameters. *Bottom*: Simulation of autocatalytic branching with preferential orientation and restriction of branching at the barbed end. (B) Modelling growth. *Top*: The two possible nucleation processes: in the case of autocatalytic branching, existing branches generate new branches near the surface; in the case of dendritic nucleation, filaments are generated at the surface, independently of existing filaments, and can then attach themselves to the comet network. *Bottom*: Force–speed relations for the two models

by exerting a stress that functions as an elastic propulsion force. When the particle moves, there is also friction between the actin gel and the object which, for its part, opposes the motion. The force–speed relation is obtained by assuming equilibrium between the internal friction and propulsion forces and the external force (see Fig. 5.11). This model can also explain the special case of saltatory motion observed with a *Listeria* mutant [32, 33] and with high-diameter microspheres in the reconstituted motility medium [34].

Note that this mesoscopic model works for a convex geometry and cannot be applied to the concave geometry of the lamellipodium membrane. The next step in the development of these biomimetic systems will be to reconstitute in vitro a system with similar geometry to a membrane protrusion.

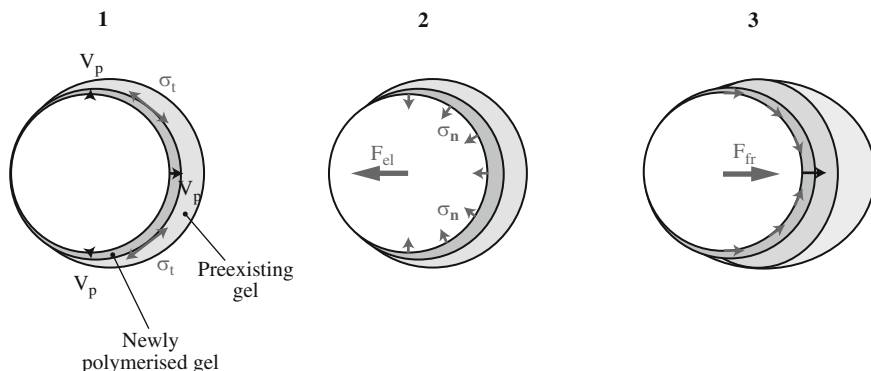


Fig. 5.11. Mesoscopic stress model. Three stages in the production of a force by an actin gel. Insertional polymerisation of actin (at rate v_p) at the interface of the actin gel and the object surface deforms the pre-existing gel, which then stores up elastic energy. The stress is maximal at the outer surface of the gel and zero at the inner surface in contact with the object (stage 1). Integration of this stress leads to a tension inducing a normal stress σ_n at the surface. The integral of these normal stresses is the elastic propulsion force F_{el} (stage 2). The forward displacement of the object to leave the gel creates friction at the surface. The resultant of this friction is an internal force F_{fr} which opposes the motion (stage 3). These two internal forces balance the external force F_{ext} applied to the system, and this determines the speed of the bead in the steady state

5.2.7 Prospects

All the studies described above attempt to measure the force generated by a population of filaments. Ideally, the density of this population should be varied in a controlled way if there is to be any hope of using this kind of result to determine a normalised force–speed curve for one filament. We have seen that, in the systems used up until now, it is still difficult to estimate the exact number of filaments. It is thus natural to seek to measure the force produced by a single filament: such an experiment can be done with formin, which initiates the polymerisation of single, unbranched filaments [14].

In the longer term, the ultimate experiment would of course be to reconstitute an artificial lamellipodium. Indeed, almost all biomimetic experiments have studied the propulsion of functionalised objects, but the problem there has been to reconstitute the movement of bacterial parasites in a host cell, which use the cell's signalling cascade to induce actin polymerisation at their surface. Since they are easier to handle than a cell protrusion, bacteria have been used as a tool to study the biochemical mechanisms of actin-based motility. Today, the molecular processes are quite well understood and can thus be exploited to develop *in vitro* biomimetic systems that bear a closer resemblance to membrane deformations.

Vesicles, or artificial liposomes, are closed bilayers used to model the cell membrane. By encapsulating the motility medium in these vesicles and

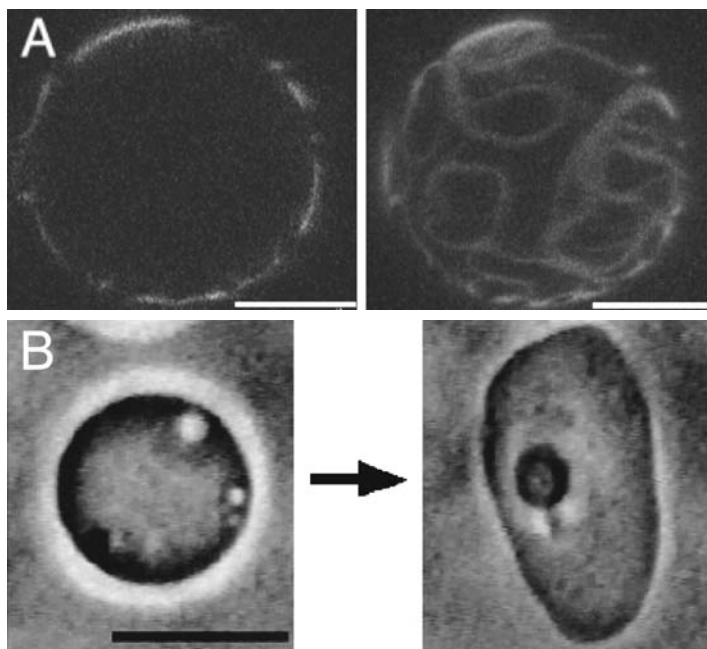


Fig. 5.12. Vesicles containing polymerised actin. Taken from [35,36]. **(A)** Formation of a cortex of filaments near the membrane, observed by confocal microscopy. **(B)** Irregular deformation of an initially spherical vesicle

functionalising their surfaces, one can imagine being able to initiate actin polymerisation against their walls, thereby generating a deformation. Some studies have already been carried out with this in mind, but they remain in their early stages. They consist for the main part in introducing actin into the liposomes and triggering its polymerisation in the bulk. The growth of confined filaments led either to the adsorption of a kind of shell of filaments on the membranes [35], or to a disorganised deformation of the liposomes [36] (see Fig. 5.12). Up to now, nobody has yet attempted to control polymerisation in a directional and localised way. The next step will thus be to functionalise the liposomes in such a way as to control the growth of filaments spatially against the membrane. This system will also share the concave geometry of the lamellipodium.

5.3 Molecular Motors: Myosins and Kinesins

5.3.1 Introduction

Molecular motors are protein nanomachines responsible for the organisation of intracellular space, the operation of muscle, and the beating of cilia and flagella. These motor proteins move along the fibres of the cytoskeleton: myosins move along actin filaments, while kinesins and dyneins move along microtubules. The study of these molecular motors, which has been under way for

many years now, is motivated in terms of both fundamental and medical research. There exist a whole range of interdisciplinary methods in chemistry, biology, and physics, which will eventually lead to a full understanding of the molecular mechanisms used by these motors, opening the way to an abundance of medical applications. In the end, we will be able to create chimeric biological motors and nanomachines, i.e., based on artificial proteins, to carry out specific tasks, both *in vivo* and *in vitro*.

5.3.2 Actin Filaments and Microtubules

Organisation of Actin Filaments and Myosin in Muscle

Muscle tissue represents some 40% of our body weight. Because it is so abundant, muscle tissue has been a subject of study for at least 50 years now, in the hope of obtaining a molecular explanation for motion. The regular transverse striations observed by optical microscopy, with a spacing of 2–3 μm , suggest that muscle fibres are highly organised on the supramolecular level, and these fibres were among the first samples to be examined at higher resolution by new methods of electron microscopy developed in the 1950s. The sarcomeres were discovered in the thin sections of myofibrils and their internal organisation was visualised. Each sarcomere, about 3 μm long, can contract to half its length by the relative slipping motion of thick and thin filaments interacting in a parallel array. The thin filaments (F-actin), about 9 nm in diameter, are made essentially of one protein, actin. The thick filaments are made of another protein, myosin.

Tubulin and Microtubules

Two-way movements were first visualised in the giant axons of squid by differential interference-contrast microscopy. This traffic was then found to be characteristic of neuron cells and other cellular extensions such as cilia and flagella. Such movements correspond to the transport of elements involved in the construction, maintenance, and operation of these long cell extensions. The traffic depends on the interaction between molecular motors and microtubules. The latter are involved in the organisation of intracellular regions during the interphase and in the motion of chromosomes during cell division. Cilium and flagellum microtubules are extremely stable, while those of the mitotic spindle assemble and disassemble in a highly dynamic way. A microtubule is a hollow cylinder of diameter around 25 nm, whose walls are made up of heterodimeric tubulin ($\alpha\beta$ -tubulin) lying head-to-tail in protofilaments aligned with the cylinder axis. In most animal cells, the region around the centrosomes plays the role of an organising center (MTOC) for microtubule nucleation. The resulting microtubule network interacts with many ligands, proteins and organelles. Among the proteins associated with microtubules, there are two motor proteins, dynein and kinesin, for which the microtubules

constitute tracks. Some ligands from plants, like taxol which is obtained from the bark of the yew tree, stabilise the microtubules, while others, like colchicine which is obtained from the meadow saffron, destabilise them. Drugs acting on microtubules have important fundamental and medical applications.

The α and β subunits of the $\alpha\beta$ -tubulin heterodimer comprise about 450 amino acid residues each. α and β associate spontaneously to form an $\alpha\beta$ dimer which is the functional form. Very often, several variants (isoforms) of the two types of tubulin coexist in eukaryotic cells. For example, there are 6α and 7β in mammals, 6α and 9β in plants, and 2α and 2β in the fungus *A. nidulans* [37]. The amino acid sequence of the proteins in each subfamily is about 60% conserved, while α -tubulin and β -tubulin exhibit about 40% sequence similarity.

Similarities Between Actin and Tubulin

The proteins, actin and tubulin, making up the cytoskeleton are among the most conserved eukaryotic proteins. Phylogenetic trees describing the evolutionary relationships between eukaryotic organisms can be constructed by comparing their amino acid sequences. Similar results are obtained by the standard method using ribosomal RNA sequences [38].

Although F-actin and microtubules are fibrous polymers, they result from the self-assembly of globular proteins, actin and tubulin. In both cases, assembly occurs via two pairs of protein–protein interactions. The first pair of interactions is polar and leads to longitudinal assembly, i.e., the formation of protofilaments. The other pair of interactions takes place laterally between protofilaments. Whereas the longitudinal actin–actin interactions lead to formation of a two-stranded helical filament, in the case of microtubules, neighbouring protofilaments associate laterally by interaction between monomers of the same type, i.e., α -tubulin with α -tubulin and β -tubulin with β -tubulin. The result is a hollow cylinder comprising 13 protofilaments. This assembly mechanism confers a structural polarity on the actin filaments and microtubules. In the case of microtubules, the β subunits are located at the plus end (rapid assembly), while the α subunits are located at the minus end. In the intracellular microtubule network, the minus ends are anchored in the organising centre, while the plus ends extend radially toward the cell walls.

The assembly of individual filaments or microtubules can be monitored in vitro by differential interference-contrast microscopy or fluorescence microscopy. When the steady state of the assembly process is reached, the polymers coexist with monomers maintained at a constant concentration. The steady state of the actin filaments is dominated by the treadmilling or head-to-tail polymerisation process, in which the barbed end of the filaments extends while the pointed end depolymerises at the same rate. The steady state in the assembly of microtubules is characterised by dynamic instability, a direct consequence of the dissipative nature of microtubule assembly, in which

the plus and minus ends alternate between rapid depolymerisation phases followed by slow polymerisation. The transitions between these two states are called catastrophe and rescue transitions. The highly dynamic nature of microtubules underlies the rapid reorganisation of the interphase radial array into a bipolar mitotic spindle, allowing segregation of chromosomes.

5.3.3 Motor Proteins

Motors Associated with Actin Filaments

Myosin was discovered and isolated from muscle as an ATPase, i.e., an enzyme catalysing the hydrolysis of adenosine triphosphate, whose activity is stimulated by actin polymerised into filaments. In muscle and in the contractile ring of dividing cells, myosin is organised into bipolar filaments. Myosin in muscle (myosin II) forms a protein complex with total molecular mass around 500 kDa. Myosin II has two heavy chains and two pairs of light chains, while the number of light chains varies from one type of myosin to another. The heavy chains of the myosin associate into dimers of characteristic shape, with two globular heads and a long rigid rod comprising two α -helices that interact to form a superhelix. Myosin is common in muscles and exists in small quantities in almost all other eukaryotic cells. By comparing the sequences of the globular domains, they can be classified into eighteen subfamilies [39]. The human genome has around forty myosin genes, either known or predicted, belonging to ten subfamilies. By comparison, only five myosins are predicted in the yeast *S. cerevisiae*. Three myosins (myosins VIII, XI and XIII) are apparently specific to the plant kingdom. Seventeen myosins (4 myosin VIII and 13 myosin XI) are predicted in the genome of the first plant to be fully sequenced, viz., *Arabidopsis thaliana*.

Genome sequencing projects continue to uncover vast numbers of myosins in eukaryotic organisms. However, their exact roles have not yet been fully understood. These so-called unconventional myosins fulfill important functions in the different muscles, but also in sensory systems such as hearing, balance, and sight. Genetic disorders are associated with mutations in the heavy and light chains. Mutations in the heavy chain of myosin II are involved in familial hypertrophic cardiomyopathy. A sporadic mutation Arg403Gly, responsible for a severe clinical phenotype, is relevant to the sudden death of young athletes by cardiac arrest. Myosins VI, VII and XV are located in the inner ear. Mutations in these proteins produce deformations of the stereocilia in sensory cells, and are involved in genetic forms of deafness.

Motors Associated with Microtubules

The brain sends electrical signals to the muscles via a network of transmission lines which are in fact the axons of the nerve cells. The axon is a very thin tube, about a million times longer than it is wide (two micrometers in diameter and

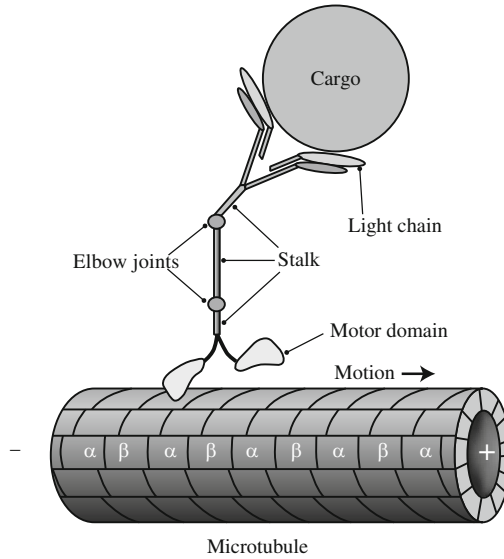


Fig. 5.13. Kinesin molecule transporting its cargo along a microtubule

about one meter long in human beings). The electrical signals and exchanges of cytoplasmic components must transit from the cell body of the neuron in the brain as far as the terminal ramifications of the axons in contact with the muscles. Random (Brownian) diffusion is totally inadequate to this task of directed transport, since several months would be required for an object of the size of a protein to travel from one end of the axon to the other. Two families of molecular motors, the dyneins (a retrograde motor for transport toward the cell body of the neuron) and kinesins (an anterograde motor), able to specifically recognise their cargos, are responsible for axonal transport. These two proteins are chemomechanical enzymes whose function is to interact selectively with microtubules and to convert chemical energy into motion by hydrolysis of ATP. These microtubular motors are needed for many biological functions: protein sorting, intracellular organisation of organelles such as the Golgi apparatus and the endoplasmic reticulum, maintenance of cell polarity, separation of the poles of the mitotic spindle, and segregation of chromosomes during mitosis.

- *Dynein.* It has long been known that the motion of cilia and flagella depends on dynein. Indeed, cilia and flagella have an organelle, the axoneme, characterised by its nine pairs of microtubules arranged in a circle around two central microtubules. The curvature of a cilium or flagellum is induced by the synchronised motion of hundreds of dynein molecules along the microtubules of the axoneme. A transport process within each flagellum is also necessary for the replacement and assembly of its hundreds of protein components. Dynein is a large protein complex comprising 9 to 12 polypeptide chains of about 2×10^6 dalton, whose 500-kDa heavy chains have a globular region interacting with the microtubules in a way that depends

on ATP. There is a three-headed form (3 heavy chains) and a two-headed form. A male hereditary disease known as Kartaneger's syndrome is associated with the dysfunction of dyneins due to a mutation. The impaired cilia and flagella cause chronic bronchitis and sterility. There is another form of dynein in many cells, namely cytoplasmic dynein. This possesses two heavy chains and several light chains, and it is involved in organelle transport along nerve cell axons, and also in cell division. The two types of dynein interact with microtubules and move toward the minus end of the microtubules by ATP hydrolysis.

- *Kinesin*. Kinesin was discovered in the 1980s. Very soon, genetic and immunological methods brought to light an increasing number of molecules in the kinesin family. Several hundred are now known, classified by sequence into 14 subfamilies [40]. Native kinesin (380 kDa) comprises two heavy chains and two light chains. The heavy chain has two globular motor heads, a linear region that is probably organised in a superhelix, and a distal globular region interacting with the light chains and specific cargos (see Fig. 5.13).

A dozen or so different kinesins are involved in cell division. These are potential targets for antimitotic drugs that are of great importance for chemotherapy. Kinesin mutations are responsible for genetic disorders such as the Charcot–Marie–Tooth disorder (a mutation of the kinesin KIFB- β , which perturbs axonal transport and leads to neurological degeneration). The kinesins move along a microtubule at speeds varying between 2 and 150 $\mu\text{m}/\text{min}$ and in a direction that also varies depending on the type of kinesin. The directionality is related to the N-terminal or C-terminal position of the motor domain in the polypeptide chain. Kinesins whose motor domain is in a central position are also known: these are then processive motors for disassembling microtubules, which play an important role in mitosis. Kinesins provide excellent models for understanding the way motor proteins work, since the size of the kinesin motor domain (around 40 kDa) is well suited to structural studies. (Myosin and dynein are three and ten times bigger, respectively.) The kinesin heavy chain can be expressed in bacteria and the pure protein obtained in a functional form.

5.3.4 Motion and Forces

The motion of myosins, kinesins, and dyneins depends on their interaction with their respective partners (actin filaments or microtubules) and the hydrolysis of ATP stimulated by this interaction. For each ATP molecule, the energy released by the reaction $\text{ATP} \longrightarrow \text{ADP} + \text{P}_i$ is $\sim 10^{-19} \text{ J}$, twenty times greater than the thermal energy $k_B T$. The main resistance to the motion of motors and their cargos is due to the viscosity of the liquid medium, and the inertia is negligible. (The Reynolds number, which expresses the relative importance of inertial and viscous forces, is very low for objects with

dimensions less than the micrometer.) To displace a molecular motor through a distance of 10 nm against a force of 1 pN (piconewton), the work done is $W = \text{force} \times \text{distance} \sim 10^{-20}$ J. Under these conditions and for a perfectly efficient system, a kinesin molecule must thus be able to move 100 nm along a microtubule for each molecule of ATP hydrolysed [6, 41].

A wide variety of experimental approaches are used to measure:

- the displacement speed,
- the displacement elementary step,
- the relation between ATP hydrolysis and motion,
- the force generated by a single molecule,
- the basic mechanism of the motion.

These approaches use optical methods that have been under constant development over the past ten years. Observation of myosin and kinesin has been greatly refined by improvements in the various methods of optical microscopy, e.g., fluorescence microscopy, dark-field microscopy, total internal reflection fluorescence microscopy, combined with the development of optical tweezers and fluorescent labelling of individual molecules.

Microtubules, with their cylindrical shape, are relatively rigid structures, and conventional kinesin (kinesin-1) is processive, i.e., it carries out several consecutive catalytic cycles on the same microtubule, covering distances in the micrometer range without detaching. These features have favoured the study of kinesin-1 in two types of experiment (see Fig. 5.14). In the first method, the motion of the microtubules is followed by interference-contrast microscopy as they slide across kinesins immobilised on the surface of a glass slide (see Fig. 5.14A). The displacement speed (around 1 $\mu\text{m/s}$ in the presence of ATP) seems to be independent of the number of kinesin molecules interacting with the microtubule and also of the monomeric or dimeric structure of the kinesin. When a microtubule interacts with a single kinesin molecule, the dimer moves processively along a protofilament of the microtubule while the monomer frequently detaches.

In the second method, an optical trap is used to bring a silica bead of diameter 1 μm carrying the kinesin into the vicinity of a microtubule. In the presence of ATP, the kinesin attached to the bead moves along the microtubule and the bead is monitored by an interference technique (see Fig. 5.14B). The random motion of the bead caused by thermal fluctuations (Brownian motion) is damped by the optical trap. Elementary displacements of 8 nm are measured, and this distance corresponds to the distance between the tubulin dimers along the protofilaments of the microtubules. Each step of 8 nm is accompanied by hydrolysis of one ATP molecule. The optical trap can also be used to apply a force to the bead and it is found that a force of 5 pN stops the motion of the kinesin [42]. More recently, accuracy has been improved by observing latex beads of diameter 0.2 μm by dark-field microscopy with laser illumination. The elementary step size of 8 nm seems to be composed of two sub-steps of 4 nm, the first rather fast and the second somewhat slower. The

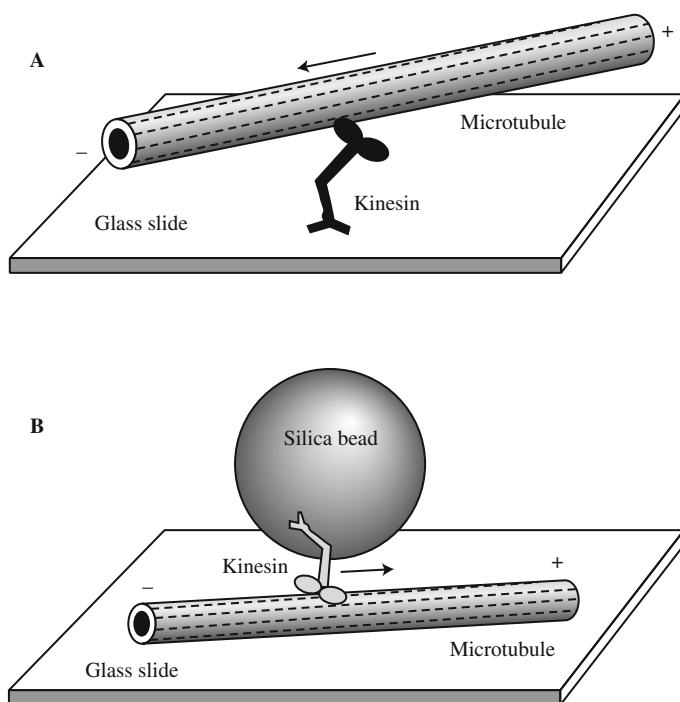


Fig. 5.14. Measuring the kinesin step size. **(A)** The kinesin is attached to a glass slide in a little observation chamber and interacts with microtubules injected into the chamber. The motion of the microtubules is monitored by interference-contrast microscopy or fluorescence microscopy in the presence of ATP. **(B)** Alternatively, microtubules are attached to the slide, and silica (or polystyrene) beads coated with kinesin molecules are injected into the chamber. The bead is often manipulated using optical tweezers and its motion is monitored by interference techniques

maximal force exerted by a kinesin molecule is around 7 pN and the average speed against a force of 1 pN is about 750 nm/s, which corresponds to about 100 individual steps per second+ [43].

Actin filaments are much more flexible than microtubules, and muscle myosin (myosin II) only attaches to actin filaments during a short part of the ATP hydrolysis cycle. To work on this system, the actin filament has to be stretched between two optical traps by attaching silica beads at each end of the filament (see Fig. 5.15). The thread is then set in contact with another bead coated with myosin molecules. Elementary displacements of the order of 5 nm and forces of a few pN were measured. An appropriate statistical approach must be used to analyse such experiments, owing to the low signal-to-noise ratio in the data.

Remarkable progress was made about ten years ago by Funatsu, who was able to visualise single fluorescent molecules of myosin–Cy3 maleimide using evanescent waves [or total internal reflection fluorescence (TIRF) microscopy], thanks to a reduction by a factor of 1,000 in the light background in the image

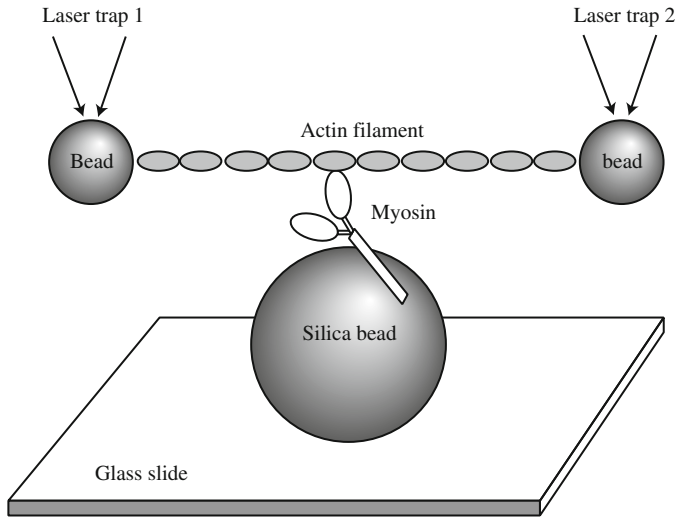


Fig. 5.15. Measuring the myosin step size. An actin filament is stretched between two optical tweezers and the motion of a bead coated with myosin is monitored by optical microscopy

[44]. The very small volume ‘illuminated’ by the evanescent wave close to the surface of a glass slide means that the behaviour of a few individual molecules can be observed in a tiny fraction of the solution. It was also possible to follow the hydrolysis of Cy3-ATP by myosin-Cy5, and the processive motion of a kinesin molecule along a microtubule over distances of around 600 nm (or 75 individual steps) [45]. This result suggests that the kinesin dimer must move along in such a way that at least one of its two motor heads is always in contact with the microtubule.

Myosin V is a processive motor with an elementary step size of about 36 nm, much longer than that of myosin II [46–48]. In principle, it should be possible to follow the motion of motors like myosin V using single fluorescent molecules, but the fundamental problem associated with the use of a single fluorescent molecule is that the resolution is limited by the small number of emitted photons compared with large fluorescent beads or nanoparticles. Note first that the optical image of a fluorescent molecule is much bigger than the molecule itself. It forms a spot of diameter about 250 nm corresponding to the resolution limit imposed by diffraction of visible light according to the Rayleigh criterion (resolution = $NA\lambda/2$, where NA is the numerical aperture of the objective and λ is the wavelength of the light). The light intensity in the spot obeys a Gaussian distribution. The position of the peak can be very accurately measured provided that enough photons are registered to obtain a high signal-to-noise ratio. A hundred successive images of a single fluorophore (total emission of around 10^6 photons) have been obtained by improving the photostability of fluorescent molecules and using a total internal reflection epifluorescence microscope equipped with a slow-scan CCD detector with low

noise [49]. The image of the fluorescent molecule is located with a precision of ± 1.5 nm and a temporal resolution of 0.5 s between successive images. A light chain of myosin V, labelled by a fluorescent molecule, is attached to the region close to one of the motor domains. During the motion of myosin V, the fluorescent spot moves through 74 nm during only one step in two, confirming a hand-over-hand mechanism. The displacement is thus through 37 nm for each elementary step.

An extension of this method (SHREC) can simultaneously visualise the fluorescence of two molecules at two different wavelengths. Hence, Spudich et al. were able to label the calmodulin on a cysteine introduced by mutation with either Cy3–maleimide or Cy5–maleimide [50]. These calmodulins are exchanged for the native light chains of the myosin V. Two filters are then used to simultaneously observe the light emitted by the Cy3 and Cy5 molecules, each located on one of the attachment domains of the light chains. In the presence of ATP, the motions of the Cy3 and Cy5 dyes are shifted with respect to one another, both in space and time, directly indicating that the myosin V moves by successive steps of each of its two motor heads.

The only myosin moving toward the minus end of actin filaments, i.e., myosin VI, still remains something of a mystery. The wild type molecule is a monomer and moves on F-actin in a non-processive way, with an elementary step of the order of 18 nm [51]. A dimeric protein is obtained by expressing the motor domain with an extension of leucine zipper type. This dimeric molecule moves in a processive, hand-over-hand manner, and each elementary step is 30 nm [52–54]. These results are remarkable:

- Do the dimers exist *in vivo*, and if so, how is the monomer–dimer transition regulated?
- The elementary displacements of myosin VI and myosin V are very similar, despite the great difference in length of their light-chain interaction regions. This region plays the part of a lever arm and is much shorter in myosin VI than in myosin V.

5.3.5 Motion and Structural Conformation

Two physical methods, X-ray crystallography and nuclear magnetic resonance (NMR), are used to visualise the structure of macromolecules. The crystallography of very large proteins and protein or nuclear-protein complexes aims to explore, on the atomic scale, the structural changes responsible for the functional mechanisms of a protein.

The structures of motor domains (catalytic domains) in myosin and kinesin have been known for over ten years now. At the present time, many structures coming from different organisms are available at the Protein Data Bank (PDB) [55]. The first myosin structure (obtained in 1993) and kinesin structure (obtained in 1996) caused great surprise at the time [56–58]. They are indeed very similar, even though the myosin motor domain (1,150 amino

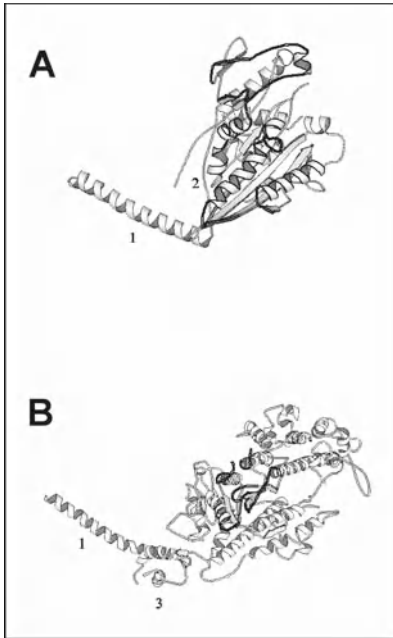


Fig. 5.16. Structural organisation of kinesin (A) and myosin (B) motor domains. The motor domains are: (1) α -helices involved in the dimerisation of kinesin, and which play the role of lever arm in myosin; (2) the region of the kinesin forming a junction between the helix and the motor domain; (3) the converter region of the myosin, which plays an analogous role

acids) is about three times bigger than the kinesin motor domain (340 amino acids) (see Fig. 5.16). In addition, their sequences do not exhibit any significant similarities. Despite these differences, the core of each motor domain comprises an eight-stranded β -sheet flanked by three α -helices on each side.

The catalytic domains (nucleotide binding sites) of the two proteins have many similarities. Two long supplementary amino acid insertions are largely responsible for the greater size of myosin compared with kinesin [56]. Many review articles detail work carried out on the structure–function relation for these two proteins [59].

Myosin Conformations

Between 1993 and 1999, the crystallographic structures of different myosins revealed conformations that seemed to be related to distinct stages in the ATP hydrolysis cycle. Then an elegant piece of work on the S1 fragment of a single myosin isoform (myosin from the bay scallop) demonstrated a clear connection between the conformation of the myosin head and the nucleotide present at the active site (see Fig. 5.17). The three structures obtained [in the absence of the nucleotide, in the presence of MgADP-VO₄ (analogue of ADP•Pi), and in the presence of MgADP] differ in particular in the positions of the so-called converter region and the lever arm [60]. Although these structures are not sufficient to deduce the behaviour of myosin in the presence of actin, they clearly demonstrate that the lever arm responsible for myosin displacement can make large movements as a result of small changes in conformation in the region interacting with the nucleotide.

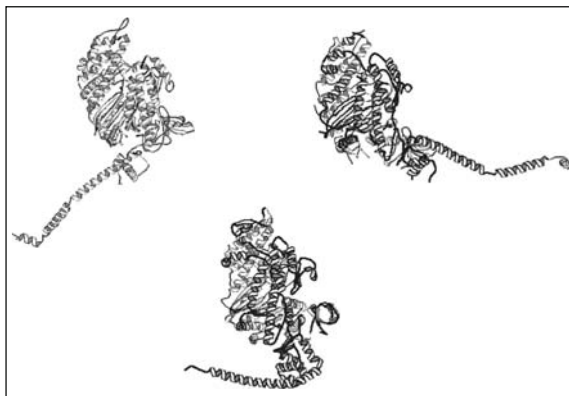


Fig. 5.17. Myosin conformations. The two structures at the *top* represent myosin from the bay scallop. The structure on the *left* was obtained in the presence of ADP and vanadate and corresponds to a transition state, while the structure on the *right*, in the presence of ADP alone, corresponds to the state in which the myosin is detached from the actin filament. The lever arm rotates through a very big angle [60]. The *lower* structure represents myosin VI which moves in the opposite direction to the other myosins [61]. Light chains and calmodulin are not shown

The structural mechanisms allowing myosin VI to move in the opposite direction to all other myosins have been examined recently by determining the crystallographic structure of bacterially expressed myosin VI [61]. This structure shows a calmodulin molecule associated with an insert of 38 amino acids in the converter region. This drives the lever arm (associated with a second calmodulin) in the opposite direction to other myosins such as myosin V (see Fig. 5.17). This explanation for the reversal of the direction of translocation along filaments is similar to the one proposed for the anterograde or retrograde motions of kinesins (see below).

Structure and Directionality of Kinesins

The crystallographic structures of motor domains in conventional kinesins (anterograde motor) and *ncd* (retrograde motor) at resolutions of 0.18 nm and 0.28 nm, respectively, are surprisingly very similar. These proteins are arrow-shaped with dimensions around $75 \text{ \AA} \times 45 \text{ \AA} \times 45 \text{ \AA}$. Inside is an 8-stranded β -sheet flanked on either side by three α -helices [56, 57]. The structure of the motor domain in conventional kinesin is shown in Fig. 5.16A. The almost exact identity of these two structures was greeted with great surprise because these two kinesins move in opposite directions along microtubules. It is thus clear that it is not the structure of the motor domain that determines the directionality of kinesins.

Crystallographic methods are not well-suited to the study of kinesin when it is interacting with a microtubule (nor to the study of myosin in interaction with an actin filament). This type of problem can be tackled by electron

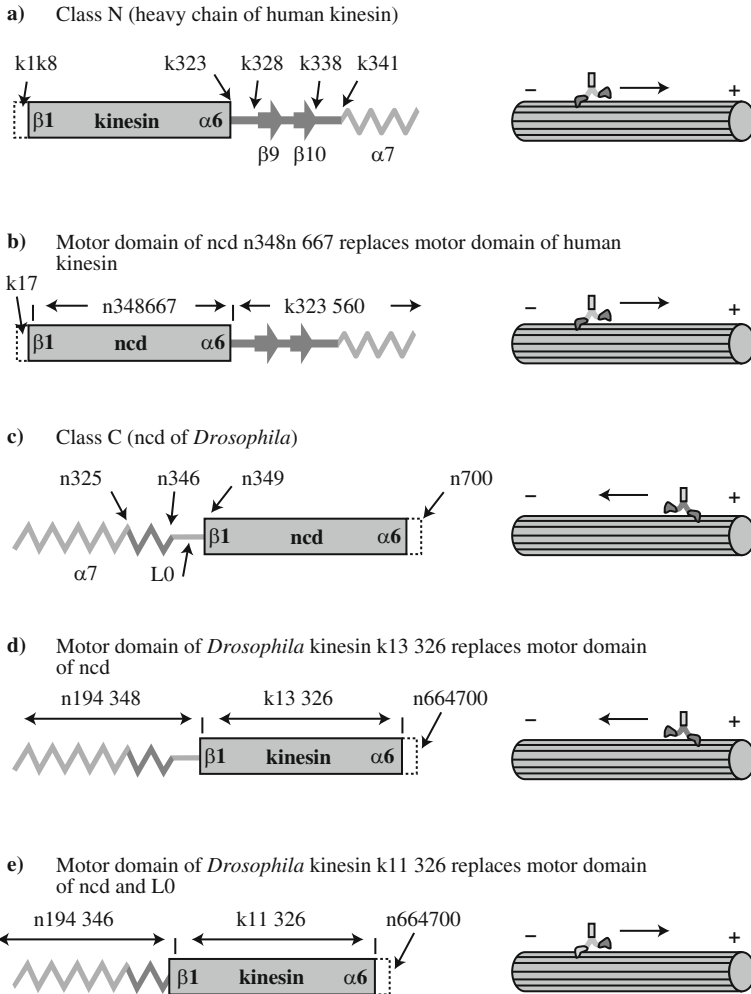


Fig. 5.18. Motion of wild-type kinesins and chimeras (artificial proteins) along microtubules. Conventional kinesin (kinesin-1) moves toward the plus end of a microtubule, while ncd (kinesin-14, the C-terminal motor domain) moves in the other direction. Chimeric proteins, built by exchanging motor domains of the two types of kinesin, can be used to show that the direction of motion depends on a region located outside the motor domain

cryomicroscopy. Three-dimensional images have been obtained of kinesin and ncd dimers interacting with microtubules at a resolution of around 3 nm by cryomicroscopy and computer-aided image reconstruction. A single motor domain of each dimer is attached to the microtubule. The free motor domain is oriented along the direction of motion, i.e., toward the plus end of the microtubule for conventional kinesin and toward the minus end for ncd [62]. This

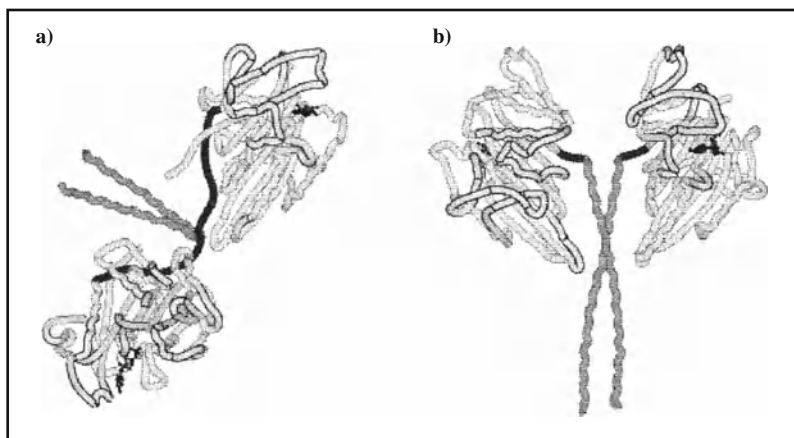


Fig. 5.19. Position of the motor domain in conventional kinesins and ncd. (a) N-kinesin: motor domain at the N-terminus of the polypeptide chain. (b) C-kinesin: motor domain at the C-terminus of the polypeptide chain

result suggests that the direction of motion depends rather on some peripheral region of the motor domain.

This idea has been put to the test in a series of experiments using chimeric (artificial) proteins [63, 64]. Kinesins with a long enough tail dimerise spontaneously. By genetic engineering, the ncd motor domain has been combined with the conventional kinesin tail. The directions of motion of the wild-type kinesin and the chimera are then observed. The chimera moves toward the plus end of the microtubule like the wild-type protein (see Fig. 5.18). In the opposite experiment, a displacement toward the minus end of the microtubule is obtained by replacing the ncd motor domain by the kinesin motor domain. In the second case, the deletion of two residues (Gly347 and Asp348) immediately adjacent to the motor domain changes the direction of motion. These experiments confirm that the direction of motion of kinesins depends on regions located close to the motor domain rather than on the motor domain itself. The crystallographic structures of conventional kinesin and ncd dimers obtained more recently clearly show that these two kinesins have very different connecting regions between the motor domain and the superhelix responsible for dimerisation (see Fig. 5.19) [65].

5.4 ATP Synthase: The Smallest Known Rotary Molecular Motor

5.4.1 Basics of ATP Synthase

ATP synthase is an enzyme complex anchored in the mitochondrial inner membrane, the chloroplast photosynthetic membrane, and the bacterial cytoplasmic membrane. It thus occurs throughout the living world, with the

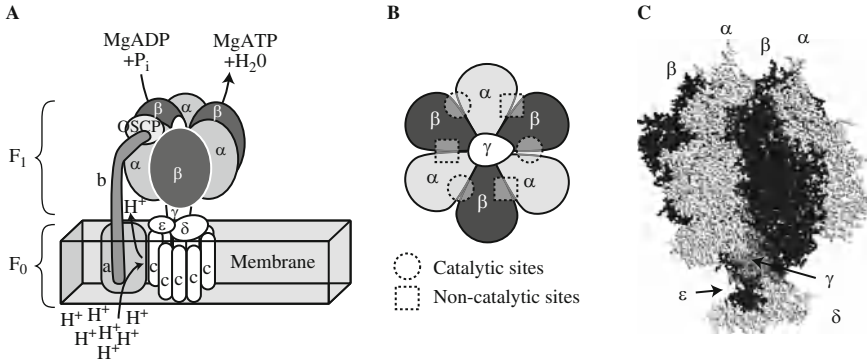
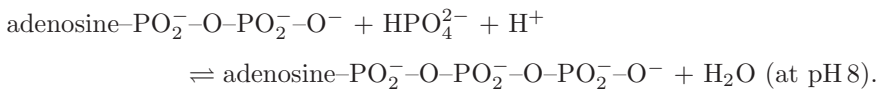


Fig. 5.20. Simplified architecture of mitochondrial ATP synthase. **(A)** Arrangement of the different subunits. All the subunits shown, except for ε , have their equivalent in bacterial and chloroplast ATP synthases. ATP synthase comprises a membrane part F_0 and an extrinsic part F_1 connected by a peripheral stalk. The structure of the extrinsic part is known at atomic resolution. It comprises a heterohexamer, i.e., an alternating arrangement of three α and three β subunits like the petals of a flower, at the interfaces of which are located the six nucleotide binding sites, and a central stalk made of γ , δ , and ε subunits. The membrane part comprises the subunit a (also called subunit 6 in mitochondria), part of subunit b (subunit 4 in mitochondria), and a ring of several copies (10 in mitochondria) of subunit c (subunit 9 in mitochondria). The peripheral stalk comprises the rest of subunit b , the OSCP subunit (mitochondrial nomenclature), and in the case of the mitochondrion, other subunits not shown here. The structure of subunits a and b is not yet accurately known. Subunits are associated in two separate subsets: the rotor ($\gamma\delta\varepsilon c_{10}$, *white* subunits) and the stator ($\alpha_3\beta_3$ [OSCP] ab , *grey* or *black* subunits). The membrane thickness of about 5 nm sets the scale. **(B)** Arrangement of catalytic sites (*circles*) and non-catalytic sites (*squares*) in the $\alpha_3\beta_3$ hexamer (*top view*). **(C)** High-resolution structure of the F_1 part. Image made with Swiss PDB viewer and atomic coordinates from file 1E79 of the Protein Data Bank

exception of viruses. It catalyses the condensation of adenosine diphosphate (ADP) and inorganic phosphate in adenosine triphosphate (ATP) according to the reaction:



This reaction is commonly called ATP synthesis. The reverse reaction, the hydrolysis of ATP, is coupled with a very great many biochemical processes and supplies the energy needed for these processes. ATP can thus be considered as the cell's fuel supply, consumed all the time and constantly recycled, for the main part by ATP synthase in the case of multicellular organisms. Every day, a human being at rest recycles a mass of ATP of the same of order of magnitude as his or her body weight.

In ATP synthase, the ATP synthesis reaction is coupled to an exergonic flow of H^+ ions resulting from a transmembrane electrochemical potential difference, itself generated by redox reactions in the membrane (respiratory or photosynthetic system). ATP synthase has a membrane part called F_0 and an extrinsic part called F_1 . The F_0 part behaves as a conducting channel for protons, while the F_1 part carries six nucleotide binding sites. Three of these sites are catalytic. The role of the three non-catalytic sites is a subject of debate. When the F_1 part is biochemically isolated, it can hydrolyse ATP. It is now known that ATP synthase is a rotary motor comprising a rotor and a stator (see Fig. 5.20).

5.4.2 How ATP Synthase Was Recognised as a Molecular Motor: A Story of Two Conceptual Leaps

A First Conceptual Leap: The Chemiosmotic Theory

For a long time, biochemists sought a hypothetical chemical intermediate between membrane redox reactions and the synthesis of ATP. This quest was brought to a halt in the middle of the 1960s with the success of the chemiosmotic theory, which established that the sought-after intermediate was in fact a transmembrane traffic of H^+ ions. This theory was confirmed in particular by showing that ATP could be synthesised with the help of a purely artificial pH difference, obtained with a total absence of redox reactions.

A Second Conceptual Leap: From Electrochemistry to Nanomechanics

Biochemical Studies

Countless biochemical and enzymological studies, combined with topological and structural investigations [66], laid the foundations for the ATP synthase mechanism [67]:

- Discovery of six nucleotide binding sites, catalytic and non-catalytic.
- Identification of the protein residues closest to the nucleotides, then the structure of the binding sites (see Fig. 5.21) and of the whole extrinsic part.
- Discovery of the cooperative functioning of the catalytic sites.
- Estimates of the stoichiometry between translocated H^+ and synthesised ATP.
- Discovery of the fact that the energy recovered by H^+ ion transfer is mainly used to bind ADP and/or phosphate and to expel ATP, rather than for the chemical synthesis of ATP.
- Demonstration of the fact that, when the enzyme is at work, a particular domain of the γ subunit interacts for an equal length of time with each of the three catalytic sites. This led to the proposal of a rotating mechanism which was subsequently demonstrated by single-molecule experiments.

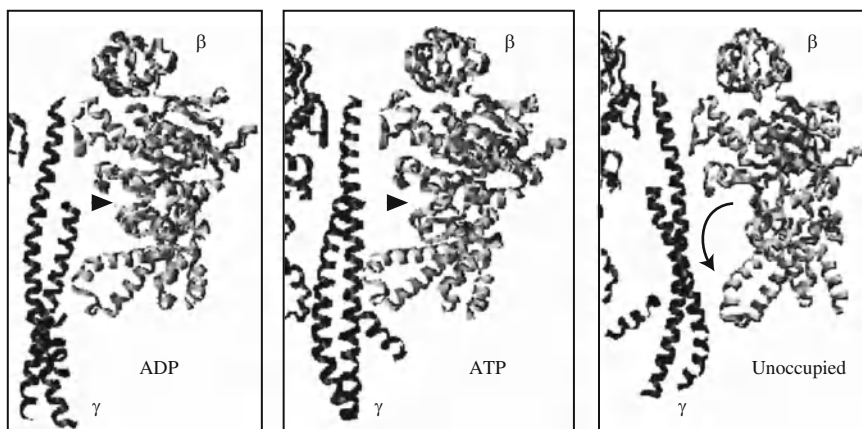


Fig. 5.21. Different conformations of the three β subunits that contribute in a significant way to the structure of the catalytic sites in the crystallised F_1 subcomplex. Each F_1 molecule has a catalytic site filled by ADP (*left*), a site filled by ATP (*center*), and an empty site (*right*). The position at which the ATP or ADP binds is shown by a *black triangle*. The lower part of the β subunit (membrane side) interacts differently with the γ depending on the occupation of the catalytic site. Note also the opening of the unoccupied catalytic site (*arrow*). This structure is supposed to give a still of the successive conformations encountered by each β subunit during the catalytic cycle [66]. Image made with Swiss PDB viewer and atomic coordinates from file 1bmf of the Protein Data Bank

Observations of Single Molecules

The first observation of the rotation of a single molecule was made by M. Yoshida and coworkers in Japan [68]. A fluorescent actin filament was grafted to the base of the γ subunit on the extrinsic part of a bacterial ATP synthase, the upper part of the $\alpha_3\beta_3$ crown being bound onto a substrate. Rotation of the actin filament induced by ATP hydrolysis was then observed by fluorescence microscopy. It occurred in the direction predicted by crystallographic analysis and was sensitive to addition of inhibitor (see Fig. 5.22).

This experiment was much refined in later experiments. To begin with, the experiments originally carried out on the isolated F_1 part were extended to the whole ATP synthase, no longer observing the rotation of the central stalk of the F_1 part, but instead the ring of c subunits which form the membrane part of the rotor (see Figs. 5.20 and 5.24). As far as detection is concerned, the main change was to replace the actin filament by a smaller object to avoid frictional limitations due to the solvent viscosity. Each improvement led to an important step forward in our understanding of the mechanism. Significant examples of such progress are illustrated in Fig. 5.23 [69–71].

Single-molecule experiments thus supply invaluable information about the catalytic mechanism, and in particular about synchronisation of events at the

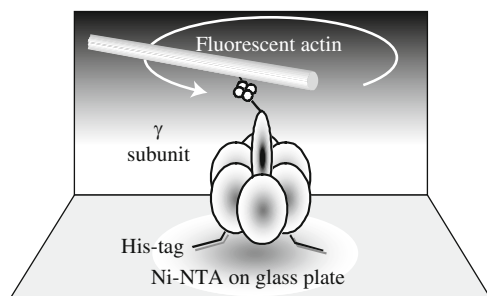


Fig. 5.22. The first single-molecule experiments on ATP synthase. The enzyme was bound to a substrate by its stator part (β subunits) and the motion of a fluorescent actin filament grafted onto the rotor (γ or ϵ subunit for the F_1 subcomplex, c subunits for the F_0F_1 complex) was observed by fluorescence microscopy. The substrate was a glass slide coated with a nickel-charged resin with a strong affinity for histidine. The enzyme anchoring was achieved by a polyhistidine tag introduced by mutagenesis at the N-terminus of the β subunits. By introducing a cysteine onto the rotor, it could be biotinylated and hence bound via a streptavidin molecule to the actin filament, itself biotinylated. This filament was between 1 and 2 μm long. In the presence of ATP, its rotation could be observed in a small proportion of the molecules (about 1%). Seen from above, it occurs in the anticlockwise direction. Successive views of the actin filament are shown schematically on the *right*. Damping due to friction between the actin filament and the solvent cause the rotation to be slow and rather continuous. It eventually stopped, probably due to friction between the actin filament and the substrate [68]

three reaction sites. Most of these experiments did not involve the synthesis of ATP, but rather the opposite hydrolysis reaction. This is due to the fact that current techniques have great difficulty in observing a single molecule anchored on a membrane across which an electrochemical proton gradient (the so-called proton-motive force) is maintained. Only two publications mention such an experiment. The first used a somewhat indirect method (FRET) to visualise changes in the rotor position [72]. The second involved disconnecting the $\alpha_3\beta_3$ hexamer from the rest of the stator, so that it could rotate at the same time as the rotor [73]. In the latter case, there is no distortion of the catalytic sites and the enzyme rotates without synthesising ATP.

How to Make ATP Without a Proton-Motive Force

A team of Japanese specialists in the nanomechanics of ATP synthase managed to get the rotor working in a completely artificial way [74]. To do this, a magnetic nanobead was grafted onto the rotor of ATP synthase, having immobilised the stator as described earlier. By applying an external magnetic field rotating in one direction or the other, the synthesis or hydrolysis of ATP was observed, depending on the direction of rotation. This experiment was not carried out on a single molecule, but on a population of ATP synthases,

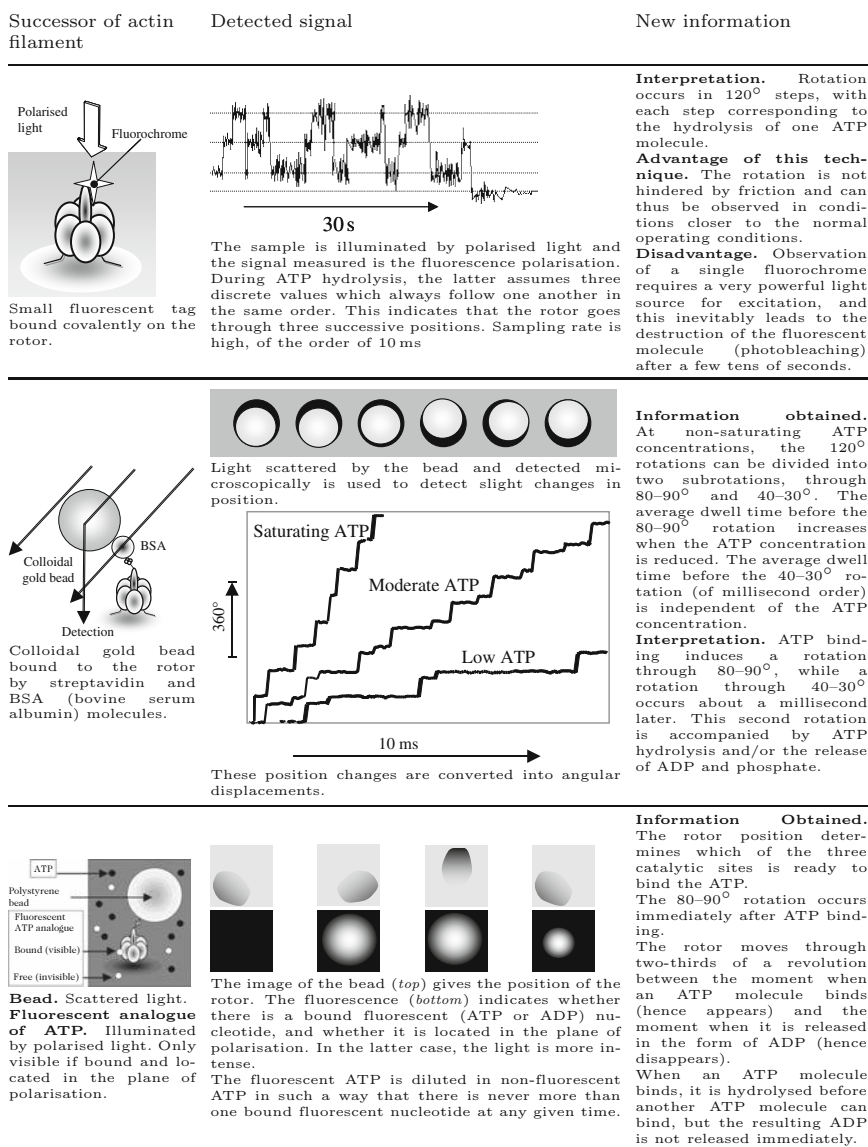


Fig. 5.23. Successive improvements in single-molecule observations. From [69–71]

the level of ATP being measured by a standard technique, viz., light emission by a luciferin–luciferase system. This was the first demonstration that ATP can be synthesised in a purely mechanical way, by removing the need for the proton-motive force.

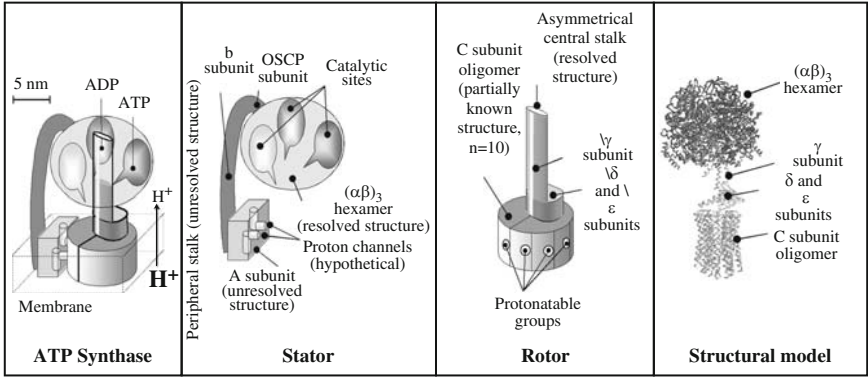


Fig. 5.24. Schematic view of ATP synthase

5.4.3 Rotation Mechanism: Current Understanding

Basic Principles

ATP synthase involves two coupled molecular motors. One is located in the extrinsic part and couples the rotation of the central stalk with the synthesis of ATP. The other one, in the membrane, is a turbine operated by the flow of protons. We shall now discuss the way these two motors work, with the help of a simplified structure (see Fig. 5.24).

Schematic View of ATP Synthase

From left to right in Fig. 5.24 are depicted the whole complex, the stator, the rotor, and a structural model obtained by crystallographic studies.

On the left is shown an element of the membrane. The latter separates two compartments between which a proton electrochemical potential difference (the proton-motive force) is maintained, as indicated by the two H⁺ symbols of different sizes. The arrow shows the direction of spontaneous migration of the H⁺ ions.

The diagram of the stator shows a membrane part, an extrinsic part, and the peripheral stalk. Two hypothetical proton channels are shown in the a subunit. Each channel is assumed to have one end flowing toward the middle of the membrane, opposite one of the protonatable groups of the rotor face (see neighbouring diagram). The other end is supposed to flow into either the upper or the lower compartment. These two channels would thus carry H⁺ ions between the compartments bounded by the membrane and the protonatable groups of the rotor, located toward the middle of the membrane.

The diagram of the rotor shows the membrane part made up of c subunits. Each c subunit carries a protonatable residue (aspartate in mitochondria and *E. coli*, glutamate in chloroplast) located in the middle of the membrane. The non-membrane part forms an axis that penetrates the α₃β₃ hexamer of the stator.

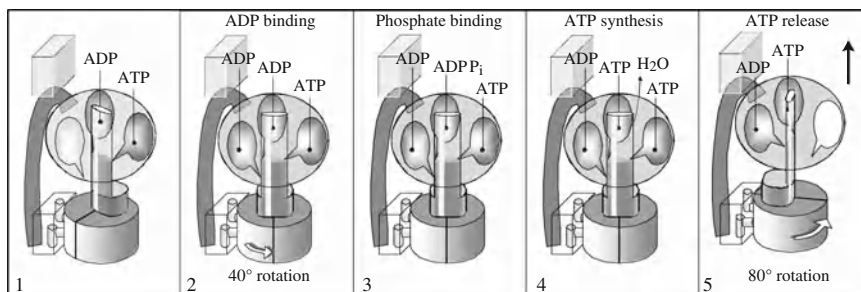


Fig. 5.25. Coupling between motion of the rotor and the different stages of ATP synthesis

In the structural model (1QO1 of the Protein Data Bank), only the α , β , γ , δ , ϵ , and c subunits are visible. The a, b, and OSCP subunits were probably eliminated by purification and/or crystallisation of the complex.

Catalytic Part

ATP synthesis is induced by rotation of the γ subunit inside the hexamer formed by the α and β subunits. The γ subunit is itself carried along by the membrane motor. Owing to its asymmetric character, the γ subunit sequentially distorts the three catalytic sites, forcing them successively to bind ADP and phosphate and expel ATP. A plausible scenario is illustrated in Fig. 5.25.

Coupling Between Motion of the Rotor and the Different Stages of ATP Synthesis

The flow of protons through the membrane part of the complex is not shown in Fig. 5.25. This flow pushes the rotor in the anticlockwise direction, but it can only actually rotate if the catalytic sites happen to be in certain configurations. The rotation modifies the interactions between the central stalk and the three catalytic sites, inducing sequential distortions in these sites which alter their affinity for the nucleotides. The lateral stalk prevents the $\alpha_3\beta_3$ hexamer from rotating at the same time as the central stalk. From enzymological and crystallographic data, and single-molecule observations of ATP hydrolysis, the following scenario can be suggested for ATP synthesis:

- In the initial state (1), one catalytic site is occupied by an ATP molecule and another by an ADP molecule, while the third is unoccupied. The first two sites have a closed configuration and interact with the asymmetric axis of the rotor. The position of the rotor is stable.
- An ADP molecule binds to the unoccupied site, changing its conformation. This may lead to conformational changes in adjacent non-catalytic sites (not shown), and even, from one site to the next, in other catalytic sites. However, this is still a long way from being demonstrated. In any case, this conformational change

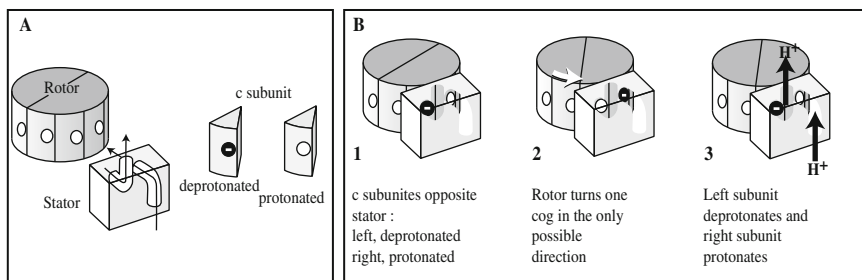


Fig. 5.26. Possible mechanism for the membrane turbine

releases the rotor which, pushed by the proton flow, rotates through 40° , leading to state (2). Interactions between the rotor and the other two catalytic sites are modified.

- With this new configuration, an inorganic phosphate molecule P_i binds to the catalytic site which originally contained the ADP, leading to state (3). ADP and P_i spontaneously form an ATP molecule with elimination of a water molecule. This stage, which is practically isoenergetic, requires no motion from the rotor. Stage (4) is thus reached.
- These new conformational changes in the catalytic sites release the rotor which rotates through another 80° under the proton flow. The interactions of the rotor with the catalytic sites are once again modified, causing an ATP molecule to be expelled. State (5) is reached, equivalent to state (1) up to a rotation of 120° . So one third of a revolution of the rotor has caused the net synthesis of one ATP molecule in the medium, and the cycle can continue.

Functioning of the Membrane Motor

Our understanding of the way the membrane turbine is run by protons is much more speculative, mainly due to the lack of structural data. However, there is one model which, although not unanimously accepted, does provide a good working hypothesis (see Fig. 5.26). There are two key features: Brownian motion and the incompatibility between electric charge and an apolar medium [75].

Possible Mechanism for the Membrane Turbine

Figure 5.26A shows the membrane part with the stator and its two proposed proton channels, together with the ring of c subunits, which can be protonated or deprotonated. The membrane (not shown here) separates a lower compartment, rich in H^+ ions, from an upper compartment, deficient in H^+ ions, which contains the extrinsic part of the ATP synthase.

Figure 5.26B shows the rotation stages. In state (1), the protonatable (carboxyl) group connected to the upper compartment is kept deprotonated, hence negatively charged. The one connected to the lower compartment is protonated, hence neutral.

The other protonatable groups, immersed in the lipid matrix, are protonated. Brownian excitation tends to make the rotor rotate randomly. However, clockwise rotation is impossible for energy reasons, because it would bring a negative charge in contact with the lipid phase with very low dielectric constant. On the other hand, anticlockwise rotation is allowed. It leads to state (2), bringing the negative charge opposite the channel connected with the lower compartment. A protonated (neutral) group is now located opposite the channel connected to the upper compartment. Then, due to the H^+ ion activities of the two compartments, the negatively charged group protonates and the neighbouring group deprotonates, giving state (3). This state is analogous to state (1) after rotation through one unit and the process continues. After a complete revolution of the turbine, a number of H^+ ions equal to the number of c subunits has crossed the membrane. Apart from the assumption concerning the existence of the two H^+ channels, this model is otherwise very economical with regard to hypotheses. However, other phenomena may play a role in the rotation, e.g., conformation changes in the c subunit depending on its state of protonation. Such changes have been observed by NMR for an isolated c subunit in monomer form.

5.4.4 Thermodynamics, Kinetics, and Nanomechanics

Mechanical Energy Produced (Consumed) by ATP Hydrolysis (Synthesis)

It is possible to calculate the maximum mechanical torque that can be produced by the hydrolysis of an ATP molecule (or, what amounts to the same thing, the torque that must be applied to synthesise an ATP molecule). The energy produced by ATP hydrolysis, and called the phosphate potential in the jargon of bioenergetics, is given by

$$\Delta G_p = \Delta G'_0 - 2.3RT \log \frac{[ATP]}{[ADP][P_i]}.$$

By convention, the energy produced is negative. Under typical experimental conditions ($[ATP] = 2 \text{ mM}$, $[ADP] = 10 \mu\text{M}$, $[P_i] = 1 \text{ mM}$), and for pH 7 at 20°C , where $\Delta G'_0 = -30 \text{ kJ mole}^{-1}$, we obtain $\Delta G = -60,730 \text{ kJ mole}^{-1}$, which represents 10^{-19} J per molecule. This energy must rotate the rotor through 120° . The torque will therefore be

$$\tau = \frac{3}{2\pi} 10^{-19} \text{ J} = 4.8 \times 10^{-20} \text{ J},$$

or 48 pN nm.

Rotation experiments with the actin filament were used to estimate, in a completely different way, the torque exerted by the hydrolysis of one ATP molecule. This was found from the average angular speed of the filament and its hypothesised hydrodynamic properties. The values obtained were a few tens of pN nm [68], which is of the same order of magnitude as the value found by thermodynamic calculations.

Energy Steps

Many enzymological experiments have suggested that, during ATP synthesis, the reaction step using the energy of the electrochemical proton gradient (proton-motive force) was not the ATP synthesis itself, but rather the binding of substrates and release of the product [67]. This was corroborated by single-molecule ATP hydrolysis experiments, where rotational motions appear to be coupled to ATP binding and ADP release, but not to the hydrolysis reaction itself, which could thus be mechanically silent [70, 71].

An Old Problem Revisited: H^+ /ATP Stoichiometry

Thermodynamic Determination

The stoichiometry between translocated H^+ ions and synthesised ATP molecules has always been considered an important mechanical parameter to be determined. The method most commonly used to estimate it is thermodynamics. At equilibrium between synthesis and hydrolysis of ATP, the total ΔG for ATP synthesis and proton transport must be zero:

$$\Delta G_p + n\Delta\tilde{\mu}_{H^+} = 0,$$

where ΔG_p is the phosphate potential defined above, $\Delta\tilde{\mu}_{H^+}$ is the proton electrochemical potential difference, and n is the desired H^+ /ATP stoichiometry.

If $\Delta G > 0$, there is a net synthesis of ATP, and if $\Delta G < 0$, there is hydrolysis of ATP. It follows that the bigger the value of n , the more ATP synthesis will be favoured thermodynamically, which is logical enough, since the system adds the energy due to the transfer of $n H^+$ ions to synthesise a single ATP molecule. The thermodynamic force $\Delta\tilde{\mu}_{H^+}$ is given by

$$\Delta\tilde{\mu}_{H^+} = F\Delta\psi - 2.3RT\Delta pH,$$

where $\Delta\psi$ is the transmembrane electrical potential difference and ΔpH is the transmembrane pH difference. Experiments to determine n , generally carried out in mitochondria or chloroplasts, but occasionally in artificial vesicles, consist in creating conditions in which the proton flow produced by the electron transfer chain is exactly balanced by membrane leakage. This means that no net flow of protons crosses the ATP synthases, which are therefore in thermodynamic equilibrium, i.e., $\Delta G = 0$. One then measures ΔG_p (which is relatively easy) and $\Delta\tilde{\mu}_{H^+}$ (which is extremely difficult), whence the value of n can be deduced. The values obtained have been bitterly debated. The most plausible values (obtained by this method or other functional methods that are just as difficult to implement) give about 3 for mitochondrial ATP synthase and about 4 for chloroplast ATP synthase.

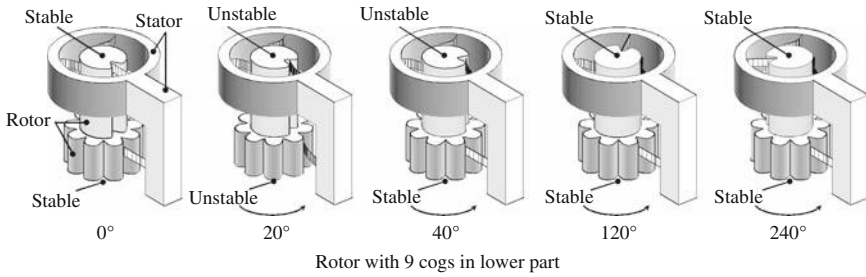


Fig. 5.27. Mechanical model of ATP synthase with stepwise rotation

H⁺/ATP Stoichiometry and Number of c Subunits

Now that the mechanism of ATP synthase is better known, the structural significance of the H^+/ATP stoichiometry is stunningly obvious. Indeed, since there are three catalytic sites, the number of ATP molecules synthesised per revolution of the wheel is three. Regarding the number of H^+ ions transported per revolution, it seems reasonably likely that it must equal the number of c subunits in the turbine. The H^+/ATP stoichiometry must therefore be equal to one-third of the number of c subunits. Now the number of c subunits has been determined by structural studies in a small number of cases. It is 10 in the mitochondria of *S. cerevisiae* [76], 14 in the chloroplast of *S. oleracea* [77], probably 10 in the bacterium *E. coli* [78], and 11 in another bacterium, *I. tartaricus*, in which ATP synthase works with Na^+ ions rather than H^+ [79]. This would give H^+/ATP (or Na^+/ATP) stoichiometries of 3.33 for the mitochondrion, 4.67 for the chloroplast, 3.33 for *E. coli*, and 3.67 for *I. tartaricus*. Comparison with the thermodynamic measurements shows that standard bioenergetics does not fall so far from the mark. Note in passing that the H^+/ATP stoichiometry deduced from the structure is never a whole number in the cases examined here.

Is There Any Reason Why the Number of c Subunits Should Not Be a Multiple of Three?

According to several authors, the fact that the number of c subunits of ATP synthase is not a multiple of 3 would increase the kinetic performance of the enzyme, by preventing the rest positions of the γ -subunit central stalk in the $\alpha_3\beta_3$ crown from coinciding too often with those of the c subunits in the membrane. Indeed, such a coincidence leads to an energy well, and hence when all is said and done to a high activation energy and a low reaction rate. This idea is illustrated in Fig. 5.27 by a simple mechanical model based on a stepwise rotation. (The distortion of catalytic sites and ATP synthesis are not shown.)

Mechanical Model of ATP Synthase with Stepwise Rotation

The stable positions causing discontinuity are created by notches in the rotor, into which an elastic slat is inserted (see Fig. 5.27). The three slats at the top correspond to catalytic sites, giving three rest positions corresponding to the pauses observed for a single molecule during ATP hydrolysis with a saturating substrate concentration. The bottom slat corresponds to the contact area between the membrane part of the stator and the ring of c subunits. The number of cogs on the lower part, which gives the number of rest positions, is equal to the number of c subunits. Starting from a position referred to as 0° where the upper (extrinsic) part and the lower (membrane) part are each stabilised by a slat–notch torque. In the present model, called the nine-cog model, where the cogs in the lower part are 40° apart, this particularly stable double rest position will recur after rotations through 120° and 240° , i.e., three times per revolution.

For a rotor with ten cogs separated by 36° , neither a rotation of the lower part through 3 cogs (108°), nor a rotation of the lower part through 4 cogs (144°) would correspond to a stable position of the upper part. One would have to wait for a whole revolution before the two stable positions could coincide again. The reduced frequency of this doubly stable position makes rotation easier. One might imagine that the force required to deform the elastic slats has an energy cost and reduces the efficiency of the operation. In fact, this is not true at all, since the existence of the rest positions only affects the time required to accomplish one revolution. One has to postulate that the deformation of these imaginary slats is generated by Brownian excitation, which takes into account the well-known fact that the speed increases with temperature.

Although the principle of kinetic optimisation looks fairly convincing, it must be treated with some caution, because to begin with the number of c subunits is only known for a small number of ATP synthases. Furthermore, it is established, at least for *E. coli*, that the peripheral stalk connecting the two parts of the stator is flexible. This implies that the extrinsic part of the stator might rotate transiently by pulling on the stalk, before being dragged back and undergoing the distortion of the catalytic sites. The central stalk may also be slightly elastic. This would suffice to desynchronise the rest positions of the membrane and extrinsic parts, whatever the number of c subunits.

H⁺/ATP Stoichiometry and Gearing. The Problem of Type V ATPases

A high H^+ /ATP stoichiometry is thermodynamically favourable for ATP synthesis to the detriment of hydrolysis. This rule is easily transposed in mechanical terms. A large number of c subunits greatly increases the torque generated by the proton-motive force, and this helps to force the distortion of the catalytic sites in the direction of ATP synthesis, even in the presence of the reaction product, whose rebinding and hydrolysis push the rotor in the opposite direction. This cycling principle also applies to a large family of type V ATPases, molecular motors that are structurally and mechanically close to ATP synthases, but whose function is to create proton gradient by hydrolysing ATP. The uses of this proton-motive force are extremely varied: accumulation

of neurotransmitters in presynaptic vesicles, production of a membrane potential in certain epithelial cells, internal acidification of endocytosis or exocytosis vesicles, and so on. What about the H^+ /ATP stoichiometry of V-ATPases? Few reliable estimates are available, but it is generally thought to be lower than for the ATP synthases, which would indeed tend to favour ATP hydrolysis. The V-ATPases generally have three catalytic sites, and it was generally assumed that their rotor had only 6 protonatable segments, compared with 10–14 for the ATP synthases. In this way, everyone was happy. Unfortunately, this comfortable state of affairs has just been disturbed by a publication showing that the Na^+ V-ATPase rotor of the bacterium *Enterococcus hirae* in fact possesses 10 protonatable segments [80]. Could this bacterial V-ATPase be an exception? It seems unlikely. We may have to conclude, at least temporarily, that the preferential direction of operation of ATP synthases and V-ATPases is not actually dictated by thermodynamics, but rather by kinetics, together with various control mechanisms (see below).

5.4.5 Conclusion

The ATP synthases (and V-ATPases) are extraordinary enzymes, because they couple an electrochemical process (translocation of H^+ ions) with a chemical reaction (ATP synthesis) via a mechanical process which long remained invisible to us. They represent a case study for anyone wishing to understand thermodynamics and kinetics at the microscopic scale, and hence to make the connection with nanomechanics. The single-molecule approach provides invaluable information and has led to major breakthroughs in our understanding of this mechanism. However, the ground had been well prepared by more than thirty years of biochemical and enzymological experimentation using techniques that are of course still relevant.

What problems remain to be solved? Here are a few of them:

- The mechanism of the F_1 part is just beginning to come to light, but the details of the catalytic cycle and the propagation of movements between the different subunits remains unknown. The role of the three non-catalytic sites is extremely controversial and thus remains a mystery.
- The F_0 part is another black box, whose structure remains hypothetical. Mitochondrial ATP synthases also possess extra subunits whose structure and topology are currently under investigation. Most of them seem to be associated with the peripheral stalk, but their role is debated.
- Finally, although ATP synthases are reversible machines, there are mechanisms which regulate, even totally forbid, the hydrolysis of ATP in the absence of an electrochemical potential. These mechanisms are not the same in mitochondrial, chloroplast, and bacterial ATP synthases. They have been under investigation for a long time now, but our current understanding compels us to interpret them in mechanical terms. In some cases, e.g., chloroplasts and some bacteria, a constitutive subunit may act

as a molecular ratchet [81]. In other cases, e.g., mitochondria, a soluble inhibitor peptide gets caught at the α/β catalytic interface when the enzyme begins to hydrolyse the ATP, and it is ejected by application of a proton-motive force. The relation between conformational changes in the enzyme, binding of the inhibitor peptide, and blocking of the rotation mechanism is extremely complex [82]. All these control mechanisms, only roughly understood on the molecular level, are being intensively studied at the present time. It will be essential to elucidate their purpose if we are to grasp the workings of these fascinating nanomachines, the ATP synthases.

References

1. Pawson, T., Nash, P.: *Science* **300**, 445–452 (2003)
2. Feynman, R.P.: Annual Meeting of the American Physical Society (1959) www.zyvex.com/nanotech/feynman.html
3. Lowe, C.R.: *Curr. Opin. Struct. Biol.* **10**, 428–434 (2000)
4. Sablin, R.P., Dawson, J.F., Van Look, M.S., Spudich, J.A., Egelman, E.H., Fletterick, R.J.: How does ATP hydrolysis control actin's association?, *Proc. Natl. Acad. Sci. USA* **99** (17), 10945–7 (2002)
5. Nogales, E.: A structural view of microtubule dynamics, *Cell Mol. Life Sci.* **56** (1–2), 133–42 (1999)
6. Bray, D.: *Cell Movements*, Garland (2000)
7. Lauffenburger, D.A., Horwitz, A.F.: *Cell* **84**, 359–369 (1996)
8. Euteneuer, U., Schliwa, M.: *Nature* **310**, 58–61 (1984)
9. Small, J.V., Herzog, M., Anderson, K.: *J. Cell Biol.* **129**, 1275–1286 (1998)
10. Ponti, A., Machacek, M., Gupton, S.L., Waterman-Storer, C.M., Danuser, G.: *Science* **305**, 1782–1786 (2004)
11. Holmes, K.C., Popp, D., Gebhard, W., Kabsch, W.: *Nature* **347**, 44–49 (1990)
12. Pantaloni, D., Le Clainche, C., Carlier, M.-F.: *Science* **292**, 1502–1506 (2001)
13. Loisel, T.P., Boujemaa, R., Pantaloni, D., Carlier, M.-F.: *Nature* **401**, 613–616 (1999)
14. Romero, S., Le Clainche, C., Didry, D., Egile, C., Pantaloni, D., Carlier M.-F.: *Cell* **119**, 419–429 (2004)
15. Gerbal, F., Laurent, V., Ott, A., Carlier, M.-F., Chaikin, P., Prost, J.: *Eur. Biophys. J.* **29**, 134–140 (2000)
16. Wiesner, S., Helfer, E., Didry, D., Ducouret, G., Lafuma, F., Carlier, M.-F., Pantaloni, D.: *J. Cell Biol.* **160**, 387–398 (2003)
17. McGrath, J.L., Eundamrong, N.J., Fisher, C.I., Peng, F., Mahadevan, L., Mitchison, T.J., Kuo, S.C.: *Curr. Biol.* **13**, 329–332 (2003)
18. Evans, E., Rawicz, W.: *Phys. Rev. Lett.* **64**, 2094–2097 (1990)
19. Upadhyaya, A., Chabot, J.R., Andreeva, A., Samadini, A., van Oudenaarden, A.: *Proc. Natl. Acad. Sci. USA* **100**, 4521–4526 (2003)
20. Giardini, P.A., Fletcher, D.A., Theriot, J.: *Proc. Natl. Acad. Sci. USA* **100**, 6493–6498 (2003)
21. Boukellal, H., Campas, O., Joanny, J.-F., Prost, J., Sykes, C.: *Phys. Rev. E* **69**, 061906-1–061906-4 (2004)

22. Marcy, Y., Prost, J., Carlier, M.-F., Sykes, C.: *Proc. Natl. Acad. Sci. USA* **101**, 5992–5997 (2004)
23. Upadhyaya, A., van Oudenaarden, A.: *Curr. Biol.* **13**, R734–R744 (2003)
24. Arnold, M., Cavalcanti-Adam, E.A., Glass, R., Blummel, J., Eck, W., Kantlehner, M., Kessler, H., Spatz, J.P.: *Chem. Phys. Chem.* **5**, 383–388 (2004)
25. Peskin, C.S., Odell, G.M., Oster, G.F.: *Biophys. J.* **65**, 316–324 (1993)
26. Goldberg, M.B., Theriot, J.A.: *Proc. Natl. Acad. Sci. USA* **92**, 6572–6576 (1995)
27. Mogilner, A., Oster, G.: *Biophys. J.* **71**, 3030–3045 (1996)
28. Mogilner, A., Oster, G.: *Biophys. J.* **84**, 1591–1605 (2003)
29. Carlsson, A.E.: *Biophys. J.* **81**, 1907–1923 (2001)
30. Carlsson, A.E.: *Biophys. J.* **84**, 2907–2918 (2003)
31. Gerbal, F., Noireaux, V., Sykes, C., Jülicher, F., Chaikin, P., Ott, A., Prost, J., Golsteyn, R.M., Friederich, E., Louvard, D., Laurent, V., Carlier, M.-F.: *PRAMANA-J. Phys.* **53**, 155–170 (1999)
32. Gerbal, F., Chaikin, P., Rabin, Y., Prost, J.: *Biophys. J.* **79**, 2259–2275 (2000)
33. Lasa, I., Gouin, E., Goethals, M., Vancompernelle, K., David, V., Vandekerckhove, J., Cossart, P.: *EMBO J.* **16**, 1531–1540 (1997)
34. Bernheim-Grosswasser, A., Wiesner, S., Golsteyn, R.M., Carlier, M.-F., Sykes, C.: *Nature* **417**, 308–311 (2002)
35. Limozin, L., Bärmann, M., Sackmann, E.: *Eur. Phys. J. E* **10**, 319–330 (2003)
36. Miyata, H., Nishiyama, S., Akashi, K., Kinoshita, K. Jr: *Proc. Natl. Acad. Sci. USA* **96**, 2048–2053 (1999)
37. Luduena, R.F.: *Int. Rev. Cytol.* **178**, 207–275 (1998)
38. Baldauf, S.L., Roger, A.J., Wenk-Siefert, I., Doolittle, W.F.: *Science* **290**, 972–977 (2000)
39. Myosin Homepage: www.mrc-lmb.cam.ac.uk/myosin/myosin.html
40. Kinesin Homepage: www.proweb.org/kinesin/
41. Howard, J.: *Ann. Rev. Physiol.* **58**, 703–729 (1996)
42. Svoboda, K., Block, S.M.: *Cell* **77**, 773–784 (1994)
43. Nishiyama, M., Muto, E., Inoue, Y., Yanagida, T., Higuchi, H.: *Nat. Cell Biol.* **3**, 425–428 (2001)
44. Funatsu, T., Harada, T., Tokunaga, M., Salto, K., Yanagida, T.: *Nature* **374**, 555–559 (1995)
45. Yildiz, A., Tomishige, M., Vale, R.D., Selvin, P.R.: *Science* **303**, 676–678 (2004)
46. De la Cruz, E.M., Wells, A.L., Rosenfeld, S.S., Ostap, E.M., Sweeney, H.L.: *Proc. Natl. Acad. Sci. USA* **96**, 13726–13731 (1999)
47. Mehta, A.D., Rief, M., Spudich, J.A., Smith, D.A., Simmons, R.M.: *Science* **283**, 1689–1695 (1999)
48. Walker, M.L., Burgess, S.A., Sellers, J.R., Wang, F., Hammer, J.A., Trinick, J., Knight, P.J.: *Nature* **405**, 804–807 (2000)
49. Yildiz, A., Forkey, J.N., McKinney, S.A., Ha, T., Goldman, Y.E., Selvin, P.R.: *Science* **300**, 2061–2065 (2003)
50. Churchman, L.S., Otken, Z., Rock, R.S., Dawson, J.F., Spudich, J.A.: *Proc. Natl. Acad. Sci. USA* **102**, 1419–1423 (2005)
51. Lister, I., Roberts, R., Schmidt, S., Walker, M., Trinick, J., Viegel, C., Buss, F., Kendrick-Jones, J.: *Biochem. Soc. Trans.* **32**, 685–688 (2004)
52. Yildiz, A., Park, H., Safer, D., Yang, Z., Chen, L.-Q., Selvin, P.R., Sweeney, H.L.: *J. Biol. Chem.* **279**, 37223–37226 (2004)

53. Rock, R.S., Ramamurthy, B., Dunn, A.R., Beccafico, S., Rami, B.R., Morris, C., Spink, B.J., Franzini-Armstrong, C., Spudich, J.A., Sweeney, H.L.: *Mol. Cell* **17**, 603–609 (2005)
54. Okten, Z., Churchman, L.S., Rock, R.S., Spudich, J.A.: *Nat. Struct. Mol. Biol.* **11**, 884–887 (2004)
55. Protein Data Base: www.rcsb.org/pdb
56. Sablin, E.P., Kull, F.J., Cook, R., Vale, R.D., Fletterick, R.J.: *Nature* **380**, 550–555 (1996)
57. Kull, F.J., Sablin, E.P., Lau, R., Fletterick, R.J., Vale, R.D.: *Nature* **380**, 555–559 (1996)
58. Rayment, I., Rypniewski, W.R., Schmidt-Base, K., Smith, R., Tomchick, D.R., Benning, M.M., Winkelmann, D.A., Wesenberg, G., Holden, H.M.: *Science* **261**, 50–58 (1993)
59. Sablin, E.P., Fletterick, R.J.: *Curr. Op. Struct. Biol.* **11**, 716–724 (2001)
60. Houdusse, A., Szent-Györgyi, A.G., Cohen, C.: *Proc. Natl. Acad. Sci.* **97**, 11238–11243 (2000)
61. Menetrey, J., Bahloul, A., Wells, A.L., Yengo, C.M., Morris, C.A., Sweeney, H.L., Houdusse, A.: *Nature* **435**, 779–785 (2005)
62. Arnal, I., Metoz, F., Debonis, S., Wade, R.H.: *Curr. Biol.* **6**, 1265–1270 (1996)
63. Henningsen, U., Schliwa, M.: *Nature* **389**, 93–95 (1997)
64. Endow, S.A., Waligora, K.W.: *Science* **281**, 1200–1202 (1998)
65. Wade, R.H., Kozielski, F.: *Nat. Struct. Biol.* **7**, 456–460 (2000)
66. Abrahams, J.P., Leslie, A.G., Lutter, R., Walker, J.E.: *Nature* **370**, 621–628 (1994)
67. Boyer, P.D.: *Biochim. Biophys. Acta.* **1140**, 215–250 (1993)
68. Noji, H., Yasuda, R., Yoshida, M., Kinosita, K., Jr: *Nature* **386**, 299–302 (1997)
69. Adachi, K., Yasuda, R., Noji, H., Itoh, H., Harada, Y., Yoshida, M., Kinosita, K., Jr: *Proc. Natl. Acad. Sci. USA* **97**, 7243–7247 (2000)
70. Yasuda, R., Noji, H., Yoshida, M., Kinosita, K., Jr, Itoh, H.: *Nature* **410**, 898–904 (2001)
71. Nishizaka, T., Oiwa, K., Noji, H., Kimura, S., Muneyuki, E., Yoshida, M., Kinosita, K., Jr: *Nat. Struct. Mol. Biol.* **11**, 142–148 (2004)
72. Steigmiller, S., Zimmermann, B., Diez, M., Börsch, M., Gräber, P.: *Bioelectrochemistry* **63**, 79–85 (2004)
73. Zhang, Y., Wang, J., Cui, Y., Yue, J., Fang, X.: *Biochem. Biophys. Res. Commun.* **331**, 370–374 (2005)
74. Itoh, H., Takahashi, A., Adachi, K., Noji, H., Yasuda, R., Yoshida, M., Kinosita, K. Jr: *Nature* **427**, 465–468 (2004)
75. Junge, W., Lill, H., Engelbrecht, S.: *Trends Biochem. Sci.* **22**, 420–423 (1997)
76. Stock, D., Leslie, A.G., Walker, J.E.: *Science* **286**, 1700–1705 (1999)
77. Seelert, H., Poetsch, A., Dencher, N.A., Engel, A., Stahlberg, H., Müller, D.J.: *Nature* **405**, 418–429 (2000)
78. Jiang, W., Hermolin, J., Fillingame, R.H.: *Proc. Natl Acad. Sci. USA* **98**, 4966–4971 (2001)
79. Stahlberg, H., Müller, D.J., Suda, K., Fotiadis, D., Engel, A., Meier, T., Matthey, U., Dimroth, P.: *EMBO Rep.* **2**, 229–233 (2001)
80. Murata, T., Yamato, I., Kakinuma, Y., Leslie, A.G., Walker, J.E.: *Science* **308**, 654–659 (2005)
81. Feniouk, B., Junge, W.: *FEBS Lett.* **579**, 5114–5118 (2005)
82. Corvest, V., Sigalat, C., Venard, R., Falson, P., Mueller, D.M., Haraux, F.: *J. Biol. Chem.* **280**, 9927–9936 (2005)

Aptamer Selection by Darwinian Evolution

F. Chauveau, C. Pestourie, F. Ducongé, and B. Tavitian

*Nothing in biology makes sense except in the
light of evolution*

Theodosius Dobzhansky (1900–1975) [1]

Any technical object is defined by a structure and certain related properties, together with a function and a way of making it. A biological object is thus a biochemical structure, either organic or organically based, which possesses one or more biological properties (recognition, structure, transformation, etc.), which carries out a specific function, and which is produced by a biological process. A biological nano-object is a biological object with nanometric dimensions, from which we understand that it is a macromolecule or an assembly of such (diameter of a hemoglobin molecule 5.5 nm).

By learning to understand and manipulate the enzymes that produce these macromolecules, biotechnology can today create or sculpt biological nano-objects using fabrication processes that closely resemble natural mechanisms of synthesis, but which do not require the presence of a living being. Although these activities are recent and still somewhat limited, our mastery of the living tool box has already produced some entities with industrial prospects, including some artificial nano-objects with quite remarkable properties, unknown in nature.

Among the biological macromolecules, the nucleic acids play a central role because they define both the species and the individual and provide the chemical support for heredity. They are also the only biological molecules we are able to reproduce identically by a simple and well understood enzyme mechanism, viz., the polymerase chain reaction (PCR) (see Chap. 15), which lends itself particularly well to mass production. The nucleic acids feature amongst the most widely used compounds in biology at the current time, e.g., as probes, amplification initiator, etc., as attested by the present market for oligonucleotides (short sequences of nucleic acids): 340 million dollars in 2003, with a predicted 776 million dollars in 2010. However, the use of nucleic acids

is generally based on the canonical Watson–Crick pairing of nuclear bases, whose sequence encodes genetic information, while their wealth of structural potential remains virtually unexploited. In contrast, natural evolution has selected many RNA for their catalytic activities or for their ability to interact with proteins or other classes of molecules.

It was on the basis of these two properties of nucleic acids, the ability of a mother sequence to generate descendents and the ability to interact with (or act on) other molecules, that the aptamer and the SELEX technology were invented in 1990. The word ‘aptamer’ is a neologism coming from the Latin *aptus*, meaning ‘apt’ or ‘appropriate for’, and the suffix ‘-mer’, indicating the basic component of a polymer, to refer to a polymer that has been adapted to a specific function.¹ More precisely, ‘aptamer’ designates a nucleic acid structure resulting from the evolution of an ancestral population by selection, at each successive generation, of those structures best suited to the given function. We recognise here the basic tenets of the Darwinian theory of evolution: generation of populations of individuals and selection of the most apt individuals.

Since its first appearance 15 years ago, the SELEX methodology has proved remarkably fertile, both conceptually and in terms of the production, currently at the industrialisation stage, of aptamers with medical or technological applications. On the conceptual level, aptamers can be considered to constitute one of the rare, if not the only direct approach validating the theory presented by Darwin 150 years ago [2]. As far as applications are concerned, the time required for the development of aptamers from their invention in the laboratory to their becoming commercially available, i.e., 13 years, is of the same order as would be required for the development of a ‘standard’ medicine, which suggests that we may be right at the beginning of these applications. For these two reasons, aptamers are exemplary biotechnological nano-objects.

6.1 Some Theoretical Aspects of Molecular Evolution

6.1.1 Darwin and the Theory of Evolution

In 1859, Charles Darwin published *On the Origin of Species by Means of Natural Selection or the Preservation of Favoured Races in the Struggle for Life* [2,3]. This book, the result of more than 20 years of reflection, has been one of the great successes of scientific publishing: the first edition sold out in one day.

Darwin’s theory was built up on the basis of a huge corpus of observations of living species, both animal and plant, wild and domesticated. Ernst Mayr [4] sums the theory up by what he considers to be its five main features, three

¹ Since most aptamers are biological ligands, it is interesting to note that Webster’s dictionary gives the literal meaning of *aptus* to be ‘fastened’.

concerning population ecology and two concerning genetics. From these five features, he infers three conclusions:

- Observation 1: The reproductive potential of each species is so great that it should lead to exponential population growth.
- Observation 2: This does not happen since populations are roughly stable.
- Observation 3: Natural resources are limited.
 - Inference 1: Only that fraction of individuals survive that have access to these resources. This is the *struggle for survival*.
- Observation 4: There is variation among the individuals making up any given population.
- Observation 5: Many of these variations are inherited.
 - Inference 2: In this struggle, survival depends (in part) on the hereditary constitution of each individual. This is the process of *natural selection*.
 - Inference 3: Over successive generations, this process engenders a gradual change in the population. This evolutionary process can lead to the appearance of a new species. This is the *origin of species*.

A theory of the evolution of species, a phenomenon that generally requires more than the lifetime of an experimenter, is not easily accessible to the experimental method. However, Darwin's theory provides such a rich framework for reflection that biologists immediately set about confronting it with their results, as Dobzhansky pointed out. It is remarkable that this theory should have stood up so well (except possibly for the gradual aspect of the changes that underlie the origin of species) to the many tests it has been subjected to, with the subsequent discovery of genetics and molecular biology. Indeed, the notion of the gene fits in perfectly with Darwin's idea of variation, as does Mendelian segregation and the inheritance of characters. The discovery of the role of nucleic acids in the inherited transmission of characters, and the determination of the double-helix structure of DNA, almost immediately provided a virtually perfect biochemical explanation for the transmission of inherited characters.

6.1.2 Molecular Evolution and Properties of Nucleic Acids

In a quite remarkable way, a strict application of the principles of Darwinian theory to populations of molecules *in vitro* does indeed lead to an evolution by selection of the most apt individuals. This is the basic principle underlying the production of aptamers. They can thus be considered as new species whose survival has been favoured, not in this case by access to resources in the natural environment, but by the survival criterion imposed by the experimenter (selection by directed evolution). This result follows from the perfect applicability of the observations stated by Darwin to the fundamental properties of nucleic acids.

Size and Diversity of Populations

Since there are four different bases making up DNA or RNA, the number of different sequences that can be obtained by random combinations of these four bases leads to very large populations of sequences as soon as the number of bases in the sequence is greater than ten. Hence, the set of possible sequences comprising a chain of

- 10 bases is 4^{10} , somewhat more than one million individuals,
- 100 bases is 4^{100} , somewhat more than 10^{60} individuals.

To be more concrete, a population containing a single representative of each possible sequence of an 80-base DNA would represent a mass greater than the mass of the Earth.

It should be noted that the diversity of sequences is not directly correlated with the diversity of structures obtained. As a general rule, although it is difficult to appreciate, the diversity of structures is lesser. It is nevertheless easy to obtain, by chemical synthesis, enough different structures to carry out efficient selection from an initial population of 10^{14} to 10^{15} sequences.

Stability of Population Size and Limitation of Resources

These principles are directly imposed by the experimenter, who deliberately limits the size of the population generated during the production of a new generation issuing from a fraction of the initial population (see the description of the method below). The limitation of resources is in fact the filter used in the stage where the sequences to be conserved in the next generation are selected. The size of the population of the following generation is determined by the conditions of fabrication of this population (PCR).

Diversity and Heritability

In contrast to what happens in the natural environment, the diversity here is mainly contained in the first population. However, a certain level of diversity is introduced at each production of a new population, due to sequence copying errors (replacement of one base by another, deletions, etc.) resulting from imperfections in the enzyme activity.

Of course, the daughter sequences produced by PCR do on the whole remain very close to the parent sequences, and there is therefore hereditary transmission at each generation.

In this chapter, we first discuss the underlying theory of the aptamer concept, and in particular the structural properties of the nucleic acids, which constitute the building blocks of the aptamers. We then describe the SELEX methodology, before tackling the properties of the aptamers themselves and their applications.

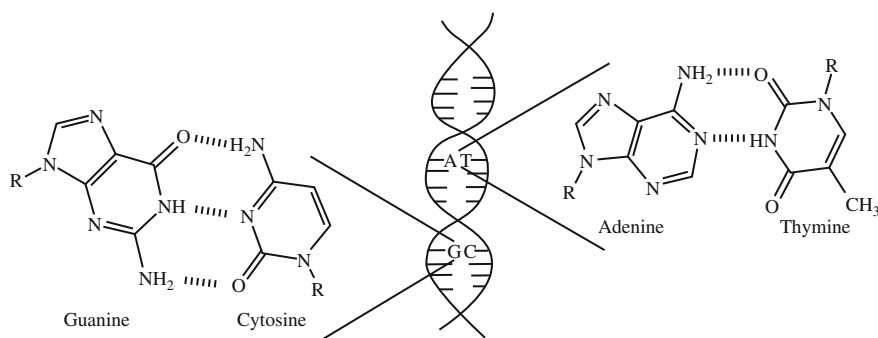


Fig. 6.1. Base pairing in a double strand of DNA according to Crick and Watson

6.2 Structural Features of Nucleic Acids

Nucleic acids are polymers of nucleotides, each comprising a ribose, a phosphate group, and a heterocyclic base of purine type, viz., adenine (A) or guanine (G), or pyrimidine type, viz., thymine/uracil (T/U) or cytosine (C). This chemical repertory, although limited to just four monomers, can provide a truly impressive structural diversity, which explains the interaction of the oligonucleotides with a large number of partners or their catalytic activities.

6.2.1 General Considerations: The Double Helix

Deoxyribose nucleic acid (DNA) like ribose nucleic acid (RNA) can form a double helix, by the canonical base pairing as established by Crick and Watson (A binds to T/U via two hydrogen bonds, and G binds to C via three hydrogen bonds) between two anti-parallel strands (see Fig. 6.1). The helix is stabilised by the interaction of π electrons in the stacked bases of each strand, forcing the ribose–phosphate backbone to the outside. DNA usually forms a type B helix, with a C-2' endo or S(outh) conformation and slightly tilted bases, whereas RNA, owing to steric hindrance induced by the hydroxyl group at the 2' position on the ribose, which imposes a C-3' endo or N(orth) conformation on it, systematically adopts a type A helical form, in which the bases are highly tilted with respect to the axis of the helix. As a consequence, the minor groove of DNA is narrow and deep, whereas the major groove is broad and shallow, and interacts strongly with surrounding water molecules. The opposite is true for RNA, which is, even so, less hydrated globally.

The double helix is a structure formed from single-strand nucleic acids by the pairing of many complementary regions (in the sense of Crick and Watson). By analogy with the terms used to describe protein structures, the helix is thus said to be the secondary structure of oligonucleotides, while tertiary structure refers to the supercoiled arrangement of the molecule. Within the helices, specific interaction sites can be formed by interaction in or out of the

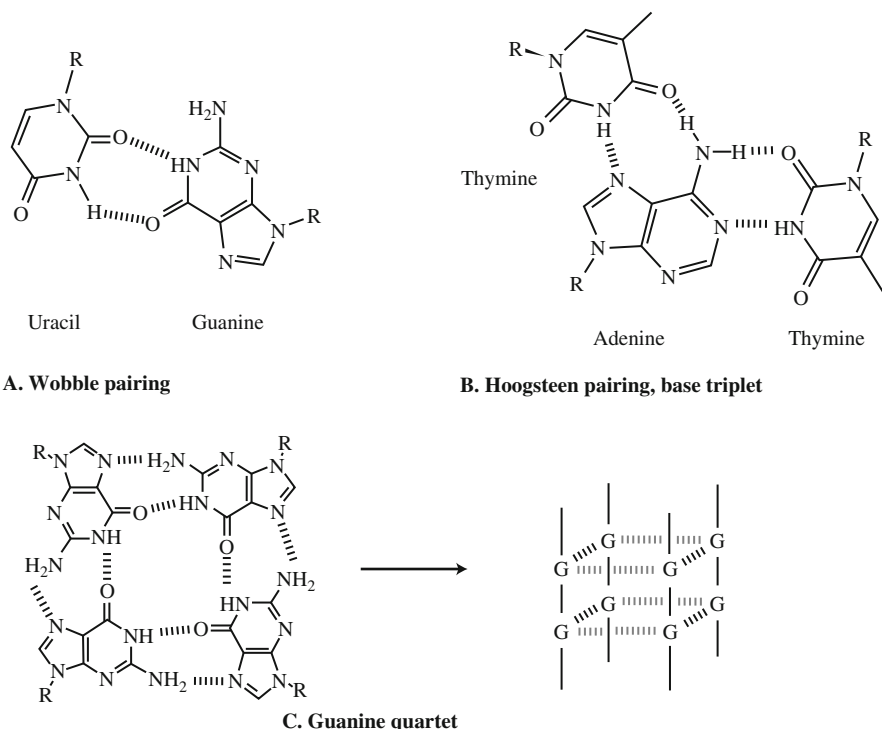


Fig. 6.2. Non-canonical base pairings and triplet/tetrad formation

plane of the heterocyclic bases [5]. However, the possibilities for interactions of this type are difficult to access, because the bases are all turned toward the interior of the helix. The grooves represent more favourable interaction areas. Hence, many transcription factors (so-called zinc finger proteins) bind to genomic DNA via the major groove.

6.2.2 Intrahelical Interaction Sites

In the plane, many base pairings other than Watson–Crick pairings have been described (see Fig. 6.2). Wobble G–U association is quite common, while others pairings require a reorientation of the bases, or indeed their enolisation or an ionisation. Such non-canonical pairings make some functional groups of the bases more accessible through a local distortion of the helical structure, thereby creating recognition sites. Hoogsteen pairings explain the formation of a triplet of bases, or even a triple helix. The association of a third strand in one of the grooves of a double strand causes the opposite groove to broaden. The association of four bases in the same plane by means of hydrogen bonds remains exceptional, apart from the case of the G tetrads (typically found in a telomer structure), where two to three planes of four associated Gs

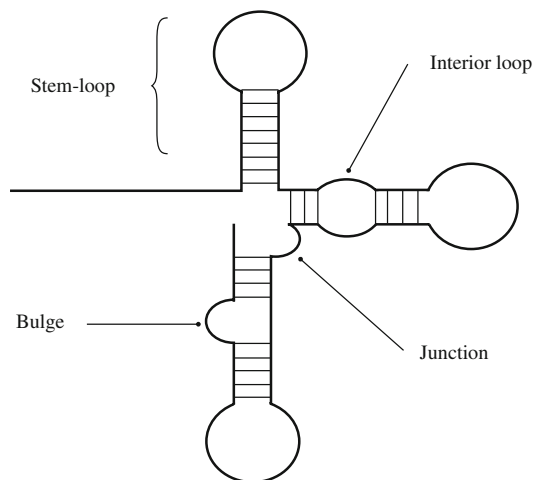


Fig. 6.3. Main structures adopted by nucleic acids

are superposed. Finally, platforms of interaction with aromatic molecules are made possible by association of two consecutive bases on the same strand.

Out of the plane, cross-strand base stacking (it is mainly intra-strand in a regular helix), referred to as a base zipper motif, can stabilise the nucleic acid structure. In the same way, one speaks of a ribose zipper motif in the case of compact RNA, where riboses interact by hydrogen bonds between their 2'-OH groups. This happens in particular in the stacking of adjacent helices of ribozyme (RNA with catalytic activity), the group 1 intron of *Tetrahymena* [6]. It contributes to the supercoil organisation of oligonucleotides by stabilising the electrostatically unfavourable juxtaposition of the helices, thereby allowing a compact 3D architecture for the molecule.

6.2.3 From Secondary to Tertiary Structure: Supercoiling

Regions of unpaired bases are all-important. They adopt a wide range of different conformations, e.g., bulges, turns, loops, pseudo-knots, etc., as shown in Fig. 6.3 [5].

One or more unpaired bases may be located within a helix, curving its orientation so as to form a protrusion or bulge. They are either oriented toward the interior of the helix and stabilised by stacking, or pushed toward the outside where they provide a privileged site for interaction due to their exposed position. This type of structure is present, for example, in the TAR RNA structure of the human immunodeficiency virus (HIV) and serves as anchoring site for the viral protein Tat which is a powerful activator for HIV gene expression [7].

The connection between differently oriented helices occurs via a junction called a turn. Different families such as the C-turn or the U-turn have been

identified, depending on the nucleotide at which the changed orientation of the strand takes place, each one being characterised by a distinct stabilising interaction. For example, a conserved motif UNR (N is any nucleotide, R is a purine) identifies a U-turn in an RNA.

When a strand folds over on itself to form a helix, this gives rise to a hairpin or stem-loop structure. The size of the unpaired region (the loop) can vary, but the conformations of the consensus sequences GNRA or UNCG are made particularly stable and rigid by wobble pairing of the first and last bases, a high level of stacking, and a network of hydrogen bonds. A loop can be internal to the helix and provide important interaction sites there too. Very close-packed pseudo-knot structures can also form by interaction between a stem-loop and a single-strand portion. In this case, there is coaxial stacking of two stalks whose loops cross a groove or span the ribose-phosphate backbone.

Stacking of helices and close-packing of the result are also the consequence of interactions between loops: loops embrace to form so-called kissing complexes by pairing of complementary residues from each, or a loop may insert itself in a helix or within a complementary receptor structure in the process known as docking. In particular, this type of structure has been identified by crystallographic study of the group 1 intron of *Tetrahymena*, which revealed the insertion of a tetraloop in an interior loop.

6.2.4 Role of Cations and Water Molecules

This high level of complexity and the astonishing variety of different structures are only possible due to the presence of cations and water molecules. Although most of the ions are delocalised within the structure, guaranteeing only the electrical neutrality of the system as a whole, there are specific coordination sites defined by local arrangement, especially tertiary, of the oligonucleotide. For example, Mg^{2+} and Ca^{2+} stabilise the stacking of helices associated with loop-loop interactions, and the presence of K^+ and Na^+ is crucial for the formation of G tetrads. Like ions, water molecules occupy electronegative cavities in the structure and sometimes form strongly bound ordered lattices, which are not necessarily extra-helical. Hydrogen bonds are thus observed between water molecules and the polar groups of unpaired bases, or bases that are not canonically paired. Likewise, the characteristic L structure of transfer RNA is stabilised by specific binding sites of metal ions, and by water molecules completing their coordination sphere. Certain cations and water molecules are thus thoroughly integrated into the final overall structure.

6.2.5 Binding of an Aptamer to Its Target: Examples of Resolved Structures

Aptamers possess the ‘natural’ structural variety of oligonucleotides for the recognition of a chemically diverse range of molecules. Whether it be a small molecule, an oligosaccharide, a peptide, or a protein, the binding of an aptamer

to its target must be considered a dynamical phenomenon, requiring some degree of rearrangement of a pre-existing structure, or indeed the adoption of a new structure integrating the target [8].

The binding to a small aromatic molecule is much more elaborate than a simple intercalation in a double-strand nucleic acid: the often cited discrimination of an RNA aptamer between theophylline and caffeine, which differ only by a methyl group, results from the establishment of a hydrogen bond with an unpaired C. The ligand is so oriented that caffeine cannot establish this bond. In a quite astonishing way, an RNA aptamer and a DNA aptamer, although differing structurally in every way possible, exhibit an identical recognition strategy by their common ligand, adenosine monophosphate (AMP): this is paired in a specific way to a G, in interaction with a non-canonical G–G pair. Likewise, RNA and DNA aptamers bind to arginine via exclusive interaction of the positive guanidinium group with heterocyclic bases, and not with a phosphate group, as is often the case for proteins binding to cell DNA. The p50 unit of the transcription factor NF- κ B thus binds a hairpin RNA aptamer with an interior loop which reproduces the local structure of naturally recognised DNA [9].

The study of aptamers recognising peptides is also revealing, since a peptide derived from the Rev protein of the human immunodeficiency virus (HIV-1), unstructured in solution, forms an α helix accommodating the major groove of the RNA aptamer, broadened by non-canonical pairings and a base triplet. An example of interaction with the minor groove of RNA aptamers is provided by the oligosaccharide antibiotics neomycin and tobramycin, genuinely encapsulated by virtue of a loop in which one of the residues closes off access to the groove. In fact, on the basis of the structures of complexes so far established (by X-ray diffraction or nuclear magnetic resonance), a fundamental principle of specific recognition by an aptamer seems to be the inclusion of the ligand within the oligonucleotide architecture: the ligand is clearly an intrinsic element of the overall structure of the complex.

6.3 SELEX

6.3.1 History

The nucleic acids come with a broad spectrum of interactions to interact specifically with a great many partners like proteins or to carry out some catalytic activity. However, it is impossible to predict their structures from the primary chain of monomers making them up, and likewise for the properties they may eventually prove to have. On the other hand, the strategies of combinative synthesis can be used to select nucleic acid structures according to some given criterion.

In 1987, Struhl and coworkers were the first to combine a random synthesis of oligonucleotides with in vitro selection [10, 11]. They synthesised double-stranded sequences containing 23 nucleotides, and the sequences retained in a

column of GCN4 (transcription factor) were cloned and amplified in vivo using bacteria. After four rounds of selection and amplification, they thus identified specific double-stranded sequences of the transcription factor GCN4. At this point, Struhl and coworkers suggested that the method could be improved by PCR amplification of selected sequences [11]. A few months later, Kinzler and Vogelstein were the first to combine selection in vitro and amplification in vitro by PCR to identify specific sequences of the factor TFIIA [12].

However, they used fragments of human DNA, rather than randomly synthesised sequences, as starting population. In fact, the full association of random oligonucleotide synthesis, followed by selection and amplification in vitro, was developed simultaneously by several research centres in 1990 [13–20].

Different names have been given to this selection principle: in vitro genetics [16], SELEX (systematic evolution of ligands by exponential enrichment) [15], or directed molecular evolution [21]. The sequences resulting from this selection process were christened aptamers by Ellington and Szostak [13]. This term derives from the Latin *aptus* meaning ‘to be apt or suitable for’.

6.3.2 General Selection Principle

The underlying principles of the in vitro selection–evolution technique for nucleic acids have been described on many occasions in the literature [22–24]. The basic idea is always the same, as illustrated in Fig. 6.4. A population of oligonucleotides called candidates is produced by chemical synthesis and includes a random sequence obtained by setting up conditions which allow one to introduce, with the same probability, an adenine, a thymine, a guanine, or a cytosine in each position. Hence, for a sequence of n nucleotides, there are 4^n possibilities. This random sequence is framed on either side by two constant sequences that are essential to the different enzymatic steps making up the selection process.

This random population (generally 10^{13} – 10^{15} different sequences) is then subjected to selection pressure. The criterion for selection may be a certain catalytic activity, or an affinity for a given target. Candidates that pass successfully through the selection process are separated from the others, then amplified by PCR to be reinjected into another selection cycle. This technique can use RNA, although a reverse-transcription step must be included before the complementary DNA thereby obtained can be amplified by PCR. The amplified DNA is subsequently transcribed in vitro to be used for a new round of selection. Most bacterial polymerases used in the amplification step are known to generate naturally a high rate of mutation, i.e., about one mutation for every 2×10^4 nucleotides incorporated). Over the last few years, efforts have been made to reduce this mutation rate by modifying these polymerases to make them more faithful. However, although such optimisation is crucial for applications such as the diagnosis of pathologies like AIDS, hepatitis, and so on, by PCR, they are of course detrimental for this type of selection, where the mutations generated by the polymerases play a key role in the Darwinian

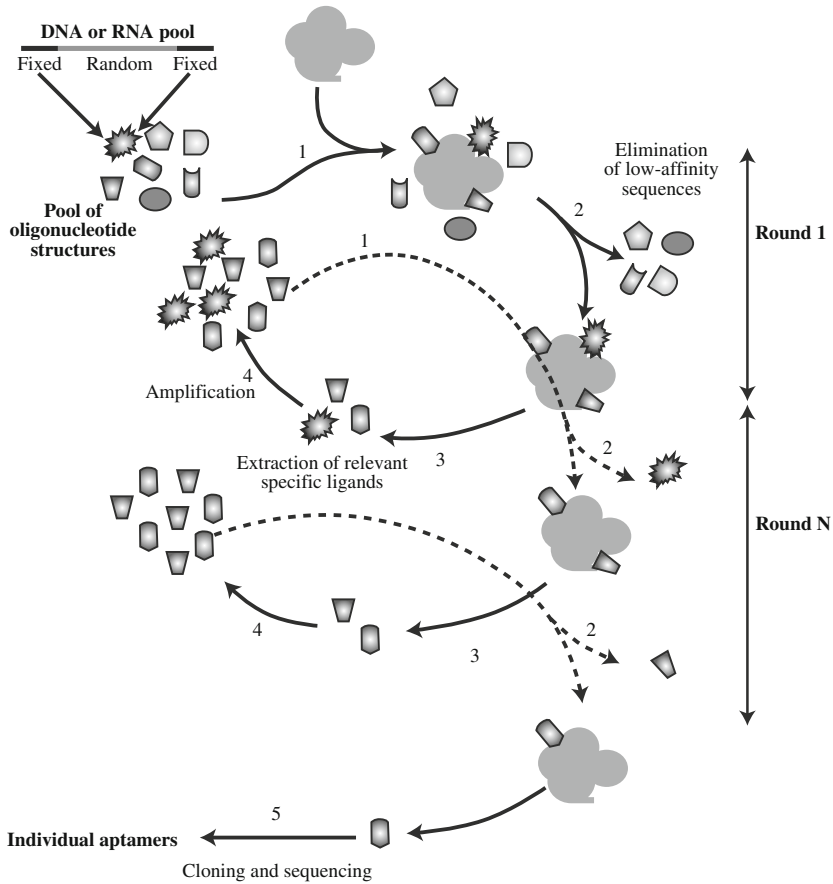


Fig. 6.4. Producing aptamers by systematic evolution of ligands by exponential enrichment (SELEX). A random population of candidate oligonucleotides is (1) synthesised, (2) incubated with the target. Sequences satisfying the selection criterion are conserved, and the others are eliminated. (3) Winning sequences are taken out and (4) amplified by PCR (or reverse transcription PCR followed by in vitro transcription when dealing with RNA pools). The selected population can then go into a new selection cycle. During the different cycles, the population evolves toward sequences that best resist the selection pressure. (5) The aptamers obtained are cloned and sequenced, then assessed for aptitude with regard to the chosen selection criterion

evolution of the sequence population. Mutations engender diversity among the selected aptamers, and this makes it possible to create better aptamers that were not present in the original pool. Some even enhance this in vitro evolutionary phenomenon by carrying out amplifications under conditions in which the fidelity of the polymerase is reduced. It is thus essential not to choose the most faithful polymerases when setting up such a selection process.

Various selection methods have been employed. Methods used to separate off sequences satisfying the selection criterion from those that do not include membrane filtration, immunoprecipitation, affinity chromatography, electrophoretic separation by the technique known as gel retardation, and magnetic beads. Regarding the selection of aptamers in terms of their catalytic activity, different separation methods are used. For example, the catalytic aptamer may be captured in a column, exploiting some covalent bond that it will itself form.

As the cycles go by, the population evolves in a Darwinian way toward sequences exhibiting the best qualities for the required function (see Fig. 6.4). Once the population has been enriched in aptamers, part of it is cloned and sequenced. By comparative analysis, sequences can often be grouped into several classes, in which the covariation and conservation of certain bases can be exploited to deduce the critical structural motifs. In most cases, consensus structural motifs can be identified. However, this type of approach assumes a unique relation between the sequence and the structure of an aptamer, and this neglects the structural pleiomorphism of oligonucleotides, i.e., their ability to adopt several conformations in equilibrium (both intramolecular and intermolecular). A technique for minimising this phenomenon, which reduces selection efficiency, consists in administering heat shock, i.e., sudden cooling after high-temperature denaturation.

The repetitive nature and the complexity of some steps in the selection procedure, which make it difficult for a single person to carry out several selections manually in parallel, have encouraged various teams to automate the selection process. Since 1998, several automated *in vitro* selection processes for high-rate selection of aptamers have been described [25], and several biotechnology companies using aptamers have been set up [26].

Although all SELEX selection methods are based on the same principle, it is important to stress that some parameters are likely to affect the final result:

- Selection pressure, e.g., washing conditions becoming more stringent during selection, changes in ion concentration, a reduction in incubation time or in the candidate and/or target concentration, or even a change in the selection system, or addition of non-amplifiable competitors.
- Elimination of artefacts such as sequences selected for their ability to bind on a filter or chromatographic surface, by including suitable counterselection steps. Hence, candidates must first be screened on a pre-chromatograph without target, and only those with no affinity for the chromatograph surface are then used for selection. Other authors also use a specific elution by virtue of the free target. Another type of specific elution consists in unlocking the aptamers from their target by adding the natural ligand of the relevant target to the reaction medium. In the end, the best solution for avoiding artefacts may be to alternate different selection methods.

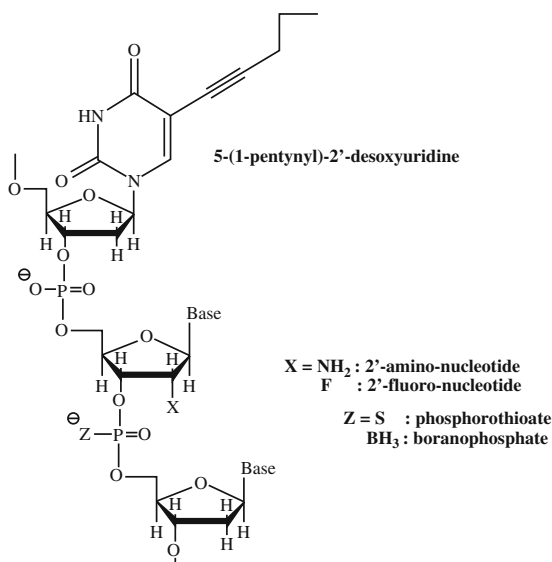


Fig. 6.5. Modified nucleotides for use during selection

6.3.3 Chemical Modifications

The chemical modification of natural nucleotides has proved to be a choice tool for overcoming the main obstacle to the *in vivo* use of aptamers: degradation by nucleases. Some chemical modifications of the ribose at 2' increase the stability of oligonucleotides and are compatible with the use of polymerases, a necessary condition for SELEX (see Fig. 6.5). The 2'-fluoro and 2'-amino nucleotides, and more recently 2'-OMe [27], can be used right at the beginning of the selection process to constitute a pool of modified oligonucleotides. Aptamers containing 2'-aminopyrimidines or 2'-fluoropyrimidines have been selected against the vascular endothelial growth factor (VEGF) [28] and the keratinocyte growth factor [29]. The aptamers identified in each case were different depending on the chemistry used. An alternative approach to replacing the hydroxyl at 2' on the ribose is to modify the backbone of the nucleic acids [30]. However, the chemical groups used are rarely integrated by the enzymes required for SELEX, with the notable exception of the phosphorothioate nucleotides and boranophosphate nucleotides, where a non-bridging oxygen of the phosphodiester group is replaced by a sulfur and a borane (BH₃) group, respectively. Finally, more drastic modifications to the phosphodiester (formacetal) linkage have sometimes been used to study aptamers. Modifications affecting the bases remain rare, although a successful selection using 5-(1-pentynyl)-2'-deoxyuridines has been reported [31].

All these nucleotides constitute steric or electronic variations on the natural motif and this is not without consequence for the structure of the resulting oligonucleotides. Their introduction into the sequence of an aptamer after it

has been selected, although it contributes to overall chemical diversification, must be carefully controlled. Experience acquired so far shows that it is difficult to predict the consequences of chemical modifications to aptamers introduced after selection, whence the construction of chimeric oligonucleotides, associating blocks of different chemical nature in order to profit from greater stability without affecting their ability to bind to their target. Another strategy for obtaining nuclease-resistant aptamers exploits the chirality of living molecules: if an aptamer is selected against a non-natural enantiomer of a target, then the enantiomer of this aptamer (L-DNA or L-RNA, or Spiegelmer) recognises the natural target, without being a substrate of the nucleases, specific to the riboses of the D enantiomeric series [32].

6.4 Applications

6.4.1 Aptamers as Research Tools

Study of Nucleic Acid–Protein Interactions

Interactions between nucleic acids and proteins often have important biological roles. Hence, gene expression is temporally and spatially regulated by the interaction between specific nucleic acid sequences and regulatory proteins. In vitro selections have often been used as tools to identify these interactions. They have been used to determine the specific double-strand DNA sequences for many transcription factors and to identify RNA structural motifs or sequences important for binding many proteins, such as the factors required for splicing, transcription termination, polyadenylation, or the elongation of translation [23].

Analysing the results produced by these investigations, several principles can be identified regarding in vitro selection. To begin with, many of the aptamers identified have turned out to be far-removed from the motif naturally targeted by these proteins. In some cases, these aptamers have a stronger affinity for the protein than the original RNA sequences. Hence, in vitro evolution can lead to selection of motifs with a greater affinity for a target than those selected in vivo by natural evolution. Surprisingly, this turns out to be generally true for in vitro selection against proteins binding to nucleic acids, although this result is surprising only in a superficial way. In fact, natural selection has optimised protein–nucleic acid interactions for biological functions involving more parameters than an in vitro selection. For example, the reversibility of the interaction is one of these parameters, as is the possibility of interaction with more than one partner. Aptamers with a high affinity for their target do not necessarily possess these properties, which may well explain why they have not been selected by nature. Finally, for the sake of reversibility, nature may have had no need to generate very tightly binding sequences.

Study of Nucleic Acids as Catalysts

The in vitro selection technique has been used to study the activity of natural catalytic RNA (ribozymes) [33]. Many in vitro selections have been carried out starting with pools of known ribozyme sequences, either partially degenerate or with completely random portions. In this way, important structural motifs or sequences for these ribozymes were quickly identified.

In vitro selection has also been used to evolve ribozymes with novel catalytic features. For instance, in vitro selection has been able to identify variants of the group I ribozyme of *Tetrahymena* which cut RNA in the presence of Ca^{2+} , whereas this ribozyme normally only functions in the presence of Mg^{2+} or Mn^{2+} [34]. Variants of this same group I ribozyme have also been identified for cutting DNA, and even for discriminating a DNA substrate from an RNA substrate [35].

Apart from identifying variants of known ribozymes, completely new activities and structures have been found by these selections. Hence, starting with a population comprising random sequences of 220 nucleotides, it has been possible to isolate ribozymes catalysing the ligation of two RNA strands [36]. Three classes of ribozymes were identified. The first catalysed the formation of a phosphodiester 3'-5' bond, while the other two catalysed a 2'-5' bond. By creating a random sequence pool derived from the first class of aptamers, Ekland and Bartel selected a ribozyme which catalyses the polymerisation of triphosphate nucleotides directed by a matrix [37]. The fact that this ribozyme was produced in this way provides a strong argument in favour of the theory of evolution of life from an RNA world, because it proves that a system comprising only RNA would have the ability to self-reproduce. Even if this type of result might encourage us to try to create self-replicating nanomachines, it should be stressed that this ribozyme is extremely slow and exhibits only low processivity (several hours to incorporate only ten nucleotides).

This work also shows that it is possible to evolve ribozymes toward more and more complex functions. Active aptamers have been selected to synthesise a nucleotide, or to form a peptide bond or a bond between two carbons. In vitro selection has also been able to isolate DNA with a catalytic activity (DNAzyme).

6.4.2 Aptamers as Purification Tools

In order to isolate molecules such as proteins, oligonucleotides, and so on, from a complex mixture by affinity chromatography, a biotinylated aptamer bound to streptavidin-coated beads can be used as the stationary phase [38]. The aptamer directed against L-selectin, for instance, has been successfully used as stationary phase [39]. Affinity chromatography has also been used to separate enantiomers by high-performance liquid chromatography (HPLC). The aptamer bound to the column was specific to the D-enantiomer of an oligopeptide and had no significant affinity for the L-enantiomer [40]. This

stereospecificity could be used to purify certain pharmaceutical molecules whose active form depends on an enantiomeric form.

Aptamers have also been used in a purification method by a system of protein chips called surface-enhanced laser desorption/ionization mass spectrometry (SELDI-MS). This purification system is often used upstream of mass spectrometry analysis to concentrate the relevant proteins. Aptamers have several advantages over other substrates associated with this type of chip, e.g., antibodies, metals, receptors, etc. They are easy to synthesise and manipulate, and their stability means that the chip can be regenerated. In addition, their small size allows a high surface coverage. This method of purification and preconcentration of proteins is highly promising and its applications could be extended to other systems [41].

Aptamer selection can also be used to purify and identify a priori unknown therapeutic targets. Hence, the selection of aptamers against microglial cells led to the identification of a new angiogenesis marker [42], and selection against glioblastom cells led to the emergence of aptamers directed against a protein of the extracellular matrix [43]. In both cases, the aptamers displayed a high affinity and good specificity for the cell type in question, and were used to isolate their target protein by affinity chromatography before analysis by mass spectrometry [42, 43].

6.4.3 Aptamers as Detection Tools

Aptamers have served as an alternative to antibodies in many detection techniques. Hence, aptamers have been used in approaches similar to the standard sandwich-type ELISA technique. Likewise, aptamers have proved to be as effective as antibodies when used as detection agents for capillary electrophoresis and flow cytometry. By their very nature, aptamers have also provided a way of developing new detection methods, and their use as a biocomponent opens up many prospects for developing biosensors (component associating a biological species and permitting recognition of a target molecule, and a transducer able to transform the recognition event into a measurable physical signal).

Detection Method Using PCR

Unlike antibodies, aptamers can be enriched by PCR. Fredriksson has suggested a convincing strategy for ultrasensitive detection of low-abundance targets (in amounts less than the zeptomole, i.e., 10^{-21} mole) [44]. The protein is detected by joint action of two capture aptamers and a sequence used for detection, produced in this case by quantitative PCR. The simultaneous binding of two aptamers (one for each protein monomer) brings them in close proximity to one another, allowing a short oligonucleotide sequence to bridge the two aptamers. Use of a specific connector brings together the free ends of the two aptamers and enzymatic ligation occurs, thereby generating a DNA

sequence that can be amplified by quantitative PCR. This method informs about the identity of the detected protein (only the homodimer leads to amplification) and also its concentration. No washing stage is required here.

Another method uses the protection of the aptamer from exonuclease I in the presence of its target [45]. Exonucleases are enzymes responsible for the degradation of oligonucleotides from the 3' end to the 5' end. The binding of an aptamer to its target can protect it from the exonuclease and thereby provide an indirect assay of the amount of target, the idea being to quantify by PCR the portion of aptamer not degraded in the presence of exonucleases.

Optical Detection Method

The conformational rearrangement of the aptamer induced by binding to its target is fully exploited in optical detection methods. Jhaveri and coworkers have shown that it is possible to introduce a fluorophore precisely in these regions undergoing conformation changes, in such a way that the bond between the target and the receptor (the aptamer) alters the fluorescence intensity. The work they present relates the selection of fluorescent RNA aptamers directed against adenosine. The selected aptamers have been assessed for their affinity with ATP, but also for their ability to enhance the fluorescent signal in the presence of this molecule [46].

Another direct detection technique is based on the use of a beacon. Classically, this is a nucleotide sequence designed to lead to the formation of a stem-loop bringing the two ends of the sequence close together. One of the ends carries the fluorophore, whose fluorescent signal will be 'absorbed' by an acceptor molecule placed at the other end. In this way, there is no detectable emission of fluorescence. However, any deformation of the stem-loop structure which causes the fluorophore to move away from its acceptor molecule will generate a fluorescent signal. An antithrombin aptamer has thus been modified to form a double-strand domain which juxtaposes a fluorescent molecule and its signal acceptor at each end. In the presence of thrombin, the aptamer reassumes its G-tetrad native structure and undoes the double-strand structure, causing the emission of a fluorescent signal proportional to the amount of target present [47].

Development of Aptamer Chips

Several groups have suggested using aptamers for high-speed screening of proteins whose expression is linked to the occurrence of certain diseases, thus providing a fast and accurate means of diagnosis. Sensitivity and specificity are the two essential criteria justifying any diagnostic test. Indeed, very low detection limits are required, and the system must be able to discriminate between physiological and pathological forms of proteins that are sometimes very similar. Aptamers seem to be perfectly suited to this task. Gold and

coworkers and the company Somalogic in Boulder, Colorado, propose the use of photoaptamers as capture agents for biomarkers [48].

The idea is to replace the weak bonds of the ligand/target complex by photoinduced covalent bonds. This allows much more stringent washing in order to improve the signal-to-noise ratio. For this purpose, a photoreactive unit, e.g., 5-bromodeoxyuridine (BrdU), must be incorporated in the aptamer sequence. Illumination at 308 nm induces, by electron transfer, the formation of a covalent bond involving a reactive amino acid (Tyr, Trp, His, Phe, etc.). This type of covalent bond formation is not *a priori* specific. However, work by Smith [49] has shown that this photoinduction of a covalent bond significantly enhances the specificity of aptamers, all the more so as their initial specificity is low (the effect on those sequences having the highest initial specificity being almost negligible). The only proteins found at the chip surface are those bound to their specific aptamers, and those proteins can then be revealed by some universal staining system.

At the present time, the development of such techniques is in full swing, and the different methods for immobilising aptamers (beads, optical fibres, etc.), for labelling, and for detection constitute a new strategy for the development of detection platforms which should provide a highly effective substitute for the use of antibodies.

Use of Aptamers for in Vivo Molecular Imaging

The low molecular weight of aptamers (10–15 kDa compared with 150 kDa for antibodies), combined with their greater capacity for tissue penetration and faster blood clearance, make these oligonucleotides good candidates for imaging purposes. The first published study in this area aimed to visualise an inflammation by means of an anti-elastase aptamer. The image obtained displayed a better signal-to-noise ratio than the reference image obtained using antibodies [50]. At the present time, there are many studies on the *in vivo* pharmacokinetics of oligonucleotides. The development of robust methods for labelling oligonucleotides with fluorine 18, combined with imaging by positron emission tomography (PET), has opened the way to the development of aptamers as *in vivo* molecular imaging tools, whose tissue concentration can be monitored quantitatively inside various organs during the hours following injection [51].

6.4.4 Aptamers as Regulatory Tools

To study the function of a gene, and hence of a protein, it is often useful to deactivate it. This deactivation may be permanent, as in the case of the so-called knock-out technique, or regulated, by placing the relevant gene under the control of an inducible promoter. However, these techniques are hard to implement and require modification of the cell genome. Several alternative approaches have been reported, using aptamers to control gene expression artificially.

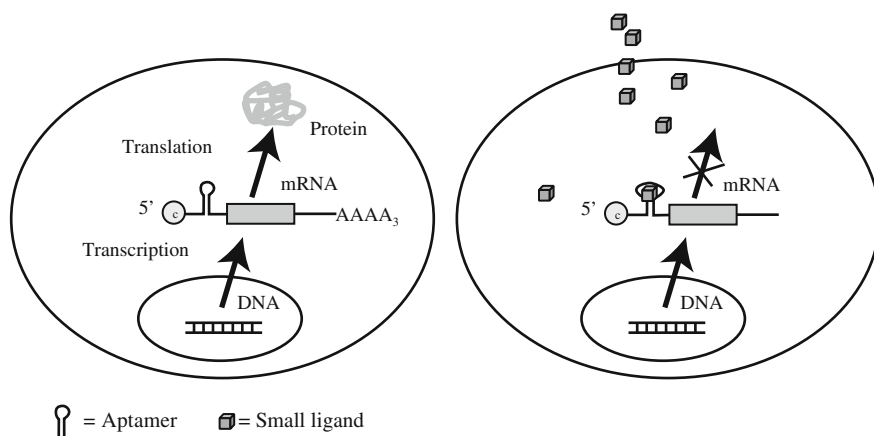


Fig. 6.6. Regulating gene expression via ligand–aptamer interaction

Controlling Genetic Expression

Several copies of aptamers can be introduced at the 5' of an mRNA. In 1998, Werstuck reported the *in vitro* control of translation by anti-tobramycin aptamers placed upstream of the mRNA to be translated [52]. The interaction of the antibiotic with the aptamer inhibits translation in a dose-dependent way (see Fig. 6.6).

Another strategy uses a ribozyme whose activity can be controlled by an aptamer. The active ribozyme can self-cleave, causing degradation of the mRNA and hence inhibition of the relevant gene expression. The interaction of the aptamer with its target leads to a conformational change such that the ribozyme is destructured and hence becomes inactive. The aptamer is then called a riboswitch. This feature is exploited to constitute a genuine ‘portable’ control system.

Inhibition of Protein Activity

Apart from their affinity, aptamers often exercise inhibiting properties over their targets. If these aptamers are expressed in a cell, their protein target can be inhibited *in cellulo*. These sequences are then called intramers.

In research on HIV-1, an aptamer directed against the protein Rev has been cloned in a dependent RNA polymerase III expression vector. This has been used to produce, in a cell model, a functional anti-Rev aptamer inhibiting the production of HIV-1 [53]. The protein Rev acts on the viral genome by binding to its recognition site (Rev binding element or RBE) and is involved in regulating the transport of viral RNA from the nucleus to the cytoplasm. Other examples concern intramers directed against endogenous nuclear proteins, like aptamers specific to RNA polymerase II, which do not interact

with RNA polymerases I and III. In yeast, the constitutive expression of this intramer placed under the dependence of the promoter of RNA polymerase III leads to a cell growth defect, which is not observed in the expression of a control sequence [54].

Shi and coworkers obtained the expression of intramers antagonistic to a regulatory protein for RNA splicing in a multicelled organism, viz., *Drosophila melanogaster*. They obtained a significant reversion of the phenotype induced by overexpression of the target protein by the regulated expression of an RNA aptamer in a transgenic animal [55].

Vuyisich and Beal have highlighted the main drawback of this method: once the aptamer is expressed in the cell and the target protein is inhibited, it is no longer possible to control the mechanism precisely in time. This problem is particularly relevant if the study concerns a protein involved in the cell cycle or early development. The solution is to include another step in the selection process, apart from the counterselection stage and the selection stage in the presence of the target protein. This new step concerns the dissociation of the aptamer and its target by adding a small molecule, in this case neomycin, known for its RNA binding properties [56]. The generalisation of this method may make it possible to inhibit a protein *in vivo* by means of a ligand-regulated RNA aptamer (LIRA) and to remove this inhibition at the desired moment by adding a small molecule, i.e., the ligand of the aptamer.

While *in vitro* selection of aptamers is relatively straightforward, the transition from the test tube to living beings is often more problematic. Concerning the use in living cells of aptamers designed to act on intracellular targets, one solution consists in expressing sequences directly within cells so that they can play their role of inhibitor. It is also interesting to obtain aptamers inhibiting extracellular proteins like membrane receptors. In this case, the target purification stage prior to selection causes significant structural modifications, since the protein loses its membrane interactions. Directing selections against targets expressed in native conditions at the cell surface provides a way of overcoming this crucial limitation. For example, aptamers have recently been selected on living cells against the Ret^{C634Y} receptor, a mutant with oncogenic tyrosine kinase activity. Among the aptamers obtained, one not only displayed an affinity for its target, but also proved able to cause a reversion of the mutant phenotype to the wild-type [57].

6.4.5 Aptamers as Therapeutic Tools

The main restriction on the use of oligonucleotides *in vivo* concerns their biodistribution. Various approaches are currently under development to get round this difficulty. However, in contrast to other therapies based on the use of oligonucleotides (antisense, interference RNA, etc.) acting on genes or messenger RNA, aptamers can carry out their task by associating with extracellular targets. These more accessible markers facilitate the transition from the use of aptamers *in vitro* to therapeutic applications *in vivo*.

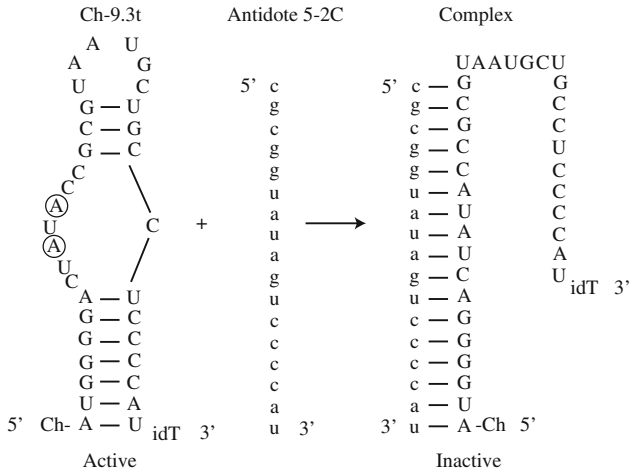


Fig. 6.7. Underlying mechanism for deactivation of the anticoagulant aptamer directed against the coagulation factor IXa (Ch-9.3t) by the antidote sequence (5-2C). Taken from [62]

The antithrombin aptamer is the first to have been selected against an extracellular protein. Thrombin is a plasma enzyme catalysing the conversion of soluble fibrinogen to its insoluble polymeric form, i.e., fibrin, and plays a key role in the process of hemostasis. After selection, the size of the antithrombin aptamer was reduced to just 15 nucleotides without loss of affinity ($K_d \approx 25$ nM). Even in its reduced form, this sequence has an inhibitory effect on its target protein [58]. Structural analysis of this sequence revealed the G tetrad functional structure of this aptamer and its interaction with two thrombin molecules [59]. One of the loops of the aptamer interacts with the heparin binding site and another with the fibrinogen binding site. This cartography illustrates how a short nucleic acid sequence can refold in an extremely ordered way to interact with its target.

Heparin is routinely used to inhibit thrombin, in such a way as to prevent the formation of clots, typical in certain pathologies, or during certain surgical operations, like those requiring extracorporeal circuits. However, the fact that its effects tend to perdure means that there is a significant risk of hemorrhage. Protamine sulfate can be used to counter this effect, but it also manifests a certain toxicity. Unlike heparin, the anticoagulation effect of the antithrombin DNA aptamer disappears almost immediately after the end of the perfusion in Cynomolgus monkeys [60]. This rapid reversion is due to the short half-life of oligonucleotides in the blood circulation system, mainly because of degradation by nucleases, which is put to use here.

It is remarkable that degradation by blood nucleases should thus be considered as advantageous. Indeed, most in vivo applications require sequences to be stable. Hence, the aptamer directed against the coagulation factor Ixa was selected from a 2'F-Py RNA pool (see Fig. 6.7) [61,62]. A cholesterol molecule

was associated with it at the 5' position after selection, in such a way as to hinder its blood clearance. One consequence of this was to greatly increase its blood circulation lifetime, from 10 min to 60–90 min, while its anticoagulant activity remained constant *in vivo* more than 1 h after a single injection with a dose of 0.5 mg/kg [62].

This is a much smaller dose than would be required to induce anticoagulation activity by constant injection of antithrombin aptamer, i.e., 0.3 mg/kg/min [60]. However, the possibility of instantaneously countering the anticoagulation effect is lost since the aptamer is stable in the blood circulation. Rusconi et al. put forward the idea of using an antidote specific to each aptamer by exploiting a characteristic property of oligonucleotides. They designed a sequence of 17 nucleotides able to associate with the aptamer as an antisense, causing a loss of conformation and consequently a loss of functionality in the aptamer. This antisense neutralises the anticoagulant effect of the aptamer by more than 95%, and this only 10 min after intravenous administration of 5 mg kg⁻¹ of antidote to pigs [62].

In the case of so-called escort aptamers, the aptamer may not itself have a therapeutic effect, but plays the role of a delivery molecule [63]. It is then modified with the aim of adding some therapeutic function, ensuring transport of the drug to its target. This type of functionalisation has been achieved with an aptamer inhibiting neutrophil elastase. The neutrophils are recruited at inflammation sites, where their role is to secrete large amounts of a serine protease, i.e., elastase. Under physiological conditions, the effect of elastase is finely regulated by inhibitors of endogenous proteases controlling the degradation of the components of the extracellular matrix. If the elastase is too active, this leads to various acute clinical manifestations. A first selection of aptamers against elastase was unable to procure an inhibiting effect on the protease [64]. However, if a low-activity elastase inhibitor is associated with the selected aptamer, its inhibiting effect is amplified by a factor of 100,000 [65].

Several dozen therapeutic aptamers are currently under development against a range of pathologies (cancer, inflammation, autoimmune diseases, etc.) and have shown positive effects on animal models [66]. A dozen or so aptamers have reached the stage of clinical trials, and one aptamer directed against the angiogenesis factor VEGF has been approved by the US Food and Drugs Administration (FDA) and is used to treat age-related macular degeneration [67].

6.5 Conclusion

Aptamers can be selected against a host of different targets, e.g., proteins, small ligands, cells, etc., which makes them serious rivals for antibodies, even capable of outdoing them in some applications (see Table 6.1). However, the properties characterising the association of aptamers with their targets make

Table 6.1. Comparing properties of aptamers and antibodies

	Aptamer	Antibody
Affinity	From nM to pM	From nM to pM
Possible targets	All	Difficulty against toxins or against weakly immunogenic targets
Selection process	In vitro	In animals
Selection condition	Determined	Difficult to impose
Storage	Unlimited	Limited
Thermal stability	Yes	Limited
Possibility of chemical synthesis	Yes	No
Interbatch variability	No	Possible
Possible chemical modifications	Wide range	Limited
Pharmacokinetic properties	Can be modified	Difficult to modify
Immunogenicity	Very low	High
Reversible inhibitor effect	Antisense = antidote	No

them rather special molecules. Like antibodies, they adopt 3D structures enabling them to associate with their target. But in the case of aptamers, this association often seems much more specific, some aptamers being able to differentiate between very similar molecules indeed. For example, the anti-theophylline aptamer can distinguish its ligand from a caffeine molecule for which its affinity is 10,000 times weaker, whereas theophylline and caffeine differ only by the replacement of one hydrogen by a methyl group [68]. (Note, however, that caffeine was used in the counterselection.) Another significant advantage of aptamers is that they are not immunogenic, even at doses 1,000 times greater than therapeutic doses. All the properties make aptamers ideal candidates for both diagnostic and therapeutic applications [26, 69].

When using SELEX, the experimenter is reproducing Darwin's principle of evolution in a nanoworld. Working on the molecular scale, the evolving species being oligonucleotide sequences here, literally billions of species can be manipulated in the space of a few microlitres. Likewise, the time scale is much shortened, so the evolution of animal species which would take centuries can be achieved in a few days with molecular species.

Their ease of use, remarkable properties, and vast range of applications mean that aptamers are sure to occupy an important position in the arsenal of new inventions available to biotechnology. The speed with which aptamers develop and the growing number of new approaches based on Darwinian molecular evolution confirm this impression. But at the same time it should remind the authors of this chapter to remain humble, since any new discoveries are soon out-of-date, given the extraordinary inventiveness driving this field of research!

To exemplify this, and to conclude, let us mention one of the many new lines of investigation recently opened up. It has been shown that aptamers can be reproduced by non-homologous recombination [70]. To do this, following PCR, the aptamer sequences are cut in different domains, which subsequently fuse together randomly. This kind of reproduction, never observed in nature, can be likened to the combination of different body parts of different species, e.g., an elephant leg with the neck of a giraffe. This could considerably enhance the evolutionary properties of aptamers, allowing faster and more effective evolution for more complex selection criteria than a simple affinity.

References

1. Dobzhansky, T.: Nothing in biology makes sense except in the light of evolution, *The American Biology Teacher* **35**, 125–129 (1973)
2. Darwin, C.: *The Origin of Species*, Penguin, London (1968)
3. Luchetta, P., Maurel, M.-C., Higuët, D., Vervoort, M.: *Evolution moléculaire: Cours et questions de révisions*, Dunod, Collection Sciences Sup (2005)
4. Mayr, E.: *The Growth of Biological Thought*, Harvard University Press, MA (1982)
5. Hermann, T., Patel, D.J.: Stitching together RNA tertiary architectures, *J. Mol. Biol.* **294**, 829–849 (1999)
6. Cate, J.H., Gooding, A.R., Podell, E., Zhou, K., Golden, B.L., Szewczak, A.A., et al.: RNA tertiary structure mediation by adenosine platforms [see comments], *Science* **273**, 1696–1699 (1996)
7. Laspias, M.F., Rice, A.P., Mathews, M.B.: HIV-I Tat protein increases transcriptional initiation and stabilizes elongation, *Cell* **59**, 283–292 (1989)
8. Hermann, T., Patel, D.J.: Adaptive recognition by nucleic acid aptamers, *Science* **287**, 820–825 (2000)
9. Ghosh, G., Huang, D.B., Huxford, T.: Molecular mimicry of the NF- κ B DNA target site by a selected RNA aptamer, *Curr. Opin. Struct. Biol.* **14**, 21–27 (2004)
10. Oliphant, A.R., Struhl, K.: The use of random-sequence oligonucleotides for determining consensus sequences, *Methods Enzymol.* **155**, 568–582 (1987)
11. Oliphant, A.R., Brandl, C.J., Struhl, K.: Defining the sequence specificity of DNA-binding proteins by selecting binding sites from random-sequence oligonucleotides: Analysis of yeast GCN4 protein, *Mol. Cell. Biol.* **9**, 2944–2949 (1989)
12. Kinzler, K.W., Vogelstein, B.: Whole genome PCR: Application to the identification of sequences bound by gene regulatory proteins, *Nucleic Acids Res.* **17**, 3645–3653 (1989)
13. Ellington, A.D., Szostak, J.W.: In vitro selection of RNA molecules that bind specific ligands, *Nature* **346**, 818–822 (1990)
14. Robertson, D.L., Joyce, G.F.: Selection in vitro of an RNA enzyme that specifically cleaves single-stranded DNA, *Nature* **344**, 467–468 (1990)
15. Tuerk, C., Gold, L.: Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase, *Science* **249**, 505–510 (1990)

16. Green, R., Ellington, A.D., Szostak, J.W.: In vitro genetic analysis of the tetrahymena self-splicing intron, *Nature* **347**, 406–408 (1990)
17. Thiesen, H.J., Bach, C.: Target Detection Assay (TDA): A versatile procedure to determine DNA binding sites as demonstrated on SP1 protein, *Nucleic Acids Res.* **18**, 3203–3209 (1990)
18. Blackwell, T.K., Kretzner, L., Blackwood, E.M., Eisenman, R.N., Weintraub, H.: Sequence-specific DNA binding by the c-Myc protein, *Science* **250**, 1149–1151 (1990)
19. Blackwell, T.K., Weintraub, H.: Differences and similarities in DNA-binding preferences of MyoD and E2A protein complexes revealed by binding site selection, *Science* **250**, 1104–1110 (1990)
20. Pollock, R., Treisman, R.: A sensitive method for the determination of protein–DNA binding specificities, *Nucleic Acids Res.* **18**, 6197–6204 (1990)
21. Joyce, G.F.: Directed molecular evolution, *Sci. Am.* **267** (6), 48–55 (1992)
22. Osborne, S.E., Matsumura, I., Ellington, A.D.: Aptamers as therapeutic and diagnostic reagents: Problems and prospects, *Curr. Opin. Chem. Biol.* **1**, 5–9 (1997)
23. Gold, L., Polisky, B., Uhlenbeck, O., Yarus, M.: Diversity of oligonucleotide functions, *Annu. Rev. Biochem.* **64**, 763–797 (1995)
24. Fitzwater, T., Polisky, B.: A SELEX primer, *Methods Enzymol.* **267**, 275–301 (1996)
25. Cox, J.C., Rudolph, P., Ellington, A.D.: Automated RNA selection, *Biotechnol. Prog.* **14**, 845–850 (1998)
26. Brody, E.N., Gold, L.: Aptamers as therapeutic and diagnostic agents, *J. Biotechnol.* **74**, 5–13 (2000)
27. Burmeister, P.E., Lewis, S.D., Silva, R.F., Preiss, J.R., Horwitz, L.R., Pendergrast, P.S., et al.: Direct in vitro selection of a 2'-O-methyl aptamer to VEGF, *Chem. Biol.* **12**, 25–33 (2005)
28. Green, L.S., Kirschenheuter, G.P., Charlton, J., Guidot, D.M., Repine, J.M.: Nuclease resistant nucleic acid ligands to vascular permeability factor/vascular endothelial growth factor, *Chem. Biol.* **2**, 683–695 (1995)
29. Pagratis, N., Bell, C., Chang, Y., Jennings, S., Fitzwater, T., Jellinek, D., et al.: Potent 2'-amino-, and 2'-fluoro-2'-deoxyribonucleotide RNA inhibitors of keratinocyte growth factor, *Nature Biotech.* **15**, 68–73 (1997)
30. Micklefield, J.: Backbone modification of nucleic acids: Synthesis, structure and therapeutic applications, *Curr. Med. Chem.* **8**, 1157–1179 (2001)
31. Latham, J.A., Johnson, R., Toole, J.J.: The application of a modified nucleotide in aptamer selection: A novel thrombin aptamer containing 5-(1-pentynyl)-2'-deoxyuridine, *Nucleic Acid Res.* **22**, 2817–2822 (1994)
32. Vater, A., Klussmann, S.: Toward third-generation aptamers: Spiegelmers and their therapeutic prospects, *Curr. Opin. Drug Discov. Devel.* **6**, 253–261 (2003)
33. Kumar, P.K.R., Ellington, A.D.: Artificial evolution and natural ribozymes, *FASEB J.* **9**, 1183–1195 (1995)
34. Lehman, N., Joyce, G.F.: Evolution in vitro of an RNA enzyme with altered metal dependence, *Nature* **361**, 182–185 (1993)
35. Tsang, J., Joyce, G.F.: Specialization of the DNA-cleaving activity of a group I ribozyme through in vitro evolution, *J. Mol. Biol.* **262**, 31–42 (1996)
36. Bartel, D.P., Szostak, J.W.: Isolation of new ribozymes from a large pool of random sequences, *Science* **261**, 1411–1418 (1993)

37. Ekland, E.H., Bartel, D.P.: RNA-catalysed RNA polymerization using nucleoside triphosphates, *Nature* **383**, 192 (1996)
38. Deng, Q., German, I., Buchanan, D., Kennedy, R.T.: Retention and separation of adenosine and analogues by affinity chromatography with an aptamer stationary phase, *Anal. Chem.* **73**, 5415–5421 (2001)
39. Romig, T.S., Bell, C., Drolet, D.W.: Aptamer affinity chromatography: Combinatorial chemistry applied to protein purification, *J. Chromatogr. B Biomed. Sci. Appl.* **731**, 275–284 (1999)
40. Michaud, M., Jourdan, E., Villet, A., Ravel, A., Grosset, C., Peyrin, E.: A DNA aptamer as a new target-specific chiral selector for HPLC, *J. Am. Chem. Soc.* **125**, 8672–8679 (2003)
41. Dick, L.W., Jr., McGown, L.B.: Aptamer-enhanced laser desorption/ionization for affinity mass spectrometry, *Anal. Chem.* **76**, 3037–3041 (2004)
42. Blank, M., Weinschenk, T., Priemer, M., Schluesener, H.: Systematic evolution of a DNA aptamer binding to rat brain tumor microvessels: Selective targeting of endothelial regulatory protein pigpen, *J. Biol. Chem.* **276**, 16464–16468 (2001)
43. Daniels, D.A., Chen, H., Hicke, B.J., Swiderek, K.M., Gold, L.: A tenascin-C aptamer identified by tumor cell SELEX: Systematic evolution of ligands by exponential enrichment, *Proc. Natl. Acad. Sci. USA* **100**, 15416–15421 (2003)
44. Fredriksson, S., Gullberg, M., Jarvius, J., Olsson, C., Pietras, K., Gustafsdottir, S.M., et al.: Protein detection using proximity-dependent DNA ligation assays, *Nat. Biotechnol.* **20**, 473–477 (2002)
45. Wang, X.L., Li, F., Su, Y.H., Sun, X., Li, X.B., Schluesener, H.J., et al.: Ultra-sensitive detection of protein using an aptamer-based exonuclease protection assay, *Anal. Chem.* **76**, 5605–5610 (2004)
46. Jhaveri, S., Rajendran, M., Ellington, A.D.: In vitro selection of signaling aptamers, *Nat. Biotechnol.* **18**, 1293–1297 (2000)
47. Hamaguchi, N., Ellington, A., Stanton, M.: Aptamer beacons for the direct detection of proteins, *Anal. Biochem.* **294**, 126–131 (2001)
48. Petach, H., Gold, L.: Dimensionality is the issue: Use of photoaptamers in protein microarrays, *Curr. Opin. Biotechnol.* **13**, 309–314 (2002)
49. Smith, D., Collins, B.D., Heil, J., Koch, T.H.: Sensitivity and specificity of photoaptamer probes, *Mol. Cell. Proteomics* **2**, 11–18 (2003)
50. Charlton, J., Sennello, J., Smith, D.: In vivo imaging of inflammation using an aptamer inhibitor of human neutrophil elastase, *Chem. Biol.* **4**, 809–816 (1997)
51. Tavitian, B.: In vivo imaging with oligonucleotides for diagnosis and drug development, *Gut* **52**, Suppl. 4, 40–47 (2003)
52. Werstuck, G., Green, M.R.: Controlling gene expression in living cells through small molecule–RNA interactions, *Science* **282**, 296–298 (1998)
53. Good, P.D., Krikos, A.J., Li, S.X., Bertrand, E., Lee, N.S., Giver, L., et al.: Expression of small, therapeutic RNAs in human cell nuclei, *Gene. Ther.* **4**, 45–54 (1997)
54. Thomas, M., Chedin, S., Carles, C., Riva, M., Famulok, M., Sentenac, A.: Selective targeting and inhibition of yeast RNA polymerase II by RNA aptamers, *J. Biol. Chem.* **272**, 27980–27986 (1997)
55. Shi, H., Hoffman, B.E., Lis, J.T.: RNA aptamers as effective protein antagonists in a multicellular organism, *Proc. Natl. Acad. Sci. USA* **96**, 10033–10038 (1999)
56. Vuyisich, M., Beal, P.A.: Controlling protein activity with ligand-regulated RNA aptamers, *Chem. Biol.* **9**, 907–913 (2002)

57. Cerchia, L., Ducongé, F., Pestourie, C., Boulay, J., Aissouni, Y., Gombert, K., et al.: Neutralizing aptamers from whole-cell SELEX inhibit the RET receptor tyrosine kinase, *PLoS Biol.* **3**, 123 (2005)
58. Bock, L.C., Griffin, L.C., Latham, J.A., Vermaas, E.H., Toole, J.J.: Selection of single-stranded DNA molecules that bind and inhibit human thrombin, *Nature* **355**, 564–566 (1992)
59. Kelly, J.A., Feigon, J., Yeates, T.O.: Reconciliation of the X-ray NMR structures of the thrombin-binding aptamer d(GGTTGGTGTGGTTGG), *J. of Biochemistry* **256**, 417–422 (1996)
60. Griffin, L.C., Tidmarsh, G.F., Bock, L.C., Toole, J.J., Leung, L.L.: In vivo anticoagulant properties of a novel nucleotide-based thrombin inhibitor and demonstration of regional anticoagulation in extracorporeal circuits, *Blood* **81**, 3271–3276 (1993)
61. Rusconi, C.P., Scardino, E., Layzer, J., Pitoc, G.A., Ortel, T.L., Monroe, D., et al.: RNA aptamers as reversible antagonists of coagulation factor IXa, *Nature* **419**, 90–94 (2002)
62. Rusconi, C.P., Roberts, J.D., Pitoc, G.A., Nimjee, S.M., White, R.R., Quick, G., Jr., et al.: Antidote-mediated control of an anticoagulant aptamer in vivo, *Nat. Biotechnol.* **22**, 1423–1428 (2004)
63. Hicke, B.J., Stephens, A.W.: Escort aptamers: A delivery service for diagnosis and therapy, *J. Clin. Invest.* **106**, 923–928 (2000)
64. Lin, Y., Qiu, Q., Gill, S.C., Jayasena, S.D.: Modified RNA sequence pools for in vitro selection, *Nucleic Acids Res.* **22**, 5229–5234 (1994)
65. Lin, Y., Padmapriya, A., Morden, K.M., Jayasena, S.D.: Peptide conjugation to an in vitro-selected DNA ligand improves enzyme inhibition, *Proc. Natl. Acad. Sci. USA* **92**, 11044–11048 (1995)
66. Pestourie, C., Tavitian, B., Ducongé, F.: Aptamers against extracellular targets for in vivo applications, *Biochimie* **87**, 921–930 (2005)
67. www.macugen.com
68. Jenison, R.D., Gill, S.C., Pardi, A., Polisky, B.: High-resolution molecular discrimination by RNA, *Science* **263**, 1425–1429 (1994)
69. Cerchia, L., Hamm, J., Libri, D., Tavitian, B., de Franciscis, V.: Nucleic acid aptamers in cancer medicine, *FEBS Lett.* **528**, 12–16 (2002)
70. Bittker, J.A., Le, B.V., Liu, D.R.: Nucleic acid evolution and minimization by nonhomologous random recombination, *Nat. Biotechnol.* **20**, 1024–1029 (2002)

Optical Tools

E. Roncali, B. Tavitian, I.e Texier, P. Peltié, F. Perraut, J. Boutet, L. Cognet, B. Lounis, D. Marguet, O. Thoumine, and M. Tramier

7.1 Introduction to Fluorescence Microscopy

Fluorescence is a physical phenomenon described for the first time in 1852 by the British scientist George G. Stokes, famous for his work in mathematics and hydrodynamics. He observed the light emitted by a mineral after excitation (absorption of light by the mineral) by UV light. He then formulated what has become known as Stokes' law, which says that the wavelength of fluorescence emission is longer than the excitation wavelength used to generate it. Some phenomena departing from this rule were later discovered, but do not in fact invalidate it. The possibility of visible excitation was subsequently developed, with the discovery of many fluorescing aromatic molecules, called fluorophores. The identification of these compounds and improved control over the physical phenomenon meant that by 1930 research tools had been developed in biology, e.g., labeling certain tissues and bacteria so as to observe them by fluorescence.

The optical microscope as it had existed since the nineteenth century thus gave rise to the fluorescence microscope: a reflection system to supply the light required to excite the fluorophores was added to the standard microscope, together with a suitable filtering system. Fluorescence microscopy soon became an important tool for biological analysis both *in vitro* and *ex vivo*, and other applications of light emission were also devised (light-emission phenomena of which fluorescence is a special case, described further in Sect. 7.2). It became possible to study phenomena that could not be observed by standard optical microscopy. Among other things, the location of molecules inside cells, monitoring of intracellular processes, and detection of single molecules all become feasible by means of fluorescence microscopy.

7.1.1 Conventional Fluorescence Microscopy

Experimental Setup

Fluorescence is a light-emission phenomenon which occurs through the de-excitation of a fluorescent molecule excited by external lighting (see Sect. 7.2)

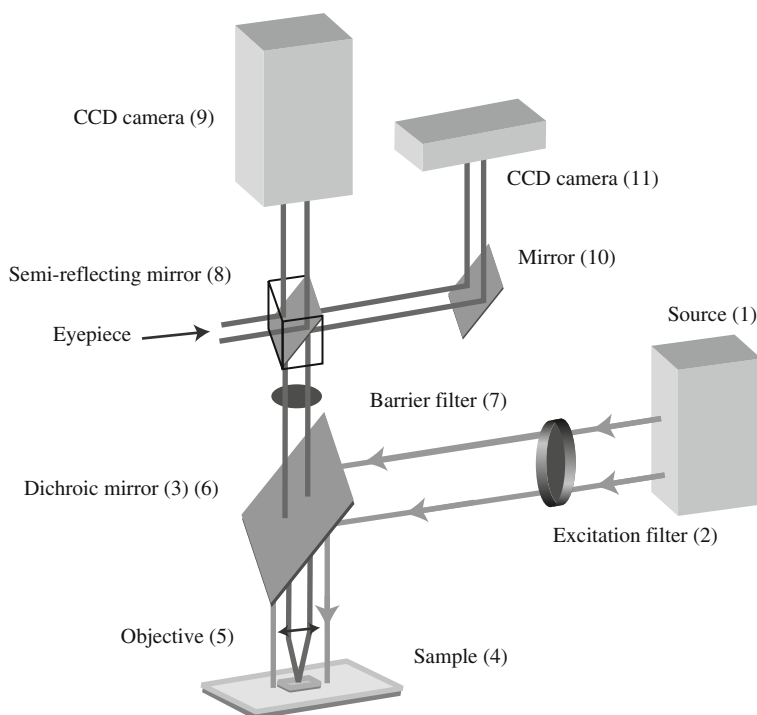


Fig. 7.1. Schematic view of reflection fluorescence microscope

[1]. A fluorescence microscope thus requires at least one excitation channel and one detection channel.

Figure 7.1 is a simplified diagram of a microscope operating by reflection. Light rays emitted by the source (1) (typically an arc lamp) cross the excitation (or exciter) filter (2), which selects a narrow optimal spectral band to illuminate the sample. The characteristics of this bandpass filter, are adjusted to suit the molecules one hopes to detect. A dichromatic mirror (or beamsplitter) reflects rays from the source perpendicularly to the incident direction (3), thus orienting the beam toward the sample. These rays then illuminate and excite the fluorescent molecules contained within the sample (4). The excited molecules emit light at a longer wavelength than the illuminating light, and this is transmitted by the dichromatic mirror (6) after passing through the objective (5). Just downstream of the dichromatic mirror is a high-pass filter called the barrier or emission filter (7), whose task is to eliminate from the detection channel any remnant of the excitation signal reflecting off the sample when it excites the fluorescent molecules. This unwanted reflection would inhibit detection of the fluorescence signal owing to its greater intensity, in fact brighter by several orders of magnitude than the fluorescent emission. It should be mentioned that the dichromatic mirror plays the same role, but cannot remove enough of the excitation light, which justifies the need for the

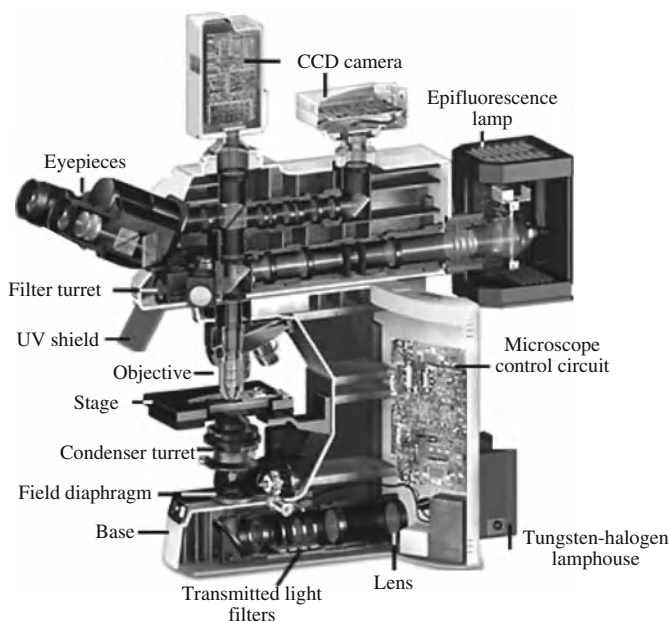


Fig. 7.2. Fluorescence microscope with two operating modes: transmission (conventional optical imaging) or reflection (fluorescence imaging) [1]

barrier filter. A CCD camera (9) detects the fluorescence image eventually formed. A second CCD camera (11) may record a similar or different fluorescence signal, recovered from the vertical channel by a semi-reflecting mirror (8) and redirected by the mirror (10). This is useful to carry out a detection using several fluorophores emitting at different wavelengths. Indeed, the semi-reflecting mirror can be dichromatic, with the task of separating several fluorescence signals emitted at different wavelengths and coming from a sample subjected to one or more excitations.

A transmission channel for exciting the sample is not shown in Fig. 7.1, but can be seen in the detailed view of Fig. 7.2. A halogen–tungsten lamp is used to illuminate the sample when carrying out conventional optical observation. It is sometimes useful to superpose a conventional ‘photograph’ of the sample on the fluorescence image in order to get a precise localisation of events.

Choice of Filter

The choice of filters and dichromatic mirror requires knowledge of the optical properties of the fluorescent molecules used, and also of the sample in which they are inserted. Figure 7.3 gives an example of a combination of filters adapted to a fluorophore, viz., Alexa Fluor 555. The absorption spectrum has a main peak at 555 nm and a secondary peak at 510 nm. The emission spectrum has a maximum at 570 nm. The excitation filter is preferably a narrow

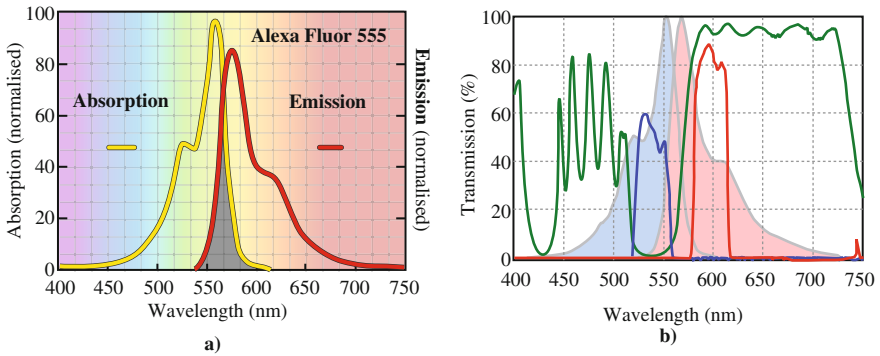


Fig. 7.3. (a) Absorption and emission spectra of the fluorophore Alexa Fluor 555, with absorption peak close to 555 nm and emission peak at 570 nm [1]. (b) Examples of transmission curves for filters and dichromatic mirrors. Here, the narrow bandpass excitation filter (*blue curve*) is effective between the two excitation peaks. The corresponding barrier filter (*red curve*) is a high-pass filter, eliminating the excitation signal. This filter backs up the imperfect cutoff by the dichromatic mirror (*green curve*) [2]

bandpass filter centered on the main absorption peak, i.e., at 555 nm. The barrier filter (also called the emission filter) will be a high-pass filter eliminating the excitation signal, for which the cutoff wavelength must therefore lie above 555 nm. A cutoff wavelength is chosen to block the maximum of the excitation signal without removing the emission peak. Here the excitation filter is a bandpass filter centered between the two absorption peaks, because the difference between the emission peak and the main absorption peak is too small to choose a filter centered on the latter. The detection filter backs up the dichromatic mirror by removing that part of the absorption spectrum that it does not eliminate and by improving the blockage in the spectral range where the dichromatic mirror is already effective.

Most fluorescence microscopes today contain what is called a fluorescence cube or block in which the excitation filter, dichromatic mirror, and barrier filter are all integrated. The microscope is then rather easily matched to another sample and another type of labelling by interchanging cubes. Four to six elements can be stored in a rotating turret which is simply twisted into position to line up the appropriate cube. This setup has the advantage that it preserves the alignment of the optical components, so the change can be made quickly. Figure 7.4 shows an example of such a fluorescence cube.

7.1.2 Examples of Biological Applications

Localisation in Cells

Fluorescence microscopy is used to detect fluorescent molecules in a sample and obtain information about the location and concentration of such

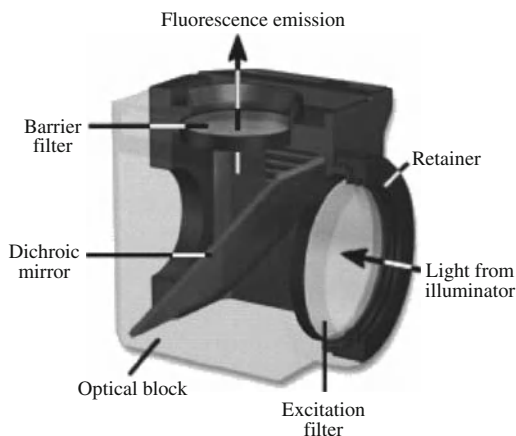


Fig. 7.4. Fluorescence cube [1]

molecules, by measuring the emitted light intensity, directly related to the fluorophore concentration. These fluorophores can thus serve as probes or labels: they reveal the presence of other molecules to which they are bound, making it possible to monitor their evolution, i.e., displacement, concentration, etc. Many biological phenomena can be studied by suitable labelling, both *in vitro* and *ex vivo*. Various light-emitting labels are described in Sect. 7.2.

There are several ways of attaching fluorophores to molecules (labelling), depending on the properties of the molecules to be monitored, and also the fluorophore. Labelling of antibodies (immunoassays) and nucleotide sequences (fluorescent *in situ* hybridization or FISH) have thus been developed.

However, when a fluorophore is anchored onto a biomolecule within a biological environment, its light-emitting characteristics are altered. For example, when a fluorophore is anchored onto a protein or a DNA sequence, these may contribute to steric hindrance which tends to reduce the quantum efficiency, i.e., the ratio of the light intensity produced to the light intensity received. Furthermore, the labelled biomolecule interacts with the environment in which it finds itself, and this may partially or totally extinguish the fluorescent light emission. There are two mechanisms here, depending on whether this extinction is reversible (a phenomenon called quenching) or permanent (the phenomenon known as photobleaching). The environment of the fluorophore can accelerate this process. Figure 7.5 shows an example of multiple labelling of different cell compartments, demonstrating that some of the fluorophores undergo photobleaching. Different parts of the cells are labelled by different fluorophores. This triple labelling reveals predominant photobleaching of the fluorophore labelling the cell nuclei.

Mobility Measurements

Some groups have investigated and exploited photobleaching phenomena in order to devise novel imaging techniques, such as fluorescence recovery after photobleaching (FRAP), in which photobleaching is deliberately induced in

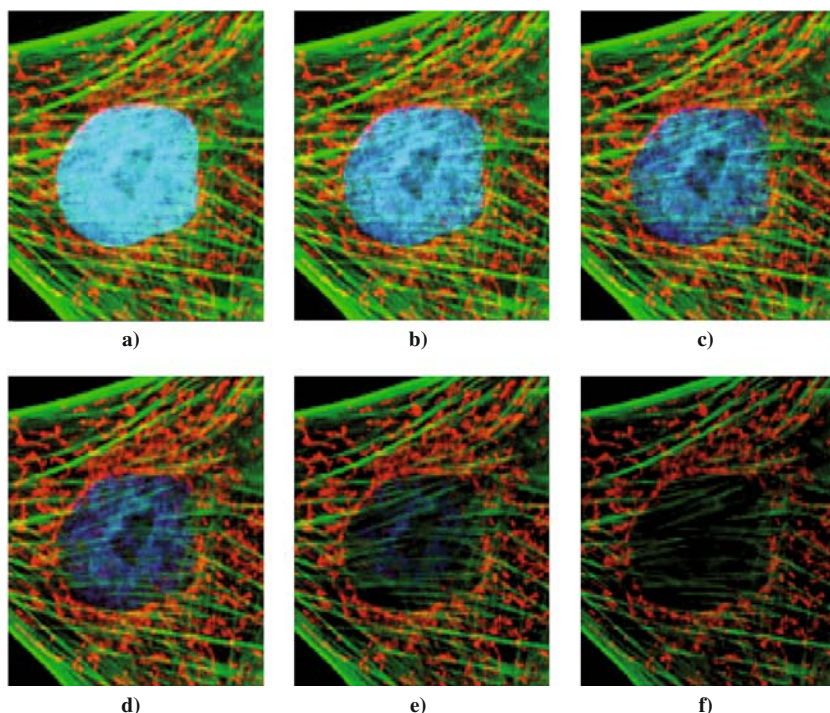


Fig. 7.5. Epidermal fibroblasts of the Indian muntjac deer. Nuclei stained with a bis-benzimidazole derivative (*blue*). Mitochondria stained with Mito Tracker Red CMXRos (*red*). Cytoskeleton stained with Alexa Fluor 488 (*green*). Images acquired every two minutes using a filter combination matched to the excitation of each fluorophore, and a suitable combination for detection. The signal from the nuclei falls off after 6–8 min, while the labels on the mitochondria and cytoskeleton resist photobleaching to some extent, but nevertheless begin to fade [1]

order to measure the mobility of fluorescent molecules returning to repopulate a bleached zone. This technique contributes in particular to the definition of the fluid mosaic model of the cell membrane, and it can also be used to determine the ability of a molecule to get through the cell membrane or interact with its components.

Interaction Measurements (FRET)

Other groups have studied fluorescence quenching effects. This has also given rise to novel imaging techniques, such as fluorescence resonance energy transfer (FRET), mainly intended for research into molecular interactions. This technique is based on the determination of the distance between two fluorophores, each labelling one of the two molecules whose interaction is under

investigation. After exciting one of the two fluorophores, called the donor, the energy emitted by this fluorophore is transmitted by resonance to the other fluorophore, called the acceptor. This transfer can only take place if the distance between the two fluorophores is less than $100 \text{ \AA}$, and it is distance dependent. The interaction and the distance between the two fluorophores, and hence the two relevant molecules, can thus be evaluated. The details of this technique are discussed in Sect. 7.5.

Other techniques, based on fluorescence quenching, photobleaching, or lifetime measurements (see the definition in Sect. 7.2) are presented in more detail in Sect. 7.5. These techniques can be used to visualise intracellular dynamical phenomena, monitor the motion of macromolecules, and even detect single molecules.

7.1.3 Confocal Microscopy

Principles

Like conventional optical microscopy, fluorescence microscopy is not resolved in depth and produces a superposed image of different layers of the specimen. This means that the background noise is high. An innovative technique called confocal microscopy was thus devised to isolate a section of the sample. This invention is attributed to Marvin Minsky [5], who presented a prototype in 1955 and the patent in 1957.

The confocal microscope replaces the arc lamp of the conventional epifluorescence microscope by a laser source. The laser is diffracted by a pinhole-type collimator, i.e., a hole or aperture with very small diameter, in order to broaden the beam and thereby illuminate the whole field of view. The beam focused by an objective then excites the fluorophores at a point of the sample, or more precisely, in a volume element centered on the relevant point.

Another pinhole collimator placed in front of the photomultiplier tube detecting the light eliminates all light coming from planes other than the focal plane. A fluorescence image of this plane can be built up point by point by scanning in two directions in the focal plane. A plane is selected by vertical translation of the sample, the effect being to modify the position of the focal plane relative to the sample. The pinhole plays a key role in the definition of the depth resolution: the latter is optimal when the aperture of the diaphragm is reduced to a diameter corresponding to the diffraction limit. This has the effect of minimising those rays coming from the volume element excited about the focal point.

One of the main advantages of the confocal microscope over the conventional microscope is the small depth of field – of micrometric order – so that one can obtain an image of a plane, a so-called optical section, which increases the signal-to-noise ratio and improves the contrast and sensitivity. By recording the different optical sections, a 3D image of the fluorescence signal can be reconstructed. The method is non-invasive, because these virtual sections

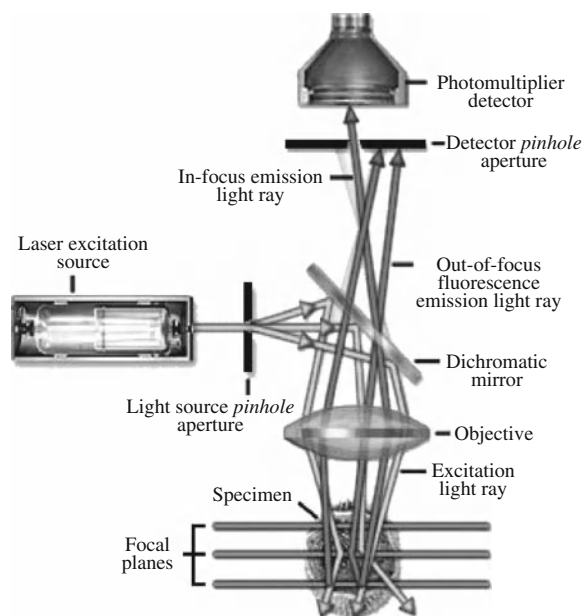


Fig. 7.6. Principal light pathways in a confocal microscope. *Light grey rays* come from the excitation source. *Dark grey rays* are those emitted by fluorescence from the fluorophores in the sample. Only the dark grey rays coming from the focal point remain after passing through the pinhole detector [3]

preserve the integrity of the sample, in contrast to conventional histological sections, so investigations can be carried out *ex vivo*.

Setup

The basic setup of a confocal microscope is shown in Fig. 7.6. The beam from the laser source is broadened by passing through a pinhole collimator, and then reflected by a dichromatic mirror in order to illuminate the sample. This beam is focused at a point of the focal plane by the objective. The focal plane is determined by the vertical position of the sample. The excited fluorophores emit light which goes through the objective and is transmitted by a dichromatic mirror. This light is partially stopped by a pinhole diaphragm of very small aperture. The effect of this is to select only those rays coming from the chosen focal plane. One must take into account the imperfect focusing and the fact that one excites a volume element and not a single point. The light is collected by a photomultiplier tube. By scanning in the plane of the sample, an optical section of the sample can be constructed. Several optical sections can be acquired by translating the sample in the vertical direction.

Example Application

We consider an example in which the confocal microscope is used for intracellular imaging. Figure 7.7 shows images of a COS7 cell (derived from kidney

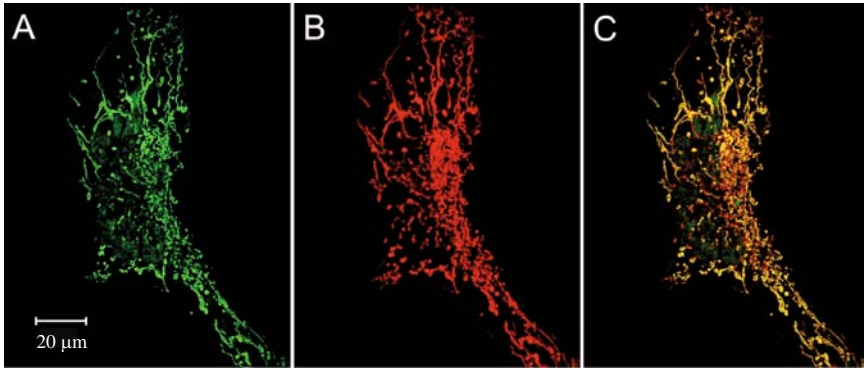


Fig. 7.7. Example of intracellular confocal imaging. (A) Confocal image of a COS7 cell transfected with GFP expressed in the mitochondria. (B) Confocal image of the same cell marked with an antibody against cytochrome C (specific protein of the mitochondrion) and a secondary antibody labelled with cyanine Cy3 (red). (C) Superposition of the two images to show the colocalisation of cytochrome C and GFP. With the kind permission of K. Rogers, Unité d'embryologie moléculaire, Institut Pasteur

cells of the African green monkey. The first image shows a COS7 cell transfected by green fluorescent protein (GFP), expressed in the mitochondria. The second shows the same cell marked with an antibody against cytochrome C (specific protein of the mitochondrion) and a secondary antibody labelled with cyanine Cy3. This study is able to locate mitochondria by the GFP and, more exactly, to locate the cytochrome C within the mitochondria.

7.1.4 Two-Photon and Multiphoton Microscopy

There are alternatives to confocal microscopy, such as two-photon microscopy or multiphoton microscopy. These techniques bring some advantages for 3D imaging and can be used to image living cells, intact brain sections, embryos, and even whole organs.

The basic idea is shown in the Jablonski diagram (see Fig. 7.8): a fluorophore that can be excited by a photon of energy E and wavelength λ can also be excited by two photons of energy $E/2$ and hence of wavelength 2λ interacting simultaneously (or rather, within a time window of 10^{-18} s) with the fluorophore.

This technique requires a very powerful laser, supplying a high enough photon density to generate pairs of photons able to create the necessary excitation. A pulsed laser is generally used. The beam is spatially focused and this means that not too many photons will be absorbed by fluorophores outside the focal point.

The absence of absorption outside the focal plane allows one to increase the penetration depth and reduce phototoxicity. These properties make

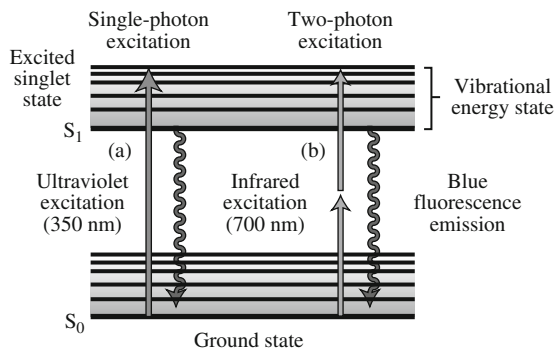


Fig. 7.8. Jablonski diagram for one- and two-photon excitation. The energy E of a photon required to get the fluorophore from the state S_0 to the state S_1 is equal to the sum of the energies $E/2$ of two photons absorbed simultaneously by the fluorophore. The same fluorescence emission occurs in both cases [4]

two-photon microscopy an innovative and powerful tool for imaging thick samples or living tissues that must not be damaged. The method can be extended to multiphoton imaging, using the idea that a molecule that can be excited by a photon of energy E can also be excited by three photons of energy $E/3$.

7.1.5 Conclusions and Prospects

Today, fluorescence microscopy has become a basic analytical technique, essential to biological research. It is relevant to molecular biology, cell biology, and even observation of cell sections. These *in vitro* and *ex vivo* studies are a preliminary step for *in vivo* research. The challenge of *in vivo* investigation has encouraged the development of processes capable of imaging the whole body of a small animal (see Sect. 7.3).

The extension of the field of application of microscopic techniques, the discovery of many molecules with suitable properties for labelling purposes, and the ever increasing need for temporal and spatial resolution have created today a context that is more propitious than ever for research in optical microscopy.

7.2 Labels

7.2.1 Introduction

Luminescence is the phenomenon whereby certain molecules raised to an excited state drop back down to their ground state by returning some of the stored energy in the form of light emission. When the molecule is excited by an external light source, one speaks of fluorescence (emission from an excited

state with the same electron spin multiplicity as the ground state) or phosphorescence (emission from an excited state with different multiplicity to the ground state) [6]. These phenomena occur to varying degrees in all molecules, at different wavelengths that are in fact specific to the molecule. When the energy causing the molecules to reach their excited state comes from a chemical or biochemical reaction, one speaks of chemiluminescence or bioluminescence. The latter was discovered in lucioles and certain marine species. The reader interested in the theoretical aspects of fluorescence should refer to the book by J.R. Lakowicz [6]. Here we limit the discussion to a definition of the specific terms used and a general description of fluorescent labels without going into the details of the underlying light-emission phenomena.

Although living beings can naturally display intrinsic fluorescence properties, due in particular to aromatic amino acids or DNA bases [6], it is easier to image tissues, cells, or small animals by using extrinsic labels as probes. Moreover, extrinsic labelling is much more flexible in allowing a choice of emission wavelength from the label.

Among these labels, the so-called exogenous labels are molecules or nanoparticles foreign to the cell, tissue, or small animal to be imaged. Typical examples are the organic fluorophores, light-emitting particles, and in general any synthetic label. Such labels are used for *in vivo* analyses, but also in many cases for *in vitro* detection, both in biology, which is the specific subject of this chapter, and in other areas such as environmental studies or the food industry, for example, for the detection of ions, pollutants, and so on. In biology, most *in vitro* studies concern biologically ‘passive’ detection systems, using the principle of molecular recognition. Detection is achieved by labelling either the probe or the target by means of a fluorescent molecule. These exogenous probes for *in vitro* or *in vivo* studies will be described in Sect. 7.2.2.

Concerning the case of *in vivo* detection (cell, tissue, or small animal), reporter genes can also be used. The organism is then genetically modified in such a way as to introduce into its genome a DNA construction which, after transcription, will produce a protein causing light emission, either by fluorescence, e.g., the family of green fluorescent proteins (GFP), or by bioluminescence in the presence of an external substrate, e.g., the family of luciferases. These genetic constructions and their *modus operandi* will be described in Sect. 7.2.3.

7.2.2 Exogenous Probes

Criteria for Selecting Light-Emitting Probes

General Criteria

Many light-emitting probes are commercially available and it is important to choose the one that is best suited to the type of fluorescence imaging required.

However, whatever probe is used, a molecule with a high molar extinction coefficient¹ must be selected, with a high quantum efficiency for emission,² and a high level of photostability or resistance to photobleaching.³

The required spectral characteristics for a given application, e.g., emission and absorption wavelengths and band widths, Stokes shift,⁴ etc., are key factors in the choice of label. For biological analysis, wavelengths shorter than about 350 nm should be avoided so as not to damage genetic material, the proteins under investigation, or surrounding tissue. Note also that the higher the Stokes shift, the easier it will be to eliminate incident photons during detection, thereby increasing the signal-to-noise ratios that can be obtained. The physical and chemical characteristics, such as sensitivity to the environment (e.g., solvent and buffer, pH, presence of ions, etc.), solubility, and luminescence lifetime,⁵ must also be taken into account when choosing the molecule.

Specific Criteria in Vivo

Some specific selection criteria for fluorescent labels must be taken into account for in vivo studies, whenever cells, tissues, or small animals are under investigation. Indeed, the ability of a biological tissue to absorb or scatter light depends on the wavelength of the photons, but also on the nature of the tissue [7]. In addition, tissues, but also cells, are autofluorescent, i.e., naturally fluorescent, as can be seen from Fig. 7.9. This issue will be tackled more specifically in Sect. 7.3, dealing with in vivo detection. Figures 7.9 and 7.10 show that in vivo fluorescence imaging is easier in the red part of the visible spectrum. Hence, Fig. 7.9 shows that autofluorescence of tissues falls off in this spectral range.

Figure 7.10 shows the optimal windows for in vivo imaging, in which absorption by tissues and water is minimal. This window in the near infrared, between 650 and 900 nm, corresponds to low tissue absorption, since tissues tend to absorb at shorter wavelengths, and low absorption by water, which tends to absorb at longer wavelengths, above 1,000 nm. These optical properties of the tissues are all the more restrictive as the region to be imaged lies deeper (see Sect. 7.3).

Chemically speaking, light-emitting probes fall into three main categories: organic fluorophores, inorganic lanthanide complexes, and systems based on nanoparticles. In the following sections, we shall discuss the properties of these different probes.

¹ This coefficient reflects the capacity of a molecule to absorb the excitation light.

² Level of conversion of absorbed light into re-emitted light.

³ Destruction of the emission properties of the molecule when it is irradiated by light.

⁴ Difference between the emission and excitation wavelengths.

⁵ Time lapse over which emission is observed from the label after its excitation by a pulsed source, of the order of the nanosecond (10^{-9} s) for an organic fluorophore, and up to several hundred μ s for lanthanide chelates (see pp. 265 and 267).

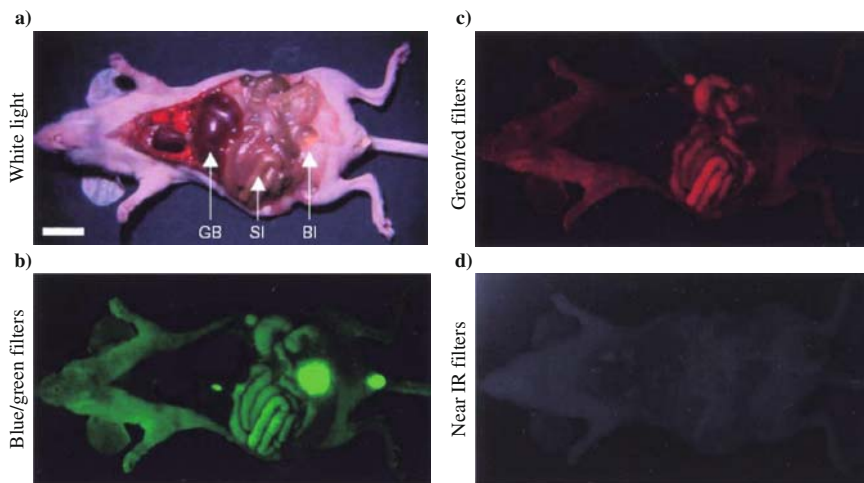


Fig. 7.9. Autofluorescence properties of tissues in different organs. A nude mouse imaged after sacrifice: (a) in white light or with different combinations of filters at an incident power of 2 mW/cm^2 , (b) blue/green (460–500 nm/505–560 nm), (c) green/red (525–555 nm/590–650 nm), (d) near infrared (725–775 nm/790–830 nm). GB gall bladder, SI small intestine, BI bladder [8]. Copyright Elsevier (2003)

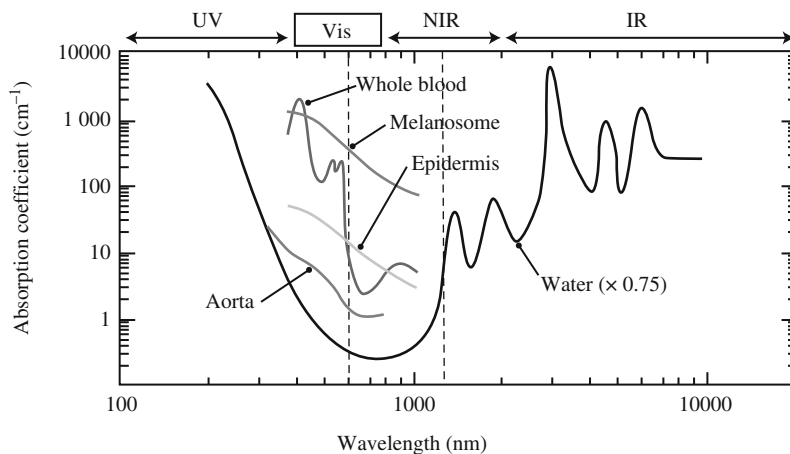


Fig. 7.10. Absorption properties of certain tissues as a function of wavelength [7]

Organic Fluorophores

Many organic molecules are commercially available [6], covering a broad spectral range from the near UV to the near infrared. Figure 7.11 shows the main fluorophore families and associated wavelength ranges: coumarins, fluoresceins, rhodamines, cyanines, bodipy (derivatives of boron dipyrromethene). It also shows some examples of the structures of these fluorophores. Some

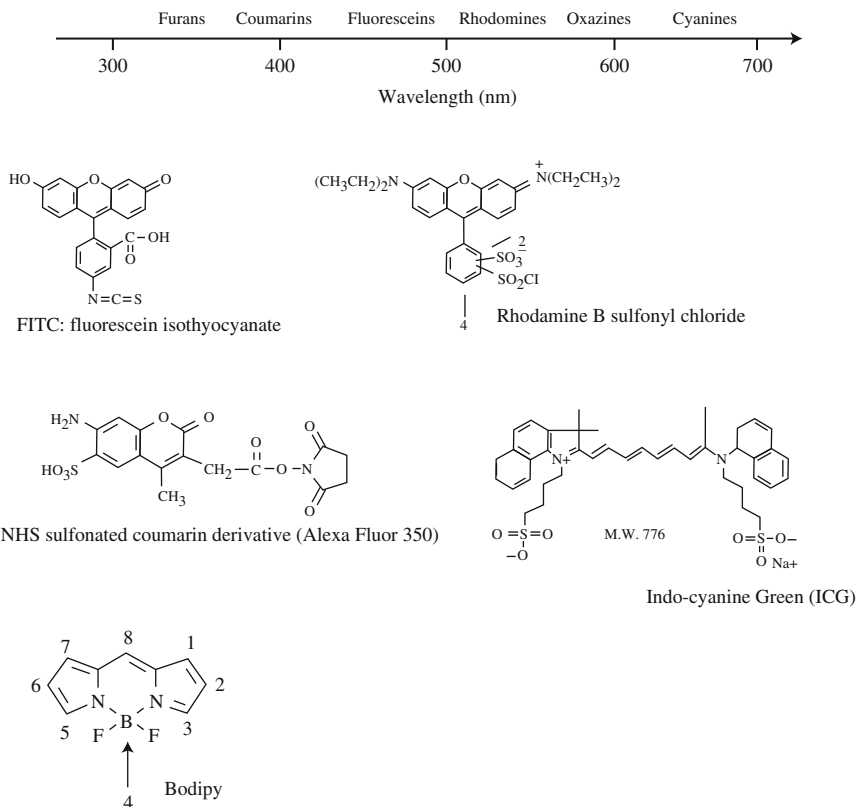


Fig. 7.11. Main families of organic fluorophores in the different optical wavelength ranges

companies commercialise series of fluorophores covering the whole visible spectrum with molecules borrowed from different families, using generic registered trade names, e.g., Alexa Fluor, Atto dyes, etc.

Fluorophores absorbing and emitting in the near UV or the visible are suitable for *in vitro* applications. Fluoresceins like FITC, FAM, JOE, etc., can exhibit some problems of sensitivity to pH, since they fluoresce only in an alkali medium, and a tendency to photobleaching. Rhodamines like ROX, TAMRA, Texas Red, etc., are more stable and less affected by the pH, but they tend to have lower fluorescence quantum yields. The optical properties of bodipys do not depend on the pH or the polarity of the solvent, and they are uncharged fluorophores.

For *in vivo* applications, especially optical imaging through tissues, it is essential to work at wavelengths at which absorption and scattering of light by intervening tissues is minimal, which means the red and near infrared spectral ranges (see p. 263). The best organic fluorophores for this type of application are the cyanines. However, organic molecules emitting in the infrared do not

possess optimal optical properties, and in particular, their quantum efficiency is low (between 5 and 40% in organic solvents) and they have a short fluorescence lifetime.⁶ In addition, these large aromatic molecules are difficult to solubilise and show a strong tendency to aggregate, in particular, to form dimers [9].

The large number of commercially available fluorophores and the fact that they are so easy to use in organic chemistry means that these labels are well suited for use as optical probes. They are very easily coupled with all sorts of biomolecules, such as oligonucleotides, peptides, saccharides, etc., using different methods of synthesis. Although only ICG (Indo Cyanine Green) and fluorescein can be injected in humans [7], many studies use these compounds in vivo for cell cultures or in small animals. Their molecular dimensions and the wide range of structures available, e.g., positively charged, negatively charged, or neutral, mean that they can be matched in vivo to suit the required pharmacokinetics when anchored to a biological probe. In addition, labels can be chosen whose emission properties, e.g., quantum efficiency and fluorescence lifetime, can be modulated by the environment (solvent, temperature, presence of ions or oxygen, etc.) and the close proximity of fluorescence quenchers. They therefore constitute choice candidates for studying molecular interactions (see the discussion of FRET in Sect. 7.5) and diffusion in biological media (see the discussion of the FRAP and FLIM techniques in Sects. 7.1 and 7.5, respectively), but also for making local measurements, notably in living beings, of physicochemical parameters such as the pH, ion concentrations, and so on. Many examples of probes and associated applications can be found in the *Handbook of Fluorescence* [10].

However, the use of these labels can be limited by their optical properties, which are not always perfectly adequate for the required applications. The most restrictive property of these compounds is their high rate of photobleaching, which makes it difficult to carry out prolonged acquisition. Their low Stokes shift can also make it difficult to filter out scattered incident light, at the excitation wavelength of the system, from the emission signal. Their broad absorption and emission bands mean that they are difficult to use for the simultaneous observation of different labels. Even over a range extending from 350 to 800 nm, it is difficult to make measurements with more than four different labels.

Luminescent Lanthanide Chelates

The structure of a lanthanide chelate comprises three elements (see Fig. 7.12): an organic chromophore motif collecting the excitation energy is anchored onto a ligand complexing an inorganic lanthanide cation (usually europium, terbium or ruthenium). The lanthanide complex can be anchored onto a biological probe, designed to target some specific molecule, via the organic ligand

⁶ Lifetime of the fluorescence-emitting excited state, typically of the order of one nanosecond for an organic fluorophore.

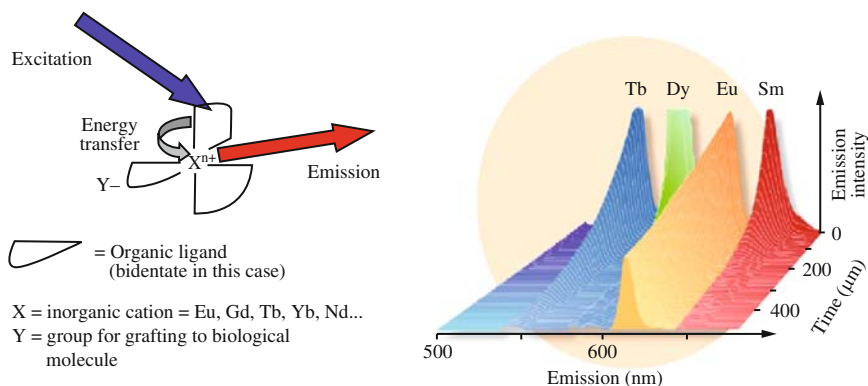


Fig. 7.12. Structure and properties of lanthanide chelates. (a) General structure of a lanthanide chelate, in this case with three bidentate organic ligands. (b) Luminescence spectra and lifetimes of lanthanide chelates in the DELFIA system proposed by Perkin Elmer

thanks to the presence of a spacer carrying an electrophilic or nucleophilic function.

The separation of the absorption function (the organic chromophore) and emission function (the cation) induces a large Stokes shift. Indeed, the organic chromophores absorb in the UV or the blue region of the visible spectrum, while the cations re-emit in the green–red, viz., 605 nm for europium and 560 nm for terbium. In addition, the cations generally have a long emission period, between several hundred nanoseconds and several hundred microseconds, which means that these complexes are particularly well suited for delayed fluorescence imaging systems.

The main disadvantage of the lanthanide chelates is that their excitation wavelength lies in the blue. Although this is not a major difficulty when developing *in vitro* analytical and diagnostic systems [11, 12], this spectral range is not particularly good for *in vivo* imaging, because autofluorescence from biological tissues is high in this region and limits detection sensitivity. The lanthanide complexes are thus not particularly well suited for *in vivo* optical imaging, unless two-photon setups are employed [13, 14].

Analyte Detection by Time-Delayed Luminescence Measurements with Lanthanide Chelates

There are several commercially available systems for analysing molecular interactions, e.g., antigen/antibody recognition, DNA hybridisation, using these labels, e.g., TRACE and HTRF developed by CisBio International. Figure 7.13 illustrates in a simplified way the method used by such systems. The two entities A and B whose interaction is to be detected are labelled, A by the lanthanide chelate and B by an organic fluorophore with short fluorescence

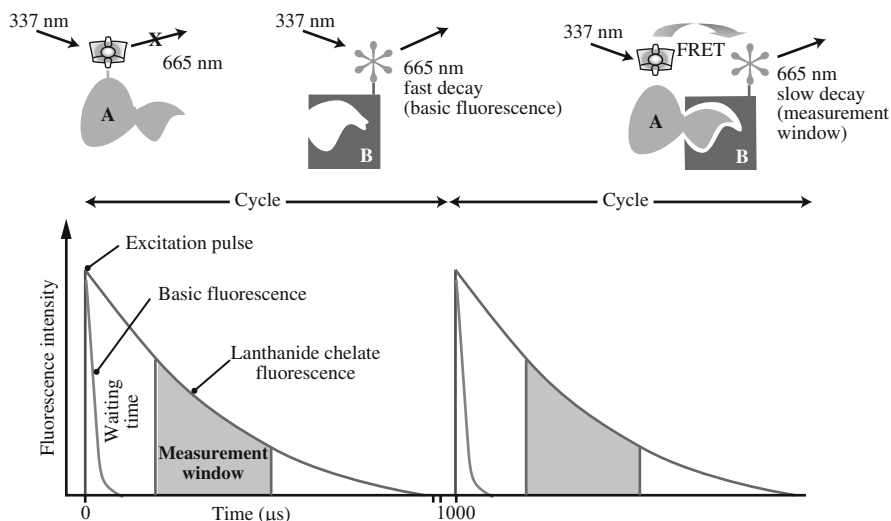


Fig. 7.13. Use of homogeneous time-resolved fluorescence (HTRF) developed by CisBio International [15]

lifetime, acceptor of the chelate fluorescence. The idea is to exploit the long luminescence lifetime of the chelate and the different emission wavelengths for the chelate (620 nm for the case shown in Fig. 7.13) and the acceptor (665 nm for the case shown in Fig. 7.13). When the entity A labelled by the chelate is alone and excited at 337 nm, the signal at 665 nm is not picked up. When the entity B labelled by the acceptor is alone and excited at 337 nm, there may be fluorescence at 665 nm, but it will have the short lifetime of the acceptor emission. When molecular recognition occurs between A and B, the chelate excited at 337 nm transfers its energy to the acceptor, which not only emits at 665 nm, but has its fluorescence lifetime extended to that of the chelate (a few hundred μs for a europium complex). Using a pulsed excitation and recording the signal in a measurement window shifted relative to these light pulses, unwanted fluorescence such as fluorescence from entity B not bound to A, or fluorescence from the buffer, are removed and only the fluorescence emitted after formation of the A–B complex is detected.

Nanoparticle Probes

Functionalised nanoparticles, especially those functionalised by biological molecules, were examined in Chap. 4. Here we shall simply summarise their main properties, together with their advantages and disadvantages when used as optical probes.

The following nanoparticles can be used as optical probes:

- light-emitting semiconductor nanocrystals (quantum dots),
- doped rare earth oxide nanocrystals,

- silica or polymer nanoparticles incorporating or functionalised by the fluorophores cited above (organic fluorophores, lanthanide complexes, quantum dots),
- metal nanoparticles.

Light-Emitting Semiconductor Nanocrystals

These new probes with particularly attractive optical properties are competing more and more with organic fluorophores, given the rather limited optical response of the latter, especially with regard to photobleaching [16–20].

Structure. Quantum dots are inorganic nanocrystals. They comprise a semiconductor core (CdS, CdSe, ZnO, InP, InAs, etc.), responsible for light emission, generally enclosed in an inorganic shell (ZnS) which increases the quantum efficiency and limits photobleaching. The size of the inorganic crystal can vary between 2 and 8 nm. If they are to be used for biological applications, they must be functionalised by organic molecules in order to graft on biomolecules (see Fig. 7.14).

Optical Properties. The recent surge of interest in quantum dots [16–20] results from their optical emission properties. These are due to the recombination of an electron–hole pair created during light excitation in the semiconductor nanocrystal core [18]. For a given semiconductor material making up the core of the nanoparticle, the emission wavelength is determined by the particle diameter, which generally varies between 2 and 8 nm. Indeed, the size of the particle directly affects the bandgap of the semiconductor nanocrystal. The larger the particles, the smaller the bandgap and the further the emission wavelength will be shifted toward the red (see Fig. 7.15). The semiconductor material chosen to form the core also determines the range of emission wavelengths, e.g., CdS for UV–blue, CdSe for the visible, and CdTe for the infrared. The so-called type II nanocrystals, for which the light-emission property is provided by a core–shell structure (e.g., CdTe for the core, CdSe for the shell), can also be used for the near infrared [23]. More recently, Nie and coworkers varied the core composition to obtain a range of emission wavelengths in the near infrared [24].

The emission band of quantum dots can be very narrow, since it is directly correlated with the size dispersion of the nanoparticles. On the other hand, their broad absorption spectrum allows for simultaneous excitation at a single wavelength of nanocrystals with different emission wavelengths (see Fig. 7.15a) [16, 21, 26]. This large Stokes shift between excitation and emission, combined with the narrowness of the emission band, means that these probes are easy to use for multiplexed analysis, as has been demonstrated with *in vitro* setups, e.g., use of colour multiplexing to label polymer beads by a kind of optical bar code (see p. 277) [27–29].

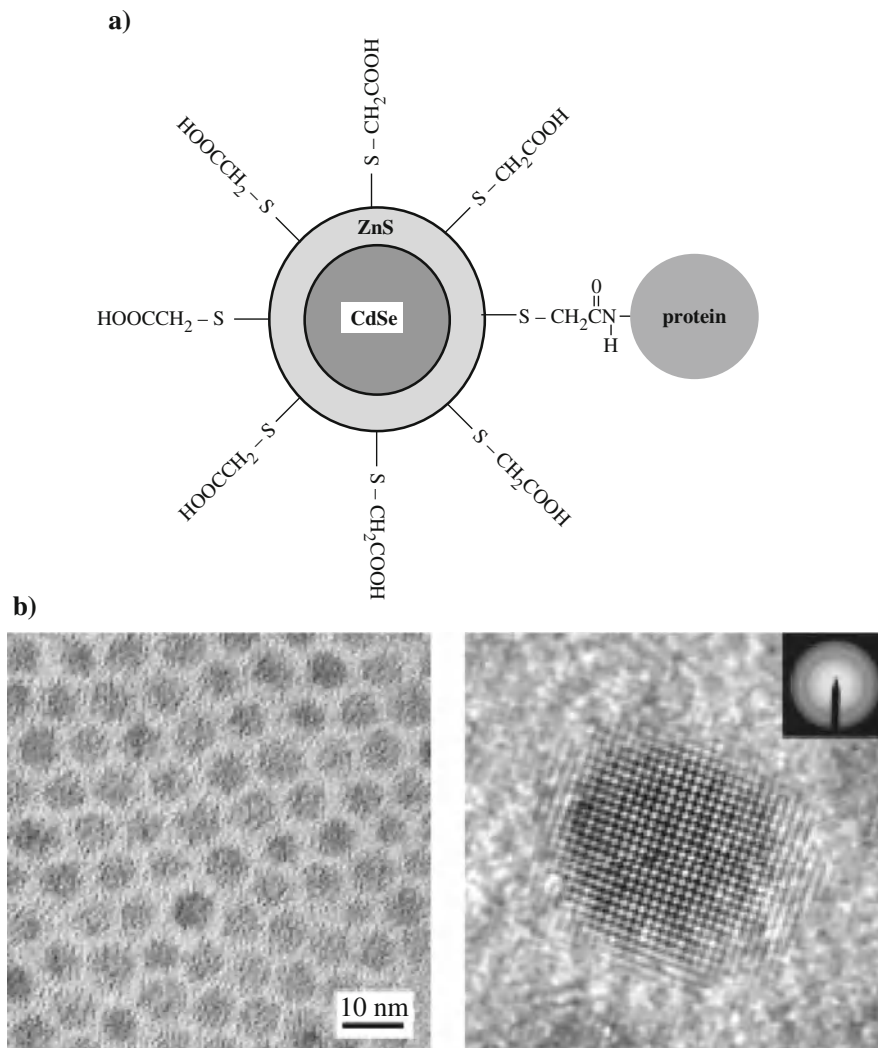


Fig. 7.14. Quantum dots. (a) Quantum dot with a CdSe core and a ZnS shell, functionalised by mercaptoacetic acid ligands for the purpose of grafting on a protein [21]. (b) High-resolution image by transmission electron microscopy (HRTEM) of PbS quantum dots, showing the crystal planes in these nanostructures [22]. Copyright Wiley-VCH Verlag. Reproduced with the kind permission of the publisher

In addition, quantum dots have very low photobleaching rates compared with organic fluorophores, and acceptable quantum emission efficiencies (10–15%) in an aqueous buffer when functionalisation has been carried out carefully [16–20, 25, 30–32]. Their emission lifetimes in the range 50–100 ns are

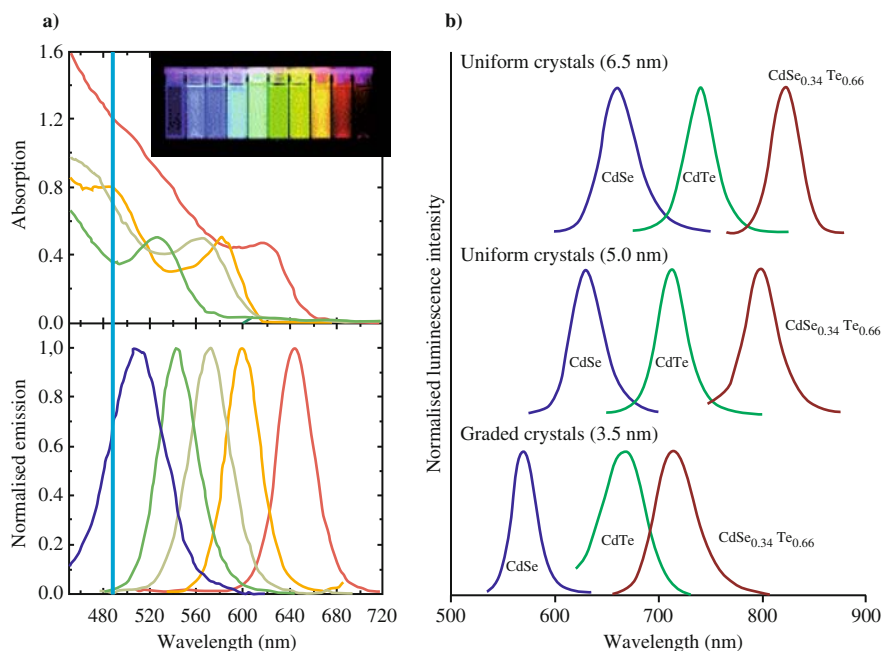


Fig. 7.15. Optical properties of quantum dots as a function of their size and composition. **(a)** Size dependence. Absorption and emission spectra of CdSe/ZnS nanocrystals encapsulated in silica. Size (CdSe core): 2.7 nm (emission 544 nm, *green curve*), 3.1 nm (emission 576 nm, *brown curve*), 4.1 nm (emission 595 nm, *orange curve*), and 4.8 nm (emission 644 nm, *red curve*) [25]. The *vertical blue line* indicates the wavelength of an argon laser at 488 nm, showing that different crystals emitting at different wavelengths can be excited simultaneously [19]. *Inset:* Emission colours of CdSe/ZnS nanocrystals of different sizes illuminated by a UV lamp. Their emission wavelengths are *from left to right*: 443, 473, 481, 500, 518, 543, 565, 587, 610, and 655 nm [16]. Reproduced with the kind permission of Elsevier. **(b)** Composition dependence. For a given size of nanocrystal, the emission wavelength depends on the nature and composition of the semiconductor material constituting the core of the quantum dot [24]

intermediate between the organic fluorophores (ns) and lanthanide complexes ($\approx \mu\text{s}$). These molecules also have two-photon absorption cross-sections three orders of magnitude greater than the organic fluorophores, making them ideal for multiphoton microscopy [33]. On the other hand, although they do not photodegrade, taken individually under continuous irradiation, their emission dies out from time to time in an unpredictable but reversible manner. This so-called blinking process limits their use to imaging single molecules, but is without consequence for the quantitative detection of a nanoparticle assembly.

Functionalisation. Semiconductor nanocrystals emitting in the near infrared are commercially available (Evident Technologies or Invitrogen) and have also

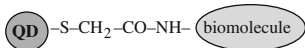
been developed by different research groups [16–20] with emission wavelengths in the range 490–950 nm, and different chemical groups or biomolecules at the surface (CO_2H , NH_2 , biotin, streptavidin, protein A, antibodies, etc.). Different methods for functionalising and solubilising these nanocrystals, initially synthesised in organic solvents, have been proposed (see Fig. 7.16) [16–20]. It nevertheless remains difficult to obtain monodispersed nanocrystals in aqueous buffers, anchored to biomolecules, and yet still to preserve all their optical properties in a satisfactory way. Indeed, since these hydrophobic inorganic compounds are synthesised in organic solvents, where they are stabilised by TOP ligands (TOP = trioctylphosphine) and TOPO ligands (TOPO = trioctylphosphine oxide), they must subsequently be transferred to an aqueous medium.

At the present time, there are three main strategies for functionalising these particles [16–20]. These are summed up in Fig. 7.16. The first (A) consists in exchanging the hydrophobic ligands used to stabilise the nanocrystals during their synthesis in an organic solvent for amphiphilic ligands onto which a biomolecule can be anchored. This strategy can lead to a significant reduction in emission properties. The second strategy (B), which is rather similar, consists in stabilising the quantum dots as soon as they are synthesised in the organic solvent by means of modified TOP ligands. The third strategy (C) consists in adsorption or encapsulation of the quantum dots in polymers or copolymers. These polymers carry functional groups on which biomolecules are already anchored or can be anchored at a later stage. This kind of strategy can raise some problems of reproducibility, for it is difficult to control the synthesis of the polymers and their adsorption at the nanocrystal surface. Moreover, poor adsorption or encapsulation of the quantum dot can have serious consequences with regard to the toxicity of these nanoparticles, since it is the passivation of the nanocrystals by the organic shell surrounding them which prevents any oxidation and subsequent leakage of the highly toxic heavy elements like Cd, Se, etc., to be found in the core of the quantum dot.

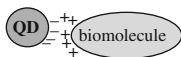
Biocompatibility. The main handicap for these labels, as compared with organic fluorophores, remains the lack of data concerning the toxicity of such particles. The heavy metals making them up, e.g., Cd, Te, Se, and so on, and the phosphine ligands used to functionalise them are all known for their acute and chronic toxicity. However, supporters of quantum dot techniques argue that, in these nanocrystals, such materials do not occur in their native form, but passivated by a layer of organic ligands which prevents them from escaping and screens their toxicity. In addition, a calculation based on the amount of nanocrystals injected to label sentinel lymph nodes in pigs showed that 400 pmol of quantum dots emitting in the near infrared (amount injected into pigs weighing 35 kg) represents 9.9 $\mu\text{g}/\text{kg}$ of Cd (dose 300 times less than the toxic daily dose for rats), 7.3 $\mu\text{g}/\text{kg}$ of Te, 2.4 $\mu\text{g}/\text{kg}$ of Se, and 4.1 $\mu\text{g}/\text{kg}$ of phosphines [34], which is well below toxic levels [34]. All studies so far carried out on cells [35–37] or on small animals [33, 34, 38–42] have shown that quantum dots can be stable and non-toxic in vivo, at least, as long as there

A. Methods based on exchange of bifunctional thiol ligands

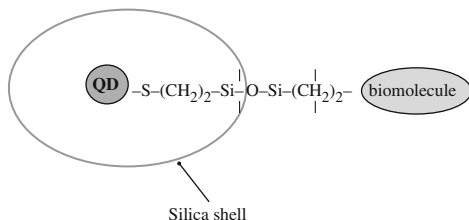
A.1 Coating with bifunctional thiols:



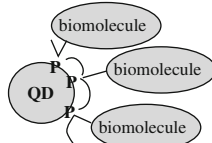
A.2 Coating with negatively charged bifunctional thiols:



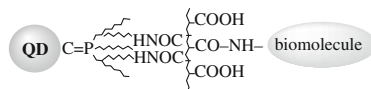
A.3 Coating with thiol silanes and formation of a silica shell

**B. Methods based on modification of phosphine ligands**

B.1 Phosphines oligomeriques

**C. Method using adsorption of amphiphilic polymers on the particle**

C.1 Adsorption of modified polyacrylic acid on the TOPO layer



C.2 Adsorption of amphiphilic phospholipids on the TOPO layer

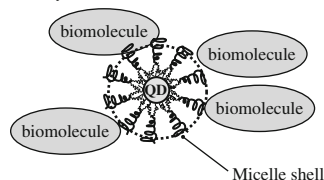


Fig. 7.16. Functionalisation of quantum dots [16]. **(A)** Ligand exchange methods using bifunctional thiols. (A1) A wide variety of thiols can be used [26, 31]. This strategy was first suggested by Nie et al. [21]. (A2) This procedure, very similar to the last, differs in the way the protein is anchored onto the layer of negatively charged organic ligands, which must be engineered to present a positively charged region. The coupling between the two entities is electrostatic and not covalent as in methods (A1) or (A3). This technique was developed by Mattoussi et al. [32]. (A3) The idea here is to form a protective silica shell to encapsulate the quantum dot, using thiol silanes. This approach was proposed by Alivisatos et al. [25, 26, 44]. **(B)** Methods based on the modification of phosphine ligands, used up to the present time by Bawendi et al. [30]. **(C)** Methods using the adsorption of amphiphilic polymers. (C1) The idea is to adsorb a polymer on the particle coated with TOP/TOPO ligands. This polymer has hydrophobic parts which interact with the hydrophobic chains of the TOPO ligands and hydrophilic parts for solubilisation in an aqueous medium, as well as groups for grafting on biomolecules. This technique is commercialised in particular by Invitrogen [45], and this kind of particle has been used to target tumours in small animals [41]. (C2) Rather than use an organic polymer, Dubertret et al. used amphiphilic phospholipids, which, after adsorption onto the TOPO ligands surrounding the quantum dot, are able to self-organise into micelles to encapsulate the nanoparticle [40]

is no leakage of heavy elements. Such leakage could be triggered by oxidation of the nanocrystal, notably by TOPO ligands, caused by oxygen or UV radiation [43]. Another study has shown that quantum dots are toxic for cells at

doses of 0.4 mg/mL, and that cell growth is hindered above about 0.1 mg/mL, even though no toxicity is observed for at least 7 months at these concentrations [42]. This reveals the fact that quantum dots are not stable in the long term in an acidic medium, conditions which could well be encountered in certain organs and vesicles [42]. The chemistry involved in functionalising quantum dots must therefore take into account this further constraint in order to avoid any leakage of heavy elements from the nanocrystal, and this in the long term in the case of in vivo administration.

Another point regarding their use as biological labels is their size, which is often greater than 10 nm after functionalising by a biomolecule. Any labelling will therefore modify the targeting or internalisation properties of the functionalised biological ligand in cells or tissues. Organic fluorophores are smaller than quantum dots and thus have less effect on these properties.

Applications. Owing to their remarkable optical properties [16–20], these new materials are more and more often used to label cells [17–21, 26, 35–37, 42, 46, 47], for in vivo fluorescence imaging of small animals [17, 33, 34, 38–41] (see Sect. 7.3), and for in vitro analysis [16, 19, 20, 27, 31]. An example application for multiplexed in vitro analysis will be given on p. 277, and an example of use in vivo for fluorescence imaging of a small animal in Sect. 7.3.4.

Doped Rare Earth Nanocrystals

Other nanocrystals made from inorganic oxides have been described as light-emitting labels for in vitro biological applications.

Structure. Inorganic oxide nanocrystals of submicron size (typically in the range 0.1–0.5 μm), are rare earth oxides or oxysulfides, such as yttrium oxide, vanadium oxide, gadolinium oxide, etc. The rare earths are strictly speaking the oxides of the lanthanide group, elements with atomic numbers between 57 and 71, from lanthanum to lutetium, but yttrium (Y) and scandium (Sc) are often associated by virtue of their similar chemical properties. To give them optical properties, these oxides are doped with absorbing and emitting ions, mainly lanthanide cations. These nanocrystals thus have very similar optical properties to those described above for the lanthanide complexes (see p. 267).

Optical Properties. Most optically interesting systems operate with metallic cations of rare earths, such as yttrium for the absorber cation, and terbium, erbium, or europium for the emitter ion [48–50]. If these ions are encapsulated in an oxide matrix, nanocrystals have emission properties that are independent of the outside environment (buffer, temperature, etc.), and a low photobleaching rate. In addition, since adsorption and emission properties are due to different dopants, these crystals have a large Stokes shift (analogous to what happens with the lanthanide complexes, as described on p. 267). Their luminescence lifetime is also long, of millisecond order, which means that a time-delayed acquisition mode can be used to get round the problem of autofluorescence by the medium (see p. 267). These materials also have very narrow emission lines [48–50].

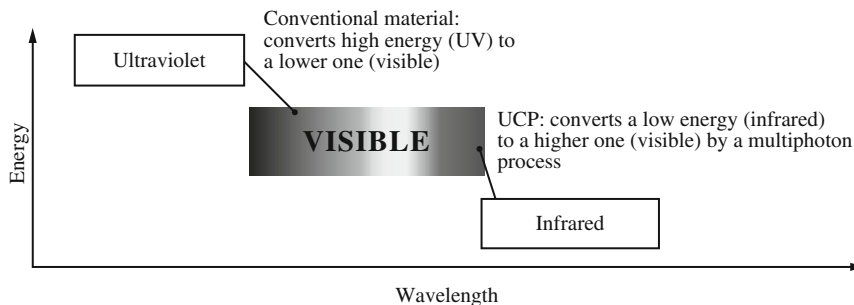


Fig. 7.17. Energy transfers in conventional materials and up-converting nanocrystals. Taken from [49]

Oxide nanocrystals are generally classified as either down-converting or up-converting. Oxide nanocrystals which absorb light at a shorter wavelength than that at which they emit are said to be down-converting (as happens with a conventional fluorophore). Oxide nanocrystals called up-converting phosphors (UCP) absorb low-energy (infrared) photons and emit higher-energy (visible) photons [48–50] (see Fig. 7.17). The origin of luminescence in up-converting or UCP nanocrystals is a multiphoton process [48–50]. By virtue of this property, up-converting nanocrystals can be used to solve the problem of autofluorescence by the buffer, which always occurs at a longer wavelength than the excitation wavelength. This leads to a very good detection sensitivity [49, 51–53].

Functionalisation. The surface of an untreated oxide nanocrystal is covered with oxides, fluorides, or oxysulfides. These ‘naked’ surfaces cannot be anchored to biomolecules by means of conventional chemistry. The methods described in the literature [49, 51–53] for functionalising this type of nanocrystal either encapsulate the nanoparticle by a silica shell and then functionalise this shell by silanisation, or else use non-specific adsorption of a multifunctional polymer at the nanocrystal surface. The organic layer thus obtained is then used to anchor biomolecules such as proteins or DNA, whence these labels can be used for the desired analytic detection.

Applications. SRI International is a company that has considerably invested in the technology of up-converting oxide nanocrystals (up-converting phosphor technology or UPT) for in vitro analysis and diagnosis [49, 51–53]. Functionalised nanocrystals are mainly used for strip tests on body fluids (urine, saliva, blood, etc.). The detection sensitivity obtained with UPT seems to be better than with conventional probes using organic fluorophores [53]. This type of test can be used for various applications: to detect antigens in cell membranes and tissues [49], to detect proteins or nucleic acids [51], and more generally to test for food or environmental contamination, and to carry out veterinary diagnoses (detection of a virus or infection).

Silica or Polymer Particles Encapsulating or Grafted onto Fluorophores

For some applications such as those described above, it may be desirable to have available not just a single fluorescent label, but an entity carrying different labels (to produce a series of different optical fingerprints) or the same label but in higher concentrations in a single micro- or nanostructure (to amplify the emission signal). For this purpose, one can use polymer or silica nanoparticles encapsulating or grafted onto the light-emitting labels discussed above, viz., organic fluorophores, lanthanide chelates, and quantum dots.

Structure and Functionalisation. These nanoparticles fall into two main families:

- Polymer nanoparticles, usually of micrometric dimensions (1–5 μm), comprise a matrix of an organic polymer such as polystyrene, poly-methyl-methacrylate (PMMA), dextran [54], etc. Fluorophores can be trapped in the matrix during synthesis, or later by swelling the polymer and allowing the fluorophores to diffuse inside [27–29, 54]. Another method is to polymerise a monomer that has already been labelled by an organic fluorophore. The biomolecules with which the particles are to be functionalised are then anchored on using the functional groups carried by the polymer [54].
- Silica nanoparticles are obtained by modified Stöber synthesis. This involves hydrolysing alkoxysilanes in the presence of the fluorophore to be encapsulated in the silica matrix, by creating a microemulsion in the presence of a surfactant, in order to control the size of the resulting particles (the size can be anywhere in the range from 20 nm to 1 μm) [54, 55]. This synthesis is easy to implement and the procedures for anchoring onto silica are well known. This means that the subsequent functionalisation of the particles obtained by organic ligands on which biomolecules can be anchored is well understood.

Optical Properties. These nanoparticles have the optical properties of the molecules which they encapsulate or which are anchored onto their surface. However, they generally have the advantage of a very high photostability compared with the fluorophore they encapsulate. Indeed, the fluorophore has less contact with oxygen dissolved in the medium, which is often the main cause of photobleaching observed with organic fluorophores [56–58]. In addition, when a biomolecule is functionalised by such a particle, it can be labelled by more than one fluorophore, thereby enhancing the signal and also the detection sensitivity. Furthermore, for single-molecule detection, such a particle combines the advantages of good detection sensitivity with low photobleaching, while avoiding the blinking problem that occurs with quantum dots.

Applications. These nanostructures are generally found in detection or complex imaging systems. We shall consider two applications: substrates identified by an optical bar code for multiplexed in vitro analysis, and the combination of an imaging function and a drug delivery function for nanomedical purposes.

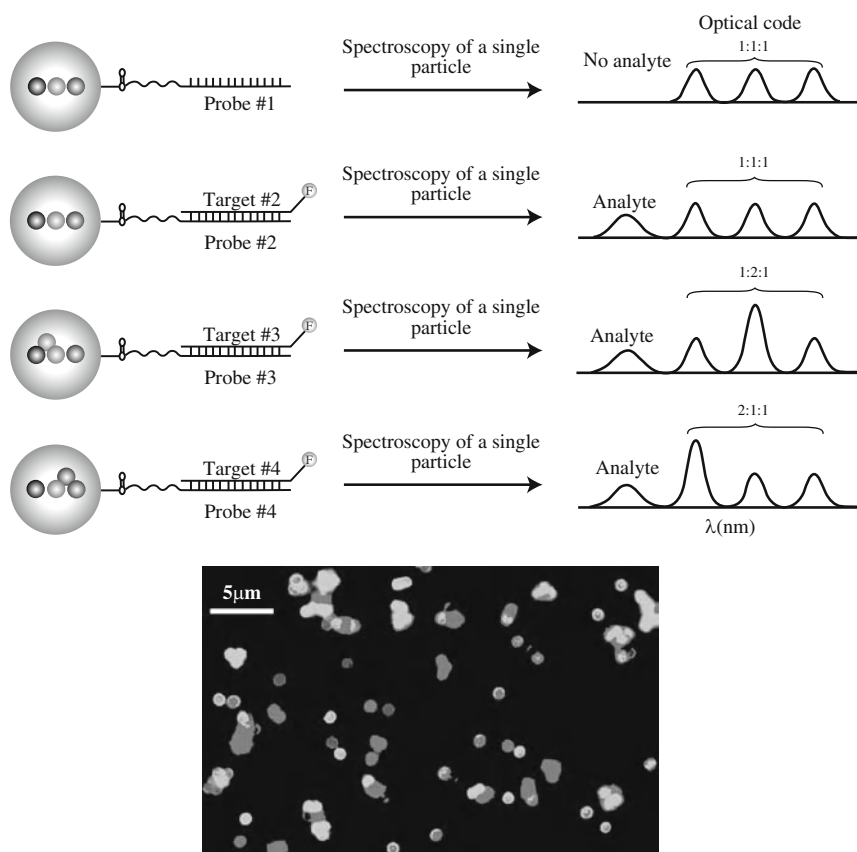


Fig. 7.18. Multiplexed analysis. For the case presented *at the top*, the polymer bead is labelled by quantum dots of three different colours and the targets by fluorescein. The system is used for recognition of DNA sequences [27]. The *bottom photograph* obtained by fluorescence microscopy shows the range of emission colours (grey-shaded here) which can be obtained using quantum dots emitting five different colours [27]

Several groups have studied the possibility of identifying substrates by an optical bar code for the purposes of multiplexed analysis [27–29, 55]. The underlying idea of this kind of analysis is illustrated in Fig. 7.18. Nanoparticles are labelled in controlled amounts by various fluorophores with distinct emission properties. One then obtains a library of beads, each one having its own optical bar code or spectral fingerprint. Each bead is functionalised by a different probe biomolecule (an oligonucleotide sequence in the case of Fig. 7.18). The specimen to be analysed (the target) is labelled by a fluorophore F emitting in a different range of the optical spectrum to the one used for the optical fingerprint of the supporting bead. After target–probe molecular recognition, the beads are examined one by one. For each probe,

identified by the optical bar code of the particle onto which it is anchored, the fluorophore F informs as to whether there has been recognition with the complementary target. The narrowness of the emission peaks of quantum dots (the possibility, for example, of having many different spectral fingerprints in the range 550–800 nm), and the fact that they can all be excited at a common wavelength (see Fig. 7.15) make them perfect candidates for this application.

Nanoparticles labelled by a fluorophore can also be used to combine the function of fluorescent label with other functionalities, which are easily anchored onto a polymer or silica substrate. We can see here the full potential of such complex structures for a whole range of different applications in biology. For instance, Prasad and coworkers are developing fluorescent silica nanoparticles, functionalised by DNA. Using these structures, one can monitor by fluorescence the transfection of genes within cell nuclei [59]. The long-term objective is to make nanoparticles that are able to deliver therapeutic molecules to localised regions which may be identified by fluorescence (nanomedical application).

Metal Nanoparticles

Although these nanoparticles are not strictly speaking fluorescent labels, they nevertheless possess quite remarkable light-scattering properties, which can be exploited for optical analyses in biology. We shall thus give a brief description here.

Structure. The metal nanoparticles used for their optical properties are mainly gold or silver nanoparticles. Today they can be produced with novel shapes and a range of dimensions, and this in a well-controlled manner (see Fig. 7.19) [60].

Optical Properties. In contrast to quantum dots, whose optical properties are due to the creation of a single electron–hole pair under the effect of light excitation (see p. 270), the optical properties of metal nanoparticles are due to a collective motion of the electrons in the plasmon band of the metal under the effect of light. These metal nanocrystals do not have particular absorption or emission properties. However, they do have quite remarkable light-scattering properties when excited by light of a wavelength that can couple with the plasmon band of the metal in the nanoparticle. The incident intensity can thus be amplified by several orders of magnitude. The position of the plasmon band depends sensitively on the size and shape of the nanocrystals, as shown in Fig. 7.19 [61]. These variable properties depending on the size and shape can be exploited for the purposes of biological analysis, as will be explained below.

Functionalisation. The functionalisation of gold nanocrystals by organic thiol molecules is well understood. The only problem is the low resistance of this functionalisation to reducing conditions.

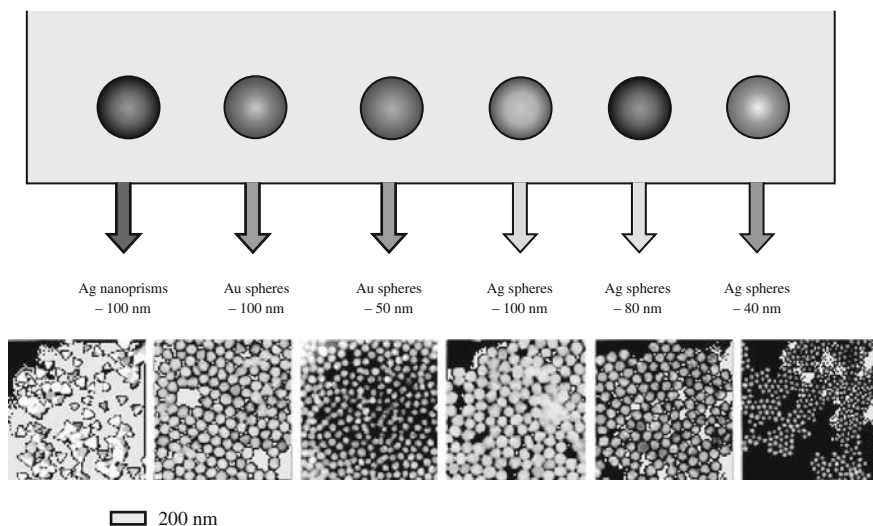


Fig. 7.19. Shapes and sizes of different metal nanoparticles, inducing different colours (different light-scattering properties) [28]

Applications. The two applications discussed here concern target–probe molecular recognition and the use of these nanostructures to enhance optical signals.

Mirkin and coworkers have carried out a great deal of work on ways of exploiting the shift in the plasmon band of these particles (and hence the change in colour of solutions) when their size increases, or when two particles are close enough together to interact. Hence, the molecular recognition between two entities, e.g., the hybridisation of two complementary DNA strands, can bring two metal particles close together, whereupon the coupling between their plasmon bands will modify the colour of the solution [62].

Other work concerns the use of these structures to enhance fluorescence signals, e.g., work by Lakowicz et al. [63], or Raman scattering by molecules adsorbed onto their surface (surface-enhanced Raman spectroscopy SERS). The adsorption of molecules at the surface of these particles means, in the case of SERS, that they can be given a spectral fingerprint that can be analysed [64].

Conclusion Concerning Exogenous Probes

In the last few sections, we have discussed many probes with very different chemical structures. Consequently, these probes have different fields of application, even though there is some overlap.

Hence, the lanthanide chelates and doped oxide nanocrystals are choice tools for *in vitro* analysis. By virtue of their special optical properties (high Stokes shift, long emission lifetime, possibility of up-conversion for

nanocrystals), they can provide interesting and innovative solutions for biochemical analysis. Indeed, the up-conversion process and long emission lifetime of these structures allow their emission to be detected with very great sensitivity, thanks to a considerable reduction in measurement noise, provided that suitable spectral and time domains are selected.

The organic fluorophores, with their extremely flexible chemistry and the very complete range of products on offer commercially, cover all wavelength ranges and all types of fluorescence application. Our good understanding of the relevant fluorescence quenching processes makes these choice systems for molecular recognition and cell dynamics applications, using techniques such as FRET, FLIM, etc., discussed in Sect. 7.5.

Quantum dots are highly innovative materials with unique optical properties. However, they have nanometric dimensions and can thus perturb measurements when functionalised by a small molecule whose location within a cell, or whose biodistribution in a small animal, is to be imaged. In this case, it may be preferable to label by an organic fluorophore of much smaller dimensions. This drawback, intrinsic to this kind of structure, means that they remain complementary labels to the organic fluorophores, which they cannot replace for all applications.

Polymer or silica nanoparticles carrying several fluorophores are useful for more complex applications in which several different types of information must be brought together in the same entity, i.e., different pieces of optical information to constitute a spectral fingerprint, or optical data combined with other functionalities.

Metal nanoparticles are not strictly used as luminescent labels, but rather as signal enhancers, by exploiting their unique scattering properties.

7.2.3 Endogenous Probes: Reporter Genes

A reporter gene is a gene coding for an easily detectable protein, artificially integrated within the genome of cells of another species and able to express itself by producing a protein whose presence indicates its expression. In the case we are concerned with here, the presence of the protein will be detectable by an optical method. The detection of reporter genes by optical imaging is an extremely powerful tool for studying gene expression *in vivo* in cells and tissues of small laboratory animals. Applications, often complementary to the use of exogenous luminescent labels described in the last section, are many and varied, concerning the *in vivo* study of gene expression and its control by natural, environmental, pharmacological, or other factors, as well as the monitoring of specific cell types (tumour proliferation, grafts, immune and inflammatory reactions, etc.) and infectious species (viruses, bacteria, parasites).

Note that the qualifier ‘endogenous’ for these probes is not strictly correct, since they are introduced artificially by genetic manipulation. It is used here to stress the fact that, in contrast to the exogenous probes described in the

last section, these probes are proteins depending on the protein-expressing machinery of the very organism or cell in which they are introduced.

Constructing a Reporter Gene System

It is important to understand that the present section makes no claim to describe the methods of molecular biology used to introduce and express a gene in a biological system, but simply to outline this field rather briefly and in a schematic way. The interested reader is referred to the many standard textbooks on molecular biology which discuss this subject in more detail.

Definitions

A gene comprises a coding sequence and other sequences which control and regulate its expression. In the construction of a reporter gene system, the coding sequence of the exogenous reporter protein is introduced under the control of regulating sequences, which are either exogenous (foreign to the receiving cell or organism) or endogenous (the same as the regulating sequences for the expression of a gene in the receiving organism).

In the first case, the reporter gene is usually placed under the control of elements regulating genes with strong constitutive expression, in order to produce a signal that is independent of the systems controlling expression in the receiving organism. This type of construction can track circulating or proliferating cells, infectious agents, etc. There is a whole range of systems of this type implementing constitutive gene expression; one of the best known is the promoter for expression of the cytomegalovirus (CMV).

In the second case, on the other hand, since the reporter gene is under the control of regulatory elements of a gene that is naturally present in the cell, its expression will depend on the same control factors as that gene. The presence of an optical signal will then reflect the fact that expression of the endogenous gene has been triggered under the influence of the various aspects of the environment, i.e., chemical, cellular, pharmacological, etc. This system is invaluable for exploring the mechanisms triggering or inhibiting expression of the relevant gene. The coding sequence for the reporter gene can either replace or be added to the coding sequence for the relevant gene.

Figure 7.20 depicts an example of the reporter gene system, in the case where the reporter gene is bound to the gene under investigation and expressed under the control of its promoter. Expression of an optical reporter gene is revealed by the production of light by a bioluminescence or fluorescence mechanism. In the case of a fluorescent protein like the green fluorescent protein (GFP), illumination at the appropriate wavelength will generate blue light that can be detected by a sensitive detector (see Fig. 7.20a). In the case of a bioluminescent protein like luciferase, light production is obtained after injecting the substrate, whose transformation by the enzyme activity of the luciferase produces photons (see Fig. 7.20b).

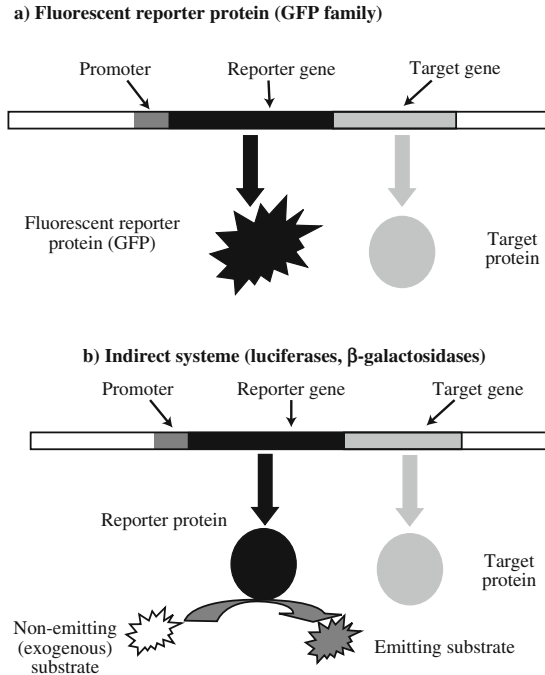


Fig. 7.20. Visualising the expression of a gene by means of a reporter gene. **(a)** The gene under investigation is associated with a fluorescent reporter gene coding for GFP. Detection of the fluorescent reporter gene (blue–green illumination) then reveals the relevant protein. **(b)** The gene under investigation is associated with a reporter coding for a luciferase, a protein involved in the enzymatic transformation of an exogenous substrate which is injected. The enzyme reaction leads to light emission

Methods of Construction: From Genes to Transgenic Animals

Cell Transfection. The reporter system can be integrated by transfection in cells that can subsequently be monitored *in vivo* in an animal after xenograft. This approach is widely used in preclinical cancer research (detection of metastases, therapeutic trials, etc.), because it provides an easy way of monitoring tumour growth [65]. It is also used to develop methods of cell and gene therapy, and to carry out research on infectious diseases. In the latter case, the reporter gene is introduced into the infectious agent, which is then used to inoculate a laboratory animal.

Rather schematically, the cell transfection technique can be broken down into four or five steps:

- The first step is to isolate the relevant gene. The method depends on the form in which the gene is available, i.e., plasmid, or bound to a promoter.
- The second step is to carry out and amplify the gene construction in such a way that the gene can express itself once introduced into the host cell.

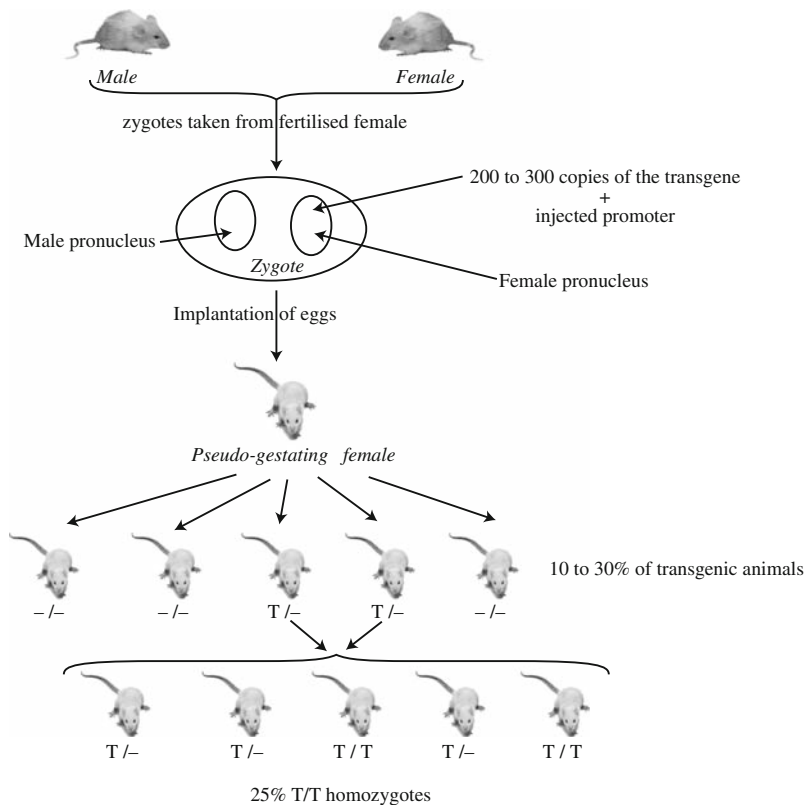


Fig. 7.21. Example of transgenesis in mice by pronuclear injection

This involves adding regulatory elements and if necessary a viral promoter, eliminating sequences irrelevant to expression, and so on.

- Once the gene construction has been achieved, there are several ways of introducing it into the cell. One can use a viral or bacterial vector, carry out homologous recombinations or liposome fusion, or use chemical transfection agents, cell stress, or electroporation.
- Several things must then be checked. (1) One must check transfer efficiency, easily achieved by optical imaging for light-emitting reporter genes. (2) One must check that transfection of the reporter gene has not had harmful effects on the phenotype of the cell in which it has been introduced. It is sometimes difficult to choose a suitable method for checking this second point.

Depending on the situation, the population of transfected cells can be used as it is, or after subcloning by the limiting dilution technique. The second approach takes longer because it introduces a further step, but it means that one can choose those cells with a good level of reporter gene expression and if necessary eliminate those with unfavourable phenotype.

Table 7.1. Enzyme–substrate combinations used in bioluminescence

Organism	Enzyme	Substrate	Cosubstrate/catalyst
Jellyfish (<i>Aequorea victoria</i>)	Aequorin	Coelenterazine	Ca ⁺⁺
Firefly (<i>Photinus pyralis</i>)			
Bacteria	Luciferase	Luciferin	ATP
Dinoflagellates			
Sea pansy (<i>Renilla reniformis</i>)	Luciferase	Coelenterazine	None

Transgenic Animal. It is a long and delicate matter to create a transgenic animal expressing a reporter gene, but the result can be extremely useful [66]. Indeed, it means that the product of the expression of the reporter gene can be tracked exactly as though it were the protein coded by the latter.

The step in which the reporter gene, placed under the control of the regulatory elements of the relevant gene, is introduced into the animal can be achieved in two ways:

- pronuclear micro-injection of genes in solution in the embryo,
- transfer via embryonic cells.

Pronuclear micro-injection is a well understood method for certain species. This method is very powerful because the target gene can be expressed in all cells of the transgenic animal. This technique is nevertheless difficult to carry out because it involves injecting a solution containing the relevant gene directly into the nuclei of culture cells. In mice, this solution can be injected directly into a fertilised egg containing the two nuclei coming from the male and female sex cells. So one of the two nuclei receives the injection before they fuse. The embryo is then transplanted into the uterus of a pseudo-gestating receptor female (see Fig. 7.21).

For some animals, transgenic blastocytes can be prepared in vitro, and subsequently transplanted in the receptor female.

Transfer by embryonic cells uses the totipotency of precocious embryonic cells. The latter can be taken from an embryo to be cultured and subsequently receive the relevant gene. The modified cells containing the reporter gene are then reintroduced into an embryo in the early stages of development. This method cannot be used to obtain transgenic animals expressing the reporter gene in all cells, but rather chimeric animals in which only a fraction of the cells express the transgene [66, 67].

Gene Reporter Systems Using Bioluminescence

This phenomenon occurs naturally in various organisms such as fireflies, some bacteria, and marine organisms such as jellyfish and sea pansies. These organisms do not use the same mechanisms to produce light, and there are several

enzyme–substrate combinations with the ability to generate a bioluminescence signal. The main combinations giving rise to optical imaging applications are listed in Table 7.1.

It should be explained that the terms ‘luciferase’ and ‘luciferin’ are generic and refer to the enzyme and the substrate involved in the bioluminescence reaction, without specifying their structure, which depends on the organism they come from. These reactions do nevertheless share a common feature on the molecular level, because they all require the presence of oxygen, as well as the formation of an intermediate luciferase–peroxide compound. The energy required for light emission is released when the bonds of the latter compound are broken.

Only luciferase from the firefly and luciferase from *Renilla* will be discussed in this section. The aequorin–coelenterazine system is examined on p. 288, where the aequorin–GFP system is treated.

Firefly Luciferase–Luciferin

Firefly luciferase [68, 69] is a 61-kDa monomeric protein that is widely used in bioluminescent imaging. This protein catalyses the oxidation of luciferin according to the reactions shown in Fig. 7.22: activation of luciferin by luciferase in the presence of ATP allows the formation of an unstable complex that emits light when it returns to its ground state.

Application to Tumour Imaging in Small Animals. Bioluminescence imaging using the luciferin/luciferase system is well suited to the detection of cancer cells by in vivo imaging in small animals. Indeed, the implantation of cancer cells transfected with the gene Luc encoding luciferase generates a tumour which, after injection of luciferin, will produce an amount of light proportional to the number of cells in the tumour, and hence the mass of the tumour. With a suitable imaging device (see Sect. 7.3), this tumour labelling can be used to assess tumour burden and characterise the effectiveness of anticancer drugs on the tumour.

Luciferin diffuses very quickly (in a few minutes) when injected intravenously or intraperitoneally. This diffusion occurs throughout the body of the animal and the luciferin enters cells very rapidly [70]. These properties make the luciferase–luciferin pair an ideal system for optical imaging, because it provides a way of monitoring many animals that is simple and easy to implement. In addition, the fast evacuation of the luciferin reduces the latency between two in vivo imaging experiments on the same animal to a few hours. Figure 7.23 shows the kinetics of light emission when luciferin is oxidised. The decrease after the peak is sometimes slower depending on the diffusion of the luciferin in the target organism. It is important to know the exact kinetics in the animal model used, firstly to obtain the maximum efficiency of the substrate (peak or plateau) in order to achieve good contrast in the resulting bioluminescence images, and secondly to obtain comparable signal intensity measurements from one experiment to the next.

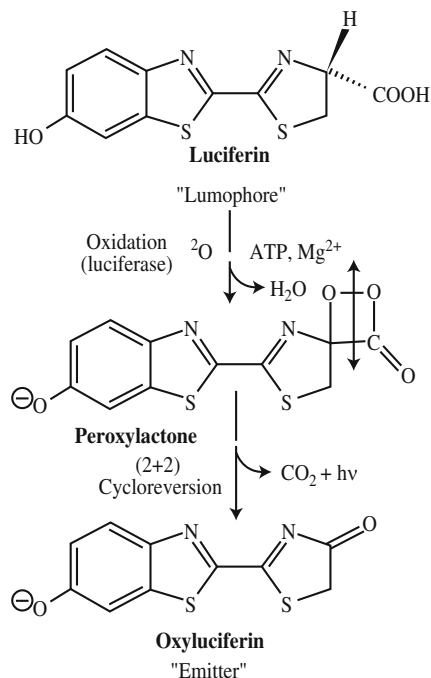


Fig. 7.22. Oxidation of luciferin by firefly luciferase. *From top to bottom:* (1) The luciferin is activated by adenosine triphosphate (ATP), involved in most energy mechanisms present in all cells. This activation occurs in the presence of Mg^{2+} ions. (2) Oxidation of the luciferin/luciferase/AMP complex in the presence of oxygen transforms the luciferin into an energetically unstable molecule carrying an oxyluciferin peroxy lactone function. (3) The unstable nature of this complex in the excited state means that it tends to drop back down to a stable state of lower energy. This transition is accompanied by decarboxylation and emission of bioluminescence photons with maximum intensity at 540 nm

Finally, no light production can occur without the presence of luciferase, luciferin, and ATP, and this eliminates any background noise of biological origins. Only instrument noise can reduce the signal-to-noise ratio.

As already stressed earlier, many other types of study on cells, tissues, or small animals can be carried out using the luciferase–luciferin system.

Renilla Luciferase–Coelenterazine

Renilla luciferase, extracted from *Renilla reniformis* (the sea pansy), is a 36-kDa monomeric protein catalysing the oxidation of coelenterazine, which produces light with an emission peak at 480 nm. This substrate also emits light by autoluminescence. Indeed, coelenterazine can oxidise in the absence of luciferase [71], and still generate light emission. This phenomenon introduces

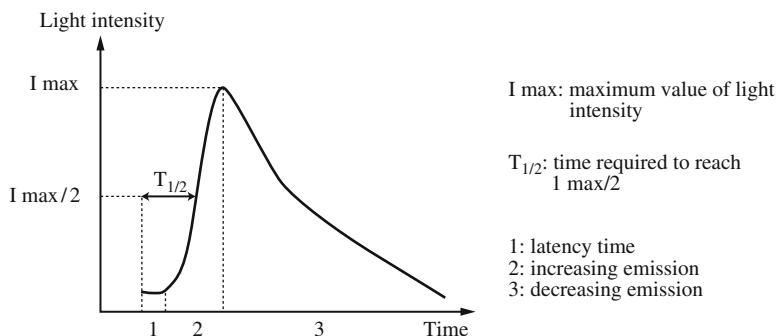


Fig. 7.23. Reaction kinetics for the oxidation of luciferin catalysed by firefly luciferase. Administration of luciferin is followed by a latency time, then an increase in light intensity. The decrease in light emission after the peak can be slower than in the example presented here. In the case of intraperitoneal injection, a plateau is observed just before the decrease in light intensity

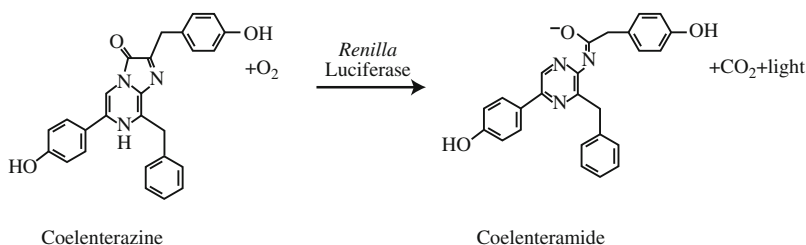


Fig. 7.24. Bioluminescence reaction catalysed by *Renilla* luciferase. The coelenterazine is oxidised by luciferase in the presence of oxygen and then forms a coelenteramide complex. The formation of this complex is accompanied by light emission

noise which reduces the sensitivity of the luciferase–coelenterazine reporter system as compared with the luciferase–luciferin system. The reaction occurring when coelenterazine is introduced into an organism expressing *Renilla* luciferase is depicted in Fig. 7.24. The luciferase plays the role of catalyst here and the reaction cannot occur in the absence of oxygen.

Systems Based on Fluorescence: GFP and Aequorin–GFP

Green fluorescent protein (GFP)

The main reporter gene application appealing to fluorescence imaging makes use of the green fluorescent protein (GFP), which produces a green light signal under ultraviolet light. First identified in the jellyfish *Aequorea victoria* [72], the GFP gene was cloned [73], then used as a reporter gene.

GFP has a molecular weight of 27 kDa and a quite remarkable architecture: the fluorophore, composed of a sequence of 3 amino acids, viz., serine–tyrosine–glycine, is housed in a cylindrical structure of diameter 30 Å and

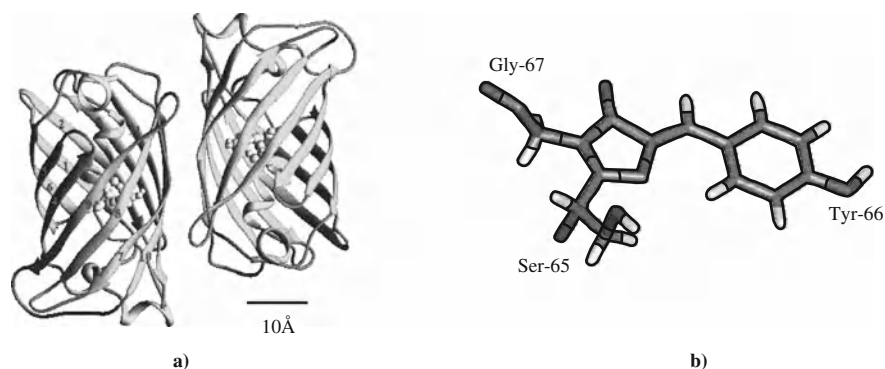


Fig. 7.25. (a) 3D structure of GFP. The central fluorophore is carried by an α -helix. This structure is located within a cylinder of 11 β -sheets. (b) Molecular structure of GFP

height 40 Å. This structure is formed from 11 β -sheets and an α -helix carrying the fluorophore (see Fig. 7.25).

Photoactivation of this protein triggers a self-catalysing reaction that causes it to fluoresce. The absorption peak of GFP lies at 470 nm, and its emission peak at 508 nm. Cyclisation between a nitrogen atom of the glycine and a carbon atom of the serine creates a chemical bond between these two amino acids. The cycle so formed spontaneously dehydrates. Neighbouring oxygen then attacks a bond of the tyrosine, which has the effect of forming a double bond between the tyrosine and the glycine. The molecule obtained is the fluorophore producing light emission. The structure of the molecules is depicted in Fig. 7.26.

Applications of GFP are many and varied. This protein can be used to target and study gene expression, to detect viral infections like HIV, or to monitor protein–protein interactions in cells by fluorescence resonance energy transfer (FRET) (see the presentation in Sect. 7.1 and detailed discussion in Sect. 7.5).

Mutations of the GFP gene [74] can be used to obtain proteins emitting at other wavelengths to the initially cloned GFP, e.g., in the red or the blue. This can be useful to carry out multiple labelling, or simply to enhance the intensity of the light emission. For example, replacing the amino acid Ser65 by Thr65 produces the mutant GFP65T, excitable at 490 nm and producing six times the intensity of GFP. Figure 7.27 shows the many excitation and emission spectra that can be obtained by modifying GFP. This protein can thus be used over a very broad spectral range.

This system nevertheless suffers from a disadvantage that is inherent in the fluorescence phenomenon: biological tissues contain fluorophores, called endogenous fluorophores, which are excited at the same time as GFP by the light source. However, it has been observed that endogenous fluorophores such

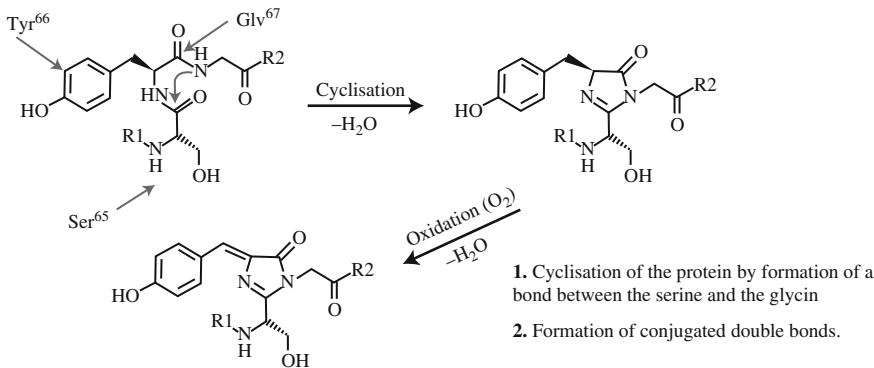


Fig. 7.26. Change in the conformation of GFP during excitation

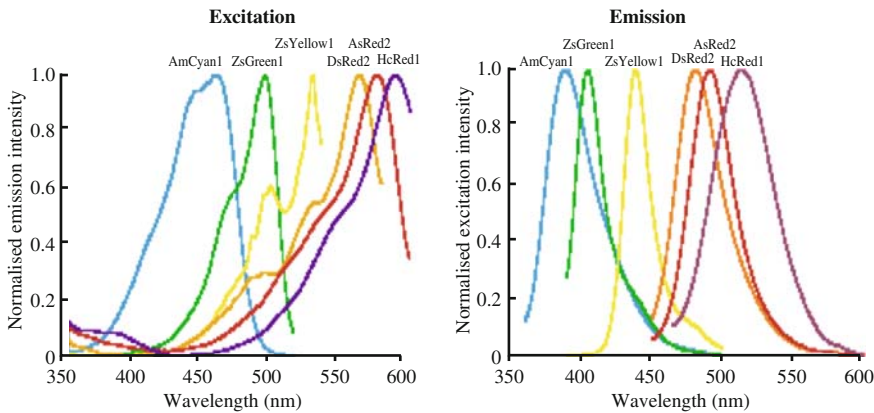


Fig. 7.27. Excitation and emission spectra of various GFP mutants

as collagen are sensitive to excitation at short wavelength, typically in the blue–green, and that autofluorescence is less intense than fluorescence [75].

Aequorin–GFP

Discovered in the jellyfish *Aequorea victoria* [75,76], aequorin is a photoprotein catalysing the production of blue light (460 nm) when Ca^{2+} and coelenterazine are present. The sensitivity of aequorin to calcium makes it a useful tool for monitoring calcium dynamics, especially for imaging neuron activity [78].

The aequorin–GFP system produces light by an energy transfer mechanism called bioluminescence resonance energy transfer (BRET), illustrated in Fig. 7.28. In the figure, so-called inactive aequorin is a complex of apoprotein, oxygen, and coelenterazine (black pentagon). When the aequorin is activated by the presence of calcium ions, it catalyses the oxidation of the coelenterazine into coelenteramide (white pentagon). The coelenteramide de-excites by

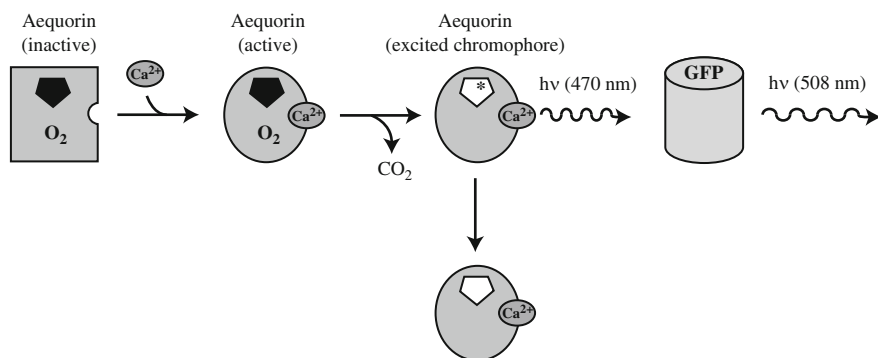


Fig. 7.28. Bioluminescence resonance energy transfer (BRET), the light production mechanism in the aequorin–GFP system

emitting light at 470 nm , corresponding to the blue light produced by purified aequorin. In vivo energy transfer between the excited state of the coelenteramide and GFP leads to green fluorescence emission by the latter [79–81].

7.2.4 Conclusion

The number and variety of the optical labels described above, together with the extraordinary range of in vivo applications they can be used for, have radically changed whole areas of fundamental and applied biomedical research over the past few years, particularly in pharmacology. It seems likely that this is just the beginning of a new approach to biology, in which biological phenomena can be monitored by non-invasive methods, thereby realising the old dream of biologists to observe the living without destroying it. However, a good understanding of the properties of these labels remains essential, whether they be endogenous or exogenous, especially with regard to their spectral characteristics (absorption and emission wavelengths), in order to optimise the associated detection systems to be described in the following sections. Finally, the spectral characteristics and behaviour of these labels remain sensitive to their environment, and in the end their use can only be fully validated in applications to living systems. Some of these properties can be handicapping for in vivo applications and restrict the domain of validity of an optical label and the equipment required for its detection. Others, however, such as sensitivity to pH or ion concentration, may turn out to be useful for making local measurements of physiological parameters in living systems.

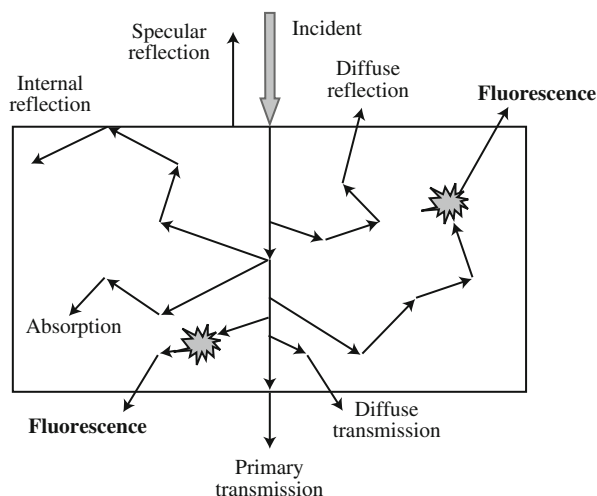


Fig. 7.29. Description of matter–photon interactions

7.3 In Vivo Detection Systems

7.3.1 Introduction to in Vivo Optical Imaging

Clinical observation always begins with a detailed visual examination of the patient. Further, microscopic observation of tissues has developed so far that events can now be observed directly on the molecular scale. The next challenge therefore consists in detecting such molecular events directly in the patient. This is the aim of molecular imaging [82, 83]. Molecular imaging provides a better understanding of mechanisms directly in vivo, e.g., monitoring of gene expression, analysis of the biodistribution of drugs, effectiveness of therapies, in a huge range of applications, such as oncology, cardiology, and so on.

Optical molecular imaging arose from the observation of two physico-chemical light-emission phenomena in certain living organisms. The first, bioluminescence, was discovered in fireflies and certain marine species. Bioluminescence imaging provides both functional and geographic information, through the specificity of molecular probes. For example, reporter gene systems can be used to monitor gene expression. The second phenomenon is fluorescence. Using fluorescent substances bound to molecules involved in some known metabolic activity (tags providing kinetic data, for example), that activity can be quantified and localised in the relevant biological tissue. It is also possible to study gene expression in a dynamic way using fluorescent reporter genes such as GFP (see Sect. 7.2).

In vivo optical imaging is a relatively recent form of molecular imaging compared with nuclear or magnetic resonance imaging (MRI). Non-invasive, i.e., not requiring tissue samples or sacrifice of the animal, in contrast to

microscope observation, it is based on the detection of low-energy photons (a few electronvolts) in the visible range 400–750 nm. The sensitivity of detection systems, their low cost, and the ease with which images can be acquired, together with the possibility of *in vivo* imaging of small animals, make optical imaging a very promising technique in the context of molecular imaging, despite limitations due to the optical properties of the biological tissues making up a complex medium. While radio-isotope imaging methods such as positron emission tomography (PET) require a cyclotron, and very strict radioprotection measures, optical imaging can be carried out with small setups and no specific protection.

It is also possible to obtain quantitative images from rather short acquisition sessions (a few minutes), owing to the high sensitivity of detection devices. Non-invasive *in vivo* imaging methods represent significant scientific progress in the field of medical research on small animals. From an ethical standpoint, non-invasive imaging is particularly interesting because it means one can limit the number of animals used. Furthermore, each animal can be imaged several times, thus offering the possibility of carrying out longitudinal studies in time. These studies are more precise, because they get around the problem of interindividual variations inherent in an experimental protocol in which several animals are compared. Finally, the non-invasive nature and sensitivity of optical imaging setups facilitates kinetic studies, e.g., biodistribution studies.

7.3.2 Basic Principles of *in Vivo* Optical Imaging

Optical Properties of Tissues

Scattering and Absorption

There are several types of interaction between photons and matter in a complex medium such as a biological tissue (see Fig. 7.29). In order to select an optimal fluorescent or bioluminescent label, all these interactions must be taken into account.

Photons can be reflected specularly (reflection at the tissue surface and at the same wavelength as the incident light) or scattered (multiple scattering in the material followed by reflection) (see Fig. 7.29). One also observes directly or diffusely scattered photons. The ability of the medium to scatter photons is characterised by a scattering coefficient denoted by μ_s and depending on the incident photon wavelengths. μ_s is generally expressed in cm^{-1} . One can define a scattering coefficient μ'_s taking into account the anisotropy of the material by the factor g :

$$\mu'_s = (1 - g)\mu_s .$$

The factor g is the average cosine of the scattering angle and hence treats the dependence of the scattering angle on the incident direction.

Some photons can be absorbed and not re-emerge from the biological tissue (see Fig. 7.29). The absorption properties of the tissue also depend on the wavelength and are characterised by a coefficient μ_a , generally expressed in cm^{-1} . The resulting intensity I obeys an exponential law of the form

$$I = I_0 e^{\mu_a z},$$

where z is the depth travelled in the medium by a photon of initial intensity I_0 .

Figure 7.29 shows the situation for fluorescence, in which incident excitation photons generate emission photons after collision with fluorophores. For light emission by bioluminescence, where there is no external illumination but a localised emission within the tissue, the diagram is still valid provided that one ignores the incident photon paths.

Consider the energy transferred by the photons as they move in a direction specified by the solid angle Ω and in a small volume element centered on the position vector $\mathbf{r}$. This energy transfer can be described by the energy radiance $L(\Omega, r, t)$ in units of $\text{W m}^{-2} \text{sr}^{-1}$.

Scattering in the medium can be isotropic or otherwise. In the case of non-isotropic scattering, one must take into account the probability $d\mu_s(\Omega', \Omega)$ that a photon is scattered from a direction Ω' into a direction Ω . This phenomenon is described by bringing in the normalised scattering phase function:

$$f(\Omega', \Omega) = \frac{d\mu_s(\Omega', \Omega)}{\mu_s}.$$

The energy balance for an infinitesimal tissue volume is

- loss by energy radiance $L(\Omega, r)$,
- radiance lost by scattering,
- radiance gained by scattering from Ω' to Ω ,
- gain from excitation source.

The variation of energy radiance in a volume element is given by

$$\frac{dL(\Omega, r)}{ds} = \sum \text{gain} + \text{losses}.$$

The photon paths in the material can be described by Boltzmann's radiative transfer equation [84] for the above energy balance:

$$\Omega \Delta L(\Omega, r) + (\mu_a + \mu_s) L(\Omega, r) = S(\Omega, r) + \mu_s \int_{4\pi} f(\Omega', \Omega) L(\Omega', r) d\Omega',$$

where L is the energy radiance, a function of the position r of the photon and its direction of propagation Ω , and S represents the source term. This equation is used to model the behaviour of photons in matter, and thus lies at the heart of many reconstruction algorithms (see p. 297).

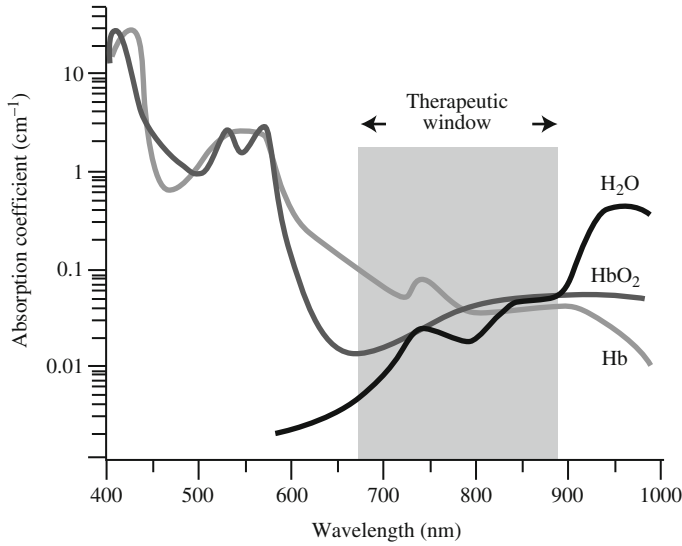


Fig. 7.30. Absorption spectra of water and hemoglobin, the main absorbers in typical tissues [86]

Order of Magnitude of Tissue Optical Parameters

The capacity of a biological tissue to absorb and scatter light will depend on the wavelength of the photons which collide with tissue particles, but also on the nature of the tissue material itself. Hence, highly vascularised hepatic tissue tends to absorb light, while bones behave rather as reflectors and other tissues scatter light.

The main absorbers in biological tissues are hemoglobin and water (see Fig. 7.30). Water absorbs a lot above 900 nm. Hemoglobin, present in the two forms Hb and HbO₂ (oxygenated form), absorbs below 650 nm. The range 650–900 nm, known as the therapeutic window, is where the absorption coefficients in tissues are at their lowest, whence it is the best range for photons to penetrate.

Average values for the optical coefficients are given in Table 7.2. The two coefficients have rather different orders of magnitude: in the near infrared, μ_a lies between 0.01 and 0.5 cm⁻¹ and μ'_s lies between 2 and 10 cm⁻¹. The table also gives some indication of the variation of these coefficients in going from one individual to another, or in going from one type of tissue to another in the same individual. Note the large variations due to the fact that the different measurements were carried out by different authors using different systems. This gives an indication of the relatively low confidence level that can be attributed to such measurements at the present time.

Table 7.2. Optical coefficients for in vitro tissues [87]. D_{SD} is the source–detector separation

Authors	Tissue	Species	λ [nm]	μ_a [OD cm ⁻¹]	μ'_s [OD cm ⁻¹]	D_{SD} [cm]
Kurth et al. (1995)	Head	Piglet	670, 830	0.10, 0.07	3.5, 3.3	3
Fantini et al. (1999)	Head	Piglet	758, 830	0.07, 0.06	4.0, 3.6	1.5–3
Hueber et al. (2001)	Head	Piglet	758, 830	0.07, 0.06	3.0, 2.6	1.5–3
Doornbos et al. (1999)	Forehead	Human	633, 700	0.04, 0.009	7.3, 6.7	0.18–1.68
	Arm			0.07, 0.04	4.0, 3.5	
	Foot			0.03, 0.02	4.9, 4.1	
Matcher et al. (1997)	Head	Human	800	0.07	4.1	4
	Forearm			0.10	3.0	
	Calf			0.07	4.08	
Toricelli et al. (2001)	Head	Human	610–1,010	0.07–0.16	7.1–4.3	2
	Abdomen			0.009–0.09	6.1–3.5	
	Arm			0.09–0.22	5.3, 2.2	
Cubeddu et al. (1999)	Arm	Human	672, 818	0.12, 0.11	4.8, 3.9	2
Farrell and Patterson (1992)	Forearm	Human	610–700	0.09–0.01	2.9, 2.5	0.1–1

Fluorescence and Bioluminescence

Comparing the Two Phenomena

As we have already seen, fluorescence only occurs following external light excitation, in contrast to bioluminescence, where emission occurs endogenously in the presence of a substrate. External illumination provides a control parameter, viz., the excitation wavelength, and hence indirectly a control over the emission, whereas the emission spectrum for bioluminescence is fixed and remains limited commercially. However, this advantage is complicated by the fact that external illumination excites fluorophores that are naturally present in the tissues as it passes through to the target. There is therefore an autofluorescence effect limiting the signal-to-noise ratio.

In addition, the fluorescence intensity depends directly on the excitation wavelength: the higher the absorption intensity at the excitation wavelength, the greater the emission intensity will be. It is thus important to ascertain the optimal excitation conditions, to which an appropriate detection system must

be coupled. If the reflected excitation signal interferes with the fluorescence signal, it must be eliminated. This is achieved by high-pass optical filtering.

Bioluminescence imaging on the other hand can be used to acquire low-noise images, since there is no light emission in the absence of the optical probe. However, it is less flexible than fluorescence imaging, which offers a wide choice of molecules that can be used in the same animal.

In both cases, the emitted signal is in principle proportional to the local concentration of labelled molecules. However, the complexity of the matter–photon interaction makes it difficult to quantify data, because several factors affect the intensity of the detected signal.

Choosing an Effective Label

The main spectral constraints when choosing a labelling system were outlined on p. 293 ff (see also Sect. 7.2). There are two main criteria determining the labelling strategy, depending on the spectral properties and biological environment:

- The first determining factor is the object to be studied. Indeed, the choice of fluorophore as exogenous label or the choice of a reporter gene system will depend on the studies carried out, as well as any specificity of the label with regard to the relevant organ. This specificity is generally difficult to devise and will constitute the main constraint on the type of labelling used. Pharmacokinetic and pharmacodynamic studies, and high-speed molecular screening will preferably be carried out using fluorescent exogenous labels, while reporter genes will be reserved for monitoring tumours or investigating gene expression.
- The second key parameter governing the choice of emission wavelength of the label will be the tissue depth to be reached. Quite generally, as discussed on p. 293 ff, the more the label emits or absorbs in the near infrared region of the optical spectrum (650–900 nm), the greater will be the depth at which tissues can be visualised. Unfortunately, bioluminescent labels offer little choice with regard to emission wavelengths, being mainly situated in the blue–green. Several types of modification are under investigation in order to devise systems with spectral characteristics shifted toward the red. For example, the replacement of an amino acid in the chromophore of GFP produces mutants emitting in the yellow and red.

Limitations of in Vivo Optical Imaging

The orders of magnitude given above clearly indicate that only rather shallow structures can be imaged by this technique, and that consequently it cannot apply to imaging the whole human body. On the other hand, these techniques are well suited to studies of small animals like rats and mice, so it is possible to carry out preclinical studies on the biodistribution of molecules, the efficiency of drugs, toxicity, and so on.

For human beings, these techniques can be used to image surface tissues, either externally accessible, or accessible by endoscopy. Hence, applications to mammography, dermatology, and examination of the buccal cavity (direct observation), or again to the detection of tumours in the uterus and oesophagus, or during surgical operations (by endoscopy) can all be carried out on human beings.

7.3.3 Experimental Setups for Fluorescence and Bioluminescence Imaging (Continuous Irradiation)

Imaging apparatus can be classified into two categories: plane imaging setups producing 2D images, and tomographic setups with multiple-view acquisition providing data in three dimensions after 3D reconstruction (see p. 299). The first example given here concerns a 2D imaging setup exploiting either bioluminescence or fluorescence.

Typical Setup

We discuss here the example of an *in vivo* optical detection system designed to study small animals, although optical mammographic devices have already been developed by various groups [88]. Figure 7.31 illustrates the basic setup for bioluminescence imaging (BLI) and fluorescence reflectance imaging (FRI).

The animal is anaesthetised and placed in the dark room. For fluorescence images, excitation is achieved by an external source that is filtered in such a way as to adjust the wavelength to the excitation wavelength of the relevant fluorophores, then scattered over the whole field of view (FOV). In contrast, bioluminescence imaging does not require an external source of excitation. The light emitted by the animal is gathered by an objective then transmitted to a CCD camera. Intensified or cooled CCD cameras are generally used in order to minimise noise and increase sensitivity, especially when detecting bioluminescence signals, which are weak. For fluorescence imaging, the light emitted by the animal is filtered by a high-pass filter located in front of the objective in order to eliminate the reflected excitation signal, which is very strong and would swamp the fluorescence signal.

Commercially Available Systems

Table 7.3 lists the various commercially available optical imaging systems. The Xenogen, Biospace, Bertold, and Hamamatsu systems are designed particularly for bioluminescence imaging (BLI), whereas the other systems are devoted rather to fluorescence imaging (FLI). In the quest for a compromise between spatial resolution, depth resolution, and sensitivity, these machines do not offer the same performance or the same scope for applications.

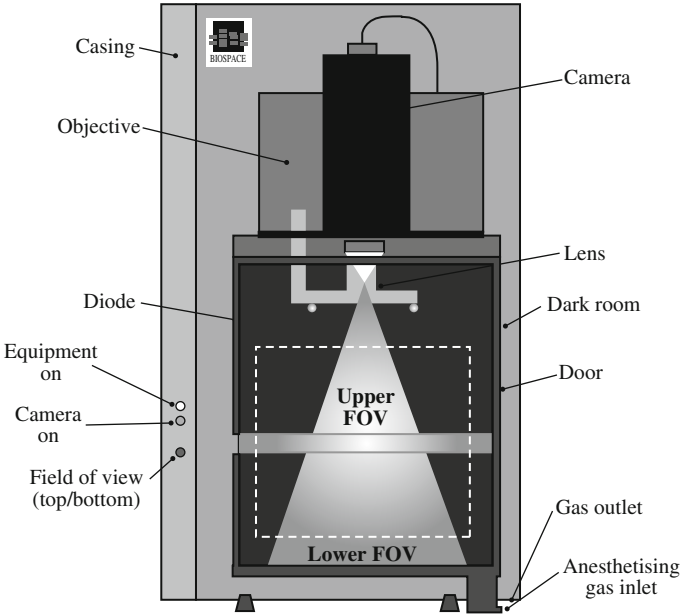


Fig. 7.31. Setup for in vivo bioluminescence imaging (BLI) and fluorescence reflectance imaging (FRI) of small animals

Table 7.3. Commercially available systems. FLI = fluorescence imaging, BLI = bioluminescence imaging

Company	System	Imaging mode
Xenogen	IVIS	BLI (+FLI)
Biospace	Photon Imager	BLI (+FLI)
Berthold	Night Owl	BLI (+FLI)
Hamamatsu	Aequoria	FLI (+BLI)
ART (GE Healthcare)	eXplore Optix	FLI
Kodak	Serie IS	BLI+FLI
Siemens	bonSAI	FLI
CRI	Maestro	FLI

Toward Optical Tomography

Imaging systems are generally divided into two categories: topographic systems and tomographic systems. The systems discussed earlier belong to the first category, because they produce surface images with little information about deeper levels. The term ‘tomography’ is used to describe systems that produce 3D images from measurements made at different points on the surface of the object.

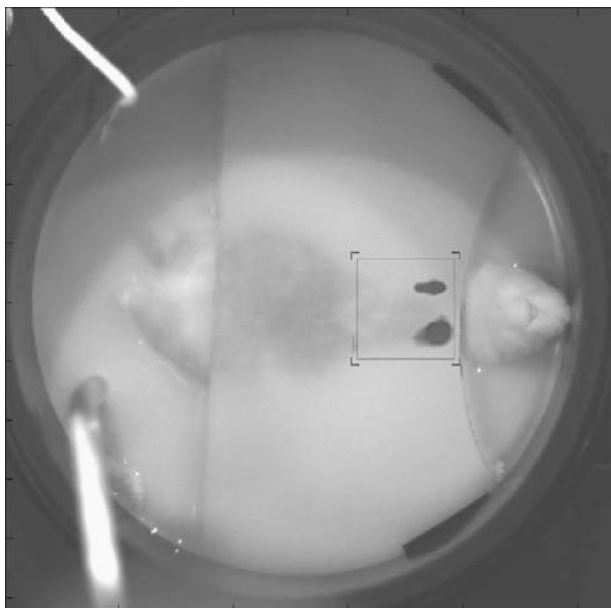


Fig. 7.32. Tomographic reconstruction of lung tumours in a mouse (CEA Grenoble, France). Projection of the reconstructed volume superposed on a white light image

The aim of these measurements is to solve the inverse or reconstruction problem (see Fig. 7.32). Starting from a set of measurements corresponding to a source–detector combination, the idea is to reconstruct a 3D absorption, scattering, or fluorescence map of the object which can distinguish between healthy and tumoral tissues, for example. Optical tomography has been developed for examination of small animals [89]. Research is being carried out on reconstruction algorithms to improve the spatial resolution of these setups.

7.3.4 Applications of Fluorescence and Bioluminescence Imaging

Detecting Lung Tumours in Mice by a (Bioluminescence) Reporter Gene System

The aim is to monitor tumour growth in nude mice. This is a hairless mutant of the mouse, which facilitates optical imaging. All bioluminescence studies have been carried out with female nude mice. These are homozygous mutants for the recessive gene *Nu*. As this gene is involved in thymogenesis, nude mice do not develop a thymus and thus have a deficiency of T lymphocytes which allows implantation of xenografts.

The cells used are PC12 MEN-2A cells which express the human proto-oncogene RET 2A. These cells are transfected in a stable way by the gene Luc cloned in a eukaryotic expression plasmid (pCDNA3+). Double insertion

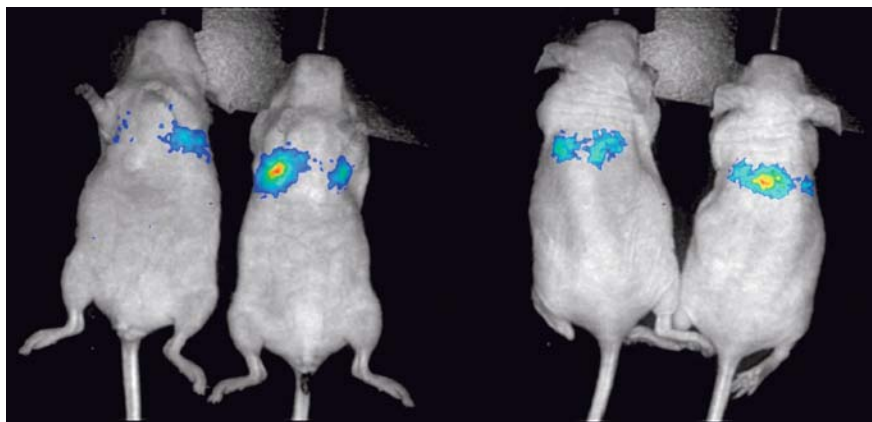


Fig. 7.33. Two mice with lung tumours resulting from injection of tumour cells expressing luciferase. Bioluminescence images were made by injecting luciferin via channel IV and carrying out acquisition for several minutes

of the oncogene and luciferase leads to the development of tumours expressing luciferase. Injecting these cells intravenously causes the growth of lung tumours. As the lungs are located at some depth, it is interesting to use a mutated Luc gene coding for a so-called red luciferase, which produces red light (in contrast to the luciferase producing a green light at 540 nm), this being less strongly absorbed by the tissues.

The lung tumours are observed, after injecting luciferin intravenously or intraperitoneally, by an *in vivo* bioluminescence imaging system. Figure 7.33 shows images made a few minutes after injecting luciferin intravenously into mice that received an injection of tumour cells coding for the Luc gene three weeks previously. Images were acquired with the Photon Imager designed by Biospace Measures. The two mice exhibit lung tumours visible by bioluminescence in dorsal and ventral positions.

The same type of acquisition can be repeated daily to monitor the development of the tumours. The quantitative nature of the bioluminescence signal means that light intensity variations can be correlated with the number of tumour cells, whence their evolution can be assessed quantitatively.

Detecting Subcutaneous Tumours in Mice with an Activatable Fluorescent Probe

This example illustrates the potential of molecular engineering through the use of activatable fluorescent probes. This concept, recently introduced for fluorescence imaging of small animals [90], uses a molecule with biological activity, e.g., a ligand, an enzyme substrate, etc., jointly labelled by a fluorophore and a fluorescence quencher. The probe is not fluorescent when injected into

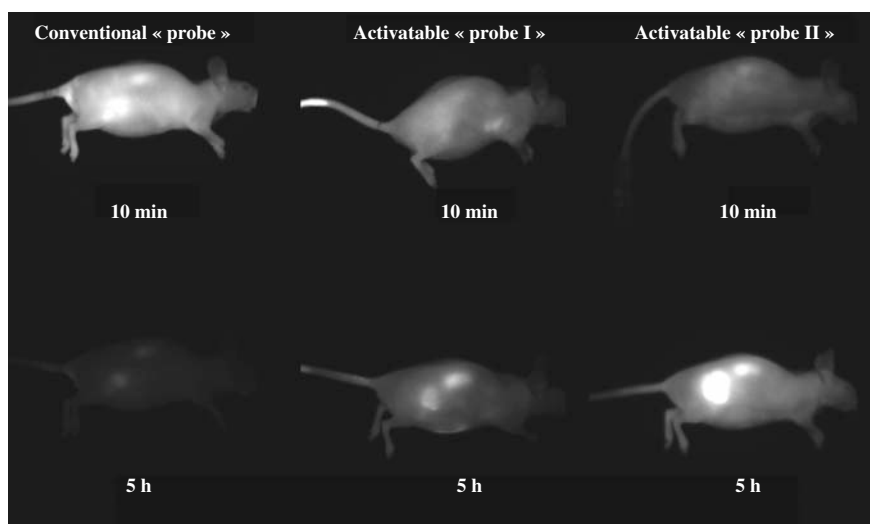


Fig. 7.34. Images of nude mice carrying subcutaneous tumours. TS/Apc model (murine breast cancer), 10^6 cells injected per mouse 10 to 15 days before imaging. Images were obtained by the FRI function of the Animage optical imaging platform. Device constructed by the CEA, Grenoble. Excitation at 633 nm by a ring of diodes producing a uniform illumination field of diameter 8 cm and intensity 50 mW/cm^2 . Detection at 670 nm by a cooled CCD camera. Images were taken 10 min (*top*) and 5 h (*bottom*) after injecting the same molecule, labelled either by Cy5 or by an activatable imaging function constructed from this same fluorophore. The amount of Cy5 injected intravenously into the tail of the anaesthetised animal was 20 nmol/animal [91,92]

the animal, but becomes so when activated by the corresponding enzyme or receptor, which induces a physical distancing (by cleavage or conformational change) of the fluorophore and the quencher. Fluorescence is then emitted, but only in the relevant imaging zone.

Figure 7.34 shows images obtained by fluorescence reflectance imaging (FRI) on nude mice carrying subcutaneous tumours, 10 min and 5 h after injecting the same molecule, labelled either by cyanine 5 (Amersham), or by an activatable probe constructed from this same fluorophore [91,92]. All images were taken under the same acquisition conditions and with the same amount of injected fluorophore. The advantage with activatable probes is clearly visible in this example. The non-specific signal outside the relevant zone is completely removed, thus considerably enhancing image contrast. Images obtained 5 h after injecting the activatable probes reveals a much better contrast. In addition, for the two activatable probes in this example, retention of the molecules in the tumour is increased, and this, combined with the progressive fluorescence activation, makes it possible to visualise the tumours

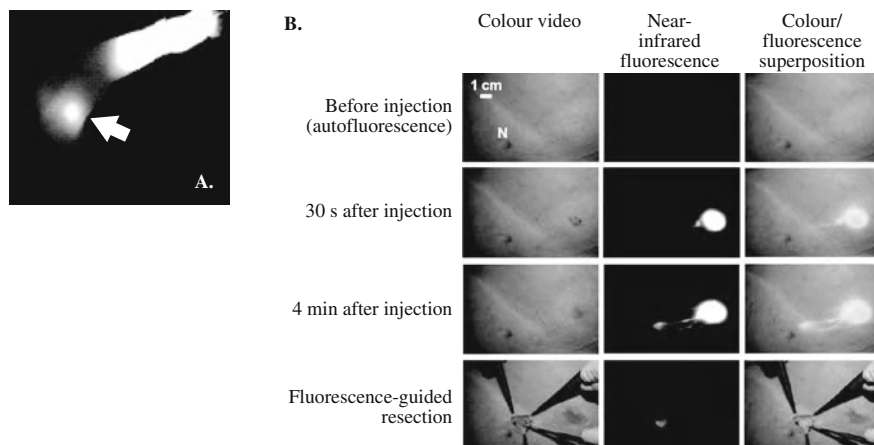


Fig. 7.35. Imaging of sentinel lymph nodes by quantum dots in mice (**A**) and in pigs (**B**) [93]

with a very good contrast up to 50 h after injection, whereas the conventional probe is quickly eliminated from the organism.

Toward Human Applications

Despite the limitations of human optical imaging, some applications can be envisaged, as demonstrated by the imaging of sentinel lymph nodes by light-emitting semiconductor nanocrystals or quantum dots (see p. 269 ff). A sentinel lymph node is the first lymph node to be affected during the metastasis of a cancer, e.g., a breast cancer. It is essential to identify this in order to guide the exeresis of affected tissues when surgery is carried out.

Groups at the Massassuchetts Institute of Technology and the Beth Israel Deaconess Medical Center have used quantum dots emitting in the near infrared to image these sentinel lymph nodes in mice and pigs (see Fig. 7.35) [93]. The quantum dots are coated with oligomeric phosphine ligands [94], but they are not functionalised to target any particular zone, and it is their dimensions (15–20 nm) which cause them to be retained in these tissues (see Fig. 7.35).

7.3.5 Time-Resolved Fluorescence Imaging

We describe here the specificities of another imaging technique, which uses a pulsed light source, in contrast to the methods discussed so far, which use a continuous source. For the moment, these techniques have been slightly less developed, but a great deal of research has been carried out for reasons to be explained below.

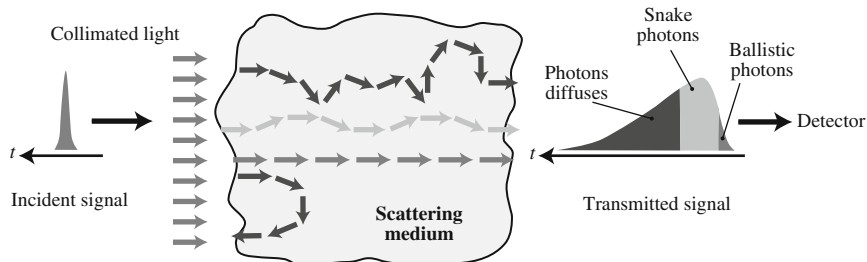


Fig. 7.36. Diagram illustrating the temporal point spread function (TPSF) in a scattering medium following a brief light pulse. Ballistic, snake, and scattered photons are distinguished [96]

Principles of Time-Resolved Measurement

Time-resolved measurements provide a way of monitoring light intensity as a function of time and wavelength. Hence, as in microscopy (see Sect. 7.1), scattering measurements allow one to probe the lifetime of different fluorophores *in vivo*.

One of the problems already mentioned, caused by the use of continuous light sources, is that the signal measured includes all scattered photons and that the paths followed are difficult to reconstruct. The time-resolved approach is a way of increasing the amount of information gathered, exploiting a source that produces very short pulses. Detectors analyse the time-of-flight or temporal point spread function (TPSF) of the light. It is then possible to distinguish between different photons, those arriving first which have followed a straight path, referred to as ballistic photons (not very common), then those that have followed an almost straight path, known as snake photons, and finally those that have suffered multiple scattering along the way (see Fig. 7.36). The integral of the signal with respect to time represents the continuous signal that would have been measured by a continuous detector. A lot more information can be extracted from the details of the scattered signal than from the continuous signal.

A third approach is a frequency method using a high-frequency modulated source. It is theoretically equivalent to the time-resolved method if all frequencies are considered. Indeed, one goes from the time to the frequency signal by carrying out a Fourier transform. We shall not discuss the associated instrumentation here, the details of which can be found in [97].

Techniques for Time-Resolved Imaging Systems

Light Sources

Laser light sources are used, either picosecond pulsed laser diodes (fixed wavelength) or pico- or femtosecond lasers (with or without tunable wavelength).

The prefixes ‘pico’ and ‘femto’ mean that these sources produce pulses with widths of the order of 10^{-12} and 10^{-15} s, respectively. The repeat rate is of the order of 50–100 MHz.

Detectors

The first is a time-correlated single-photon counting (TCSPC) system [98]. The detector is often a photomultiplier (PM) tube. The idea is to detect at most one photon per light pulse. The time elapsed between the synchronisation pulse and the arrival of each photon is recorded by a specific electronic system. Acquisition is carried out over many laser pulses, whereupon the temporal point spread function (TPSF) can be reconstructed.

A streak camera [99] can also be used to register TPSF simultaneously at several points with a time resolution of 1–10 ps. Finally, the recently developed intensified high-speed cameras can produce time-resolved images, but with lower resolution (acquisition with a 200-ps time window) [100].

Making Use of the Time Signal

Diffusion, Absorption and Fluorescence

The shape of the TPSF depends on the optical properties of the object crossed by the light. For example, it is more attenuated when the object is more absorbent. The more scattering occurs inside the object, the more the photon path is extended by successive scattering events, and this increases their average arrival time and therefore broadens the TPSF. Using a suitable model for the propagation of the light signal [101], it is thus possible to recover the optical coefficients of a medium from the TPSF [102, 103].

On top of the scattering and absorption effects, there may also be fluorescence from the tissues and from the probe labelled by fluorescent molecules. Simple fluorescence is characterised by an exponential decay lasting a time estimated by the fluorescence lifetime τ of the fluorophore. For a fluorophore alone, the intensity is given by

$$I(t) = I_0 e^{-t/\tau} ,$$

where I_0 is the incident intensity and τ quantifies the decay time of the fluorescence emission by the fluorophore.

Total Signal

The total signal is thus the convolution of the scattering and the absorption with the fluorescence signal (if there is one). Figure 7.37 shows the fluorescence signal from an inclusion of Cy5 (lifetime about 1.2 ns) in a liquid phantom. The first part of the signal is due to the path in the scattering medium, the second is due to the exponential fluorescence decay (shown by a straight line on the graph, which has a logarithmic scale on the vertical axis).

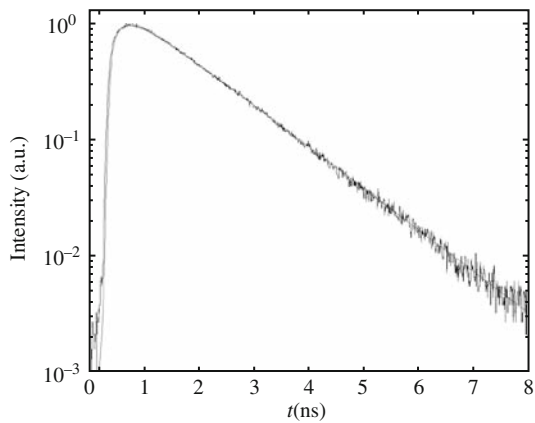


Fig. 7.37. TPSF measured for a liquid scattering medium containing a 1 mm³ inclusion of Cy5, located 2 mm below the source–detector fibres (interfibre distance 2.5 mm) [104]

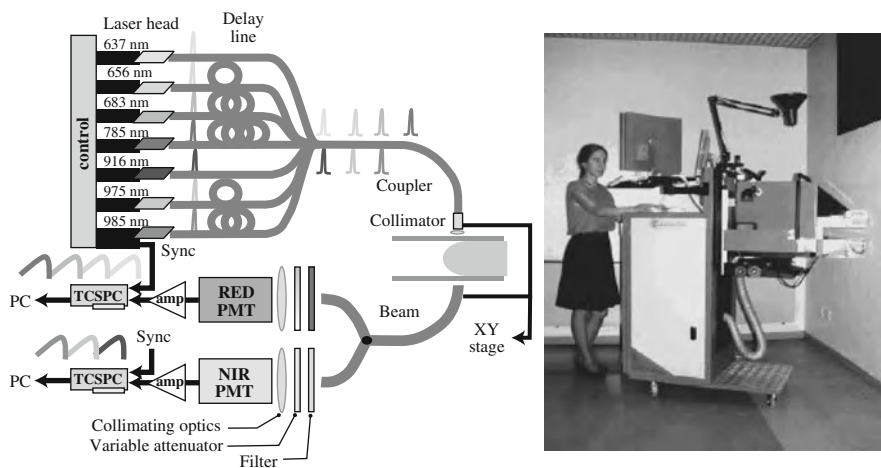


Fig. 7.38. Optical mammography. (a) Diagram and (b) photograph of MAMMOT-4 [105]

The main difficulty is to find the best way to exploit a signal that is so rich in information. Work on absorption and scattering is more advanced than work on fluorescence, but investigations continue in both areas to improve data processing and resolution.

Applications

The main applications, as for continuous-source imaging, concern cancer research and the investigation of brain activity. The method was originally developed using changes in the absorption and scattering properties of tumoral

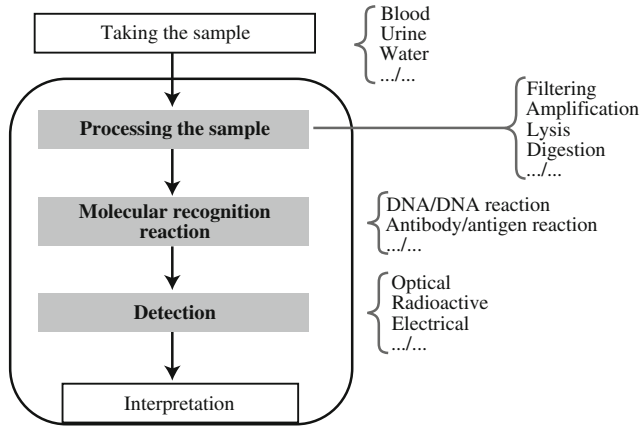


Fig. 7.39. Schematic view of biological analysis

tissues or an activated zone as a source of optical contrast. The development of fluorescent labels then made it possible to enhance the optical contrast and devise specific labelling that would assist the detection of small tumours. It should nevertheless be emphasised that tissues are themselves fluorescent (the phenomenon known as autofluorescence), which adds noise, even if the lifetime of the signal associated with autofluorescence differs from that of the labelled probes.

However, in-depth imaging of tissues remains problematic. Hence the main applications to humans are optical mammography (see Fig. 7.38) [105] and surface exploration, e.g., brain topography, dermatology [106], and endoscopy (bronchial tubes, colon, cervix).

Research continues to develop effective models and reconstruction algorithms [107] and to adapt instrumentation, either to small animals [108] for research purposes, or to hospital use for medical applications.

7.4 In Vitro Detection Systems

7.4.1 Introduction to Biochips and Microarrays

Definition

The word ‘biochip’ is a generic term for any microsystem designed for applications in biology. The purpose of these microsystems is to transform a biological sample by purification, separation, phase change, DNA amplification, and so on. For detection by molecular recognition, these biochips are usually referred to as microarrays, and this is the term that will be used throughout this section.

In the definition given by M. Pirrung [109], a microarray is a monolithic plane surface carrying a few hundred or a few thousand molecular recognition

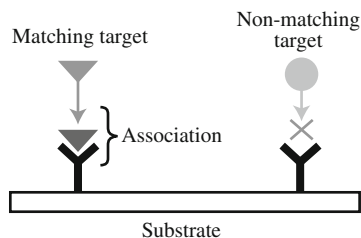


Fig. 7.40. Molecular recognition. *Left:* The lock and key match and a duplex forms. *Right:* No match, and a duplex does not form

regions generating, after reaction with a sample of given composition, a signal that can be detected by a measuring device. This signal is usually produced by scattering of light by fluorescence (see Sect. 7.4.2) but other transduction methods are also used, e.g., plasmon resonance, interferometry, electrochemical detection, etc., discussed in Sect. 7.4.3.

In the context of biological analysis, the microarray is used in the penultimate step (see Fig. 7.39). The idea of the microarray is to allow target molecules, such as antigens, antibodies, peptides, DNA fragments, etc., present in a liquid, to react with a surface carrying small rectangular or circular spots. Each spot is endowed with a molecular recognition element, i.e., a well-determined family of molecules able to capture specific targets by molecular recognition.

Molecular Recognition

The molecular recognition reaction can be defined as the property of two molecules to associate and form a duplex, e.g., an antibody/antigen pair or two complementary strands of DNA. The analogy with a lock and key, introduced by Fisher in the context of immunology in 1894 [110], is often used to explain the specific nature of this interaction (see Fig. 7.40).

Hence, an antibody steered toward a given substance will be able to associate in a more or less specific way with certain molecular groups of the substance. Likewise, a single DNA strand will be able to associate with another such strand provided that these two strands are complementary.

In molecular biology, the DNA molecule immobilised at the surface is called the capture probe. This term is not usually used in immunoassays, although it is perfectly well suited. For simplicity, we shall use it for all fields of application.

Applying this idea to a microarray, i.e., depositing capture probes of a given type on each spot, targets can then be captured in a localised manner (a specific target for each position). One can thus determine the species present in the specimen along with their locations (see Fig. 7.41).

In most cases, there is a labelling stage during the reaction process or the preparation of the sample. This operation consists in binding a molecule called a label or a tag on each target. This molecule usually has optical properties that can be used to detect it, with the help of suitable equipment. These labels

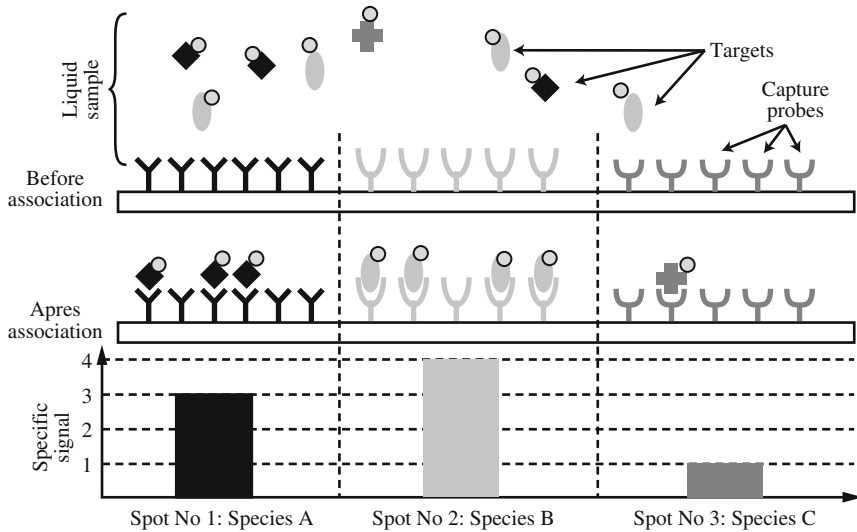


Fig. 7.41. Molecular recognition using a microarray. Three spots are represented. Each of the spots 1, 2, and 3 corresponds to a type of target, recognising species A, B, and C, respectively. The signal measured for each spot is then proportional to the number of targets of each type

can be radioactive, electroactive, or magnetoactive (magnetic nanoparticles). An optical detection method is most commonly used, usually with fluorescent, but also absorbent, chemiluminescent, or electroluminescent labels.

After reading with a suitable instrument, data is obtained in the form of an image, in which the intensity level of each spot informs as to the type of target captured and, in some cases, the concentration of each species in the original sample.

A Brief History of the Microarray

The biochip was born from work carried out as early as the 1970s [111]. The analysis of biological samples using a spotted slide was described with spots measuring around 8 mm [112]. Apart from the scale factor, the system did indeed display the main features of a biochip, with an array of spots arranged in a grid pattern for anchoring capture probes, and fluorescence detection using a microscope.

This multiplexed analysis structure on a flat substrate was taken up again with the aim of carrying out analyses in parallel. From 1981, several articles and patents describe structures bearing a strong resemblance to the microarray concept [113–116]. It would seem that the word ‘microarray’ was first used by R. Ekins [114]. It was in this decade that the size of the microarray spots went below the millimeter.

In 1991, M. Pirrung and coworkers published an article describing a light-directed in situ method for synthesising peptides [117]. This technique, applied

to DNA, gave rise to the Affymetrix arrays, where the working surface carries several hundred thousand spots.

Since then, many research groups have constructed and/or used microarrays for research purposes, but also for devising experimental protocols and diagnoses. The microarray concept is thus not a new one, and can be considered as a logical extension of work on reaction supports, although it also exhibits several important advantages to be discussed below.

Special Features of Microarrays

In a certain sense, a microarray differs little from a microplate (96 wells of diameter 7 mm distributed over an area of $12 \times 8 \text{ cm}^2$), especially as some arrays are not flat, but incorporate dips. Pushing the analogy to the extreme, a microarray is just like a miniaturised test-tube holder or an immunofluorescence slide. However, there are differences and some of them are significant.

In contrast to conventional reaction supports, e.g., tubes or microplates, the microarray can carry out analyses with very small sample volumes. By virtue of today's microfabrication techniques, the size of each individual recognition element can be very small indeed, from a few hundred to a few micrometers. For example, Affymetrix mass-produce microarrays for DNA sequencing in which the spots measure $9 \mu\text{m}$ across for a spacing of $11 \mu\text{m}$. On an area of just 25 mm^2 , one can thus arrange 200,000 spots, while a microplate is limited to 96 to 380 wells and an immunofluorescence slide only about ten. Mysiakos and coworkers in Greece showed that it was possible to achieve widths below the micrometer by using an electron beam method for localised surface activation [118]. Parallel analysis can thus be implemented on a grand scale insofar as multiple reactions are possible. Hence, DNA chips allow a massively parallel analysis of samples, with a significant saving of time. It has even been shown that it may be possible to analyse a complete human gene in a single experiment [119–123].

In general, a microarray is housed in a very small reaction chamber. The time required for the molecules of the liquid to diffuse to the chamber walls is thus very short, thereby reducing the time required for the test and also the sample volume. In a few minutes, it is therefore possible to detect picomolar target concentrations (10^{-12} mole/litre) in immunoassays [124, 125] and in molecular biology [126]. Reducing the volume of the reaction chamber leads to a reduction in the volume of reagents producing the reaction, e.g., enzymes, fluorophores, detection probes, etc., whence a reduction in the cost.

Depending on the configuration of the reaction chamber, e.g., capillary chips [127], the whole sample can be passed over all the molecular recognition elements, whereas in a conventional system, the sample has to be distributed in all wells. This distribution generally leads to a dilution which does not make the test any faster, or any more sensitive.

Finally, detection microarrays are compatible with the microsystem commonly called a lab-on-a-chip, which should be able to carry out all processing

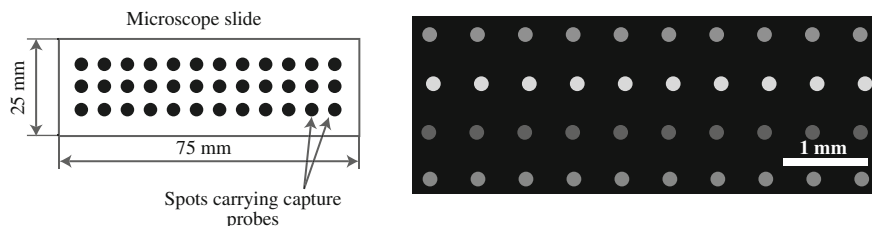


Fig. 7.42. Microarray on a microscope slide and image obtained after reading by scanner. The diameter of the spots here is $150\text{ }\mu\text{m}$. Each row contains the same capture probes, while each column contains different capture probes. The colour level indicates the density of hybridised targets after reaction, from blue to red depending on the density

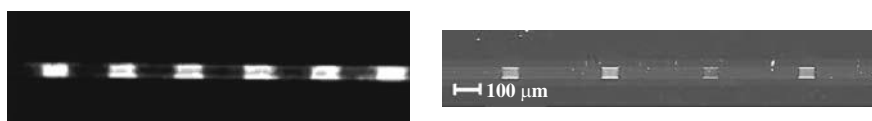


Fig. 7.43. Microarrays in capillary tubes. *Left:* Antibody microarray for immunoassay [27]. *Right:* DNA chip (CEA/DSV/DRDC/Biopuces and CEA/LETI)

operations on a sample in an automatic and compact way [128]. To this end, reaction surfaces must be developed that are compatible with this idea of integrated laboratory, because it would be hard to imagine ending the analysis of a sample prepared in a microsystem by a microplate detection.

These arguments in favour of microassays explain why, over the past few years, a considerable amount of effort has been invested in applying them to a variety of different fields. Hence many research centers and companies now propose different microarray setups for molecular recognition and detection.

Examples of Microarrays

The most widespread geometry is the microarray on a microscope slide (see Fig. 7.42), for which capture probes are deposited robotically (see p. 312) and read by scanner (see Sect. 7.4.2).

Some microarrays are made inside glass capillary tubes with very small cross-section, typically $100\text{--}500\text{ }\mu\text{m}$ (see Fig. 7.43).

For this kind of microarray, each bright region corresponds to a spot reacting differently with the sample. This type of microarray is made by a photo-immobilisation procedure (see p. 312). Some microarrays are fabricated on a micropatterned silicon substrate including functions for detecting or immobilising capture probes (see Fig. 7.44).

As yet, there are no ‘universal’ chips, capable of application to all situations. Each field of application and each specific kind of use requires a corresponding type of chip. Microarrays on microscope slides are widely used

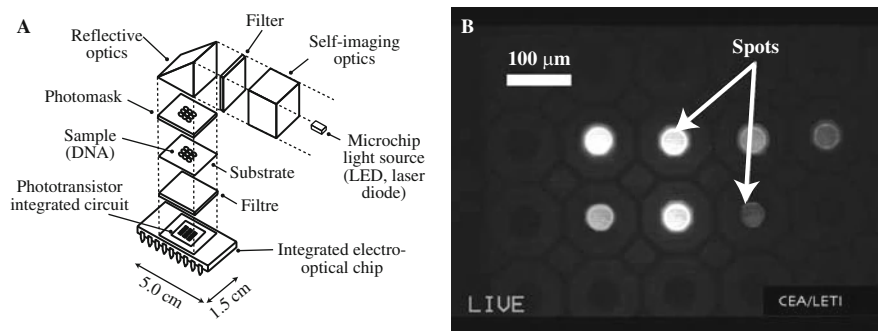


Fig. 7.44. Microarray on a silicon substrate. (A) Microarray fabricated directly on a fluorescence imaging sensor [173]. (B) MICAM fluorescence microarray after hybridisation, in which capture probes are immobilised by electropolymerisation (see p. 312)

in research, while other systems are already destined for more practical uses in diagnosis, pharmacology, continuous analysis, and so on.

Fields of Application

There are many fields of application for microarrays. Naturally, microarrays occupy an important place in clinical diagnosis, i.e., identification of pathogens in human beings. Genomics is another area in which this type of technology is much sought after, for genome mapping, the identification of new genes, the study of gene functions, and the sequencing of DNA molecules.

Note also that the screening of molecules in pharmacy is an important field of application for microarrays. Indeed, given the very large number of molecules to be tested, assays must be carried out on a massively parallel scale, and this is precisely what the microarray can achieve.

Finally, studies are under way for uses outside the laboratory, e.g., for military applications in the prevention of bioterrorism (civilian protection), but also the protection of troops during military operations [129, 130]. Other studies aim to apply this technology to environmental surveillance, e.g., for the detection of pollutants like hormones, pesticides [131] and the like, or any form of natural or industrial contamination. In the latter case, the compactness and rapidity of these microarrays when associated with integrated systems will be a determining factor.

Main Methods of Fabrication

The aim here is just to give a brief overview of the techniques used to fabricate microarrays. The interested reader is referred to the more complete view by M. Pirrung [109].

Microarray fabrication methods are largely a question of chemistry. First, the surface has to be prepared before discrete deposition of capture probes. For silica-based surfaces like glass or silicon derivatives, the chemistry of silanes is most commonly employed. For conducting surfaces like gold, silver or platinum, anchoring can be carried out by electropolymerisation of monomers, e.g., pyrrole, bound to probe strands [132]. This approach applies very well to precise localisation on a scale of a few micrometers if the substrate carries an array of individually addressable microelectrodes. An intermediate dextran layer can be used. For gold, thiols can be used as surface/probe intermediates. For plastic, probe molecules are usually adsorbed, especially for protein microarrays. In some cases, hybrid substrates are used, e.g., a plastic base, a thin film of silica, and then silane chemistry [133].

The main instrumental method for fabricating a microarray is the deposition of microdroplets by a robot. The idea is very simple and resembles the automated preparation of microplates or industrial drawing plotters. An arm moving along the three space axes takes a sample of liquid from a container, then deposits it at a well-determined point on the substrate. This operation is repeated and, between each deposit, a different solution is sampled. Today, many commercial devices are available and service companies offer tailor-made fabrication of microarrays by this technique.

The method coming closest to the techniques used in microelectronics is undoubtedly the one used by Affymetrix [134]. It combines stages of light exposure and in situ synthesis of DNA-type capture probes. The elementary nucleotides are bound where the surface has been exposed to light. The capture probes are thus assembled nucleotide by nucleotide, in successive steps. There are variations on the technique used by Affymetrix, some concerning the chemistry [135] or the exposure method, e.g., not using a physical mask, but directing the light by means of a scanner [136] or a digital micromirror array similar to those used by video projectors [137]. These techniques are based on the light deprotection of an active molecular group (i.e., the light removes some form of protection), which can be used to immobilise proteins via a neutravidin with a photolabile bond [138].

There do exist other light-directed methods for anchoring probes, such as surface photoactivation (photoactivatable silane [139]), or photoimmobilisation of a molecular group (the light contributes to the bond), e.g., a photobiotin [140–142] or a benzophenone [143].

7.4.2 Conventional Read Instruments

The first labels used were radioactive. This technique, still widely used, and for good reason, is taken as a reference in this area. Indeed, provided one has the time, particle emission (often β particles) can always be detected, and there is little interference from other sources, even though the label is present only in very small amounts, and this approach is ideal for quantification.

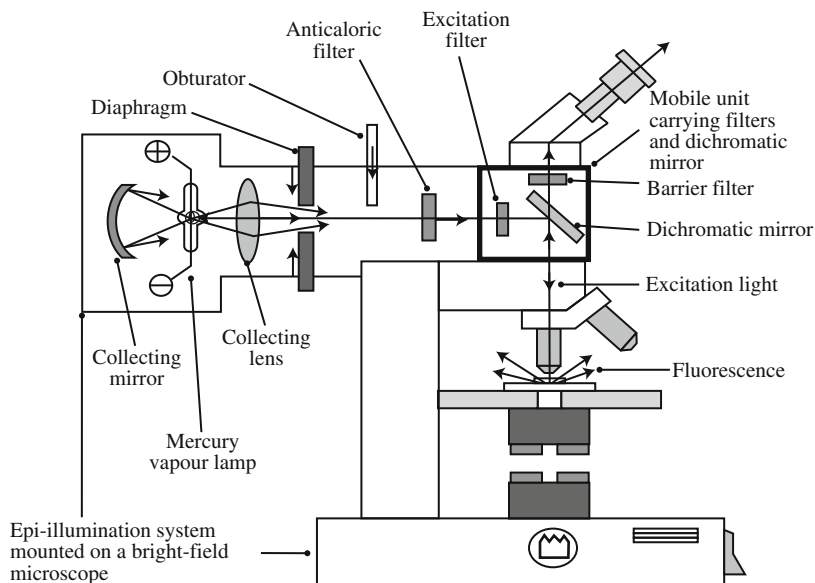


Fig. 7.45. Epifluorescence microscope

However, this type of detection has been superseded nowadays by optical detection of fluorescent labels, because they are much easier to handle than radioactive elements (fewer demands on traceability, no real safety problems), and fluorescence has given rise to far less costly detection devices than those required for tracking radioactive labels. Moreover, scanning systems can achieve comparable levels of detection sensitivity to radioactive sensors.

There are two main families of device available commercially:

- systems using epifluorescence microscopes,
- laser scanning systems (confocal or otherwise).

Epifluorescence Microscopes

These are altogether conventional microscopes, usually operating by reflection because many microarrays are not transparent or only slightly transparent, illuminated by a powerful xenon or xenon/mercury lamp (a 100 W arc lamp to provide a suitable point source), filtered in the wavelength of the fluorophore used (see Fig. 7.45). One speaks of epi-illumination because the paths of the excitation light and the fluorophore emission coincide in the microscope objective.

As can be seen from Fig. 7.45, these light signals are separated by a dichroic cube, comprising an excitation filter which isolates a spectral band around the optimal wavelength of the fluorophore, a dichromatic mirror reflecting the excitation wavelength and transmitting the fluorophore emission, and a

barrier filter eliminating all unwanted light, in particular light scattered at the excitation wavelength.

The body of the microscope is usually made by one of the major microscope manufacturers, viz., Olympus, Leica, Zeiss, and Nikon, which share the fluorescence microscopy market. In order to cover variable fields, these microscopes are equipped with a turret carrying between 4 and 6 objectives (traditionally $5\times$, $10\times$, $20\times$, 40 or $50\times$, $60\times$, $100\times$), dry objectives, or oil-immersed objectives to increase the aperture (provided that the oil in contact with the object does not alter the properties of the microarray).

Likewise, if several fluorophores are to be used, the setup is equipped with a turret carrying 4 to 6 dichroic cubes. Today, such a microscope is equipped with a trinocular head (two eyepieces and a video outlet), together with a camera for image acquisition. This type of setup is a tool not just for observation, but also for quantification.

This is a good point to say something about the cameras used. The (optical) fluorescence signal observed is extremely weak, since less than 10^{-6} of the excitation power will eventually enter the objective. (In general, from a 100-W arc lamp, one has an illumination of 100 mW in the blue or green in the plane of the sample, so the amount of fluorescence entering the objective is of the order of 100 nW at the very most over the whole field, if everything is fluorescent.) There is thus no choice but to use cooled, slow-scan cameras: cooled to reduce dark noise generated by thermal vibration of photoelectrons, and slow scan because scanning produces significant read noise in a standard video camera. Hence, acquisition rates of a few frames per second or fewer are typical, and integration times can be in the range 10 ms to 10 s or more.

The microarray to be observed is placed on a microscope stage with X, Y translation. This stage, like the focusing system and the objective- and cube-holding turrets, can be motorised, in which case the work station is completely run by computer. Along with the camera control software, a program for image processing and analysis is often proposed. One then has a genuinely automated tool for fluorescence quantification.

Depending on the degree of automation and the sophistication of the processing and analysis software, the price of such a system can vary between 10 and 100 keuros. This price often depends on the performance of the camera, the objectives, and the number of dichroic cubes.

The four main microscope manufacturers mentioned above offer pick-and-choose solutions for assembly by the user, while some biotechnology companies propose completely integrated ready-to-use solutions, taking into account the specific features of microarrays (especially regarding image analysis tools and acquisition parameters). One should also mention Apibio (see Fig. 7.46), or Imstar which offers a very complete solution with at least four colours. Today this type of equipment provides a good standard, while remaining versatile (easy to change magnifications, fluorophores, etc.) and at a reasonable price.

The main drawback with the conventional microscope is limited resolution and a high level of light interference since the whole object field is illuminated.

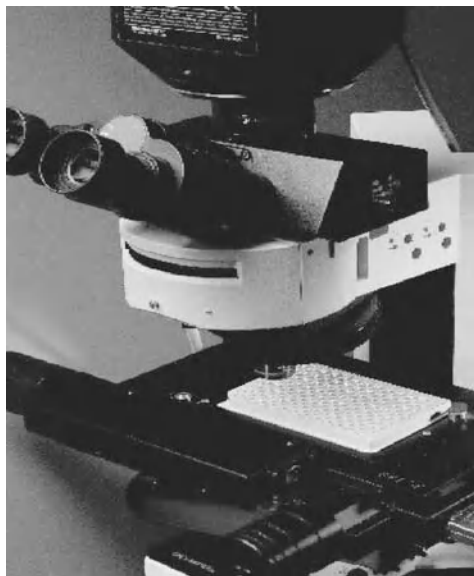


Fig. 7.46. Apimager system made by Apibio. DNA microarrays are placed at the bottom of wells

Laser Scanning Systems

For a long time now, microscope manufactures have been offering laser scanning systems, essentially confocal microscopes, designed mainly for 3D imaging. Let us look briefly at this kind of setup.

Conventional microscopes use a lamp to illuminate the whole object field in a fairly uniform manner. This is therefore incoherent (Köeler) illumination. In contrast, in laser scanning microscopy, a low power laser beam (a few mW) is focused on a very small point of the sample surface ($< 10\text{ }\mu\text{m}$ for an objective of $10\times$ or more) and scanned across the surface point by point (or else the sample is displaced using a motorised translation stage). This kind of illumination is then coherent. If in addition there is a diaphragm in the detection plane, located at the point conjugate to the source with respect to the objective (whence the term confocal), the setup becomes totally coherent (see the theory developed by Wilson in his treatise [122]). Wilson has demonstrated that the resolution is improved, because the point source response is narrower. Furthermore, any light not originating from the focal plane is eliminated by the diaphragm, as can be seen from Fig. 7.47.

This is what is called the depth of focus, not to be confused with the depth of field: the latter is the lateral distance from the focusing point over which the image is not blurred, according to some standard of sharpness, while the depth of focus is the distance from the focal plane along the focusing axis beyond which the detector receives no more light. Hence, for a conventional microscope, the depth of focus is very large, while the depth of field is defined by the magnification of the objective being used, i.e., moving out of the focal plane, the image becomes blurred, but the overall amount of light entering the

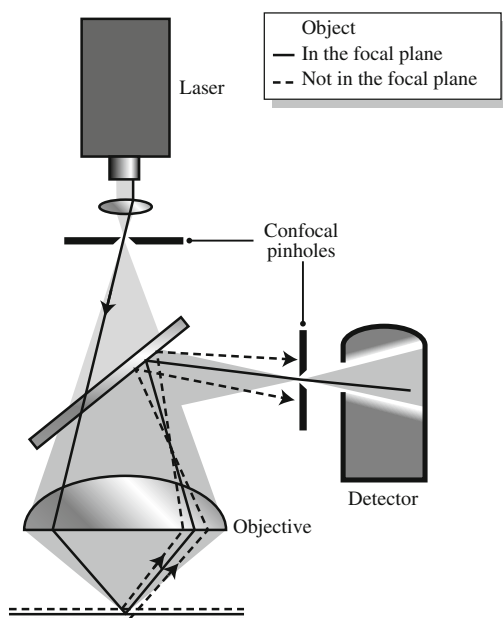


Fig. 7.47. Schematic view of a confocal microscope

objective is unchanged. For a confocal microscope, the depth of field is identical, but when one moves out of the focal plane, the light does not get through the diaphragm and all signals from outside the focal plane are removed.

With the development of the microarray, some manufacturers developed very high performance confocal fluorescence microscopes, with a resolution of several micrometers or less, and a much reduced depth of focus, at less than 10 micrometers, which means that a fluorescent film can be visualised under several mm of potentially fluorescent liquid or glass that would interfere with the useful signal. These microscopes use laser sources. The wavelengths of the lasers correspond to the main organic fluorophores used. Hence there are argon lasers (blue and green lines at 488 and 514 nm, respectively), green and red HeNe lasers (543 and 633 nm), frequency-doubled YAG lasers (532 nm), and even argon/krypton lasers and UV or IR lasers.

The beam scans the object through the objective by means of two galvanometric mirrors, or else the object is displaced by means of motorised translation stages. The excitation and emission beams are separated by means of conventional dichroic cubes. The basic ideas were described by Marvin Minsky in 1953, but the first publication on this subject is doubtless that of Slomba in 1970, and it was not until 1980 that this kind of system became commercially available.

It is no surprise to find the four main microscope manufacturers offering this kind of system, as does another company called BioRad. Note also that Leica has patented and developed an original idea for replacing the emission filters by a prism and adjustable slit allowing one to shift the spectral emission

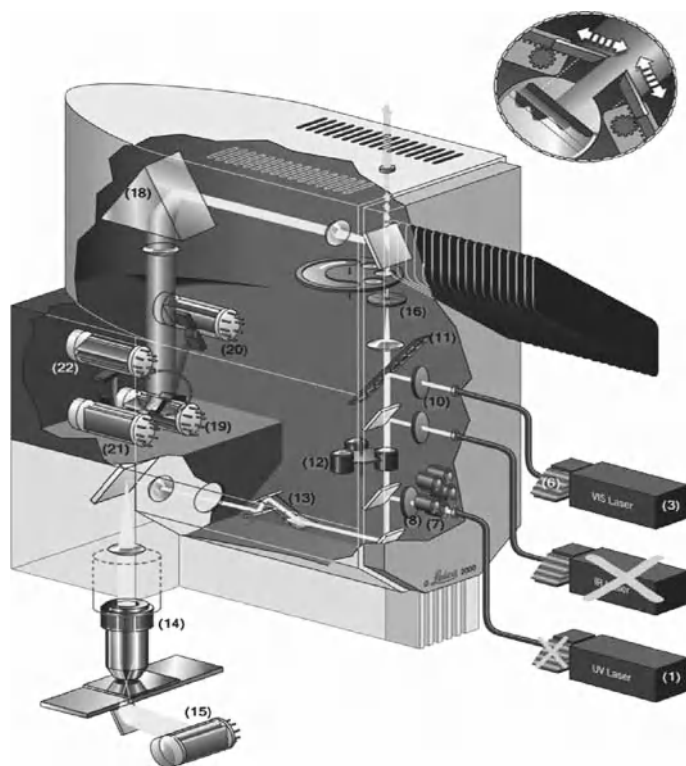


Fig. 7.48. Confocal microscope proposed by Leica. (1) UV laser 351 and 354 nm. (3) Visible range lasers: Argon 458, 476, 488, 514 nm, HeNe green 543 nm, HeNe red 633 nm. (6) Visible range acousto-optical tunable filter (AOTF). (7) UV adaptation optics for 40 $\times$ and 63 $\times$. (8) UV excitation pinhole. (10) Visible excitation pinhole. (11) Main dichroic filter DD458/514 nm, DD488/543 nm, TD488/543/633 nm, RSP 500, RT30/70%. (12) Adjustable illumination pupil at 3.6 nm. (13) K scanner with rotator. (14) Direct or reverse microscope and objective. (15) Transmitted light detector + DIC. (16) Confocal detection pinhole. (18) Prism spectrophotometer. (19–22) Photomultiplier channels 1–4

band and alter its width. All these microscopes are of course entirely automated and equipped with software for image processing and analysis, as well as 3D reconstruction. The price corresponds to the level of performance, lying in the range 150–200 keuros depending on the configuration. Figure 7.48 shows the Leica system.

For planar microarrays in the style of a microscope slide, on which fluorescent droplets are deposited, 3D reconstruction is not absolutely necessary. Hence, some manufacturers have developed specific machines in which the laser beam is fixed and the slide under examination is displaced over a reduced field corresponding to the useful part of the slide, i.e., roughly 20 $\times$ 70 mm, and

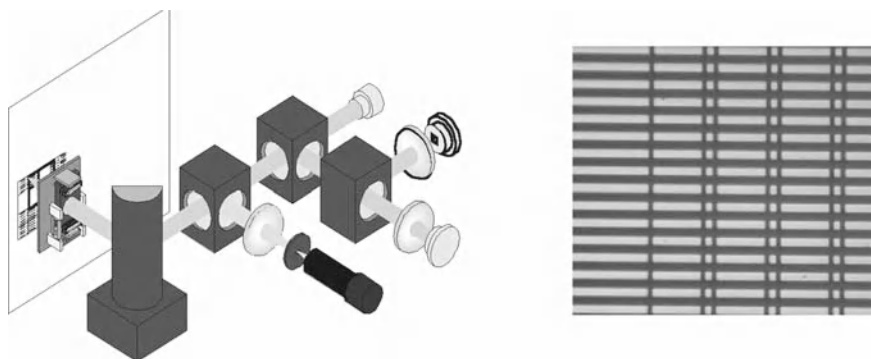


Fig. 7.49. Confocal microscanner and microarray structure (allowing focusing and tracking) [172]

where the light source contains 2, 3, or 4 distinct wavelengths, e.g., HeNe 543 and 633 nm for the combination Cy3/Cy5. Hence, General Scanning has designed a relatively cheap scanner compared to those mentioned above (around 100 keuros), with two colours.

To read the Affymetrix type of microarray, Agilent Technologies, formerly Hewlett Packard (HP), and Molecular Dynamics have also developed dedicated scanners. Finally, there are more recent systems such as GenePix (Axon Instruments) and GeneTac.

The Leti/bioMérieux team under the auspices of the company bioMérieux has developed a confocal miniscanner called lightscan, specially designed for Affymetrix microarrays, which is without doubt the state-of-the-art in its generation. Using very cheap components (CDROM read system, with associated self-focusing and tracking), this scanner should be commercialised at a very attractive price, well below the prices indicated above (by a factor of at least three). The only restriction is that, on the glass substrate, one must format the transparent tracks for dynamic focusing and tracking (see Fig. 7.49).

This section would not be complete without mention of two-photon microscopy. This approach, mainly devoted to molecular and cellular imaging, lies almost outside the range of this chapter. However, these techniques can be applied to cell chips which are beginning to appear on the scene.

The two-photon microscope uses an ultrashort pulsed laser (100 fs), with near-IR wavelength (690–1,000 nm), which is focused on the object under examination. At the focal point, the superposition of two near-IR photons can excite the fluorophore with a comparable energy to a single photon of half the wavelength. This fluorophore will thus emit in a lower spectral band than the excitation band of the femtosecond laser, which is a clear distortion of Stokes' law. The advantage is that the system is naturally confocal, since the effect will only occur statistically at the focal point of the light. In addition, excitation

is in the near IR, for which objects are naturally less fluorescent than in the visible (glass slides, cells, biological tissues, buffers used in microarrays, etc.), and it is much easier to separate the excitation and emission signals in this configuration.

7.4.3 Detection by Surface Plasmon Resonance (SPR)

Surface plasmon resonance (SPR) is a detection technique which grew in the middle of the 1990s to become the standard method for studying molecular interactions, especially protein interactions, on surfaces. Indeed, it provides a way of detecting the presence of molecules in the immediate vicinity of a surface (within 100–300nm) without the need to label the molecules. Reactions can be monitored in real time, so reaction kinetics can be recorded and one can determine whether two families of molecules react together or not, as well as the binding force, association and disassociation constants, and after calibration, the concentrations of different molecules.

The method became the reference following developments at Pharmacia at the end of the 1980s [144]. Several techniques were available to study interactions, but there was no standard mass-produced setup that could be used to make rigorous comparisons between the results obtained in different research centers. Examples are ellipsometry [145], evanescent wave fluorescence spectroscopy (with excitation by evanescent waves), and fluorescence recovery after photobleaching (FRAP) [146]. The first does not permit real-time measurements, while the other two do. By proposing an instrument that could handle liquids in a fully automated way, control the temperature, and analyse results by means of dedicated software, together with a microfluidic system, viz., the SensorChip (perhaps the first commercial microsystem for biology) and anchoring procedures, Pharmacia provided equipment that biological research teams could use to focus on their own field of research without having to develop specific instrumentation themselves. Biacore products have gradually become standard laboratory equipment on a par with the spectrophotometer, and the SPR technique is now the reference for studying molecule/molecule interactions. Later, Biacore became an independent company of the Pharmacia group and was recently bought by General Electric. It is interesting to analyse these developments. Indeed, each of the various detection techniques was intrinsically as powerful as any of the others, but it was the development of tools, e.g., by Biacore, which made these tests routine that allowed one technique to prevail over the others.

Physical Basis

SPR is an optical method using measurements of the reflectance of a thin metal film (between ten and a hundred nanometers thick), deposited on a glass surface. Gold or silver can be used but gold is preferred because it is more stable under general conditions of use. If the metal surface is illuminated

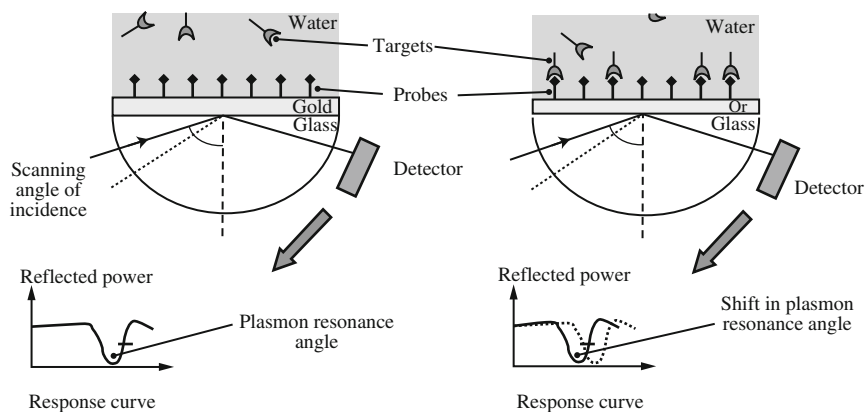


Fig. 7.50. Modification of the optical index of a medium by binding of biomolecules

through a glass substrate and if the reflectance is measured along the angles of incidence, it is observed that, for a very narrow range of angles, the reflectance is weak, even zero. The photons of the incident light excite free electrons in the lattice structure of the metal and these electrons begin to resonate. There is therefore a transfer of light energy to maintain the electron vibrations. This vibration produces the so-called plasmon wave, whose wave vector lies parallel to the surface. It can interact over several tens of nanometers of the probed medium in an analogous way to an optical evanescent wave, whose properties and characteristics it shares. If the refractive index of the probed medium changes in the immediate vicinity of the surface, the angle of incidence of the light for which resonance occurs will also change, and this can be exploited to detect the adsorption of molecules, for example.

Interaction with Surface Molecules

When biomolecules are bound to a surface, this modifies the optical refractive index of the medium, and this in turn changes the angle at which the surface plasmon resonance occurs (see Fig. 7.50). The result is a shift of the resonance in the response curve. The optical index is linearly related to the concentration of surface molecules, and this simplifies both measurement and calculation. This explains why the state of the surface, and in particular the functionalisation of the sensor, are essential features of an SPR measurement.

Functionalisation of the Sensor

Gold is usually used because it is inert, particularly with regard to oxidation, and can be functionalised by biological molecules. The film thickness is around 50 nm. The choice of substrate is less important. It can be glass, polymethylmethacrylate, or polycarbonate.

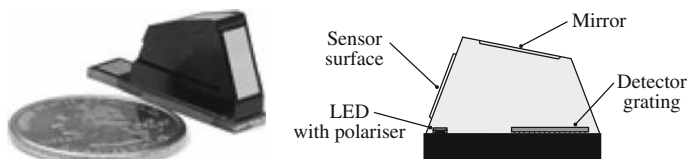


Fig. 7.51. Nomadics Spreeta component made by Texas Instruments

Biological molecules are not directly deposited on the gold, because they would not bind. The first step is to add a sulfur radical (thiol) to a dextrose (a linear chain of glucoses). It is the thiol that anchors to the gold in a self-assembled monolayer (SAM). The binding is such that the sulfur atom becomes a constituent of the gold crystal lattice. One speaks of self-assembly because the repulsive van der Waals forces between the dextrose molecules compel them to line up vertically like ears of wheat. Functionalisation is accomplished by binding the desired biomolecule to the dextrose by a peptide bond (COOH-NH_2).

Now that the biosensor is ready for use, let us consider the different measurement arrangements.

Measurement Configurations

There are three different configurations for generating the plasmon and measuring the angle:

- Attenuated total reflection, discussed in the last section.
- Laser coupling in a waveguide.
- Coupling in a grating. The incident light is guided along an optical fibre to an etched grating (ideal sinusoidal shape with 2,400 lines/mm) coated with a thin metal film. At the end of the waveguide, a photodiode measures the attenuation of the wave by the plasmon resonance.

The source is usually a LED, a laser diode, or a P polarised (parallel to the plane of the surface) solid-state laser. The detector is a photodiode array or a CMOS or CCD camera in the case of multispot systems, or a single detector mounted on a goniometric arm. The Nomadics Spreeta made by Texas Instruments is a good example of a low-cost miniature sensor (see Fig. 7.51).

The light source is a polarised LED illuminating a gold surface. The reflected light is measured by a photodiode array. The whole object is moulded into a plexiglas matrix. The user must provide a microfluidic chamber which is brought into contact with the sensor to convey and evacuate samples. This chamber is often made from polydimethylsiloxane (PDMS) and its dimensions depend on the application (required flow rate, need for heat control, and so on).

Among the manufacturers of industrial SPR equipment are Biacore, Autolab SPR, Genoptics, Plasmonic, Reichert, and Ibis Biosensor. Each of these also offers biological protocols designed for different target applications.

Applications

This type of equipment has applications to food safety and defense (rapid detection of pathogens or chemical contamination), but also water quality and medical diagnosis.

More precisely, these systems can detect the following interactions:

- protein/protein,
- DNA/DNA,
- protein/DNA,
- lipid/protein,
- hybrid molecular systems,
- non-biological surfaces (chemical products or gases dissolved in a liquid phase).

Since these interactions can be measured by many other methods, one ought to ask what advantages and disadvantages set SPR apart from conventional approaches.

Advantages and Disadvantages of SPR

The main drawback with this type of analysis is the need to work directly in contact with the gold surface. The main approach to anchoring molecules onto the gold is binding via thiols (see p.312), but this type of binding is not very resistant and molecules immobilised on the surface are observed to desorb during analysis. Another method involves polymerising a pyrrole monomer (see p.312), but this technique is not standardised and requires significant technical knowhow. Finally, a layer of dextran can be used, but this layer is rather thick and its porosity means that interactions cannot always be considered to occur at the surface. An alternative method [147, 148] is to deposit a thin film of silica on the gold layer, and anchor molecules by means of silane binding. The resonance is then observed to shift, but without too much attenuation. This technique is nevertheless only rarely used, because the Biacore kits with thiol or dextran binding are so well established and so well suited to users' needs.

Some manufacturers, including Texas Instruments, have designed throw-away SPR biosensors. For their part, reusable sensors require accurate and rigorous cleaning.

SPR biosensors do have one huge advantage, however, in that targets do not have to be labelled and the protocol for their use is extremely simple. As an example, this means that more samples can be analysed in a given time, and for a lesser cost, so that, in certain cases, the alert can be given more quickly.

Although it is less sensitive than methods using fluorescent labeling, it is nevertheless sensitive enough for many applications (typically 1 nM for antigen/antibody reactions). The robustness and compactness of existing systems are also major arguments in favour of the technique.

7.4.4 Fluorescence Enhancement

Fluorescence enhancement refers to the increased signal observed in the detector when fluorescent molecules are placed in the immediate vicinity of a surface (within a few hundred nanometers at the most), as is the case for microarrays. When fluorescent molecules are placed within a few nanometers or a few hundred nanometers from a surface, the emission pattern (distribution of light energy in space) can be significantly modified and the ability of these molecules to transform the excitation energy can increase (very short relaxation time). These two phenomena can contribute to increasing the amount of light reaching the detector by a factor of as much as 1,000 [149]. A great deal of work has been carried out to exploit these properties in fluorescence-enhanced microarrays, with the aim of improving detection levels. The mechanisms underlying this enhancement are rather complex and their description goes beyond the scope of this book. The interested reader is referred to the very complete review in [150].

These effects arise for example with a simple glass slide. The amount of light emitted through the glass is greater than that emitted in the other medium, e.g., air or a liquid. Most of the energy is thus emitted into the medium with the highest refractive index [151]. The phenomenon can be even more intense if the surface is coated with a metal or a thin dielectric film (a few hundred nanometers thick) with a higher refractive index than the glass or silica substrate [152].

Some devices involve fabricating a mirror tuned to the fluorescent molecules by depositing a thin dielectric layer onto a silicon substrate [153, 154]. Silicon can be replaced by a metal such as silver or gold [155]. An enhancement by a factor of ten is then observed experimentally.

Another more complex configuration involves depositing a stack of thin films often called a Bragg mirror on a substrate, transparent or otherwise. This Bragg mirror acts rather like an interference filter tuned to some narrow spectral interval. Tuning between the filter, which functions as a cavity, and the fluorescent molecules serves to favour fluorescent emission in certain directions, hopefully toward the detector. The enhancement obtained can be as much as a factor of 30. Figure 7.52 shows an example of fluorescence enhancement.

An alternative to thin films is to etch a grating on the substrate (with period a few hundred nanometers, etch depth a few tens of nanometers) and then deposit a thin film (around a hundred nanometers) of metal oxide on top, e.g., Ta_2O_5 . The enhancement observed with this structure is by a factor of around 100 [156]. This structure has the advantage of being compatible with epi-illumination read systems without the need to modify them, but etching the grating is a costly technological step, somewhat incongruous with the idea of throw-away biosensors.

The most efficient system yet presented is thus able to enhance the amount of light emitted into the detector by a factor of 1,000 [149]. It is by associating

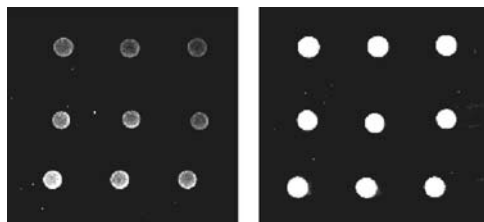


Fig. 7.52. Microarray before (*left*) and after (*right*) fluorescence enhancement

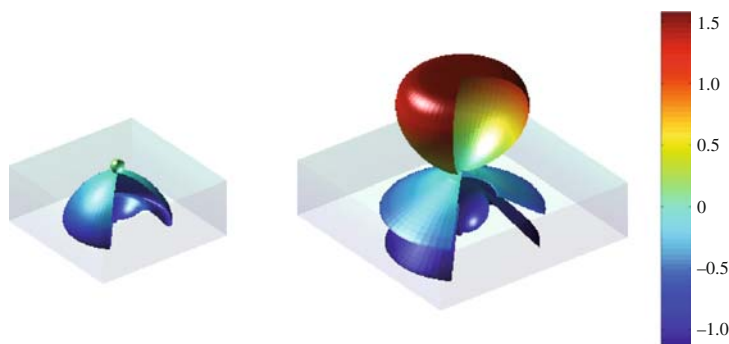


Fig. 7.53. Scattering indicators before (*left*) and after (*right*) fluorescence enhancement

all the techniques so far mentioned that such a result can be obtained, viz., deposition of a Bragg mirror onto a grating.

These very encouraging results have had few practical repercussions in the field of biosensors. Indeed, although the fluorescence of fluorescent labels is effectively enhanced for test slides, the results are much less convincing in the case of ‘biological’ slides, i.e., after a biological reaction with a sample including several target families. As S. Quake and coworkers have shown [157], single fluorescent molecules can be detected with a modest amount of material provided that the substrate is extremely clean. The detection limit is imposed by autofluorescence from the materials making up the substrate, biological buffer, and functionalisation layers, and also by non-specific adsorption, i.e., biological targets at the measurement point without molecular recognition by capture probes having taken place. The fluorescence of targets absorbed non-specifically cannot be distinguished from that emitted by hybridised targets because they have the same characteristics. The enhancement mechanism thus acts in the same way for both types of molecule, specifically adsorbed or otherwise. The signal-to-noise ratio is not therefore enhanced in the same proportion as the raw signal. In addition, when a mirror of any kind is used, however it is made, it will necessarily return a significant proportion of the excitation light toward the detector in an epi-illumination setup of the kind equipping microscopes and most scanners. To get round this problem, LETI has developed a dedicated reader reviving an oblique-incidence lighting

arrangement first proposed by Mirzabekov and coworkers, which is well suited to this application [158] (see Fig. 7.53).

Another heavy restriction related to the use of enhanced fluorescence microarrays arises because the enhancement depends on the distance and orientation of the molecules at the surface. Indeed, a modulation of the enhancement factor is observed depending on the height of the molecules. These parameters affect the radiation pattern, the fluorescence lifetime, and the gathering efficiency of the objectives [159]. It is therefore particularly difficult to compare results obtained with different coupling chemistry, biological models, or instruments. Interpretation has to be carried out with great caution. Curiously, biologists are less concerned about this restriction than physicists, because they are used to working by comparison, while physicists have some difficulty applying theoretical models.

Although fluorescence enhancement has not yet had any particularly striking applications (note the commercial activities of Genewave, a French start-up company which sells this type of product, and with a longer record, Diagnostic Products Inc. with the product Immunogold), this phenomenon is present in most microsystems for biology, due to the coexistence of different materials, e.g., silicon, metal, dielectric, and so on. It is thus essential to take it into account when designing these systems and interpreting measurement results. This is a field where there should be a significant interaction between biology, chemistry, optics, and instrumentation.

7.4.5 Current Trends in Biological Instrumentation

Introduction

Today microarray read systems have reached a level that seems to make the best possible use of instrumental techniques. The intrinsic limits of read equipment are those imposed by physics, namely, the diffraction limit for the separation and definition of microarrays, and the photon limit for detection. Today, it is not difficult to detect single fluorescent molecules [157], because the energy emitted to generate the information is sufficient. The true limits of these read systems do not in fact concern the physical measurement, but are related rather to the biochemical quality of the microarray, i.e., effects like non-specific adsorption of molecules and repeatability of the preparation and reaction, and the constraints affecting use, such as read rate, reaction time for very low concentrations, size of equipment, or ergonomics. This is why current developments focus rather on improving these points.

New Restrictions

Over the last few years, new restrictions have arisen, more closely related to the environment in which the system is used. In particular, there is a strong tendency today to incite those involved in this field to develop solutions allowing analyses to be made outside the laboratory:

- analysis at the doctor's surgery, e.g., diagnosis of bacterial angina,
- analysis in health centers that may not be carried out by a specialist,
- analysis in the home, like pregnancy tests,
- analysis in the ambulance to prepare for admittance to the emergency center at the hospital.

Two other fields of application are also concerned by systems that can be used outside the laboratory: military applications in the field of operations (troop protection) but also in the home country (civilian protection), and environmental monitoring (detection of industrial or natural pollutants). One could mention the system developed by the Lawrence Livermore Laboratory in the United States for the 2002 Winter Olympic Games in Salt Lake City [160], or the RIANA system developed by the University of Tübingen in Germany for detecting pesticides and hormones in the water supply [131].

For all such applications outside the laboratory, some of the constraints imposed are quite new, e.g., the very small size of the device (it must be portable), the device must be easy to use because it has to be accessible to non-specialists, it must be autonomous (possibility of operating with batteries) and also robust, in terms of both instrumentation and presentation of results (automatic analysis without the help of a specialist).

Despite the considerable effort invested over the past few years to reduce the bulk of conventional instruments, improve the ergonomic aspects of use, and automate data processing, they still cannot be used under these conditions.

One line of research to achieve these objectives is to integrate as far as possible the detection functions into the microarray itself, to avoid problems of malfunctioning (making systems more robust), to avoid optomechanical interfaces (to make the system more robust and reduce bulk), and to remove the need for conventional optical components. Finally, by choosing unit components with low electricity consumption such as light-emitting diodes and MOS photodiodes, it should be possible to reduce electrical consumption.

Work on biosensors was already moving in the direction of integrating detection, but these sensors could not seriously claim to be biochips (rather monoplexes). The first publications describing the principles of highly integrated systems appear in the patents of H. Marsoner, who established the first examples of biochips fabricated directly on a sensor array [161]. This idea was then taken up by various research groups (see p. 328).

Marsoner also suggested simultaneously integrating the excitation sources and detector into the biochip with fluorescence detection and evanescent wave excitation [162], an idea that was taken up again some twenty years later by German and Greek research groups [163]. The variations on this theme will be described in the next section concerning optical and electrical detection.

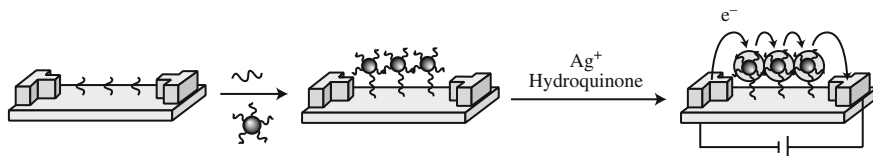


Fig. 7.54. Electrical detection using nanoparticles. Taken from [164]

Detection Integration

Electrical and Electrochemical Detection

A highly promising solution for fully integrated detection is to use electrical detection: electrical signals probe the surface of the microarray and the result is given in the form of an electrical signal. The attraction of such methods is easy to understand, because it removes the need for optical stages, so one can envisage designing systems that look more like microelectronic components, i.e., highly integrated and compact. This can be done with or without labeling.

Techniques involving labels will implement electrochemical labels, e.g., ferrocene, enzymes transforming a substrate into an electrically active product, or electrical labels, e.g., gold nanoparticles. An example is provided by the ingenious work carried out by C. Mirkin and coworkers, who label targets with gold nanoparticles, then carry out a silver enhancement. The quality of the electrical connection between the two electrodes will depend on the density of hybridised targets at the surface, and the resulting current will thus inform as to the composition of the sample (see Fig. 7.54).

Detection without labeling will use some specific measurement method, e.g., impedance measurement [165], transformation of hybridised targets after exposure to a reagent, e.g., oxidation of guanines for DNA [166].

The most advanced electrical detection system yet produced on a microarray is without doubt a product of the company Combimatrix in the USA [128]. The microarray comprises 1,000 spots using CMOS technology, each spot being associated with a circular microelectrode of diameter 30 μm . Detection is achieved at several spots by electrochemical measurements of the activity of enzymes (horse radish peroxidase or HRP) used as labels. The targets are toxins (ricins), bacilli such as *Bacillus globigii*, or glycoproteins. These tests show that it is possible to detect the presence of five different analytes, added simultaneously with incubation times of 60 min. The detection limit is very good, e.g., for ricin, it is given as 0.3 ng/mL after 60 min of incubation, comparable with the best optical systems dedicated to this type of target.

Despite some successes and a burst of interest at the end of the 1990s, electrical detection methods for microarrays are today facing stiff competition from optical methods, because several groups have been able to demonstrate the integration potential of such methods.

Detection by Magnetic Labeling

Another non-optical process with a good potential for integrating detection uses magnetic nanoparticles as labels. The capture probes are then located above microcoils (one spot per coil). The electrical signal from the coils depends on the presence or otherwise of nanoparticles bound to the targets. This technique can be used to detect DNA concentrations close to the fM [167].

Detection by Nanostructures

A method using nanostructures has recently been applied to biology. The idea is to anchor the capture probes onto a silicon or carbon nanostructure. Such structures are read electrically and the measured signal depends on the local chemical environment. When targets associate with the probes, this environment is sufficiently modified to cause a change in the measured signal, even in the absence of labels. Work on nanostructures promises extremely high performance. For example, Li and Hahm have shown that it is possible to detect from 1 pM [168] down to a few fM [169] of targets in a time close to one minute. These results have been cautiously contested by a group at the Naval Reserach Laboratory who, using theoretical results, show that the probability of a spot having captured a molecule after a few minutes at concentrations of a few fM is extremely small.

Integrated Optical Detection

The non-optical transduction methods described so far reveal their limitations as soon as they are used for concrete applications, where the initial sample is variable. The main criticisms of optical transduction are the fragility of the instruments, their bulky proportions and weight, and the time required to reconfigure them. Several groups are trying to get around these constraints and have produced highly promising structures without sacrificing detection performance.

Optical Detection with Sensors Integrated into the Chip. The idea originally introduced by Marsoner [161] consists in anchoring capture probes directly to the photosensors. The first attempt with a single spot was published in 1996 by Lu et al. [170]. This system could achieve detection in two different spectral bands. The development of array-based systems was suggested by the Genometrix group [171], anchoring capture probes directly onto a CCD sensor. This application was designed to carry out multiplex fluorescence or chemiluminescence detection. However, the cost of the biochip was equal to the cost of the CCD array. This technology could not compete with chips made on glass, especially as cheap and compact read systems were beginning to appear [172].

The problem of cost is disappearing today thanks to progress with CMOS image sensors, also called active pixel sensors, found in the cheap WebCam products. Vo-Dinh suggested using these sensors for fluorescence detection,

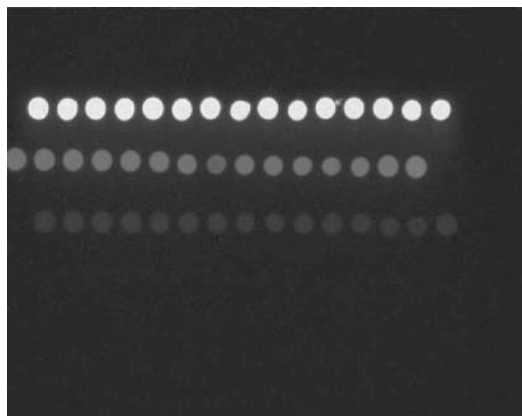


Fig. 7.55. Image made with a CMOS multispot chip

but the system remains complex due to the fact that light sources are required to excite labels [173]. In addition, the need to use highly selective filters to separate the fluorescence signal from its excitation rules out the possibility of a monolithic chip. Marsoner [162] has suggested several ways of getting around this difficulty (excitation by evanescent wave, use of total reflection to separate excitation and emission), but they have never been implemented.

The most successful development in this direction uses enzymes as labels, viz., enzymes causing light to be generated in the presence of some particular product. The labels are themselves the light sources here. There is thus no longer any source required for excitation, and no filters to separate energies. F. Mallard and coworkers have obtained very interesting results with a multispot DNA chip (see Fig. 7.55) [126].

Simplifying the Light-Gathering Optics. A lot of work has been carried out to adapt guided optical systems to biological applications. In most cases, these sensors, often called bio-optodes, couple light in a waveguide, e.g., an optical fibre or planar waveguide (see Fig. 7.56). This light probes the biological medium. Detection is usually by fluorescence [130, 174], but sometimes by light-scattering from nanoparticles [175]. In some cases, the fraction of fluorescent light emitted by hybridised targets and coupled in the guide is used for detection [176, 177].

The main difficulty with this kind of system is to couple the excitation light in the optical guide, whose thickness varies from a few micrometers to a few hundred nanometers. To get round this problem, one alternative is to illuminate a chip in which each spot is deposited above a waveguide [178]. The measurement is made on the light at the output of each guide. In this case, spatial information, i.e., concerning which spot is being read, is given by the position of the guide at the chip output. No mechanical scanning is then necessary, and no sophisticated optics.

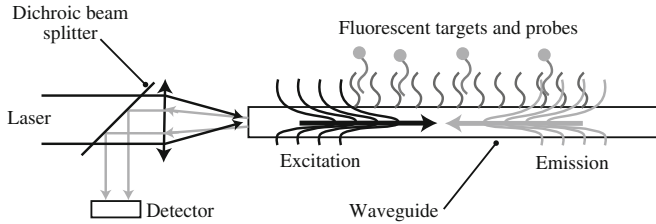


Fig. 7.56. Detection by a waveguide with excitation by evanescent waves and measurement of the back-scattered signal coupled in the guide

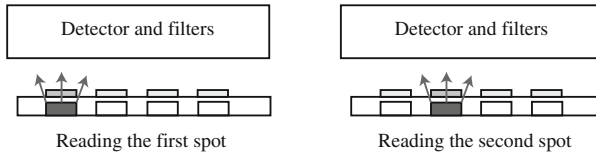


Fig. 7.57. Non-mechanical detection using a chip with integrated sources

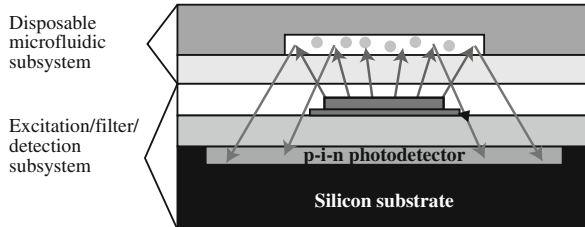


Fig. 7.58. Microsystem with integrated optical detection. Taken from [183]

Optical Detection with Integrated Sources. Another approach that is currently being explored is to associate a source with each biological spot and to switch on the sources one after the other, thus achieving scanning without mechanical means (see Fig. 7.57).

The first step toward achieving this system was to illuminate the whole chip with an excitation source and reproduce the scanning by means of an array of liquid crystals placed in front of the detector, in which only one pixel is transparent [179].

However, arrays of silicon light-emitting diodes are rather costly and the risk is that these devices may not lead to industrial applications. Furthermore, it is no easy matter to filter out the excitation light in the configuration depicted in Fig. 7.57. A solution to the cost problem was suggested by an American research group [180]. This alternative consists in using organic LEDs or polymer LEDs (OLEDs and PLEDs, respectively) which are potentially much cheaper to produce.

A first application with a PLED integrated into a biological device was demonstrated in capillary electrophoresis [181]. The results obtained exhibit similar performance to detection by epifluorescence microscopy.

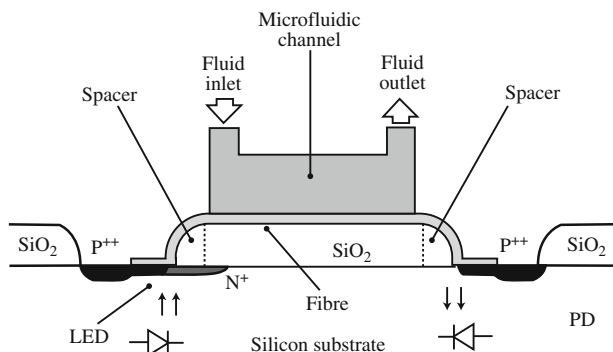


Fig. 7.59. Integrated detection by Misiakos et al. [163]

The completely integrated system for multiplex detection is prefigured by a chemical sensor for detection of oxygen: to the system illustrated in Fig. 7.44a is added a film containing OLED sources, pixelated or otherwise, which are placed between the CCD detector and the bioreactive surface [182]. This setup, which has not been used for experiments in biology, remains monoplex.

Integrating Sources and Detector. Complete integration of fluorescence optical detection is shown in Fig. 7.58 for a single spot and the detection of fluorescent nanoparticles. A photodiode (detection), a filter (to block excitation), a blue PLED light-emitting film (excitation), and a microfluidic chamber are used to detect nanoparticles adsorbed onto the surface of the biosensor [183].

This setup solves the problem of filtering off the excitation light. In fact, this is achieved by a CdS layer. The whole arrangement with source, filter, and detector is monolithic. With regard to fabrication, this device is compatible with a multispot structure.

The most advanced work so far is without doubt the system presented by groups at the Institute of Microelectronics in Greece and the Fraunhofer Institute in Germany [163]. The device comprises a capillary moulded in PDMS (700 μm long, 240 μm wide, and 140 μm high), then fixed on the reaction substrate with its capture probes and detection system. The latter is made in a silicon substrate, on which a LED optical source, an optical fibre, and an optoelectronic detector are fabricated (see Fig. 7.59). The LED and photodiode are made by suitably doping the substrate. The optical fibre is fabricated by depositing a layer of silicon nitride 25 μm wide and 150 nm thick on top of a silicon layer which serves as a spacer to maintain a certain distance between fibre and substrate. This fibre is in fact a planar waveguide. By fabrication, the LED and detector are self-aligned with the fibre. The fabrication yield in the laboratory is given as 95%.

The capture probes (biotins and/or antibodies) are anchored onto the fibre by O₂ plasma activation followed by silanisation with a hydrophilic silane like APTES. Targets are labelled before reaction with the substrate, by gold nanoparticles of diameter 8–10 nm. Their binding to the probes is detected

by measuring the reduction in light caused by decoupling of the light guided in the fibre due to the presence of the nanoparticles. This decoupling is significant, because a plasmon wave is created owing to the metallic nature of the particles. This measurement technique is thus similar to an optical density measurement, with all the advantages that represents, i.e., no photoextinction, very little influence of the composition of the medium on the measurement, no temperature dependence, and high stability. The system can be used in kinetic mode, because the decoupling of the light in the waveguide is only caused by the target molecules present in the volume occupied by the evanescent light field (height ~ 200 nm), and therefore only by the targets bound to probes. The intrinsic performance of the sensor is very good, since it is limited only by fluctuations in the source, given as 0.5% over a long period.

Tests have been carried out on labelled avidin/immobilised biotin. After 30 min, the detection limits were calculated from kinetics recorded with a target concentration of 104 pM, at 3.8 pM (or 3 fmoles, or again 230 pg/mL) and 13 pM (or 0.3 amoles, or again 0.8 ng/mL), depending on whether there is circulation or not. A silver enhancement procedure (20 min) can reduce this detection limit to 20 fM (or 16 amoles, or again 1.2 pg/mL, or again, relative to the total volume of analysed sample, 810 μ L, about 10 million molecules), which is quite remarkable.

Multiplexed detection has been attempted with this device, simultaneously analysing labelled streptavidin and anti-mouse antibodies on two spots (two fibres) with a concentration of 52 pM for the streptavidin and 1 nM for the antibodies.

The present device contains nine distinct waveguides and one integrated fluidic circuit, all packed into a space with linear dimension less than 2 cm, with flexible input and output tubes, plus all the electrical connections for a standard connector, and state-of-the-art electronic card interface. This work demonstrates the feasibility of complete detection integration into a consumable without sacrificing performance.

Conclusion. Optical detection can be integrated using robust, compact, high-performance devices, as has been demonstrated by the groups at the Fraunhofer Institute and the Greek Institute of Microelectronics (IMEL) [163], or CEA/LETI and bioMérieux [126]. These very impressive first demonstrations only exploit a few of the possibilities offered by optical detection.

By bringing together knowhow from chemistry, biology, optics, and microtechnology, it should be possible to satisfy all the new requirements for the use of biosensors.

7.5 Other Detection Systems.

Dynamics of Molecular Interactions

Molecular interactions lie at the heart of all biological systems: they play an essential role in the transfer and storage of information, during the assembly

of macrocomplexes, and in the regulation of physiological processes. In order to investigate this extraordinary dynamics of life, a whole set of tools has been specifically built up using the phenomenon of fluorescence. These include fluorescence recovery after photobleaching (FRAP), fluorescence resonance energy transfer (FRET), not to mention fluorescence correlation spectroscopy (FCS) and single particle tracking (SPT). Although all these methods serve to describe the diffusion mode of molecules, they differ from one another in the way they quantify their inherent parameters. In this sense, they are often complementary, without necessarily answering the same fundamental questions. For a recent review of dynamic imaging techniques, the reader is referred to [187].

7.5.1 Fluorescence Recovery after Photobleaching (FRAP) and Associated Techniques

Fluorescence recovery after photobleaching first emerged at the end of the 1970s, when laser technology was introduced into the field of microscopy. This technique was for some time reserved for the exclusive use of a small number of specialised research centers, but it gradually became more widespread in cell biology with the arrival of the confocal microscope, which integrates several laser systems. FRAP and associated techniques consist in measuring the recovery (or decay) of a signal after localised extinction (or activation) of a fluorescent species, thus providing a way of probing the local dynamics of proteins or lipids in different cell compartments. Analysis of such data supplies information about the structure of these compartments and the role of any molecular interactions that may have occurred.

Fluorescent Labelling

FRAP exploits the photobleaching of a fluorescent species. The relevant molecules must therefore be labelled in some way (see Fig. 7.60). This can be done with an antibody or a ligand conjugated with organic fluorophores such as fluorescein, rhodamine, or cyanines, and directed against a protein expressed at the cell surface, or by microinjecting a labelled purified protein. To label lipids, one can directly use the lipid conjugated with an organic fluorophore, or labelled ligands, e.g., the cholera toxin, which has a high affinity for GM1 lipids. Here, the valency of the probe must be taken into account. For example, antibodies have two recognition sites, and the cholera toxin is pentameric, which can induce aggregation effects that may bias the measurement. The relevant protein can also be labelled by using genetic engineering to fuse the sequence coding for this protein to the sequence coding for a fluorescent protein like GFP or DsRed, then getting the whole gene to express itself in the cells. Likewise, for membrane proteins, one can add an extracellular peptide sequence (myc, HA, binding site of the bungarotoxin) and use an antibody or ligand with high affinity recognising this sequence. In these

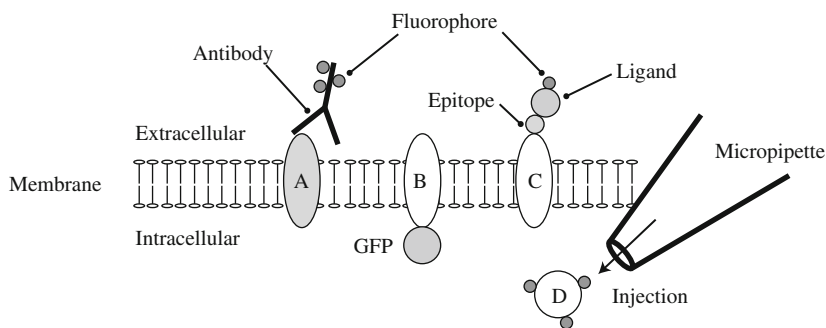


Fig. 7.60. Different possibilities for fluorescent labeling of a protein. (A) Antibody or ligand recognising an extracellular region of the endogenous receptor, coupled with fluorophores. (B) Transfection of a GFP-fused protein. (C) Transfection of a fused protein with specific epitope at the N-terminus, recognised by a labelled ligand or antibody. (D) Injection of a purified cytosolic protein conjugated to fluorophores

approaches using transfection, it should be checked that the damage caused to cells is minimal, and that the exogenous proteins are faithful labels for the endogenous proteins, in terms of localisation and function.

Photobleaching

Photobleaching, or fluorescence extinction, is a complex process related to the fact that a fluorescent molecule can only go through a finite number of photon absorption and emission cycles, this number being specific to each fluorophore (typically between 10^4 and 10^6). Extinction can be explained by a rearrangement of the chemical structure of the fluorophore, due to the fact that covalent bonds are broken by an oxidation process. Oxygen is therefore necessary for the reaction, and anti-oxidising agents are in fact used in the mounting media for fluorescent samples to reduce photobleaching.

If a fluorophore solution is illuminated at constant intensity, the emitted fluorescence is observed to fall off gradually. Precise measurements show that the decay is exponential, just like a radioactive process. This is due to the fact that extinction is a random process, so that the fluorophores do not all switch off at the same time, but one by one, and this at variable times. Let $N(t)$ be the total number of emitting fluorophores in a sample at time t . The probability that a number dN of fluorophores emitting at time t go out during an infinitesimal time interval dt is proportional to $N(t)$, whence

$$-\frac{dN}{dt} = kN(t),$$

where k is a time constant with units s^{-1} . Writing $N_0 = N(0)$ for the number of fluorophores initially present, this equation integrates to give

$$N(t) = N_0 \exp(-kt).$$

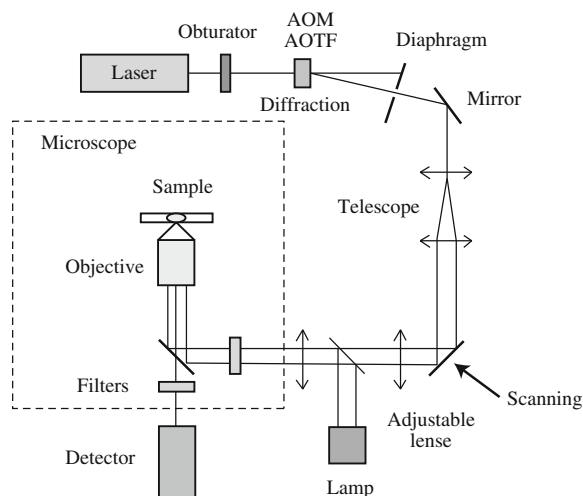


Fig. 7.61. Experimental setup for FRAP and associated techniques. Electronic components (AOM, obturators, galvanometric mirrors) are computer controlled

This does indeed reflect the exponential decay mentioned above, with a half-life equal to $\ln(2)/k$. The higher the illuminating power, the more photon absorption and emission cycles will occur per unit time, and the faster the photobleaching effect will be.

Optical Setup

In the FRAP technique, this phenomenon is exploited to quickly and locally photobleach fluorophores. One then measures the fluorescence recovery due to diffusion of fluorescent molecules from neighbouring regions into the initially extinguished zone (see Fig. 7.61). A laser with power a few tens of mW is used, with a wavelength adjusted to the absorption spectrum of the desired fluorophore. The laser beam is dilated by means of a telescope into a parallel beam a few mm across and enters a microscope objective with high numerical aperture (see Fig. 7.62). The most practical setup involves the use of an epifluorescence port. A lense at the microscope input, mounted on a triaxial translator, is used to center and focus the beam. The laser goes through an acousto-optic modulator (AOM) to alternate rapidly between a weak illumination intensity for reading the fluorescent signal and a high intensity for photobleaching. These optical components are already present in a conventional confocal microscope (see Sect. 7.1). The diameter of the photobleached zone depends on the amount of backfilling and the numerical aperture of the objective. In confocal microscopy, regions of variable size and geometry are delimited by scanning with the laser beam. For detection, a cooled digital camera or photomultiplier is used. For fast diffusion processes, (typically a membrane lipid or cytosolic protein, with D of the order of $1\mu\text{m}^2/\text{s}$), it is

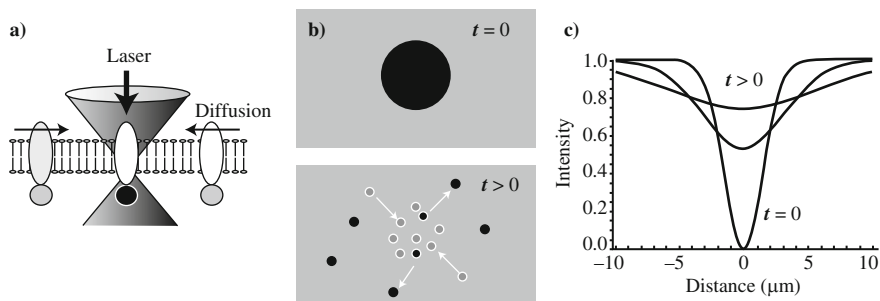


Fig. 7.62. General principle of a FRAP experiment. (a) A focused laser beam is used to photobleach the fluorophores. (b) At time $t = 0$, almost all fluorescence is extinguished in a narrow region. It then gradually returns due to the diffusion of fluorescent molecules from neighbouring regions ($t > 0$). (c) The intensity profile varies with time

better to use the signal coming from low-intensity laser illumination in order to be able to read without background noise, and fast detectors (cascade diode, photomultiplier). For slower processes, during which the observed specimen may move spontaneously, as often happens with living samples, an image must be formed. To do this, one uses either wide-field illumination (xenon or mercury vapour fluorescence lamp), or laser scanning, which is intrinsically slower.

Experimental Precautions

The illumination and acquisition parameters must first be selected. One always chooses the lowest light intensity to probe the fluorescence, sacrificing image quality if necessary. Indeed, one must at all costs avoid photobleaching the sample during acquisition, because this would pervert the subsequent measurement. One acquires a base line of a few points before photobleaching, then photobleaches as quickly as possible to avoid the diffusion of molecules during the photobleaching phase. The depth of photobleaching depends on the laser intensity and diffusion coefficient of the studied species. The photobleached fraction must be sufficient to measure a reasonable recovery, but not necessarily 100%. Acquisition of the recovery phase must be adjusted to the kinetics, and should include more points in fast phases than in slow ones, to allow good fitting later on. Recording is stopped when fluorescence recovery reaches a plateau.

Interpreting the Data

Curves can be interpreted using different models, which depend on the experimental context and biological situation. In the simplest case, one can

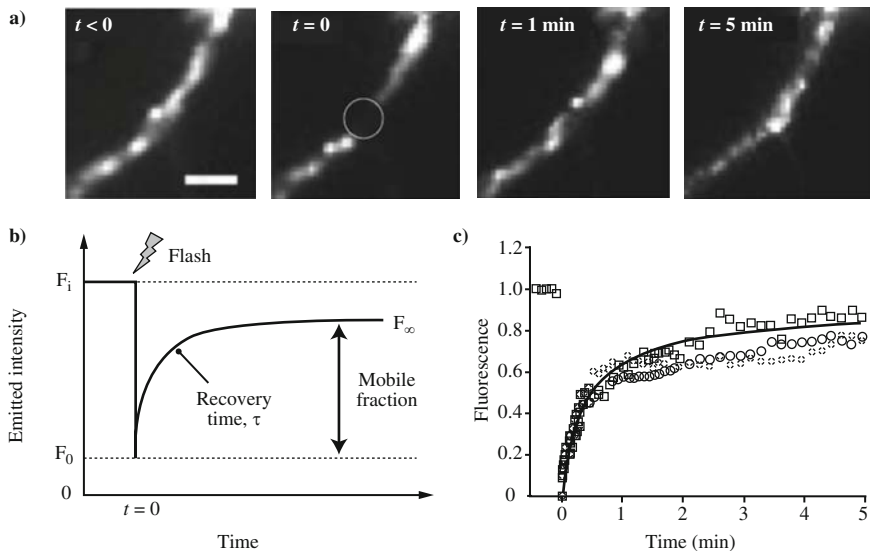


Fig. 7.63. A specific FRAP experiment. (a) Time sequence. Neurons transfected with an adhesion protein, N-cadherin fused to GFP, which is expressed inside the cell and also at the cell surface. A small region of one neurite is photobleached at time $t = 0$ (circle). Scale bar $5\mu\text{m}$. (b) Theoretical analysis of the experiment. (c) Curves obtained for three different cells for the the experiment described above, illustrating the variability of results. Data are normalised between 0 and 1, and one series of points is fitted to a non-constrained 1D diffusion model (*continuous curve*). In this case, the mobile fraction is 100%

distinguish a fraction of immobile molecules, and a fraction of completely mobile molecules to which a constant diffusion coefficient D can be attributed (see Fig. 7.63). Since one is concerned with a large number of molecules, a macroscopic approach can be used, starting with the diffusion equation. In one dimension, this is written

$$\frac{\partial C}{\partial t} = D \frac{\partial^2 C}{\partial x^2},$$

where C is the concentration of fluorescent molecules, t is the time, and x is a space variable. Initial conditions are laid down ($F = F_0$ at $t = 0$, immediately after photobleaching), as are boundary conditions $F = F_\infty$ as $x \rightarrow \infty$, if we assume an infinite reservoir of fluorophores, or $\partial C / \partial x = 0$ at $x = \pm L$, where L is half the length of the cell, plus a continuity condition at the edge of the photobleached region $C(x = a_-) = C(x = a_+)$, where a is the radius of the photobleached region. These equations can be solved by analytical methods, in particular, using Fourier–Laplace transforms [188]. Their solutions are generally expressed as an infinite series of terms combining exponential functions of time and trigonometric functions of the space variable x [189]. By truncating

the series after a few tens of terms, a satisfactory approximate solution can be obtained. The diffusion coefficient D can then be extracted for the relevant molecule. For unbiased Brownian diffusion, D is constant and there is a simple relation between the characteristic time τ for fluorescence recovery, and the radius a of the photobleached region, viz., $D = a^2/\tau$. However, if the diffusion is anomalous, D will depend on the radius and the observation time [190].

In combination with the constructions of molecular biology (truncated proteins in cytoplasmic or extracellular regions), FRAP has been used to show that the diffusion of membrane proteins depends essentially on their structure and the way they anchor to the plasma membrane [191]. Measurements based on variation of the radius of the extinguished zone have revealed that the cell surface is compartmented into domains [192]. The nature and size of these domains, related for example to lipid assemblies or the structural parts of the cytoskeleton, are still controversial [193]. Concerning the FRAP curves, the presence of confinement domains causes two time scales to appear: at short times, the proteins diffuse freely, but at long times, they display anomalous diffusion, due to the difficulty in going from one compartment to another. Using proteins fused with GFP, it is now possible to probe the lateral diffusion in intracellular compartments and study transport mechanisms [194] between compartments, or identify the association mechanisms of molecular complexes [195].

Associated Techniques: iFRAP, FLIP, FLAP, PAF

FRAP has several variants that can probe slightly different processes. Fluorescence loss in photobleaching (FLIP) involves photobleaching some zone at repeated intervals (see Fig. 7.64a). One does not observe the fluorescence recovery in this zone, which remains extinguished all the time, but rather the reduction of fluorescence in adjacent regions, since this tests whether there are connections between different cell compartments.

Inverse FRAP (iFRAP) consists in photobleaching the whole of a sample, except for the region under investigation, which is allowed to fluoresce (see Fig. 7.64b). One then measures the fluorescence decay as time goes by, this corresponding to the leakage of fluorophores from the targeted zone. In both types of experiment, care must be taken with regard to photodestruction processes due to the fact that the sample is much more intensely illuminated than during a FRAP experiment.

Fluorescence localisation after photobleaching (FLAP) is an analogue of FRAP which requires expression of a molecule labelled by two different fluorophores, either on the same molecule or on coinjected or cotransfected molecules. FLAP involves selectively photobleaching just one of the fluorophores over a narrow region, then measuring the difference between the signal corresponding to the non-photobleached species and the signal corresponding to the photobleached species, pixel by pixel. If the two fluorophores

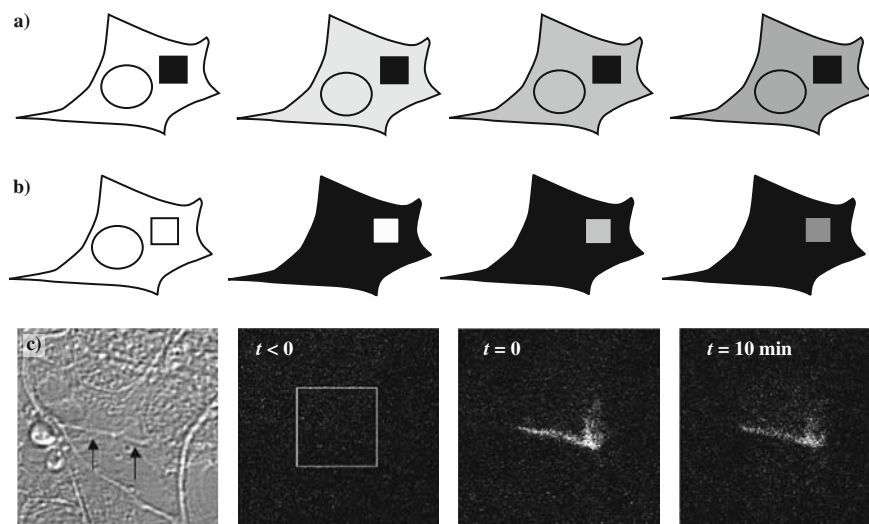


Fig. 7.64. Variants of FRAP. (a) FLIP. The *black square* indicates the zone under investigation, which is repeatedly photobleached. Fluorescence decays over the whole cell as time goes by, revealing the continuity of the compartments. (b) iFRAP. The *square* represents the non-photobleached region, which loses fluorescence as time goes by. (c) Example of photoactivation. Glial cells transfected with an adhesion protein (N-cadherin) which accumulates at junctions (*arrows* on the transmission image). The protein is fused with a photoactivatable variant of GFP, invisible before photoactivation ($t < 0$). At time $t = 0$, a $6 \times 6 \mu\text{m}$ square is illuminated by a two-photon laser beam at 800 nm, which exposes the fluorescence in this region. The latter falls off as time goes by due to the fact that adhesive bonds are renewed in these junctions

are colocalised, an entirely black image is obtained. In the photobleached region, a positive signal is obtained, which decreases as time goes by as the labelled molecules diffuse or exchange places. This technique, like the PAF technique described below, has been used in particular to probe the exchange dynamics of globular actin and polymerised actin [196–198].

Photoactivation of fluorescence (PAF) uses activatable organic fluorophores or GFP variants, which have the property of only emitting fluorescence when they have been excited by strong illumination in the near ultraviolet, typically between 350 and 400 nm [199]. This radiation causes an irreversible structural rearrangement of the fluorophore which renders it active. The PAF technique involves illuminating a narrow zone at time $t = 0$, then recording the fluorescence decay in this zone, due to diffusion or exchange processes. The curves obtained are generally the mirror images of those obtained using FRAP (see Fig. 7.64c). One drawback with photoactivatable GFP is that it is not fluorescent before photoactivation, so that it is difficult to identify the transfected cells. (One has to cotransfect with a red reporter such as DsRed.) There are

now photoconvertible variants, which change emission wavelength, e.g., from blue to green, after photoactivation at 400 nm, and which can be used to make ratiometric measurements [200].

It is also worth mentioning fringe pattern photobleaching, in which fluorescence is not destroyed on a circular region, but rather on an interference fringe pattern. This provides a way of probing the whole cell surface and obtaining a better signal-to-noise ratio [201]. Furthermore, by varying the fringe separation, several distance scales can be explored [202].

Advantages and Disadvantages

FRAP, which is intrinsically based on kinetic analysis, cannot be used to monitor temporal variations in the diffusion properties of a species. In addition, the parameter called the immobile fraction is poorly defined, and difficult to correlate with specific structural or functional properties. Moreover, FRAP can be used to probe both the cell surface and intracellular compartments, but since fluorophores are photobleached along the whole of the optical axis, fluorescence recovery can include a contribution from molecules outside the focal plane, which is difficult to evaluate. This problem is less urgent in confocal mode, which can be used to probe relatively thin optical sections. However, confocal FRAP is still much more complex to interpret than spot FRAP. In addition, the size of the photobleached region often becomes non-negligible compared with the reservoir of fluorescent molecules. This point is particularly underestimated by experimenters. Finally, a population of fluorophores is probed, allowing a fairly reliable average measurement, but hiding possible heterogeneities on the individual level. Other techniques to be discussed now can measure the mobility of membrane proteins and lipids by tracking only a very small number of molecules.

7.5.2 Fluorescence Correlation Spectroscopy (FCS)

Fluorescence correlation spectroscopy (FCS) is an experimental method used to investigate the mobility of molecules in biological media, but also to measure local concentrations or the degree of multimerisation of a molecular complex [203–206]. This method is conceptually similar to FRAP, in the sense that it can also be used to probe relaxation processes operating locally in an experimental system. It differs in that, in FCS, it is fluctuations in the signal, rather than the intensity of the signal, which become the source of information. These fluctuations continue all the time and are usually treated as stochastic noise perturbing the recorded signal. However, the temporal fluctuations of a signal around its average value can inform us about the dynamics of the system. The renewed and increasing interest in this approach is due to technological progress, which provides a way of analysing phenomena directly in complex biological media, even in the cell medium [207–209]. In its current application, it is fluorescence fluctuations that are analysed. The formalism

was described at the beginning of the 1970s to measure the diffusion and interaction dynamics of a DNA intercalating agent [210]. Today, the method can gather data in real time concerning intramolecular dynamic processes, and the dynamics of diffusion and/or association of molecular complexes. FCS remains a method for obtaining statistical information about an ensemble, which must be carried out at low molecular concentrations. In this sense, it is similar to recording a signal produced by individual molecules, when the average number of molecules detectable in the observation volume is less than unity.

It is therefore useful here to summarise the potential of FCS, now that this method has reached maturity. We shall discuss, in this order, the basic principles, experimental implementation, precautions required for a proper application, and interpretation of the data obtained in its various modes of operation.

Principles, Theoretical Concepts, and Main Features of FCS

FCS is used to quantify a certain number of observables which, through the diffusion properties of specifically identifiable elements, viz., fluorescent molecules, can lead to a better understanding of the environmental parameters regulating the observed system. In order to get a clearer picture of the benefits of FCS, let us begin with an analogy. Consider two people: the first, A, is standing in the middle of the railway station at the rush hour trying to decide which way to go, while the other, B, is lost in the middle of the desert. How will these two people move? Will they interact with their environment? Will they tend to follow favoured directions? Will they join a group? From this analogy it is immediately obvious that the movements of A will be constrained compared with those of B. Likewise the interactions of A with his or her environment will be much more significant than those of B. One would also imagine that, when observing the variations in the number of people in the station, the more people there are, the less the variations due to people coming and going will be perceptible. It is thus easier to identify local fluctuations in a set of molecules if the average number of molecules is small. This last property is one of the fundamental features of FCS: it becomes possible to describe the molecular dynamics of a system without significantly perturbing the biology, because it is not necessary to overexpress a molecular label in order to establish reliable observables from a statistical point of view.

Basic Principles

In solution, fluorescent molecules diffuse by Brownian motion, entering and leaving the region of observation in a random manner. When they pass through the laser light, this results in emission of photons (see Fig. 7.65). These signals come together to produce a fluctuating global signal. The raw data thus correspond to a temporal record of the flux of photons emitted by fluorescent

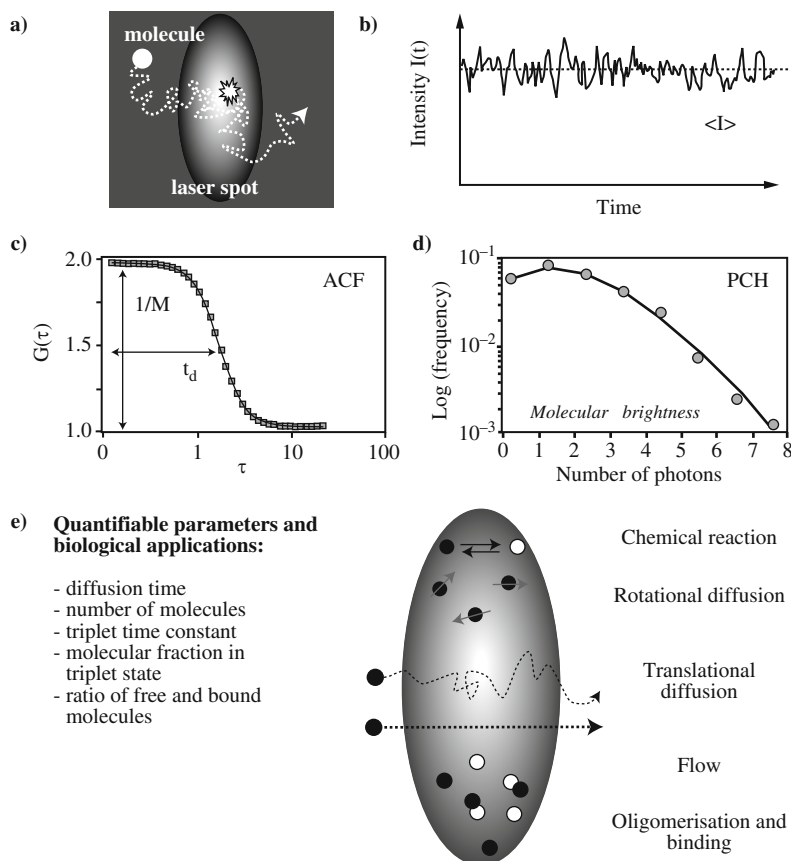


Fig. 7.65. Fluorescence correlation spectroscopy. Principles, data analysis, and biological applications

molecules present in the excitation volume. For a given system, the number of photons emitted per unit time depends on extrinsic parameters (global efficiency of excitation of the fluorophore, detection efficiency for photons emitted by the probe, geometry of the observation region), and also intrinsic parameters (diffusion time of the molecules). It remains only to relate the statistical properties of the fluctuating signal to the underlying dynamics of the emitting molecules. The tools most commonly used analyse either the temporal fluctuations of the signal (autocorrelation function ACF) or the amplitude fluctuations (probability distributions). Finally, other analytical methods have been implemented, such as statistical moment analysis. The first order moment is the average of the distribution, while the second order moment describes the variance of the distribution. However, it should be remembered that these formalisms only apply if it can be assumed that the signal is stationary (steady state) and ergodic, and this is difficult to check experimentally in biology.

Autocorrelation Function

The basic idea is thus to use the fluctuations $\delta F(t)$ of the fluorescence about its average value $\langle F \rangle$ to obtain the diffusion characteristics of the molecules. Assuming that the signal intensity is steady-state and ergodic, the time average coincides with the ensemble average. In this case, the autocorrelation function $G(\tau)$ of the fluorescence intensity depends only the lapse of time τ between two detections:

$$G(\tau) = \frac{\langle \delta F(t) \delta F(t + \tau) \rangle}{\langle F(t) \rangle^2}.$$

The most widely used analytical solution assumes a 3D Gaussian distribution for the detected light [211]. In the ideal case of a single fluorescent species undergoing translational diffusion, the function $G(\tau)$ has the form

$$G(\tau) = 1 + \frac{1}{N} \frac{1}{1 + \tau/\tau_0} \left[\frac{1}{1 + (\omega_0^2/z_0^2)(\tau/\tau_0)} \right]^{1/2},$$

where ω_0 and z_0 are the radii of the beam at $1/e^2$ in the plane and along the optical axis, respectively. The ratio ω_0/z_0 is traditionally called the structure parameter S of the collection volume, while τ_0 is the relaxation time of the translational diffusion, given by the time at half-height of the ACF and related to the translational diffusion coefficient D by [212]

$$\tau_0 = \frac{\omega_0^2}{4D}.$$

As the time τ between two detections tends to infinity, fluctuations can no longer be correlated and the function $G(\tau)$ tends to unity. On the other hand, when the time τ tends to zero, all signals autocorrelate and the function $G(0)$ then takes its maximal amplitude defined by

$$G(0) = 1 + \frac{1}{\langle N \rangle}.$$

Hence, $G(0)$ provides a direct measurement of the average number of molecules $\langle N \rangle$ in the observation volume. Knowing the effective observation volume V_{eff} , the molecular concentration $\langle C \rangle$ can then be deduced from

$$\langle C \rangle = \frac{\langle N \rangle}{V_{\text{eff}}}.$$

Probability Distribution

While the autocorrelation function describes the temporal fluctuations of the signal, it is the probability distribution that characterises the amplitude of the fluctuations. The latter depends directly on the brightness and number of

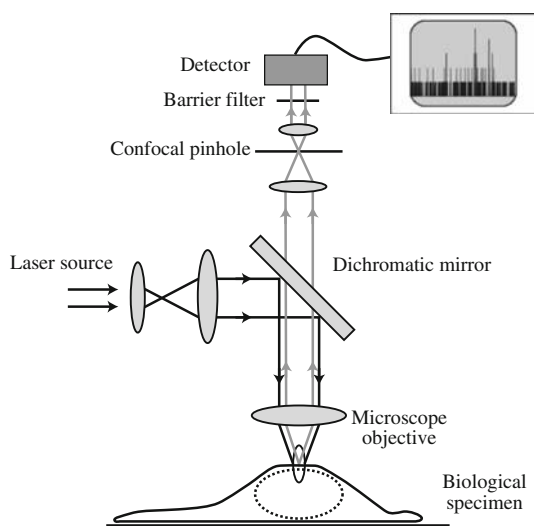


Fig. 7.66. FCS setup in excitation mode with a conventional laser source

molecules. In this case, the problem is to describe the probability of counting k photons per unit time when sampling the signal fluctuation data obtained by FCS. This probability is determined experimentally by the photon counting histogram (PCH). A theoretical expression for the PCH is obtained from the theory of photon detection [207].

Two important observables are thereby accessible. The first corresponds to the average number of molecules $\langle N \rangle$ in the observation volume, and the second estimates the brightness ε of the molecules, defined as the expected value of the number of photons detected per time sample and per molecule. This approach has been generalised to cases where several molecular species coexist (multimerisation of the fluorescent label).

To sum up, the ACF is used to distinguish different molecular species, but this is only possible in cases where a substantial change in the diffusion coefficient distinguishes the different molecular sub-populations. PCH analysis provides a powerful alternative, by analysing the differences in molecular brightness related, for example, to multimerisation of a molecule. The two methods are highly complementary. This is particularly true in the sense that, in many biological systems, the contrast related to differences in brightness has greater amplitude than the contrast resulting from a difference in diffusion coefficients.

Experimental Setup

Technological developments have profoundly transformed FCS instrumentation, making the method accessible to microscopy. Experimentally, the fluorescence coming from a very small volume compared with the total volume of the sample is recorded with an ultrasensitive sensor. This is done using a fluorescence microscope in the confocal configuration (see Fig. 7.66). The

volume in which the fluorophore is excited is then delimited by the focusing of the laser beam by the microscope objective and a variable diaphragm located in the image plane, blocking the fluorescent signal that does not come directly from the focal plane. The fluorescent molecules crossing this volume emit photons, which are gathered by the objective, then focused in a detector capable of counting the photons individually, e.g., a cascade photodiode or photomultiplier tube. Finally, the detector is connected to an input channel of the correlator which calculates the correlation function of the delivered signals in real time.

Spatiotemporal Resolution and Experimental Precautions

Depending on the data required, there are several options for obtaining a suitable spatiotemporal resolution for the relevant biological system. Here we shall summarise the options concerning in particular the excitation laser source and the characteristics of the observation objective, and those related to the correlation mode of the signal.

One-Photon/Two-Photon (1P/2P) Excitation

Classically, the fluorescence from a molecule results from absorption of a photon bringing exactly the right amount of energy to cause a transition. This is the 1P excitation mode. However, if this transition energy is not supplied by a single photon, but rather by two photons each carrying half the required energy (2P excitation), the fluorophore can be sent into the excited state leading to the emission of a fluorescence photon, exactly as though it had been excited by a single photon. Two excitation modes can be produced depending on the availability of a pulsed laser source. The 2-photon (2P) excitation mode requires the almost simultaneous absorption of two photons (time interval less than 10^{-15} s) with twice the wavelength required for conventional excitation by single-photon (1P) absorption.

These two modes do not offer the same possibilities [213]. If a single molecular species is being tracked, the 1P mode is certainly preferable because it is easier to implement, costs less, and works with a wide range of fluorophores. The 2P mode is especially useful in making available fluorophores that are normally excitable by ultraviolet light. Since the latter can cause significant damage, they would not preserve the vital characteristics of biological samples. Two-photon excitation removes this problem because it uses infrared wavelengths. Generally speaking, the 2P excitation profile of most fluorophores is much more extensive than in 1P mode. Several fluorophores that are normal spectrally separate in 1P mode can then be excited simultaneously. As we shall see below, this property, often viewed as a constraint, can also bring experimental benefits.

With regard to photobleaching of the fluorophore, the effective volume of 1P excitation is significantly greater than that for 2P excitation. As a

consequence, the fluorescent molecules thereby exposed are in danger of photodestruction before entering the effective volume for collection of the signal. The efficiency of excitation mechanisms involving two photons is no longer proportional to the intensity of the excitation source (as in the standard 1P mode), but rather to the square of the intensity. This means that 2P excitation can only occur in the focal plane, where the photon density will be high enough for simultaneous absorption of two photons by the fluorophore to become reasonably probable. There is thus no absorption outside the excitation volume as such. However, the excitation powers then required can outweigh the advantages, by leading to the apparent disappearance of the observed emitter before it has physically left the observation point.

Size of Observation Volume and Spatial Resolution

In solution, one must simply work with a small enough analysis volume relative to the volume of solution to be able to treat the latter as an inexhaustible reservoir of molecular labels. In a cell medium, this parameter becomes critical if one is to be able to justify considering the cell as an infinite reservoir of labels. In fact, the size of the observation spot depends on the excitation wavelength λ (the linear dimension of the volume is inversely proportional to λ), the numerical aperture of the objective (increased laser focusing and signal collection), and adjustment of the optical setup (by almost completely filling the rear aperture of the objective, it is possible to obtain an optimal volume whose diameter is limited by light diffraction). In typical configurations, the spot has diameter about 500 nm in the focal plane and measures 1.2 μm along the optical axis. This defines an observation region with volume of femtolitre order, corresponding to a quarter of the volume of a bacteria. Moreover, since biological samples are kept alive in an aqueous medium, it is therefore preferable to make measurements with a water-immersed objective.

Correlators and Time Resolution

With modern correlators, FCS can characterise, in a single measurement, relaxation times between fifty or so nanoseconds and several tens of milliseconds. When the phenomena under investigation occur on shorter time scales than the dead time of the counting modules (of the order of a few tens of nanoseconds), the time resolution must be extended using two detectors in cross-correlation configuration, and this after splitting the signal using a semi-reflecting cube. Note that, to analyse correlation scales of microsecond order, it suffices to use a single detector, and the correlator then functions in autocorrelation mode.

Photostability of Fluorophores

Since the signal fluctuations are supposed to reflect those of the molecular probe, it is important to understand its photophysical properties, e.g., triplet

generation, dark states, photon production, in order to avoid misinterpreting the results. For example, during an experimental measurement with FCS, the observation spot is continuously illuminated and this can last from a few seconds to several minutes. The emitting molecular source can therefore be extinguished before it has even physically left the observation spot. The average diffusion time deduced from the autocorrelation function will not therefore be correct. This effect becomes all the more important as the molecules diffuse slowly in the sample.

A good fluorophore is thus characterised not only by a high quantum yield and a large absorption cross-section, but also in a more crucial way by a very high photostability. For this reason, fluorescein is not recommended, and there are new organic molecules with equivalent spectral properties and a greatly increased photostability. The newcomer to this field is therefore recommended to refer to the literature before making a choice.

Furthermore, synthetic fluorophores have the major disadvantage that they must first be coupled directly or indirectly to the relevant molecule. It is a tedious and often complex matter to control this stage, when the target molecule is intracellular. Today, autofluorescing proteins can be used to label almost any protein while preserving its functionality. There are many variants, although the most common are derived from the green fluorescent protein (GFP). For some variants, the photophysics has been documented, precisely by FCS measurements [214–216]. These are therefore a good choice when the generation of recombinant proteins is possible.

Measurement Accuracy with FCS

The measurement accuracy depends on the number of fluorescent molecules present on average in the observation spot. Given the sensitivity of the method, there is therefore a compromise to be made between the spot size and the molecular concentration of the label, in order to optimise the desired spatiotemporal resolution with respect to the signal-to-noise ratio, the minimal measurement time, and the kinetics of the reported events. For a volume of femtolitre order, this corresponds to concentrations of nanomolar order for the analysed molecular species. This property can be both an advantage and a difficulty. Very often, the concentration of the label can reach levels much greater than the nanomole per litre in a cell medium. But on the other hand, this reflects the possibility of incorporating the molecule of interest in a cell at extremely weak doses, which are therefore without consequence for the functionality of the biological system.

It is now possible to make measurements with high molecular concentrations [217]. In this case, a very small observation volume can be created (of the order of 10^{-18} to 10^{-21} L) by using an observational support containing holes of diameter 30 to 400 nm, which is thus several orders of magnitude smaller than the size usually obtained in confocal microscopy. Using this idea, measurements can now be made in cell media for which molecular concentrations can reach the micromolar level [218,219].

Experimental Applications

Molecular Aggregation and Concentration Measurements

It was shown earlier that the amplitude of the function $G(0)$ corresponds to the reciprocal of the number of molecules N in the detection volume, and that as a result, this measurement could be used to give a very precise definition of the absolute local concentration if the effective observation volume were precisely known. However, this approach is not so trivial from an experimental standpoint. Indeed, FCS measurements must be made at low concentrations. Now, molecules often have a natural tendency to adsorb non-specifically onto surfaces. This phenomenon, which is of no consequence at high concentrations, becomes prohibitive at the concentrations required for FCS by reducing the concentration of free molecules.

In a certain number of cases, it is enough to determine the relative local concentration, e.g., when checking the expression of intracellular fluorescent proteins. This value is much more precise since it no longer depends on the size of the observation spot.

Molecular brightness is also an extremely useful parameter. It is calculated by dividing the average count rate by the number of molecules in the observation volume. To begin with, this parameter can be used to determine the performance and evolution of an optical setup by comparing the brightness of standard reference fluorophores. In a complex system, knowing the molecular brightness allows one to assess the environmental parameters affecting the properties of the fluorophore (quenching), or an enhancement of the fluorescence due to an aggregation effect.

Dwell Time, Molecular Interaction, and Mobility

With regard to temporal dynamics, the performance of FCS typically ranges between the microsecond and a hundred or so milliseconds. The primary objective of FCS analyses in solution has thus been to determine the diffusion parameters of biologically relevant molecules. Now that it has become possible to delimit a very small observation volume, FCS can also be applied in the cell medium. It is then important to identify which process is likely to explain the observed diffusion times and to eliminate at the outset any artefact associated with the use of a particular fluorophore. Many fluorophores are highly lipophilic and tend to associate with membranes, which greatly reduces diffusion, or even induces a significant deviation from purely Brownian diffusion. A comparative study of FCS measurements of the same fluorescent molecule in solution and in a cell medium reveals the importance of the environment on its diffusion properties. Thus, the time at half-height of the ACF varies over several orders of magnitude between molecules in a buffer solution or in cytosol, or again when the molecules are anchored to the plasma membrane. New strategies for adapting to the time scales of the processes under investigation are currently being developed [205, 220, 221].

In the cell medium, the diffusion of a target molecule will generally differ from purely Brownian behaviour. Several models have been proposed to account for such deviations, e.g., anomalous diffusion, heterogeneity of the observed population, active transport, etc. The reader will find an abundant literature on this subject, which goes beyond the scope of this chapter. It is important to bear in mind that, although FCS measurements are easy to implement, theoretical models are absolutely essential to fit experimental autocorrelation curves.

How do molecular interactions manifest themselves through FCS measurements? To be observable, the time scales on which these molecular interactions/modifications take place must be significantly greater than the dwell time of the molecules in the observation spot. Secondly, the modifications induced must be perceptible by a factor of 2 in terms of the modification of the diffusion time: it is only possible to identify an interaction of a fluorescent molecular component F with a molecule M if the mass ratio between the complex FM and the free molecule F is at least equal to 8 (D goes as the cube root of the molecular mass). Experimentally, this is achieved by consecutive FCS measurements with sufficiently short integration times. By decomposing the ACF, it is even possible to quantify the percentage of the different populations contributing to the fluctuations and hence to evaluate the progress of a biological phenomenon. For example, it is possible to analyse chemical equilibria, or to describe the kinetics of the interaction between a ligand and its receptor.

Recently, it has been shown that diffusion measurements at variable length scales could be used to identify the mechanisms restricting the free diffusion of a molecular species [205]. These measurements are made by modifying the observation volume: when the rear aperture of the microscope objective is partially filled by the laser beam, the less well focused observation volume increases as a consequence [222]. Then by analysing the average dwell time of the molecules as a function of the size of the observation volume, it is possible to observe whether the mobility of the various components of the plasma membrane deviates from a Brownian process and to deduce the origin of molecular confinement.

Cross-Correlation

We have seen that interactions will not be analysable by FCS if the mass change resulting from formation of a molecular complex is not big enough. To get around this problem, without necessarily having to face the constraints of molecular geometry inherent in FRET interaction measurements, two species A and B can be tracked simultaneously if each carries a different fluorescent emitter. When these species are associated or bound into the same molecular complex/aggregate, they will jointly emit a signal when they cross the observation spot. In this case, analysis will focus on the cross-correlation, frequency, and amplitude of the fluorescence signals, informing as to the level of

molecular association. This method, known as fluorescence cross-correlation spectroscopy (FCCS), is extremely sensitive in principle. However, it is more difficult to implement. Briefly, to begin with the emission spectra of the two fluorophores must overlap as little as possible. Typically, the two species are excited by different laser sources, which implies that the two excitation spots must completely overlap in order to be sure that an absence of cross-correlation does indeed reflect an absence of molecular association. An excellent alternative is to use a 2P excitation source. Since the excitation spectra of the fluorophores are broader in 2P mode than for a 1P excitation, the two fluorophores can be excited simultaneously with the same wavelength, which in itself ensures the exact overlap of the observation spots.

7.5.3 Tracking Single Molecules and Particles

Experimental Features

The technique known as single-particle tracking (SPT) was first used in the 1990s [223]. The idea is to use optical microscopy to track the motion of small particles coupled with membrane lipids or proteins at the cell surface. The particles commonly used are latex (200–500 nm), colloidal gold (40 nm), and more recently quantum dots (a few nm). The particles are functionalised with reactive chemical groups such as sulfate, carboxylate, or amine, allowing covalent grafting of antibodies or ligands against a specific receptor (Fig. 7.67a). The cells are incubated with a diluted solution of particles to obtain a point-by-point labelling of receptors (Figs. 7.67b and c). The particles then move in two dimensions at the cell surface (Fig. 7.67d).

Interpreting Trajectories

The membrane environment is so viscous that the motion of the particle reflects the intrinsic mobility of receptors, rather than the opposite. This assumption is all the more valid as the particle used is made smaller. Particles are identified by image analysis using algorithms that are often not commercially available, producing particle trajectories over long periods of time (see Fig. 7.67e). The theory of Brownian motion is used to interpret the results, calculating the mean square displacement (MSD) as a function of time. In two dimensions,

$$\text{MSD}(n\Delta t) = \sum_{i=1, \dots, N} \left[(x_{i+n} - x_i)^2 + (y_{i+n} - y_i)^2 \right],$$

where x_i and y_i are the coordinates of the particle at time $t = i\Delta t$, and Δt is the time lapse between two acquisitions. The shape of the MSD can be reliably used to identify the type of motion. For example, $\text{MSD} = 4Dt$ for a random walk, $\text{MSD} = 4Dt^\alpha$ with $\alpha < 1$ for anomalous diffusion, $\text{MSD} = v^2t^2$

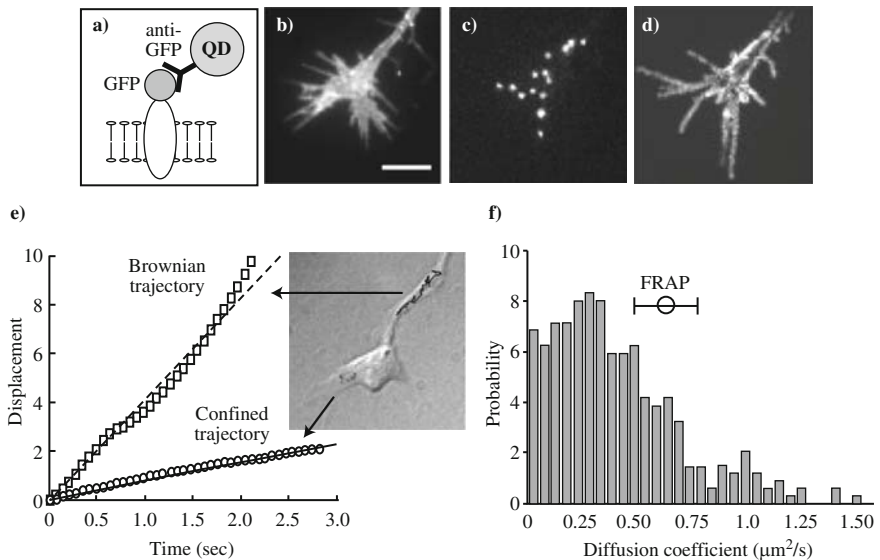


Fig. 7.67. Illustration of single-particle tracking. (a) A neuronal adhesion protein (NrCAM) fused with GFP outside the cell is expressed at the neuron surface. It is recognised by quantum dots (QD) conjugated with an antibody against GFP. (b) NrCAM–GFP fluorescent signal in a growth cone at the motile end of the neuronal extensions. Scale bar 5 μm. (c) Image of the QDs, which bind specifically to NrCAM–GFP, labelling a small number of receptors. (d) The receptors diffuse to the cell surface, carrying the QDs with them. Note that almost all the surface of the growth cone is covered. (e) Trajectories of 1 min for two QDs, and computation of the mean square displacement (MSD). The form of the MSD can be used to classify the type of trajectory (confined or Brownian). A linear fitting of the first points is used to calculate an instantaneous diffusion coefficient. (f) Distribution of diffusion coefficients for about a hundred QDs. This is a GFP protein anchored to the membrane by a glycosphosphatidyl inositol group, for which FRAP measurements are available [193]

for directed motion at speed v . Other more complex expressions have been formulated, e.g., for diffusion in a context with obstacles [223]. This technique has been able to reveal a wide range of different behaviour with a single receptor population [224]. This is reflected by a very broad distribution of diffusion coefficients (see Fig. 7.67f). Even better, SPT has been able to correlate certain types of motion with the presence and activity of specific structural components. For example, the anchoring of an adhesion receptor to retrograde actin flow in motile structures (e.g., lamellipodium, growth cone) manifests itself by a directed motion [225], while association with lipid domains is characterised by confined diffusion [226]. The use of high-speed cameras makes it possible to distinguish transitions between different types of motion and thereby reveal the highly dynamical structure of biological membranes [227].

Advantages and Disadvantages

SPT measures the mobility of molecules essentially at the cell surface. (Although probe particles can be injected into cells, their mobility is significantly reduced due to hindrance by the cytoplasm, and it is hard to label a specific component.) On the other hand, particles may suffer endocytosis in cells within a few minutes, and this can artificially create the appearance of a fraction of low-mobility particles. A reduction in the mobility of particles as a function of their size (in the range 40–500 nm) has been clearly demonstrated [228]. Furthermore, antibodies or ligands grafted onto particles can bind to several receptors, thus causing a local aggregation effect which may bias the mobility of the particle. If the antibody concentration is lowered, this cross-linking effect is reduced, at the risk of losing the specificity of the labelling. These arguments mean that diffusion coefficients measured in SPT are generally smaller than those obtained using FRAP (see Fig. 7.67f).

Detecting Single Fluorescent Molecules

Principles and Recent Observations

Detection of single biomolecules involves transposing SPT methods, which use labels with dimensions greater than about 50 nm, to nanoscale labels such as fluorophores and nanoparticles. The first observations of single biomolecules by fluorescence microscopy in living cells were carried out at the end of the 1990s [229–232]. Since then, this field has been expanding rapidly.

The detection of single molecules in biology can gather several types of information that can be obtained by no other method. In the first place, it provides a way of knowing the complete distribution of a parameter rather than just an ensemble average. Hence, even minority sub-populations can be studied, and correlations between different parameters can be recorded on the different sub-populations. As an example, one could mention studies carried out on neurotransmitter receptors in the synapses of living neurons. Not only have these revealed the existence of several sub-populations in terms of lateral diffusion in the cell membrane, but they have also been used to study the diffusion characteristics of mobile receptors [233]. Secondly, studies carried out on single objects provide access to the dynamic fluctuations of a parameter and hence also to all the steps in its time development. Indeed, it is not then necessary to synchronise the molecules under investigation (fortunately, because this would be impossible in practice in living organisms). In the last example, it was demonstrated for the first time that neurotransmitter receptors are exchanged between cell compartments (synaptic gaps and extrasynaptic regions). Finally, there is a third benefit brought by the optical detection of single molecules. This concerns the sub-wavelength accuracy with which the spatial position of a molecule can be determined. It is entirely determined by the signal-to-noise ratio with which the molecule is detected [234, 235]. Hence,

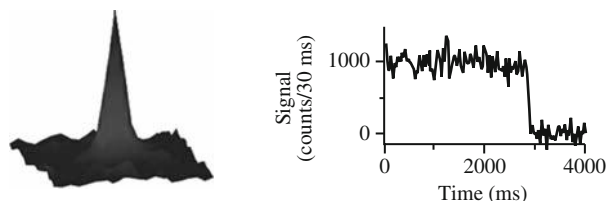


Fig. 7.68. 3D representation of the signal emitted by a single Cy5 molecule (*left*) and its time development (*right*). Digital photobleaching of the molecules interrupts the collected signal

returning to the last example, the detection of individual neurotransmitter receptors has been used to measure the motion of receptors in the synapse of neurons with diameter (~ 300 nm) of the order of the optical wavelength. The pointing accuracy of the molecules was of the order of 45 nm.

Experimental Details

The main difficulties in obtaining signals from single fluorophores in living cells arise from the fact that, at room temperature, the fluorophores only emit a limited number of photons before photobleaching sets in (see Fig. 7.68). It follows that a single molecule cannot be imaged over an arbitrarily long period, and that autofluorescence from the cell can produce a detrimental background signal. In practice, the best fluorophores can typically be imaged over periods of a few seconds in a physiological medium, containing oxygen among other things, and up to a hundred or so seconds in a medium without oxygen.

For efficient detection, fluorophores must satisfy the following conditions:

- They must emit enough photons to be identified in the relevant time window. It is thus important to use an optical setup that efficiently collects the emitted photons (microscope objectives with high numerical aperture, high quality optical filters) and very low noise detectors to reduce the number of emitted photons required to identify a single fluorophore. It should be noted that the saturation phenomenon with fluorescence emission means that the emission rate of a fluorophore cannot be increased at will by simply raising the intensity of the exciting laser. Hence, even using the most sensitive detectors, integration times typically longer than one millisecond must be used in order to detect single fluorophores. This limits the imaging rate for single molecules to order kHz.
- They must emit more photons than the molecules in their environment in the relevant spectral band. One must thus use lasers, high quality optical filters, and special detection geometries (confocal and evanescent wave microscopy [236], but also simply wide-field microscopy [234, 237]).

Metal Nanoparticles and Quantum Dots

To get around the photobleaching problem, light-emitting semiconductor nanoparticles (quantum dots) are more and more widely used. In particular, quantum dots composed of CdSe emit radiation in the visible, which means they can be directly used in the biologist's microscope [238]. Compared with organic molecules, the main advantages of quantum dots come from their broad excitation spectrum, and the narrowness and tunability of their emission spectrum, not to mention their brightness and high resistance to photobleaching. It follows that they are easily imaged singly and for periods of several minutes, making them extremely effective labels for single biomolecules in cell medium applications [239]. On the other hand, there are some drawbacks. To begin with, they have a tendency to blink which can make it difficult to identify moving objects. Then, functionalised to make them biocompatible, they reach dimensions exceeding ten nanometers, so they are distinctly more bulky than organic molecules. This problem means that quantum dots (and hence also latex particles) cannot reach intercellular junctions, which measure only a few tens of nanometers.

Labelling by metal particles is widely used in biology (electron microscopy, particle tracking). The smallest of these (< 20 nm) were until recently exclusively detected by electron microscopy. However, far-field optical methods based on the photothermal effect have now been developed to detect nanoparticles measuring only 1.4 nm [240, 241]. The use of photothermal methods to image single biomolecules seems to be a promising alternative to fluorescence since it combines the photostability of the signal with the use of nanoscale labels [242].

7.5.4 Fluorescence Resonance Energy Transfer (FRET)

Introduction

Formalised by Förster in 1948 [243], FRET has been used as a tool for biochemical characterisation of macromolecular assemblies since the 1950s. It was only much later, in the 1990s, that FRET was introduced into cell biology to characterise and monitor the interactions of macromolecules, especially proteins, in cells [245]. This development resulted from the use of GFP as an endogenous protein label [244], but also thanks to various technological advances in the field of fluorescence microscopy. Other applications of FRET have also been devised, using molecular engineering to construct probes that carry the donor and acceptor and have some tendency to change conformation in response to cell activity.

Theory

FRET exploits the electromagnetic interaction occurring between two neighbouring molecules, separated by a few nanometers. One is a fluorescent

molecule in an excited state (the donor), and the other is a non-excited molecule whose absorption spectrum partly overlaps the emission spectrum of the donor (the acceptor). A non-radiative energy transfer, i.e., without emission of a photon, then occurs, and the donor returns to its ground state, while the acceptor absorbs the energy from the donor (see Fig. 7.69a). In most cases, the acceptor is also a fluorescent molecule. FRET is then characterised by the fluorescence intensities emitted by the donor and by the acceptor. When there is no transfer, only the donor emits fluorescence. However, when transfer occurs, the fluorescence of the donor falls off and the fluorescence of the acceptor simultaneously grows (see Fig. 7.69b). The efficiency of FRET corresponds to the quantum efficiency of the transfer, i.e., the probability that an excited donor molecule returns to its ground state by transferring its energy to an acceptor rather than decaying by fluorescence. Apart from a factor related to the relative orientation of the two molecules, this efficiency goes as $1/R^6$, where R is the distance separating donor and acceptor (see Fig. 7.69c). The Förster distance R_0 is defined as the distance at which the FRET efficiency E is 50%. Then,

$$E = \frac{1}{1 + (R/R_0)^6}.$$

Each pair of donor/acceptor molecules is characterised by its value of R_0 , which generally lies between 3 and 10 nm. FRET can characterise a distance of the order of a few nanometers between donor and acceptor. By labelling the relevant proteins with fluorescent probes, e.g., GFP fusions, CFP/YFP pairs, or GFP/DsRed pairs, or by tagging the proteins with organic fluorophores such as FITC/TRITC pairs, Cy3/Cy5 pairs, etc., then introducing them into cells, the FRET measurement informs about the interactions between proteins, given that the distance for which transfer occurs (< 10 nm) is small compared with the size of the proteins (Fig. 7.69d). Conformational variations of a protein carrying the donor and the acceptor can also be revealed by FRET measurements, giving rise to many probes of cell activity (calcium probe [246], GDP/GTP activity [247], etc.) (see Fig. 7.69e).

Measurement Methods for FRET

Using Fluorescence Intensity

The basic method for measuring FRET is to measure the fluorescence intensities of the donor and acceptor, using wide-field or confocal fluorescence microscopy. The idea is to measure the fluorescence intensity for three filter combinations: (1)/(2) (donor filter), (1)/(4) (FRET filter), and (3)/(4) (acceptor filter) (see Fig. 7.70). These intensities do not separate the donor and acceptor. [For example, in the FRET filter, one measures the intensity of the acceptor after exciting the donor, added to the intensity of the donor passing through filter (4).] To correct, one must measure the donor alone and

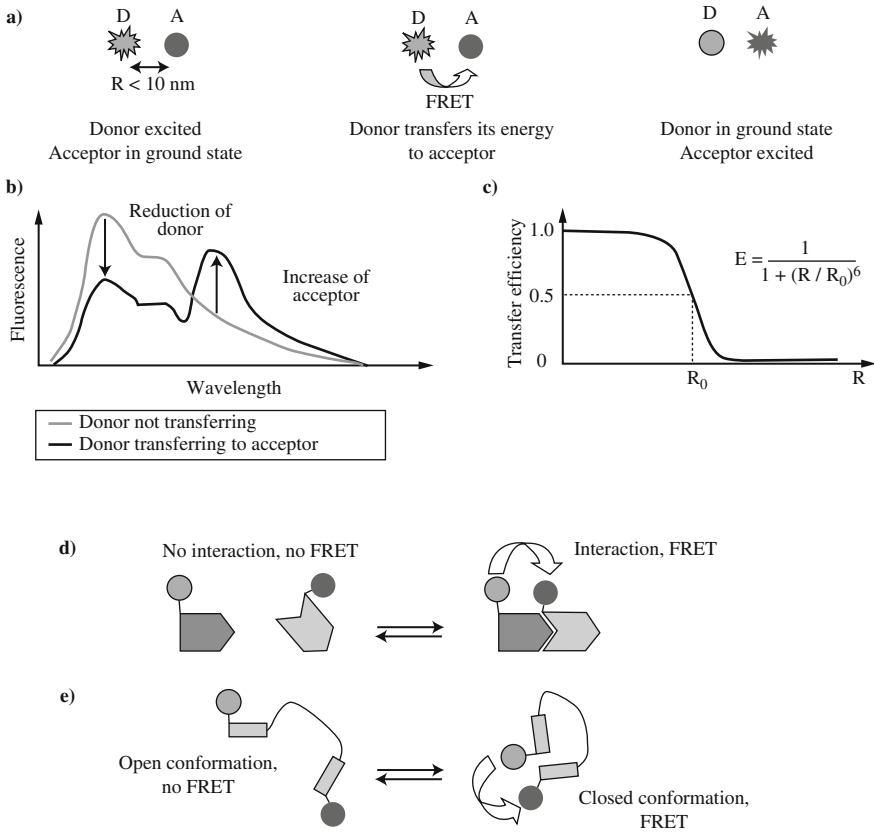


Fig. 7.69. FRET theory. (a) FRET decay diagram for the donor. (b) Emission spectra of the donor and acceptor in the absence and presence of FRET. (c) FRET efficiency as a function of donor-acceptor separation distance. (d) Intermolecular FRET to study protein-protein interactions. (e) Intramolecular FRET to study protein conformations

the acceptor alone. A qualitative measure of the reduction in intensity of the donor and the increase in intensity of the acceptor is then obtained [248]. But in order to make a quantitative measurement of the FRET efficiency, other measurements are required to take concentrations into account [249]. While this method has been broadly validated for in vitro measurements, the fact that there is no control over donor and acceptor concentrations makes it more delicate in cells, except in the special case of probes carrying both donor and acceptor, where the concentrations are necessarily the same and where the fluorescence ratio between the FRET filter and donor filter corresponds to a FRET measurement [246, 247]. Other measurement methods have thus been devised.

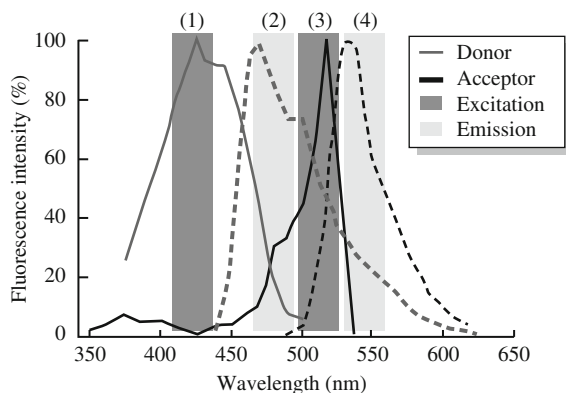


Fig. 7.70. Spectral characteristics of donor, FRET, and acceptor filters for FRET measurement by intensity under the microscope

Using Photobleaching of the Acceptor

This involves measuring the donor fluorescence in the presence of the acceptor and then measuring it again after photodestruction of the acceptor by powerful irradiation of the sample at the acceptor excitation wavelength, and hence in the absence of the acceptor (see Fig. 7.71). Measurements are generally made using a confocal microscope since then, by increasing the laser power, the acceptor can be specifically photobleached. The images of the donor before and after photobleaching the acceptor can be used to calculate an image of the FRET efficiency in the region where the acceptor has been photodestroyed. This gives

$$E = 1 - \frac{F_{\text{before}}}{F_{\text{after}}},$$

where F_{before} and F_{after} correspond to the donor fluorescence intensity before and after photobleaching of the acceptor. This approach allows one to ignore variations in the donor concentration, since the fluorescence intensity of the donor is measured without and with FRET on the same sample [250]. Many experimental difficulties nevertheless remain, especially with regard to the photodestruction of the acceptor, which can generate a photoconversion biasing the measurement (for a review of the photophysical properties of GFP, see [251]), but also with regard to the irreversibility of photobleaching, which rules out monitoring over time.

Using the Fluorescence Lifetime of the Donor

Fluorescence is a transient phenomenon. A fluorescent molecule remains in the excited state for a few nanoseconds only. This fluorescence lifetime is characteristic of the various phenomena affecting its de-excitation. In the case

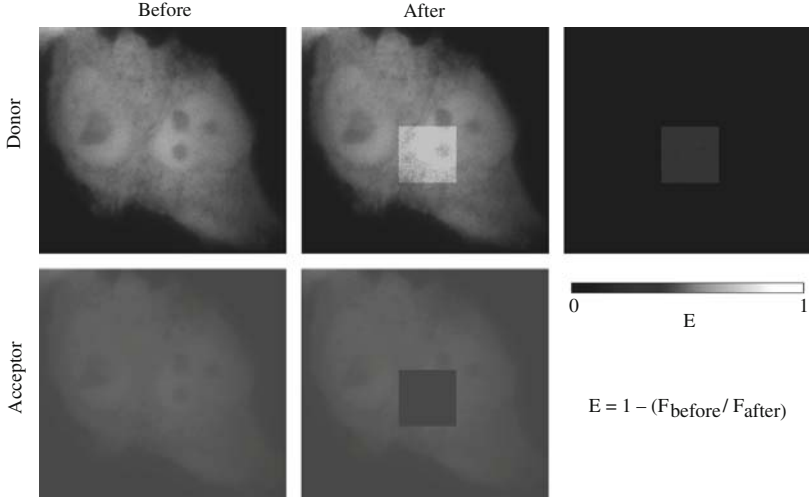


Fig. 7.71. FRET measurement by photobleaching the acceptor. The intensity of the donor is measured before and after photodestruction of the acceptor. The FRET image in the photobleached region is calculated from the donor intensity images before and after photobleaching

where there is no FRET, the fluorescence lifetime τ_D of the donor is given by the reciprocal of the fluorescence decay constant k_{fluo} , viz.,

$$\tau_D = \frac{1}{k_{\text{fluo}}} .$$

In contrast, when there is FRET, the fluorescence lifetime τ_{DA} of the donor in the presence of the acceptor is reduced by the de-excitation of the donor due to FRET:

$$\tau_{DA} = \frac{1}{k_{\text{fluo}} + k_{\text{FRET}}} .$$

The reduction of the donor fluorescence lifetime in the presence of the acceptor as compared with its lifetime in the absence of the acceptor is one of the signatures of FRET. The transfer efficiency is then given by

$$E = 1 - \frac{\tau_{DA}}{\tau_D} .$$

One advantage of the fluorescence lifetime measurement is that it is independent of the concentration. One sample of the donor alone and another sample of the donor in the presence of the acceptor are compared without regard for concentrations. In the example shown in Fig. 7.72, the fluorescence lifetime measurement is carried out in imaging mode (fluorescence lifetime imaging microscopy or FLIM), using the technique of counting time- and space-correlated

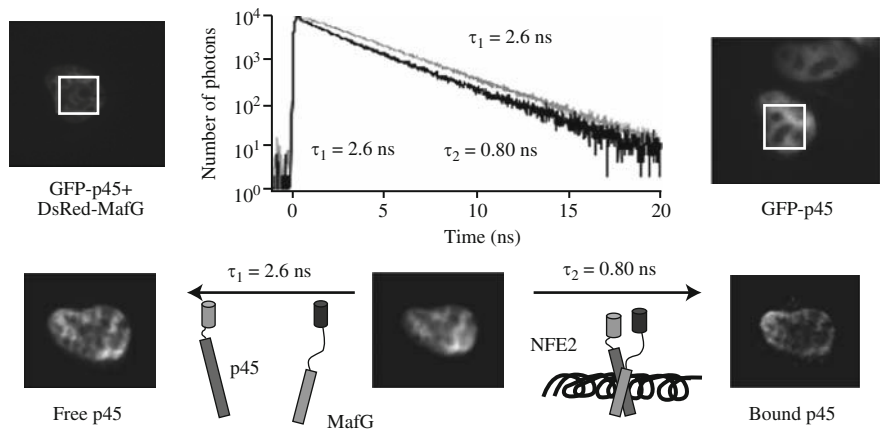


Fig. 7.72. FRET measurement by donor fluorescence lifetime. Using FLIM, the fluorescence decay of the donor can be measured alone and in the presence of the acceptor (*green and red curves*). In this example, the protein p45 fused with GFP is expressed alone or in the presence of the protein MafG fused with DsRed. By analysing these decays, one can separate the lifetime of the GFP alone (2.6 ns) from that of the GFP close to DsRed (0.8 ns or 4.1 nm). Using FLIM, it is possible to image the respective contributions from these two lifetimes, this corresponding to the image of free p45 and p45 bound to MafG. p45 and MafG are the two subunits of the transcription factor NFE2

single photons [252,253]. Apart from the fact that this method is particularly well suited to monitoring over time, it is the only method allowing accurate quantification of the FRET phenomenon. Indeed, when the sample is not uniform, e.g., when studying a protein–protein interaction and only a fraction of the proteins labelled with the donor interact with their partner, a measurement of FRET efficiency by intensity or photobleaching of the acceptor will yield an average over the different species (donors which transfer and which do not transfer), whereas these species can be separated by measuring fluorescence lifetimes (see Fig. 7.72) [254]. Now, the biologically relevant quantity when studying interactions between proteins is the proportion of proteins that actually interact, and only the fluorescence lifetime method can deliver this.

Conclusion

The quantification of fluorescent signals involves many difficulties and FRET is by its very nature an approach requiring accurate quantification. There is no doubt that donor fluorescence lifetime measurements are the most promising for applications to cells under the microscope. Many commercial solutions exist today to carry out such measurements, opening up good prospects for this tool in biology.

Acknowledgements

The authors would like to thank Daniel Choquet who supported this work, Christelle Breillat, Françoise Rossignol, and Delphine Bouchet for the neuron cultures and molecular biology, Edouard Saint-Michel for the QD experiments, Julien Falk for the NrCAM-GFP plasmids, but also the CNRS, French Ministry of Research, and the Aquitaine regional council for financial support.

References

Section One. Introduction to Fluorescence Microscopy

1. www.microscopyu.com/articles/fluorescence/fluorescenceintro.html (25/01/06)
2. www.omegafilters.com/front/curvomatic/spectra.php (25/01/06)
3. www.microscopyu.com/articles/confocal/confocalintrobasics.html (25/01/06)
4. www.microscopyu.com/articles/fluorescencemultiphoton/multiphotonintro.html (25/01/06)
5. Minsky, M.: Microscopy Apparatus, U.S. Patent 301467, issued December, 1961. The 1956 patent application for the Confocal Scanning Microscope is partially reproduced in the scanning article

Section Two. Labels

6. Lakowicz, J.R.: *Principles of Fluorescence Spectroscopy*, 2nd edn., New York, Kluwer Academic/Plenum Publishers (1999)
7. Vo-Dinh, T.: *Biomedical Photonics Handbook*, D.C. Press (2003)
8. Frangioni, J.V.: In vivo near-infrared fluorescence imaging, *Current Opinion in Chemical Biology* **7**, 626–634 (2003)
9. Metelev, V., Weissleder R., Bogdanov, A.: Synthesis and properties of fluorescent NF- κ B-recognizing hairpin oligodeoxyribonucleotide decoys, *Bioconjugate Chem.* **15** (6), 1481–1487 (2004)
10. Haugland, R.: *Handbook of Fluorescent Probes and Research Products*, 9th edn., Molecular Probes Inc. (2002)
11. Hemmilä, I., Harju, R.: In: *Bioanalytical Applications of Labeling Technologies*, ed. by I. Hemmilä, T. Stahlberg, P. Morttram, Walac Oy and EG&G Cie Pb. (1995) Chap. 5
12. Hemmilä, I., Webb, S.: Time-resolved fluorometry: An overview of the labels and core technologies for drug screening applications, *Drug Discovery Today* **2** (9), 373–381 (1997)
13. Piszczek, G., et al.: Multi-photon sensitized excitation of near infrared emitting lanthanides, *Journal of Fluorescence* **12** (1), 15–17 (2002)
14. White, G.F., et al.: Multi-photon excited luminescence of a lanthanide ion in a protein complex: Tb³⁺ bound to transferrin, *Photochem. Photobiol. Sci.* **3**, 47–55 (2004)
15. www.htrf-assays.com

16. Chan, W.C., et al.: Luminescent quantum dots for multiplexed biological detection and imaging, *Current Opinion in Biotechnology* **13**, 40–46 (2002)
17. Gao, X., et al.: In vivo molecular and cellular imaging with quantum dots, *Current Opinion in Biotechnology* **16**, 63–72 (2005)
18. Smith, A.M., Gao, X., Nie, S.: Quantum dot nanocrystals for in vivo molecular and cellular imaging, *Photochemistry and Photobiology* **80**, 377–385 (2004)
19. Michalet, X., et al.: Quantum dots for live cells, in vivo imaging, and diagnostics, *Science* **307**, 538–544 (2005)
20. Green, M.: Semiconductor quantum dots as biological imaging agents, *Angew. Chem. Int. Ed.* **43**, 4129–4131 (2004)
21. Chan, W.C., Nie, S.: Quantum dot bioconjugates for ultrasensitive nonisotopic detection, *Science* **281**, 2016–2018 (1998)
22. Sargent, E.: Infrared quantum dots, *Adv. Mater.* **17** (5), 515–522 (2005)
23. Kim, S., et al.: Type II quantum dots: CdTe/CdSe (core/shell) and CdSe/ZnTe (core/shell) heterostructures, *J. Am. Chem. Soc.* **125**, 11466–11467 (2003)
24. Bailey, R.E., Nie, S.: Alloyed semiconductor quantum dots: Tuning the optical properties without changing the particle size, *J. Am. Chem. Soc.* **125**, 7100–7106 (2003)
25. Gerion, D., et al.: Synthesis and properties of biocompatible water-soluble silica-coated CdSe/ZnS semiconductor quantum dots, *J. Phys. Chem. B* **105**, 8861–8871 (2001)
26. Bruchez, M.P., et al.: Semiconductor nanocrystals as fluorescent biological labels, *Science* **281**, 2013–2016 (1998)
27. Han, M., et al.: Quantum-dot-tagged microbeads for multiplexed optical coding of biomolecules, *Nature Biotech.* **19**, 631–635 (2001)
28. Rosi, N., Mirkin, C.: Nanostructures in biodiagnostics, *Chemical Review* **105** (4), 1547–1562 (2005)
29. Gaponik, N., et al.: Toward encoding combinatorial libraries: Charge driven microencapsulation of semiconductor nanocrystals luminescing in the visible and near IR, *Adv. Mater.* **14** (12), 879–882 (2002)
30. Kim, S., Bawendi, M.G.: Oligomeric ligands for luminescent and stable nanocrystal quantum dots, *J. Am. Chem. Soc.* **125**, 14652–14653 (2003)
31. Pathak, S., et al.: Hydroxylated quantum dots as luminescent probes for in-situ hybridization, *J. Am. Chem. Soc.* **123**, 4103–4104 (2001)
32. Mattoussi, H., et al.: Self-assembly of CdSe–ZnS quantum dot bioconjugates using an engineered recombinant protein, *J. Am. Chem. Soc.* **122**, 12142–12150 (2000)
33. Larson, J.P., et al.: Water-soluble quantum dots for multiphoton fluorescence imaging in vivo, *Science* **300**, 1434–1436 (2003)
34. Kim, S., et al.: Near-infrared fluorescent type II quantum dots for sentinel lymph node mapping, *Nature Biotechnology* **22** (1), 93–97 (2004)
35. Parak, W.J., et al.: Cell motility and metastatic potential studies based on quantum dot imaging of phagokinetic tracks, *Advanced Materials* **14** (12), 882–885 (2002)
36. Jaiswal, J.K., et al.: Long-term multiple color imaging of live cells using quantum dot bioconjugates, *Nature Biotechnology* **21**, 47–51 (2003)
37. Derfus, A.M., Chan, W.C., Bhatia, S.N.: Intracellular delivery of quantum dots for live cell labeling and organelle tracking, *Advanced Materials* **16** (12), 961–966 (2004)

38. Akerman, M.E., et al.: Nanocrystal targeting in vivo, *PNAS* **99** (20), 12617–12621 (2002)
39. Ballou, B., et al.: Non invasive imaging of quantum dots in mice, *Bioconjugate Chem.* 79–86 (2004)
40. Dubertret, B., et al.: In vivo imaging of quantum dots encapsulated in phospholipid micelles, *Science* **298**, 1759–1762 (2002)
41. Gao, X., et al.: In vivo cancer targeting and imaging with semiconductor quantum dots, *Nature Biotech.* **22** (8), 969–976 (2004)
42. Hoshino, A., et al.: Applications of T-lymphoma labeled with fluorescent quantum dots to cell tracing markers in mouse body, *Biochem. Biophys. Res. Comm.* **314**, 46–53 (2004)
43. Derfus, A.M., Chan, W.C., Bhatia, S.N.: Probing the cytotoxicity of semiconductor quantum dots, *Nano Letters* **4** (1), 11–18 (2004)
44. Gerion, D., et al.: Sorting fluorescent nanocrystals with DNA, *J. Am. Chem. Soc.* **124**, 7070 (2002)
45. Wu, X., et al.: Immunofluorescent labeling of cancer marker Her2 and other cellular targets with semiconductor quantum dots, *Nature Biotechnology* **21**, 41–46 (2003)
46. Jaiswal, J.K., et al.: Use of quantum dots for live cell imaging, *Nature Methods* **1** (1), 73–78 (2004)
47. Mitchell, P.: Turning the spotlight on cellular imaging, *Nat. Biotechnol.* **19**, 1013–1017 (2001)
48. Auzel, F.: Upconversion and anti-Stokes processes with *f* and *d* ions in solids, *Chem. Rev.* **104**, 139–173 (2004)
49. Zijlmans, H., et al.: Detection of cell and tissue surface antigens using up-converting phosphors: A new reporter technology, *Analytical Biochemistry* **267**, 30–36 (1999)
50. Ostermayer, F.: Preparation and properties of infrared-to-visible conversion phosphors, *Metall. Trans.* **752**, 747–755 (1971)
51. Corstjens, P., et al.: Lateral-flow and up-converting phosphor reporters to detect single-stranded nucleic acids in a sandwich-hybridization assay, *Analytical Biochemistry* **312**, 191–200 (2003)
52. Niebdala, R.S., et al.: Detection of analytes by immunoassay using up-converting phosphor technology: A new reporter technology, *Analytical Biochemistry* **293**, 22–30 (2001)
53. Van de Rijke, F., et al.: Up-converting phosphor reporters for nucleic acid microarrays, *Nature Biotechnology* **19**, 273–276 (2001)
54. Santra, S., et al.: Luminescent nanoparticle probes for bioimaging, *J. Nanosci. Nanotech.* **4** (6), 590–599 (2004)
55. Battersby, B., et al.: Optical barcoding of colloidal suspensions: Applications in genomics, proteomics and drug discovery, *Chem. Commun.* 1435–1441 (2002)
56. Santra, S., et al.: TAT conjugated, FITC doped silica nanoparticles for bioimaging applications, *Chem. Commun.* 2810–2811 (2004)
57. Ow, H., et al.: Bright and stable core-shell fluorescent silica nanoparticles, *Nano Letters* **5** (1), 113–117 (2005)
58. Li, Z.F., Ruckenstein, E.: Water-soluble poly(acrylic acid) grafted luminescent silicon nanoparticles and their use as fluorescent biological staining labels, *Nano Letters* **4** (8), 1463–1467 (2004)

59. Roy, I., et al.: Optical tracking of organically modified silica nanoparticles as DNA carriers: A nonviral, nanomedicine approach for gene delivery, *PNAS* **102** (2), 279–284 (2005)
60. Sastry, M., et al.: New approaches to the synthesis of anisotropic, core-shell and hollow metal nanostructures, *J. Mater. Chem.* **15**, 3161–3174 (2005)
61. Kelly, K., et al.: The optical properties of metal nanoparticles: The influence of size, shape and dielectric environment, *J. Phys. Chem. B* **107**, 668–677 (2003)
62. Elghanian, R., et al.: Selective colorimetric detection of polynucleotides based on the distance-dependent optical properties of gold nanoparticles, *Science* **277**, 1078–1081 (1997)
63. Aslan, K., et al.: Metal-enhanced fluorescence: An emerging tool in biotechnology, *Current Opinion in Biotechnology* **16**, 55–62 (2005)
64. Cao, Y., Jin, R., Mirkin, C.: Nanoparticles with Raman spectroscopic fingerprints for DNA and RNA detection, *Science* **297**, 1536–1540 (2002)
65. Paroo, Z., et al.: Validating bioluminescence imaging as a high-throughput, quantitative modality for assessing tumor burden, *Molecular Imaging* **3** (2), 117–124 (2004)
66. Ciana, P., et al.: Engineering of a mouse for the in vivo profiling of estrogen receptor activity, *Mol. Endocrinol.* **15**, 1104–1113 (2001). For a review: Maggi, A., Ciana, P.: Reporter mice and drug discovery and development, *Nat. Rev. Drug Discov.* **4** (3), 249–255 (2005)
67. Clemens, T.L., et al.: Analysis of osteocalcin expression in transgenic mice reveals a species difference in vitamin D regulation of mouse and human osteocalcin genes, *J. Bone Miner. Res.* **12**, 1570–1576 (1997)
68. De Wet, J.R., et al.: Cloning of firefly luciferase cDNA and the expression of active luciferase in *Escherichia coli*, *Proc. Natl. Acad. Sci. USA* **82**, 7869–7873 (1985)
69. Wood, K.V., et al.: Synthesis of active firefly luciferase by in vitro translation of RNA obtained from adult lanterns, *Biochem. Biophys. Res. Comm.* **124**, 592–596 (1984)
70. Contag, C.H., et al.: Visualizing gene expression in living mammals using a bioluminescent reporter, *Photochem. Photobiol.* **66** (4), 523–531 (1997)
71. Matthews, J.C., Hori, K., Cormier, M.J.: Purification and properties of *Renilla reniformis* luciferase, *Biochemistry* **16**, 85–91 (1977)
72. Shimomura, O.: The discovery of aequorin and green fluorescent protein, *J. Microsc.* **217**, 1–15 (2005)
73. Chalfie, M., et al.: Green fluorescent protein as a marker for gene expression, *Science* **263**, 802–805 (1994)
74. Shaner, N.C., et al.: Improved monomeric red, orange and yellow fluorescent proteins derived from *Discosoma* sp. red fluorescent protein, *Nat. Biotechnol.* **22** (12), 1567–1572 (2004)
75. Troy, T., et al.: Quantitative comparison of the sensitivity of detection of fluorescent and bioluminescent reporters in animal models, *Molecular Imaging* **3** (1), 9–23 (2004)
76. Ridgway, E.B., Ashley, C.C.: Calcium transients in single muscle fibers, *Biochem. Biophys. Res. Commun.* **29** (2), 229–234 (1967)
77. Inouye, S., Tsuji, F.I.: Aequorea green fluorescent protein. Expression of the gene and fluorescence characteristics of the recombinant protein, *FEBS Lett.* **341** (2–3), 277–280 (1994)

78. Rogers, K.L., al.: Visualization of local Ca^{2+} dynamics with genetically encoded bioluminescent reporters, *Eur. J. Neurosci.* **21** (3), 597–610 (2005)
79. Wilson, T., Hastings, J.W.: Bioluminescence, *Annu. Rev. Cell. Dev. Biol.* **14**, 197–230 (1998)
80. Johnson, F.H., et al.: Extraction, purification and properties of aequorin, a bioluminescent protein from the luminous hydromedusan, *Aequorea*, *Physiol.* **60**, 85–103 (1962)
81. Johnson, F.H., al.: Extraction, purification and properties of aequorin, a bioluminescent protein from the luminous hydromedusan, *Aequorea*, *J. Cell. Comp. Physiol.* **59**, 223–239 (1962)

Section Three. In Vivo Detection Systems

82. Weissleder, R.: Scaling down imaging: Molecular mapping of cancer in mice, *Nature Reviews* **2**, 1–8 (2002)
83. Massoud, T.F., Gambhir, S.S.: Molecular imaging in living subjects: Seeing fundamental biological processes in a new light, *Genes & Development* **17**, 545–580 (2003)
84. Ishimaru, A.: *Wave Propagation and Scattering in Random Media*, Academic Press, New York (1978)
85. Cheong, W.F., Prael, S.A., Welsh, A.J.: A review of the optical properties of biological tissues, *IEEE J. Quantum Electron.* **26**, 2166–2185 (1990)
86. Weissleder, R.: A clearer vision for in vivo imaging, *Nature Biotechnology* **19**, 316–317 (2001)
87. Hollis, V.S.: Non invasive monitoring of brain tissue temperature by near infrared spectroscopy, Thesis, University of London (2002) pp. 65–66
88. Taroni, P., Torricelli, A., Spinelli, L., Pifferi, A., Arpaia, F., Danesini, G., Cubeddu, R.: Time-resolved optical mammography between 637 and 985 nm: Clinical study on the detection and identification of breast lesions, *Physics in Medicine and Biology* **50**, 2469–2488 (2005)
89. Ntziachristos, V., Weissleder, R.: Charge-coupled-device based scanner for tomography of fluorescent near-infrared probes in turbid media, *Medical Physics* **29**, 803–809 (2002)
90. Weissleder, R., et al.: In vivo imaging of tumors with protease activated near infrared fluorescent probes, *Nature Biotech.* **17**, 375–378 (1999)
91. Texier, I., et al.: Vecteur biologique ciblé fonctionnalisé par une fonction d'imagerie, Patent EN 05 07784 (2005)
92. Razkin, J., Josserand, V., Boturnyn, D., Jin, Z., Dumy, P., Favrot, M.C., Coll, J.-L., Texier, I.: Activatable fluorescent probes for tumour-targeting imaging in live mice, *ChemMedChem*, **1** (10), 1069–1072 (2006); Jin, Z., Razkin, J., Josserand, V., Boturnyn, D., Grichine, A., Texier, I., Favrot, M.-C., Dumy, P., Coll, J.-L.: In vivo non invasive optical imaging of receptor-mediated RGD internalization using self-quenched Cy5-labeled RAFT-c(-RGDfK)₄, *Molecular Imaging* **6** (1), 43–55 (2007)
93. Kim, S., et al.: Near-infrared fluorescent type II quantum dots for sentinel lymph node mapping, *Nature Biotechnology* **22** (1), 93–97 (2004)
94. Kim, S., Bawendi, M.G.: Oligomeric ligands for luminescent and stable nonocrystal quantum dots, *J. Am. Chem. Soc.* **125**, 14652–14653 (2003)

95. Ntziachristos, V., Weissleder, R.: Charge-coupled-device based scanner for tomography of fluorescent near-infrared probes in turbid media, *Medical Physics* **29**, 803–809 (2002)
96. Dunsby, C., French, P.M.W.: Techniques for depth-resolved imaging through turbid media including coherence-gated imaging, *Journal of Physics D: Applied Physics* **36**, R207–R227 (2003)
97. Godavarty, A., Sevick-Muraca, E.M., Eppstein, M.J.: Three-dimensional fluorescence lifetime tomography, *Medical Physics* **32**, 992–1000 (2005)
98. Becker, W., Bergmann, A., Biscotti, G., Rück, A.: Advanced time-correlated single photon counting technique, presented at: *Commercial and Biomedical Applications of Ultrafast Lasers IV* (2004)
99. Zint, C.V., Uhring, W., Torregrossa, M., Cunin, B., Poulet, P.: Streak camera: A multidetector for diffuse optical tomography, *Applied Optics* **42**, 3313–3320 (2003)
100. D’Andrea, C., Comelli, D., Pifferi, A., Torricelli, A., Valentini, G., Cubeddu, R.: Time-resolved optical imaging through turbid media using a fast data acquisition system based on a gated CCD camera, *Journal of Physics D: Applied Physics* **36**, 1675–1681 (2003)
101. Patterson, M.S., Chance, B., Wilson, B.C.: Time resolved reflectance and transmittance for the non invasive measurement of tissue optical properties, *Applied Optics* **28**, 2331–2336 (1989)
102. Liebert, A., Wabnitz, H., Grosenick, D., Moller, M., Macdonald, R., Rinneberg, H.: Evaluation of optical properties of highly scattering media by moments of distributions of times of flight of photons, *Applied Optics* **42**, 5785–5792 (2003)
103. Cubeddu, R., Pifferi, A., Taroni, P., Torricelli, A., Valentini, G.: Experimental test of theoretical models for time-resolved reflectance, *Med. Phys.* **9**, 1625–1633 (1996)
104. Laidevant, A., Da Silva, A., Berger, M., Dinten, J.-M.: Experimental study of time-resolved measurements on turbid media: Determination of optical properties and fluorescent inclusion characterization, *Proceedings of SPIE* **5859** (2005)
105. Taroni, P., Torricelli, A., Spinelli, L., Pifferi, A., Arpaia, F., Danesini, G., Cubeddu, R.: Time-resolved optical mammography between 637 and 985 nm: Clinical study on the detection and identification of breast lesions, *Physics in Medicine and Biology* **50**, 2469–2488 (2005)
106. Cubeddu, R., Pifferi, A., Taroni, P., Torricelli, A., Valentini, G., Rinaldi, F., Sorbellini, E.: Fluorescence lifetime imaging: An application to the detection of skin tumors, *IEEE Journal of Selected Topics in Quantum Electronics* **5**, 923–929 (1999)
107. Lam, S., Lesage, F., Intes, X.: Time domain fluorescent diffuse optical tomography: Analytical expressions, *Optics Express* **13**, 2263–2275 (2005)
108. Ntziachristos, V., Ma, X., Yodh, A.G., Chance, B.: Multichannel photon counting instrument for spatially resolved near infrared spectroscopy, *Review of Scientific Instruments* **70**, 193–201 (1999)

Section Four. In Vitro Detection Systems

109. Pirrung, M.C.: How to make a DNA chip, *Angewandte Chemie International Edition* **41** (8), 1276–1289 (2002)

110. Zuber, E.: Approche biométrique de la dynamique de réactions antigène-anticorps, PhD Thesis, Université Claude Bernard Lyon 1 (1997)
111. Wang, S.P., Grayston, J.T.: Immunologic relationship between genital TRIC, lymphogranuloma venereum, and related organisms in a new microtiter indirect immunofluorescence test, *Am. J. Ophthalmol.* **70** (3), 367–374 (1970)
112. Gallo, G., Riggs, J.L., Schachter, J., Emmons, R.W.: Multiple-antigen slide test for detection of immunoglobulin M antibodies in newborn and infant sera by immunofluorescence, *J. Clin. Microbiol.* **13** (4), 631–636 (1981)
113. Hawkes, R.: Identification of concanavalin A-binding proteins after sodium dodecyl sulfate-gel electrophoresis and protein blotting, *Analytical Biochemistry* **123** (I 1), 143–146 (1982)
114. Ekins, R., Chu, F., Biggart, E.: Development of microspot multi-analyte radiometric immunoassay using dual fluorescent-labelled antibodies, *Analytica Chimica Acta* **227**, 73–96 (1989)
115. Hirschfeld, S.: Assaying for a multiplicity of antigens or antibodies with a detection compound, US4 514 508 (1982)
116. Khrapko, K., Lysov, Yu., Khorlin, A., Shick, V., Florentiev, V., Mirzabekov, A.: An oligonucleotide hybridization approach to DNA sequencing, *FEBS Letters* **256**, 118–122 (1989)
117. Fodor, S.P.A., Read, J.L., Pirrung, M.C., Stryer, L., Lu, A.T., Solas, D.: Light-directed, spatially addressable parallel chemical synthesis, *Science* **251**, 767–773 (1991)
118. Glezos, N., Misiakos, K., Kakabakos, S., Petroub, P., Terzoudic, G.: Electron beam patterning of biomolecules, *Biosensors & Bioelectronics* **17** (4), 279–282 (2002)
119. DeRisi, J.L., Iyer, V.R., Brown, P.O., Exploring the metabolic and genetic control of gene expression on a genomic scale, *Science* **278**, 680–686 (1997)
120. Lashkari, D.A., DeRisi, J.L., McCusker, J.H., Namath, A.F., Gentile, C., Hwang, S.Y., Brown, P.O., Davis, R.W.: Yeast microarrays for genome wide parallel genetic and gene expression analysis, *Proc. Natl. Acad. Sci.* **94**, 13057–13062 (1997)
121. Wastiaux, G.: *La microscopie optique moderne*, Tec et Doc Lavoisier (1994)
122. Wilson, T.: *Confocal Microscopy*, Academic Press (1990)
123. Wodicka, L., Dong, H., Mittmann, M., Ho, M.H., Lockhart, D.J.: Genome-wide expression monitoring in *Saccharomyces cerevisiae*, *Nature Biotechnology* **15**, 1359–1367 (1997)
124. Narang, U., Gauger, P.R., Ligler, F.S.: Capillary-based displacement flow immunosensor, *Anal. Chem.* **69** (10), 1961–1964 (1997)
125. Cesaro-Tadic, S., Dernick, G., Juncker, D., Buurman, G., Kropshofer, H., Michel, B., Fattinger, C., Delamarche, E.: High-sensitivity miniaturized immunoassays for tumor necrosis factor using microfluidic systems, *Lab on a Chip* **4** (6), 563–569 (2004)
126. Mallard, F., Marchand, G., Ginot, F., Campagnolo, R.: Opto-electronic DNA chip: High performance chip reading with an all-electric interface, *Biosensors and Bioelectronics* **20** (9), 1813–1820 (2005)
127. Ligler, F.S., Breimer, M., Golden, J.P., Nivens, D.A., Dodson, J.P., Green, T.M., Haders, D.P., Sadik, O.A.: Integrating waveguide biosensor, *Anal. Chem.* **74** (3), 713–719 (2002)

128. Dill, K., Montgomery, D.D., Ghindilis, A.L., Schwarzkopf, K.R., Ragsdale, S.R., Oleinikov, A.V.: Immunoassays based on electrochemical detection using microelectrode arrays, *Biosensors and Bioelectronics* **20** (4), 736–742 (2004)
129. Yang, J.M., Bell, J., Huang, Y., Tirado, M., Thomas, D., Forster, A.H., Haigis, R.W., Swanson, P.D., Wallace, R.B., Martinsons, B., Krihak, M.: An integrated, stacked microlaboratory for biological agent detection with DNA and immunoassays, *Biosensors and Bioelectronics* **17** (6–7), 605–618 (2002)
130. Rowe, C.A., Tender, L.M., Feldstein, M.J., Golden, J.P., Scruggs, S.B., Mac-Craith, B.D., Cras, J.J., Ligler, F.S.: Array biosensor for simultaneous identification of bacterial, viral, and protein analytes, *Anal. Chem.* **71** (17), 3846–3852 (1999)
131. Tschmelaka, J., Proll, G., Riedt, J., et al.: Automated water analyser computer supported system (AWACSS) Part I: Project objectives, basic technology, immunoassay development, software design and networking, *Biosensors and Bioelectronics* **20** (8), 1499–1508 (2005)
132. Livache, T., Roget, A., Dejean, E., Barthet, C., Bidan, G., Téoule, R.: Preparation of a DNA matrix via an electrochemically directed copolymerization of pyrrole and oligonucleotides bearing a pyrrole group, *Nucleic Acids Res.* **22** (15), 2915–2921 (1994)
133. Holt, D.B., Gauger, P.R., Kusterbeck, A.W., Ligler, F.S.: Fabrication of a capillary immunosensor in polymethyl methacrylate, *Biosensors and Bioelectronics* **17** (1–2), 95–103 (2002)
134. Pease, A.C., Solas, D., Sullivan, E.J., Cronin, M.T., Holmes, C.P., Fodor, S.P.A.: Light-generated oligonucleotide arrays for rapid DNA sequence analysis, *PNAS* **91**, 5022–5026 (1997)
135. Beier, M., Hoheisel, J.H.: Production by quantitative photolithographic synthesis of individually quality checked DNA microarrays, *Nucleic Acids Res.* **28** (4), e11 (2000)
136. Dontha, N., Nowall, W.B., Werner, G., Kuhr, W.G.: Generation of biotin/avidin/enzyme nanostructures with maskless photolithography, *Anal. Chem.* **69** (14), 2619–2625 (1997)
137. Singh-Gasson, S., Green, R.D., Yue, Y., Nelson, C., Blattner, F., Sussman, M.R., Cerrina, F.: Maskless fabrication of light-directed oligonucleotide microarrays using a digital micromirror array, *Nature Biotechnology* **17** (10), 974–978 (1999)
138. Silzel, J.W., Cercek, B., Dodson, C., Tsay, T., Obremski, R.J.: Mass-sensing multianalyte microarray immunoassay with imaging detection, *Clinical Chemistry* **44** (9), 2036–2043 (1998)
139. Conrad, D.W., Davis, A.V., Golightley, S.K., Bart, J.C., Ligler, F.S.: Photoactivatable silanes for the site-specific immobilization of antibodies, *Proc. SPIE Int. Soc. Opt. Eng.* **2978**, 12–21 (1997)
140. Roost, F.W.D.: *Quantitative Fluorescence Microscopy*, Cambridge University Press (1991)
141. Rouessac, F., Rouessac, A.: *Analyse chimique, méthodes et techniques instrumentales modernes*, Masson, Paris (1994)
142. Sabanayagam, C.R., Smith, C.S., Cantor, C.R.: Oligonucleotide immobilization on micropatterned streptavidin surfaces, *NAR* **28** (8), E33–00 (2000)
143. Seliger, H., Hinz, M., Happ, E.: Arrays of immobilized oligonucleotides – Contributions to nucleic acid technology, *Current Pharmaceutical Biotechnology* **4** (6), 379–395 (2003)

144. Ivarsson, B., Ulf, J., Sjolander, S., Stanghlberg, R., Sjodin, H.: Optical biosensor system, US 5313264 (1989)
145. Berg, O.G., Blomberg, C.: Association kinetics with coupled diffusional flows: Special application to the lac repressor-operator system, *Biophysical Chemistry* **4** (I4), 367–381 (1976)
146. Thompson, N.L., Burghardt, T.P., Axelrod, D.: Measuring surface dynamics of biomolecules by total internal reflection fluorescence with photobleaching recovery or correlation spectroscopy, *Biophysical Journal* **33** (3), 435–454 (1981)
147. Benner, R.E., Dornhaus, R., Chang, R.K.: Angular emission profiles of dye molecules excited by surface plasmon waves at a metal surface, *Optics Communications* **30** (2), 145–149 (1979)
148. Attridge, J.W., Daniels, P.B., Deacon, J.K., Robinson, G.A., Davidson, G.P.: Sensitivity enhancement of optical immunosensors by the use of a surface plasmon resonance fluoroimmunoassay, *Biosensors and Bioelectronics* **6**, 201–214 (1991)
149. Sullivan, K.G., King, O., Sigg, C., Hall, D.G.: Directional, enhanced fluorescence from molecules near a periodic surface, *Applied Optics* **33** (13), 2447–2454 (1994)
150. Chance, R.R., Prock, A., Silbey, R.: Molecular fluorescence and energy transfer near interfaces. In: *Advances in Chemical Physics*, Vol. XXXVII, ed. by I. Prigogine, S.R. Rice (1978) pp. 1–65
151. Enderlein, J., Ruckstuhl, T., Seeger, S.: Highly efficient optical detection of surface generated fluorescence, *Applied Optics* **38** (14), 724–732 (1999)
152. Perraut, F., Chaton, P., Pouteau, P.: Amplification d'un signal de fluorescence émis par un échantillon surfacique, WO0140778 (2000)
153. Bras, M., Dugas, V., Bessueille, F., Cloarec, J.P., Martin, J.R., Cabrera, M., Chauvet, J.P., Souteyrand, E., Garrigues, M.: Optimisation of a silicon/silicon dioxide substrate for a fluorescence DNA microarray, *Biosensors and Bioelectronics* **20** (I4), 797–806 (2004)
154. Stambouli, V., Labeau, M., Matko, I., Chenevier, B., Renault, O., Guiducci, C., Chaudouët, P., Roussel, H., Nibkin, D., Dupuis, E.: Development and functionalisation of Sb doped SnO₂ thin films for DNA biochip applications, *Sensors and Actuators B: Chemical* **113** (I2), 1025–1033 (2006)
155. Lakowicz, J.R., Malicka, J., Gryczynski, I., Gryczynski, Z., Geddes, C.D.: Radiative decay engineering: The role of photonic mode density in biotechnology, *Journal of Physics D: Applied physics* **36**, R240–R249 (2003)
156. Neuschäfer, D., Budach, W., Wanke, C., Chibout, S.D.: Evanescent resonator chips: A universal platform with superior sensitivity for fluorescence-based microarrays, *Biosensors and Bioelectronics* **18** (I4), 489–497 (2003)
157. Unger, M., Kartalov, E., Chiu, C.S., Lester, H.A., Quake, S.R.: Single molecule fluorescence observed with mercury lamp illumination, *Biotechniques* **27** (5), 1008–1014 (1999)
158. Yershov, G., Barsky, V., Belgovskiy, A., Kirillov, E., Kreindlin, E., Ivanov, I., Parinov, S., Guschin, D., Drobishev, A., Dubiley, S., Mirzabekov, A.: DNA analysis and diagnostics on oligonucleotide microchips, *PNAS* **93** (10), 4913–4918 (1996)
159. Hellen, E.H., Axelrod, A.: Fluorescence emission at dielectric and metal-film interfaces, *JOSA* **14**, 337–350 (1987)

160. Hindson, B.J., Brown, S.B., Marshall, G.D., McBride, M.T., Makarewicz, A.J., Gutierrez, D.M., Wolcott, D.K., Metz, T.R., Madabhushi, R.S., Dzenitis, J.M., Colston, B.W.: Development of an automated sample preparation module for environmental monitoring of biowarfare agents, *Anal. Chem.* **76** (13), 3492–3497 (2004)
161. Marsoner, H., Karpf, H., Leitner, A.: Sensor element for determination of concentration of substances, US 4755667 (1988)
162. Marsoner, H., Kroneis, H., Karpf, H., Wolfbeis, O., List, H., Leitner, A.: Sensor element for determination of concentration of substances, US 5039490 (1991)
163. Misiakos, K., Kakabakos, S.E., Petrou, P.S., Ruf, H.H.: A monolithic silicon optoelectronic transducer as a real-time affinity biosensor, *Anal. Chem.* **76** (5), 1366–1373 (2004)
164. Park, S., Taton, T.A., Mirkin, C.A. Array-based electrical detection of DNA with nanoparticle probes, *Science* **295** (5559), 1503–1506 (2002)
165. Singhal, P., Kuhr, W.G.: Ultrasensitive voltammetric detection of underivatized oligonucleotides and DNA, *Anal. Chem.* **69** (23), 4828–4832 (1997)
166. Popovich, N.D., Eckhardt, A.E., Mikulecky, J.C., Napier, M.E., Thomas, R.S.: Electrochemical sensor for detection of unmodified nucleic acids, *Talanta* **56** (5), 821–828 (2002)
167. Sheehan, P.E., Whitman, L.J.: Detection limits for nanoscale biosensors, *Nano Letters* **5** (4), 803–807 (2005)
168. Li, Z., Chen, Y., Li, X., Kamins, T.I., Nauka, K., Williams, R.S.: Sequence-specific label-free DNA sensors based on silicon nanowires, *Nano Letters* **4** (2), 245–247 (2004)
169. Hahm, J., Lieber, C.M.: Direct ultrasensitive electrical detection of DNA and DNA sequence variations using nanowire nanosensors, *Nano Letters* **4** (1), 51–54 (2004)
170. Lu, G.N., Ben Chouikha, M., Sou, G., Sedjil, M.: Colour detection using a buried double p–n junction structure implemented in the CMOS process, *Electron. Lett.* **213**, 594–596 (1996)
171. Eggers, M., Hogan, M., Reich, R.K., Lamture, J., Ehrlich, D., Hollis, M., Kosicki, B., Powdrill, T., Beattie, K., Smith, S., et al.: A microchip for quantitative detection of molecules utilizing luminescent and radioisotope reporter groups, *Biotechniques* **17** (3), 516–525 (1994)
172. Perraut, F., Lagrange, A., Pouteau, P., Peyssonneaux, O., Puget, P., McGall, G., Menou, L., Gonzalez, R., Labeye, P., Ginot, F.: A new generation of scanners for DNA chips, *Biosensors and Bioelectronics* **17**, 803–813 (2002)
173. Vo-Dinh, T., Alarie, J.P., Isola, N., Landis, D., Wintenberg, A.L., Ericson, M.N.: DNA biochip using a phototransistor integrated circuit, *Anal. Chem.* **74** (2), 358–363 (1999)
174. Lim, D.V.: Detection of microorganisms and toxins with evanescent wave fiber-optic biosensors, *Proceedings of the IEEE* **91** (6), 902–907 (2003)
175. Stimpson, D.I., Hoijer, J.V., Hsieh, W.T., Jou, C., Gordon, J., Theriault, T., Gamble, R., Baldeschwieler, J.D.: Real-time detection of DNA hybridization and melting on oligonucleotide arrays by using optical wave guides, *PNAS* **92** (14), 6379–6383 (1995)
176. Duveneck, G.L., Pawlak, M., Ehrat, M., Neuschäfer, D., Bär, E., Budach, W., Pielies, U.: Novel bioaffinity sensors for trace analysis based on luminescence excitation by planar waveguides, *Sensors and Actuators B: Chemical* **38** (1–3), 88–95 (1997)

177. Hofmann, O., Voirin, G., Niedermann, P., Manz, A.: Three-dimensional microfluidic confinement for efficient sample delivery to biosensor surfaces. Application to immunoassays on planar optical waveguides, *Anal. Chem.* **74** (20), 5243–5250 (2002)
178. Labeye, P., Pouteau, P., Perraut, F., Ginot, F.: Dispositif de lecture de fluorescence intégrée, FR 2846420 (2002)
179. Lundgreen, J.S., Neal Watkins, A., Racz, D., Ligler, F.S.: A liquid crystal pixel array for signal discrimination in array biosensors, *Biosensors and Bioelectronics* **15**, 417–421 (2000)
180. Choudhury, B.J., Shinar, R., Shinar, J.: Luminescent chemical and biological sensors based on the structural integration of an OLED excitation source with a sensing component. In: *Organic Light-Emitting Materials and Devices VII*, ed. by Z.H. Kafafi, P.A. Lane, Proceedings of the SPIE **5214**, 64–72 (2004)
181. Edel, J.B., Beard, N.P., Hofmann, O., deMello, J.C., Bradley, D.D.C., deMello, A.J.: Thin-film polymer light emitting diodes as integrated excitation sources for microscale capillary electrophoresis, *Lab on a Chip* **4** (2), 136–140 (2004)
182. Savvate'ev, V., Chen-Esterlit, Z., Aylott, J.W., Choudhury, B., Kim, C.H., Zou, L., Friedl, J.H., Shinar, R., Shinar, J., Kopelman, R.: Integrated organic light-emitting device/fluorescence-based chemical sensors, *Applied Physics Letters* **81** (24), 4652–4654 (2002)
183. Chediak, J.A., Luo, Z., Seo, J., Cheung, N., Lee, L.P., Sands, T.D.: Heterogeneous integration of CdS filters with GaN LEDs for fluorescence detection microsystems, *Sensors and Actuators A: Physical* **111** (11), 1–7 (2004)

General Reading

184. Kamoun, P.: *Appareils et méthodes en biochimie et biologie moléculaire*, Flammarion Médecine-sciences, Paris (1997)
185. Lacowicz, J.R.: *Principles of Fluorescence Spectroscopy*, Kluwer Academic, Dordrecht (1999)
186. Péré, J.P.: *La microscopie*, Nathan, Paris (1994)

Section Five. Other Detection Systems

187. Lippincott-Schwartz, J., Snapp, E., Kenworthy, A.: *Nat. Rev. Mol. Cell. Biol.* **2** (6), 444–456 (2001)
188. Crank, J.: *The Mathematics of Diffusion*, Oxford University Press, New York (1975)
189. Axelrod, D., et al.: *Biophys. J.* **16** (9), 1055–1069 (1976)
190. Feder, T.J., et al.: *Biophys. J.* **70** (6), 2767–2773 (1996)
191. Edidin, M., Zuniga, M.C., Sheetz, M.P.: *Proc. Natl. Acad. Sci. USA*, **91** (8), 3378–3382 (1994)
192. Edidin, M., Stroynowski, I.: *J. Cell. Biol.* **112** (6), 1143–1150 (1991)
193. Kenworthy, A.K., et al.: *J. Cell. Biol.* **165** (5), 735–746 (2004)
194. Cole, N.B., et al.: *Science* **273** (5276), 797–801 (1996)
195. Marguet, D., et al.: *Immunity* **11** (2), 231–240 (1999)
196. Dunn, G.A., et al.: *J. Microsc.* **205** (Pt 1), 109–112 (2002)
197. Mallavarapu, A., Mitchison, T.: *J. Cell. Biol.* **146** (5), 1097–1106 (1999)
198. McGrath, J.L., et al.: *Biophys. J.* **75** (4), 2070–2078 (1998)

199. Patterson, G.H., Lippincott-Schwartz, J.: *Science* **297** (5588), 1873–1877 (2002)
200. Chudakov, D.M., et al.: *Nat. Biotechnol.* **22** (11), 1435–1439 (2004)
201. Munnely, H.M., et al.: *Biophys. J.* **75** (2), 1131–1138 (1998)
202. Davoust, J., Devaux, P.F., Leger, L.: *EMBO J.* **1** (10), 1233–1238 (1982)
203. Sanchez, S.A., Gratton, E.: *Acc. Chem. Res.* **38** (6), 469–477 (2005)
204. Bacia, K., et al.: *Biophys. J.* **87** (2), 1034–1043 (2004)
205. Lenne, P.-F., et al.: *EMBO J.* **25** (14), 3245–3256 (2006)
206. Margeat, E., et al.: *J. Mol. Biol.* **306** (3), 433–442 (2001)
207. Chen, Y., et al.: *Biophys. J.* **77** (1), 553–567 (1999)
208. Hess, S.T., et al.: *Biochemistry* **41** (3), 697–705 (2002)
209. Rigler, R., Elson, E.: *Fluorescence Correlation Spectroscopy: Theory and Applications*, Springer Series in Chemical Physics, ed. by F.P. Schäfer, W. Zinth, J.P. Toennies, Vol. 65, Springer, Berlin Heidelberg New York (2001)
210. Magde, D., Elson, E.L., Webb, W.W.: *Phys. Rev. Lett.* **29**, 705 (1972)
211. Rigler, R., et al.: *Eur. Biophys. J.* **22**, 169–175 (1993)
212. Elson, E.L., Magde, D.: *Biopolymers* **13** (1), 1–27 (1974)
213. Schwille, P., et al.: *Biophys. J.* **77** (4), 2251–2265 (1999)
214. Schwille, P., et al.: *Proc. Natl. Acad. Sci. USA* **97** (1), 151–156 (2000)
215. Haupts, U., et al.: *Proc. Natl. Acad. Sci. USA* **95** (23), 13573–13578 (1998)
216. Chen, Y., et al.: *Biophys. J.* **82** (1), 133–144 (2002)
217. Levene, M.J., et al.: *Science* **299** (5607), 682–686 (2003)
218. Wenger, J., et al.: *Biophys. J.* **92** (3), 913–919 (2007)
219. Edel, J.B., et al.: *Biophys. J.* **88** (6), L43–L45 (2005)
220. Ruan, Q., et al.: *Biophys. J.* **87** (2), 1260–1267 (2004)
221. Digman, M.A., et al.: *Biophys. J.* **89** (2), 1317–1327 (2005)
222. Wawrezinieck, L., et al.: *Proc. SPIE* **5462**, 92–102 (2004)
223. Kusumi, A., Sako, Y., Yamamoto, M.: *Biophys. J.* **65** (5), 2021–2040 (1993)
224. Simson, R., et al.: *Biophys. J.* **74** (1), 297–308 (1998)
225. Felsenfeld, D.P., Choquet, D., Sheetz, M.P.: *Nature* **383** (6599), 438–440 (1996)
226. Dietrich, C., et al.: *Biophys. J.* **82** (1 Pt 1), 274–284 (2002)
227. Vereb, G., et al.: *Proc. Natl. Acad. Sci. USA* **100** (14), 8053–8058 (2003)
228. Kucik, D.F., Elson, E.L., Sheetz, M.P.: [published erratum appears in *Biophys. J.* **76** (3), 1720 (Mar 1999)] *Biophys. J.* **76** (1 Pt 1), 314–322 (1999)
229. Ueda, M., et al.: *Science* **294**, 864–867 (2001)
230. Schutz, G.J., et al.: *EMBO J.* **19** (5), 892–901 (2000)
231. Seisenberger, G., et al.: *Science* **294**, 1929–1932 (2001)
232. Harms, G.S., et al.: *Biophys. J.* **81** (5), 2639–46 (2001)
233. Tardin, C., et al.: *EMBO J.* **22** (18), 4656–4665 (2003)
234. Schmidt, C.E., et al.: *J. Neurosci.* **15** (5 Pt 1), 3400–3407 (1995)
235. Thompson, R.E., Larson, D.R., Webb, W.W.: *Biophys. J.* **82** (5), 2775–2783 (2002)
236. Axelrod, D.: *Methods Cell. Biol.* **30**, 245–270 (1989)
237. Dickson, R.M., et al.: *Science* **274**, 966–969 (1996)
238. Bruchez, M., Jr, et al.: *Science* **281**, 2013–2016 (1998)
239. Dahan, M., et al.: *Science* **302**, 442–445 (2003)
240. Boyer, D., et al.: *Science* **297**, 1160–1163 (2002)
241. Berciaud, S., et al.: *Phys. Rev. Lett.* **93** (25), 257402 (2004)

- 242. Cognet, L., et al.: *Proc. Natl. Acad. Sci. USA* **100** (20), 11350–11355 (2003)
- 243. Förster, T.: *Ann. Phys.* **2**, 55–75 (1948)
- 244. Tsien, R.Y.: *Ann. Rev. Biochem.* **67**, 509–544 (1998)
- 245. Wouters, F.S., Verveer, P.J., Bastiaens, P.I.H.: *Trends Cell. Biol.* **11**, 203–211 (2001)
- 246. Miyawaki, A., et al.: *Nature* **388**, 882–887 (1997)
- 247. Pertz, O., Hahn, K.M.: *J. Cell Sci.* **117**, 1313–1318 (2004)
- 248. Gordon, G.W., et al.: *Biophys. J.* **74**, 2702–2713 (1998)
- 249. Stockholm, D., et al.: *J. Mol. Biol.* **346**, 215–222 (2005)
- 250. Wouters, F.S., et al.: *EMBO J.* **17**, 7179–7189 (1998)
- 251. Subramanian, V., et al.: *Methods Enzymol.* **360**, 178–201 (2003)
- 252. Tramier, M., et al.: *Biophys. J.* **83**, 3570–3577 (2002)
- 253. Emiliani, V., et al.: *Appl. Phys. Lett.* **83**, 2471–2473 (2003)
- 254. Tramier, M., et al.: *Methods Enzymol.* **360**, 580–597 (2003)

Nanoforce and Imaging

C. Le Grimallec, P.-E. Milhiet, E. Perez, F. Pincet, J.-P. Aimé, V. Emiliani, O. Thoumine, T. Lionnet, V. Croquette, J.-F. Allemand, and D. Bensimon

8.1 Molecular and Cellular Imaging Using AFM

8.1.1 Introduction

The invention of the scanning tunneling microscope by Binnig and coworkers [1], who won the Nobel Prize for Physics in 1986, marked the beginning of a new type of microscopy, called near-field or local probe microscopy. These microscopes are powerful tools for studying the surface properties of samples. They all involve scanning the sample surface with a probe or tip at a distance of nanometric order and determining point by point the value of some physical quantity, e.g., electron transfer (scanning tunneling microscope STM), photon transfer (scanning near-field optical microscope SNOM), or an interaction force (atomic force microscope AFM) [2].

AFM can achieve atomic resolution on crystal samples in air or vacuum, but its development in structural biology comes from the fact that it can be made to work in a liquid medium. It is used to observe the surface of biological samples ranging from complex biological structures like plasma membranes in eukaryotic cells to single molecules. At the present time, it is the only technique able to obtain subnanometric resolutions in a physiological environment. Used initially for imaging, it can also serve as a tool for dissection and manipulation on a molecular scale, or for measuring intra- and intermolecular interaction forces (see Sect. 8.3). Through these wide-ranging applications, AFM has become an indispensable tool in the development of nanobiotechnology.

8.1.2 Atomic Force Microscopy

AFM imaging consists in scanning a sample, adsorbed onto a substrate, row by row with a finely tapered tip mounted on a very flexible cantilever and brought into contact or within a few angstrom units from the sample. This tip is generally made from silicon nitride, and the end of the tip has a radius

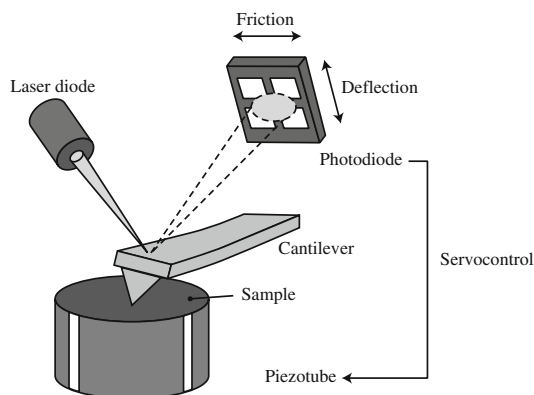


Fig. 8.1. Principle of the atomic force microscope. A laser beam directed at the end of the cantilever is reflected onto a photodiode with two or four quadrants. The displacement of the beam on the photodiode detects the motion of the cantilever. The tip is positioned in and above the sample plane with subatomic precision during scanning. This is done by means of piezoceramics. The system is controlled by a computer, which reconstitutes the data gathered by the tip in the form of a 3D image

of curvature of a few nanometers. The deflection of the cantilever, caused by the tip-sample interaction, is detected by an optical system (four-quadrant photodiode), and this signal is used to minimise the force applied to the sample via a feedback system using piezoelectric ceramics which adjust the height of the sample (see Fig. 8.1). These vertical displacements are then exploited to reconstruct the surface topography of the sample. During scanning, the force of interaction between the sample and the tip is held constant. In a liquid medium, this force results mainly from long-range attractive or repulsive electrostatic forces, attractive van der Waals forces and repulsive interatomic forces.

There are two main imaging modes in biology: contact mode and tapping mode (see below for more details). Contact mode can obtain high-resolution images on reasonably smooth objects (see the review in [3]). However, when studying samples only weakly adsorbed onto a substrate, or with highly irregular topography, like cells, contact mode can cause friction likely to damage or detach the sample. The tapping (intermittent contact) mode was developed precisely to overcome these effects. In this mode, the tip oscillates at high frequency, minimising the contact between tip and sample.

AFM Modes

In Figs. 8.2 and 8.3 the sample is shown by grey shading, deposited on a substrate that is often mica.

- *Contact Mode.* The mode in which the tip and surface remain in permanent contact or quasi-contact is called contact mode. The sample is scanned while

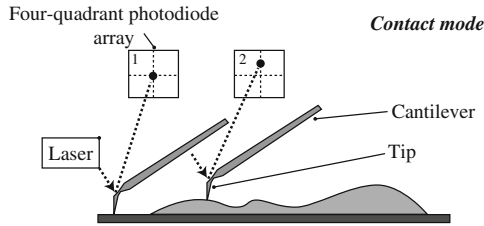


Fig. 8.2. Contact mode atomic force microscopy

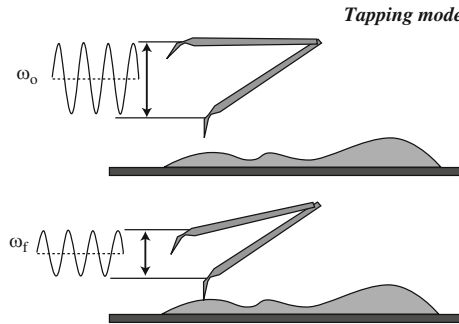


Fig. 8.3. Tapping mode atomic force microscopy

maintaining a constant force between tip and sample. During scanning, a laser beam illuminates the back of the cantilever above the tip, and is reflected toward a photodiode with 4 quadrants (see Fig. 8.2).

In quadrant 1, when the tip is in contact with the substrate, the deflection signal (vertical displacement) and friction signal (horizontal displacement) are both zero. If the tip follows the sample topology (quadrant 2), the deflection signal increases. This signal is used for the servo-control, and the sample height is modified to reposition the laser beam centrally once again (similar to quadrant 1) and maintain the smallest possible force between tip and sample. Three types of image can be acquired simultaneously:

1. Topographic (isoforce) images or height images, corresponding to z displacement of the sample after servo-control.
 2. Deflection images corresponding to the derivative of movements to readjust the relative tip-sample position, which reveal surface details while height information is lost.
 3. Friction images resulting from twisting of the cantilever.
- *Tapping Mode.* The second mode is the tapping mode, in which the tip oscillates at the resonant frequency of the cantilever and only touches the surface in a transient manner. It is the amplitude signal of the oscillation that serves as feedback signal. When the tip is not touching the sample, it oscillates with the so-called free amplitude ω_0 . When the tip is in contact with the sample, the oscillation amplitude decreases and a working amplitude ω_w is chosen, below which the sample height is adjusted by the servosystem (see Fig. 8.3). Four types of image can be acquired:

1. Topographical images or height images.
2. Amplitude images, obtained by recording root mean square (RMS) values of ω .
3. Phase images, obtained by recording a phase shift in the oscillation signal when the tip contacts the sample. This signal depends on the energy dissipated during each transient contact and gives access to local viscoelastic properties.
4. The deflection signal.

AFM imaging requires the sample to be adsorbed onto a plane substrate and to stick sufficiently well there for it not to be displaced while the AFM tip is scanning. The most commonly used substrates, especially for high-resolution imaging of isolated molecules or molecular complexes, are mica for hydrophilic samples and highly ordered pyrolytic graphite (HOPG) for hydrophobic samples, since these provide broad and smooth plane regions on the atomic scale. Glass slides are generally used as substrate for imaging intact cells or sub-cellular extracts, sometimes treated by an agent like polylysine to assist in anchoring the sample. A key feature to be taken into account with AFM is the size and shape of the objects to be analysed. The piezoelectric elements controlling the displacement of the tip or sample currently restrict the scanning range in the plane to some 100 μm , with a maximum vertical clearance of 7–10 μm . In contrast to the situation in electron microscopy, it is therefore impossible to carry out a preliminary visualisation at low magnification in order to select some relevant region of the sample. Finally, the acquisition time of an image varies from 1 to 10 min depending on the scan range and the type of sample. Compared with applications in physics and chemistry, the specificity of AFM in biology lies in the need to use the smallest scanning forces possible to avoid damaging the biological sample, which is generally rather soft (stiffness constant of a few millinewtons).

8.1.3 Imaging Soluble Molecules

One of the first applications of AFM in biology was to the imaging of nucleic acids. There has been growing interest in this technique in the past few years, with more than 160 publications in the PubMed data base over the last 18 months. AFM can be used to study the structure of DNA and RNA (size and secondary structure), and also the condensation or bending of DNA (see the reviews in [4, 5]). AFM has also proven itself in the study of DNA–protein interactions. The binding of the RNA polymerase of *Escherichia coli* with its target DNA [6] and the binding of p53, a tumour-suppressing protein, to a specific target have also been visualised [7]. In the latter case, AFM also revealed the dynamics of this interaction. The dynamics of DNA–protein complexes has been revealed using AFM when visualising the degradation of DNA by the endonuclease DNaseI [8]. The experiments described in these three publications also have the particularity of having been carried out in liquid media, whereas most experiments imaging nucleic acids have been carried out in standard atmospheric conditions. More complex structures such as

nucleosomes and the dynamics of their reconstitution have also been imaged by AFM in vitro (see, for example, [9]).

The quality of AFM imaging depends to a large extent on the way the sample is adsorbed onto a plane substrate. In a liquid medium, this adsorption must concern most of the molecule to prevent parts of it from swaying from side to side in the buffer, a phenomenon that reduces resolution. Indeed, this is one of the problems encountered when imaging nuclear protein complexes. Apart from a few cases discussed below, this constraint certainly explains the rather disappointing images obtained with soluble proteins.

The first soluble proteins for which AFM was able to obtain high-resolution images (about 1 nm) were GroEL and GroES, two chaperonins of the bacterium *Escherichia coli* [10]. However, this image quality was obtained with a cryo-AFM, which is a highly sophisticated experimental setup developed by Zhifeng Shao and coworkers at the University of Virginia (USA). This same type of equipment has been used to investigate changes in position of myosin heads, thiophosphorylated or otherwise [11], and the structure of actin filaments [12]. Other elements of the cytoskeleton such as microtubules have been observed in a liquid medium with a resolution of a few nanometers, using a more standard AFM microscope [13]. More generally, it transpires that proteins arranged into biopolymers, such as the nucleic acids mentioned above, are better suited to this type of imaging, undoubtedly because of a certain rigidity, as compared with globular protein. The structure and assembly of collagen fibrils by lateral fusion have been characterised with nanometric resolution [14, 15]. More recently, M. Miles and coworkers in Bristol UK have imaged the various conformations of a glycopolymer (mucin) and demonstrated the heterogeneity of the population [16]. The ability of AFM to characterise different species within a complex mixture represents a significant advantage over other techniques used in structural biology, such as NMR or electron crystallography.

It may be that the structure of a soluble molecule adsorbed onto a substrate for AFM imaging will differ from its structure in solution. However, this criticism is unjustified when this adsorption mimics a natural mechanism, as in the case of certain coagulation proteins or proteins in the extracellular matrix. As an example, AFM has been used to characterise the structure of laminin and heparan sulfate chains, as well as the structural dynamics of this molecule in an aqueous environment [17].

AFM seems particularly well suited to studying peptides and small proteins forming fibrils or amyloid deposits associated with Alzheimer's disease, prions, and certain types of diabetes. The formation and structure of amyloid fibrils can be observed in real time (see the review in [18]). Most published studies have been carried out in air, but some recent work has succeeded in studying the bundling conformation of filamentary structures in a physiological buffer [19]. These experiments have also revealed a marked difference between the conformations of structures observed in air and liquids.

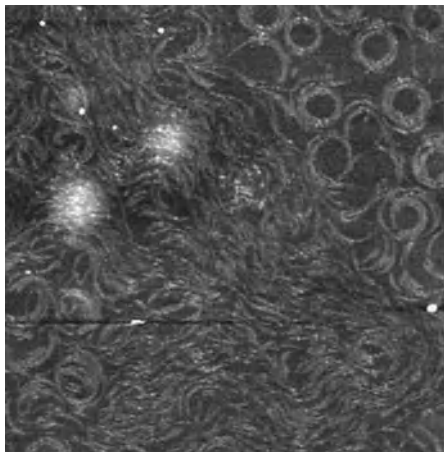


Fig. 8.4. AFM imaging of vector peptides in a monolayer. A monolayer of the vector peptide hCT(9-32), a derivative of calcitonin, is deposited, using a Langmuir trough, and compressed at 20 mN/m, before being transferred to mica. The sample is examined in air in contact mode [20]. The image was obtained with a scan of 5 μm by 5 μm and the height scale is 10 nm. The peptides arrange themselves into filaments, coiled into a spiral with a height of 1 nm. Reproduced with the kind permission of the Biophysical Society

AFM can also provide very important structural data in the case of vector peptides spread in a monolayer with a Langmuir balance. For hCT(9-32), a fragment of human calcitonin with the ability to translocate across the plasma membrane of eukaryotic cells, AFM can be associated with other biophysical techniques such as Fourier transform infrared spectroscopy to characterise the structural arrangement of this peptide [20]. In solution, hCT(9-32) is not structured, whereas it is arranged into helices when located at the water–air interface. Under such experimental conditions, AFM has shown that this peptide forms filamentary structures arranged in spirals (see Fig. 8.4).

8.1.4 Membrane Imaging

Model Membranes

Membrane systems, lipid monolayers and bilayers, whether or not a peptide or protein is inserted or interacting with the membrane, constitute useful models for investigating structure–function relations within biological membranes. The various types of model membranes, their fabrication, analysis, and main uses were discussed in Chap. 2.

Model membranes have been widely used to study the lateral arrangement of lipids (see the review in [21]). In such systems, AFM can be used to visualise lipid phase separations and supply a lot of data that is not accessible by

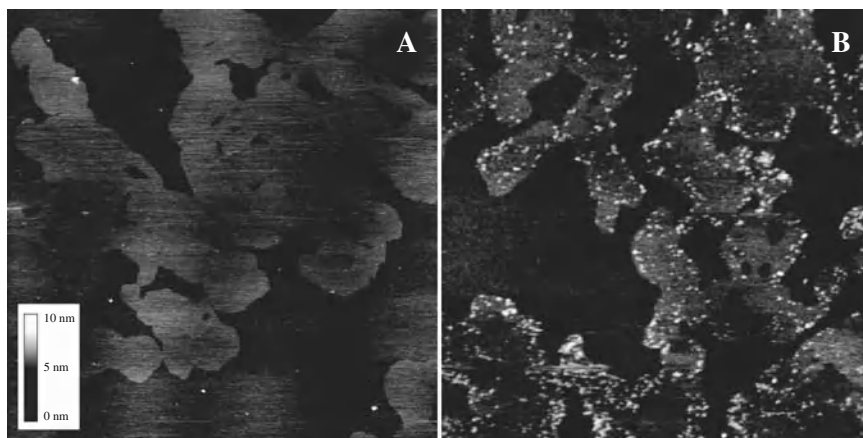


Fig. 8.5. Spontaneous insertion of alkaline phosphatase in a model membrane. The model membrane is a bilayer composed of a binary mixture of dioleoylphosphatidylcholine (DOPC) and dipalmitoylphosphatidylcholine (DPPC) (1:1), supported on mica. The bilayer was obtained by fusing small unilamellar vesicles. **(A)** Control image, in which gel phase domains (DPPC, *lighter shade*) alternate with fluid phase domains (DOPC, *darker shade*). The gel phase is located about 1 nm above the fluid phase. **(B)** When the sample is incubated with alkaline phosphatase, a protein with a glycosylphosphatidylinositol lipid anchor, it is observed to insert (*white spots* measuring 15–50 nm) in the lipid domains in the gel phase [67]. Copyright Wiley-VCH Verlag. Images obtained by scanning $6\mu\text{m} \times 6\mu\text{m}$. Height scale 10 nm

other techniques, regarding the existence, shape, and size of membrane microdomains, especially for raft microdomains, i.e., membrane microdomains enriched in sphingolipids and cholesterol, which resist solubilisation by Triton X-100.

These artificial membrane systems can also be used to visualise the topography of transmembrane proteins, once reconstituted in a lipid bilayer. Under favourable conditions, these proteins can form crystals confined in two dimensions on which the lateral resolution can reach 5 Å (see the review in [22]). Using supported bilayers, it is also possible to visualise proteins in a non-crystalline form, anchored onto or associated with these membranes. Topographic details have been observed with a lateral resolution of 1 nm with the B chain of the cholera toxin, associated with the membrane via the ganglioside GM1 [23], and with the protein Cry1Aa, a pore-forming insect toxin [24]. In studies of raft-type membrane microdomains, we have been able to characterise the preferential insertion of alkaline phosphatase, a protein with glycosylphosphatidylinositol anchoring, in ordered lipid phases (see Fig. 8.5) [25,26]. Even though the resolution obtained was not sufficient to yield structural data for this protein, probably due to its bulky extracellular domain, AFM was able to visualise its distribution in a lipid mixture mimicking the phase separation

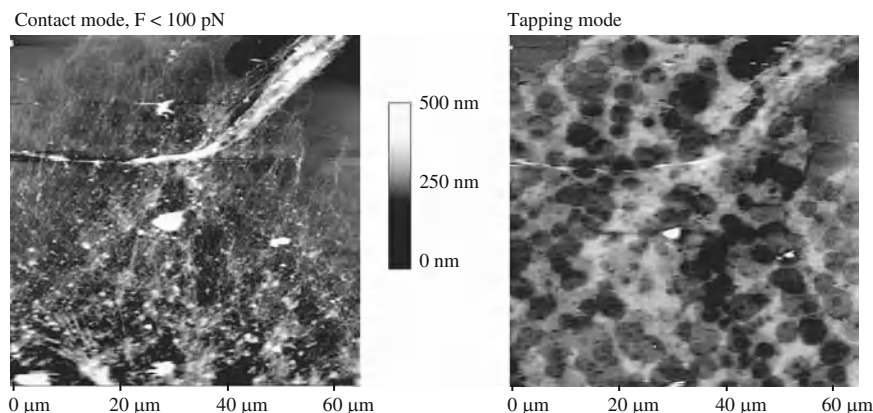


Fig. 8.6. AFM imaging of membrane fragments resisting extraction by detergent (DRM). CV-1 cells (monkey fibroblasts) were treated at 4°C for 30 min with 1% Triton X-100 solubilised in Hank's buffer. They were then immobilised with 2% paraformaldehyde and imaged by AFM in both contact and tapping modes. In contact mode (*left*), even at forces below 100 pN, the tip flattens the sample, revealing elements of the cytoskeleton (filamentary structures). However, in tapping mode (*right*), these features are much less visible and the membrane fragments resisting the detergent extraction (*brighter shade*) alternate with zones extracted by the Triton X-100

of plasma membranes in eukaryotic cells. The use of model membranes allows one to exercise tight control over the membrane cholesterol level, this being one of the constituents of raft microdomains.

Subcellular Imaging and Native Membranes

The membrane surrounding eukaryotic cells contains zones rich in specific lipids, cholesterol and sphingolipids, which play an important role in many cell functions. Under the effect of detergents, these membrane regions group together to form large domains which resist the treatment. Figure 8.6 shows how AFM can be used to visualise these extremely fragile structures in situ [27]. It also illustrates the superiority of the tapping mode, with very flexible cantilevers, over the contact mode when it comes to 3D imaging of samples.

More generally, the topography of membrane fragments, purified by subcellular fractioning methods and adsorbed onto mica, generally reveals a surface from which many structures emerge, usually globular, and these very likely correspond to membrane proteins [28]. Other structures frequently observed, in the form of rings of diameter 10–20 nm surrounding a central dip, suggest something like a transmembrane pore. As for intact cells, the way this new data concerning the molecular-scale organisation of membranes is exploited will depend on the development of methods for local identification (see Sect. 8.1.6). In certain fortunate cases, the purified native membrane

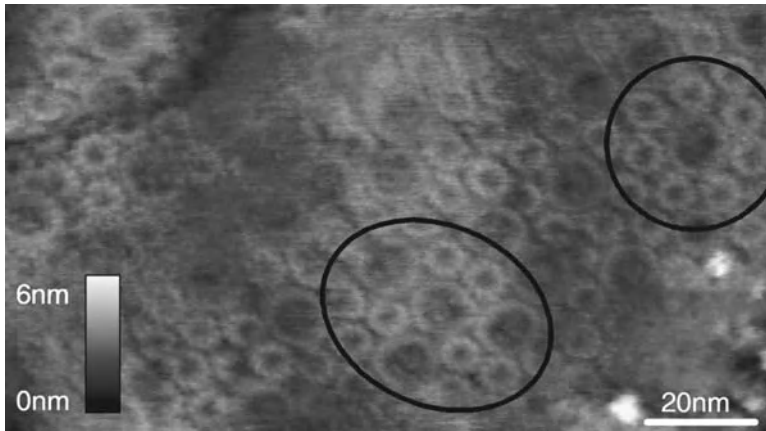


Fig. 8.7. High-resolution imaging of photosynthetic complexes in *Rhodospirillum rubrum*. Image kindly proposed by Simon Scheuring (Institut Curie, France)

contains only a limited number of different protein types. This is the case for the membrane fractions of the photosynthetic apparatus in certain bacteria. AFM is then a unique tool for gathering information about the organisation of supramolecular complexes in physiological conditions, and this at a resolution close to 10 Å (see Fig. 8.7) [29]. Some membrane proteins, such as those of the communicating junctions in eukaryotes or bacteriorhodopsin in *Halobacterium halobium*, self-organise spontaneously into a 2D lattice in order to carry out their function. The communicating junctions are purified in the form of plates comprising two superposed membranes deposited on mica. The cytoplasmic membrane faces, examined in the buffer, appear to have little structuring. The tip, used as a nanodissection tool, can remove the upper membrane and expose the extracellular face of the second membrane. The hexagonal arrangement of the individual channels or connexons forming the junction, which protrude 1.4 nm above the surface, can then be observed with a lateral resolution of 2.5 nm [30]. The trimeric organisation of bacteriorhodopsin, a proton pump with 7 transmembrane helices, is revealed with a lateral resolution as good as 5 Å and a vertical resolution in the range 1–2 Å [22]. The dimeric organisation of rhodopsin in rod membranes has also been demonstrated recently [31].

Apart from this remarkable work on single molecules, one of the great strengths of AFM is the possibility of imaging, in physiological conditions, organelles and subcellular structures on a mesoscopic scale in order to investigate structure–function relations. Indeed, this size range lies outside the scope of optical characterisation. Consider, for example, the structural evolution of the excised nuclear envelope of xenopus oocytes, where AFM results have revealed the existence of very small pores on the periphery of the complexes forming the large nuclear pore complexes (NPC). The extent to which these pores open depends on the calcium and ATP concentrations and may provide

a preferred channel for ion transfer, with the NPC being reserved for transport of macromolecules [32].

8.1.5 AFM and Cells:

Cell Imaging, Mechanical Properties, and Adhesion

One of the main aims of AFM in biological applications is to obtain 3D images of cell surfaces in a physiological environment. The first images, naturally produced in contact mode, were obtained for dried lymphocytes and red blood cells, and the bacteria *Escherichia coli* and *Halobacterium halobium* [33]. Very soon, it became possible to image the same blood cells immobilised by glutaraldehyde in a physiological buffer, with detection of surface features of the order of 10 nm [34]. The observation of cells normally suspended in solution already raised the problem of how to immobilise them on a substrate whose surface properties were modified to favour their adsorption. Later, having identified some of the factors limiting the topographical characterisation of cells and subcellular structures by AFM, new applications became possible with near-field microscopes, such as the determination of local mechanical properties [35] and adhesion forces between cells or between a cell and its substrate [36], and the localisation of certain cell functions by recognition imaging [37].

Topography of Intact Cells

The results obtained for eukaryotic cells [38, 39], prokaryotic cells [40], and plant cells [41], all of which have been reviewed recently [42, 43], show that this microscope can obtain topographic data concerning the cell surface, in a buffer, with a resolution that is often better than an electron scanning microscope. On the mesoscopic scale, which lies between the lower limit of optical microscopy ($\lambda/2$, or ~ 250 nm) and the molecular scale (≤ 15 nm), AFM can provide a unique 3D view of the topography of living cells in their natural environment. This is illustrated by a visualisation of a virus budding at the cell surface, and the appearance over time of pits with diameters between 100 and 180 nm, responsible for exocytosis at the surface of living pancreatic acini stimulated by mastoparan [44, 45].

However, topographical analysis of eukaryotic cells in contact mode requires the use of scanning forces at the limit of experimental feasibility: the stiffness constant, i.e., the force required to compress or stretch a spring by unit length, of the most flexible AFM cantilevers, of the order of 10 mN/m, is significantly greater than the 'stiffness constant' k_{pm} of a plasma membrane, usually close to mN/m. Imaging conditions thus involve a vertical deformation of the surface, which can nevertheless be limited to ~ 10 nm for scanning forces of the order of 20 pN [46]. Under these conditions, the presence of globular structures measuring a few nm, associated with the absence of visible elements of the cytoskeleton over whole regions of the surface of CV1 cells,

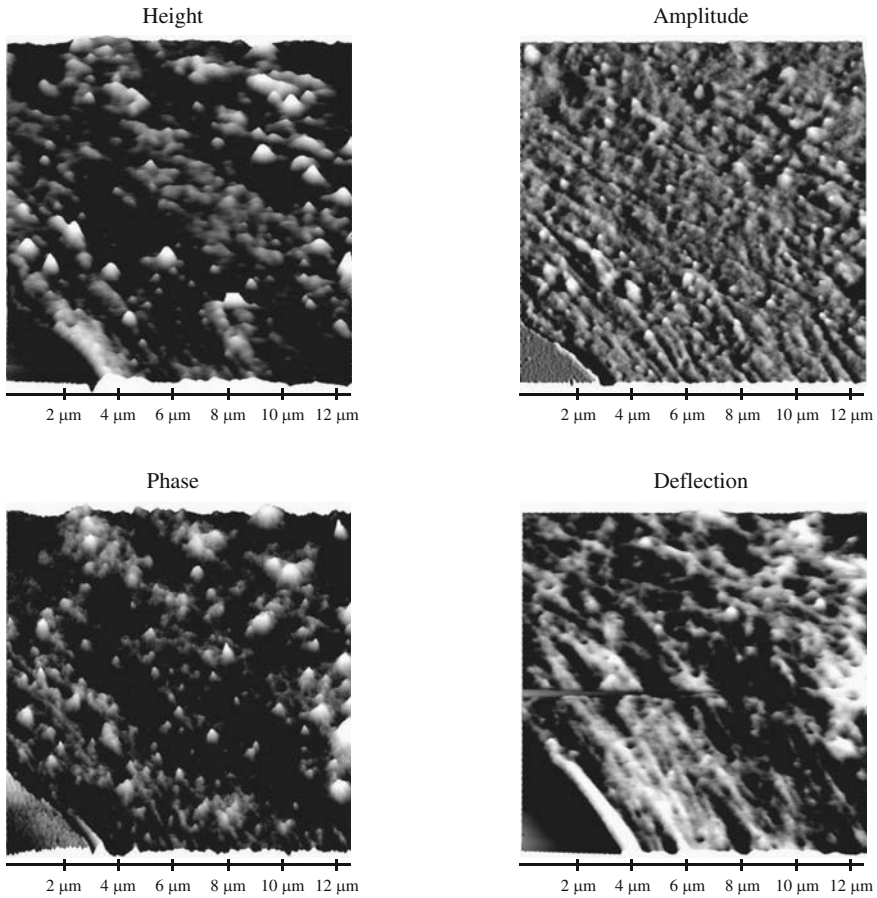


Fig. 8.8. Imaging CV-1 cells in tapping mode. 3D images of CV-1 cells in a physiological buffer. Height and amplitude images reveal topographical properties of the surface. Phase and deflection images (DC to AC signal) are strongly influenced by local mechanical properties

suggests that it may be possible to gain access in contact mode AFM to the very structure of the plasma membrane in living cells. When studying the ultrastructure of bacterial cell surfaces [43], the values of k_{pm} are higher by an order of magnitude than those encountered in eukaryotes [47], which means that contact mode imaging can be carried out without deforming the subject of investigation. This advantage is nevertheless counterbalanced to some extent by the need to obtain a perfect immobilisation of the bacteria on the substrate.

A very effective alternative for visualising the softest cells has been to use tapping mode AFM with cantilevers having a low stiffness constant

(≤ 0.06 N/m) (see Fig. 8.8). Under these conditions, the deformation of structures is minimised compared with contact mode at the limiting force (see the last example, illustrated in Fig. 8.6). The most likely explanation for these observations is that, under the chosen conditions of minimal damping ($< 5\%$ of the oscillation amplitude), the tip probes the surface in the attractive regime and hence images the surface without actually touching it. A better understanding of the behaviour of these cantilevers in tapping mode in a liquid medium should clarify this point.

The presence of a glycocalyx (and possibly mucus), a sort of polymer made from complex (branched) sugars, occurring to differing degrees depending on the type of cell, and the high mobility of membrane constituents (lipids and proteins) represent further significant obstacles to a truly molecular visualisation of the membrane surface of living cells. Immobilisation by paraformaldehyde or glutaraldehyde is still very widely used in biology when studying structure–function relations in cells. It leads to cell death associated with immobilisation of the cell constituents. It also causes a significant increase in the value of k_{pm} , which is multiplied by a factor anywhere between 3 and 10. The 3D imaging of the surface of cells from the inner ear, immobilised and observed in a buffer with a resolution of 3 nm [48], corroborates the hypothesis that these experimental conditions are the most favourable for characterising membrane structures *in situ*.

These same images raise the question of how to identify the many structures present at the surface and revealed by AFM. Recently developed near-field optical microscopy of fluorescent labels and recognition imaging techniques (see Sect. 8.1.6) are likely to be good candidates for answering this question. Another interesting idea is scanning ion conductance microscopy [49, 50], but the need to keep the tip more than 50 nm from the surface of living cells will limit resolution to the upper bound of the mesoscopic scale.

Mechanical Properties of Cells. Adhesion Forces

AFM can be used to determine local mechanical properties [51] by measuring the indentation of the sample by the AFM tip under the effect of applied forces inducing this deformation. This measurement is used to evaluate local elastic properties with a lateral resolution of a hundred or so nanometers. The value of Young's modulus (elastic constant which, for a homogeneous and isotropic material, relates the stress to the strain) obtained in this way for eukaryotic cells varies between ~ 0.2 and 200 kPa depending on the type of cell and the distance from the periphery of the cell [35, 52]. Local mechanical properties of cells also vary with the physiological conditions [53] and under the effect of various drugs [54]. One recent study suggests using the model by Chen and Tu as an alternative to the Hertz model used up until now for determining Young's modulus [55]. There is still some debate as to how to interpret the elastic modulus of a cell as determined by AFM.

The cohesive forces binding cells together play a key role in cell migration, the binding of lymphocytes to the vascular wall, and the formation and maintenance of tissues. These forces involve interactions between proteins present at the surface of each cell making up the tissue or organ. The direct measurement of these interaction forces in individual cells (of the order of a few tens of piconewtons) is now possible using AFM. The cells must first be arranged on the AFM cantilever, either by cultivating them there, or by tethering them onto it by means of proteins. They are then brought into contact with the surface of cells cultivated on a solid substrate under the usual conditions. The influence of environmental factors (ions, pH, temperature, etc.), and also of the modification by directed mutagenesis of proteins involved in intercellular interactions, on the forces required to separate cells previously in contact is a unique source of information for understanding interactions between cells [36, 56].

8.1.6 Current Limits and Future Developments

The more and more widespread use of AFM in structural biology (~ 220 publications in 2000 in the PubMed data base, compared with ~ 460 in 2004) gives a better idea today of the limits of this technique. Here we discuss some proposals for further developments.

Reducing the Force Applied to the Sample

Most biological structures are relatively soft and delicate and can be deformed by applying forces as small as one nanonewton. Under these conditions, one of the main improvements expected in AFM is the development of systems minimising the forces applied to the sample. Several ideas are currently being explored, most of which are related to the cantilever itself. The most logical solution to this problem is to use cantilevers with very low stiffness constants (small cantilevers are currently under development). However, these are likely to be unstable and/or to have very long equilibration times. Another solution consists in balancing the applied force by the radiation pressure of spontaneous cantilever motions at very low forces [57]. Wider use of the tapping mode would also provide a way of reducing the applied force by increasing the quality factor Q of the cantilever oscillating in a liquid medium [58]. The ideal solution would be to image in non-contact mode, but this regime is currently difficult to control with complex samples in a liquid medium. A novel solution has been put forward by H. Hörber and coworkers in the USA, who have developed a new form of near-field microscopy called photonic force microscopy (PFM) [59]. This uses a very small bead trapped by a laser beam instead of the standard AFM tip, and it can apply forces in the piconewton range.

Increasing the Resolution

The resolution can in principle be improved by making cantilevers with highly tapered tips, although this point is debatable since resolutions of 5 Å are already attainable. At the present time, carbon nanotubes provide an interesting possibility, but they are expensive and it is difficult to ensure reproducibility in their fabrication.

Increasing the Scan Rate

To be able to monitor the dynamics of biological processes or obtain images of living cells with significant membrane dynamics (see Sect. 8.1.5), the rate at which the sample is scanned must be increased. The development of small cantilevers with low stiffness constants, suitable for high scan rates, is currently the main channel of investigation among AFM manufacturers, with the aim of reaching the video scan rate of 25 frames per second.

Identifying the Observed Structures

In order to improve the performance of AFM when it is used on complex systems such as native membranes or intact cells, it seems essential to develop recognition imaging. A relatively simple system involves combining AFM with a fluorescence microscope. Several such setups described recently [60–62] have been able simultaneously to obtain a topographical image with nanometric resolution and to identify the sample constituents by fluorescence. This type of setup is nevertheless limited by the resolution that can be obtained with an optical microscope, which is governed by the laws of diffraction (~ 250 nm). One advantage of scanning near-field optical microscopy (SNOM) (see Chap. 7) is that it does not obey the same physical laws. Recent progress with this technique, allowing detection and localisation of fluorescence emitted by a single molecule, should make it possible in the near future to combine topography and molecular identification with a resolution of the order of ten nanometers.

The system developed by P. Hinterdorfer and coworkers (Linz, Austria) couples an antibody covalently to the microscope tip and simultaneously carries out topographic imaging and antigen mapping (see Fig. 8.9) [37]. This elegant strategy remains to be tested on complex systems.

Finally, amylase molecules, secreted by a fusion pore in a pancreatic cell have been identified by AFM using an antibody coupled to colloidal gold beads, the spherical shape of the beads being identifiable by the tip [63].

Combining AFM with Other Biophysical Techniques

It also seems important to combine AFM with other biophysical techniques, such as fluorescence correlation spectroscopy (FCS), able to assess membrane dynamics with excellent time resolution, or other surface techniques such

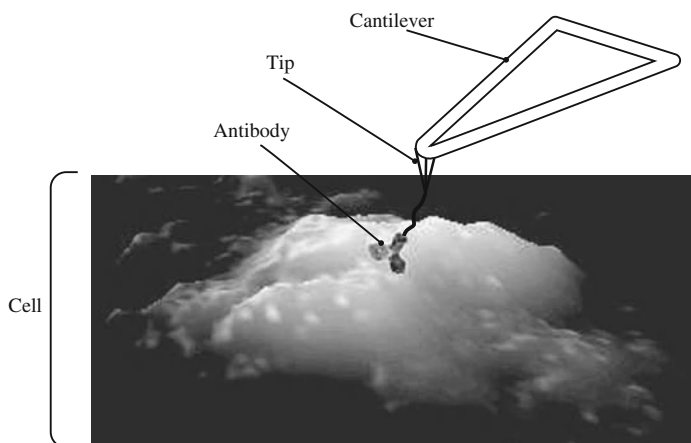


Fig. 8.9. Recognition imaging. A cell, represented by this 3D height image obtained by tapping mode AFM, is scanned by a tip on which an antibody recognising the relevant protein has been covalently coupled. During scanning, antigen–antibody recognition will stress the cantilever and modify its amplitude signal. This system can simultaneously obtain the topology and the antigen distribution of a protein. The z colour scale for the cell is $3\mu\text{m}$

as ellipsometry or the quartz crystal microbalance (see Sects. 9.2 and 9.6 of Chap. 9).

8.1.7 Developments in Nanobiotechnology and Medicine

Development of AFM over the past few years has led to significant progress in a wide range of scientific disciplines, in physics, chemistry, and biology. This characteristic makes it one of the most useful tools in the development of nanobiotechnology.

AFM is an extremely effective tool for nanodissection. The first nanodissection experiments were made on DNA by increasing the applied force [64], and it has been shown that AFM can extract genetic material from chromosomes with very great precision [65]. Engel and coworkers in Basel, Switzerland, used the AFM tip as a mechanical scalpel to remove proteins associated with the plasma membrane (see the review in [3]). The controlled manipulation of molecules selected by force spectroscopy (see Sect. 8.2) coupled with high-resolution imaging forms a powerful approach for obtaining information about biomolecular assemblies.

More recently, an operation was effected by AFM on a living cell using a nanoneedle [66], suggesting the possibility one day of carrying out nanosurgery.

In the long term, AFM may be used to fabricate nanosystems, either purely biological, or in association with other components, e.g., electronic components.

8.2 Surface Force Apparatus and Micromanipulation

8.2.1 Surface Force Apparatus (SFA)

Surface force apparatus or SFA [68] measures the force between two surfaces as a function of their separation by combining precision mechanics and interferometry on a scale of 0.1 nm. It is on this length scale that one can observe the forces governing the properties of colloids [69,70]. (A colloid is a suspension of solid or liquid particles in a liquid.) The DLVO theory, named after Derjaguin, Landau, Verwey and Overbeek [71], established in the 1940s, explains the stability of colloids via two forces, one attractive, the van der Waals force, and the other repulsive, screened electrostatic forces. After that, some progress was made in describing these forces theoretically, but experimental measurements were always limited. Then with the advent of SFA, it became possible to make precise comparisons between measurement and theory. It was also possible to give a detailed description of other contributions to these forces, such as the forces due to the structuring of a liquid in the vicinity of a solid surface, steric interactions due to the presence of a polymer, forces involved in the recognition between a receptor and its ligand, and several others. This technique can also monitor the deformation of surfaces when they adhere, and quantify this adhesion. Some versions of SFA can measure frictional forces between two surfaces separated by a liquid film of nanometric thickness and obtain information about the dynamics of confined liquids [72].

Surface force apparatus includes a mechanical device combining translation stages with several levels of precision, able to vary the separation between two surfaces placed face to face: in general, a micrometric translation stage, a differential spring mechanism made of two springs, one of which is a thousand times more rigid than the other (accuracy 1 nm), and a piezoelectric tube (accuracy 0.1 nm) (see Fig. 8.10 and below) [68]. When mica is cleaved, one can obtain atomically smooth plane surfaces over several square centimeters, and this is why it is often used in SFA experiments.

Surface Force Apparatus (SFA)

Surface force apparatus includes a mechanical translation with several levels of precision, used to adjust the separation between two mica surfaces. The upper cylinder is connected to a translation stage and used to position the lower surface with micrometric accuracy. The lower cylinder, driven by a micrometric translation stage, compresses a helical spring. The latter pushes against a spring consisting of a deformable parallelogram a thousand times more rigid, carrying the lower surface. A displacement of the lower cylinder through one micrometer thus leads to a displacement of the lower surface through one nanometer. The upper surface is fixed onto a piezoelectric tube which expands by 0.7 nm per volt. The lower surface is fixed onto the deformable parallelogram by a leaf spring, which measures the force between the two surfaces. The bending of the spring is equal to the difference between the

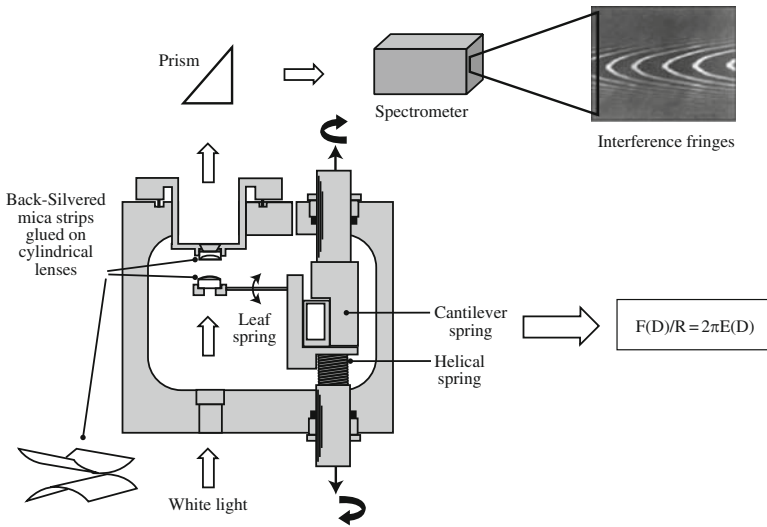


Fig. 8.10. Surface force apparatus. See text for explanation

motion of the base of the spring, imposed by the calibrated translations, and the change in distance between the surfaces measured optically.

The interferometric measurement of the separation between the two surfaces is obtained by illuminating with white light the interferometer formed by the two surfaces (see Fig. 8.11) and analysing the transmitted spectrum. The light enters the apparatus from the bottom and leaves by the top, after having crossed the surfaces. An objective focuses the image of the interferometer on the entry slit of a spectrometer, in such a way that the point where the surfaces are closest is centered on the slit. The interference fringes observed in the spectrometer are used to measure the distance between the surfaces and, independently, the refractive index of the medium that separates them. They also allow one to measure the shape of the surfaces in contact during the interaction, which provides full information about the geometry of the interacting surfaces. The surfaces are curved and the crossed cylinder geometry is equivalent to a sphere/plane geometry when the distance between the surfaces is much less than their radius of curvature. Derjaguin's approximation then relates the force measured between curved surfaces to the interaction potential energy per unit area between plane surfaces [see (8.5)].

Two sheets of mica of strictly equal thickness between 2 and 5 μm and dimensions $8 \times 8 \text{ mm}^2$ are stuck onto cylindrical glass lenses and arranged face to face in the SFA, with a crossed cylinder geometry. When the radius of curvature R of the surfaces is much greater than the shortest distance D between them, this geometry is equivalent to a sphere/plane geometry. One of the lenses is placed on a leaf spring, which serves to measure the force between the surfaces. The mica sheets are silvered on one face and arranged so as

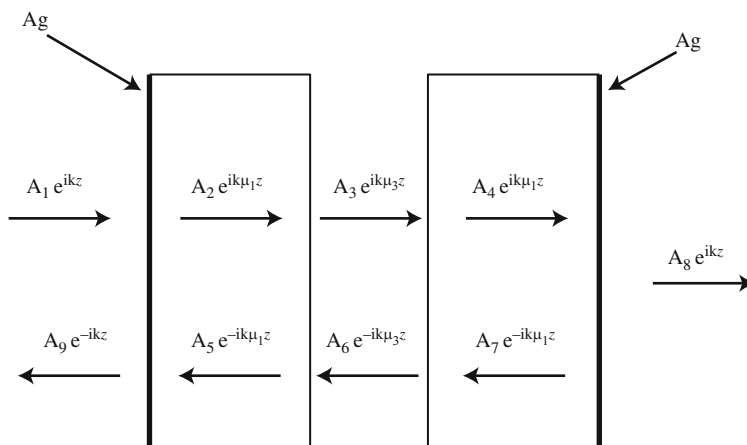


Fig. 8.11. Multibeam interferometry. Two mica sheets of equal thickness, silvered on their outer faces in such a way as to transmit 1–5% in the visible, form a Fabry–Pérot interferometer. A wave incident on this interferometer will reflect on the various dioptric surfaces. When white light is normally incident, the transmitted wavelengths correspond to constructive interference and are called fringes of equal chromatic order (FECO). These wavelengths are calculated from the continuity equations of the electromagnetic field and the reflection and transmission coefficients at the different dioptric surfaces. A plane interferometer gives FECO corresponding to a discrete number of wavelengths, which appear as discrete lines in the spectrometer. If the separation between the surfaces is altered, the FECO wavelengths will also change. When the mica surfaces are curved and the separation therefore changes at each point of the interferometer, the FECO are then curved (see Fig. 8.10). Their shape informs as to the geometry of the surfaces during their interaction

to form a Fabry–Pérot interferometer, with the silvered surfaces on the outside of the interferometer, and the bare mica surfaces placed face to face (see Fig. 8.11). The silver film has a transmission coefficient of 1–5% in the visible. For plane surfaces, if such an interferometer is illuminated with white light, the different wavelengths will go through and reflect at the various dioptric surfaces, and the transmitted spectrum will be composed of those wavelengths for which there is constructive interference. (These correspond roughly to the normal vibrational modes of the light waves between the two silver mirrors.) In the spectrometer, these wavelengths form interference fringes called fringes of equal chromatic order (FECO), which depend on the thickness of the mica sheets, their separation, and the refractive indices of the various media. The position of the FECO can be used to obtain, simultaneously and independently, the separation and the refractive index of the medium separating the two surfaces [73] (see Table 8.1 which gives the measurement accuracies):

$$D = \frac{\lambda_n}{2\pi\mu_3} \arctan \frac{2\bar{\mu} \sin(n\pi F_n \Delta\lambda_n/\lambda_n)}{(1 + \bar{\mu}^2) \cos(n\pi F_n \Delta\lambda_n/\lambda_n) \pm (\bar{\mu}^2 - 1)}, \quad (8.1)$$

Table 8.1. Accuracies for parameters obtained by various techniques. SFA surface force apparatus, BFP biomembrane force probe

Technique	Force [N]	Adhesion energy [J m ²]	Separation [nm]
SFA	10 ⁻⁷	10 ⁻⁶	0.1–0.01
BFP	10 ⁻¹²		5
Cell micromanipulation	10 ⁻⁹		
Vesicle micromanipulation		10 ⁻⁶	

where λ_n is the wavelength of the fringe of order n ,

$$\Delta\lambda_n = \lambda_n - \lambda_n^0, \quad (8.2)$$

n is deduced from the wavelengths λ_n^0 of the fringes when the mica surfaces are in contact,

$$\frac{\lambda_{n-1}^0 - \lambda_n^0}{\lambda_{n-1}^0} = \frac{1}{nF_n}, \quad (8.3)$$

with

$$F_n = 1.024 + \frac{1}{n} \quad (8.4)$$

a factor accounting for light absorption and phase shift on reflection by the silver films. μ_3 is the refractive index of the medium separating the two mica sheets, and $\bar{\mu}$ is equal to μ_3/μ_1 , where μ_1 is the refractive index of the mica. In the denominator of (8.1), the plus sign corresponds to odd order fringes and the minus sign to even order fringes.

The possibility of simultaneously measuring the separation and the refractive index between the two mica surfaces can be understood by analogy with a vibrating string. Consider the normal modes of a vibrating string between two walls. If we add a piece of string of different line density in the middle, the modes with a vibrational node in the center will not be perturbed in the same way as those having an antinode there. The same is true for the fringes. In addition, these fringes can be used to determine the shape of the surfaces. The Derjaguin approximation [74] gives a linear relation between the force $F(D)$ between the two curved surfaces, as a function of their separation D , and the energy $E(D)$ between plane, parallel surfaces when the force decreases fast enough (at least as $1/D^2$):

$$F(D) = 2\pi RE(D). \quad (8.5)$$

This technique was used to obtain the first measurements of van der Waals forces in the non-retarded limit [75]. The van der Waals interaction, named after the person who discovered it by studying the behaviour of real gases in comparison with the behaviour of perfect gases, is an attractive force in most cases, occurring between atoms or molecules, but also between macroscopic

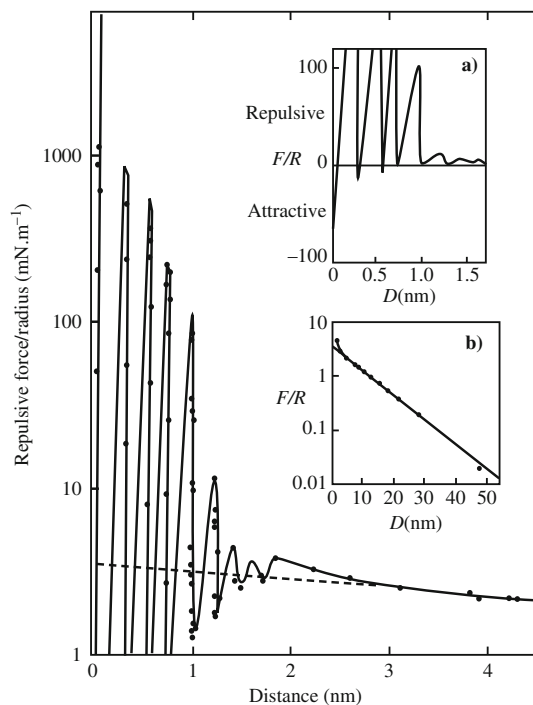


Fig. 8.12. Force normalised by the radius of curvature (equal to $2\pi E$, where E is the interaction energy per unit area for the same distance) between mica surfaces immersed in a KNO_3 solution at 10^{-3} M. (a) Force at separations below 2 nm. (b) Force at large separations, greater than 2 nm. The *continuous line* indicates the electrostatic double-layer force for a surface potential of -78 mV. Taken from [76]

objects. It results from the correlation between the instantaneous or permanent dipoles of the different objects, and varies with their separation. For two media separated by distance d , the short-range interaction energy (typically < 5 nm in water) goes as $1/d^2$. When the distance is large enough, the time taken by the electromagnetic waves to go from one medium to the other can no longer be neglected, and the interaction energy goes as $1/d^3$ [68, 75].

The SFA technique was also used to obtain the first demonstrations that an ordinary liquid like water, when confined, can adopt a layered structure (see Fig. 8.12) [76]. This structuring is revealed by measuring the force between the surfaces. The force oscillates as a function of the distance, and the spatial period of the oscillations is 0.25 nm, i.e., the size of a water molecule.

In the study of cadherins, cell adhesion proteins, the distance resolution of the SFA can reveal the properties of different extracellular domains of the protein (see Fig. 8.13) [77]. Lipid bilayers are deposited on the mica surface. The extracellular parts of the cadherins, made up of five domains of size 5.0 nm, are adsorbed onto the bilayers by their polyhistidine tag which binds to the functionalised lipid heads in an appropriate manner. Forces are measured

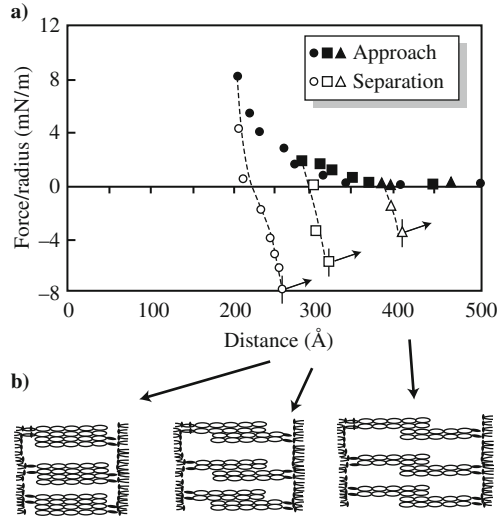


Fig. 8.13. (a) Force normalised by the radius of curvature, showing three energy minima between layers of cadherin with surface density 3.94×10^4 cadherin/ μm^2 . *Black symbols* indicate approach and *white symbols* indicate separation. *Dashed lines* are to guide the eye. (b) Relative positions of proteins at the three attractive minima. Taken from [77]

as a function of the distance and reveal adhesive forces, with minima indicating that the cadherins can bind according to several modes, either by all their extracellular domains, or by a subset of them.

By measuring the adhesion forces with the SFA, it has been possible to establish experimentally the relation between the free binding energy e_b and the free adhesion energy W_a [78] in the case where each surface carries a compact layer of molecules able to form a bond with the molecules of the other surface. In these experiments, lipids whose polar head is a DNA base are deposited in a monolayer on the mica surface, with the bases turned toward the aqueous medium. The surfaces thus coated with a compact layer of adenine (A) or thymidine (T) are then set in contact and their separation force measured (see Fig. 8.14). This separation force is proportional to the adhesion energy according to the Johnson–Kendall–Roberts (JKR) theory [78], which describes the adhesion between deformable surfaces. The force F required to separate a sphere of radius R from a plane is

$$F = \frac{3\pi RW_a}{2}, \quad (8.6)$$

where W_a is the free energy of adhesion of the surfaces. Given the surface density ρ of A or T on the surfaces, and the adhesion energy W_a obtained by measurement, it can be shown that the adhesion energy is the energy of the A/T bonds formed per unit area:

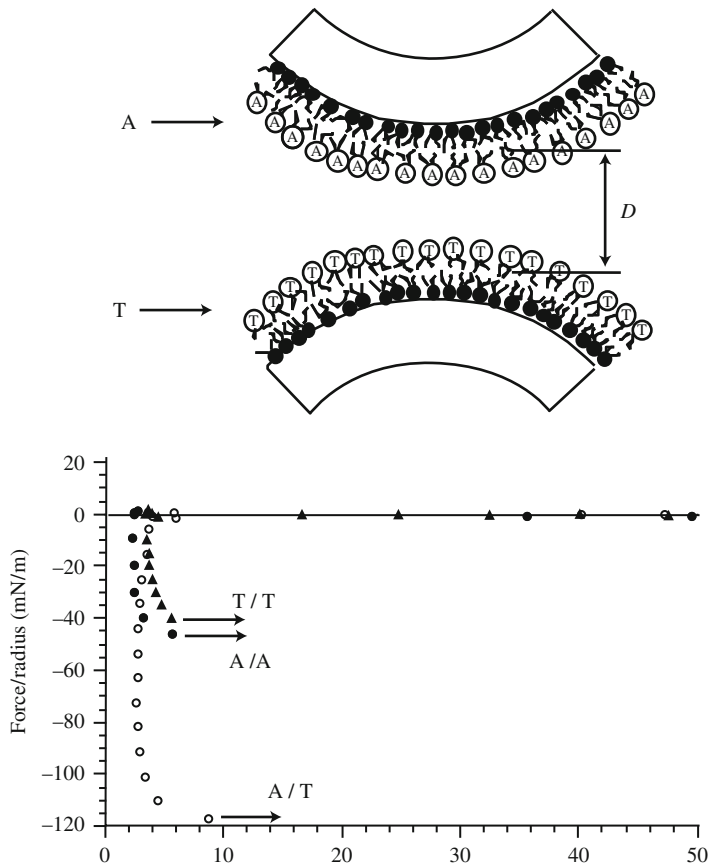


Fig. 8.14. Force normalised by the radius of curvature as a function of the distance between mica surfaces coated with a compact layer of adenosine (A) and thymidine (T). A positive force corresponds to repulsion and a negative force to attraction. *Triangles* and *black circles* represent the experiments with A/A and with T/T, respectively. *White circles* represent the experiment with A/T. *Arrows* indicate the separation force of the surfaces (pull-off force)

$$W_a = \rho e_b \frac{\exp(e_b/k_B T)}{1 + \exp(e_b/k_B T)}. \tag{8.7}$$

Since the AFM became standard equipment in research centers, several authors have had the idea of sticking a microbead onto the AFM cantilever, as shown in Fig. 8.15, and using it to measure the forces between surfaces [80]. This has the advantage of being easy to implement, but the drawback is that it gives neither an absolute measurement of the distance between the surfaces, nor a measurement of their deformation during the interaction.

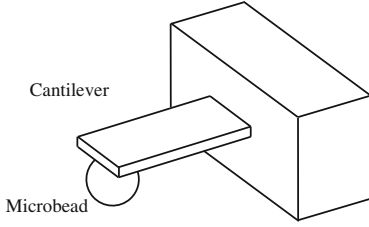


Fig. 8.15. A microbead is glued onto the cantilever of an atomic force microscope, which is used to measure the interaction force between the bead and a plane, thereby giving the same type of information as the SFA

8.2.2 Micromanipulation

Today the micromanipulation of small soft objects using micropipettes can be used to study a wide range of phenomena relating to biological processes, such as cell adhesion and membrane fusion, but also interactions between individual molecules. It does not have the same spatial resolution as SFA, but it can be used to investigate phenomena in which the forces and energies are very small. It can be used to manipulate lipid vesicles with sizes of a few micrometers and to control the tension in their membrane by aspiration. It is thus possible to adhere or fuse functionalised vesicles. Micropipettes can also manipulate cells, placing them in contact with other entities and measuring their adhesion. The red blood cell held by aspiration in a micropipette constitutes an ideal spring for measuring forces between individual molecules, insofar as its stiffness can be changed by several orders of magnitude by simply changing the aspiration pressure. By using some appropriate means to fix receptors on the red blood cell and ligands on another surface, one can make dynamic force spectroscopic measurements of the rupture force of a single bond.

The Biomembrane Force Probe (BFP) and the Bond-Breaking Force Between Two Molecules

A red blood cell has a membrane joined to a 2D cytoskeleton of spectrin which gives it remarkable elastic properties. These can be used to study single bonds between two molecules [81, 82]. When a red blood cell is held by aspiration ΔP in a micropipette of inner radius R_p , its membrane acquires a tension τ , as predicted by Laplace's law which relates a pressure difference to the surface tension and curvature of a membrane:

$$\Delta P = \tau \left(\frac{1}{R_1} + \frac{1}{R_2} \right), \quad (8.8)$$

where R_1 and R_2 are the two radii of curvature of the surface. If a microbead is glued onto the free side of the red blood cell as shown in Fig. 8.16, the pipette–red blood cell–bead system behaves like a spring of stiffness k given by

$$k \approx \frac{\Delta P}{C}, \quad (8.9)$$

where C is a parameter depending on τ and the geometry of the red blood cell–pipette system according to

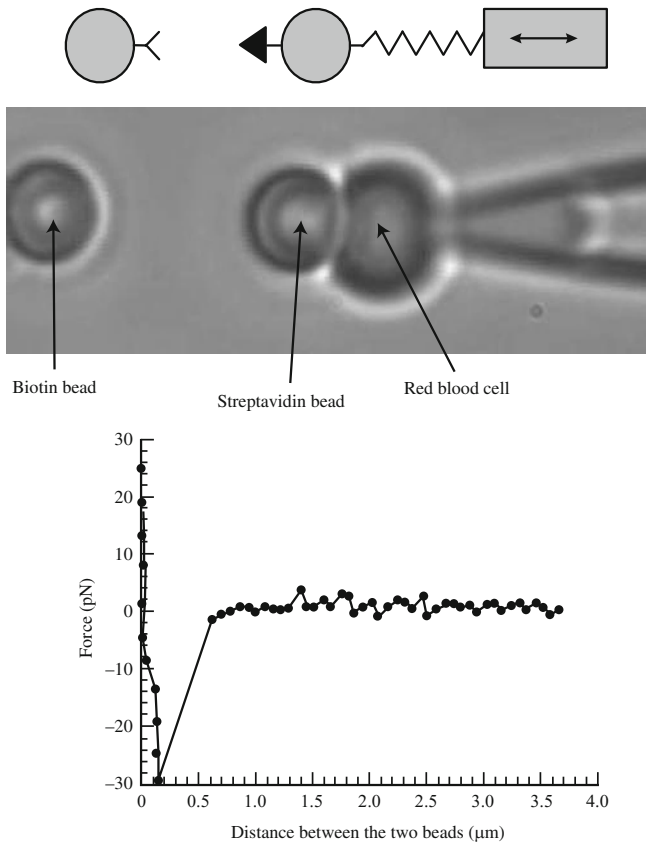


Fig. 8.16. The biomembrane force probe (BFP) can be used to measure piconewton forces by means of a spring made from a red blood cell held by aspiration in a micropipette. It is the astonishing elastic properties of the red blood cell that make this feasible. The surface tension of the membrane can be adjusted by aspiration pressure in the micropipette according to Laplace's law [see (8.8)]. The greater the surface tension of the red blood cell, the more rigid it is. In this way, its rigidity can be varied by as much as three orders of magnitude. To use it as a dynamometer, one marks the red blood cell in order to be able to measure its extension. This is done by sticking a bead onto the surface to one side of the red blood cell. Any change in the distance between the bead and the micropipette corresponds to an extension or compression of the spring, and hence to a force. To measure the force required to break a single bond between a receptor and a ligand, the receptor is anchored onto the bead and the ligands are fixed onto another surface. The micropipette is brought down to the surface (here a bead) and the bead is set in contact with the other surface. The micropipette is then separated from the other surface, while measuring the separation force. In order to be sure that the measurement refers to a single bond, the surface density of the receptors and ligands is adjusted to a sufficiently low value, so that the probability of making a bond during each contact is only 0.1

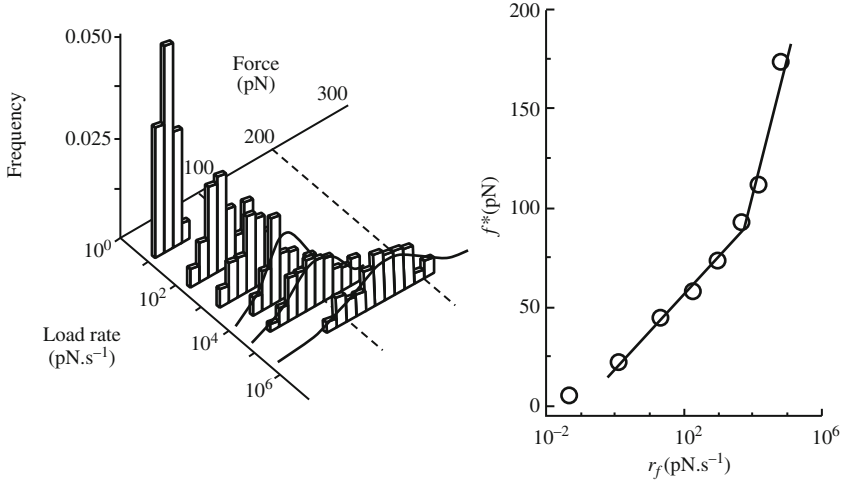


Fig. 8.17. Most probable breaking force for a single bond between streptavidin and biotin as a function of the load rate r . The rupture force is measured many times at the same load rate with the biomembrane force probe. The most probable force corresponds to the maximum of the resulting distribution. It is shown in the *right-hand graph* as a function of the load rate. Taken from [82]

$$C = \frac{\left[\ln(2R_{\text{rbc}}/R_p) + \ln(2R_{\text{rbc}}/R_b) \right] (1 - R_p/R_{\text{rbc}})}{R_p}, \quad (8.10)$$

with r_p , r_{rbc} , and r_b the radii of curvature of the pipette, red blood cell, and contact region between the red blood cell and the bead, respectively. This spring is called the biomembrane force probe (BFP). It can be used to make dynamometric measurements of the force required to break a single bond between a receptor anchored onto the bead and a ligand immobilised on another surface [81, 82]. One of the strong points of this technique is to be able to vary the stiffness constant of the spring through several orders of magnitude. This means that one can measure forces between 1 and 10³ pN. The extension of the spring made in this way is measured by monitoring the bead–pipette separation. The position of the bead can be measured by image processing at 1–3 nm (see Sect. 8.5) and the micropipette can be placed on a piezotranslator with similar accuracy (see Table 8.1 for measurement accuracies).

Nature uses weak bonds to carry out important processes in living matter. This device can be used to probe the force involved in such a bond, e.g., between a single receptor and ligand. At this energy scale, thermal fluctuations play an important role and lead to dispersion in the results for measurements of the force required to break a given bond, quite independently of the error in the measurement. A large number of measurements thus produces the distribution of this force, from which one can obtain the most probable breaking

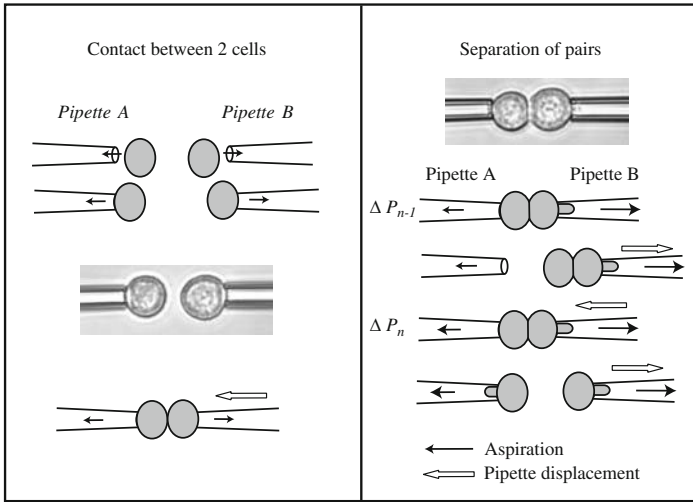


Fig. 8.18. Measuring the adhesion force between living cells. The right-hand cell is held in a micropipette by strong aspiration, while the one on the left is held by weak aspiration. The pipettes are brought together until the cells come into contact, then separated. The pair of cells remains in the right-hand pipette. The left-hand pipette is then repositioned and the aspiration pressure increased by a small increment. The pipettes are once again separated, the pressure increased, and so on, until the aspiration is just enough to separate the cells. This pressure multiplied by the cross-sectional area of the pipette gives the force required to separate the cells

force f^* . By pulling slowly on the bond, the measured breaking force will be small, even zero, because a thermal fluctuation will eventually break the bond. The situation is quite different when the measurement is made while pulling quickly on the bond. In a laboratory experiment, one usually applies a force f that varies linearly with time t , with a load rate r given by $f = rt$. The time scale over which the force is applied determines the most probable value f^* of the breaking force. Theory predicts that f^* should vary linearly with $\ln r$ when the form of the receptor/ligand interaction potential contains a single potential well. In the case where there are several wells and several energy barriers for separation, the graph of f^* as a function of $\ln r$ is composed of a series of straight line segments (see Fig. 8.17), but the theoretical analysis is more involved and remains controversial. The shape of the breaking force distributions depends on the shape of the interaction potential. This technique can thus be used to obtain information about the shape of the potential.

Adhesion Force Between Living Cells

Micropipettes can also be used to measure the adhesion force between living cells (see Table 8.1 for measurement accuracies), and hence to determine the dynamics of adhesion between cells, to find out whether this adhesion

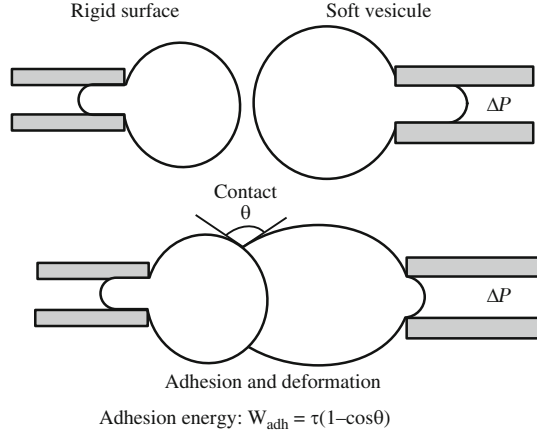


Fig. 8.19. Measuring the free energy of adhesion of two vesicles by micromanipulation. The vesicle on the left is made very rigid by a high aspiration pressure. The one on the right is held in the micropipette by a moderate aspiration, producing a membrane tension τ . When the two vesicles are set in contact by displacing the micropipettes, the contact angle θ between the two membranes of the vesicles reaches an equilibrium value that can be used to evaluate the adhesion energy $W_a = \tau(1 - \cos \theta)$

is triggered by signals of some kind, and to identify the origins of certain pathologies. This works as follows (see Fig.8.18) [83,84].

Two cells are held by aspiration, one weakly (pressure ΔP_1) and the other strongly, by means of micropipettes. They are then brought into contact. When the two micropipettes are separated, the two adhering cells remain in one of the pipettes. The lower aspiration pressure is increased (to pressure ΔP_2), the cells are brought into contact, and the pipettes are separated once again. If the cells remain in contact, the cycle is repeated until the cells separate, each remaining in its pipette, for some aspiration pressure ΔP_n . The force separating the pipettes is then

$$\frac{1}{2}\pi R_p^2(\Delta P_{n-1} + \Delta P_n). \quad (8.11)$$

Micromanipulation and Adhesion Energy of Vesicles

Lipid vesicles are bilayers of lipids with nanometric thickness (~ 5 nm), which close in on themselves. Their adhesion energy can be measured if one knows the membrane tension, controlled by the aspiration pressure and given by (8.8), and the angle of contact θ they make (see Fig.8.19 and Table 8.1 which gives the measurement accuracies) [85–87]. The work required to displace unit area of the membrane under tension τ is $\tau(1 - \cos \theta)$. The change in interface

energy is W_a . At equilibrium, the free energy is minimal, i.e., $\delta F = 0$, which implies

$$W_a = \tau(1 - \cos \theta). \quad (8.12)$$

This approach can be used to measure the parameters describing adhesion between two model membranes.

8.3 Atomic Force Microscopy in Contact and Tapping Modes

8.3.1 Introduction

The invention of the scanning tunneling microscope (STM) in 1981 [88] was a major event in physics because this instrument can examine matter on the atomic scale. For the first time, it became possible to view atoms individually in real space on a plane surface. G. Binnig and H. Rohrer, who invented STM, obtained the Nobel Prize for Physics with the inventor, E. Ruska, of the scanning electron microscope (SEM). A great many metals and semiconductors were then analysed. Some results even got beyond the confines of the scientific community to reach the ears of the general public, e.g., the manipulation of individual atoms by R. Eigler [89].

The STM is only able to produce images of conducting surfaces, and this limits its application to metals and semiconductors. This restriction was finally removed in 1985, when Binnig, Quate and Gerber invented the atomic microscope (AFM) [90]. As for the STM, a finely tapered tip is brought toward the surface to be analysed, but instead of applying a potential difference and measuring a tunnel current, one measures the force between the tip and the surface. AFM can thus be applied to study any kind of surface. Atomic force microscopes have a remarkably simple design: the central element is a flexible cantilever embedded in a support at one end, while the free end is equipped with a tip (see Fig. 8.20). This simplicity, combined with the ability to produce images over a wide range of length scales and for a wide range of systems, explains the immediate success and many applications of AFM.

This embedded cantilever can be put to use in two different ways. In static mode, in which case one measures the distance moved through by the free end of the cantilever, giving the degree of bending. For small amounts of bending, the force is simply obtained by multiplying the bending stiffness by the distance moved by the free end. The alternative is to make use of dynamical modes. The dynamical or tapping mode of atomic force microscopy consists in causing the tip to oscillate in the vicinity of the sample object. One then has an oscillator that is completely specified by its resonance frequency and damping coefficient. This works as follows. The properties of this oscillator, e.g., the correspondence between excitation amplitude and oscillation amplitude, or the relation between amplitude and frequency, will change when

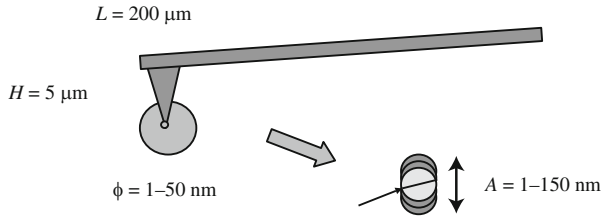


Fig. 8.20. Central part of a local force probe: cantilever and tip. It is easy to show that this system behaves just like a spring onto which a sphere has been fixed. H is the height of the tip, L the length of the cantilever, and ϕ the diameter of the end of the tip

the oscillator enters into interaction with another object. It is through these changes in the behaviour of the oscillator when it interacts that one can gain information about the other objects, e.g., its shape, mechanical properties, and even chemical properties as the case may be.

This oscillator comprises a cantilever with a rectangular cross-section of a few micrometers and a length of several hundred micrometers. These are relatively large dimensions, implying that the cantilever is made up of some 10^{15} to 10^{16} atoms, and it is with this object that one measures the behaviour of a single atom, or an assembly of just a few atoms. The main reason why this is possible is that one can treat the whole ensemble as a system with a single degree of freedom, which means that it can be described as a simple harmonic oscillator. This description remains valid over a rather wide range of oscillation amplitudes, from one to a few hundred nanometers.

This view of things means that we can forget the almost macroscopic structure of our oscillator, reducing it simply to the very end of the tip. For a comparison on the human scale, a similar approach would lead us to take Mont Blanc in the French Alps, turn it upside-down, then have it oscillate up and down above our heads. In this thought experiment, the evolution of the oscillations of the mountain turned upside-down in this way would only depend on the interaction between the last few meters at its summit and the population under investigation.

The effective size of the end of the tip is determined by the region of its surface that actually interacts with the sample. By treating the tip end as a sphere, one ends up with tip diameters varying between the nanometer and about ten nanometers. In some specific situations, e.g., the study of silicon in ultrahigh vacuum, it can be shown that the image obtained results from the variations of the interaction between a few atoms in the tip and a few atoms in the surface, where the latter can be counted individually.

We end this introduction with a graphical representation of the different methods available to image biological systems (see Fig. 8.21). The general idea is to have recourse to a whole panoply of tools for analysing on every length and time scale, from the angstrom to the meter, and over as broad a time scale as possible.

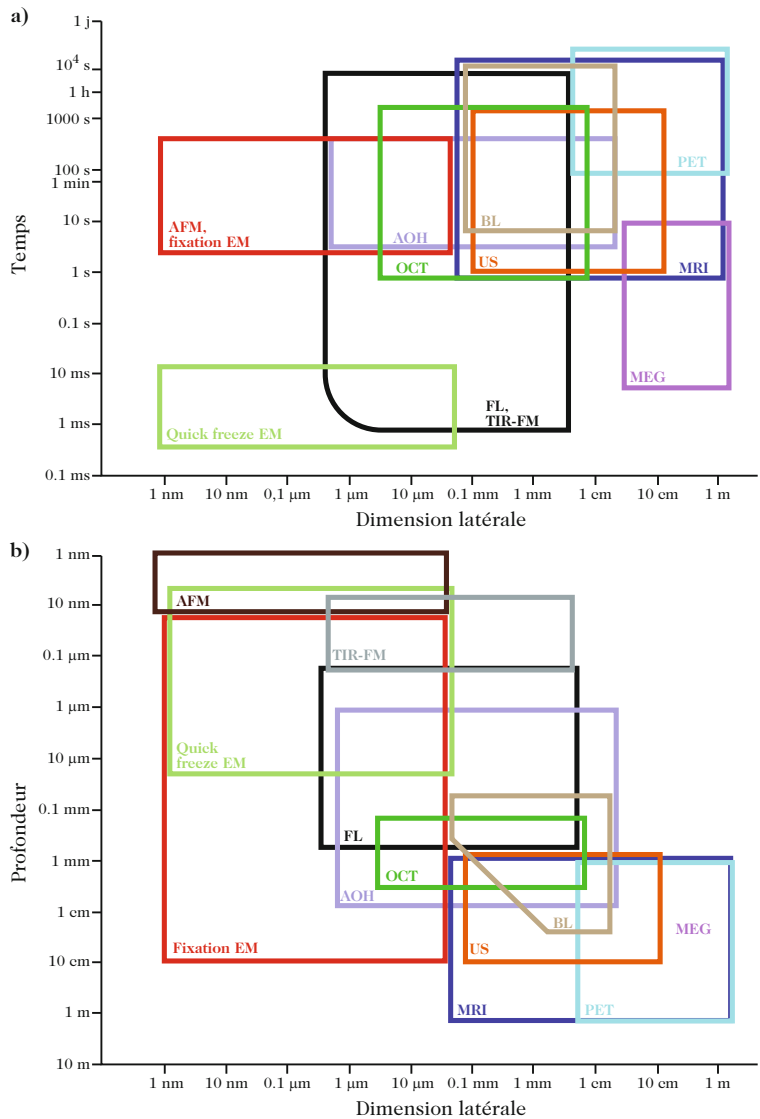


Fig. 8.21. Different methods for examining living matter, taken from [91]. FL fluorescence microscopy, EM electron microscopy, AFM atomic force microscopy, PET positron emission tomography, OCT optical coherence tomography, BL bioluminescence, MRI magnetic resonance imaging, AOH all-optical histology, TIR-FM total internal reflection fluorescence microscopy, MEG magnetoencephalography, US ultrasound. *Squares* give a rough indication of time and length scales, and require some comments in the case of AFM. For the time range, it is difficult to reach 500–1,000 s due to drift in the piezoceramics, and even in servo-controlling the position of the surface. However, for short time scales, time resolutions of millisecond order are feasible

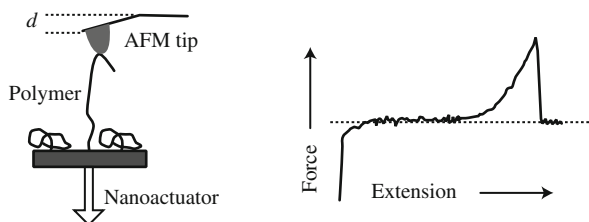


Fig. 8.22. *Left:* AFM tip that has picked up a macromolecule which is stretched under vertical retraction of the surface. *Right:* Variation of the force corresponding to the extension imposed on the polymer chain. The force is obtained by multiplying the cantilever stiffness by the distance moved through by the end of the cantilever

The idea of visualising objects on the scale of molecular assembly, or even single molecules, and being able to observe these systems in action, is not a new one, and lies at the heart of the work carried out to develop AFM. Some examples are presented below. The first part is devoted to force measurements made in static mode, and the second to imaging and the AFM tapping mode.

8.3.2 Force Measurements in Contact (Static) Mode

Measuring the Separation Force for Avidin–Biotin Systems

The underlying idea of a force measurement is rather simple. A fixed point in the plane of the surface is displaced vertically to bring it near the tip. When the tip comes into contact with the surface and the displacement continues still further, the cantilever is deflected to an extent that depends on the compression of the equivalent spring, determined by the mechanical response of the embedded cantilever. In general, since the cantilever has finite stiffness, adhesion between the tip and surface means that there is hysteresis between the force curve measured during approach and the one measured during retraction. When the motion is reversed by withdrawing the surface, provided there is no object between the tip and surface, one measures the force required to unstick the tip.

In the case where there is a macromolecule or ligand–receptor system situated between the tip and surface, one measures either the elastic behaviour of the macromolecule, or the binding force between the ligand and the receptor.

Figure 8.22 shows a typical experiment to measure the elasticity of a polymer chain using the AFM tip. One end of the cantilever is embedded in a support, while the other, carrying the tip, is left free. It can be shown that, in the case where the deflection d is small compared with the length L of the cantilever, the force applied at its free end varies linearly with the deflection, i.e., $F = k_L d$, where k_L is the stiffness constant of the cantilever. This constant depends on the elastic modulus E of the cantilever, its size, and its shape. For a cantilever with rectangular cross-section, with width W , thickness e , and length L , the stiffness constant of a cantilever embedded at one end is

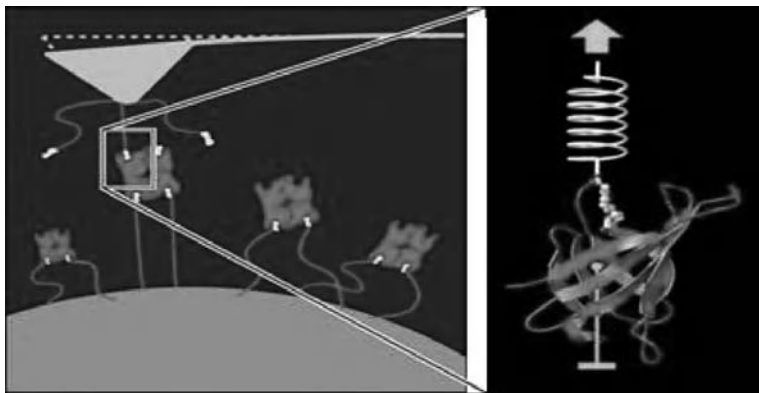


Fig. 8.23. Schematic view of separation force measurements for the avidin–biotin pair using an atomic force microscope. Avidin is a tetramer containing a cavity or receptacle in which biotin inserts itself. The AFM is represented by a spring *on the right* [119]

$$k_L = E \frac{We^3}{4L^3}.$$

For a cantilever with a circular cross-section of radius r , the stiffness constant is

$$k_L = E \frac{3\pi r^4}{4L^3}.$$

There have been many measurements of the elasticity of DNA and other biological macromolecules, using optical tweezers, magnetic tweezers, and micropipettes. These are described in detail in Sects. 8.4 and 8.5. In the present section, we have selected several studies centered on measuring specific interactions in biological systems. These measurements primarily involve force measurements, which is what distinguishes AFM from other techniques, but also imaging, with contrast on the nanoscale, in parallel with these force measurements. Hence, AFM can in principle be used to run experiments in which, for example, proteins are unfolded with visualisation of the resulting modifications.

The strong interaction between the avidin receptor and the biotin ligand is well known to biologists and often used to build molecular assemblies or to functionalise peptide sequences. The study of the binding force between avidin and biotin thus inspired the first experiments carried out on the separation forces between chemical species using a cantilever (see Fig. 8.23). The main exploratory work in this field was carried out by H.E. Gaub and coworkers in Munich [114, 117, 118].

Although non-specific interactions, typically van der Waals interactions, can determine the adhesion between molecular assemblies, the selectivity of the interaction between ligand and receptor induces the specificity required

Table 8.2. AFM separation force measurements for different ligand–receptor pairs [117]

Ligand–receptor	F_s [pN]
Avidin–biotin	160 ± 20
Avidin–iminobiotin	85 ± 10
Streptavidin–biotin	257 ± 25
Avidin–desthiobiotin	94 ± 10
Streptavidin–iminobiotin	135 ± 15

to control the organisation and activity of multicomponent systems. Since biological systems are often controlled by weak forces, it will be difficult to establish a classification between specific and non-specific interactions. To tackle this problem, the Munich group studied variants of the base pairing in the streptavidin–biotin pair (see Table 8.2).

From these measurements, the authors deduced an effective range for the binding potential between ligand and receptor. This effective range is found to be the same for all the pairs, with a value around $r_{\text{eff}} = 9.5 \text{ \AA}$ [117]. This common value suggests that these systems have very similar structures for the receptacle containing the biotin. However, as we shall see below when discussing the unfolding of these polymers, it turns out that there is a critical load rate below which dissociation is frequent enough for the most probable force to be close to zero (see Fig. 8.17).

Over long periods of time, thermal fluctuations are sufficient to mean that the bond is likely to break. This has been shown with biomembranes used as force sensors (see Sect. 8.2). In fact, the influence of the load rate differentiates between the avidin–biotin and streptavidin–biotin pairs. In the case of the avidin–biotin pair, for load rates below about ten pN per second, the most probable breaking force is close to zero.

Force–Extension Curves for Different Polymers

Experiments carried out with titin provide a good example of how the AFM cantilever can be used to study protein unfolding. Analysis of these mechanical properties of titin is interesting for another reason, because this protein is located in muscle tissue and is subject to large changes in tension. Its role is to maintain the structure in the presence of stresses. From this point of view, the way in which a protein folds plays a crucial role in determining its function. However, it is hard to determine the energy landscape of complex peptide sequences. As a consequence, it is often impossible to predict the arrangement of sequences from knowledge of the chemical composition of peptides, and likewise when it comes to inferring properties such as their resistance to unfolding. Therefore the only possible course of action is to measure the forces that can bring about this unfolding. In this context, titin, also

called connectin, provides an exemplary system owing to the fact that its mechanical properties are essential for its biological functions. The passive stress developed by the sarcomere muscle under a stretching action is mainly due to this protein. It is organised in pairs (tandem Ig) in the immunoglobulin domain. This tandem is an extendible chain which resists stresses due to large extensions of the sarcomere.

The experiments described in [118] to measure force–extension curves refer to titin and a macromolecule, in fact a polysaccharide, able to produce very fast conformational changes for large extensions. The authors of this study were thus concerned with two systems. In the first, the characteristic relaxation times were of the order of the experimental time, while in the second, they were much shorter than the stretching time.

Under extension, a polymer chain exhibits different strain regimes. For small forces (a few pN), such as those exerted by optical or magnetic tweezers (see Sects. 8.4 and 8.5), measurement results agree well with standard models for entropic elasticity in polymers. These results are exemplified by a great many studies carried out on DNA. In intermediate force ranges (a few tens of pN), when the elastic response of the polymer backbone must be taken into account, some adjustments to the ideal or semi-rigid chain models are required. When the forces are much stronger (a few hundred pN), as can be reached with an atomic force microscope, the elastic regime is no longer relevant. Significant changes appear in the conformations of biopolymers. This is the case when a protein unfolds. When conformational changes involve a large number of atoms, these conformational changes lead to long characteristic times for reorganisation. If these times are comparable with or less than experimental times, the shape of the force–extension curves will depend on the rate at which stretching is effected. In this case, one must take into account kinetic features during conformational change.

Force measurements on the polysaccharide polymer and on a sequence of 8 tandem Ig segments are shown in Figs. 8.24a and b. On the polysaccharide, for large forces around 700 pN, there is a plateau in the force–extension curve. For a relatively large increment in the extension of the chain, there is only a very slight increase in the force (see Fig. 8.24a). This plateau is interpreted as a consequence of the rotation of a C–C bond in the sugar ring. Despite these strong forces and the change in conformation that accompanies them, the stretching and relaxation curves are identical. These curves thus exhibit no hysteresis, implying that the rotation of the C–C bond is an adiabatic transformation.

The measurements on titin concern 8 units of Ig domains. A characteristic sawtooth behaviour is observed. Each tooth corresponds to the uncoiling of an Ig domain. The first uncoiling occurs in the mechanically weakest domain. The forces measured and the shape of the curves depend on the rate at which extension is imposed, i.e., the stretch rate (see Fig. 8.25).

When the conformation of the polymer backbone deviates from that of the ideal chain described by a random walk, the elastic response of entropic origin

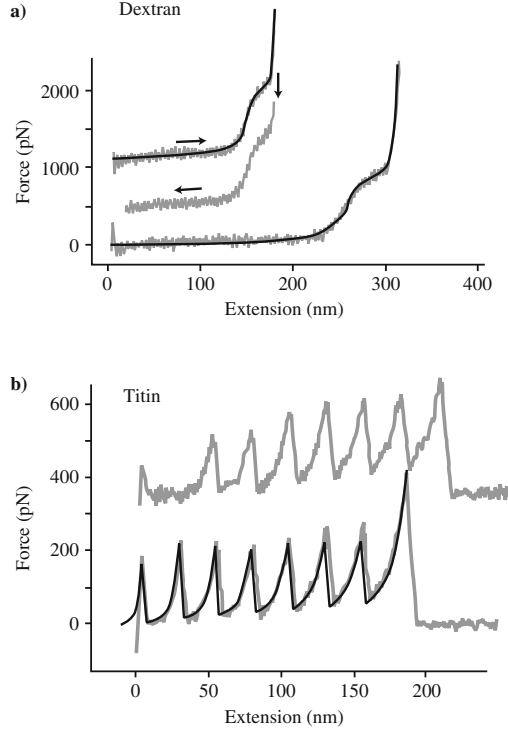


Fig. 8.24. Characteristic force–extension curves for two types of polymer. (a) Three curves obtained for two dextran chains of different lengths. These curves show a plateau near 700 pN, corresponding to the rotation of a C–C bond. The curves for the shorter chain have been shifted along the ordinate axis for better visualisation. The first curve was obtained during stretching and the second during retraction. *Black continuous curves* correspond to the results of numerical simulations including the elastic response of a semi-rigid polymer [see (8.13) and Fig. 8.26]. (b) Force curves obtained for a sequence of eight immunoglobulin domains of the muscle protein titin. Each event corresponds to the unfolding of one domain. The weakest domain unfolds first. The *black continuous curve* corresponds to the result of a numerical simulation [118]

must take into account the locally rigid nature of the polymer. In this case, the polymer is semi-rigid. The best example here is DNA, with a persistence length of 50 nm, characterising the correlation length of the orientation of the osculating plane along the chain. The characteristics of semi-rigid polymers described by Kratky and Porod are obtained using the wormlike chain model. This gives the following relation between the force F and the extension x of the chain [111]:

$$F(x) = \frac{k_B T}{p} \left[\frac{1}{4(1 - x/L)^2} - \frac{1}{4} + \frac{x}{L} \right], \quad (8.13)$$

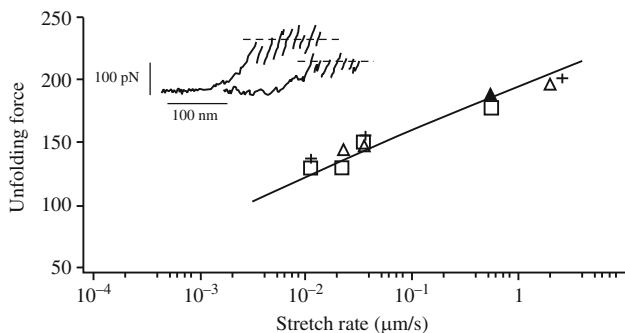


Fig. 8.25. The force required to uncoil a domain depends on the stretch rate. In the *inserts*, a segment of titin 400 nm long is stretched at a rate of $0.5 \mu\text{m/s}$. At this rate, the average force required to unfold is of the order of 190 pN. When the rate is $0.01 \mu\text{m/s}$, the average force decreases to 130 pN. The semi-logarithmic graph shows three different experiments. When the stretch rate is reduced by three orders of magnitude, the average force decreases to 100 pN

where p is the persistence length which describes the angular correlation of the tangent along the chain, L is the contour length, and k_B is the Boltzmann constant. Equation (8.13) gives the elastic response of the semi-rigid chain. By construction, this model is no longer sufficient to describe the observed curves, since the shape of the force curves required to unfold the Ig pairs depends on the stretch rate (see Fig. 8.25).

In a similar way to what was done when considering the binding force of avidin–biotin pairs, in order to account for the mechanical response, a term must be added to the expression for the force given by (8.13) to account for thermal fluctuations. This term can be introduced by considering a model in which two states are separated by a potential barrier. In the case of the polysaccharide dextran, these two states are supposed to describe the positions of the carbon bond. For the Ig domain, the two states correspond to the folded and unfolded states of the protein, I_f and I_u , respectively. The possibility of going from one state to the other will depend on the symmetry of the system and the respective heights of the potential barriers.

The unfolding of a protein by an external force can be described as a process in which the increase of the external force reduces the potential barrier. Thus, during the stretching experiment, thermal fluctuations will be more effective in assisting the unfolding transition as the potential barrier is lowered. As a consequence, by using a two-state model in which the transition between the states can be significantly assisted by thermal fluctuations, one expects the force required to unfold the protein to depend on the stretch rate and, in particular, one expects it to decrease when the stretch rate is diminished. From this point of view, the behaviour illustrated in Fig. 8.25 is in complete agreement with the prediction by E. Evans et al., which stipulates that the unfolding force should be a logarithmic function of the stretch rate [113].

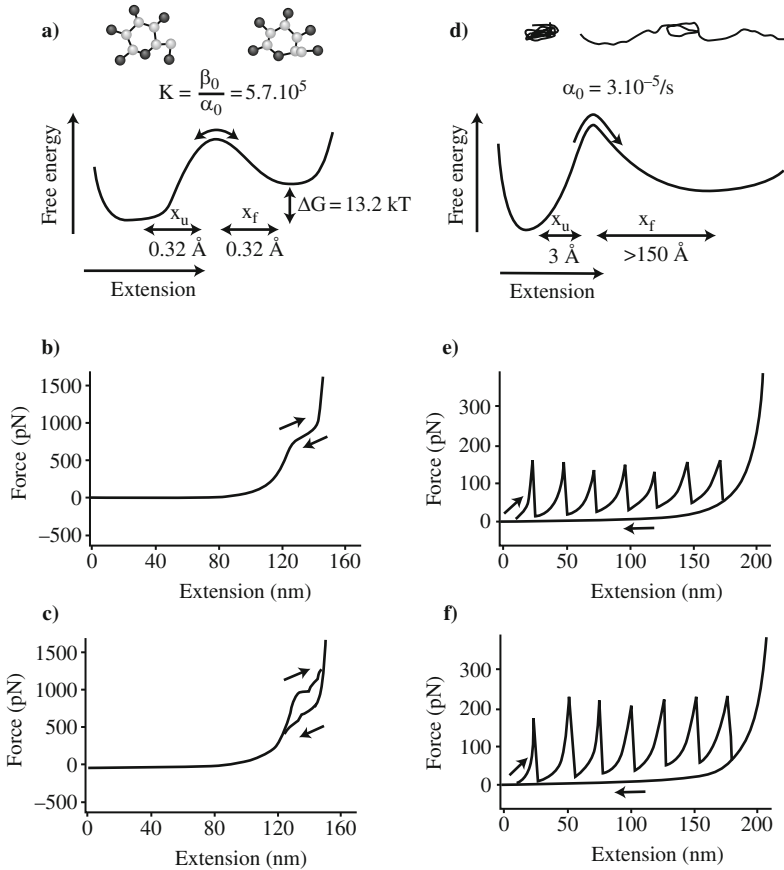


Fig. 8.26. (a) Parameters of the two-state model for conformational change in dextran used for numerical simulation (see also Fig. 8.24). (b) Results of a simulation of force–extension curves with stretching and relaxation, using a stretch rate of $1 \mu\text{m/s}$. The transition occurs at thermodynamic equilibrium. (c) Results of a simulation of the same curves for a stretch rate of 1 cm/s . At this rate, hysteresis is observed, indicating non-adiabatic behaviour. (d) Parameters of the two-state model for conformational change in Ig domains of titin (see also Fig. 8.24). The potential is asymmetric, with the spatial extent of the potential in the folded state some two orders of magnitude smaller than in the unfolded state. (e) Simulation for a very low stretch rate ($0.01 \mu\text{m/s}$). (f) Simulation for a higher stretch rate ($1 \mu\text{m/s}$), corresponding to the measurements given in Fig. 8.24

The reaction diagrams for the changes in conformation of the tandem Ig and polysaccharide are shown in Fig. 8.26. As can be seen from these results, in the case of dextran, the force–extension curves are the same for stretching and relaxation. This reversibility indicates that the reversal of the C–C bond occurs much faster than the experimental time. It is only above a stretch rate

of 1 cm/s that hysteresis begins to occur between stretching and retraction. As an indication, given that the reversed orientation corresponds to a deformation of the order of 0.65 Å, this means that, at this speed, the experimental time during which the deformation occurs on this scale is a few nanoseconds. It follows that, for these experimental conditions, which cannot be reached in atomic force microscopy, the angular reversal of the bond is non-adiabatic on this time scale.

According to the model proposed by G. Bell [110] and developed by E. Evans [113], the two-state model shown in Fig. 8.26 leads to an unfolding rate given as a function of the force by

$$\alpha(F) = \omega e^{-(\Delta G_u - Fx_u)/k_B T} = \alpha_0 e^{Fx_u/k_B T}, \quad (8.14)$$

and a folding rate given as a function of the force by

$$\beta(F) = \omega e^{-(\Delta G_f + Fx_f)/k_B T} = \beta_0 e^{-Fx_f/k_B T}, \quad (8.15)$$

where ω is the characteristic frequency of the system at the bottom of the well, and α_0 and β_0 are the unfolding and folding rates at thermodynamic equilibrium.

The two-state model which describes folding and unfolding of Ig domains is highly asymmetric. The width of the activation barrier for forcing unfolding is 3 Å. Furthermore, for an unfolded peptide chain, with a force as weak as 10 pN, the polymer will be stretched by half its contour length. As a consequence, for a chain of total contour length 30 nm, corresponding to an Ig domain, and for this force, the width of the activation barrier will be of the order of 15 nm. Using (8.14) and (8.15), it follows that the folding rate will be at least e^{-35} smaller than the unfolding rate. In other words, folding is blocked. Consequently, on the time scale of an AFM experiment, the unfolding of proteins in the titin domain is a non-equilibrium process.

Protein Unfolding and Images

We end this section with an image showing the result of protein unfolding. These results were obtained by A. Engel and coworkers of Biozentrum in Basel, Switzerland, pioneers in the field of AFM protein imaging in liquid media, and the reference in this area today [125, 127–131]. The S layers of bacteria are very simple membrane structures. The inner part of these S layers is made from proteins arranged hexagonally and is called the hexagonally packed intermediate (HPI) layer.

Each protein complex making up the hexagon comprises six monomers. The Basel group achieved the feat of unwinding one of these monomers, then imaging the result. To be more precise, five of the six monomers were unfolded and removed, as shown in Fig. 8.27.

There are other examples, e.g., with polypeptide rings of bacteriorhodopsins, obtained by the same group in Basel, and a great many discussions and images of proteins in liquid media. This last point is treated briefly in Sect. 8.3.4.

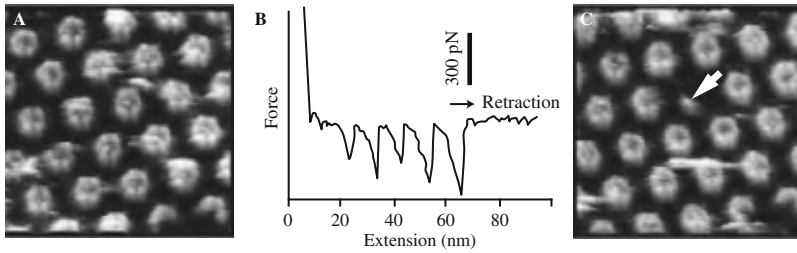


Fig. 8.27. Uncoiling a protein complex comprising several monomers. (A) Inner surface of an S layer (HPI layer) of the archaeobacterium *Deinococcus radiodurans*. The six protomers making up the protein complex and arranged in a hexagon are clearly visible. (B) Force–extension curves similar to those obtained with titin, showing the uncoiling of five monomer units of the protein complex. (C) Image made in the same region after the force measurement. At the site of the monomer complex, indicated by a *white arrow*, only one monomer unit remains [129]. Copyright Elsevier (1997)

8.3.3 AFM Oscillating Modes: Introduction and Definitions

Dynamic AFM: An Oscillating Nanotip

There are two methods for measuring the behaviour of a nanotip oscillating in the vicinity of a surface, in non-contact and intermittent contact situations:

- The first consists in fixing the phase of the oscillator at $-\pi/2$ and holding the oscillation amplitude constant. This is called frequency modulated AFM (FM-AFM). The signals extracted are the shift $\Delta\nu$ in the resonance frequency and the excitation amplitude a_{exc} that must be applied to hold the oscillation amplitude constant. The shift in the resonance frequency depends only on conservative interaction forces. The excitation amplitude a_{exc} depends only on energy losses from the interacting oscillator. The signal extracted from the servo loop which holds the oscillation amplitude A_0 constant is a measure of the damping constant of the cantilever.
- The second method consists in fixing the frequency and amplitude of the excitation. This is called amplitude modulated AFM (AM-AFM). This is what is commonly called tapping mode. The signals measured are the variations in the amplitude A and phase ϕ . The two signals mix up conservative interactions and energy losses. Hence, the amplitude variation results from instantaneous interactions between the oscillator and the sample, but also the extra damping due to the interaction, since the measurement is carried out at a fixed excitation value a_{exc} . The phase also contains both amplitude variations and variations in the damping coefficient. Table 8.3 sums up the main characteristics of the two modes.

Table 8.3. Characteristics of FM- and AM-AFM

FM-AFM	AM-AFM
Constant phase $\phi = -\pi/2$	Constant excitation frequency ω_{exc}
Constant oscillation amplitude A_0	Constant excitation amplitude a_{exc}
Measurement	Measurement
(1) Shift in resonance frequency $\Delta\nu(A_0, D) = \nu(A_0, D) - \nu_0$	(1) Oscillation amplitude A
(2) Excitation amplitude $a_{\text{exc}}(A_0, D)$ or damping coefficient $\gamma_{\text{tot}} = \gamma_0 + \gamma_{\text{int}}(A_0, D)$	(2) Phase ϕ , with the relation $\sin \phi = -\frac{\omega_{\text{exc}} A}{\omega_0 A_0} \left[1 + \frac{\gamma_{\text{int}}(A, D)}{\gamma_0} \right]$
(3) Servo-control of vertical position of surface. A fixed value of the frequency shift is generally used to servo-control the tip-sample distance, with $\Delta\nu(A_0, D) < 0$	(3) Servo-control of vertical position of surface. A fixed value of the drop in oscillation amplitude A is generally used to servo-control the tip-sample distance, with $\Delta A(\omega_{\text{exc}}, a_{\text{exc}}, D) < 0$

Sensitivity and Noise. An Example in FM-AFM

Originally, this mode was designed to provide a way around the response time of the oscillator imposed by amplitude detection in ultrahigh vacuum conditions [93]. The frequency shift is instantaneous, and hence should in principle provide a way round this drawback. However, this is not perfectly true. For one thing, measurement of the frequency shift requires a time of millisecond order. For another, if this measurement is to be fully significant, the oscillation amplitude must be held constant, and this in part restores the effect of the quality factor.

FM-AFM really came into its own in 1995 after it was demonstrated that it could achieve atomic scale contrast on the 7×7 reconstruction of silicon (111) [100, 104, 106] and on InP [105]. Since then, many experiments have been done, mainly by German and Japanese groups, leading to the creation of an annual conference devoted to this technique in 1998 (Osaka, 1998).

In STM, the distance between the tip and the sample is controlled by the tunnel current I_z :

$$I_t(z) = I_0 \exp(-2\kappa z), \quad (8.16)$$

where I_0 is a function of the applied potential and the electron densities of states of the tip and sample, and κ is the characteristic length of the evanescent wave in the insulator (the vacuum between tip and sample), given by

$$\kappa = \frac{\sqrt{2m\Phi}}{\hbar}, \quad (8.17)$$

with m the electron mass, $\hbar$ the reduced Planck constant, and Φ the work function of the junction. For a metal, $\Phi \approx 4\text{ eV}$ gives $\kappa \approx 1\text{ \AA}^{-1}$. When

z increases by one angstrom, the current changes by an order of magnitude. The noise in the distance z is related to the noise in the current measurement. Defining the level of noise in the quantity x by

$$\delta x = \sqrt{\langle (x - \langle x \rangle)^2 \rangle},$$

we have

$$\delta z = \frac{\delta I_t}{|\partial I_t / \partial z|}. \quad (8.18)$$

The tunnel current is measured using a current–voltage converter and the noise is dominated by the noise from the resistance in the converter [103]:

$$n_R = \sqrt{4k_B T R}.$$

At 300 K, taking a tunnel current $I_t = 100$ pA, a bandwidth $B = 1$ kHz for the measurement, and a resistance $R = 100$ M Ω , the noise in the voltage V is obtained as

$$\delta V = \sqrt{4k_B T R B} = 40 \mu\text{V},$$

which leads to a noise δI_t in the current of

$$\delta I_t = 0.4 \text{ pA}.$$

The noise in the tip–surface distance will be

$$\delta z = \frac{\sqrt{4k_B T B / R}}{2\kappa |I_t|} \approx 0.2 \text{ pm}. \quad (8.19)$$

This very low intrinsic noise and an exponential function sensitive to distance variations of angstrom order explains the relative ease with which atomic scale images can be obtained.

For FM-AFM, the tip–sample distance is controlled by the frequency shift of the oscillator. For the interaction force, we use the expression corresponding to the case of van der Waals forces with a sphere–plane interaction. This is an interaction that varies slowly with distance and must be treated as a long-range interaction as far as atomic scale contrast is concerned:

$$\nu(z) \approx \nu_0 \left(1 - \frac{H R}{6k A^{3/2} z^{3/2}} \right), \quad (8.20)$$

where k is the cantilever stiffness, A is the oscillation amplitude, H is the Hamaker constant (in joule), and R is the tip radius. Using our STM notation, we denote the closest distance by z . It will subsequently be denoted by Δ in the usual way. The noise in z is given by

$$\delta z \approx \frac{\delta \nu}{|\partial \nu / \partial z|}. \quad (8.21)$$

The noise in the frequency is given by [93, 101]

$$\delta\nu \approx \nu_0 \sqrt{\frac{2k_{\text{B}}TB}{\pi^3 k A^2 \nu_0 Q}}. \quad (8.22)$$

At 300 K, so that $k_{\text{B}}T \approx 4 \times 10^{-21}$ J, with $k = 10 \text{ Nm}^{-1}$, $B = 1 \text{ kHz}$, $A = 5 \text{ nm}$, $\nu_0 = 200 \text{ kHz}$, $Q = 20,000$, this implies $\delta\nu_0 \approx 10^{-1} \text{ Hz}$. The derivative of the frequency shift is given by

$$\frac{\partial\nu(z)}{\partial z} \approx \nu_0 \frac{HR}{4kA^{3/2}z^{5/2}}. \quad (8.23)$$

With $H = 10^{-19} \text{ J}$, $R = 1 \text{ nm}$, and a distance of $4 \text{ \AA}$, equation (8.23) yields 44 Hz/angstrom , and with (8.21) to a noise in z of $\delta z \approx 0.2 \text{ pm}$. In terms of forces, this implies force variations of the order of 0.1 pN .

A difficulty that is specific to atomic force microscopes is the possibility of long-term drift due to the high mechanical sensitivity of the embedded cantilever. Hence, environmental variations, e.g., in the temperature and humidity, are likely to modify image acquisition conditions. From this point of view, a vacuum or a controlled atmosphere are experimental conditions that improve measurement stability. For example, a temperature change leads to non-negligible drift. The resonance frequency of the cantilever is given by

$$\nu_0 \approx \frac{e}{L^2} \sqrt{\frac{E}{\rho}} = \frac{e}{L^2} v_s,$$

where e is the thickness, L the length, E the elastic modulus, and v_s the speed of sound. With α the coefficient of thermal expansion, the relative frequency shift is given as a function of temperature by

$$\frac{1}{\nu_0} \frac{\partial\nu_0}{\partial T} = \frac{1}{v_s} \frac{\partial v_s}{\partial T} - \alpha.$$

For silicon, $\alpha = 2.6 \times 10^{-6} \text{ K}^{-1}$ and $(1/v_s)\partial v_s/\partial T = -5.5 \times 10^{-5} \text{ K}$, and we obtain a shift $\Delta\nu/\nu_0 \approx -5.8 \times 10^{-5} \text{ K}^{-1}$. For a resonance frequency of 200 kHz , this corresponds to a shift by 6 Hz for a temperature change of 0.5 K . This is a large drift [see (8.23)], which can lead to height variations of several angstroms.

Local Rheological Properties

A good example of the use of the frequency modulation mode is provided by the study of triblock copolymers comprising a central elastomer part and two thermoplastic sequences (see Fig. 8.28). These materials exhibit a phase separation on the nanoscale with very different mechanical properties [107], the central part being much softer than the outer sequences.



Fig. 8.28. A triblock copolymer comprising a central elastomer sequence, viz., a poly(isooctylacrylate) for the experiments discussed here, to which are fixed two thermoplastic sequences of poly(methylmethacrylate), better known as PMMA

It is no easy matter to synthesise these materials with well-defined sequence lengths. The main idea is to have materials that are easy to shape, while combining the mechanical properties of the various constituents. Hence, networks have been made in which the nodes are no longer chemical as in vulcanised materials, but are in fact glassy. These systems have the advantage that they are easier to recycle by simple heating, in contrast with chemically crosslinked networks. Height and dissipation images are shown in Fig. 8.29.

Figure 8.29a shows no relief, i.e., the copolymer surface is flat. The image of dissipated energy clearly shows the phase segregation between the elastomer and thermoplastic sequences. Brighter zones indicate that, when the tip is located above elastomer regions, the AFM oscillator consumes more energy to hold the oscillation amplitude constant than when it is located over thermoplastic regions [96]. Observed contrasts are a direct consequence of the different mechanical properties, resulting from the viscoelastic response of the sample under the action of the oscillating tip. The diagrams in Fig. 8.30 sum up the way in which contrasts of mechanical origin can be enhanced by adjusting the average distance between the end of the oscillating tip and the sample surface. This average distance is controlled by the magnitude of the shift in the resonance frequency of the oscillator. In the attractive interaction region, the more negative the frequency shift, the smaller the average separation between the tip and sample. As a consequence, the tip-sample interaction increases.

AM-AFM Mode: Influence of the Quality Factor and Investigation of Soft Materials

AM-AFM mode was first used to control the approach of a tip in order to be able to carry out measurements in the optical near field [102]. Even at this early stage, the nonlinear behaviour of the cantilever was already more or less established. Then came commercial AFM systems, leading to the rapid development of AM-AFM under the name of tapping mode AFM. The virtually immediate success of this mode is due to the undeniable facility with which soft materials can be imaged, without causing irreversible damage. The main reason, as compared with the static contact mode, is the reduction of shear forces between the tip and sample.

As we saw above for the FM mode, the quality factor is a parameter that controls not only the noise level, but also the sensitivity. The availability of this parameter in amplitude modulation can become an important criterion in

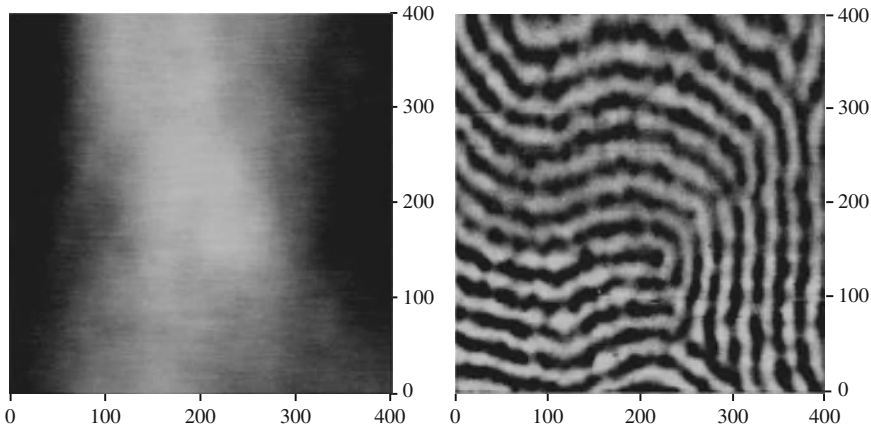


Fig. 8.29. *Left:* Height image obtained by holding fixed a shift in the resonance frequency of the AFM oscillator. The height contrast is 3 nm. *Right:* Dissipation image recorded simultaneously with the height image. *Bright zones* indicate that the energy dissipated by the oscillator is greater than in *dark zones*. These zones correspond to the elastomer segment

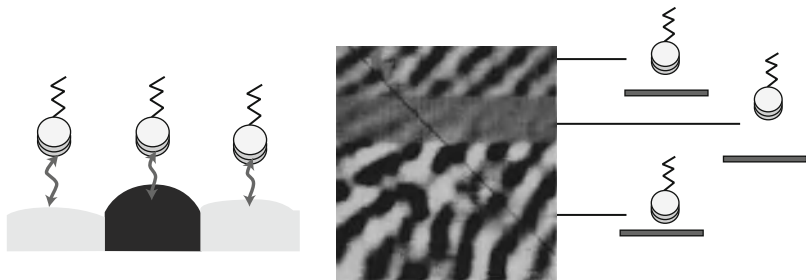


Fig. 8.30. *Left:* Interaction of the tip with different domains. *Right:* Dissipation image showing that the contrast varies with the tip–sample separation. When the shift in the resonance frequency is small, corresponding to a weak interaction force between the tip and surface, the dissipation contrast fades out

the choice of measurement depending on the properties of the sample. When studying soft materials, it is a determining factor. In order to understand its influence, let us begin by discussing the relation between the amplitude variation and the interaction with the sample in a somewhat simplified context. In fact, we shall consider that the whole resonance curve of the oscillator is shifted without change of shape.

In other words, the perturbation is weak enough to mean that the interaction between the AFM oscillator and the sample can be adequately described in the framework of a linear response. In reduced coordinates, with $u = \omega/\omega_0$, $a = A/A_0$, $d = D/A_0$, where D is the distance between the cantilever at rest

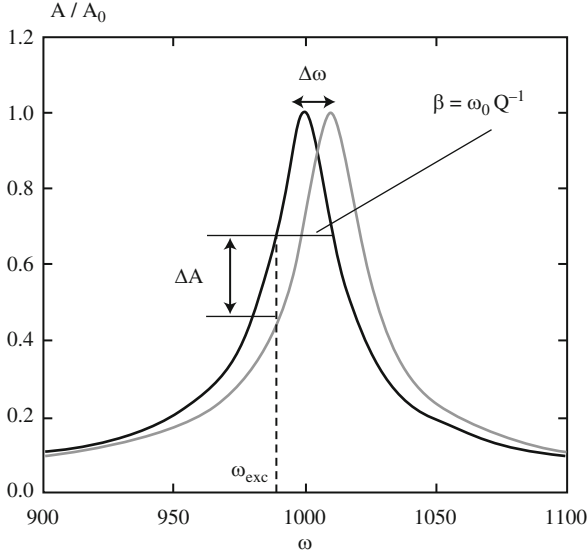


Fig. 8.31. Shift of the resonance curve of the oscillator under the effect of a small perturbation. The Lorentzian shape of the resonance curve is conserved. The width at half-height is inversely proportional to the quality factor Q . To a first approximation, the variation ΔA of the amplitude at fixed angular frequency ω_{exc} , for a shift $\Delta\omega$, can be calculated using (8.27)

and the surface, the expressions for the harmonic oscillator are

$$a(u) = \frac{1}{\sqrt{Q^2(1-u^2)^2 + u^2}}, \quad (8.24)$$

$$\varphi(u) = \arctan \frac{u}{Q(u^2 - 1)}. \quad (8.25)$$

The presence of a force gradient will lead to a translation along the frequency axis whose sign will depend on the type of interaction (see Fig. 8.31):

$$u'_F = u \pm \Delta u_F \approx u \pm \frac{\nabla F(d)}{2k}. \quad (8.26)$$

For completeness, note that the tip-sample interaction generally leads to nonlinear oscillator behaviour, and this will usually distort the resonance curve [92, 94, 95, 109]. For small frequency shifts, the amplitude variation corresponding to the fixed angular frequency is evaluated using the relation

$$\Delta a_F \approx \frac{da}{du} \Delta u_F, \quad (8.27)$$

with the slope of the amplitude variation given as a function of frequency by

$$\frac{da}{du} = -\frac{u[2Q^2(u^2 - 1) + 1]}{[Q^2(u^2 - 1)^2 + u^2]^{3/2}}. \quad (8.28)$$

This expression simplifies for excitation frequencies u situated at $-\pi/4$ or $-3\pi/4$, which are also close to the maximum sensitivity of the oscillator. At these excitation frequencies, we have $Q(u - 1) = 1$, and with $Q \gg 1$, which means that we may use $u \sim 1$,

$$\frac{da}{du} \approx Q. \quad (8.29)$$

Near the maximum sensitivity of detection of the amplitude variation, the slope is proportional to the quality factor. From a practical standpoint, this is an important property for obtaining images with as small a perturbation of the sample as possible. Let us examine the case in which the surface has an elasticity which leads to a finite contact local stiffness.

To simplify the discussion, we follow the reasoning in [108]. The sample surface is treated as being simply elastic, without attractive interaction or viscous behaviour. Furthermore, we consider a linear response of the elastic force of the surface as a function of the indentation by the tip. This simplification amounts to ignoring the intrinsic difficulty of contact mechanics, which gives a nonlinear variation of the elastic response as a function of the indentation. Put briefly, the linear response approximation is not correct. However, this simplification has the significant advantage of leading to analytic solutions adequate for describing the main features of the oscillator behaviour.

The parameter that interests us here is the variation of the oscillation amplitude as a function of the indentation depth δ_e of the tip in the sample. Let k_e be the contact stiffness characterising the elastic response of the sample. The contact stiffness is simply the product of the diameter of the contact area and the elastic modulus of the surface. Recall first the result for the static mode, which relates the displacement of the cantilever to the vertical displacement of the surface. If k is the stiffness of the cantilever, the slope of the change δ_L in the cantilever deflection is given by

$$p(\delta_L) = \frac{1}{1 + k/k_e}. \quad (8.30)$$

In the dynamic mode, the slope $p(A) = \delta A / \delta D$ of the variation of the oscillation amplitude is calculated from [108]

$$p(A) = \frac{1}{1 + b(A)(k/Qk_e)^{2/3}}, \quad (8.31)$$

where $b(A)$ is a function of the amplitude. This remains more or less constant over a good part of the amplitude variation [108]. The main result comes by comparing the slopes calculated for the static and oscillating modes. In the

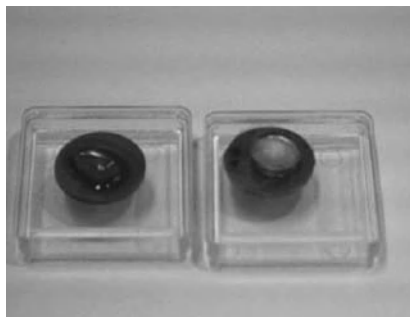


Fig. 8.32. Samples of polyisoprene with molecular weights of 10^6 and 200,000 dalton. The latter is placed in a basin, in fact the crucible of a differential calorimeter, to avoid the flow of the lower molecular weight

case of the static mode, only the ratio of the two stiffnesses is relevant. If the surface is soft and has quasi-liquid behaviour, the ratio of the stiffnesses is large and the slope is almost zero. The cantilever will not be deflected whatever the tip penetration depth. At the other extreme, for a hard surface leading to a stiffness ratio close to zero, the slope is equal to unity and leads to a cantilever deflection equal to the displacement of the surface. In the oscillating mode, it is not just the contact stiffness that comes in, but the product of this quantity with the quality factor. As a consequence, the quality factor can increase significantly under the influence of the elastic properties of the surface. For example, consider a contact stiffness of 10^{-2} Nm^{-1} and a cantilever of stiffness 1 Nm^{-1} . In the static mode, the cantilever deflection will be almost zero, of the order of one percent of the surface displacement. In oscillating mode, with a quality factor of $Q = 500$, the slope becomes of the order of 80%.

The results obtained for a polymer melt, using polyisoprene, provide a good illustration of this (see Figs. 8.32 and 8.33). A polymer melt behaves as a liquid over long time scales. The characteristic time giving the limit between the rubber plateau and liquid behaviour is a function of the molecular weight of the polymer chain. In the time range shorter than the characteristic time, corresponding to the elastic plateau, the contact stiffness is small. For example, for an elastic modulus of 10^7 N m^{-2} and a contact diameter of 10 nm, the stiffness is $k_e = 0.1 \text{ N m}^{-1}$. A simple variation of the quality factor is obtained by varying the pressure in a pressure chamber. By reducing the pressure, the viscous contribution of the gas is reduced. At different pressures, one can observe the evolution of the slope of the oscillator variation. At 10^{-7} torr, with a quality factor of $Q = 23,000$, the polymer melt appears as hard as a silica surface, and this despite the high cantilever stiffness, of the order of 30 N m^{-1} [97].

These results show why it can be interesting to use an oscillator with a high quality factor when studying soft materials. This is one of the fundamental problems when using the oscillating mode in a liquid medium. In the

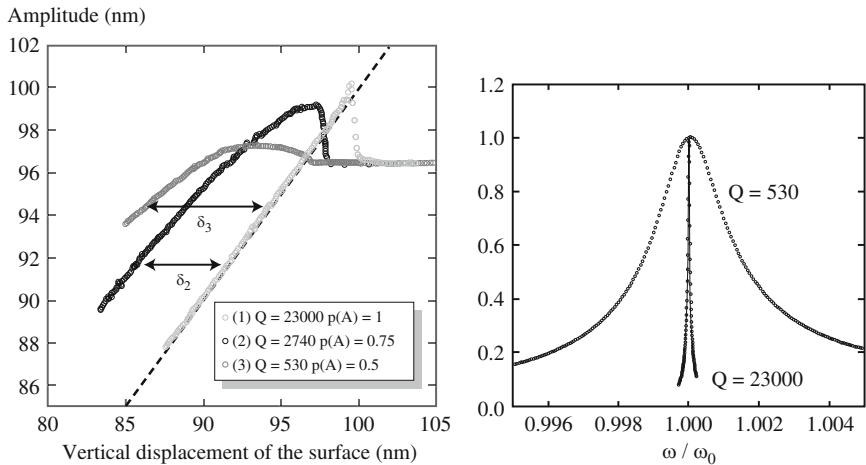


Fig. 8.33. Change in the oscillation amplitude for intermittent contact on a polymer melt (polyisoprene). At atmospheric pressure, $Q = 530$. At $p = 10$ mbar (7.5 torr), $Q = 2,740$ and at $p = 2 \times 10^{-7}$ torr, $Q = 23,000$. The *dashed line* gives the position of the surface. *Arrows* give the maxima of the indentation depth $\delta_3 = 8$ nm and $\delta_2 = 6$ nm for $Q = 530$ and $Q = 2,740$, respectively [97]. The resonance curves at atmospheric pressure and at 2×10^{-7} torr are shown *on the right*

next section, we shall discuss several attempts to produce images in a liquid medium using an oscillating mode. We shall give examples to illustrate the general development of these attempts, without entering into any analysis of the images obtained. The latter are discussed in more detail in Sect. 8.1.

8.3.4 Oscillations in a Liquid Medium

The cantilevers generally available are usually of the order of $150\mu\text{m}$ long, $20\text{--}30\mu\text{m}$ wide, and a micrometer or more thick. In contrast to measurements in ultrahigh vacuum, or in air, the approximation wherein one forgets the cantilever as a whole and considers only the end of the tip interacting with the sample surface becomes untenable in a liquid medium. In this context, the whole cantilever must be taken into account in order to give an adequate description of the movement and the influence of the fluid carried along with the displacement of the cantilever. The main difficulty is therefore that the cantilever is always a few tens of micrometers wide and a few hundred micrometers long, which inevitably leads in water to hydrodynamic forces of the order of a few nanonewtons. In comparison with the vacuum, or even air, there is no way of escaping the fact that it is no longer just the tip-sample interaction which governs the oscillating behaviour of the cantilever. Over the past ten years or so, a great deal of experimental and theoretical work has been carried out to get around this difficulty [141,143,144]. From the experimental standpoint, one aims either to reduce the dimensions of the cantilever, or to

exploit some kind of electronics that can servo-control the cantilever motion in quadrature and thereby reduce the influence of the viscous motion.

When a cantilever of length L and width W oscillates in a liquid medium, it carries with it an extra mass of water and is subject to additional damping due to the viscosity of the liquid. The additional mass and the damping coefficient can be simply expressed in the following way [139]:

$$\begin{cases} m_{\text{add}} = \rho_{\text{fluid}} \frac{\pi}{4} W^2 L \left(a_1 + a_2 \frac{\delta}{W} \right), \\ \gamma = \rho_{\text{fluid}} \frac{\pi}{4} W^2 L \omega \left[b_1 \frac{\delta}{W} + b_2 \left(\frac{\delta}{W} \right)^2 \right], \end{cases} \quad (8.32)$$

with numerical coefficients $a_1 = 1.0553$, $a_2 = 3.7997$, $b_1 = 3.8018$, and $b_2 = 2.7364$, and δ a length given by

$$\delta = \sqrt{\frac{2\eta}{\rho_{\text{fluid}}\omega_{\text{fluid}}}},$$

where η is the viscosity and ρ_{fluid} the density of the liquid. δ is a sort of skin thickness over which the speed imposed by the oscillation of the solid in the fluid vanishes. Hence, around the oscillating solid, there is an evanescent wave that characterises the velocity field in the fluid. Taking the water viscosity as 1 cp (at 20.20°C), a thickness $\delta \approx 1.4 \mu\text{m}$ is obtained in water and $5.3 \mu\text{m}$ in air. Using the expression (8.17) for the damping, for a cantilever of stiffness 30 N m^{-1} , oscillating at the resonance frequency 150 kHz with width $20 \mu\text{m}$ and length $150 \mu\text{m}$, quality factors of $Q_{\text{air}} \approx 420$ and $Q_{\text{water}} \approx 2.4$ are obtained in air and water, respectively. These values agree well with measured values of the quality factors, showing that damping arises mainly from the viscosity of the fluid.

In view of what has been presented so far, it is clear that a quality factor of 2.4 virtually eliminates all sensitivity from the AFM oscillator. In a liquid medium, the point that differentiates the oscillating mode from the contact mode is that, in practice, the lateral forces remain small. However, it remains to be unambiguously proven that there is some real advantage in the present state of the art.

Several approaches have been put forward to try to solve these problems. A considerable effort has been made to use interferometric detection at very small amplitudes [138]. Right at the beginning of studies in liquid media, a lot of work was done to develop very small cantilevers [143, 144]. As can be seen from (8.32), the hydrodynamic drag varies as the product $W^2 L$, so reducing the width can significantly reduce the hydrodynamic force. However, this reduction requires one to modify the beam by reducing the size of the spot on the back of the cantilever. Typically, this means going from a spot diameter of ten micrometers or so down to 2 to 3 micrometers, and this requires a specific setup [143]. The second point concerns the excitation process. The

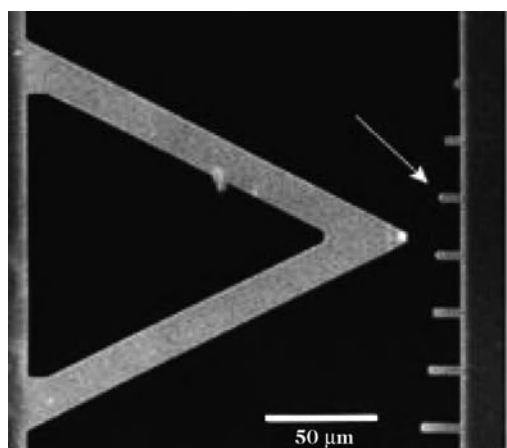


Fig. 8.34. The tiny cantilevers used in these experiments are shown by an *arrow*

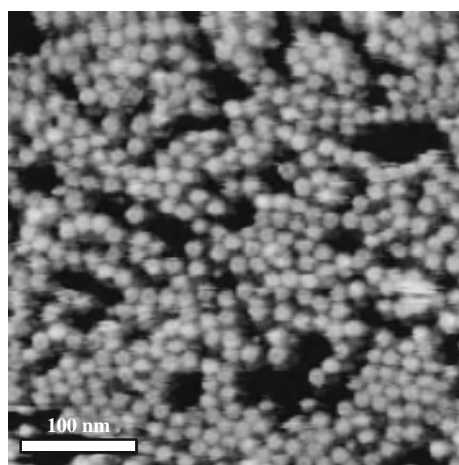


Fig. 8.35. Height image of GroEL adsorbed onto mica in a solution. The total height scale is 15 nm. The central channel of the GroEL is visible as a black spot on the protein [135]

simplest and most widely used excitation mode in commercial devices is by piezoactuator. The main drawback is that it leads to a large number of modes which bear no relation to the normal mechanical modes of the cantilever being set in resonance via the acoustic excitation. There are other methods, either by magnetic excitation, or by electrostatic excitation. Magnetic excitation is in principle the best suited. However, it requires the deposition of a cobalt bead measuring several micrometers in diameter. As a consequence, this method can only be used with wide cantilevers.

Among the many attempts to use oscillating modes in a liquid medium, let us consider one which seems to us the most interesting with regard to the goals fixed for studying protein activity [135]. Viani et al. aimed to use very

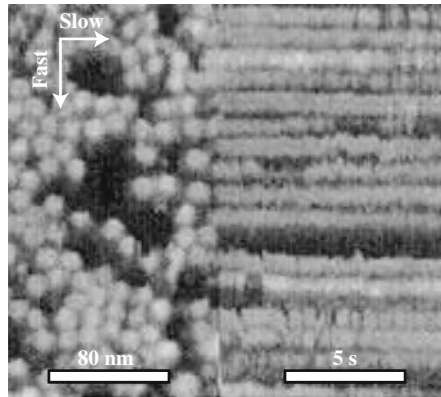


Fig. 8.36. Time study. *On the left* is a classic X,Y height image. *On the right*, the displacement along the X axis has been stopped. The X axis is now a time axis. The image is of a single row, with scanning only along the Y axis. Each row corresponds to the evolution of a single protein at position X,Y where the X scan has been stopped. The time between two measurements on the same protein is 100 ms instead of 25 s in the classic X,Y image

small cantilevers to observe the interactions between proteins in real time. The chosen protein assemblies were the chaperone proteins GroES which pair up and dissociate GroEL proteins. GroEL is a protein which repairs proteins that have been incorrectly synthesised or proteins containing conformational defects. The tiny cantilevers are rectangular and measure about ten micrometers long and a few micrometers wide (see Fig. 8.34). The microscope optics can produce 3- μm spots, compared with those in commercial microscopes which measure 20 μm or more. The high resonance frequencies (300–500 kHz) allow the tip to scan the surface quickly, typically 10 Hz per row.

The first images were taken without magnesium and ATP, or GroES (see Figs. 8.35 and 8.36). Under these conditions, the height dispersion of the GroEL did not exceed 0.5 nm (see Fig. 8.37). However, when GroES was added with Mg-ATP, the authors observed a steplike variation in height. The amplitude of the steps was 3.6 ± 1 nm. This height variation agrees with what is expected if GroES binds onto GroEL. These height variations are only observed when both GroES and Mg-ATP are present in the solution.

If one wishes to analyse the dynamics of association and dissociation when GroES pairs with GroEL, it is essential to overcome the problem of the acquisition time required to form the image. Even for a scan as fast as 10 Hz per line, for an image of 512 lines, the time interval of more than 50 s between each image is obviously unacceptable. One way of improving the temporal resolution by reducing the lapse of time between each measurement is to stop scanning along one axis (see Fig. 8.36), or even both, and to record the variations in the cantilever oscillating above the GroEL.

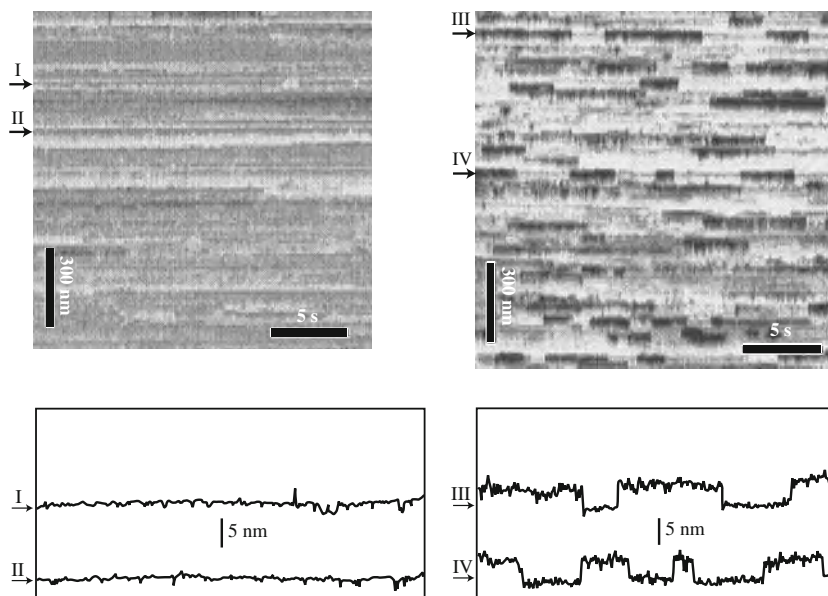


Fig. 8.37. Association and dissociation of GroES–GroEL complexes. When the X scan is stopped, ‘tubes’ are produced in the image, and the X axis becomes a time scale. The cross-sections of rows I and II show no particular features. In the *right-hand image*, GroES and Mg-ATP have been added. The cross-sections III and IV show steps whose heights suggest that GroES–GroEL has formed

This is what was done in these experiments. Figures 8.36 and 8.37 show images in which the X axis is the time axis. Since one row takes 25 s to scan, with 256 points per row, the time interval between each measurement is 100 ms. On the left of Fig. 8.37, corresponding to the deposition of GroEL, the cross-sections reveal rather flat profiles without any particular features. This result shows that the tip remains on the protein and that the height of the protein does not exhibit any significant fluctuations. When GroES and Mg-ATP are added, steps are observed in the cross-sections, with heights of 3.6 ± 1 nm. These values agree with the height differences measured between GroEL and the complex formed by GroEL and GroES.

The technical developments of oscillating modes in liquid media are currently attracting a good deal of attention. In particular, many groups are working on the idea of piezoactuator excitation, including several Japanese groups in Tokyo, Osaka, and Kyoto. One could mention the work by Kawakatsu using an oscillator that oscillates at frequencies of several MHz, or work using phase-locking in order to ensure that the cantilever oscillation occurs at the mechanical resonance.

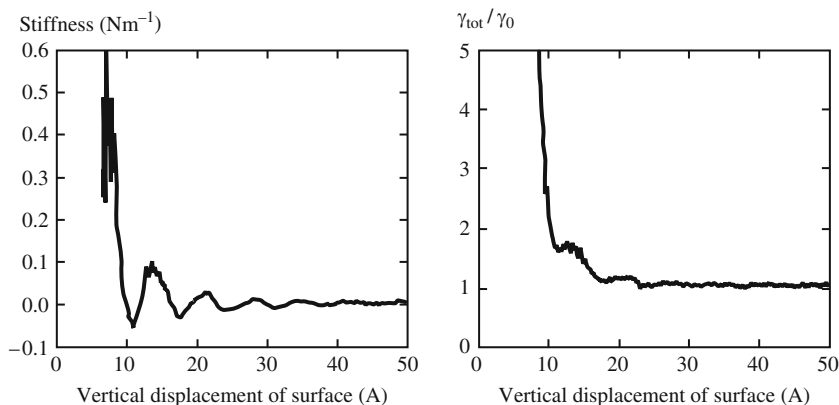


Fig. 8.38. *Left:* Dependence of the stiffness of confined OMCTS molecules on the tip–surface separation. The period is $7.5 \text{ \AA}$, corresponding to the size of the molecule [140]. These measurements were obtained with acoustic excitation and an oscillation amplitude of $2 \text{ \AA}$. *Right:* Corresponding variation of the dissipation function. One can also see oscillations resulting from structuring of the liquid

Furthermore, technical improvements are currently being made which allow one to control the mechanical resonance of the cantilever [136]. Among other results, they allow one to study confinement in a liquid medium by using very small amplitudes [138, 140]. Figure 8.38 shows the variations in stiffness and dissipation governed by ordering of fluid molecules confined between an oscillating tip and a surface. These results are similar to those obtained with the surface force apparatus [137].

8.3.5 Force Measurements and Height Images. DNA Measurements

To Touch or Not to Touch DNA with an Oscillating Nanotip

Tapping mode force microscopy measures force variations. Consequently, however the tip–surface distance is controlled, height variations obtained using an image will always be the result of a balance of interactions between the tip and different regions of the surface. To illustrate this claim, we shall discuss several experimental results obtained with DNA [132]. Hundreds of publications can be found describing measurements made on DNA that has been deposited on different substrates. The aims are many and varied, including investigations of the interaction between a substrate and DNA, comparative studies of different techniques in microscopy, analysis of chain conformations in 2D, and manipulation of DNA by an AFM tip [134, 135].

The idea of the experiment is summarised in Fig. 8.39. Depending on the choice of experimental conditions, the tip either oscillates in such a way as to touch the DNA in an intermittent manner, or oscillates at a certain distance

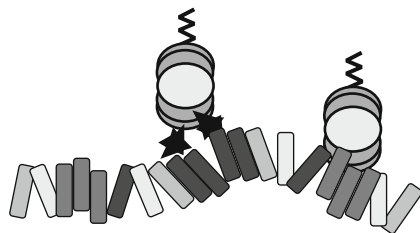


Fig. 8.39. The elementary building blocks A, T, C, and G of DNA are represented by *four different shades of grey*. The two situations of contact and non-contact between the oscillating tip and the DNA are shown. In both cases, the local structure of the sequences can lead to different interactions with the tip, leading in turn to different relief in the image

from the chain. In the case of intermittent contact, the image will reveal relief and possibly mechanical properties of the subject, e.g., elastic response and relaxation. In the second case, we measure specific properties of the chemical entities arising from long-range interactions, such as electron polarisability or the distribution of electrical charges in the molecule.

In this experiment, the control parameter is the evolution of the oscillation amplitude as a function of the tip-sample separation. An excitation frequency is fixed close to the resonance frequency, and the amplitude is measured when the tip is very far from the surface. Let A_∞ be this value, where the subscript ∞ indicates that the measurement was made very far from the surface. A reference value of the amplitude $A_c < A_\infty$ is then fixed. As long as the oscillation amplitude does not reach the value A_c , the oscillating tip is brought toward the surface. When the surface is reached, the surface is scanned horizontally and an electronic feedback loop servo-controls the vertical position of the surface in such a way as to hold the value of the oscillation amplitude equal to A_c . It is these variations in the vertical positions that are registered and produce the images presented. Note also that one can simultaneously record other images containing complementary information about the evolution of properties of the oscillator, such as phase variations.

It is easy to control the contact and non-contact situations by varying the excitation amplitude of the oscillator, which corresponds to varying the value of A_∞ . This procedure can be understood qualitatively in the following way. In most cases, as long as the tip is not at a distance of nanometric order, the attractive interaction between tip and sample is negligible. Hence, unless it actually touches, for oscillation amplitudes of a few tens of nanometers, the tip will only have a significant interaction with the surface for a fraction of the period which depends on the value of this amplitude. The bigger the amplitude, the less the interaction related to non-contact situations will be significant. At very large amplitudes, since the time spent near the surface becomes much less than the period, the attractive interaction will have a

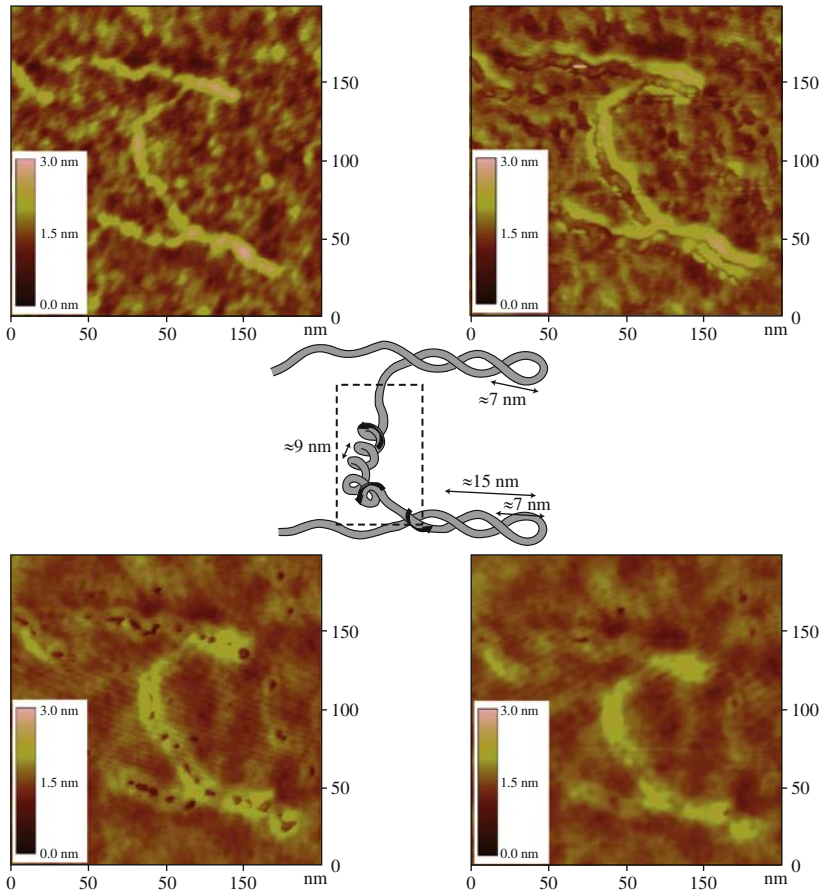


Fig. 8.40. Images of DNA, moving progressively from situations in which the tip touches to others in which it passes over without contact. (a) $A_\infty = 49$ nm. (b) $A_\infty = 11$ nm. (c) $A_\infty = 9$ nm. (d) $A_\infty = 6$ nm

negligible effect on the properties of the oscillator. On the other hand, if a small amplitude is used, these interactions will be favoured.

The balance between these situations also depends on the intensity and the nature of the interactions, and on the size of the end of the tip. If the different chemical constituents lead to spatial distributions of the dipole moments, one expects to find regions of the image which correspond either to non-contact situations or to contact situations, for the same reference value of the amplitude. This is a way of producing images which reflects the heterogeneity of the physicochemical properties of the sample.

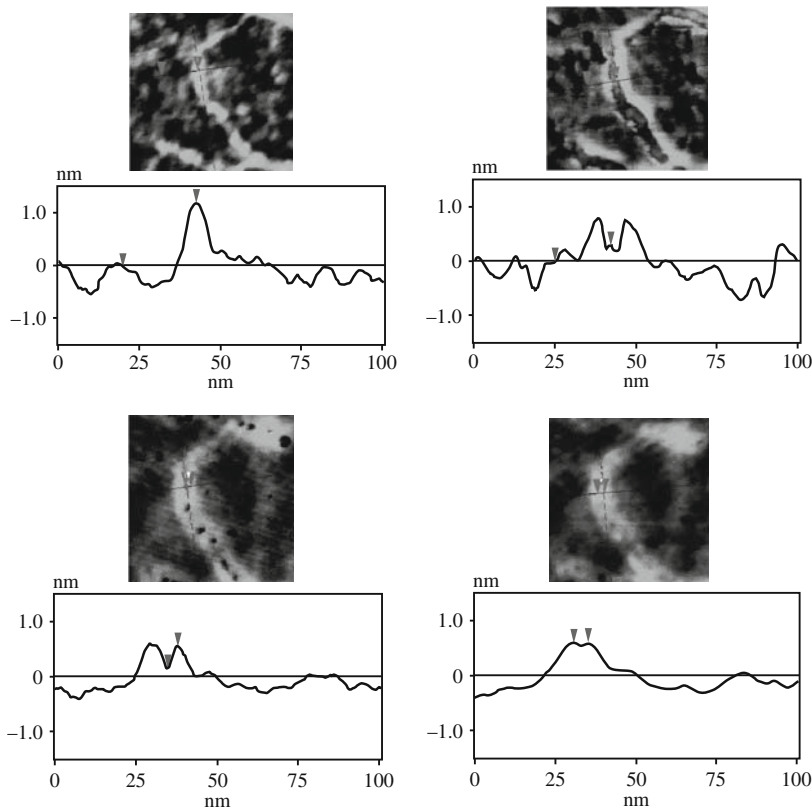


Fig. 8.41. Cross-sections of images obtained for decreasing oscillation amplitudes. From *left to right* and from *top to bottom*: 49 nm, 11 nm, 9 nm, 6 nm. The smaller the oscillation amplitude, the more the attractive interaction between tip and sample will dominate the oscillator behaviour. For example, for the *bottom right* image obtained with an oscillation of amplitude 6 nm, the DNA appears broader and structureless. The richest image in terms of information is the one for amplitude 11 nm (*top right*), where superstructures can be observed, revealing a stronger local attractive interaction along the chain

To sum up, at small amplitudes, the behaviour of the oscillator will be determined by attractive interactions or non-contact situations, whereas at large amplitudes, its behaviour will be dominated by repulsive interactions or contact situations.

Figure 8.40 shows images of DNA obtained using different oscillation amplitudes, moving progressively from situations in which the tip touches the DNA intermittently to others in which the tip flies over it without contact. All zones are the same, only the conditions of analysis are modified.

The central part of the figure shows the shape of the object and the dashed rectangle indicates the region where cross-sections were taken (see Fig. 8.41). The DNA studied here is a sequence of 2,500 base pairs with arbitrary

arrangement of the nucleotides. These images show the ‘height’ variations of the surface which are required to hold the reference amplitude constant. The brightest zones correspond to the highest regions of the surface if the variations in the oscillation amplitude are entirely due to the relief. To simplify, if we omit the contribution from the mechanical response of the DNA, this is effectively the case in Fig. 8.40a. In this first figure, the tip intermittently touches every point of the surface, including the DNA. Over the whole surface, the oscillator behaviour is dominated by repulsive interactions. In this case a topographical survey of the surface is obtained.

In Fig. 8.40b, the oscillation amplitude has been reduced and the figure shows significant differences. To begin with, part of the DNA is no longer illuminated, i.e., the DNA has lost height. A still more remarkable result is the over-illumination appearing along the macromolecule. This over-illumination corresponds to a greater ‘height’ than the DNA. In practice, this means that the oscillating nanotip feels the surface at a greater distance when it is near the macromolecule. It can be shown [132] that this type of situation is encountered if the attractive interaction is locally strong enough compared with the other zones. For this choice of amplitude, the conditions of analysis are particularly discriminating. On the grafted surface and over most of the DNA, the tip touches intermittently, whereas in the vicinity of the DNA it no longer touches.

This over-illumination indicates the presence of a particularly strong attractive interaction around the periphery of the DNA. When the amplitude is reduced still further, this phenomenon is enhanced. In Fig. 8.40c, only a few zones (the small dark regions) of the DNA are not overflown by the tip.

For the smallest amplitude (Fig. 8.40d), the tip passes over the surface at a greater distance. As a consequence, the contrast is less clear and we observe fewer details than in Fig. 8.40b. This evolution is particularly clear when Fig. 8.40c is compared with Fig. 8.40b.

A slightly different way of visualising these results is to carry out cross-sections in the same region of the image for the different measurement conditions (see Fig. 8.41). In going from the amplitude of 49 nm to an amplitude of 11 nm, the DNA is seen to sink (central arrow), or to be precise, the abutments along the edge of the DNA are seen to grow, in such a way that the very importance of these zones masks the DNA itself. When the effect due to the attractive interaction is amplified, the DNA disappears and these abutments tend to overlap (see Fig. 8.41, bottom). In order to explain this evolution, an additional interaction must be introduced over a very localised zone. One possible explanation for a locally stronger interaction in the vicinity of the DNA is the formation of rows of dipoles. The origin of these dipoles can be explained as follows. DNA, which is an acid, carries one anion per elementary unit at pH 7. At this pH, the amine-grafted silica we use has positively charged amine groups (Figs. 8.42 and 8.43).

The presence of rows of dipoles substantially increases the attractive interaction between the tip and the surface, while preserving the localised nature of the interaction. Under experimental conditions for which the oscillator is

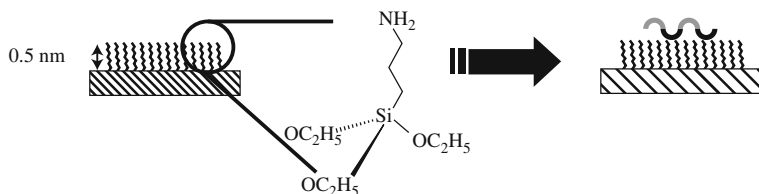


Fig. 8.42. AFM requires the sample to be immobilised on a surface. In the case of DNA, which is an acid, the best way to do this is to prepare an alkaline surface, e.g., silica grafted with silane molecules terminated with an amine group. At pH 7, the latter are protonated

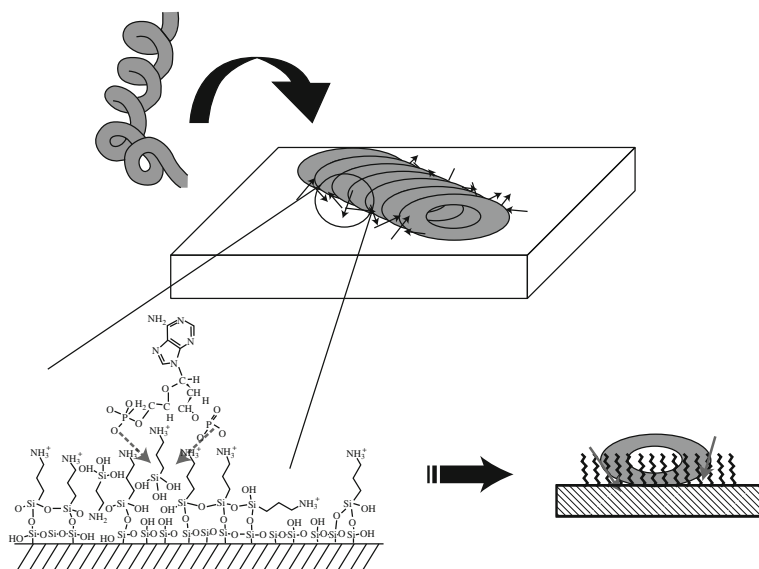


Fig. 8.43. This diagram shows the approach developed to understand the over-illumination observed along the DNA molecule in the attractive regime. We consider two rows of randomly oriented dipoles along the main structure

at the boundary between the repulsive and attractive regimes, when the tip passes over the row of dipoles, a significant increase in surface height should be observed. Figure 8.44 shows a numerical simulation which illustrates the effect of these rows of dipoles in more detail. The result is compared with the experimental results. These experiments thus show the extent to which the notion of height in force microscopy is above all the result of a balance in the interactions between tip and sample on the one hand, and between tip and substrate on the other.

Since dynamical AFM is above all a form of microscopy that is very sensitive to variations in the interaction imposed by entities of nanometric dimensions or smaller, this method can reveal the various structures of DNA

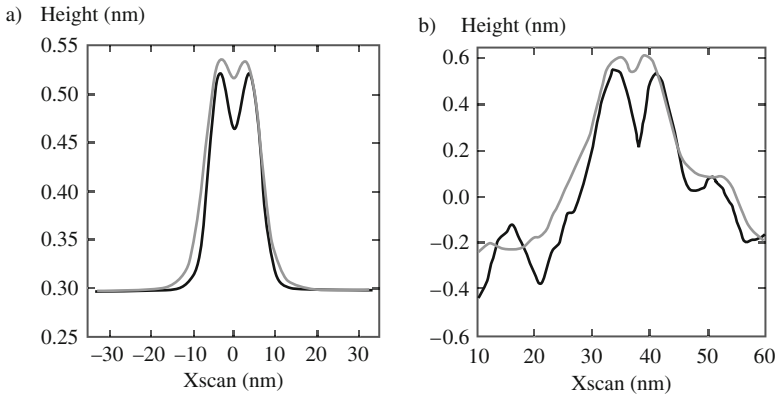


Fig. 8.44. (a) Numerical simulations of the influence of rows of randomly oriented dipoles on the height feedback control in a dynamical force microscope for two oscillation amplitudes (15 nm and 7 nm). (b) Magnified profiles observed for the oscillation amplitudes 9 nm and 6 nm, corresponding to the two figures at the bottom of Fig. 8.41

(topographic mapping) and at the same time identify specific local properties by appropriately exploiting the interaction regimes between tip and sample.

DNA Viewed from the Chromosome to the Nucleosome

For many years now, Eric Le Cam and coworkers at the Gustave Roussy Institute have been studying the structure and dynamics of the interaction between DNA and proteins using electron microscopy and AFM. Here we describe a study devoted to the hierarchical organisation of DNA packing within the nucleus (see Fig. 8.45) [126].

DNA is a macromolecule with contour length over one meter. It must therefore be compacted within the cell nucleus, whose diameter does not exceed ten micrometers or so. In order for DNA to enter such a confined space, various levels of hierarchical organisation are required. The top left image of Fig. 8.45 shows all the chromosomes in a human cell at the time of the metaphase in cell division (image size 50 μm). The second image on the top row shows the two parts of the chromosome containing the same genetic information. During cell division, the two parts will separate into the two daughter cells (image size 7 μm). The third image shows a partial unwinding of the chromosome. Only part of the arm is visible on the right. The DNA is compacted by proteins that can be observed on this scale (image size 5 μm). The fourth image (bottom left) is a magnification of the previous image. The structure is less compact and one can begin to make out the coiling of the DNA around proteins (image size 1.2 μm). To study these structures more closely, the group at the Gustave Roussy Institute isolated the proteins causing the DNA to coil – these proteins are called histones – in order to place them in contact with

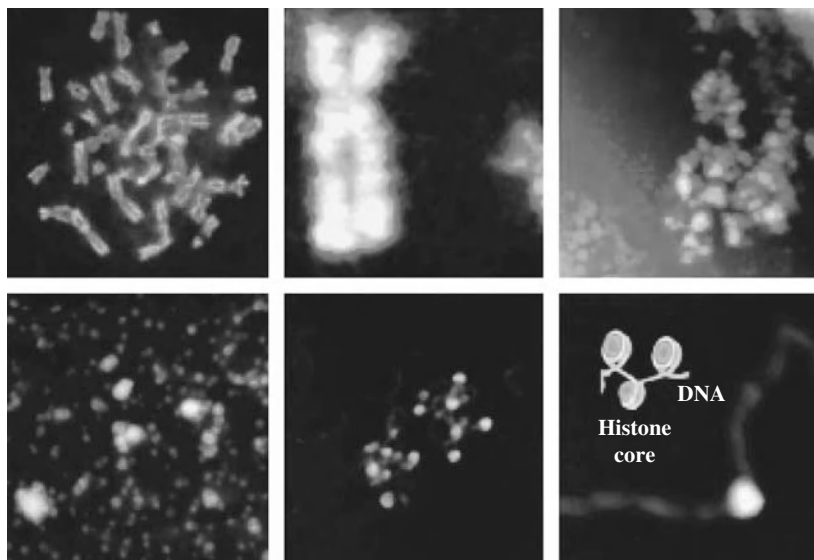


Fig. 8.45. Visualisation of the hierarchical organisation of DNA within the nucleus

DNA. The fifth image shows how the DNA coils up in the presence of histones (image size 400 nm). The last image is a magnification of the close-packing between DNA and the histone.

Acknowledgements

We would like to thank E. Perez for his comments on measurement of binding forces in avidin–biotin pairs.

8.4 Optical Tweezers

8.4.1 Basic Principles and Main Parameters

Optical tweezers are a form of microscopy using a highly focused laser beam. This technique is used to trap and manipulate dielectric particles by exerting forces of a few piconewtons on them. The first experimental demonstration of optical trapping dates back to 1986 and is due to Ashkin et al. [145]. These authors showed that, when refracting particles are lit by a focused laser beam, they are attracted toward the regions of maximal laser intensity.

The basic principle of optical trapping can be understood in a simple way through the laws of geometric optics. Each time a light ray crosses the interface between the dielectric medium and the particle, a momentum change is produced by the refraction phenomenon. This momentum change is characterised

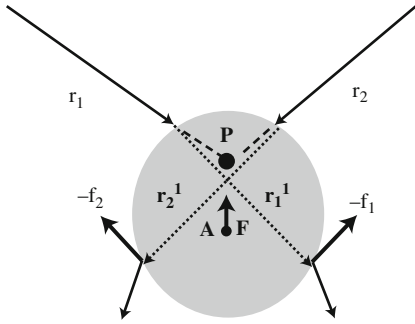


Fig. 8.46. Optical tweezers. The restoring force F acts from the center of the bead (A) to the position (P) where the two beams intersect when the sphere is not there

by the change in speed and direction of the light. Momentum conservation implies that, if the momentum of a light ray changes by the vector quantity Δp , the dielectric must suffer the opposite change in momentum, viz., $-\Delta p$, and hence a force f equal to the change in momentum per unit time.

The total force exerted by the laser beam on a dielectric sphere can be calculated by representing the beam as an ensemble of independent rays and summing the contributions from each ray. Qualitatively, one can see what direction the trapping force will point by considering the change in momentum due to just two incident rays r_1 and r_2 which are refracted into rays r_1^1 and r_2^1 , as shown in Fig. 8.46. The point P would be the point of intersection of the two rays if the sphere were not there. The vectors $-f_1$ and $-f_2$ represent the forces due to the refraction of the two rays. By symmetry, the resultant force F lies along the optical axis and points toward P.

In addition, one must consider a dispersion force related to absorption and reflection of light by the dielectric particle. This force lies along the optical axis and in the direction of propagation of the beam. The equilibrium position of the particle depends on the resultant of these two forces (restoring force and dispersion force) and will thus be located close to the focal point, slightly shifted in the direction of propagation of the laser beam. Under these conditions, the particle is trapped and lateral or axial scanning of the laser beam can be used to displace the particle in the three space directions. One can also work with a fixed laser beam and displace the sample.

The quality of the optical trap depends largely on three parameters. To begin with, there is the difference of refractive index between the dielectric particle, e.g., 1.57 for glass beads and 1.35 for latex beads, and the surrounding medium (typically water or a weakly saline medium, with refractive index 1.33), which increases the refraction effect. Second, the trapping force increases linearly with the laser intensity. In practice, laser powers between 10 mW and 1 W are used, corresponding to forces in the range 10–1,000 pN and to a light intensity of 10^6 – 10^8 W/cm² in the focal plane. When optical tweezers are used in biology, it is important not to damage the sample by heat or photodegradation generated at such power levels. For this reason, lasers with wavelengths in the near infrared (700–1,300 nm) are preferred, because it is

in this range that biological tissues are most transparent. Finally, at a given laser power, trapping is made more effective by focusing the laser beam. For this, microscopes with high numerical aperture (typically 1.3 or higher) must be used.

8.4.2 Estimating the Stiffness Constant of the Trap

To a first approximation, an optical trap behaves like a spring, i.e., the force $\mathbf{F}$ exerted on a bead of micrometric dimensions is proportional to the distance x between the center of the bead and the center of the trap, so that $F = -kx$. The most relevant parameter to be characterised is thus the equivalent stiffness k of the trap. There are several ways of estimating the value of k [146–148].

A first method is to superpose a lateral friction force on a microsphere trapped in a liquid medium. To do this, the sample is displaced with a motorised translation stage and the trap held fixed, or a laminar flow chamber is used. Stokes' relation then gives the imposed force as

$$F = -6\pi\eta r v,$$

where η is the fluid viscosity, r is the radius of the microsphere, and v is the speed of flow around the bead. Under the effect of this force, the bead, still trapped, will adopt an equilibrium position that is shifted a distance Δx relative to the center of the trap. By measuring the values of Δx for different speeds v , it is easy to determine the value of k . Beyond a critical speed which depends on the laser intensity, the bead will escape from the trap. This speed determines the maximal force of the trap, called the escape force.

Another way of calibrating the stiffness constant of the trap is to measure the frequency spectrum of the position fluctuations. As described in Sect. 8.5.2, the motion of a Brownian particle inside an optical trap of stiffness k in a fluid of viscosity η can be approximated by Langevin's equations (quoted here in one dimension for simplicity):

$$\gamma \frac{dx}{dt} + kx = F(t).$$

The first term is the force due to viscous friction, where $\gamma = 6\pi\eta r$ is the coefficient of viscosity. The second term is the restoring force. Both are counterbalanced by forces F_t due to Brownian fluctuations.

The Fourier transform of the Langevin equation can be used to deduce the Fourier transform $X(f)$ of the variable $x(t)$, from which the spectral density $S_x(f) = |X(f)|^2$ is obtained as

$$S_x(f) = \frac{k_B T}{\gamma \pi^2 (f_c^2 + f^2)},$$

where $f_c = k/2\pi\gamma$ is the cutoff frequency. To make these measurements, one must use a position detector with better temporal resolution than this cutoff frequency. The fluctuations in the position x are measured and the corresponding spectrum $S_x(f)$ is calculated. The last function is fitted to the curve

obtained experimentally. On this curve, for $f \ll f_c$, one extrapolates the value of $S_x(0)$, from which one extracts the stiffness constant of the trap via

$$k = \frac{2k_B T}{\pi S_0 f_c}.$$

Alternatively, the stiffness constant can be obtained directly from the measurement of f_c and the relation $k = f_c 2\pi\gamma$, if γ is known.

Finally, another method is to use the principle of equipartition of energy. The stiffness constant of the trap is deduced from the expression

$$\frac{1}{2}k \langle x^2 \rangle = \frac{1}{2}k_B T,$$

where x is the displacement of the particle from its average position $x = 0$ and angle brackets denote the average of x over a large number of positions.

8.4.3 Different Types of Optical Tweezers

Multiple Beam Optical Traps

More recently multiple beam optical traps have been developed [149–152]. In these setups, the laser beam crosses several positions in a cyclic manner, using a high-speed scanning system, e.g., involving galvanometric mirrors or acousto-optic deflectors (AOD). One can thereby create an array of several traps with variable geometry. When the beam returns to a position after travelling over the whole array, the particle trapped at this position has not had time to move sufficiently far away, and it is captured again. The use of high-speed deflectors and an intense laser source (in principle, for n traps, one requires n times the power of a single trap) is of great importance here, because the time needed to cover all positions must be less than the characteristic diffusion time (typically around 200 ms for a 1- μm particle). In addition, by adjusting the positions picked out by the laser beam, the multiple trap array can be reconfigured in a dynamic way (see Fig. 8.47).

Multiforce Optical Tweezers

A further improvement of the optical tweezer system is provided by multiforce optical tweezers. In this case, a multiple trap array can be constructed as before, but this time by controlling the laser power delivered individually to each element of the array, since this determines the trapping force at this position [152]. The simplest way to achieve this is to adjust the number of passes made by the laser beam at each position. When the laser beam scans a given position more often than it scans neighbouring positions, the average intensity at this position will be higher, and this will create a stronger trap. This means that the force of each trap can be chosen independently, and one can thereby construct multiforce arrays like the one shown in Fig. 8.48a. Figure 8.48b shows the laser intensity and trapping force profiles for this array [152].

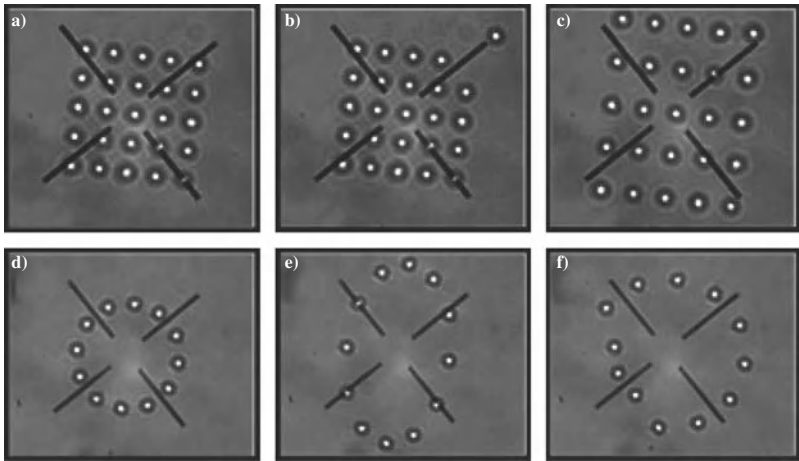


Fig. 8.47. Beads of diameter $2\mu\text{m}$ positioned in a square lattice (a)–(c) or circular lattice (d)–(f). The two geometries can be stretched or deformed in any direction

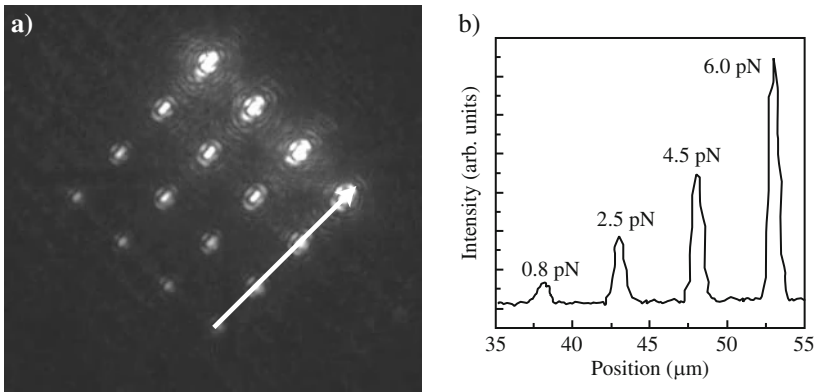


Fig. 8.48. (a) Laser intensity distribution for an array of 4×4 points, reflected on a glass slide and imaged by a CCD camera. The intensity at a point of the array is determined by the average time the laser spends at this point. (b) Laser intensity profile corresponding to the path shown by the *arrow* in (a), indicating the independently measured value of the escape force at each peak

3D or Holographic Optical Tweezers

Another way of creating multiple optical tweezers is to use a holographic method which can simultaneously generate several focal points. Briefly, a diffracting optical element (DOE) is inserted along the optical path in such a way as to modulate the wave front of the laser beam, thereby producing optical points arranged according to arbitrary geometries. The most delicate

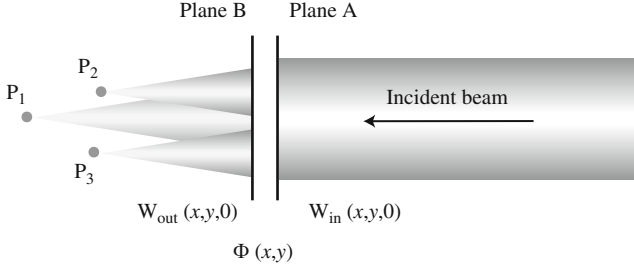


Fig. 8.49. To calculate the profile of the diffracting element (DOE), the starting point is to determine the spatial distribution one would like to create, then work backwards along the light propagation to find the distribution $W_{\text{out}}(x, y)$ of the electromagnetic field that would be generated in plane A by spherical waves propagating from the traps P_1 , P_2 , P_3 . $W_{\text{in}}(x, y)$ is the electromagnetic field distribution due to the trapping laser beam in plane B and $\Phi(x, y)$ is the phase modulation to be imposed on the incident beam to reproduce the distribution $W_{\text{out}}(x, y)$ at the DOE output and to converge on the three points P_1 , P_2 , and P_3

aspect in this approach is to fabricate the DOE. The starting point is to choose the type of multiple trap array one would like to create, e.g., circle, straight line, or any other more sophisticated geometry. One then determines the light distribution which corresponds to waves propagating from the traps. One can thereby calculate the spatial distribution of the phases and intensities in a given plane upstream (profile A). Likewise, one can calculate the light distribution of the trapping laser in a plane B (profile B) separated from plane A by a small volume (see Fig. 8.49).

Once the two distributions have been determined, one can calculate the modification of the wave front that must be imposed on the incident beam crossing the volume contained between planes A and B, in such a way as to reproduce profile A. Hence, the laser beam passing through the DOE will produce the desired array of focal points. Experimentally, the phase modulation is obtained by controlling locally the thickness of a transparent material, e.g., by etching a glass plate using conventional lithographic techniques [153, 154]. The phase of the laser beam is then spatially modulated, because its various components must follow different optical paths through the DOE. As a consequence, the wave front is distorted in a static way.

This system of holographic optical tweezers with static phase modulation has recently been improved by using a computer guided spatial light modulator (SLM). This can modulate the light beam dynamically. Here the phase modulation is achieved by locally controlling the orientation of molecules in a liquid crystal film. Thanks to the fast relaxation of the liquid crystal array, the position of the focal point can be controlled in real time (less than the millisecond) [155–157]. Holographic optical tweezers present several advantages over conventional optical tweezer systems:

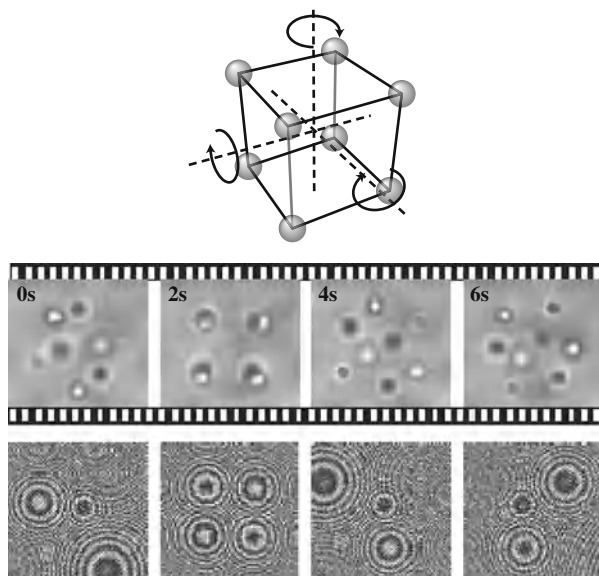


Fig. 8.50. Rotation of a cubic structure made of silica beads with diameter $2\mu\text{m}$, placed $5\mu\text{m}$ apart from one another, and corresponding holograms [158]

- Generation of multiple focal points is no longer restricted to the focal plane of the objective, but can be extended to a 3D lattice (see Fig. 8.50 top). As an example, Fig. 8.50 (bottom) shows eight beads arranged in a cubic structure and set in rotation, along with the corresponding phase profile imposed on the SLM [158].
- The mode of the trapping laser can be changed [159–161]. The most common laser mode, called the transverse electric mode (TEM_{00}), corresponds to a circular Gaussian beam. With a holographic optical element, a TEM_{00} mode can for example be transformed into a Laguerre–Gauss mode. This is characterised by a helical wave front (see Fig. 8.51). Whereas a Gaussian beam forms a single point after focusing, the helical mode gives a torus. This phenomenon leads to two interesting properties. Firstly, the centre of the focal region is dark. Secondly, the toroidal light distribution carries orbital angular momentum. Hence toroidal focusing, often called an optical vortex, can be used to trap reflecting or absorbing particles, or particles with low refractive index, which would otherwise be damaged or repelled by a conventional optical trap. By transferring the angular momentum to the trapped object, this effect, known as an optical spanner, can also be used to apply a torque to it and thereby cause it to rotate [162].

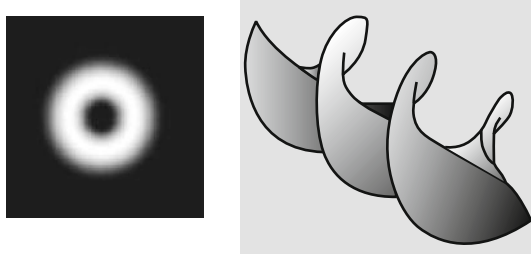


Fig. 8.51. Helical mode of a Laguerre–Gauss beam and toric profile obtained after focusing

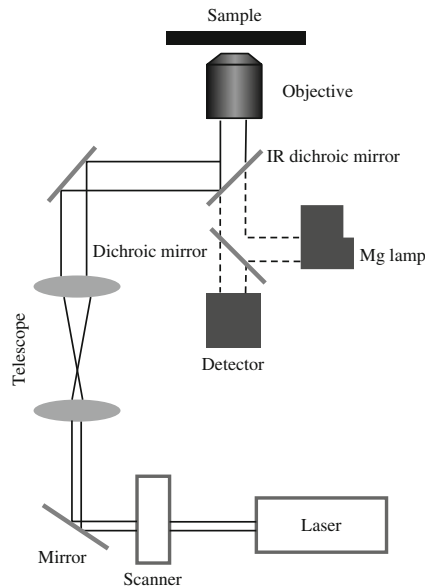


Fig. 8.52. Experimental setup for optical tweezers

8.4.4 Experimental Setup

The standard optical tweezer setup (see Fig. 8.52) is usually built around an inverted microscope, in such a way as to leave the upper part free for manipulating the sample. Once the beam has been passed through mirrors, deflectors, SLM, and so on, and then expanded, it is directed via a dichroic mirror reflecting the infrared (IR) into an objective with high numerical aperture. To maximise the gradient component with respect to dispersion forces, the expansion factor of the laser is adjusted to slightly exceed the input pupil of the objective. In most applications, it is useful to be able to manipulate the sample with optical tweezers and carry out fluorescence imaging at the same time. To do this, the dichroic mirror is positioned between the epifluorescence filter cube and the objective (see Fig. 8.52). Another solution is to use

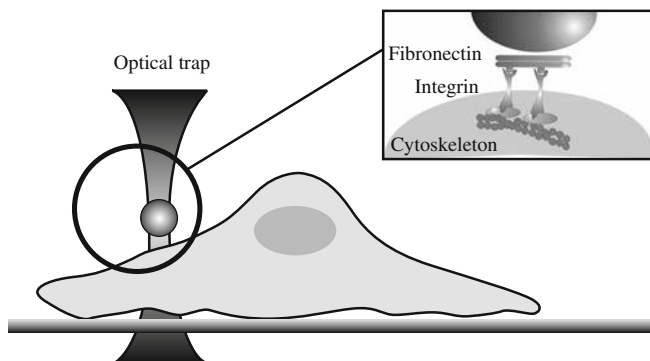


Fig. 8.53. Typical mechanotransduction experiment. A bead coated with fibronectin is placed at the surface of a fibroblast using an optical trap. The recruitment of specific receptors (integrins) at the contact site is used to form a trans-membrane linkage with the underlying actin cytoskeleton

a double dichroic mirror in the cube, reflecting both infrared and the excitation wavelength of the desired fluorophore. Hence, the light coming from the epi-illumination system can be focused on the sample with the same object used to trap, and one can simultaneously gather fluorescence images. Recently, new experimental strategies have been put forward to combine optical tweezers with evanescent wave techniques (TIRF) (see Sect. 7.2) or 3D scanning [163, 164].

To measure small shifts in position (and hence small forces), good spatial resolution is required. For such applications, the best detector is a quadrant diode array [165], which can give a spatial resolution of a few nanometers and a temporal resolution of microsecond order. The diode is situated above the condenser. The laser beam scattered by the particle is projected onto it, with a feedback loop to the translation stage or deflectors. This system can for example be used to servo-control the trap in a certain position relative to the trapped object, to work at constant force.

8.4.5 Biological Applications of Optical Tweezers

Since their discovery, optical tweezers have gained an ever larger following in the biological research community. In this section, we shall discuss some of their applications. For a more exhaustive list, the reader is referred to [166].

Cell Mechanotransduction

Cells are sensitive to chemical signals, but also to their immediate surroundings. The latter can be connected to general processes of adhesion and cell migration or, in some tissues, to specific forces, e.g., flow in blood vessels, acoustic waves in the inner ear. The way cells react and adapt to these mechanical stresses, e.g., by modulating gene expression and the local recruitment of

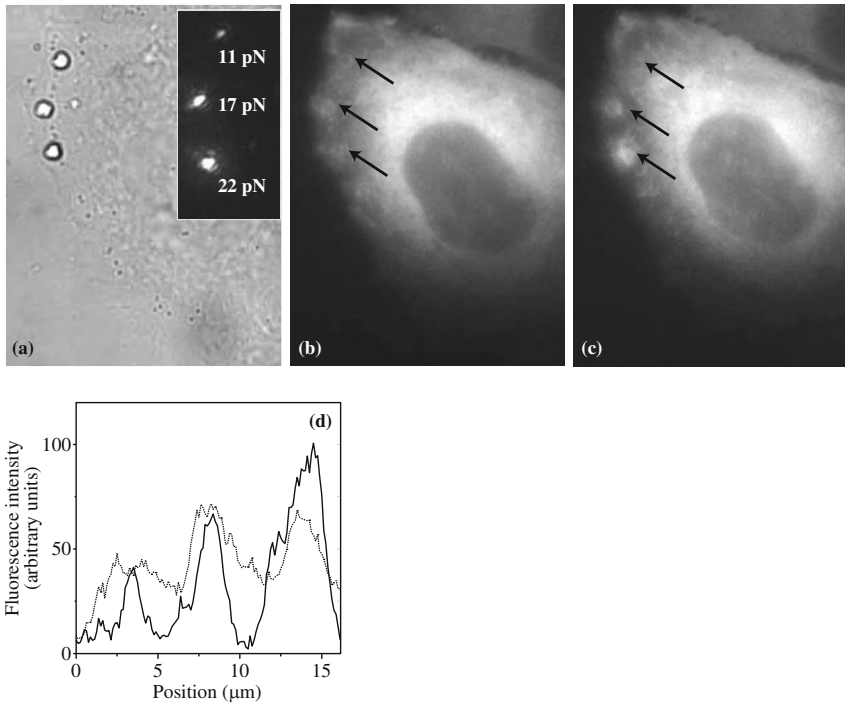


Fig. 8.54. Strengthening of adhesive contacts in response to an external force. Three beads coated with fibronectin have been placed on the cortex of a HeLa cell transfected by vinculin–GFP. (a) DIC image of the cell. Using a multiforce optical tweezer system, the beads are captured by three traps of different stiffness (*insert*). (b) and (c) Fluorescence distribution of vinculin–GFP at times $t = 5$ and 15 min. Note the signal accumulation, showing the selective recruitment of vinculin–GFP on the beads subjected to the greatest forces. (d) Fluorescence intensity profiles at times $t = 5$ and 15 min, showing that the level of recruitment of vinculin–GFP is proportional to the force applied to the bead

proteins, is known as mechanotransduction. The mechanical environment thus modulates the key cell functions, such as morphology, proliferation, and differentiation. To study the processes of mechanotransduction, the cell is usually subjected to an external force reproducing as closely as possible the stresses that can exist

in vivo, and the various mechanical or biochemical parameters are subsequently measured.

To understand the transduction of adhesive forces, an experiment exploiting optical tweezers has been widely used. The idea is to place microspheres coated with specific extracellular ligands on the dorsal surface of migrating cells, thereby imitating point adhesion sites (see Fig. 8.53). Then, using the optical trap, the beads are subjected to forces of different strengths and the

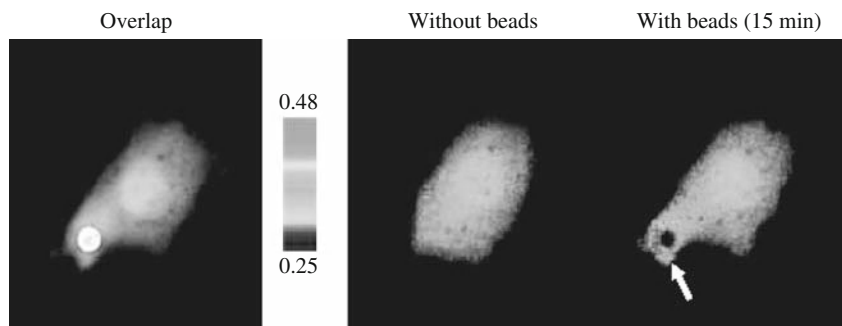


Fig. 8.55. Activation of Src in the vicinity of microbeads coated with fibronectin and placed on the cortex of human endothelial cells. The colour scale codes the FRET signal, related to the Src probe used [171]

reaction of the cell is measured by means of different tests. This type of approach has been used to show that the application of a local force during the initial contact between microspheres coated with fibronectin (a protein from the extracellular matrix) and the lamellipodium of a fibroblast (extension of the cell, made mainly of actin, which the cell forms to be able to anchor itself to the substrate during migration) generates a strengthening of the linkages between integrin receptors and the actin cytoskeleton [167]. A combination of optical tweezers and fluorescence observation of the distribution of GFP-fused proteins was used to show that this strengthening effect is related to an accumulation of associated proteins (vinculin and paxilin) and is controlled by enzyme processes [168]. For example, it has been shown that the amount of vinculin–GFP recruited around beads coated with fibronectin is proportional to the applied force [168,169]. The multiforce optical tweezer system provides the advantage here of being able to measure simultaneously the cell response to a range of different forces. Figure 8.54 shows the recruitment of vinculin–GFP by beads subjected to different forces on the same cell, as observed using a multiforce system. It illustrates the fact that the recruitment of vinculin at adhesion sites is indeed proportional to the applied force [169].

Recently, a more sophisticated setup, allowing the simultaneous manipulation by optical tweezers and measurement of FRET signals (see Sect. 7.3), has been used to relate mechanical forces to activation of enzymes such as Rac, Rho, or Src [170, 171]. These proteins have the particularity of being able to occur in an active state (bound to guanosine triphosphate GTP) or an inactive state (bound to guanosine diphosphate GDP). Labelling them with a single fluorophore, one cannot distinguish these two states. However, if the same protein is associated with two fluorophores that are FRET partners, such as CFP and YFP, and the activation of this protein is associated with a conformational change, it is possible to detect a FRET signal due to the two fluorophores coming close together, thereby characterising the active state of

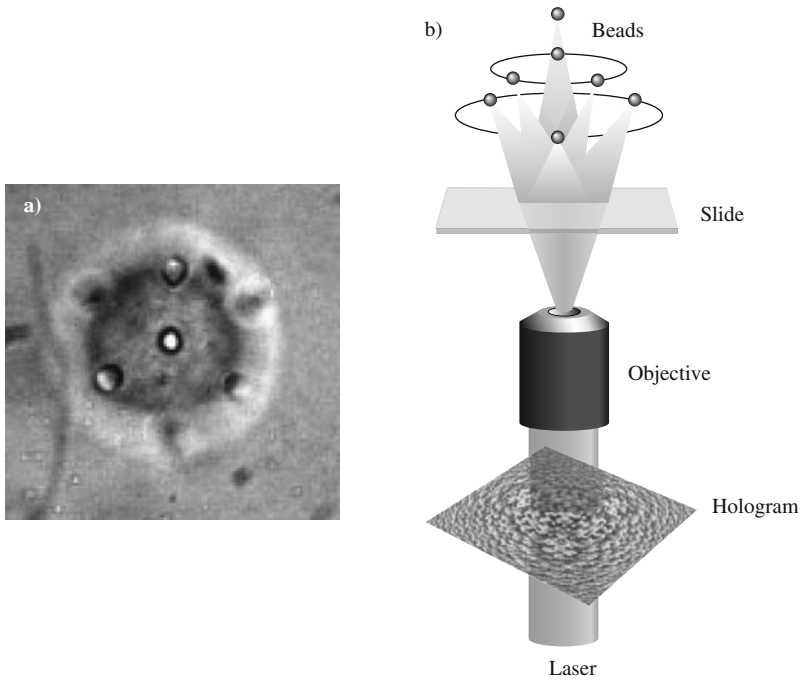


Fig. 8.56. (a) 3D array of microspheres of diameter $2\mu\text{m}$ attached to the cortex of a HeLa cell. (b) Schematic view of the geometry of the array [172]

the molecule. An example is given in Fig. 8.55 [171]. This shows cells on which fibronectin-coated microspheres have been placed. After 15 min, a FRET signal is recorded around the beads subjected to a force, and this unambiguously characterises activation of the enzyme Src.

The recent development of holographic optical tweezers should make it possible to improve the experimental methods described above, thanks to dynamical adjustment of the optical trap arrangement and hence also the distribution of mechanical stimuli applied to the cell. For example, Fig. 8.56 shows a 3D array of beads attached to the surface of a round cell [172].

Manipulation of Whole Cells

It is much more difficult to trap whole cells (diameter $5\text{--}15\mu\text{m}$) than synthetic microspheres. There are several reasons for this:

1. The difference in refractive index between a living cell (on average around 1.35) and the physiological medium (1.33) is small.
2. They are much bigger than the wavelength of the trapping laser. In fact, capture conditions are considered to be optimal when the subject has dimensions of the same order as the wavelength.

3. The cell often has an irregular morphology, which results in an anisotropic force distribution when it is illuminated by a beam with axial symmetry.
4. The cell is made from living material, and this limits the light power needed to compensate for the difficulties raised by the first three points.

Several types of optical micromanipulation have nevertheless been implemented with some success on whole cells.

Optical tweezers have been used as a sorting tool, e.g., to study the regeneration of synapses between a photoreceptor and the cells onto which it projects [173]. To do this, a two-compartment culture system was devised. One of the compartments contained the adhesive cells, and the other, cells in suspension. The cell groups were assembled by using the optical trap to bring the cells in suspension toward the adhesive cells.

Optical tweezers have also been used to impose forces on the scale of whole cells. For example, individually manipulated fibroblasts are allowed to form brief contacts with a glass surface coated with fibronectin. The cell is then pulled by moving the optical trap vertically. This demonstrates the formation of adhesive linkages between the fibronectin and its receptor, and characterises the kinetic constants and the force of this interaction [174].

Other studies on red blood cells and nucleated cells have characterised the viscoelastic properties associated with cell deformation when the cell is stretched between two optical traps [175] and in a system with two antiparallel lasers (optical stretcher) [176].

Finally, individual bacteria can be held in an optical trap to study their adaptation to changes in nutrient concentration, inside confined microchambers [177].

Optical Measurement of Picoforces in Biology

One application for which optical tweezers have proved themselves to be the perfect tool is the measurement of very weak forces between biological molecules. For example, the forces exerted by molecular motors on the polymers of the cytoskeleton have been detected on the level of individual molecules [178–180]. The type of setup used for these experiments is shown in Fig. 8.57. Typically, microbeads coated with a small number of specific molecules, viz., myosin and kinesin, are brought into contact with an actin filament or a microtubule, respectively, sedimented on a glass surface. Once the interaction between the two molecules has occurred, and in the presence of an energy supply, i.e., hydrolysis of adenosine triphosphate (ATP), the motor tends to move along the filament, pulling the microsphere out of the trap. The motor stops when the bead has been shifted a distance Δx far enough from its original position ($x = 0$, the center of the trap) for the restoring force of the trap to exactly balance the maximal force F generated by the motor. This equilibrium position is then given by the relation $F = k\Delta x$, where k is the stiffness constant of the trap, defined in Sect. 8.4.2 and calibrated in

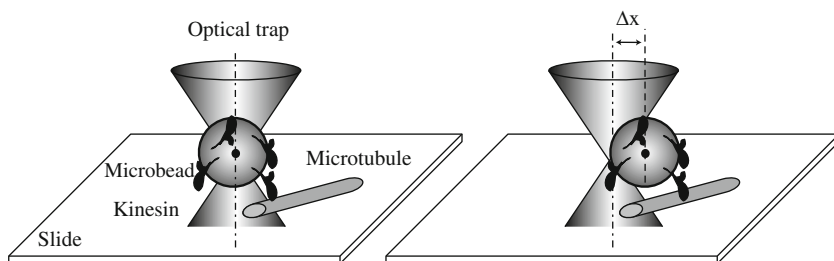


Fig. 8.57. Typical optical tweezer experiment for studying molecular motors. A microbead coated with kinesin is placed in contact with a microtubule, visualised by interference contrast. During its interaction with the microtubule, the kinesin displaces the bead a distance Δx from its initial position

the experiment. A measurement of Δx thus gives the force F generated by the motor, which is not known a priori, and more generally, the force–speed diagram for the motor.

An analogous experiment has been devised to measure the interactions between adhesion molecules. In this case, a substrate is coated with a known density of receptors, and a bead coated with the corresponding ligand is scanned over the surface. Single adhesion events can be detected when the bead moves away from the center of the trap, and the corresponding bond-breaking forces can be estimated. The latter are characterised by the fact that the bead returns suddenly to the trap [180].

8.5 Magnetic Tweezers

8.5.1 General Idea

The micromanipulation of molecules or biological entities requires some way of applying a force in the piconewton (pN) range and then measuring the effects. As described in the other sections of this chapter, there are several ways of applying such a force: using a flexible glass fibre, the cantilever of an atomic force microscope, or an optical trap (optical tweezers). Finally, magnetic forces also provide a way of applying a force using magnetic beads with diameters from a few hundred nanometers to a few micrometers. This kind of bead has long been used by biologists, who know how to treat their surface in such a way as to coat them with proteins or antibodies.

The basic idea of magnetic tweezers is thus very simple. Once the magnetic beads have been attached to the relevant biological entity, one simply brings along a magnet or switches on an electromagnet in order to exert a force on the beads (see Fig. 8.58). By adjusting the distance between magnet and bead (or adjusting the current in the electromagnet), it is easy to modulate the applied force. In addition, the direction or orientation of these forces can be

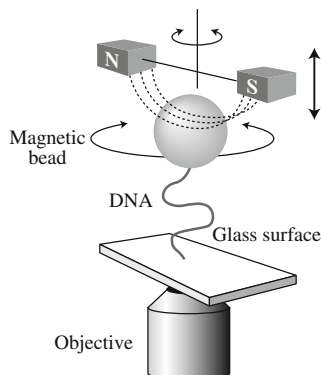


Fig. 8.58. Magnetic tweezers. A magnetic bead is attached to one end of a DNA molecule, while the other end is immobilised on the glass slide holding the sample. Two magnets placed above the sample exert a vertical force that pulls the bead upward. A microscope objective is used to observe the image of the bead from below. The drawing is not to scale. The bead usually measures something like 1 micrometer, the DNA molecule a few micrometers, the magnets a few millimeters, and the air gap separating the magnets about 1 mm

changed. The technical points that need to be controlled include the design of magnets or electromagnets able to apply significant forces, the accurate measurement of the position of the magnetic bead, and the measurement of the applied force.

The magnet design is a technical problem that depends on the desired use. Most often, the possible geometries are limited by various aspects of the observation device, for example, the optical path used for illumination, observation by an immersed objective, etc. Permanent magnets made from samarium–cobalt or neodymium–iron–boron alloys are suitable for high-performance devices. Electromagnets are interesting here because they provide a simple way of modulating the magnetic force rapidly in time. However, they require a cooling system and also a way of measuring the magnetic field directly, because the pole pieces make the relationship between the magnetic field and the control current hysteretic.

There can be no doubt that magnetic tweezers provide the simplest experimental means for applying a force to a single molecule. On the other hand, as for the other micromanipulation systems, measurement of the force that is actually applied is a more delicate matter. Indeed, the force varies non-linearly with the distance between the magnets and the magnetic bead, and this necessitates a good calibration scheme. Fortunately, the Brownian motion observed in all very small objects provides an ideal way of doing this, which we shall now outline.

Micromanipulation devices can be used in two different ways: some impose the position of the subject while others impose the applied force. Micropipettes, AFM sensors, or optical tweezers belong to the first category.

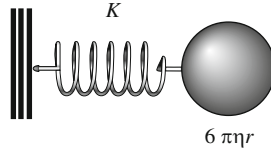


Fig. 8.59. Modelling a force sensor by a spring associated with a dissipative element, represented here by a bead. The sensor is held by a spring of stiffness K . It is immersed in a medium (usually water) exerting viscous friction proportional to its instantaneous speed. For a bead of radius r , the coefficient relating the viscous force to the speed is given by $\mu = 6\pi\eta r$, where η is the fluid viscosity, e.g., for water $\eta \approx 10^{-3} \text{ N s m}^{-2}$

They can position objects very accurately in space. The applied force is then measured by evaluating the deflection of the micropipette or the AFM cantilever, or the shift in position of the bead relative to the center of the optical trap. For magnetic tweezers made with permanent magnets, one now works with a constant force and observes the position adopted by the bead under this effect. Depending on the use, one operating mode is often better suited than the other, and this then the criterion for choosing the appropriate technique. Curiously enough, most micromanipulation applications require a constant force rather than a constant position, and very often those who use optical tweezers servo-control the position of their trap to obtain a system producing a constant force. Since this approach is rather complex experimentally speaking, the use of magnetic tweezers provides a much simpler way of achieving this goal. By replacing the permanent magnets by electromagnets, the magnetic tweezer system can be transformed into a micromanipulation device imposing the position of the subject, as we shall describe at the end of this chapter.

8.5.2 A Mechanical Model for a Force Sensor: A Bead Attached to a Spring

Almost all force sensors used in micromanipulation comprise a marker associated with an elastic element, viz., a spring. In AFM, for example, the cantilever stands in for both elements at the same time. For optical tweezers, the marker is a bead, while the role of the spring is played by the restoring force toward the center of the optical trap. In the case of magnetic tweezers, the marker is the magnetic bead, while the spring is the molecule attached to the bead. In a simplified approach, these sensors can be modelled by a bead attached to a spring, as shown in Fig. 8.59.

If the position of the bead is denoted by x , we can write down a differential equation for the motion of the bead. The latter is immersed in water which, on the micrometric scale, is extremely viscous. The forces acting on the bead are therefore the force $F_{\text{spring}} = Kx$ exerted by the spring and the viscous force proportional to the speed of displacement of the bead. This viscous force

imposes a limiting speed V_{lim} on a free bead on which a constant force F is applied. The limiting speed is given by $V_{\text{lim}} = F/6\pi\eta r$. On the microscopic scale, the effects of viscosity are so considerable that the bead's own inertia is often negligible. As an example, if one suddenly cuts off the force dragging a bead of diameter $1\text{ }\mu\text{m}$, the bead will come to a halt in only one microsecond! To simplify our model, we thus neglect inertial terms in the equation of motion, which therefore becomes

$$6\pi\eta r \frac{dx}{dt} + Kx = F. \quad (8.33)$$

This equation is easy to solve if we choose a force F varying sinusoidally in time:

$$X(\omega) = \frac{F(\omega)}{K \left(1 - i \frac{\omega}{\omega_0}\right)}, \quad (8.34)$$

where

$$\omega_0 = 2\pi f_0 = \frac{K}{6\pi\eta r} = \frac{1}{\tau}.$$

This type of response is Lorentzian, characterised by a uniform response for low frequencies $f < f_0$ and a response which decreases with frequency when the latter is greater than the frequency f_0 . The frequency f_0 is a characteristic of the sensor, defining its pass band.

Brownian motion can be modelled by a random force F_L whose amplitude is independent of the frequency (up to very high frequencies). Since $F_L(\omega)$ is a random variable with zero mean and constant variance, the spontaneous fluctuations in the position of the sensor (which define the ultimate limit to its sensitivity) have a Lorentzian frequency spectrum as response function, a consequence of the fluctuation–dissipation theorem. It remains to define the amplitude of $F_L(\omega)$. To do this, we may use the energy equipartition theorem [181]. Since our sensor is bathed in an environment at temperature T , each of its degrees of freedom has an average energy of $k_B T/2$, where k_B is Boltzmann's constant with $k_B = R/N_a$, R is the perfect gas constant, and N_a is Avogadro's number. For our sensor, the equipartition theorem gives

$$\frac{1}{2} K x^2 = \frac{1}{2} k_B T. \quad (8.35)$$

This equation can also be expressed in the frequency space, whereupon $F_L(\omega)$ may be calculated. We thereby obtain

$$F_L^2 = 24\pi\eta r k_B T. \quad (8.36)$$

This relation may seem surprising, because it does not involve the stiffness K of the sensor, as might be expected by considering the equipartition theorem. In fact, it is indeed the right relation for evaluating the sensor noise, because

the fluctuation–dissipation theorem says basically that the fluctuations arise from the dissipative terms in the equation of motion. The only dissipative term here is the one relating to the viscosity, appearing in the expression $6\pi\eta r$. The spring is not dissipative and does not appear directly in the formula for the noise.

The stiffness K of the sensor nevertheless has a role to play, namely, determining the pass band. In addition, we have just mentioned the force fluctuations of the sensor, but we are often interested in the position fluctuations which for their part do involve K .

8.5.3 Measuring the Bead Position with Nanometric Resolution

In magnetic tweezers, the magnets apply a force to the magnetic bead, whose displacements must be detected with great accuracy. Although the optical microscope has resolution limited to a fraction of the wavelength (optical resolution of the microscope), this limitation does not apply to the displacements of an object. Indeed, the diffraction induced by the objective imposes a theoretical limit on the separation of two very close objects. However, an object can itself be positioned with a precision that is only limited by the value of the intensity emitted by this object. With a simple video camera connected to an optical microscope, the displacements of a micrometric bead can be measured with nanometric accuracy. This operation is intuitively obvious in the observation plane, but less so in the third direction z along which the light propagates.

Tracking a Bead in the Observation Plane

The image of the bead obtained with a microscope and camera is a 2D array of pixels, each of which represents a certain level on the grey scale over a region equivalent to 100×100 nm. Measuring the bead position in the observation plane amounts to determining the position of the center of the image of the bead obtained by video. There are several ways to do this. The simplest is just to calculate the barycenter of the pixels in the image of the bead, for which one must be able to discriminate a boundary defining the interior and exterior of the image. This is done by thresholding the grey levels, a rather easy method to put into practice, but often with somewhat unsatisfactory results because it is very sensitive to the light conditions affecting the bead (this is the thresholding problem). We prefer a method that is more sophisticated in terms of computation, but which proves to be more robust. Two consecutive images of the same bead differ by a slight shift in position which can be evaluated by calculating the maximum of the correlation product for the two images [182]. This method has several advantages: it requires no hypotheses concerning the appearance of the observed object, which can be dark, bright, or something more complex, and it is largely insensitive to lighting fluctuations.

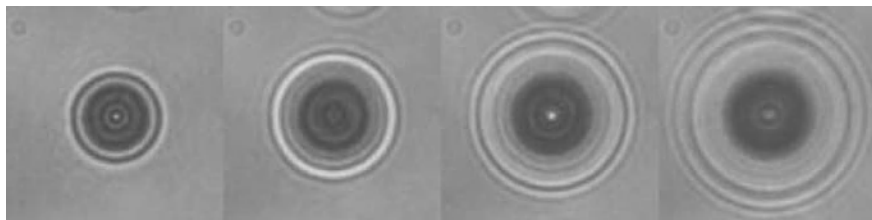


Fig. 8.60. Images of a fixed bead observed with parallel illumination for four heights differing by $1\text{ }\mu\text{m}$ between each image. The shape of these rings is used by the program to determine the height of the bead

In practice, we do not calculate the correlation product of two images, because this would require too much computation time. We average over about ten rows and ten columns centered on the previous position of the bead and we calculate the 1D correlation product of these profiles, averaged with their mirror image obtained by reversing the order of the points in these profiles. The correlation functions are calculated using the fast Fourier transform (FFT), which speeds up the calculation. We thus obtain twice the shift of the bead in x and in y from the position obtained in the previous image. The correlation function is interpolated between the pixels by approximating the profile by a quadratic polynomial whose maximum is calculated. By using the FFT, the average profiles can be filtered to avoid problems of numerical moiré, which is the main source of error in this interpolation [183]. The whole calculation takes only $100\text{ }\mu\text{s}$ on a 2-GHz processor. It is thus possible to track the bead in real time for most cameras.

Tracking a Bead in the Direction Normal to the Observation Plane

It is a slightly more delicate matter, but nevertheless interesting, to measure the position of the bead in the third direction. We have all had the experience of turning the knob on a projector to focus an image. During this operation, we know whether we are getting closer to or moving away from the focal point, but when the image is fuzzy, it is difficult to say in which direction we should turn the knob. To measure the vertical position of the bead, we are faced with the same problem. If we slightly defocus the microscope, we can then say when the bead approaches or moves away from the focal point, but the defocusing must always have the same sign. For a better appreciation of this effect, we have observed that illumination parallel to the sample causes diffraction rings to appear, which decorate the image of the bead [184] and whose apparent diameter grows when the bead moves away from the focal point, as shown in Fig. 8.60. These rings are clearly visible when the distance separating the bead from the focal point does not exceed $10\text{--}15\text{ }\mu\text{m}$.

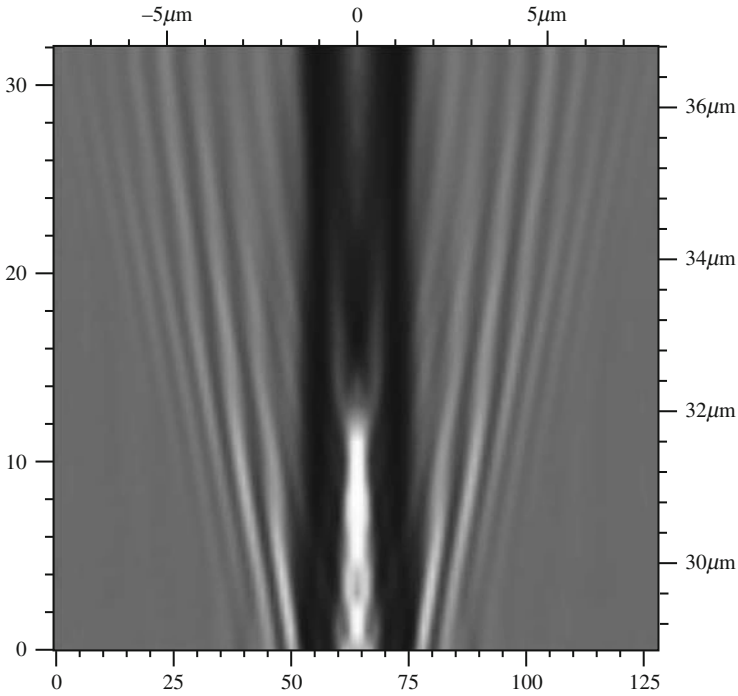


Fig. 8.61. Calibration image for a bead

When tracking the position of the bead center in the observation plane, we can then calculate the radial profile of the image of the bead which reflects these rings. In Fig. 8.61, we have plotted these radial profiles as a function of the distance from the bead to the focal point (situated at the bottom of the image). From this image, the process just described can be reversed. If we take the radial profile of the bead at an unknown height, it can be compared with this image and one can say how high the bead was. The radial profile is compared with this calibration image using a least squares method, whereupon one can say which profile of the calibration image most closely resembles the observed profile. One can then interpolate between the profiles of the calibration image and obtain the position of the bead to an accuracy of a few nanometers.

More processor time is required to calculate the vertical position of the bead than to track the horizontal position. The longest stage is the one required to obtain the average radial profile. Indeed, for each of the 10^4 pixels surrounding the bead, one must calculate its distance from the center of the bead, and this involves the calculation of 10^4 square roots. This requires about $500\mu\text{s}$ on a 2-GHz processor (using instructions to carry out operations in parallel). However, this is still good enough to be able to track several beads in real time using a video camera.

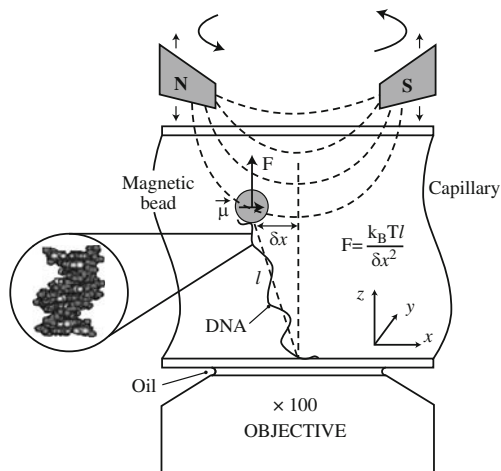


Fig. 8.62. Measuring forces by monitoring Brownian motion. The bead attached to a glass slide by a DNA molecule behaves like an upside-down pendulum. Brownian motion perturbs the pendulum from its equilibrium position. The bigger the magnetic force, the more rigid the pendulum and the smaller its fluctuations. The equipartition theorem can be used to relate these fluctuations to the stiffness of the pendulum. By inverting this relation, a formula is obtained for determining the force applied to the bead from its mean square fluctuations, the length of the molecule, and the temperature

8.5.4 Calibrating the Force Measurement by Brownian Motion

The magnets apply a constant and reproducible force on the bead. Indeed, the field gradient responsible for this force varies on a millimetric length scale (the air gap). On the scale of the DNA molecule, viz., a few micrometers, this gradient is constant. The force on the bead is given by the product of this gradient (which depends only on the magnets) and the magnetisation of the bead, which varies by 30% from bead to bead, within a given type of bead. We have devised a measurement of the force acting on the bead which does not depend on the magnetic properties of the magnets and bead. The system comprising the bead attached by a molecule of length l and on which a vertical force F is acting is equivalent to an upside-down pendulum, permanently disturbed by the impacts of water molecules against the bead (see Fig. 8.62). The motions of the bead in the (x, y) plane normal to the force F are subjected to a force that can be treated as a spring of stiffness $K = F/l$. In order to determine the force, one measures the horizontal Brownian fluctuations $\langle \delta x^2 \rangle$ and the extension l of the molecule it is attached to [185]. These quantities are related to F by the equipartition theorem, which gives

$$F = \frac{k_B T l}{\langle \delta x^2 \rangle},$$

where T is the absolute temperature.

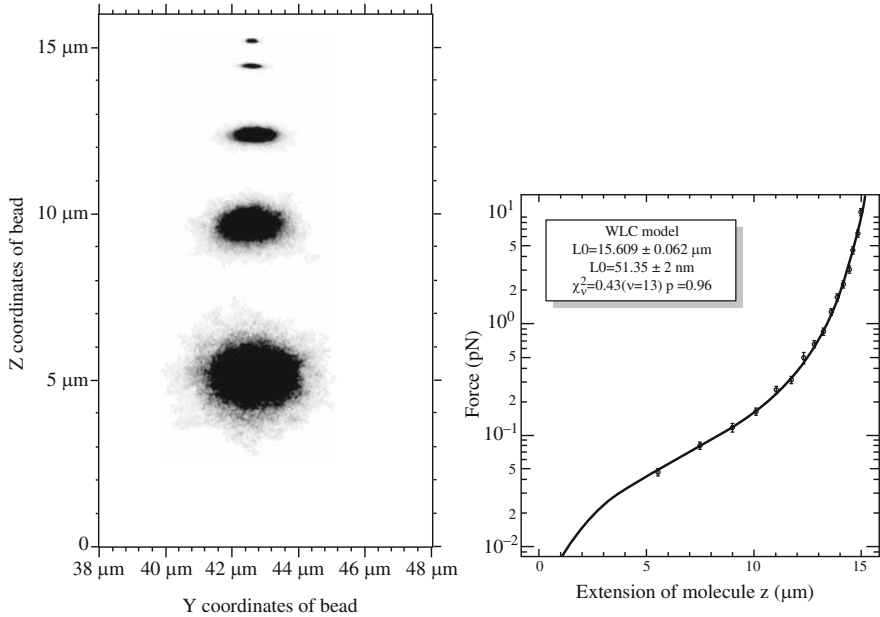


Fig. 8.63. Force measurement made by analysing the Brownian motion of the bead. *Left:* Clouds of measurements representing the position of a bead attached to a molecule $15.6\ \mu\text{m}$ long and subjected to different forces. For small forces, fluctuations are significant and almost isotropic, and the molecule is very slightly stretched. For bigger forces, fluctuations are much less significant and anisotropic, and the molecule is considerably extended. By applying the method described in the last figure, the force actually applied can be deduced. *Right:* Force–extension curve for a DNA molecule, compared with the wormlike chain model, which describes the elasticity of a polymer [187, 188]

To carry out this measurement, the magnets are positioned and then the horizontal fluctuations of the bead are recorded over a few minutes and the extension l of the molecule is determined by tracking the vertical position of the bead. This operation is repeated by varying the distance of the magnets for five or six points, chosen to produce a significant variation in the extension of the molecule. We thus obtain a force curve for the relevant DNA molecule (see Fig. 8.63) [185, 186]. This curve is very well described by the wormlike chain model [187, 188]. To begin with, one can find out whether there really is just one molecule attached to the bead, because two molecules would be twice as difficult to stretch as one. Furthermore, one can establish a correspondence between the position of the magnets and the force exerted on the molecule for the given bead. After this calibration, which takes about half an hour, the time has come to measure the forces directly. The traction force is chosen by

imposing the position of the magnets according to the correspondence just established.

In practice, the mean squared fluctuations $\langle \delta x^2 \rangle$ are measured in the spectral domain, whence it is possible to check its validity and eliminate low-frequency drift due to temperature drift, which leads to overestimation of $\langle \delta x^2 \rangle$.

The spectrum of position fluctuations δx has the Lorentzian shape already described, characterised by a cutoff frequency f_0 . However, the sampling frequency f_{camera} restricts frequency measurements to values below $f_{\text{camera}}/2$, and this establishes an upper bound for measurable forces. In addition, measurement of f_0 provides a way of checking that the viscous term has the value expected. Finally, measurement in the Fourier space allows one to evaluate the error in the measurement of $\langle \delta x^2 \rangle$.

8.5.5 Magnets Used for Magnetic Tweezers

The magnets are of course a key element of the magnetic tweezer system. In general, one tries to obtain the largest force possible, which means the highest magnetic field gradient possible. Since the beads themselves are paramagnetic, their magnetisation increases with the magnetic field, and this must also be maximised. Magnets made from samarium–cobalt and neodymium–iron–boron alloys are the best for this application. Moreover, geometric constraints are determined by the presence of the microscope objective and the fact that light arrives vertically from above. For this reason, the magnets are arranged in such a way that the magnetic field is horizontal, produced by two magnets separated by a thin air gap allowing light to pass through. In the air gap, the field reaches about 1 tesla, then falls off rapidly in the z direction in moving away from the magnets. It is easy to measure the function $B(z)$ giving the magnetic field as a function of the distance from the magnets. This function is rather close to an exponential. To a first approximation, one has

$$B(z) = B_0 \exp(-z/z_0),$$

where z_0 is given by the size of the air gap. The magnetisation of the bead obeys a Langevin law, i.e., it is proportional to the magnetic field, so that $M = \mu H$, when it is small, and saturates at M_s beyond a characteristic field H_0 . The quantities μ and H_0 depend on the beads used. Typically, H_0 is about 0.1 T. The magnetic force acting on the bead is proportional to the product of the field gradient and the magnetisation. By differentiating the function $B(z)$ and using the magnetisation curve provided by the bead manufacturer, the theoretical value of the applied force can be obtained. This value is in good agreement with the experimental value measured by analysing the Brownian motion of the bead, as described above. Any discrepancies arise due to variations in the size (and hence magnetisation) of the beads and the error in the measurement of the derivative of the magnetic field produced by the magnets. These errors are of the order of 20–30%.

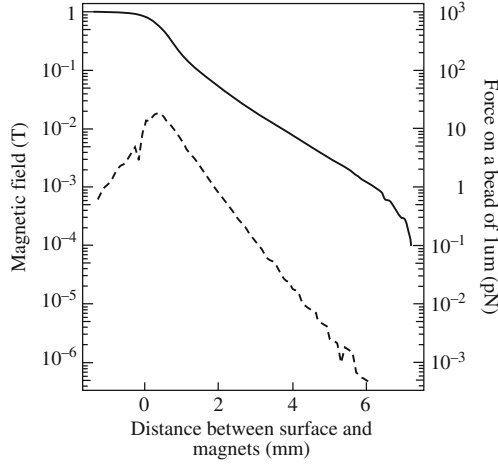


Fig. 8.64. Magnetic field (*continuous curve*) and force exerted on a bead of diameter $1\text{ }\mu\text{m}$ (*dashed curve*) by a pair of permanent NeFeB magnets ($5 \times 5 \times 6\text{ mm}$) separated by an air gap of 0.6 mm

With the simplifying assumption $B(z) = B_0 \exp(-z/z_0)$, the force is given by

$$F(z) = 2\mu \frac{B_0^2}{z_0} \exp(-2z/z_0),$$

in the linear magnetisation regime, while for strong fields,

$$F(z) = M_s \frac{B_0}{z_0} \exp(-z/z_0).$$

The applied force varies rapidly with the distance z . The size z_0 of the air gap gives the characteristic length for the z -dependence of the force. To increase the force, z_0 must be reduced. However, it cannot be reduced too much because there must be enough space for light to pass through, and in addition, in order to profit from the increased force, it must be possible to go very close to the magnets on the scale of the air gap. A compromise must therefore be found. In practice, z_0 is a fraction of a millimeter. This parameter also determines the distance over which the force can be treated as constant. If the force is required to vary by less than 5% when the bead moves through $10\text{ }\mu\text{m}$, then z_0 must be greater than $200\text{ }\mu\text{m}$.

With an air gap of 0.6 mm (see Fig. 8.64), the maximal forces vary significantly with the diameter of the bead, in fact, as the cube of the bead diameter. For Dynal beads measuring $1\text{ }\mu\text{m}$ (MyOne), this maximal force is 10 pN , but it exceeds 100 pN for Dynal beads of diameter $2.8\text{ }\mu\text{m}$ (M280) and approaches nN for beads of $4.5\text{ }\mu\text{m}$ (M450). On the other hand, it is very easy to obtain small amplitude forces since the force decreases exponentially with distance,

the only limitation being the weight of the beads (0.3 pN for beads of 4.5 μm , and a few tens of femtonewtons for the M280 and MyOne beads).

8.5.6 Advantages of Magnetic Tweezers

Twisting a Molecule

One of the advantages of magnetic tweezers is that they offer the possibility of rotating the bead on its own axis. As the magnetic field is horizontal and its gradient vertical, the simple fact of rotating the magnets about the vertical axis will maintain a constant vertical field gradient while rotating the magnetic field itself. The magnetic susceptibility of the paramagnetic beads is anisotropic and hence follows the direction of the field, just like a compass. As soon as there is an angle between the bead magnetisation and the magnetic field direction, a strong magnetic torque tends to close this angle. The order of magnitude of the torque is given by the applied force and the radius of the bead, i.e., a few pN per micrometer. The torques exerted by biological molecules are of the order of a few pN per nm, i.e., a thousand times weaker. They are therefore negligible in practice. Magnetic tweezers therefore impose the angular direction of the bead, by adapting the necessary magnetic torque, whereas they work at constant force. For this reason, magnetic tweezers can control the angular position of the beads on which they act. However, it is difficult to measure the torque they produce directly.

If we attach a DNA molecule by several points at each end, rotation of the bead controls the torsion of the DNA molecule [185, 189]. This torsional stress only exists if the molecule is fixed to the surface at several points, but especially if there is no ‘nick’, i.e., no break in one of the strands of the DNA double helix. Indeed, this type of defect provides a way of relaxing all the torsional stress, since the molecule can then freely rotate about the simple chemical bonds of the intact strand.

Although the possible role played by the traction force on an *in vivo* DNA molecule has not yet been clarified, the torsional stress for its part is a fundamental biological parameter. To characterise it, one compares the number of turns n added or removed from the molecule with the number of turns Lk_0 exhibited by the double helix in the absence of the stress, i.e., one turn for every 10.5 base pairs. The ratio $\sigma = n/Lk_0$ of these two numbers provides a way of comparing molecules of different sizes.

The effect of torsion on a DNA molecule has been carefully studied by White [190]. It is easy to understand by analogy with a rubber tube. If the tube is steadily twisted, holding the traction force constant, it will oppose this with a torque proportional to the twist angle. If the traction force is now released but holding the twist angle constant, an instability will soon appear in the form of little loops, well known to all of us who remember the twisted wires of telephone receivers. When they are forming, the torque is significantly reduced. These loops absorb a large part of the stress. In fact, this

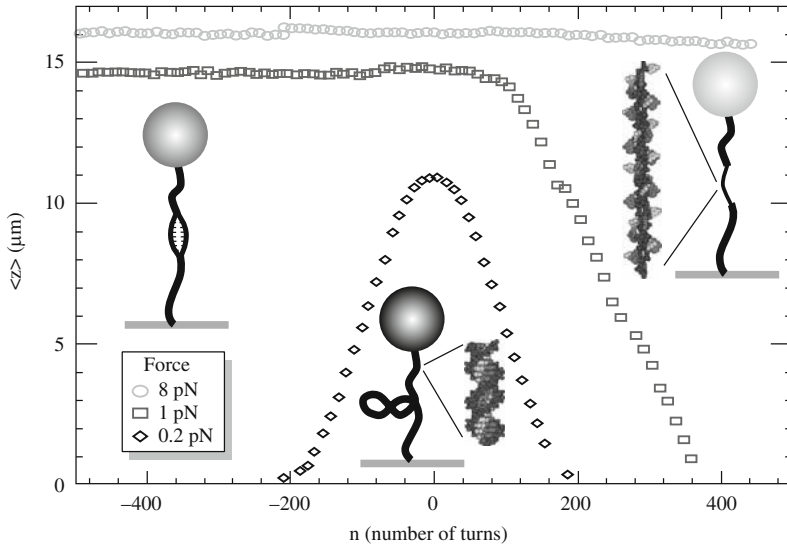


Fig. 8.65. Extension curves for a DNA molecule as a function of the number of turns applied at constant force. For weak forces, the DNA molecule forms loops, similar to those that sometimes appear in the wire of a telephone receiver. These shorten the molecule as soon as it is twisted in one sense or another (*curve with diamonds*). For intermediate forces, the same phenomenon occurs but only when further turns are added. For $n < 0$, the torque is sufficient to denature the bases (*curve with rectangles*) [191]. For strong forces, a new phase of DNA called P-DNA appears when n is positive [193]

instability results from competition between two mechanisms, whose energy contributions come from the pure twist and the curvature. For low stresses, the tube minimises its curvature energy by staying straight, but for high stresses, the torsion energy grows too big, and it becomes profitable to pay the energy required to curve the tube in order to form loops which reduce the torsion. The behaviour of the DNA molecule is very similar. However, the effects of thermal fluctuations must be included. These soften the appearance of loop or plectoneme instabilities. Plectonemes shorten the molecule. So if we apply a weak force to a DNA molecule and if we plot its length as a function of the number of turns n applied to it, we observe a bell-shaped curve. The extension is maximal for zero stress, then falls off linearly when n is increased, and this whatever its sign. This behaviour is shown by the symmetric curve in Fig. 8.65, which agrees extremely well with theoretical predictions [12]. In addition, it can be used to monitor very closely the changes induced in the torsion of the molecule by external elements (enzymes).

When this measurement is repeated with a stronger traction force (1 pN), we observe a non-symmetric behaviour with respect to σ . For $\sigma > 0$, the molecule shortens as before, but for $\sigma < 0$, this length no longer seems to



Fig. 8.66. Crossing over of two molecules induced by rotating the bead. In some cases, the bead can be attached to the slide by two molecules. The latter are separated by a variable distance e , comparable on average to the bead radius. A rotation of one turn applied to the bead causes a single cross-over of the two molecules. This state is easily observed by the resulting reduction in height

depend on σ . This asymmetry relative to the sign of the twist is predictable since the DNA double helix is chiral. The transition we have described corresponds to the appearance of a ‘denaturation bubble’. By unwinding the double helix, we cause the appearance of a small region in which the hydrogen bonds holding together the complementary bases of the two strands give way under the stress (see Fig. 8.65) [192]. The same phenomenon is observed when we unwind a two-stranded rope. For strong forces, a new phase of DNA appears when n is positive, known as P-DNA [193].

Using Magnetic Tweezers to Determine the Presence of Nicks or to Cross Two Molecules at a Single Point

The possibility of rotating the bead on its own axis provides a quick way of selecting beads in terms of the torsional properties of the molecule anchoring them to the glass substrate. One characteristic feature of magnetic tweezers is their parallelism, i.e., they apply a parallel traction force on many beads at the same time, given that the magnets are several millimeters wide. By displacing the sample across the field of view of the microscope, it is easy to select the molecule on which one would like to work among several hundred possibilities. This is very convenient because neither the number nor the type of molecules attached to the bead can be controlled. For example, in some situations, one seeks beads attached by a single DNA molecule with at least one nick [194]. In order to identify these, it suffices to apply a force of about 1 pN (which significantly stretches the molecule) and rotate the magnets clockwise. Only those beads attached by a single DNA molecule with a nick do not change length with rotation of the magnets. Molecules without a nick form plectonemes which shorten the molecule as the magnets are rotated. Beads attached by several molecules also shorten when the molecules coil around one another. The situation for two molecules is depicted in Fig. 8.66. Here the system behaves like a swing. The observed extension is maximal when the two molecules are parallel and shortens suddenly as soon as the bead rotates through a half-turn in either sense [195]. This type of test takes a few seconds

to carry out. It can thus be applied to a large number of beads in order to choose the most suitable one.

8.5.7 Examples of Studies Using Magnetic Tweezers

Magnetic tweezers can be used to stretch a DNA molecule and measure its extension with a resolution of a few nanometers. Now many enzymes act on DNA and produce a change in length of this molecule. Some directly cause a change in the structure of the molecule, such as RecA which covers the DNA and changes both its overall length and the length of each turn [196]. But in many cases, the sensitivity of the experiment is good enough to measure the action of a single enzyme. Here we shall outline a few examples and invite the reader to consult the literature for more detail.

Revealing the DNA Loop Formed by GalR

The first example we have chosen concerns the galactosidase repressor (GalR). This enzyme is the analogue of the lactose repressor (LacR) studied by Jacob and Monod when they investigated the lactose operon. GalR inhibits the synthesis of galactosidase (the enzyme which digests galactose) by preventing the RNA polymerase from reaching the corresponding promoter gene. To do this, GalR forms a DNA loop containing the promoter by binding to two specific sequences located on either side of it. In the presence of galactose, GalR detaches, the promoter becomes accessible, and the galactose is degraded.

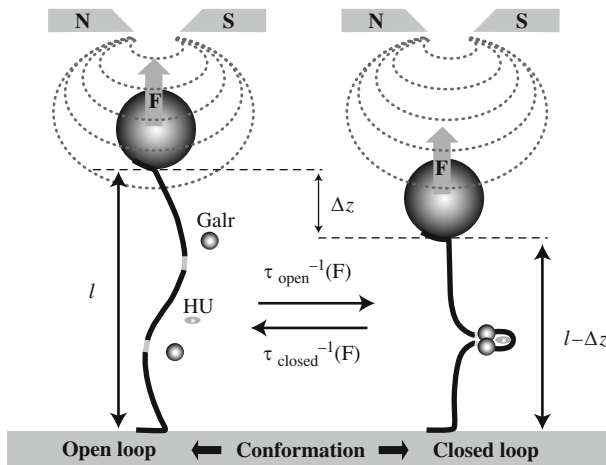


Fig. 8.67. Activity of the galactosidase repressor GalR. A DNA molecule containing two sequences able to bind to GalR is subjected to a traction force of 1 pN using a magnetic bead. When GalR is present in the solution, with the help of Hu, it forms a loop that can be detected by the consequent shortening of the DNA molecule

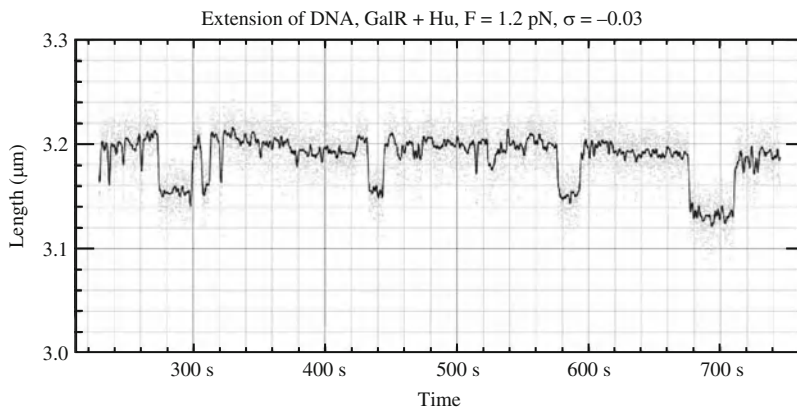


Fig. 8.68. Telegraph signal illustrating the attachment and detachment of GalR as time goes by. The stronger the traction force, the longer the loop remains open

In order to visualise this mechanism, we prepared a DNA molecule containing the two specific sequences separated by 119 bp and attached it between a magnetic bead and a glass slide (see Fig. 8.67) [197]. For weak forces, adding the enzyme GalR to the solution together with its cofactor Hu, we observed a shortening of the molecule, i.e., a loop had formed. Obviously, if we pull hard on the molecule, the loop will open out completely. There is an intermediate force range in which the applied force just balances the GalR anchoring force. In this situation, the length of the molecule fluctuates significantly, going from the open loop state to the closed loop state in a random way over periods of a few seconds, as can be seen in Fig. 8.68.

Observing the Separation of Two DNA Strands by the Helicase UvrD

The structure of DNA, in which genetic information is encoded by sequences of bases within a double helix, constitutes both an advantage and a disadvantage as far as this information is concerned. The advantage is that the bases are protected from possible degradation, and the disadvantage is that the information they carry is not so easily accessible to the agents essential to the life of the cell. Many cell processes therefore require the double helix to be opened up temporarily, in order to be able to read, copy, translate, or repair the DNA bases. It is the helicases that carry out this crucial operation within the cell. These enzymes insert themselves into the double helix, then move along a strand cutting the hydrogen bonds that hold the two strands together. These molecular motors have been the subject of an intense effort to understand the relevant mechanisms. Indeed, these enzymes, or more precisely, their failure to function correctly, lie at the heart of many serious genetic disorders, leading for example to premature ageing of the patient.

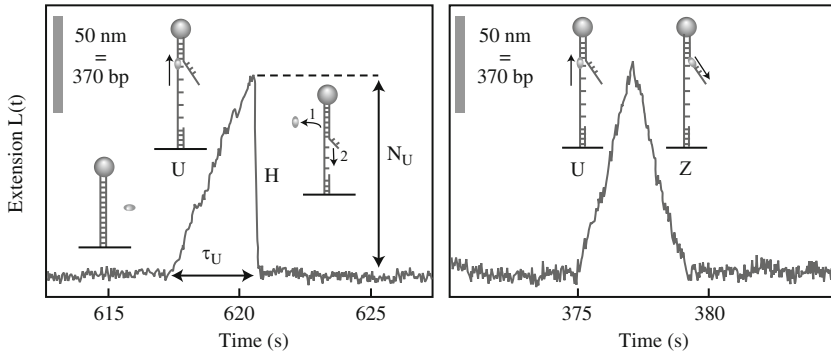


Fig. 8.69. Action of the helicase UvrD on a DNA molecule with a single-strand break. Time-dependence of the extension of a molecule subjected to a constant force of 30 pN. When the enzyme is working, the molecule extends at a steady rate because single-strand DNA is longer than double-strand for this force. The action of the helicase often stops suddenly, causing a rapid reduction in length corresponding to the reformation of double-strand DNA (*left*). A gradual return to the original length of the molecule is often observed (*right*). This unexpected behaviour is due to the fact that the helicase switches from one strand of the DNA to the other during its action. This type of phenomenon cannot be detected in conventional experiments carried out with helicases. The helicase concentration is 0.5 nM, while that of ATP is 0.5 mM. The solution also contains Tris 25 mM, MgCl_2 3 mM, and DTT 1 mM

The task of the helicases can be seen as the transformation of double-stranded DNA into two simple strands. To detect the action of a helicase, we apply a force of 30 pN to a double-stranded DNA molecule with a single-strand break required by the enzyme to begin its work [194]. At low enzyme concentrations, we observe phases in the activity of a helicase by an extension of the molecule as the DNA strands are gradually separated (single-stranded DNA being longer than double-stranded for this force). As can be seen from Fig. 8.69, the enzyme proceeds rather steadily. We can also measure the rate of action in real time, the number of unlinked bases, and many kinetic parameters that are difficult to monitor in test tube experiments. These experiments are made possible because helicase is processive, i.e., it carries out a large number of reaction cycles before detaching itself from the DNA, each cycle giving rise to one step of the protein by a few base pairs along the substrate. Here, it is not possible to distinguish each step, but the step size can be evaluated from the observed rate fluctuations.

Unknotting of the DNA Molecule by Topoisomerase

Each of our cells must confine within its nucleus, which measures at most ten micrometers across, a complete set of chromosomes, each of which measures some ten centimeters long when fully extended. The extreme density of DNA

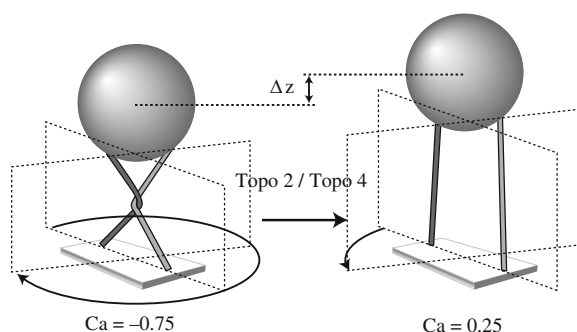


Fig. 8.70. Action of topoisomerase on the cross-over of two DNA molecules. By rotating the bead through one turn, we go from a situation in which the molecules do not cross over (*right*) to one in which they do (*left*). The topoisomerase then binds to the cross-over point and passes one molecule through the other, thereby unknotting the cross-over (*right*). At this point, the enzyme can no longer act. The change in height between the two configurations is used to determine the instant when the enzyme acts

is a major challenge to the cell. In particular, during replication of these long molecules, a tangle of knots is almost inevitable. However, these are quickly undone by astonishing enzymes in the topoisomerase family, which, as their name suggests, are able to modify the topology of the DNA molecule. This family is divided into two groups, the type I topoisomerases, which are able to temporarily cut a single strand of the DNA, and the type II topoisomerases, which cut both strands. By opening up a break in a single strand of the double helix, the type I topoisomerases allow the molecule to rotate about the other strand and hence release torsional stress. For their part the type II topoisomerases bind to the junction point of two DNA molecules, cut both strands of one of them, and pass the intact molecule through the break, while holding on to the two ends which they then proceed to stick perfectly back together again after the translocation. This operation, which undoes knots and untangles the molecules, requires the presence of an energy cofactor, i.e., ATP. Topoisomerase II plays a key role in the cell cycle. Its inhibitors are often used as anticancer agents. They make it impossible to untangle daughter chromosomes during replication, causing the suicide of the cell. The cytotoxic effect is aimed primarily at tumour cells, because they frequently undergo division.

We are still a long way from a complete understanding of the molecular mechanisms brought into play during the catalytic cycle of the topoisomerases. Experiments used to characterise them are always delicate, because it is difficult to alter the torsion of a molecule in a test tube.

When two molecules are crossed over by rotating the bead in Fig. 8.70 through one turn, the molecular configuration recognised by the topoisomerases is created [195]. If the solution contains ATP (the energy supply

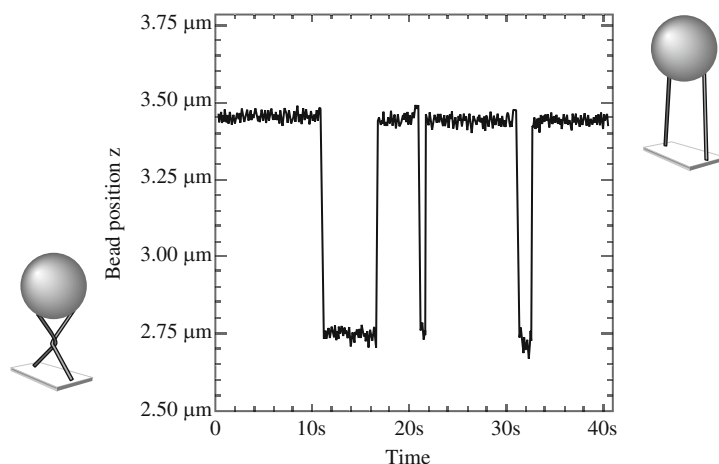


Fig. 8.71. Experimental signal showing the action of a topoisomerase. When the extension of the system reaches $3.45\text{ }\mu\text{m}$, the two molecules are parallel. When they cross over, the distance reduces to $2.75\text{ }\mu\text{m}$. If there are topoisomerases in the solution, no change in length is observed when the molecules are parallel. However, each time the magnets are rotated through one turn, the bead first comes closer to the wall, but when the enzyme undoes the cross-over, the extension suddenly increases again. Here the operation has been repeated three times, and each time, the enzyme has acted. However, the time required to do so seems to vary significantly

needed by most enzyme operations), an enzyme can attach itself to the cross-over and undo it, whence the extension can return to its initial value. The bead can then be rotated through another turn, generating another cross-over that the enzyme will rush to undo, and so on. Each event corresponds to a single enzyme cycle, which is easy to detect since the length change is a fraction of a micrometer, as can be seen from Fig. 8.71.

This type of experiment is able to detect enzyme cycles one by one, and can thereby analyse the way the enzyme operates as time goes by. Furthermore, it can also measure the crossing angle of the molecules, a parameter that conventional enzyme experiments are unable to control.

Magnetic tweezers also provide a way of studying the elasticity of single-strand DNA molecules [198], RNA [199], chromatin [200], and double-strand DNA coated with different enzymes [201]. One can also study the pairing energy of the two strands in the double helix [202]. Moreover, the activities of many enzymes have been studied on the single-molecule level by means of magnetic tweezers. One could mention the DNA polymerases [203], the RNA polymerases [204], type I topoisomerases [205], type II topoisomerases [206], helicases [194], the RuvAB complex [207], condensins [208], EcoR124 [209], FtsK [210], and many others.

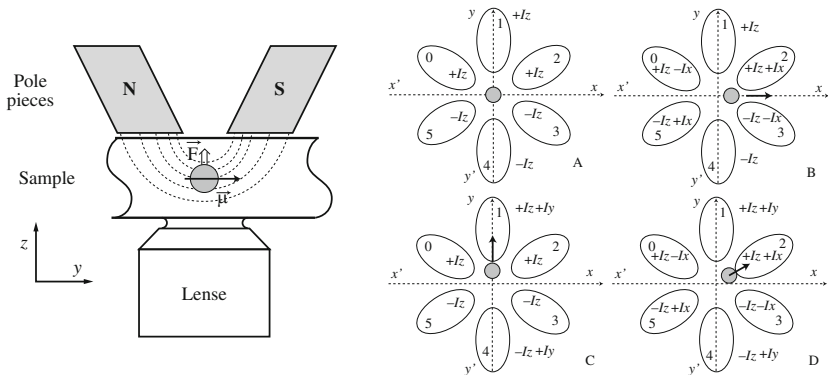


Fig. 8.72. Magnetic tweezer system able to manipulate a bead in the three space directions. This device uses six electromagnets placed above the sample, distributed along the axes of a hexagon. The idea is to create a horizontal principal magnetic field using coils 0, 1, and 2 as north pole and 3, 4, and 5 as south pole. The vertical field gradient is used to pull the bead in the z direction. If the currents in the coils are made asymmetric with respect to the x axis, a horizontal force is produced which drags the bead in the y direction. By applying opposite currents in coils 1 and 4, a force is produced in the y direction. To manipulate the bead, its position must be measured in three dimensions and the currents in the coils must be modulated to displace the bead to the desired position

8.5.8 Manipulating an Object with Magnetic Tweezers

The magnetic tweezer method so far described is used to apply a vertical force and a rotation. Although this device is ideal for probing the elasticity of a DNA molecule, it cannot genuinely manipulate a bead or cell, as optical tweezers are able to do. But it can be arranged with the help of a few modifications that we shall now outline. The result is particularly useful for probing the viscoelastic properties of a cell or its nucleus.

As magnetic systems tend to exert a constant force, the displacement of a bead in the three space directions requires two modifications: it must be possible to modulate the direction of the applied force and it must be possible to control the position of the bead by rapidly adjusting the strength of the force. To do this, it is better to use electromagnets than permanent magnets. In fact, several are required in order to be able to change the direction of the force.

There are several ways of arranging the coils of the electromagnets. Figure 8.72 shows a hexagonal configuration, able to apply a force in the three space directions. In order to achieve a genuine micromanipulation, this system is associated with a position measurement of the bead in the three directions and a feedback circuit, which modulates the currents in the electromagnets in such a way as to hold the bead at a definite position [183]. The six-fold

symmetry of the device means that the bead can be rotated in steps of 60° . The forces accessible to this device (20 pN for Dynal M450 beads) are weaker than can be obtained with permanent magnets.

Many magnetic tweezer setups are hybrid arrangements of the two types we have been discussing here (constant force and constant position), especially those designed to study cells. Indeed, the viscoelasticity of the cytoplasm or the membrane is much greater than that experienced in water. Strong forces must therefore be available to obtain significant displacements which occur at slow rates. Electromagnets with an air gap of a few tens or hundreds of micrometers can produce the forces of more than 100 pN required for this motion. Furthermore, the speed of the beads is low enough for the operator to handle these displacements without servo-control [211].

References

Section One. Molecular and Cellular Imaging Using AFM

1. Binnig, G., et al.: *Phys. Rev. Lett.* **49**, 57–61 (1982)
2. Binnig, G., Quate, C.F., Gerber, C.: *Phys. Rev. Lett.* **56** (9), 930–933 (1986)
3. Fotiadis, D., et al.: *Micron* **33** (4), 385–397 (2002)
4. Hansma, H.G.: *Annu. Rev. Phys. Chem.* **52**, 71–92 (2001)
5. Hansma, H.G., Kasuya, K., Oroudjev, E.: *Current Opinion in Structural Biology* **14** (3), 380–385 (2004)
6. Bustamante, C., et al.: *J. Biol. Chem.* **274** (24), 16665–16668 (1999)
7. Jiao, Y.K., et al.: *J. Mol. Biol.* **314** (2), 233–243 (2001)
8. Abdelhady, H.G., et al.: *Nucleic Acids Res.* **31** (14), 4001–4005 (2003)
9. Virstedt, J., et al.: *J. Struct. Biol.* **148** (1), 66–85 (2004)
10. Mou, J.X., et al.: *FEBS Lett.* **381** (1–2), 161–164 (1996)
11. Sheng, S.T., et al.: *J. Biol. Chem.* **278** (41), 39892–39896 (2003)
12. Shao, Z., Shi, D., Somlyo, A.V.: *Biophys. J.* **78** (2), 950–958 (2000)
13. Vinckier, A., et al.: *Ultramicroscopy* **57** (4), 337–343 (1995)
14. Gale, M., et al.: *Biophys. J.* **68** (5), 2124–2128 (1995)
15. Christiansen, D.L., Huang, E.K., Silver, F.H.: *Matrix Biology* **19** (5), 409–420 (2000)
16. Round, A.N., et al.: *J. Struct. Biol.* **145** (3), 246–253 (2004).
17. Chen, C.H., Hansma, H.G.: *J. Struct. Biol.* **131** (1), 44–55 (2000)
18. Ding, T.T., Harper, J.D.: *Methods Enzymol.* **309**, 510–525 (1999)
19. Moreno-Herrero, F., et al.: *Biophys. J.* **86** (1Pt1), 517–525 (2004)
20. Wagner, K., et al.: *Biophys. J.* **87** (1), 386–395 (2004)
21. Milhiet, P.E., Giocondi, M.C., Le Grimallec, C.:
www.thescientificworld.com **3**, 59–74 (2003)
22. Engel, A., Gaub, H.E., Muller, D.J.: *Curr. Biol.* **9** (4), R133–136 (1999)
23. Milhiet, P.E., et al.: *Single Mol.* **2**, 109–112 (2001)
24. Vie, V., et al.: *J. Membr. Biol.* **180** (3), 195–203 (2001)
25. Saslowsky, D.E., et al.: *J. Biol. Chem.* **277** (30), 26966–26970 (2002)
26. Milhiet, P.E., et al.: *EMBO Reports* **3** (5), 485–490 (2002)
27. Giocondi, M.C., et al.: *J. Struct. Biol.* **131** (1), 38–43 (2000)

28. Le Grimmellec, C., et al.: Scanning Microsc. **9** (2), 401–411 (1995)
29. Scheuring, S., Rigaud, J.L., Sturgis, J.N.: EMBO J. **23** (21), 4127–4133 (2004)
30. Hoh, J.H., et al.: Biophys. J. **65** (1), 149–163 (1993)
31. Fotiadis, D., et al.: Nature **421** (6919), 127–128 (2003)
32. Shahin, V., et al.: Faseb. J. **15** (11), 1895–1901 (2001)
33. Gould, S.A.C., et al.: J. Vac. Sci. Technol. **A8**, 369–373 (1990)
34. Butt, H.J., et al.: J. Struct. Biol. **105** (1–3), 54–61 (1990)
35. Radmacher, M.: IEEE Eng. Med. Biol. Mag. **16** (2), 47–57 (1997)
36. Benoit, M., et al.: Nat. Cell Biol. **2** (6), 313–317 (2000)
37. Stroh, C., et al.: Proc. Natl. Acad. Sci. USA **101** (34), 12503–12507 (2004)
38. Le Grimmellec, C., et al.: Biophys. J. **67** (1), 36–41 (1994)
39. Radmacher, M., et al.: Science **257** (5078), 1900–1905 (1992)
40. Dufrene, Y.F.: Micron **32** (2), 153–165 (2001)
41. Crèvecoeur, M., et al.: Protoplasma **212**, 46–55 (2000)
42. Lesniewska, E., et al.: Methods in Cell Biology **68**, 51–65 (2002)
43. Dufrene, Y.F.: Nature Reviews Microbiology **2** (6), 451–460 (2004)
44. Horber, J.K., et al.: Scanning Microsc. **6** (4), 919–929; discussion 929–930 (1992)
45. Schneider, S.W., et al.: Proc. Natl. Acad. Sci. USA **94** (1), 316–321 (1997)
46. Le Grimmellec, C., et al.: Biophys. J. **75** (2), 695–703 (1998)
47. Velegol, S.B., Logan, B.E.: Langmuir **18**, 5256–5262 (2002)
48. Le Grimmellec, C., et al.: J. Comp. Neurol. **451** (1), 62–69 (2002)
49. Hansma, P.K., et al.: Science **243** (4891), 641–643 (1989)
50. Shevchuk, A.I., et al.: Biophys. J. **81** (3), 1759–1764 (2001)
51. Burnham, N.A., Colton, R.J.: J. Vac. Sci. Technol. **A7**, 2906–2913 (1989)
52. Weisenhorn, A., et al.: Nanotechnology **4**, 106–113 (1993)
53. Matzke, R., Jacobson, K., Radmacher, M.: Nat. Cell Biol. **3** (6), 607–610 (2001)
54. Rotsch, C., et al.: Ultramicroscopy **86** (1–2), 97–106 (2001)
55. Mahaffy, R.E., et al.: Biophys. J. **86** (3), 1777–1793 (2004)
56. Zhang, X., et al.: Am. J. Physiol. Heart Circ. Physiol. **286** (1), H359–367 (2004)
57. Tokunaga, M., et al.: Biochem. Biophys. Res. Commun. **231** (3), 566–569 (1997)
58. Tamayo, J., et al.: Biophys. J. **81** (1), 526–537 (2001)
59. Jeney, S., Florin, E.L., Horber, J.K.: Methods Mol. Biol. **164**, 91–108 (2001)
60. Burns, A.R.: Langmuir **19** (20), 8358–8363 (2003)
61. Gradinaru, C.C., et al.: Ultramicroscopy **99** (4), 235–245 (2004)
62. Kassies, R., et al.: J. Microscopy Oxford **217**, Part 1, 109–116 (2005)
63. Cho, S.J., et al.: Cell. Biol. Int. **26** (1), 35–42 (2002)
64. Hansma, H.G., et al.: Science **256** (5060), 1180–1184 (1992)
65. Thalhammer, S., et al.: J. Struct. Biol. **119** (2), 232–237 (1997)
66. Obataya, I. et al.: Nano Letters **5** (1), 27–30 (2005)
67. Milhiet, P.E., et al.: Single Mol. **3** (2–3), 136–141 (2002)

Section Two. Surface Force Apparatus and Micromanipulation

68. Israelachvili, J.N., Adams, G.E.: J. Chem. Soc. Faraday I **74**, 975–1001 (1978)
69. Cabane, B., Hénon, S.: *Liquides, solutions, dispersions, émulsions, gels*, Belin, Paris (2003)

70. Verwey, E.J.W., Overbeek, J.Th.G.: *Theory of the Stability of Lyophobic Colloids*, Elsevier, Amsterdam (1948)
71. Israelachvili, J.N.: *Intermolecular and Surface Forces*, Academic Press Limited, London (1991)
72. Homola, A.M., Israelachvili, J.N., Gee, M.L., McGuiggan, P.M., J. Tribology: Transactions of the ASME **111**, 675–682 (1989)
73. Israelachvili, J.N.: J. Colloid. Interface Sci. **44**, 259–272 (1973)
74. Derjaguin, B.V.: Kolloid-Z **69**, 155 (1934)
75. Israelachvili, J.N., Tabor, D.: Proc. Roy. Soc. London Series A: Mathematical and Physical Sci. **331**, 19–38 (1972)
76. Israelachvili, J.N., Pashley, R.M.: Nature **306**, 249–250 (1983)
77. Sivasankar, S., Brieher, W., Lavrik, N., Gumbiner, B., Leckband, D.: Proc. Natl. Acad. Sci. **96**, 11820–11824 (1999)
78. Pincet, F., Perez, E., Bryant, G., Lebeau, L., Mioskowski, C.: Phys. Rev. Lett. **73**, 2780–2784 (1994)
79. Johnson, K.L., Kendall, K., Roberts, A.D.: Proc. Roy. Soc. London A **324**, 301 (1971)
80. Nalaskowski, J., Veeramasuneni, S., Hupka, J., Miller, J.D.: J. Adh. Sci. Nd. Tech. **13**, 1519–1533 (1999)
81. Evans, E., Ritchie, K., Merkel, R.: Biophys. J. **68**, 2580–2587 (1995)
82. Merkel, R., Nassoy, P., Leung, A., Ritchie, K., Evans, E.: Nature **397** (6714), 50–53 (1999)
83. Chu, Y.S., Thomas, W.A., Eder, O., Pincet, F., Perez, E., Thiery, J.-P., Dufour, S.: J. Cell. Biol. **167**, 1163–1194 (2004)
84. Chu, Y.S., Dufour, S., Thiery, J.-P., Perez, E., Pincet, F., Phys. Rev. Lett. **94**, 028102 (2005)
85. Evans, E.: Colloid Surfaces **43**, 327–347 (1990)
86. Pincet, F., Perez, E., Loudet, J.C., Lebeau, L.: Phys. Rev. Lett. **87**, 178101–178114 (2001)
87. Gourier, C., Pincet, F., Perez, E., Zhang, Y.-M., Mallet, J.-M., Sinaÿ, P.: Angew. Chemie **44** (11), 1683–1687 (2005)

Section Three. Contact and Tapping Mode Force Microscopy

General References for Near-Field and Atomic Force Microscopy

88. Binnig, G., Rohrer, H.: Helv. Phys. Acta **55**, 726 (1982); Nobel Lecture, Rev. Mod. Phys. **59**, 615 (1987)
89. Eigler, D.M., Schweitzer, E.K.: Nature **344**, 524 (1990)
90. Binnig, G., Quate, C.F., Gerber, C.: Phys. Rev. Lett. **56**, 930 (1986)
91. Tsien, R.Y.: Imagining imaging's future, Nature Cell Biology **5**, S16–S21, September (2003)
92. Aimé, J.P., Boisgard, R., Nony, L., Couturier, G.: Phys. Rev. Lett. **89**, 3388 (1999)
93. Albrecht, T.R., Grütter, P., Horne, D., Rugar, D.: J. Appl. Phys. **69**, 668 (1991)
94. Anczycowsky, B., Krüger, D., Fuchs, H.: Phys. Rev. B **53**, 15485 (1996)
95. Boisgard, R., Michel, D., Aimé, J.P.: Surf. Sci. **401**, 199 (1998)
96. Dubourg, F., Marsaudon, S., Aimé, J.P., Leclère, Ph., Lazzaroni, R.: Eur. Phys. J. E **6**, 387 (2001)

97. Dubourg, F., Aimé, J.P., Couturier, G., Salardenne, J.: Eur. Phys. Lett. **62**, 671 (2003)
98. Dürig, U.: Appl. Phys. Lett. **76**, 1203 (2000)
99. Garcia, R., Perez, R.: Surf. Sci. Reports **47**, 197 (2002)
100. Giessibl, F.J.: Science **267**, 68–71 (1995)
101. Giessibl, F.J., Bielefeldt, H., Hembacher, S., Mannhart, J.: Appl. Surf. Science **140**, 352 (1999)
102. Gleyzes, P., Kuo, P.K., Boccard, A.C.: Appl. Phys. Lett. **58**, 25 (1991)
103. Horowitz, P., Hill, W.: *The Art of Electronics*, Cambridge University Press (1989)
104. Kitamura, S., Iwatsuki, M.: Jpn. J. Appl. Phys. Part 2 **34**, L145 (1995)
105. Sugawara, Y., Ohta, M., Ueyama, H., Moriat, S.: Science **270**, 1646 (1995)
106. Lüthi, R., Meyer, E., Bammerlin, M., Baratoff, A., Lehmann, T., Howald, L., Gerber, Ch., Güntherodt, H.J.: Z. Phys. B **100**, 165 (1996)
107. Marsaudon, S., Dubourg, F., Leclère, P., Lazzaroni, R., Aimé, J.P.: Langmuir **16**, 8432 (2000)
108. Nony, L., Boisgard, R., Aimé, J.P.: J. Chem. Phys. **111**, 1615 (1999)
109. Wang, L.: Appl. Phys. Lett. **73**, 3781 (1998)

Force Measurements

110. Bell, G.: Science **200**, 618 (1978)
111. Bustamante, C., Marko, J.F., Siggia, E.D., Smith, S.: Science **265**, 1599 (1994)
112. Clausen-Shaumann, H., Seitz, M., Krautbauer, R., Gaub, H.E.: Force spectroscopy with single biomolecules, Current Opinion in Chemical Biology **4**, 524 (2000)
113. Evans, E., Ritchie, K.: Biophys. J. **72**, 1541 (1997)
114. Florin, E.-L., Moy, V.T., Gaub, H.E.: Science **264**, 415 (1994)
115. Merkel, R., Nassoy, P., Leung, A., Ritchie, K., Evans, E.: Nature **397**, 50 (1999)
116. Merkel, R.: Force spectroscopy on single passive biomolecules and single biomolecular bonds, Phys. Rep. **346**, 343–385 (2001)
117. Moy, V.T., Florin, E.-L., Gaub, H.E.: Science **266**, 257 (1994)
118. Rief, M., Fernandez, J.M., Gaub, H.E.: Elastically coupled two-level systems as a model for biopolymer extensibility, Phys. Rev. Lett. **81**, 4764 (1998)
119. www.mpibpc.gwdg.de/abteilungen/071/strept.html
120. Kratky, O., Porod, G.: Rec. Trav. Chim. **68**, 1106 (1949)
121. Nony, L., Boisgard, R., Aimé, J.P.: J. Chem. Phys. **111**, 1615 (1999)
122. Rief, M., Oesterhelt, F., Heymann, B., Gaub, H.E.: Single molecule force spectroscopy on polysaccharides by atomic force microscopy, Science **275**, 1295 (1997)
123. Oesterhelt, F., Oesterhelt, D., Pfeiffer, M., Engel, A., Gaub, H.E., Müller, D.J.: Unfolding pathways of individual bacteriorhodopsins, Science **288**, 143 (2000)
124. Zhuang, X., Rief, M.: Single-molecule folding, Current Opinion in Structural Biology **13**, 88–97 (2003)

Biological Imaging

125. Engel, A., Lyubchenko, Y., Müller, D.: Atomic force microscopy: A powerful tool to observe biomolecules at work, Trends Cell Biol. **9**, 77–80 (1999)

126. Michel, D., Le Cam, E.: La microscopie de force atomique permet l'exploration du matériel génétique, *Pour la Science* **272**, 110 (2000)
127. Müller, D.J., Schabert, F.A., Büldt, G., Engel, A.: Imaging purple membranes in aqueous solutions at subnanometer resolution by atomic force microscopy, *Biophys. J.* **68**, 1681–1686 (1995)
128. Müller, D.J., Büldt, G., Engel, A.: Force-induced conformational change of bacteriorhodopsin, *J. Mol. Biol.* **249**, 239–243 (1995)
129. Müller, D.J., Schoenenberger, C.A., Schabert, F., Engel, A.: Structural changes in native membrane proteins monitored at subnanometer resolution with the atomic force microscope: A review, *J. Struct. Biol.* **119**, 149–157 (1997)
130. Müller, D.J., Fotiadis, D., Engel, A.: Mapping flexible protein domains at subnanometer resolution with the atomic force microscope, *FEBS Letters* **430**, 105–111 (1998)
131. Möller, C., Allen, M., Elings, V., Engel, A., Müller, D.J.: Tapping mode atomic force microscopy produces faithful high-resolution images of protein surfaces, *Biophys. J.* **77**, 1150–1158 (1999)
132. Nony, L., Boisgard, R., Aimé, J.P.: *Biomacromolecules* **2**, 827 (2001)
133. Pastré, D., Pietrement, O., Fusil, S., Landousy, F., Jeusset, J., David, M.O., Hamon, L., Le Cam, E., Zozime, A.: *Biophys. J.* **85**, 2507–2518 (2003)
134. Rivetti, C., Guthold, M., Bustamante, C.: *J. Mol. Bio.* **264**, 919 (1996)
135. Viani, M.B., et al.: *Nature Structural Biology* **7**, 644 (2000)

Technical Aspects of Cantilever Oscillation in a Liquid Medium

136. Fukuma, T., Kimura, M., Kobayashi, K., Matsuhige, K., Yamada, H.: *Rev. Sci. Instrum.* **76**, 53704 (2005)
137. Israelachvili, J.: *Intermolecular and Surface Forces*, 2nd. edn., Academic Press, London (1998)
138. Jeffery, S., Hoffmann, P.M., Pethica, J.B., Ramanujan, C., Ozer, H.O., Oral, A.: *Phys. Rev. B* **70**, 54114 (2004)
139. Maali, A., Hurth, C., Boisgard, R., Jai, C., Cohen-Bouhacina, T., Aimé, J.-P.: *J. Appl. Phys.* **97**, 074907 (2005)
140. Maali, A., Cohen-Bouhacina, T., Aimé, J.P., Couturier, G.: Unpublished results (2005)
141. Sader, J.E.: *J. Appl. Phys.* **84**, 64 (1998)
142. Tuck, E.O.: *J. Eng. Math.* **3**, 29–44 (1969)
143. Viani, M.B., Schäffer, T.E., Chand, A., Rief, M., Gaub, H.E., Hansma, P.K.: *J. Appl. Phys.* **86**, 2258 (1999)
144. Walter, D.A., Cleveland, J.P., Thompson, N.H., Hansma, P.K., Wendman, M.A., Gurley, G., Elings, V.: *Rev. Sci. Instrum.* **67**, 3583 (1996)

Section Four. Optical Tweezers

145. Ashkin, A., et al.: Observation of a single-beam gradient force optical trap for dielectric particles, *Opt. Lett.* **11**, 288 (1986)
146. Gittes, F., Schmidt, C.F.: In: *Methods in Cell Biology*, ed. by M.P. Sheetz, Academic, New York (1998) p. 129
147. Capitano, M., Romano, G., Ballerini, R., Giuntini, M., Dunlap, D., Finzi, L.: Calibration of optical tweezers with differential interference contrast signals, *Rev. Sci. Instrum.* **73**, 1687 (2002)

148. Bechhoefer, J., Wilson, S.: Faster, cheaper, safer optical tweezers for undergraduated laboratory, *Am. J. Phys.* **70**, 393 (2002)
149. Sasaki, K., et al.: Pattern formation and flow control of fine particles by laser-scanning manipulation, *Opt. Lett.* **16**, 1463 (1991)
150. Visscher, K., Brakenhoff, G., Krol, J.J.: Micromanipulation by multiple optical traps created by a single fast scanning trap integrated with the bilateral confocal scanning laser microscope, *Cytometry* **14**, 105 (1993)
151. Visscher, K., Gross, S.P., Block, S.M.: Construction of multiple-beam optical traps with nanometer-resolution, *IEEE J. Sel. Top. Quantum Electron.* **2**, 1066 (1996)
152. Emiliani, V., et al.: Multi-force optical tweezer to generate gradients of force, *Opt. Express* **12**, 3906 (2004)
153. Dufresne, E., Grieret, D.G.: Optical tweezer arrays and optical substrates created with diffractive optics, *Rev. Sci. Instrum.* **69**, 1974 (1998)
154. Cojoc, D., et al.: Design and fabrication of diffractive optical elements for optical tweezers arrays by means of e-beam lithography, *Microel. Eng.* **61–62**, 963 (2002)
155. Grier, D.G.: A revolution in optical manipulation, *Nature* **424**, 810 (2003)
156. Curtis, J.E., Koss, B.A., Grier, D.G.: Dynamic holographic optical tweezers, *Opt. Comm.* **207**, 169 (2002)
157. Melville, H.G.F., Milne, G.F., et al.: Optical trapping of three dimensional structures using dynamic holograms, *Opt. Express* **11**, 3562 (2003)
158. Leach, J., Sinclair, G., Jordan, P., Courtial, J., Padgett, M.J., Cooper, J., Laczik, Z.J.: 3D manipulation of particles into crystal structures using holographic optical tweezers, *Opt. Express* **12**, 220 (2004)
159. Gahagan, K.T., Swartzlander, G.A., Jr: Optical vortex trapping of particles, *Opt. Lett.* **21**, 827 (1996)
160. Curtis, J.E., Koss, B.A., Grier, D.: Dynamic holographic optical tweezers, *Opt. Commun.* **207**, 169 (2002)
161. Rodrigo, P.J., Daria, V.R., Gluckstad, J.: Real-time interactive optical micro-manipulation of a mixture of high- and low- index particles, *Opt. Express* **12**, 1417 (2004)
162. O'Neil, T., Padgett, M.J.: Three-dimensional optical confinement of micron-sized metal particles and the decoupling of the spin and orbital angular momentum within an optical spanner, *Opt. Comm.* **185**, 139 (2000)
163. Lang, M.J., Fordyce, P.M., Engh, A.M., Neuman, K.C., Block, S.M.: Simultaneous, coincident optical trapping and single-molecule fluorescence, *Nat. Meth.* **22**, 133 (2004)
164. Emiliani, V., Cojoc, D., Ferrari, E., Garbin, V., Durieux, C., Coppey-Moisán, M., Di Fabrizio, E.: Wave front engineering for living cells microscopy, *Opt. Express* **13**, 1405 (2005)
165. Rohrbach, A., Stelzer, E.H.K.: *J. Appl. Phys.* **91**, 5474 (2002)
166. Sheetz, M.P. (Ed.): *Methods in Cell Biology*, Academic, New York (1998)
167. Choquet, D., Felsenfeld, D., Sheetz, M.P.: Extracellular matrix rigidity causes strengthening of integrin-cytoskeleton linkages, *Cell* **88**, 39 (1997)
168. Galbraith, C.G., Yamada, K.M., Sheetz, M.P.: The relationship between force and focal complex development, *J. of Cell Biol.* **159**, 695 (2002)
169. Emiliani, V., Sanvitto, D., Coppey, M., Durieux, C.: Submitted

170. Del Pozo, M.A., Kiosses, W.B., Alderson, N.B., Meller, N., Hahn, K.M., Schwartz, M.A.: Integrin regulated GTPRac localized effector interaction through dissociation of Rho-GDI, *Nature Cell Biol.* **4**, 232 (2002)
171. Wang, Y., Botvinick, E.L., Zhao, Y., Berns, M.W., Usami, S., Tsien, R.Y., Chien, S.: *Nature* **434**, 1040 (2005)
172. Di Fabrizio, E., Cojoc, D., Emiliani, V., Cabrini, S., Coppey-Moisán, M., Ferrari, E., Garbin, V., Altissimo, M.: Microscopy of biological sample through advanced diffractive optics from visible to X-ray wavelength regime, *Microscopy Research and Techniques* **65**, 252 (2004)
173. Townes-Anderson, E. et al.: Micromanipulation of retinal neurons by optical tweezers, *Molecular Vision* **4**, 12 (1998)
174. Thoumine, O., Kocian, P., Kottelat, A., Meister, J.J.: Short term binding of fibroblasts to fibronectin: Optical tweezers experiments and probabilistic analysis, *Eur. Biophys. J.* **29**, 398 (2000)
175. Lenormand, G., Henon, S., Richert, A., Simeon, J., Gallet, F.: Direct measurement of the area expansion and shear moduli of the human red blood cell membrane skeleton, *Biophys. J.* **81**, 43 (2001)
176. Guck, J., Ananthakrishnan, R., Mahmood, H., Moon, T.J., Cunningham, C.C., Kas, J.: The optical stretcher: A novel laser tool to micromanipulate cells, *Biophys. J.* **81**, 767 (2001)
177. Umehara, S., Wakamoto, Y., Inoue, I., Yasuda, K.: On-chip single-cell microcultivation assay for monitoring environmental effects on isolated cells, *Biochem. and Biophys. Research Com.* **305**, 534 (2003)
178. Finer, J., Simmons, R.M., Spudich, J.A.: Single myosin molecule mechanics: Piconewton forces and nanometer steps, *Nature* **368**, 113 (1994)
179. Svoboda, K., Block, S.: Force and velocity measured for single kinesin molecules, *Cell* **77**, 773 (1994)
180. Coppin, C.M., Finer, J.T., Spudich, J.A., Vale, R.D.: Detection of sub-8-nm movements of kinesin by high-resolution optical trap microscopy, *Proc. Natl. Acad. Sci. USA* **93**, 1913 (1996)

Section Five. Magnetic Tweezers

181. Kubo, R.: *J. Phys. Soc. Japan* **12**, 570 (1957); Kubo, R.: The fluctuation-dissipation theorem, *Rep. Prog. Phys.* **29**, 255–284 (1966)
182. Gelles, J., Schnapp, B., Sheetz, M.: Tracking kinesin-driven movements with nanometre-scale precision, *Nature* **331**, 450–453 (1988)
183. Gosse, C., Croquette, V.: Magnetic tweezers: Micromanipulation and force measurement at the molecular level, *Biophys. J.* **82**, 3314 (2002)
184. Ovrn, B.: Three-dimensional forward scattering particle image velocimetry applied to a microscopic field of view, *Experiments in Fluids (Suppl.)* S175–S184, Springer, Berlin Heidelberg New York (2000)
185. Strick, T.R., Allemand, J.-F., Bensimon, D., Bensimon, A., Croquette, V.: The elasticity of a single supercoiled DNA molecule, *Science* **271**, 1835 (1996)
186. Smith, S.B., Finzi, L., Bustamante, C.: Direct mechanical measurements of the elasticity of single DNA molecules by using magnetic beads, *Science* **258** (5085), 1122 (1992)
187. Bustamante, C., Marko, J.F., Siggia, E.D., Smith, S.: Entropic elasticity of lambda-phase DNA, *Science* **265** (5178), 1599–1560 (1994)

188. Bouchiat, C., Wang, M.D., Allemand, J.-F., Strick, T.R., Block, S.M., Croquette, V.: Estimating the persistence length of a worm-like chain molecule from force-extension measurements, *Biophys. J.* **76**, 409–413 (1999)
189. Strick, T.R., Allemand, J.-F., Bensimon, D., Croquette, V.: Behavior of supercoiled DNA, *Biophys. J.* **74**, 2016 (1998)
190. White, J.: Self linking and the Gauss integral in higher dimensions, *Am. J. Math.* **91**, 693–728 (1969)
191. Bouchiat, C., Mézard, M.: Elasticity theory of a supercoiled DNA molecules, *Phys. Rev. Lett.* **80**, 1556–1559 (1998); Moroz, J., Nelson, P.: Torsional directed walks, entropic elasticity and DNA twist stiffness, *Proc. Natl. Acad. Sci. USA* **94**, 14418–14422 (1998)
192. Strick, T.R., Croquette, V., Bensimon, D.: Homologous pairing in stretched supercoiled DNA, *Proc. Natl. Acad. Sci. USA* **95**, 10579 (1998)
193. Allemand, J.-F., Bensimon, D., Lavery, R., Croquette, V.: Stretched and overwound DNA forms a Pauling-like structure with exposed bases, *Proc. Natl. Acad. Sci. USA* **95**, 14152 (1998)
194. Dessinges, M.N., Lionnet, T., Xi, X.G., Bensimon, D., Croquette, V.: Single-molecule assay reveals strand switching and enhanced processivity of UvrD, *Proc. Natl. Acad. Sci. USA* **101**, 6439 (2004), doi:10.1073/pnas.0306713101
195. Charvin, G., Bensimon, D., Croquette, V.: Single-molecule study of DNA unlinking by eukaryotic and prokaryotic type-II topoisomerases, *Proc. Natl. Acad. Sci. USA* **100** (17), 9820–9825 (2003)
196. Fulconis, R., Bancaud, A., Allemand, J.-F., Croquette, V., Dutreix, M., Viovy, J.L.: Twisting and untwisting a single DNA molecule covered by RecA protein, *Biophys. J.* **87** (4), 2552 (2004)
197. Lia, G., Bensimon, D., Croquette, V., Allemand, J.-F., Dunlap, D., Lewis, D.E.A., Adhya, S.C., Finzi, L.: Supercoiling and denaturation in Gal repressor/heat unstable nucleoid protein (HU)-mediated DNA looping, *Proc. Natl. Acad. Sci. USA* **100**, 11373 (2003), doi:10.1073/pnas.2034851100
198. Dessinges, M.N., Maier, B., Zhang, Y.H., Peliti, M., Bensimon, D., Croquette, V.: Stretching single stranded DNA, a model polyelectrolyte, *Phys. Rev. Lett.* **89**, 248102 (2002)
199. Abels, J.A., Moreno-Herrero, F., van der Heijden, T., Dekker, C., Dekker, N.H.: Single-molecule measurements of the persistence length of double-stranded RNA, *Biophys. J.* **88** (4), 2737–2744 (2005). Epub Jan 14 2005
200. Zlatanova, J., Leuba, S.H.: Magnetic tweezers: A sensitive tool to study DNA and chromatin at the single-molecule level, *Biochem. Cell Biol.* **81**, 151 (2003); Leuba, S.H., Karymov, M.A., Tomschik, M., Ramjit, R., Smith, P., Zlatanova, J.: Assembly of single chromatin fibers depends on the tension in the DNA molecule: Magnetic tweezers study, *Proc. Natl. Acad. Sci. USA* **100**, 495, PNAS publications (2003)
201. Ali, B.M.J., Amit, R., Braslavsky, I., Oppenheim, A.B., Gileadi, O., Stavans, J.: Compaction of single DNA molecules induced by binding of integration host factor (IHF), *Proc. Natl. Acad. Sci. USA* **98**, 10658 (2001); Skoko, D., Wong, B., Johnson, R.C., Marko, J.F.: Micromechanical analysis of the binding of DNA-bending proteins HMGB1, NHP6A, and HU reveals their ability to form highly stable DNA-protein complexes, *Biochemistry* **43**, 13867 (2004)
202. Danilowicz, C., Kafri, Y., Conroy, R.S., Coljee, V.W., Weeks, J., Prentiss, M.: Measurement of the phase diagram of DNA unzipping in the temperature-force plane, *Phys. Rev. Lett.* **93**, 078101 (2004); Danilowicz, C., Coljee, V.W.,

- Bouzigues, C., Lubensky, D.K., Nelson, D.R., Prentiss, M.: DNA unzipped under a constant force exhibits multiple metastable intermediates, *Proc. Natl. Acad. Sci. USA* **100**, 1694, PNAS publications (2003)
203. Maier, B., Bensimon, D., Croquette, V.: Replication by a single DNA polymerase of a stretched single-stranded DNA, *Proc. Natl. Acad. Sci. USA* **97**, 12002 (2000), doi:10.1073/pnas.97.22.12002
 204. Revyakin, A., Ebright, R.H., Strick, T.R.: Promoter unwinding and promoter clearance by RNA polymerase: Detection by single-molecule DNA nanomanipulation, *Proc. Natl. Acad. Sci. USA* **101**, 4776, PNAS publications (2004)
 205. Koster, D.A., Croquette, V., Dekker, C., Shuman, S., Dekker, N.H.: Friction and torque govern the relaxation of DNA supercoils by eukaryotic topoisomerase IB, *Nature* **434** (7033), 671–674 (2005); Dekker, N.H., Rybenkov, V.V., Duguet, M., Crisona, N.J., Cozzarelli, N.R., Bensimon, D., Croquette, V.: The mechanism of type IA topoisomerases, *Proc. Natl. Acad. Sci. USA* **99**, 12126 (2002), doi:10.1073/pnas.132378799
 206. Strick, T.R., Croquette, V., Bensimon, D.: Single-molecule analysis of DNA uncoiling by a type II topoisomerase, *Nature* **404**, 901 (2000), doi:10.1038/35009144; Crisona, N.J., Strick, T.R., Bensimon, D., Croquette, V., Cozzarelli, N.R.: Preferential relaxation of positively supercoiled DNA by E-coli topoisomerase IV in single-molecule and ensemble measurements, *Genes Dev.* **14**, 2881 (2000)
 207. Dawid, A., Croquette, V., Grigoriev, M., Heslot, F.: Single-molecule study of RuvAB-mediated Holliday-junction migration, *Proc. Natl. Acad. Sci. USA* **101**, 11611 (2004), doi:10.1073/pnas.0404369101; Amit, R., Gileadi, O., Stavans, J.: Direct observation of RuvAB-catalyzed branch migration of single Holliday junctions, *Proc. Natl. Acad. Sci. USA* **101**, 11605, PNAS publications (2004)
 208. Strick, T.R., Kawaguchi, T., Hirano, T.: Real-time detection of single-molecule DNA compaction by condensin I, *Curr. Biol.* **14**, 874 (2004)
 209. Seidel, R., Van Noort, J., Van Der Scheer, C., Bloom, J.G.P., Dekker, N.H., Dutta, C.F., Blundell, A., Robinson, T., Firman, K., Dekker, C.: Real-time observation of DNA translocation by the type I restriction modification enzyme EcoR124I, *Nat. Struct. Mol. Biol.* **11**, 838 (2004)
 210. Saleh, O.A., Peral, C., Barre, F.X., Allemand, J.-F.: Fast, DNA-sequence independent translocation by FtsK in a single-molecule experiment, *EMBO J.* **23**, 2430 (2004); Saleh, O.A., Bigot, S., Barre, F.X., Allemand, J.-F.: Analysis of DNA supercoil induction by FtsK indicates translocation without groove-tracking, *Nat. Struct. Mol. Biol.* **12** (5), 436–440 (2005)
 211. Matthews, B.D., LaVan, D.A., Overby, D.R., Karavitis, J., Ingber, D.E.: Electromagnetic needles with submicron pole tip radii for nanomanipulation of biomolecules and living cells, *Appl. Phys. Lett.* **85** (14), 2968–2970 (2004); Amblard, F., Yurke, B., Pargellis, A., Leibler, S.: A magnetic manipulator for studying local rheology and micromechanical properties of biological systems, *Rev. Sci. Instrum.* **67**, 1–10 (1996); Hosu, B.G., Jakab, K., Banki, P., Toth, F.I., Forgacs, G.: Magnetic tweezers for intracellular applications, *Rev. Sci. Instrum.* **74**, 4158 (2003)

Surface Methods

D. Altschuh, S. Ricard-Blum, V. Ball, M. Gaillet, P. Schaaf, B. Senger, B. Desbat, P. Lavalle, and J.-F. Legrand

9.1 Biosensors Based on Surface Plasmon Resonance: Interpreting the Data

9.1.1 Introduction

Definition of a Biosensor

A biosensor is an analytical device which physically associates a biological receptor and a transducer [1]. The transducer transforms the interaction between the receptor and its target into an interpretable signal. No other step, such as the separation of free and complexed molecules or the addition of a further reagent, is required.

Biosensors have been classified in terms of the transduction mode or the nature of the biological receptor [2–4]. The modes of transduction applicable to all molecular interactions, without labelling the molecules and whatever their nature, detect either a change in mass on a surface (surface plasmon resonance, acoustic biosensors), or a heat transfer (calorimetry) [6]. Among the devices using these transduction modes, the most widely used in biology at the present time are based on surface plasmon resonance (SPR). Since the commonest SPR biosensors are manufactured by Biacore AB, a company based in Uppsala, Sweden [5] (Biacore is a registered trademark), we begin by discussing the interpretation of SPR data obtained in this configuration.

The Biacore Technology

Figure 9.1 shows schematically the configuration and underlying principle of the Biacore instruments. One of the molecules, called the ligand here, is immobilised on a biospecific interface called the sensor surface (Fig. 9.2). The other molecule, called the analyte here, is injected onto this surface at a continuous flow rate using a microfluidic system (Fig. 9.1b). The surface is illuminated by a beam of monochromatic polarised light, in such a way that all the light is

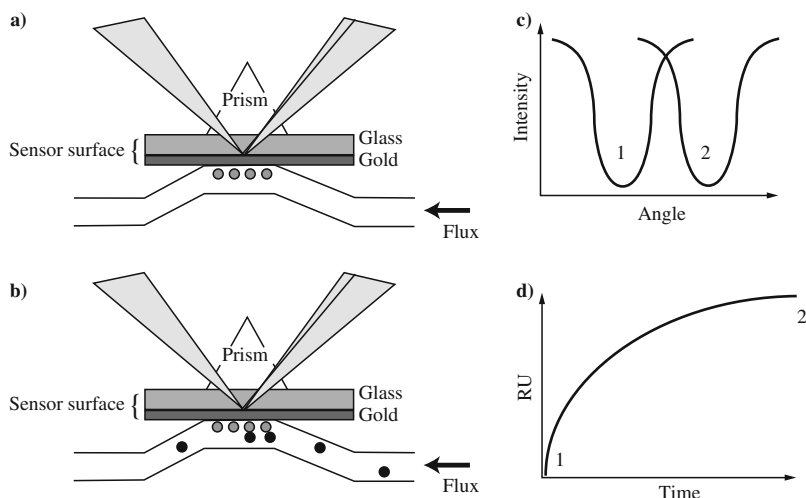


Fig. 9.1. Biacore technology. **(a)** A ligand has been immobilised covalently on the sensor surface. The resonance angle lies in position 1. **(b)** The analyte is injected onto the surface at a continuous flow rate. If it binds to the immobilised ligand, the resonance angle shifts to position 2. **(c)** The detector measures the position of this angle (minimum light intensity) at each instant of time. **(d)** The displacement of the angle, expressed in resonance units (RU), is plotted as a function of time, thereby generating a sensorgram

reflected (conditions of total internal reflection). The phenomenon of surface plasmon resonance (SPR) causes a drop in the intensity of the light reflected at a definite angle, called the resonance angle. This resonance angle depends on the refractive index near the surface, which itself depends on a certain number of factors, including the mass of surface molecules. By recording the displacement of the resonance angle over time, it is possible to monitor the association and dissociation of the analyte and the immobilised ligand (see Fig. 9.1c). The variation of the resonance angle (or SPR signal), expressed in arbitrary units called resonance units (RU), is plotted as a function of time. The resulting interaction curves are called sensorgrams (see Fig. 9.1d). A variation of 1,000 RU corresponds to a deviation of 0.1° in the resonance angle and to the binding of 1 ng of protein per mm^2 of the surface. The relation between the molecular mass (MM) per unit area and the variation of the signal is the same for all proteins, but differs for other molecular species, such as nucleic acids, carbohydrates, lipids, and so on.

The SPR data can be interpreted qualitatively (detection of binding). However, we shall be particularly interested in quantitative information that can be extracted by mathematical analysis of these sensorgrams, viz., affinity and rate parameters at the interaction equilibrium, active concentration of analyte

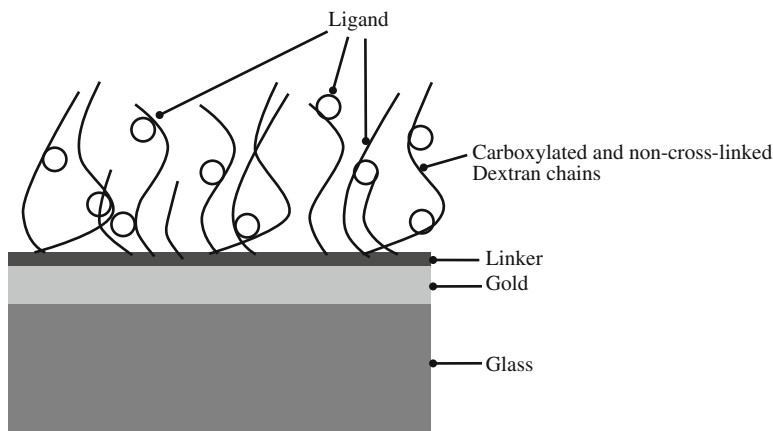


Fig. 9.2. The sensor surface comprises a glass substrate coated with a film of gold on which the molecular interactions take place. Indeed, the SPR phenomenon occurs when a thin metal film (gold here), rich in free electrons, is deposited at the interface between two media with different refractive indices (here, glass and the aqueous medium). The surfaces commercialised by Biacore are usually grafted with a hydrogel of dextran (a polysaccharide made up of glucose units), carboxylated and not crosslinked, with an average thickness of 100 nm. The dextran matrix allows covalent binding of the ligand to carboxyl groups, and also an increase in the sensitivity of the response because more molecules can be grafted in a volume of gel than onto a plane surface. The dextran chains are relatively mobile because they are not crosslinked. Hence, the interaction does not take place on a solid phase, but in a semi-liquid phase

and binding stoichiometry. Indeed, few other systems allow such a detailed mathematical description of the experimental data.

Experimental Data or Sensorgrams

A sensorgram typically comprises three phases (see Fig.9.3). The analyte-injection and post-injection phases are used to evaluate the data. The ligand-analyte interaction is reversible and obeys the law of mass action. The association and dissociation events occur during both the injection and post-injection phase. The predominance of one or the other event depends on the respective concentrations of free and bound molecules at or near the surface and the rate parameters of the interaction. We begin by describing the equations used to evaluate the Biacore data, then the experimental conditions that must be respected in order to produce interpretable data using these equations.

9.1.2 Evaluating the SPR Data

The equations in this section were taken from [7–12].

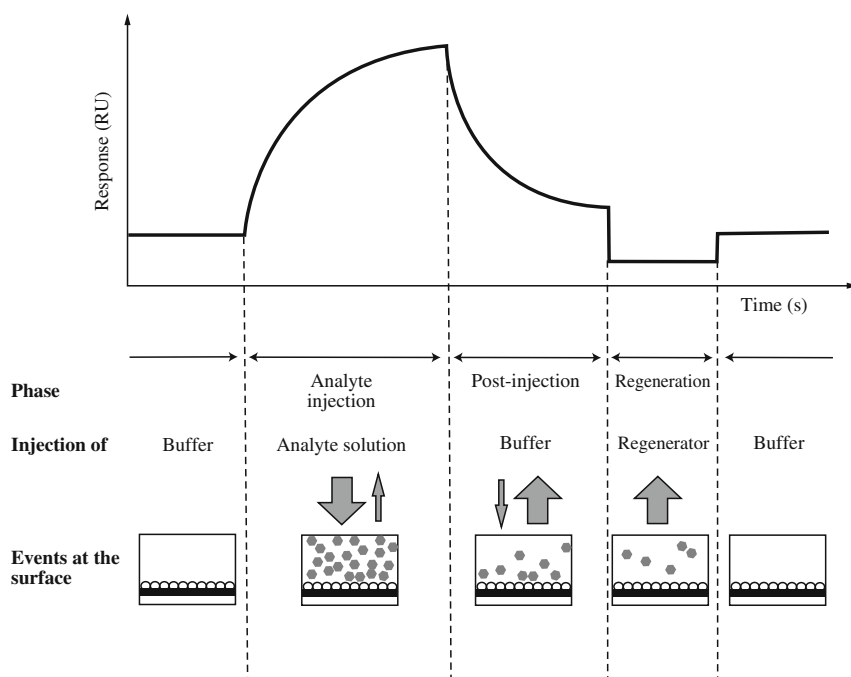
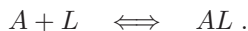


Fig. 9.3. Sensorgram phases. The surface shown schematically here carries a co-variantly immobilised ligand. It is permanently exposed to a flow of solvent (buffer alone, solution containing the analyte, regeneration solution). The response (RU) is initially stable. Analyte injection phase: the flow of buffer is replaced by a flow of analyte and the response increases if the analyte interacts with the immobilised ligand and the mass of molecules on the sensor surface increases. Post-injection phase: the analyte flow is replaced by a buffer flow and the response falls off if the complex dissociates. Regeneration phase: a regeneration solution is injected to eliminate any analyte still bound to the ligand. The surface carrying the ligand can now be used for another experiment

Interaction in Solution

Consider a reversible interaction between a monovalent ligand L and analyte A , both in solution:



According to the law of mass action, the rate of a reaction is proportional to the concentrations of each of the reacting substances. The rate of formation of the complex depends on the concentrations of free analyte (A_{free}) and free ligand (L_{free}), and also on the association rate constant k_a , sometimes denoted k_{on} ($k_a \times A_{\text{free}} \times L_{\text{free}}$). The dissociation rate of the complex depends on the concentration AL of the complex and the dissociation rate constant k_d , also

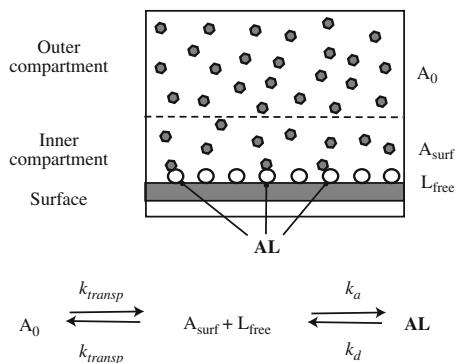


Fig. 9.4. Schematic view of the two-compartment model. A_0 and A_{surf} are the analyte concentrations in the inner and outer compartments, respectively. L_{free} is the concentration of free ligand at the surface. AL is the concentration of the complex

denoted $k_{\text{off}}(k_d \times AL)$. The reaction rate is then the difference between these two quantities:

$$\frac{dAL}{dt} = (k_a \times A_{\text{free}} \times L_{\text{free}}) - (k_d \times AL) . \quad (9.1)$$

Interaction on a Surface

The interactions taking place on a surface that carries one of the two interacting partners are influenced by the diffusive transport of free molecules toward the surface. If the binding process is fast compared with this transport, a gradient of analyte concentration will form beyond the surface. The standard model used to describe this type of interaction, e.g., an analyte binding to receptors present at the surface of a cell, divides the volume beyond the surface into two compartments [7], close to and far from the surface, called here the inner and outer compartments, respectively (see Fig. 9.4). The reaction rate observed then depends on two phenomena. The first is the transport of analyte molecules between the two compartments, which itself depends on the transport coefficient k_{transp} . The other phenomenon is the binding process between the analyte and the immobilised ligand, which depends on the rate parameters of the interaction. The concentration of free analyte in the outer compartment is denoted A_0 , while that in the inner compartment, close to the surface and therefore available for binding to the immobilised ligand, is denoted A_{surf} .

Interaction in a Biacore Flow Cell

In the Biacore configuration, the analyte is transported toward the surface by diffusion and convection, because it is injected into a continuous flow. Furthermore, the ligand–analyte interaction does not take place on a plane surface,

but rather in a volume of gel (usually a dextran matrix). Various models have been developed to take these phenomena into account and then confronted with experimental data, in order to determine the equations describing the molecular interaction under these specific conditions and/or the experimental conditions to be respected in order to make use of the data. It has been shown that the two-compartment model gives an adequate description of the system in most cases [7, 9, 10].

In this model the concentrations of free analyte in the two compartments are treated as being uniform in space. The free analyte concentration A_0 in the outer compartment is constant in time (continuous flow of analyte) and equal to the injected concentration. The free analyte concentration A_{surf} in the inner compartment varies in time because the analyte is transported from one compartment to the other, and because it can associate and dissociate with the immobilised ligand. The equations used to interpret the data describe the rate of formation of the complex [see (9.2)] and also the change in concentration of the free analyte near the surface as a function of time [see (9.3)]:

$$\frac{dAL}{dt} = (k_a \times A_{\text{surf}} \times L_{\text{free}}) - (k_d \times AL) , \quad (9.2)$$

$$\frac{dA_{\text{surf}}}{dt} = (k_{\text{transp}} \times A_0) - (k_{\text{transp}} \times A_{\text{surf}}) - (k_a \times A_{\text{surf}} \times L_{\text{free}}) + (k_d \times AL) . \quad (9.3)$$

The transport coefficient k_{transp} depends on the diffusion constant of the analyte, the geometry of the flow cell, and the flow rate of the solvent. This binding mode, expected for an interaction between a monovalent analyte and a ligand attached on a surface, is equivalent to the so-called Langmuir model. This model is widely used to describe particle adsorption phenomena on a surface when the number of adsorption sites on the surface is fixed, each site can only adsorb one particle, and interactions between particles can be neglected.

Transposing to Biacore Data

The correspondence between the quantities in (9.2) and (9.3) and the Biacore data are given in Table 9.1 together with measurement units. In Biacore data, (9.2) and (9.3) become, respectively,

$$\frac{dR}{dt} = k_a \times A_{\text{surf}} \times (R_{\text{max}} - R) - k_d \times R , \quad (9.4)$$

$$\frac{dA_{\text{surf}}}{dt} = (k_t \times A_0) - (k_t \times A_{\text{surf}}) - \left[k_a \times A_{\text{surf}} \times (R_{\text{max}} - R) \right] + (k_d \times R) . \quad (9.5)$$

The concentration A_{surf} is not constant over time, but changes slowly, so one can make the approximation $dA_{\text{surf}}/dt = 0$ (quasi-equilibrium state [7, 9, 10]), whereupon

$$A_{\text{surf}} = \frac{(k_t \times A_0) + (k_d \times R)}{k_t + \left[k_a \times (R_{\text{max}} - R) \right]} . \quad (9.6)$$

Table 9.1. Correspondence between the usual notation and Biacore data. $R_{\max}$ is the total concentration of ligand sites, given by the SPR response when the surface is saturated by the analyte. $R_{\max}$ and R thus have the same units, viz., RU of bound analyte. Note that the word ‘parameter’ is often used for the association and dissociation constants when one of the two partners in the interaction, immobilised on a surface, is unable to diffuse freely. In the last row, $k_t = r \times k_{\text{transp}}$, where r is a factor converting the response into a surface concentration

Usual notation	Biacore notation	Definition
AL [M]	R [RU]	Complex concentration: the response R is proportional to the mass of analyte bound to the immobilised ligand, and hence to the complex concentration
A_0 [M]	A_0 [M]	Injected concentration of free analyte
A_{surf} [M]	A_{surf} [M]	Concentration of free analyte near the surface
L_{free} [M]	$R_{\max} - R$ [RU]	Free ligand concentration: difference between total number of sites $R_{\max}$ and number of bound sites R
k_a [$\text{M}^{-1}\text{s}^{-1}$]	k_a [$\text{M}^{-1}\text{s}^{-1}$]	Association rate constant (or parameter)
k_d [s^{-1}]	k_d [s^{-1}]	Dissociation rate constant (or parameter)
k_{transp} [m/s]	k_t [RU/M s]	Mass transport coefficient: describes analyte transport between the two compartments

Substituting A_{surf} into (9.4) yields

$$\begin{aligned} \frac{dR}{dt} = & \frac{k_a}{1 + k_a \times (R_{\max} - R)/k_t} \times A_0 \times (R_{\max} - R) \\ & - \frac{k_d}{1 + k_a \times (R_{\max} - R)/k_t} \times R. \end{aligned} \quad (9.7)$$

Equation (9.7) is used to describe the reaction rate observed in the Biacore configuration. It can also be written

$$\frac{dR}{dt} = k_+ \times A_0 \times (R_{\max} - R) - k_- \times R, \quad (9.8)$$

with the effective rate constants k_+ and k_- given by

$$k_+ = \frac{k_a}{1 + k_a \times (R_{\max} - R)/k_t}, \quad (9.9)$$

$$k_- = \frac{k_d}{1 + k_a \times (R_{\max} - R)/k_t}. \quad (9.10)$$

The order of magnitude of the different parameters in (9.7) is illustrated in the following sections by means of simulated and experimental sensorgrams.

Complex Interaction Models

If the kinetic experimental data cannot be interpreted using the above equations, there may be three reasons for this: poor experimental conditions, complexity (heterogeneity, aggregation) of the molecules, or complexity of the binding process. Once the first two possibilities have been eliminated by carefully checking the molecules and experimental conditions (see Sects. 9.1.3 and 9.1.4), complex interaction models can be applied for interpreting the data. These models take into account things like partial mass transport, bivalent analytes, heterogeneity of the ligand, conformational change following binding, and so on [8].

9.1.3 Measurements Under Mass Transport or Kinetic Conditions

Different information is extracted from the experimental data depending on whether the measured reaction rate is controlled by the rate parameters of the interaction (so-called kinetic conditions, illustrated in Fig. 9.5a) or by the transport of molecules toward the surface (so-called mass transport conditions, illustrated in Fig. 9.5b).

Kinetic Experimental Conditions

Information about the rate parameters of the molecular interaction are obtained when transport of the analyte toward the surface is not the limiting factor, so that the measured reaction rate dR/dt depends only on the rate parameters k_a and k_d of the interaction.

Equation (9.7) shows that these conditions are fulfilled when $k_a \times R_{\max} \ll k_t$. Schematically, it could be said that any analyte molecule forming a complex with the immobilised ligand is replaced by a molecule of free analyte brought by the flow of solute, and that any analyte molecule that dissociates from the surface is washed by the flow of buffer. Consequently, during the analyte injection phase, the concentration of free analyte near the surface is equal to the injected concentration, i.e., $A_{\text{surf}} = A_0$ (see Fig. 9.5a). This concentration is constant during the whole time that the analyte is being injected and throughout the whole flow cell. The two-compartment model is no longer necessary. The only difference compared with an interaction in solution is that the ligand cannot diffuse freely, which can lead to underestimation of the rate parameters [9]. However, the ligand still has a certain degree of freedom if bound to a flexible polymer, such as a non-crosslinked dextran gel.

If $k_a \times R_{\max} \ll k_t$, equation (9.7) becomes equivalent to (9.4) with $A_{\text{surf}} = A_0$:

$$\frac{dR}{dt} = k_a \times A_0 \times (R_{\max} - R) - k_d \times R, \quad (9.11)$$

so that, after integrating,

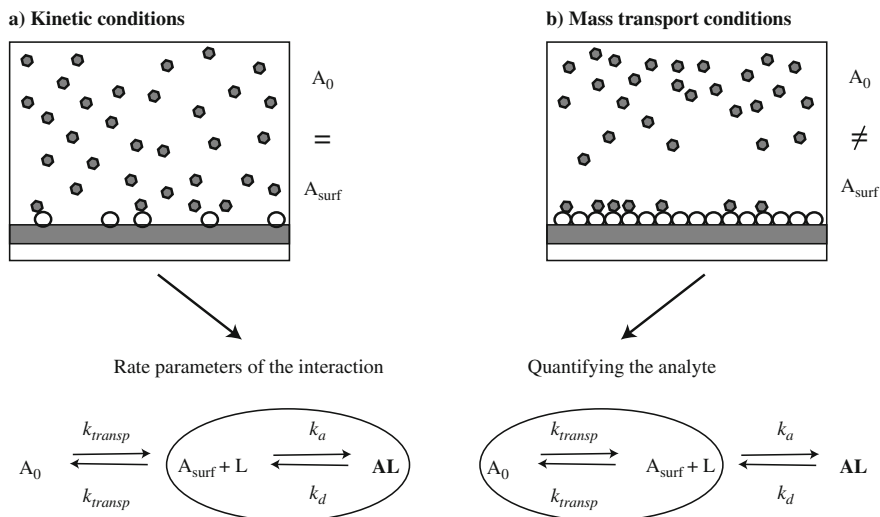


Fig. 9.5. Schematic representation of kinetic conditions (a) and mass transport conditions (b) during the analyte injection phase

$$R = \frac{k_a \times A_0 \times R_{\max} \times [1 - e^{-(k_a \times A_0 + k_d)t}]}{k_a \times A_0 + k_d}. \quad (9.12)$$

This relation describing the time dependence of the response function when kinetic experimental conditions hold sway can be used to calculate the values of k_a and k_d by non-linear regression. In experimental practice, analyte solutions are injected at different concentrations onto a surface carrying the ligand, and the fitted parameters are k_a , k_d , and $R_{\max}$ (see Fig. 9.6). Numerical integration methods are used to seek the value of each of these parameters that are compatible with the whole set of experimental data (injection and post-injection phases for all injected analyte concentrations). Statistical parameters can then assess the quality of the fit.

Specific Conditions

During the post-injection phase, under kinetic conditions, $A_{\text{surf}} = 0$ and

$$\frac{dR}{dt} = -k_d \times R, \quad \text{whence } R = R_0 e^{-k_d(t-t_0)}, \quad (9.13)$$

where R_0 and t_0 are the response and the time, respectively, at the beginning of the evaluated post-injection phase.

At equilibrium,

$$\frac{dR}{dt} = 0, \quad \text{whence, by (9.11), } R = \frac{K_a \times A_0 \times R_{\max}}{K_a \times A_0 + 1}. \quad (9.14)$$

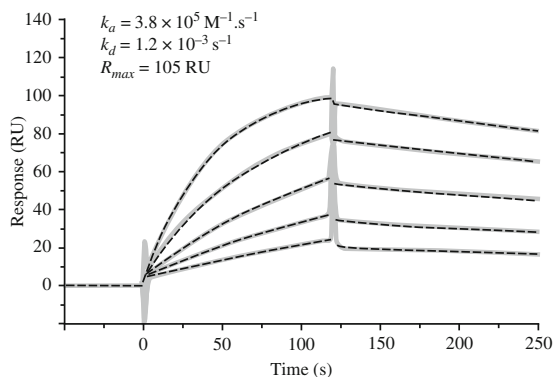


Fig. 9.6. An experiment to make kinetic measurements. A protein ligand has been immobilised on the sensor surface. The analyte has been injected at five different concentrations (4, 8, 16, 32 and 64 nM), followed by injection of a 50 mM hydrochloric acid solution to regenerate the surface. The figure shows the superposition of 5 sensorgrams (*thick grey lines*), but the regeneration stage is not shown. The best possible fit between experimental and theoretical data has been found by numerical integration and global processing of the data [11], delivering values for k_a , k_d , and $R_{\max}$ that are compatible with the full set of experimental data. *Black dashed lines* indicate theoretical sensorgrams simulated using (9.12) together with these values

For a low-affinity interaction ($K_a \leq 10^6 \text{ M}^{-1}$), the equilibrium state is reached during the first few seconds after the injection phase, a short period during which the mixing of buffer and analyte solution does not always allow interpretation of the kinetic data (see Figs. 9.7a and c). Likewise, the post-injection phase cannot always be interpreted. Equation (9.14) can be used in this case to calculate the affinity at equilibrium ($K_a = k_a/k_d = 1/K_d$), by plotting the equilibrium response R_{eq} as a function of the injected analyte concentration (see Figs. 9.7b and d).

Total Mass Transport Conditions

Information about the active analyte concentration is obtained when the observed reaction rate dR/dt is controlled by transport of the analyte toward the surface and is independent of the rate parameters [9]. Equation (9.7) shows that these conditions, called total mass transport conditions, are fulfilled when $k_a \times R_{\max} \gg k_t$, and R is negligible (beginning of the analyte injection phase). In this case, the reaction rate is proportional to the concentration of injected analyte A_0 . Schematically, it could be said that each molecule of analyte approaching the surface is captured in the form of a ligand-analyte complex, and that the free analyte concentration A_{surf} near the surface is zero (see Fig. 9.5b).

To evaluate the active concentration of analyte in an unknown sample, a calibration curve is established relating the reaction rate shortly after injecting

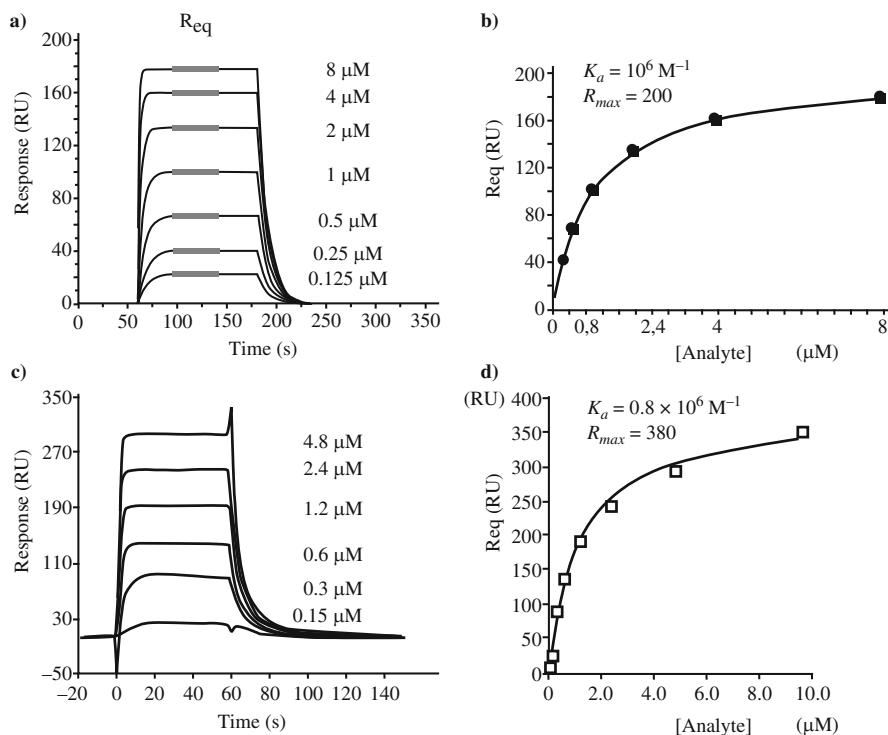


Fig. 9.7. Affinity measurements at equilibrium. (a) and (b) show simulated data. (a) Expected sensorgrams for the injection of 7 analyte concentrations (between 125 and 8,000 nM) on a ligand surface ($R_{max} = 200$ RU). The rate parameters used for the simulation are $k_a = 10^5 \text{ M}^{-1}\text{s}^{-1}$, $k_d = 0.1 \text{ s}^{-1}$, whence $K_a = 10^6 \text{ M}^{-1}$. (b) The equilibrium response as a function of analyte concentration is interpreted using (9.14). The resulting values of K_a and R_{max} are indicated. (c) and (d) Experimental data adapted from [13]. (c) The surface carries a protein ligand. Analyte solutions have been injected, followed by a regeneration solution. The figure shows the superposition of six sensorgrams (omitting the regeneration stage). (d) Equilibrium response plotted against analyte concentration, interpreted using (9.14), and used to find the values of K_a and R_{max} .

the analyte (R negligible compared with R_{max}) to the analyte concentration. Figure 9.8 simulates the injection phases of an analyte under conditions close to total mass transport. Note that the reaction rate is constant during the analyte injection period (at least initially), in contrast to what is observed under kinetic conditions, where dR/dt falls off as fewer free ligand sites remain available (see Fig. 9.6).

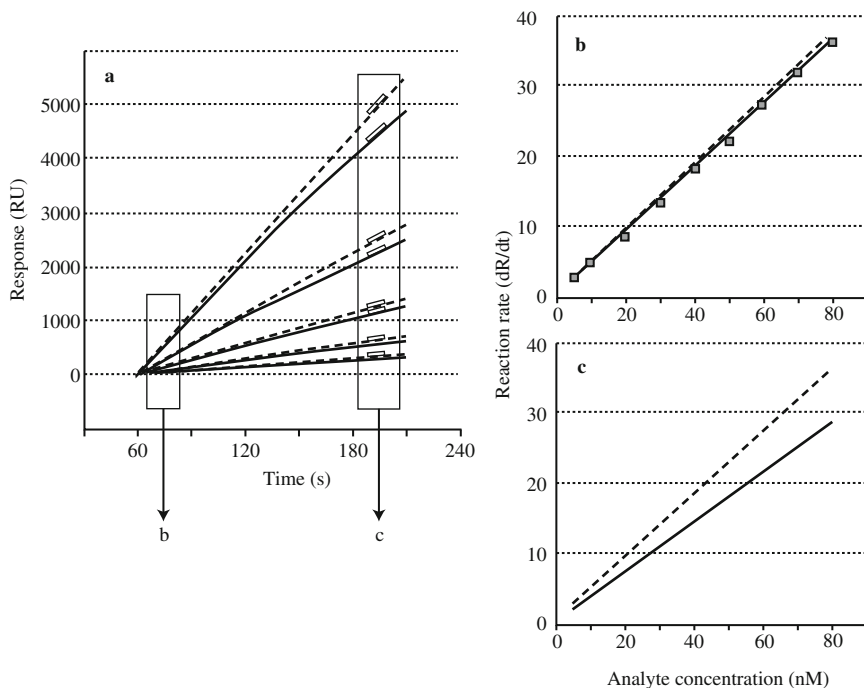


Fig. 9.8. Calibration curve. Data are simulated for two interactions, with affinities $K_a = 10^6 \text{ M}^{-1}$ (continuous curves) and 10^{10} M^{-1} (dashed curves), with $k_a = 10^5 \text{ M}^{-1}\text{s}^{-1}$, and $k_d = 10^{-1}\text{s}^{-1}$ and 10^{-5}s^{-1} , respectively. (a) Superposition of corresponding sensorgrams under injection of 5 solutions (at 5, 10, 20, 40, and 80 nM) of analyte (MM 50,000 Da) on a surface carrying a high density of ligand ($R_{\text{max}} = 30,000 \text{ RU}$), at a flow rate of $10 \mu\text{L}/\text{min}$ and at 25°C . The reaction rate is calculated either 10 s or 190 s after the beginning of the analyte injection. (b) Reaction rate dR/dt , calculated 10 s after injection, plotted as a function of the injected concentration for interactions with affinities 10^6 M^{-1} (continuous curves) and 10^{10} M^{-1} (dashed curves). The reaction rate is proportional to the concentration of injected analyte and the two straight lines can be superposed. White squares represent real experimental data obtained under these conditions. This calibration curve is used to measure the active analyte concentration in an unknown sample. (c) Reaction rate dR/dt , calculated 190 s after injection, plotted as a function of the injected concentration. The curves obtained for interactions with affinities 10^6 M^{-1} and 10^{10} M^{-1} can no longer be superposed, because total mass transport conditions are no longer fulfilled for the low-affinity interaction

Adjusting Experimental Conditions

Experimental conditions are controlled by adjusting the ratio $k_a \times R_{\text{max}}/k_t$. We have seen that the transport coefficient k_t depends on the diffusion coefficient of the analyte, the geometry of the flow cell, and the flow rate of the solvent.

Among these parameters, those that can be adjusted are the temperature and viscosity of the solvent, which affect diffusion of the analyte, and the flow rate. The first two parameters are generally fixed in biological measurements, to maintain uniformity between experiments. k_t depends on the cube root of the flow rate. Increasing the flow rate from 10 to 100 $\mu\text{L}/\text{min}$ only increases the transport coefficient by a factor of 2.15 at 25°C. It is thus the quantity $k_a \times R_{\text{max}}$ that is usually adjusted in order to obtain either kinetic or mass transport conditions. The density of immobilised ligands, which determines R_{max} , is reduced to obtain kinetic conditions and increased to obtain mass transport conditions.

Two methods are used to determine whether the interaction is affected by mass transport. The first involves injecting an analyte solution onto the ligand surface at different flow rates. Under kinetic conditions, the observed reaction rate (RU/s) is independent of the flow rate, i.e., independent of k_t , whereas in mass transport conditions, it varies with the flow rate. Alternatively, the effect of mass transport can be calculated if the rate parameters of the interaction are known. Figure 9.9 shows a simulation, illustrating the sensorgrams that would be obtained for different values of k_a and k_d at low ligand density ($R_{\text{max}} = 200 \text{ RU}$).

9.1.4 Other Experimental Adaptations

Quantitative assessment of SPR data means obtaining curves that strictly reflect the specific interaction of a homogeneous ligand and analyte. The design of an SPR experiment to generate such data has been described in detail [8,14]. We shall summarise the main features, i.e., elimination of all non-specific signals and control over the state and concentration of the molecular partners.

Eliminating Non-Specific Signals

Reference Surface

The SPR response depends on the mass of the molecules at the sensor surface, but also on other factors, such as the pressure, the temperature, and the buffer composition. The reference surface is used to remove the signal generated by these factors (the non-specific signal) from the total signal, and thereby obtain a specific signal corresponding to the analyte–ligand interaction. An ideal reference surface mimics the physicochemical properties of the ligand surface (mass of the molecules, charge, hydrophobicity, etc.), but not its binding properties with the analyte. This ideal situation can rarely be achieved. The reference is often a surface that has undergone the same treatment as the ligand surface, apart from the immobilisation of the latter. When the ligand surface contains a significant mass of molecules (>500–1,000 RU), it is better to produce a reference surface containing an equivalent amount of material, but inactive with respect to the relevant analyte. For example, an

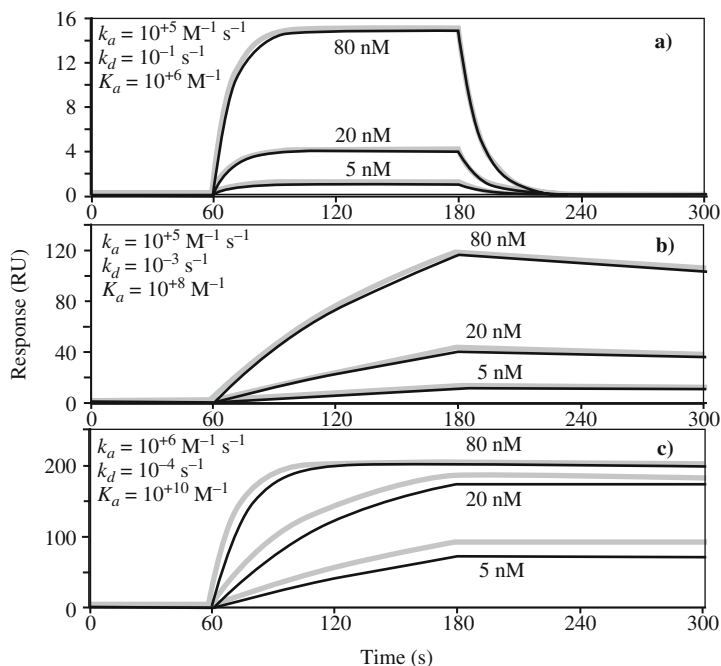


Fig. 9.9. Simulated sensorgrams. Curves simulated using the BIA simulation software [12] represent the superposition of sensorgrams corresponding to injection of three different concentrations of analyte (5, 20, and 80 nM) on a ligand surface ($R_{\text{max}} = 200 \text{ RU}$). The analyte injection and post-injection phases last 2 min. The analyte has an MM of 25,000 Da, the flow rate is 30 $\mu\text{L}/\text{min}$, and the temperature is 25°C. *Black curves* take into account mass transport and correspond to the data expected in a Biacore experiment. *Grey curves* illustrate what the reaction rate would be under kinetic conditions. A superposition of grey and black curves indicates that kinetic conditions are fulfilled for this experiment, which is the case for examples (a) and (b). However, an interaction characterised by $k_a = 10^6 \text{ M}^{-1} \text{ s}^{-1}$ and $k_d = 10^{-4} \text{ s}^{-1}$ would be affected by mass transport, as in case (c). The concentration of immobilised ligand must then be reduced in order to achieve kinetic conditions

antibody with a different specificity to that of the ligand surface, or a nucleic acid with a different sequence from that of the nucleic acid used as ligand.

Figure 9.10 shows a complex experiment, designed to analyse the ability of a series of peptides to bind to the oncoprotein E6 [13]. The protein cannot be bound covalently to the sensor surface owing to its fragility (deactivation by binding and/or regeneration solutions). The covalent binding of the peptides would involve using a large number of sensor surfaces. The peptides have thus been expressed in recombinant form, fused with glutathione-S-transferase (GST), which can be captured by anti-GST antibodies immobilised covalently on the surface. The regeneration stage eliminates the analyte and

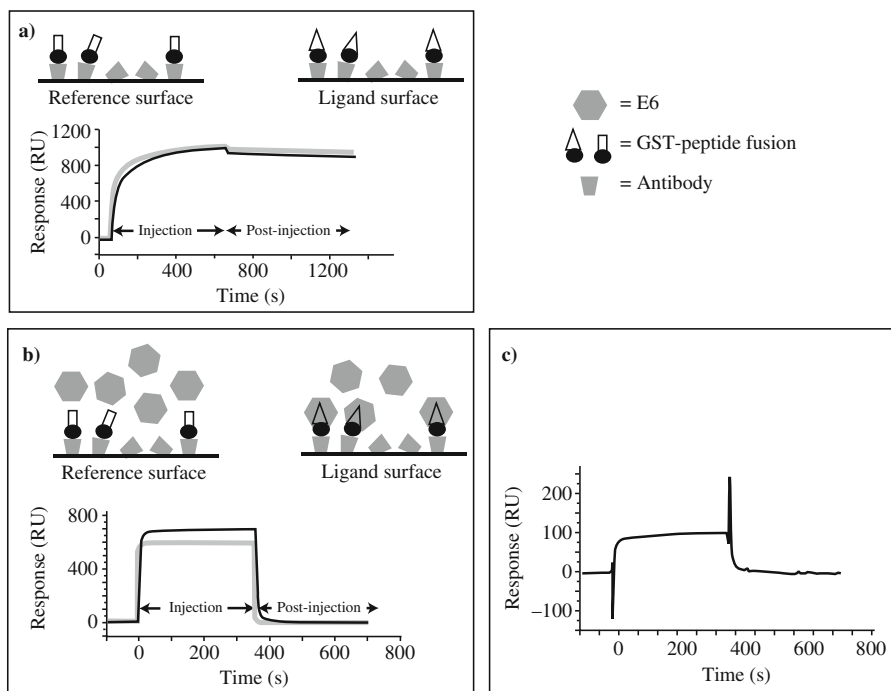


Fig. 9.10. Subtracting the signal obtained on the reference cell. Anti-GST antibodies have been immobilised on the surface. (a) A blank GST-peptide and a positive GST-peptide are injected onto reference and ligand surfaces, respectively. Superposition of the corresponding sensorgrams shows equivalent response levels (mass of proteins at the surface) of about 1,000 RU. (b) The analyte, at a concentration of $10\ \mu\text{M}$, is injected onto the two surfaces. The superposition of the sensorgrams shows a larger response on the ligand surface than on the reference surface. (c) After deduction of the ligand and reference signals, a sensorgram reflecting the interaction between the analyte and the positive peptide-GST is obtained

the GST-peptide ligand, so that the whole library of GST-peptide fusions can be screened on the same surface. In these experiments, the surfaces contain a relatively large amount of proteins: 15,000–18,000 RU of antibodies and 500 to 1,000 RU of GST-peptide. A surface without protein, or carrying only the antibody, does not provide a suitable reference. The latter has been realised by injecting the analyte onto an antibody surface that has captured a ‘blank’ GST-peptide fusion, the latter corresponding to a sequence that is not recognised by the oncoprotein E6. This example illustrates how to generate high quality kinetic data like those in Fig. 9.7c with a difficult molecular system (fragility of the protein, low affinity of the interaction).

Double Referencing

A single reference surface is sufficient in most cases to obtain interpretable data. However, the data can be further improved by subtracting a second reference, namely, the signal obtained by injecting a solution without analyte onto the ligand surface. This procedure can be used to eliminate low noise levels generated by instrumentation, e.g., the motion of valves, effects that are extremely reproducible for a given method and cell (injection time and volume, flow rate, and number of identical injections) [14].

DMSO Calibration

Dimethyl sulfoxide or DMSO is widely used to solubilise small molecules, thereby generating a weak signal when they bind to the ligand. DMSO causes a significant change in the refractive index compared with the expected weak specific signal. The change in refractive index due to the DMSO is also slightly different on different surfaces. A DMSO calibration involves measuring the change in refractive index as a function of the injected DMSO concentration, and this on each surface used. The data obtained on the ligand surface are doubly corrected: by subtraction of the signal recorded on the reference surface, and by compensation of the difference in DMSO response on the two surfaces [15, 16].

Non-Specific Interactions

The signals from non-specific interactions are generally weaker in a biosensor than in conventional assays, which involve the addition of a certain number of substances for revealing complex formation, with the consequence that the background noise gets amplified. If a non-specific signal, coming for example from ionic or hydrophobic bonds, is observed in the reference channel, it is important to apply simple measures for cancelling or at least reducing it. Precise quantitative interpretation of the data will then be optimal, even for a weak specific signal. For example:

- Purify the analyte if the non-specific signal comes from contaminating molecules.
- Avoid high molecular concentrations (risk of protein denaturation or aggregation).
- Reduce negative charges in the matrix (when the injected solution contains positively charged molecules).
- Adapt the buffer, e.g., increase the salt concentration to screen charges, add a detergent to prevent molecules from aggregating or adsorbing onto the channels and substrates, etc.

These procedures have been applied to the E6-peptide interaction [13] illustrated in Figs. 9.7 and 9.10, to eliminate any non-specific binding in the reference channel. The SPR signal observed on this channel in Fig. 9.10b is due solely to the change of buffer.

Controlling the Molecules

In order for the above equations to be valid, the ligand and analyte must be homogeneous and the analyte monovalent. The presence of molecular subpopulations with different binding affinities will generate complex kinetics. Multivalence or aggregation of the analyte leads to avidity effects: the analyte anchored to several points on the surface with repeated ligand sites tends not to dissociate, and this leads to an apparent affinity up to 500 times higher than the actual affinity [17]. The ligand must also be pure if it is covalently immobilised, whereas purification of the analyte is not required.

The characterisation and control of the molecular species under investigation plays a major role in the collection of interpretable data. The rules for preparing samples are specific to each molecular system and go beyond the scope of this chapter. However, two of the points discussed below, immobilisation and regeneration of surfaces, contribute to the homogeneity of ligand sites. Finally, quantitative measurements require adjustment and/or knowledge of the molecular concentrations.

Surface Regeneration

Regeneration prepares the ligand surface for reuse, which is desirable not only for practical reasons, e.g., saving time and materials, but also because it provides a way of carrying out experiments under similar conditions. The regeneration stage must eliminate any substance adsorbed onto the surface, and this without affecting the activity of the ligand (see Fig. 9.11). Identifying a suitable regeneration solution may be a limiting factor when determining experimental conditions, e.g., if the immobilised ligand is sensitive and/or when the interaction is difficult to dissociate. When no suitable regeneration solution is identified, an alternative experimental design consists in capturing the ligand via a first molecule, usually an antibody, that has been immobilised covalently, as shown in Fig. 9.10. In this case, the regeneration stage eliminates the captured ligand and the analyte, and each new cycle must therefore include a ligand capture stage before injection of the analyte.

The regeneration solutions most commonly used are acidic, basic, saline, chaotropic, detergent, and denaturing solutions. To avoid damaging the immobilised ligand and/or the instrumentation, these solutions are injected in pulses (contact time < 2 min). The use of mixtures of compounds has been suggested [18]. In all cases, it is essential to check for compatibility between injected solutions and instrumentation, following the recommendations of the manufacturer.

Immobilising the Ligand

Oriented immobilisation of the ligand is recommended because it ensures a uniform arrangement of binding sites. For example, a peptide can be immobilised via a single N- or C-terminal cysteine (thiol coupling), or by the

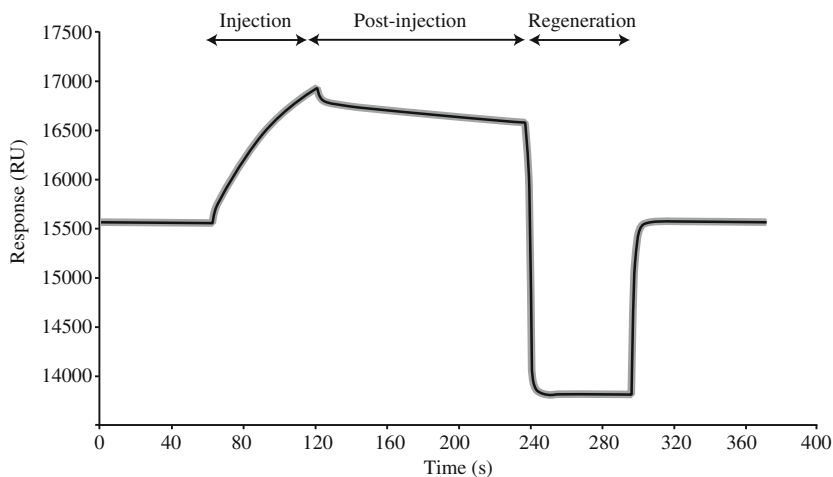


Fig. 9.11. Reproducibility of data by adequate regeneration. This figure shows the superposition of 4 sensorgrams corresponding to injection, repeated 4 times, of an analyte and a regeneration solution. The experimental conditions are the same: flow rate, analyte concentration, time and volume of injection of analyte and regeneration solution. The 4 sensorgrams overlay one another, illustrating the accurate reproducibility of the recorded data. The base line is stable (total dissociation of the analyte after regeneration) and the injection and post-injection phases are superposed (the activity of the ligand is conserved after regeneration)

N-terminus if it contains no lysine. Biotinylated molecules in a single position can be coupled via streptavidin. A ligand captured by a first molecule can also be oriented, if the capture molecule is itself homogeneous, e.g., a monoclonal antibody.

Regarding the degree of immobilisation of the ligand, since the response is proportional to the mass of bound molecules, the relation between the maximal SPR response expected at saturation ($R_{\max}$) and the amount of immobilised ligand (R_{ligand}) is for proteins

$$\frac{R_{\text{ligand}}}{R_{\max}} = \frac{\text{MM}_{\text{ligand}}}{\text{MM}_{\text{analyte}}}, \quad (9.15)$$

where $\text{MM}_{\text{ligand}}$ and $\text{MM}_{\text{analyte}}$ are the molecular masses of the ligand and the analyte, respectively. This relation is used to assess the degree of immobilisation of the ligand, R_{ligand} , that must be achieved to obtain a given maximal analyte response. It also provides a way of extracting a certain number of things from the data, such as an evaluation of the interaction stoichiometry, or the state of the molecules. For example:

- An experimental $R_{\max}$ less than the calculated $R_{\max}$ suggests that a proportion of the ligand is inactive or inaccessible, or that the binding stoichiometry is not equal to unity.

- An experimental $R_{\max}$ greater than the calculated $R_{\max}$ suggests that the analyte binds in multimer form (aggregated analyte or binding stoichiometry not equal to unity).

Active Concentration of Analyte

This must be known in order to solve (9.11) when making kinetic measurements. Standard methods for measuring concentrations, e.g., optical density, Bradford, etc., give the total concentration of molecules. The molecular preparations can contain a non-negligible fraction of inactive molecules [19], and this leads to an overestimate of the analyte concentration and an underestimate of the value of k_a . There are few methods for measuring active concentrations that are independent of an internal reference, i.e., prior knowledge of the concentration in a preparation assumed to be 100% active. One possibility is to use (9.7) to simulate the initial reaction rate for different analyte concentrations, under total mass transport conditions, and at a given flow rate. The theoretical calibration curve, as shown in Fig. 9.8b, relating the reaction rate to the active concentration is then used to measure the active concentration in a sample injected at the same flow rate, on a surface under total mass transport conditions.

9.1.5 Applications

Since the SPR response is based on a change in molecular mass, this technology can in principle be applied to all molecular interactions and any type of molecule, e.g., proteins, nucleic acids, sugars, lipids, or small molecules. Many applications of Biacore technology have been described since the first device was put on the market (reviews by Myszkka [20], Rich and Myszkka [21–25], Van Regenmortel [26], Homola [27], Karlsson [28]), and in many different fields, such as virology [29, 30], bacteriology [31], the food industry [32], host–pathogen interactions [33], immunology [34], drug design [35], or proteomics [36, 37]. SPR can characterise biomolecular interactions by determining their stoichiometry, kinetics, thermodynamics, and affinities. It is also used to map epitopes, to study protein self-assembly mechanisms, and to fish for ligands in a complex biological medium, i.e., to capture the partners of a given molecule and identify them using mass spectrometry (BIA-MS). This is one of the qualitative or semi-quantitative applications of SPR, which includes assays to probe interactions, screening of libraries of small molecules, and the study of interactions between cells and an immobilised ligand. Current developments concern four main areas: the food industry, immunogenicity, proteomics, and drug discovery [28]. SPR is also used as a detection method for protein arrays (see Chap. 18).

The aim here is not to give a comprehensive overview of all the fundamental and applied research carried out using the Biacore technology, all of which the reader can find in recent reviews [20–28]. Most of this section is devoted to recent developments in different areas of application of this technology.

Protein–Protein Interactions

Many studies have focused on protein–protein interactions for proteins with different sizes and functions. Although many recent developments with SPR analysis concern small molecules, this technology can also be used to determine rate and affinity constants for interactions involving macromolecules such as the procollagen molecule which has a molecular mass of around 450 kDa [38]. A lot of work has been devoted to antigen–antibody interactions, and to the characterisation of antibodies. The development of immunological assays based on SPR is reviewed in [34]. Protein–protein interactions play a crucial role in many reactions of the immune system, and the SPR technique has shown for example that the protein C1q of the blood complement recognises the immobilised prion protein and that, in this immobilised form, the prion protein undergoes a conformational change similar to the change that occurs when the normal form of the prion protein converts to the pathogenic form [39]. SPR has been used to analyse the kinetics and thermodynamics of the interaction between an enzyme and an inhibitor [40], enzymes being biological catalysts that are generally proteins. Mapping interaction sites on macromolecules requires the immobilisation on sensor chips, not of whole molecules, but rather their domains, to locate the binding site(s) in the macromolecule [41]. The amino acids involved in the interactions are identified by studying mutated proteins [42]. The effect of the state of oligomerisation of the proteins on their ability to interact with their ligands can also be determined by SPR [43].

Supramolecular assemblies such as amyloid fibrils, which form in the brains of patients suffering from Alzheimer's disease [44–46], require protein–protein interactions. Kinetic data have been obtained with amyloid fibrils immobilised on a sensor chip. The amyloid peptide A β (1–40) is injected in soluble form and binds on the sensor chip. The results obtained show that fibril extension is described by a polymerisation model involving three stages, of which the first is reversible [46]. SPR has also been used to identify small molecules likely to inhibit aggregation and cell toxicity of the peptide A β and to characterise its interactions with different ligands. It has been shown that the ligands of the peptide A β which have the strongest affinity for the peptide affect its aggregation and confer the greatest protection against cell toxicity [47]. Applications of SPR to the interactions involved in the formation of amyloid fibrils and their impact on molecular mechanisms in Alzheimer's disease can be found in two review papers [48, 49].

SPR and Protein Structure

One application of the Biacore technology is in the study of structure–function relationships in proteins [26]. Several groups have used SPR and in particular the Biacore technique to analyse conformational changes in immobilised proteins. Sota et al. [50] developed a methodology using SPR to monitor the

conformational changes arising during denaturation of an immobilised protein subjected to an acidic pH. The combined use of SPR and monoclonal antibodies can detect conformational changes in the protein ovalbumin when it is subjected to irradiation [51]. The change in conformation induced by the chaotropic agent guanidinium chloride has been demonstrated by SPR using Triton X100 as a hydrophobic probe to monitor protein unfolding, i.e., the loss of native structure. This detergent adsorbs onto the hydrophobic part of the protein, which is exposed during unfolding, and SPR measurements of its adsorption are used to evaluate the state of unfolding of the protein [52]. The refolding of a protein can also be monitored by SPR, as has been demonstrated for the enzyme luciferase [53]. The Biacore technology may be used to optimise the refolding conditions of an immobilised protein. Mannen et al. [54] have studied the effect of buffers of different pH on the behaviour of immobilised proteins and have shown that modifications in the SPR signal are related to the overall charge of the protein and to conformational changes occurring at different pH. It should nevertheless be noted that, according to the work of Paynter and Russell [55], the SPR signal recorded when immobilised proteins are exposed to different pH is due rather to electrostatic interaction effects. Using circular dichroism and Fourier transform infrared spectroscopy, May and Russell [56] have established that modifications in the SPR signal can be correlated with the secondary structure of immobilised polypeptides (poly-L-lysine, poly-L glutamic acid and a protein, concanavalin A). A decrease in the SPR signal corresponds to a high level of β structures, turns, or disordered structures, whereas an increase in the signal reflects a significant level of α helices [56]. SPR can also monitor in real time the slow reversible conformational transition that occurs in an enzyme during the binding of its substrate. This has been shown for the binding of phenylalanine on phenylalanine hydroxylase [57].

Nucleic Acid–Protein Interactions

Interactions between therapeutic agents and target DNA molecules or transcription factors are particularly important for drug design and development, because these interactions control gene expression. Transcription factors are proteins regulating the transcription of a gene, i.e., the synthesis of an RNA molecule from a DNA molecule, by binding onto a promoter. The review published by Gambari [58] describes SPR-based approaches used to develop oligonucleotides forming a triple helix, analogues of DNA (peptide nucleic acids), drugs binding to DNA, and decoy molecules able to interfere with transcription.

RNA–protein complexes are dynamic and their kinetics, which is important for the assembly of ribonuclear protein complexes and the ability of RNA to migrate from one cell compartment to another, has been monitored by SPR [59]. Biacore technology is also used to characterise DNA–DNA and DNA–RNA interactions. SPR is also used to monitor hybridisation between

a target in solution and a probe immobilised on the sensor chip, and thereby detect a DNA sequence amplified by polymerase chain reaction (PCR), i.e., amplification of a small DNA fragment by action of a polymerase, specific primers, and nucleotides in a thermocycler [60].

Protein–Sugar Interactions

A review describing the use of SPR in glycobiology reports work based on the immobilisation of sugars, glycoconjugates, or lectins, which are proteins with a non-catalytic domain for reversible binding to specific mono- or oligosaccharides [61]. SPR provides information about the specificity of lectins and the structure of oligosaccharides involved in the interactions. SPR is used to characterise interactions of proteins with simple sugars and complex sugars like glycosaminoglycans, which are linear polymers with disaccharide units. Several studies have been carried out with heparin or heparan sulfate to determine the minimal size and structural characteristics of the glycosaminoglycans needed for their interaction with proteins [62,63]. The characterisation of the interactions of several enzymes with sulfated polysaccharides has shown a good correlation between affinities determined by SPR and the inhibiting activity of these polysaccharides measured by a functional test [64]. Methods have also been devised to characterise the interactions between proteins and sugars associated with lipids (glycolipids) in an artificial membrane environment using liposomes to mimic their natural environment, which is the lipid bilayer of the plasma membrane [65]. The relevant proteins are injected at the surface of the liposomes captured at the surface of a sensor chip and containing different glycolipids.

Interactions in a Membrane-Mimicking Environment

SPR is also used to study the binding of proteins to their receptors [66] and more generally to investigate ligand–receptor interactions [67]. The interactions between membrane receptors and their ligands, together with the interactions between small molecules and membrane constituents such as lipids, are of great interest for the development of new therapeutic molecules. The importance of biosensors based on SPR in drug design has been stressed in a recent review [35]. Several recent papers report on interactions between drugs and immobilised liposomes [68,69]. One of these concerns 78 drugs, including inhibitors of kinases, thrombin, and carbonic anhydrase [69]. The interactions of protein kinases, which are major therapeutic targets, with small molecules which are potential inhibitors, have also been investigated [70]. The SPR study carried out with immobilised enzymes has confirmed the inhibitive action of the molecules analysed, characterised the kinetics of their interaction with the protein kinases, and identified competitive inhibitors by carrying out competition experiments with ATP [70].

SPR can also provide useful data for assessing the intestinal permeability of drugs and predicting their oral absorption by humans, important parameters in drug development. Kim et al. [71] have developed an SPR method for investigating interactions between a drug and the specialised plasma membrane of intestinal epithelial cells, called the brush border membrane. They immobilised brush border vesicles isolated from tissue on an L1 sensor chip and injected 14 drugs on the surface. A hundred compounds can be tested at a single concentration in 24 h under the experimental conditions used [71].

Interactions with Micro-Organisms and Eukaryotic Cells

The ligand immobilised on a sensor chip must be purified, but the analyte can be a purified molecule, a complex biological medium such as serum, urine, hybridoma supernatants [72], subcellular fractions [73], a virus, a bacterium like *Salmonella* [31], or living cells. Indeed work has been carried out in which whole cells are injected onto immobilised ligands. Under these conditions, it is not possible to determine the rate and affinity constants of the interactions, because only a fraction of the cells is located in the evanescent field, but this qualitative approach can determine whether the receptors present at the surface of the injected cells are able to bind to the immobilised ligand. Interactions between cells and sugars have been studied, e.g., Jurkat cells expressing L-selectin, an adhesion molecule, with a derivative of galactose [74]. Interactions between cells and proteins have also been characterised, e.g., between polynuclear neutrophils and collagen I [75], between a neuronal cell line and the soluble form of the β -amyloid precursor protein [76], or again between erythrocytes and M specific immunoglobulins [77] to detect blood group antigens. We are currently studying the interaction of Chinese hamster ovary (CHO) cells with polysaccharides and with receptors of the integrin family (Ricard-Blum et al., unpublished results).

Micro-organisms can be immobilised on a sensor chip in order to study their interactions with different partners. Abad et al. [78] have immobilised adenoviruses and checked by electron microscopy that the viruses were intact after immobilisation. They were used to monitor the immune response of patients treated by gene therapy based on an adenovirus vector and to carry out isotyping of the detected antibodies. This approach could be applied to study the interaction of a virus with its receptor, or a virus with an antiviral drug.

RaPID Plot Isoaffinity Curves

Rate parameters determined by Biacore technology for a series of compounds, e.g., potential inhibitors, can be represented in the form of 2D diagrams in which the horizontal axis shows the values of the dissociation rate constants and the vertical axis those of the association rate constants. Compounds with

the same affinity constant are located on the same diagonal [79]. This representation of the data is called a RaPID plot, short for rate plane with isoaffinity diagonals. It has been used to detect variants of the heavy chain of antibodies directed against the prostate-specific antigen [80] and to characterise camel antibodies at different temperatures [81]. This representation is useful for selecting molecules with the desired kinetic characteristics among molecules with a given affinity.

Concentration Measurements

There are many applications in the food industry for concentration measurements using Biacore technology, especially for detecting additives or molecules like hormones, vitamins, and antibiotics. This approach gives very fast results, once appropriate experimental conditions have been optimised, and this is a decisive advantage in this field. For example, SPR is used to measure the concentration of vitamins B5 [82] and B12 [83] in foods, and proteins binding folate in milk [32]. It can also be used to detect cows' milk in the milk of ewes or goats [84], pathogenic agents like the one causing listeria, i.e., *Listeria monocytogenes* [85], hormones in milk [86] and in animal tissues such as meat [87], β -lactam antibiotics in milk [88], or another antibiotic, chloramphenicol, in foodstuffs [89]. This technology can also be used for medical purposes to determine the active concentration of recombinant proteins under investigation in clinical studies [90] or the concentration in biological fluids of molecules involved in pathological processes. Interleukin-8, a cytokin, has been assayed by SPR in the saliva of patients suffering from cancer, with a detection limit of 184 nM [91].

9.2 Ellipsometry

9.2.1 Introduction

Measurements on the angstrom scale require extremely accurate techniques. Ellipsometry involves measuring the change in polarisation state of light when it reflects off a surface, and constitutes a very accurate technique for analysing ultrathin layers. It is used to probe monolayer and multilayer thicknesses from a few angstroms to several tens of micrometers. A broad range of properties of materials is also accessible using this technique, such as optical constants, the optical gap, chemical composition, crystallinity, and the depth and surface uniformity of films (roughness, porosity, interface, index gradient, anisotropy, and so on).

Owing to its popularity as a laboratory technique, manufacturers have greatly contributed to its simplification and automation. Ellipsometry has been part of the industrial scene for a decade or so now and is exploited in various stages of production, from optimisation and validation of processes to

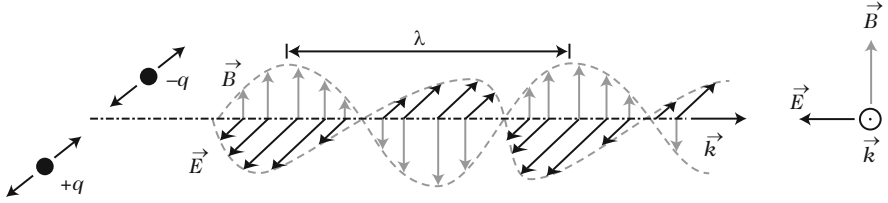


Fig. 9.12. Electromagnetic wave

quality control on the production line. Used in areas as varied as electronics, telecommunications, flat screens, and optical films, it is a powerful and versatile tool for applications in bio- and nanotechnologies.

As a non-destructive technique, its sensitivity is unequalled, with spot sizes as small as $10\mu\text{m}$. It is a modular technique, working in situ and adapting easily to any type of reactor, where it can monitor growth or etching of films and carry out fast kinetic measurements.

9.2.2 Theory of Light and Polarisation

Description of Electromagnetic Waves

Light is an electromagnetic wave that propagates through space. It comprises an electric field $\mathbf{E}$ and a magnetic field $\mathbf{B}$, both perpendicular to the direction of propagation (see Fig. 9.12). Since the electric and magnetic fields are related by Maxwell's equations, the wave can thus be represented by just one of these fields, generally the electrical field.

The general equation for a plane wave is

$$\mathbf{E}(\mathbf{r}, t) = e^{i(\omega t - \mathbf{k} \cdot \mathbf{r} + \varphi)} \mathbf{E}_0,$$

where $\mathbf{r}$ is the position vector of the field point, $\mathbf{k}$ is the wave vector, with magnitude $2\pi c/\lambda$, c is the speed of light, λ is the wavelength, and φ is the phase at the origin.

Assuming that $\mathbf{k}$ is oriented along the z axis, the electric field components of the wave have the form

$$E_x(z, t) = A_x \cos(\omega t - kz + \varphi_x), \quad E_y(z, t) = A_y \cos(\omega t - kz + \varphi_y).$$

Properties of Electromagnetic Waves

Different Polarisation States

Polarisation is one property of light waves. The polarisation of a wave refers to the behaviour of the electric field vector representing the wave as time goes by,

at a given point in space. Decomposing it into its two orthogonal components, its time dependence is strictly sinusoidal. Polarising a wave means giving a definite trajectory to the end of the electric field vector. The polarisation state is determined by two parameters, the relative phase and amplitude of the components of the electric field $\mathbf{E}$:

- *Linear Polarisation.* Light is linearly polarised if the phase difference

$$\varphi = \varphi_y - \varphi_x$$

is $\varphi = 0$ or $\varphi = \pi$. E_x and E_y are then proportional:

$$E_y = \frac{A_y}{A_x} E_x ,$$

for $\varphi = 0$ and

$$E_y = -\frac{A_y}{A_x} E_x ,$$

for $\varphi = \pi$, which is the equation for a straight line.

- *Circular Polarisation.* In the special case where the phase difference is 90° , $\varphi = \pm\pi/2$, and the two components have the same amplitude, $A_x = A_y = A$, the polarisation is circular. Then

$$E_x = A \cos(\omega t - kz + \varphi_x) ,$$

$$E_y = A \cos(\omega t - kz + \varphi_x \pm \pi/2) = \pm A \sin(\omega t + \varphi_x) .$$

The direction of rotation is left-handed about the z axis if $\varphi = -\pi/2$ and right-handed if $\varphi = +\pi/2$. In the first case, the polarisation is said to be left circular, and in the second, right circular.

- *Elliptical Polarisation.* If A_x and A_y are different, but φ remains equal to $\pm\pi/2$, then

$$E_x = A_x \cos(\omega t - kz + \varphi_x) , \quad E_y = \pm A_y \sin(\omega t - kz + \varphi_x) .$$

Hence,

$$\frac{E_x^2}{a^2} + \frac{E_y^2}{b^2} = 1 ,$$

setting $a^2 = A_x^2$ and $b^2 = A_y^2$, where $2a$ and $2b$ are the lengths of the major and minor axes of the ellipse, respectively. The polarisation is right elliptical if $\varphi = \pi/2$ and left elliptical if $\varphi = -\pi/2$. The general case is obtained for

$$E_x = A_x \cos(\omega t - kz + \varphi_x) , \quad E_y = A_y \cos(\omega t - kz + \varphi_y) .$$

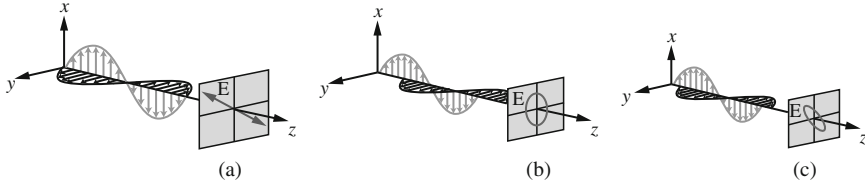


Fig. 9.13. Different polarisation states of light: (a) linear, (b) circular, and (c) elliptical

In the general case where $\varphi \neq \pm\pi/2, \pm\pi$, the components $E_x(z, t)$ and $E_y(z, t)$ of the light satisfy

$$\left(\frac{E_x}{A_x}\right)^2 + \left(\frac{E_y}{A_y}\right)^2 - 2\frac{\cos\varphi}{A_x A_y} E_x E_y = \sin^2\varphi, \quad \varphi = \varphi_y - \varphi_x.$$

This is the equation for an ellipse relative to the x and y axes. As can be seen from the illustrations in Fig. 9.13, the end of the electric field vector then traces out an ellipse, which can become a circle or flatten into a straight line.

These three patterns define the three polarisation states of the wave, which is said to be elliptically polarised (Fig. 9.13c), circularly polarised (Fig. 9.13b), or linearly polarised (Fig. 9.13a).

Light–Matter Interaction. Wave Behaviour

In a homogeneous and isotropic medium, the wave propagates in a straight line. When the propagation medium changes, there is reflection and refraction. Part of the electromagnetic wave returns to the original medium: this is reflection. The other part of the wave propagates into the second medium with a change in direction: this is refraction. If the medium is absorbent, the wave will be attenuated as it propagates. The Snell–Descartes laws describe the behaviour of light at the interface between two media (see Fig. 9.14).

Each medium is characterised by its ability to slow light down, as modelled by its refractive index n , which is given by

$$n = \frac{c}{v},$$

where v is the speed of light in this medium and c is the speed of light in vacuum.

The law of reflection states that:

- the reflected ray lies in the plane of incidence defined below,
- the angles of incidence θ_1 and reflection θ_2 satisfy $\theta_2 = -\theta_1$.

The law of refraction states that:

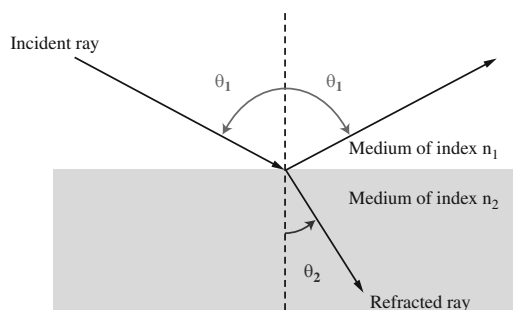


Fig. 9.14. Interaction of light with a plane interface separating two different media

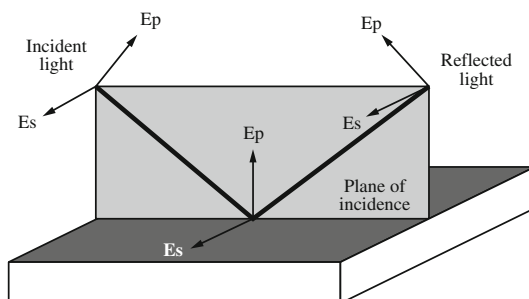


Fig. 9.15. Definition of the plane of incidence and components of the incident and reflected electric fields

- the refracted ray lies in the plane of incidence,
- the refractive indices n_1 and n_2 of each of the media and the angles of incidence θ_1 and refraction θ_2 are related by the Snell–Descartes $n_1 \sin \theta_1 = n_2 \sin \theta_2$.

Reflection of Light

Defining the Plane of Incidence

Ellipsometry is a technique based on analysis of the light reflected from a surface. When reflection occurs following oblique incidence, the plane of incidence is defined by the incident beam and the normal to the sample surface (see Fig. 9.15). The incident vibration can be decomposed into two components E_p^i and E_s^i , parallel (p) and perpendicular (s) to the plane of incidence, respectively. These two vibrations undergo reflection and yield the reflected components E_p^r and E_s^r , respectively.

Fresnel Reflection Coefficients

The Fresnel coefficients describe reflection and refraction effects at the interface between two media with different refractive indices. They are obtained

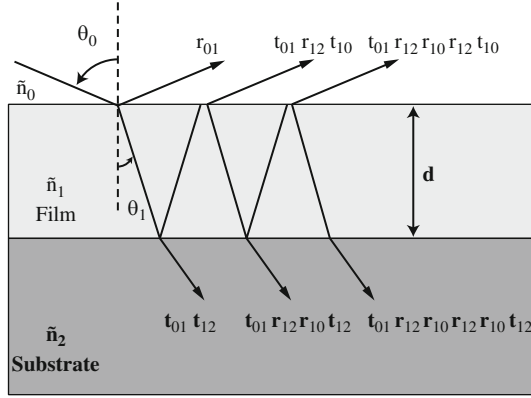


Fig. 9.16. Multiple reflection of light at the surface of a substrate coated with a film of thickness d

by considering the relations expressing continuity of the electromagnetic field across the interface:

$$r_p = \frac{n_2 \cos \theta_1 - n_1 \cos \theta_2}{n_2 \cos \theta_1 + n_1 \cos \theta_2}, \quad r_s = \frac{n_1 \cos \theta_1 - n_2 \cos \theta_2}{n_1 \cos \theta_1 + n_2 \cos \theta_2},$$

with the respective characteristics of incidence and refraction, (n_1, θ_1) and (n_2, θ_2) . In a practical situation, most materials are absorbent and their refractive index is complex, i.e., $\tilde{n} = n - ik$.

The Fresnel coefficients can also be expressed in the form of a ratio between the amplitudes of the reflected and incident electric fields:

$$r = \frac{E_r}{E_i}, \quad t = \frac{E_t}{E_i},$$

where E_i , E_r , and E_t are the amplitudes of the incident, reflected, and transmitted fields, respectively.

Reflection of Light by a Thin Film

The above treats the simple case when light is reflected at an air/substrate interface. The following model (see Fig. 9.16) treats a three-phase system including a substrate with index $\tilde{n}_2$, coated with a film of thickness d and index $\tilde{n}_1$, immersed in a surrounding medium of index $\tilde{n}_0$. We assume the ideal case in which each of the media is homogeneous and isotropic.

The overall reflection coefficient of this system is the sum of all the multiple reflections at the interfaces 1 and 2, i.e., r_{01} and r_{12} :

$$r_{\text{total}} = \frac{r_{01} + r_{12} \exp(-2i\beta)}{1 + r_{01}r_{12} \exp(-2i\beta)},$$

where

$$\beta = \frac{2\pi d}{\lambda} \tilde{n}_1 \cos \theta_1$$

is the phase shift introduced by reflection between two consecutive rays.

Taking into account the polarisation of the incident wave along the s or p axis, it follows that

$$r_{\text{total}}^s = \frac{E_{s\text{ total}}^r}{E_s^i} = \frac{r_{01}^s + r_{12}^s \exp(-2i\beta)}{1 + r_{01}^s r_{12}^s \exp(-2i\beta)} ,$$

$$r_{\text{total}}^p = \frac{E_{p\text{ total}}^r}{E_p^i} = \frac{r_{01}^p + r_{12}^p \exp(-2i\beta)}{1 + r_{01}^p r_{12}^p \exp(-2i\beta)} .$$

This single-layer model can be generalised to a sample comprising n layers. The problem is treated with the help of the Abélès matrix formalism, used to calculate the overall Fresnel coefficient for this type of multilayer, expressing the result in terms of the dielectric function of each layer. It is based on the fact that the equations for the propagation of a wave in a medium are linear and the continuity of the fields at the interface between two isotropic media can be expressed via a transfer matrix.

Effective Medium Theory

In the above, we have been considering homogeneous and isotropic layers, bounded by parallel, plane surfaces, an ideal case that does not often correspond to the reality of thin films. The materials and surfaces encountered in real situations exhibit heterogeneities that may manifest themselves in different ways: phase mixing, oxidation, roughness, etc. Making the assumption that these defects can be treated as small compared with the wavelength, the optical response of these materials can be described by the theory of effective media.

The idea is to associate a dielectric constant locally with the heterogeneities and then, taking a spatial average, to treat the medium as being macroscopically homogeneous, thereby attributing an effective dielectric constant.

The simplest effective medium model is obtained by linear interpolation between the optical responses of materials, such that

$$\tilde{\epsilon} = f_A \tilde{\epsilon}_A + f_B \tilde{\epsilon}_B + f_C \tilde{\epsilon}_C ,$$

where $\tilde{\epsilon}$ is the complex effective dielectric constant, f_A , f_B , and f_C are the volume fractions, and $\tilde{\epsilon}_A$, $\tilde{\epsilon}_B$, and $\tilde{\epsilon}_C$ are the dielectric constants of each material making up the effective medium. The total of the volume fractions must of course be equal to unity, and the equation is valid for a three-component model, or a two-component model when f_C is equal to zero.

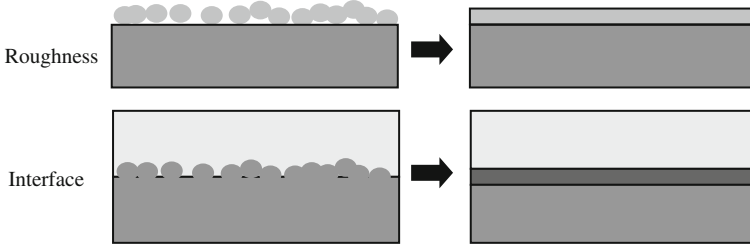


Fig. 9.17. Schematic view of a rough sample and a rough sample with an interface, together with their equivalents in the effective medium theory

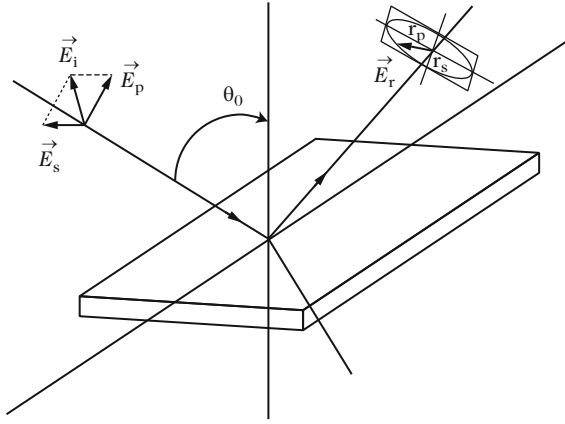


Fig. 9.18. Electric field reflected by a plane surface

There are two main effective medium theories: the Maxwell–Garnett model and the Bruggeman model, the choice being made according to the way the materials are arranged.

The Maxwell–Garnett model allows for the inclusion of one or two materials denoted by B and C , totally surrounded by a host medium denoted by A :

$$\frac{\tilde{\epsilon} - \tilde{\epsilon}_A}{\tilde{\epsilon} + 2\tilde{\epsilon}_A} = f_B \frac{\tilde{\epsilon}_B - \tilde{\epsilon}_A}{\tilde{\epsilon}_B + 2\tilde{\epsilon}_A} + f_C \frac{\tilde{\epsilon}_C - \tilde{\epsilon}_A}{\tilde{\epsilon}_C + 2\tilde{\epsilon}_A}.$$

When the inclusions are uniformly distributed, the dielectric constant of the effective medium is given by

$$\tilde{\epsilon} = \tilde{\epsilon}_h \frac{\tilde{\epsilon}_i(1 + 2f) + 2\tilde{\epsilon}_h(1 - f)}{\tilde{\epsilon}_i(1 - f) + \tilde{\epsilon}_h(2 + f)},$$

where $\tilde{\epsilon}_h$ is the dielectric constant of the host medium, $\tilde{\epsilon}_i$ is the dielectric constant of the inclusions, and f is the volume fraction of the inclusions. This model only works for inhomogeneities with granular shape distributed with a small volume fraction in the host medium.

The second model, the Bruggeman model, is not limited by the number of components in the mixture. It treats each component in an equivalent way, taking into account only its proportion relative to the whole. The dielectric constant of such a medium must satisfy

$$\sum_i f_i \frac{\tilde{\varepsilon}_i - \tilde{\varepsilon}}{\tilde{\varepsilon}_i + 2\tilde{\varepsilon}} = 0 ,$$

where $\tilde{\varepsilon}_i$ and f_i are the dielectric constant and relative concentration of each constituent, respectively (so that i is now treated as an index). This model is the most widely used in ellipsometry. It defines a globally heterogeneous medium, with no particular arrangement of the components.

In practice, the theory of effective media is very widely used to model three types of layer (see Fig. 9.17):

- Rough layer, comprising equal proportions of the material of the layer and air.
- An interface, comprising $x\%$ of material 1 and $(1 - x)\%$ of material 2.
- A polycrystalline material, where the layer comprises $x\%$ of the crystalline material and $(1 - x)\%$ of the amorphous material.

9.2.3 Basic Principles and Possibilities of Ellipsometry

Underlying Principles of Ellipsometry

Ellipsometry is an optical analytical method, exploiting measurements of the change in polarisation state of polarised light after reflection at oblique incidence from the surface under investigation (see Fig. 9.18). When a light wave interacts with a material, its polarisation state is modified. By measuring the change in polarisation between the incident and reflected wave, one can deduce something about the properties of the sample.

Reflection by a plane surface involves coefficients r_p and r_s , which characterise the new polarisation state of the beam. In the general case, the reflection coefficients are complex, having the form

$$r_p = \frac{E_p^r}{E_p^i} = |r_p| e^{i\delta_p} , \quad r_s = \frac{E_s^r}{E_s^i} = |r_s| e^{i\delta_s} .$$

In practice, the ellipsometer delivers the ratio ρ of these two coefficients, viz.,

$$\rho = \frac{r_p}{r_s} = e^{i\Delta} \tan \Psi ,$$

where $\tan \Psi = |r_p|/|r_s|$ and $\Delta = \delta_p - \delta_s$. This is the fundamental equation of ellipsometry. Ψ and Δ are the two parameters measured by an ellipsometer and which represent the change in polarisation state due to reflection. They

are called the ellipsometric angles. In the general case, the range of definition of these angles is $0 \leq \Psi \leq 90^\circ$ and $0 \leq \Delta \leq 360^\circ$. The function $\tan \Psi$ giving the ratio of the moduli of the two coefficients r_p and r_s represents the change in amplitude, and more exactly, the attenuation of the electric field components, while Δ represents the phase shift between the s and p components of the electric field introduced by reflection from the sample.

Note that the ratio ρ is used to define the complex dielectric function of the bulk material by

$$\tilde{\epsilon} = \tilde{n}_1^2 = \tilde{n}_0^2 \sin^2 \theta_0 \left[1 + \left(\frac{1 - \rho}{1 + \rho} \right)^2 \tan^2 \theta_0 \right].$$

The dielectric function depends on the angle of incidence θ_0 , the refractive index n_0 of the medium in which the measurement is carried out, and the angles Ψ and Δ through ρ . In conclusion, by measuring the ellipsometric angles Ψ and Δ for a substrate, the complex refractive index $\tilde{n}_1 = n_1 + ik_1$ of the material can be determined directly by inversion.

Possibilities of this Technique

In contrast to most optical methods, ellipsometry measures two quantities at the same time, viz., the modulus $\|\rho\| = \|r_p/r_s\| = \tan \Psi$ and the phase $\Delta = \delta_p - \delta_s$. This 2D feature is a key point, making this a highly sensitive technique. For thickness measurements, ellipsometry is often compared with reflectometry. Spectroreflectometers deliver a spectroscopic measurement of the ratio of reflected and incident intensities. They have the advantage of simple design and good spatial resolution, but they are limited by the fact that only one quantity is actually measured, namely an intensity. Furthermore, measurements are affected by fluctuations in the source and can characterise a minimal thickness of the order of 10 nm.

Ellipsometry provides a very accurate way of characterising thin films, surfaces, and interfaces. It delivers three types of information:

- Thicknesses between a few angstroms and a few tens of micrometers. The sample can be a simple monolayer deposited on a substrate or a stack of complex multilayers. This highly sensitive technique can easily detect a native layer, an interface, or surface asperities. In the latter case, many papers refer to the very good correlation between AFM and ellipsometric measurements.
- Optical constants:
 - refractive index n ,
 - extinction coefficient k .
- Properties of the material such as:
 - The composition of type III–V or type II–VI semiconductor alloys.
 - Microstructure, e.g., density or porosity of the layer, with the possibility of estimating the empty volume within a porous layer.

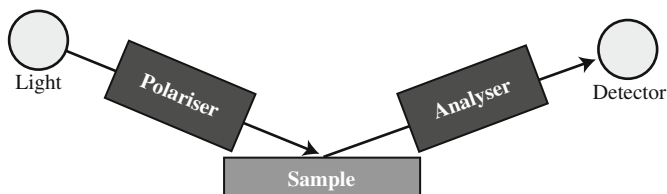


Fig. 9.19. Optical setup in an ellipsometer

- Crystallinity, with the most representative example being silicon, whose three types (crystalline, semi-crystalline, or amorphous) possess widely different optical constants.
- The optical gap E_g .
- The homogeneity of surface or buried layers. Films produced by deposition, synthesis, annealing, etc., are often inhomogeneous. Ellipsometry can characterise inhomogeneities via index gradients, anisotropy, or depolarisation phenomena due for example to excessive roughness.

The main advantages of ellipsometry are:

- it is non-invasive,
- it is highly sensitive, down to a few atomic monolayers,
- it can be used to control growth or etching in real time.

9.2.4 Instrumentation

Ellipsometer Configurations

Introduction

Three main ellipsometer technologies are available on the market. All three are based on an optical arrangement like the one shown in Fig. 9.19, with a light source, polariser, sample, analyser, and detector.

The idea of ellipsometry is to send a light beam of known polarisation (after passing through a polariser) at oblique incidence onto the sample, and then to analyse the polarisation of the reflected beam (by passing it through an analyser). Recall that the change in polarisation is due to the interaction with the sample, and it is through this that one can reconstitute its properties. Different elements can be adjoined to the setup, e.g., modulator, compensator, and this greatly contributes to the sensitivity of ellipsometric analysis. Apart from these differences in the general configuration, there are two main families of ellipsometer:

- *Laser Ellipsometer.* This measures the parameters Ψ and Δ at a single wavelength, generally that of the helium–neon laser at 632.8 nm. For two measured parameters, two quantities can be determined, such as the thickness and the refractive index of a transparent monolayer deposited on a

known substrate. These ellipsometers have the advantage of being highly accurate, but they are limited to very simple applications.

- *Spectroscopic Ellipsometer*. This uses a white light source, whence it is possible to cover a broad spectral range, from the far ultraviolet (FUV) to the near infrared (NIR), i.e., typically from 190 nm to 2,000 nm. The ellipsometric angles Ψ and Δ are measured at each wavelength, so complex structures can be characterised, e.g., stacks of multilayers, inhomogeneities in materials, etc.

Note that spectroscopic ellipsometers include a wavelength selection system which can be one of two types, either a monochromator carrying out sequential acquisition, wavelength by wavelength, or a CCD carrying out simultaneous acquisition.

Nulling Ellipsometer

These ellipsometers exploit extinction of the signal to determine the ellipsometric angles Ψ and Δ . They use a polariser followed by a compensator (usually a quarter-wave plate), which transform the linear polarisation into an elliptical polarisation. The role of the compensator is to cancel the delay introduced by reflection from the sample, in such a way as to make the polarisation linear once again. The compensator thus plays a symmetrical role with respect to the sample. Successive adjustment of the polariser and the analyser leads to extinction of the signal.

This classic technique is very accurate, but measurement is slow and difficult to automate. For this reason, these ellipsometers have been supplanted by modulation ellipsometers, which extract information from the changing intensity at the detector.

Rotating Element Ellipsometer

The polarisation can be modulated in three different ways: rotating the polariser, the analyser, or the compensator. Ellipsometers with rotating polarisers and analysers have been around for at least fifteen years now. They are well-suited to spectroscopic studies, since the response of all the elements, apart from the sample surface, is independent of the wavelength. Furthermore, these systems are sensitive to the residual polarisation of the source or the detector, but also to inhomogeneities in the polariser/analyser due to rotation of the beam. These imperfections can be reduced to some extent by calibrating the system, but they can give rise to significant errors. The rotation frequency is of the order of a few hundred hertz.

By calculating the ellipsometric angles for these two types of system, one obtains $\tan \Psi$ and $\cos \Delta$. There remains an indeterminacy with regard to the sign of Δ , which is only known up to addition of 180° . The accuracy in Δ is therefore poor in regions where Δ is equal to 0 or 180° , which correspond to applications with a transparent substrate, e.g., glass or plastics. Finally, Δ

is the most sensitive parameter to small variations, such as ultrathin films or films with small index contrast. These two technologies are therefore inaccurate for this type of advanced application.

A rotating compensator ellipsometer comprises the same elements as the nulling ellipsometer just described, except that the quarter-wave plate is motorised, generally rotating at a few hundred hertz. This type of system overcomes all the polarisation constraints in the source and detector, but requires spectral calibration of the compensator, a source of systematic error in the measurement. The signal equation of these ellipsometers delivers $\tan \Delta$ and $\tan \Psi$, so the ellipsometric angles are unambiguously defined.

Phase Modulation Ellipsometer

The optical setup includes a photoelastic modulator placed before the analyser. All elements are fixed, and the polarisation is modulated by a birefringent modulator. This type of ellipsometer does not therefore require any special characteristics with regard to the polarisation for the source or the detector. Measurements can be made over a broad spectral range and at a high acquisition rate. The use of a photoelastic modulator does not involve very accurate alignment (no rotating elements). However, this technology does need high-performance electronics, able to carry out acquisition and analysis of harmonics of order 0, 1, and 2 in a signal modulated at a frequency of 50 kHz. Moreover, the modulator is a chromatic element that must be calibrated with respect to the wavelength.

This ellipsometer measures $\tan \Delta$ and $\sin 2\Psi$. By using two measurement arrangements (rotating the modulator), the ellipsometric angles are unambiguously defined, since the combination of two measurements yields $\tan \Delta$ and $\tan \Psi$. Further description of this system is given in the next section.

Detailed Description of the Phase Modulation Ellipsometer

Optical Elements

Figure 9.20 depicts the optomechanical sequence of the phase modulation ellipsometer. It comprises the following elements:

- *Light Source.* This is a xenon arc lamp with a spectrum from the near infrared to the ultraviolet (190–2,100 nm).
- *Polariser.* This is a device which, when illuminated by arbitrary incident light, delivers light of a well-defined polarisation. It is in this case a linear polariser, characterised by an extinction rate of the order of 10^{-5} . A Glan polariser is used to cover a visible spectral range (240–830 nm), and a Rochon polariser to go down to the ultraviolet (190 nm and below).
- *Photoelastic modulator.* This is a silica rod in the shape of a parallelepiped, exploiting the photoelastic effect (see Fig. 9.21). Its optical properties are modified when it is put under strain. While the rod is optically isotropic

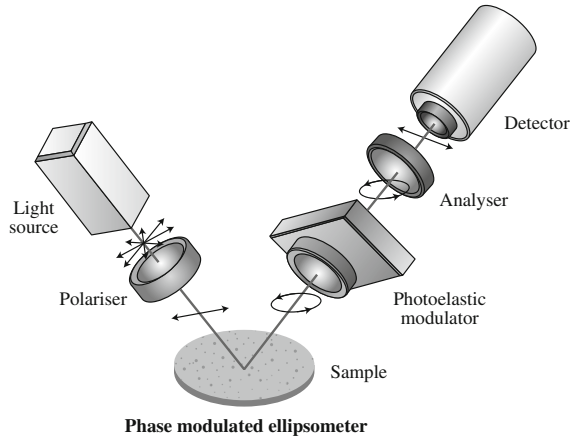


Fig. 9.20. Optomechanical setup in the phase modulation ellipsometer UVISEL

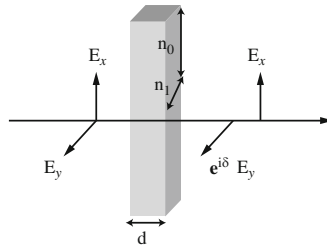


Fig. 9.21. Principle of the photoelastic modulator

with index n_0 in its equilibrium state, it becomes birefringent under uniaxial strain. The strain axis defines a proper axis of index n_1 . In a photoelastic modulator, the strain is varied sinusoidally. The birefringence is then modulated with angular frequency ω . The modulation of the rod is obtained by setting up a resonance that is maintained by one or more piezoceramics connected to the control electronics and producing this shear stress. The resonance frequency of the rod is $f = \omega/2\pi = 50 \text{ kHz}$, which allows very high acquisition rates, up to 1 ms/point. The phase difference introduced between the two components of the electric field $\mathbf{E}$ is then given by

$$\delta(t) = A \sin \omega t ,$$

where $A = 2\pi d(n_1 - n_0)/\lambda$.

- **Analyser.** This is located in the output arm, just after the photoelastic modulator, and is the same as the input polariser. It analyses the polarisation state of the reflected beam.
- **Monochromator.** This is placed after the analyser to select the wavelengths emitted by the xenon source. A motorised grating is used to separate the different wavelengths of the beam and direct the beam of

chosen wavelength onto the output slit. The measurement time covering the whole spectral range 190–2,100 nm is around 5 min. The advantage of the monochromator is that it can control the acquisition interval and the spectral resolution, giving a highly accurate and reproducible measurement. Its position in the optical setup can lead to light interference that may add to the ellipsometer beam. To overcome this problem, the background noise is estimated at each acquisition step by obscuring the source, whence the interference can be subtracted in the form of a continuous component.

- **Detector.** This measures the intensity of light reflected from the sample at each wavelength. The monochromator integrates one or more detectors depending on the spectral range covered by the ellipsometer. The photomultiplier is used to detect UV–visible spectra in the range 240–830 nm. Its gain can be adjusted to obtain a very good signal-to-noise ratio (good linearity and high dynamic range). An InGaAs photodiode is used to cover the near infrared (850–2,100 nm).

Making the Measurement

Here we describe the underlying measurement principle for the phase modulation ellipsometer. To do this, we consider the output light intensity $I(\omega, t)$ collected in the detector. It is proportional to the square of the amplitude of the electric field $\mathbf{E}$ emerging after passing through the whole optical setup of the ellipsometer. In the Jones formalism, this is given by

$$I(\omega, t) \propto E_t E_t^* = I \left[I_0 + I_S \sin \delta(t) + I_C \cos \delta(t) \right], \quad (9.16)$$

where

$$I_0 = 1 - \cos 2\Psi \cos 2A + \cos 2(P - M) \cos 2M (\cos 2A - \cos 2\Psi) \\ + \cos 2(P - M) \sin 2A \sin 2M \sin 2\Psi \cos \Delta,$$

$$I_S = \sin 2(P - M) \sin 2A \sin 2\Psi \sin \Delta,$$

and

$$I_C = \sin 2(P - M) \left[\sin 2M (\cos 2\Psi - \cos 2A) + \sin 2A \cos 2M \sin 2\Psi \cos \Delta \right].$$

Measurement configurations correspond to simplified expressions: $P - M = 45^\circ$ and $A = 45^\circ$, configuration II $M = 0^\circ$, configuration III $M = 45^\circ$. We then have

$$I_0 = 1, \quad I_S = \sin 2\Psi \sin \Delta, \\ I_C = \sin 2\Psi \cos \Delta, \quad \text{configuration II},$$

$$I_{C'} = \cos 2\Psi, \quad \text{configuration III.}$$

The periodic signal S from the photomultiplier has the form

$$S(t) = S_0 + S_1 e^{i\omega t} + S_2 e^{2i\omega t}. \quad (9.17)$$

The measurement protocol involves identification of (9.16) and (9.17), in the following stages:

- harmonic analysis of the signal, extracting the continuous component S_0 , the fundamental S_1 (50 kHz) and the harmonic S_2 (100 kHz),
- calculation of the various corrections,
- determination of I_0 , I_S , and I_C ,
- deduction of the ellipsometric angles Ψ and Δ .

To sum all this up, a linearly polarised incident wave is obliquely incident on the sample. The modulated and polarised reflected wave is then analysed by a second analyser. Harmonic analysis of the signal is carried out relative to the frequency of modulation of the polarisation. Knowing the amplitudes of the different harmonics, one can then deduce the ellipsometric angles Ψ and Δ .

9.2.5 Ellipsometric Data and Its Use

General Approach

The greatest difficulty for the user is not so much in the instrumentation as in processing the data. Ellipsometry is an indirect technique in the sense that the measured parameters Ψ and Δ are not the physical properties of the sample that one aims to determine, these being a thickness, a refractive index, and so on. A mathematical model is required to determine these.

There are four stages in the ellipsometric analysis of a sample (see Fig. 9.22):

1. Measurement of the ellipsometric data Ψ_{exp} and Δ_{exp} for different wavelengths.
2. Construction of a model to describe the sample. The theoretical sample comprises a number of discrete and well-defined layers, described by values of thickness and optical constants. The initial values are thus used to calculate the theoretical ellipsometric angles Ψ_{th} and Δ_{th} for the sample. At this stage, the unknown properties of the sample are treated as fitting parameters.
3. The fitting process involves adjusting the selected parameters of the theoretical sample so as to minimise the quantity χ^2 defined by

$$\chi^2 = \min \sum_{i=1}^n \left[\frac{(\Psi_{\text{th}} - \Psi_{\text{exp}})_i^2}{\Gamma_{\Psi,i}} + \frac{(\Delta_{\text{th}} - \Delta_{\text{exp}})_i^2}{\Gamma_{\Delta,i}} \right],$$

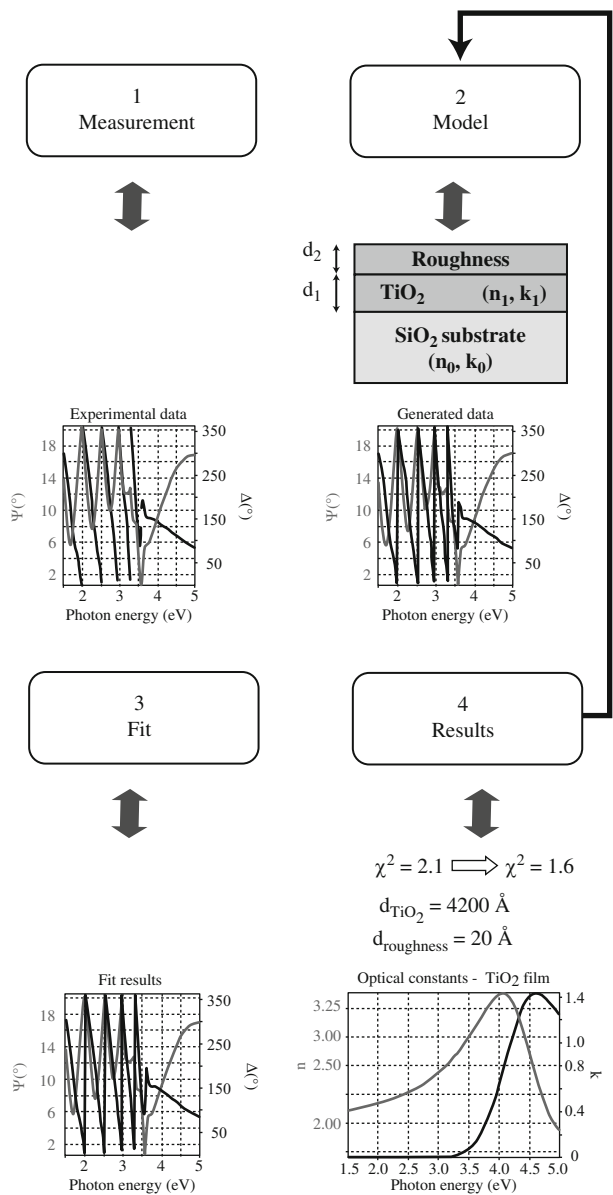


Fig. 9.22. Four steps in the ellipsometric analysis of a sample

Table 9.2. Example of an analysis report obtained for an SiN film deposited on c-Si

χ^2 minimization on I_S, I_C					
$I_S = \sin(2\Psi) \times \sin(\Delta), I_C = \sin(2\Psi) \times \cos(\Delta)$					
$\chi^2 = 0.272912$					
Iterations Number = 52					
Parameters					
1) L1 Thickness [Å] = 633.677 ± 1.192					
2) sin_new_2 n^∞ = 2.062, 450, 0 ± 0.022, 592, 1					
3) sin_new_2 ωg = 4.051, 184, 0 ± 0.036, 112, 8					
4) sin_new_2 fj = 0.295, 343, 2 ± 0.010, 873, 6					
5) sin_new_2 ωj = 4.804, 370, 0 ± 0.159, 089, 4					
6) sin_new_2 Γj = 2.262, 315, 0 ± 0.0751, 214					
Correlation matrix					
=1=	=2=	=3=	=4=	=5=	=6=
1.000	0.398	0.114	0.042	−0.529	−0.017
	1.000	0.506	0.584	−0.929	0.427
		1.000	0.844	−0.180	0.543
			1.000	−0.284	0.882
				1.000	−0.197
					1.000
Initial data for fit					
Model File: wafer8_lpcvd-nitrid.mdl					
Fitting Choice: Default					
Fitting Routine: Simple Fit					
Experimental File: wafer8.spe					
Spectrum					
Range: 1.5000–4.7000 eV					
Increment: 0.0500 eV					
Points Number: 65					

where Γ is the experimental error in the ellipsometric angles. This minimisation procedure is based on mathematical algorithms (e.g., Levenberg–Marquardt, Simplex) included in the software.

4. Determination of the sample properties depends on the results of the fitting process. If the latter results are not satisfactory, i.e., if χ^2 is high, a new model or new parameters must be used until a better description of the sample is obtained. In the example depicted in Fig. 9.22, addition of surface roughness in the TiO₂ film greatly improves the value of χ^2 , which drops from 2.1 to 1.6.

The analysis of ellipsometric data has generated a considerable number of mathematical and computing tools, e.g., a library of optical constants of

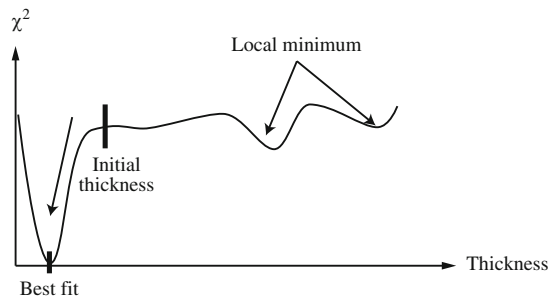


Fig. 9.23. Explaining the behaviour of the algorithm, which, starting from an initial thickness, will either find the best fit as indicated by the lowest value of χ^2 , or evolve toward a local minimum which can give a good approximation to the result, but which is not the best solution

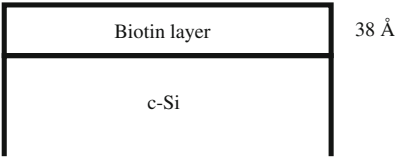


Fig. 9.24. Model used for analysis of the biotin/c-Si sample

materials, regression algorithms, matrix computation, and dispersion and mixing models. However, all this complexity remains hidden to the user, who merely selects the tools to be used via the software.

In practice, the main problem when modelling is to initialise the values of the fitting parameters so that they are not too far off the final solution. This requires a minimum amount of information about the sample, e.g., number of layers, materials, deposition methods, and so on. Indeed, the more complete the initial estimates, the easier it will be to set up the model and the easier it will be to develop it in the right direction, for example, by adding roughness, an interface, or an index gradient.

Goodness of Fit

At the end of the fitting routine, the value of χ^2 obtained is the first indicator of the quality of the model, along with the graph comparing the experimental measurements with the curve representing the theoretical model. The regression algorithm gives a correlation matrix, together with an estimate of the uncertainty in each of the parameters (see Table 9.2 for an example).

A strong correlation between the fitted parameters (value close to 1) indicates that the solution is not unique, but depends explicitly on the initial data. There may then be a tendency toward a local minimum as explained in Fig. 9.23. In this case, there are two possible approaches:

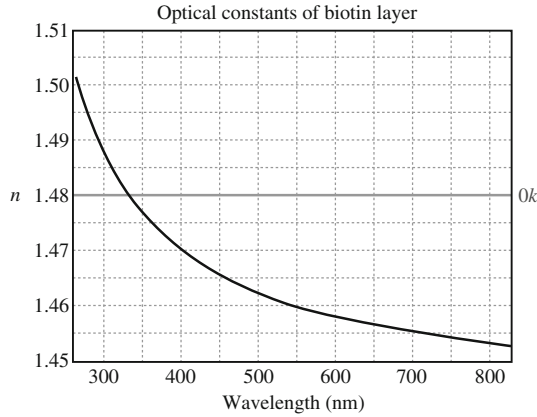


Fig. 9.25. Optical response of the biotin layer held fixed during the fitting routine

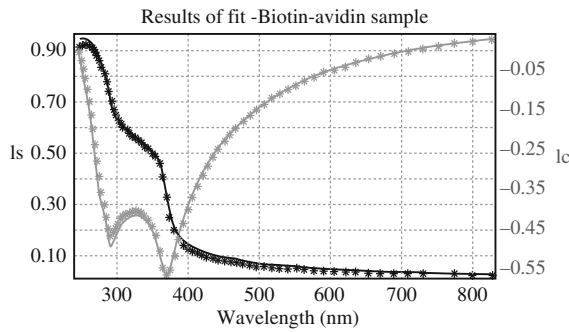


Fig. 9.26. Results of the fitting routine

- Reduce the number of fitted parameters. Convergence of the algorithm is more difficult when there are many unknowns. In general, this approach can be used to increase the repeatability of the fitting routine. It may also be useful to identify the highly correlated parameters and to fix one or more amongst them.
- Increase the number of independent measurements and combine them. The more experimental data available for the sample, the lower will be the correlation between the various fitted parameters. There are different ways to increase the amount of experimental data. Varying the angle of incidence consists in carrying out measurements at several angles of incidence. A transparent sample can also be measured in transmission, with subsequent combination of Ψ_{exp} , Δ_{exp} , and T_{exp} .

Finally, a basic rule is that the solution must make sense physically. Mathematically, a good result may be obtained, but which is not realistic, such as a null or negative thickness, or unphysical optical constants.

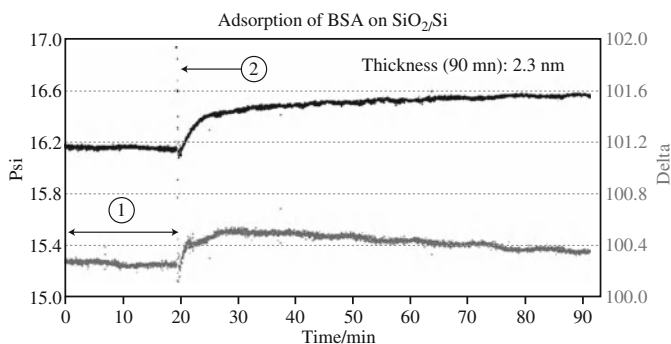


Fig. 9.27. Kinetic monitoring of the adsorption of the protein BSA on an SiO_2 interface

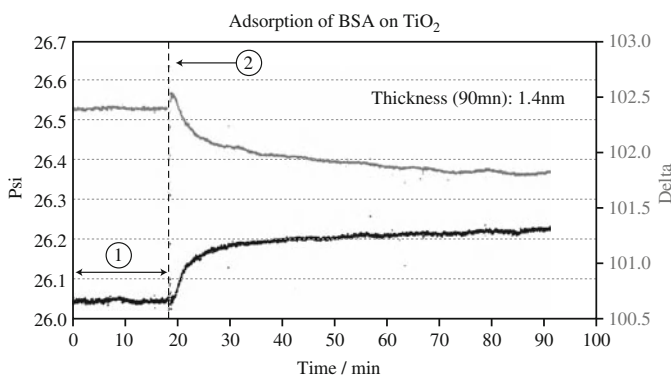


Fig. 9.28. Kinetic monitoring of the adsorption of the protein BSA on a TiO_2 interface

9.2.6 Applications

Characterising the Adsorption of Protein on a Surface

The biotin–avidin complex has been widely studied in the context of DNA chips. These large proteins each have four independent recognition sites which are able to easily immobilise a large number of biomolecules like proteins or enzymes, without the need for chemical agents.

Ellipsometry has been used in two stages to characterise the thickness of the deposited biotin layer, and then for biotin–avidin assembly. The measurement was made over the range 260–830 nm at an angle of incidence of 70° . The model here is a biotin monolayer deposited on a c-Si monolayer, as shown in Fig. 9.24. A thickness of 38 Å is found after fitting. The optical properties are held fixed during the fitting process (see Fig. 9.25).

The biotin–avidin complex was characterised in an aqueous medium. The measurements were made inside a cell containing the sample immersed in

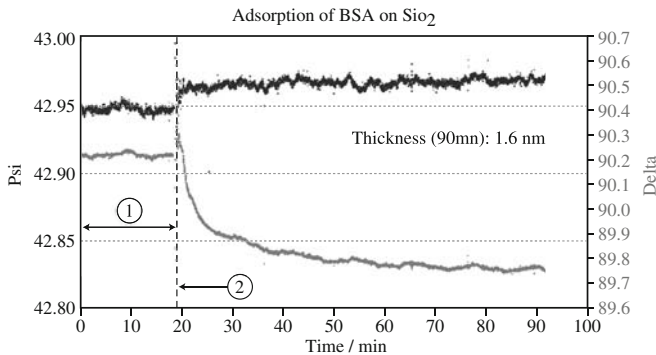


Fig. 9.29. Kinetic monitoring of the adsorption of the protein BSA on an Au interface

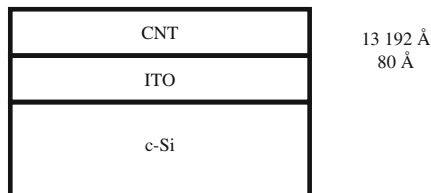


Fig. 9.30. Model used to analyse a sample of CNT/ITO/c-Si

distilled water. A monolayer model was also used, accounting for the different surrounding medium (here distilled water). The thickness of the layer was found to be 62 Å, which demonstrates the high affinity of biotin for avidin. The result of the fit shows the good agreement between the model used (curve) and the experimental measurements (stars), shown in Fig.9.26.

Kinetic Monitoring of the Adsorption of the Protein BSA on Different Surfaces

The liquid cell contains the substrate and is filled with a phosphate buffered saline (PBS) solution (denoted by 1 in Figs.9.27, 9.28, and 9.29). Then the solution of the bovine serum albumin (BSA) is introduced in such a way as to obtain a final protein concentration of 0.1 mg/L (denoted by 2 in Figs.9.27, 9.28, and 9.29). Kinetic monitoring of the angles ψ_{exp} and Δ_{exp} is able to characterise the way the protein adsorbs onto the surface. This study showed a preferred interaction between the protein and a surface coated with a layer of silica.

Characterising a DNA Layer Deposited on Gold

In this case the samples were thin layers of DNA deposited on gold. The DNA chains were composed of different numbers of bases (15, 25, or 35 bases) and

Table 9.3. Thickness of the DNA layer for different values of the refractive index

Refractive index n of the layer at 633 nm	Thickness of the DNA layer [Å]		
	15 bases	25 bases	35 bases
1.4	35	27	15
1.45	33	25	13
1.47	32	24	13
1.50	30	23	12

had different lengths. The thickness of the DNA layer must be very accurately known to make an optical biosensor. This type of application involves two main difficulties, viz., ultrathin layers and an unknown refractive index. Fitting the two parameters simultaneously often leads to a high correlation, indicating that the solution is not unique.

Table 9.3 shows the thicknesses found for different DNA chains as a function of the refractive index, treated as fixed during the fitting routine while only the thickness was allowed to vary.

Characterising a Carbon Nanotube Sensor

The structure here comprises a thin film of indium tin oxide (ITO), serving as an electrode, on which is deposited a thick layer of carbon nanotubes (CNT) as shown in Fig. 9.30. Owing to their tubular shape, carbon nanotubes introduce anisotropy into the layer. This is in fact a uniaxial anisotropy normal to the surface. Note that the refractive index is particularly low, a characteristic of carbon nanotubes, and there is non-negligible absorption for the extraordinary axis in the near infrared. The optical response of the CNT layer (as shown in Fig. 9.31) was modelled using an absorbing Lorentz oscillator with the following formula:

$$\varepsilon = \varepsilon_\infty + \frac{(\varepsilon_s - \varepsilon_\infty)\omega_t^2}{\omega_t^2 - \omega^2 + i\Gamma_0\omega} \; .$$

Figure 9.32 shows the very good agreement between experimental data and model.

Characterising a Photosensitive Langmuir–Blodgett (LB) Film

This kind of analysis is useful for obtaining a better understanding of physicochemical effects involved in certain biological phenomena, such as the formation of cancerous lesions in DNA induced by UV radiation, or the formation of secondary protein structures.

The sample is a thin Langmuir–Blodgett film deposited on quartz. This film is photosensitive beyond 2.6 eV. A very good fit is obtained ($\chi^2 = 0.22$) over the range 0.6–2.3 eV for a two-layer model with a dense interface. The

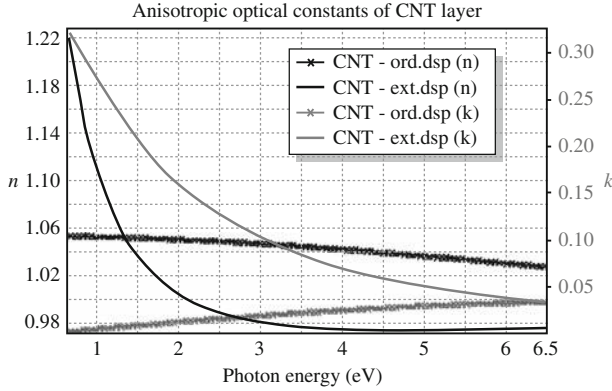


Fig. 9.31. Optical response of a carbon nanotube layer

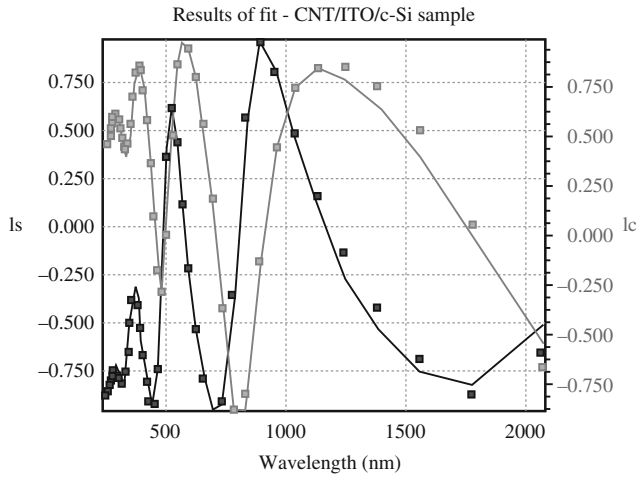


Fig. 9.32. Results of fitting procedure

total thickness of the film is 555 Å (see Fig. 9.33). The parameters are decorrelated and a better description of the absorption peaks of the film is obtained by simultaneously fitting three pieces of experimental data, viz., I_S , I_C , and T (see Fig. 9.34).

An absorbing quadrupole Lorentz oscillator was used to determine the optical constants of the Langmuir–Blodgett film (see Fig. 9.35).

9.2.7 Conclusion

Applications in bionanotechnology often require non-invasive techniques to characterise interface phenomena between substrates and biological materials. In this respect, ellipsometry has many advantages, e.g., thickness detection in

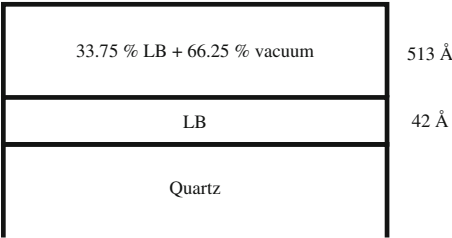


Fig. 9.33. Model used to analyse an LB/quartz sample

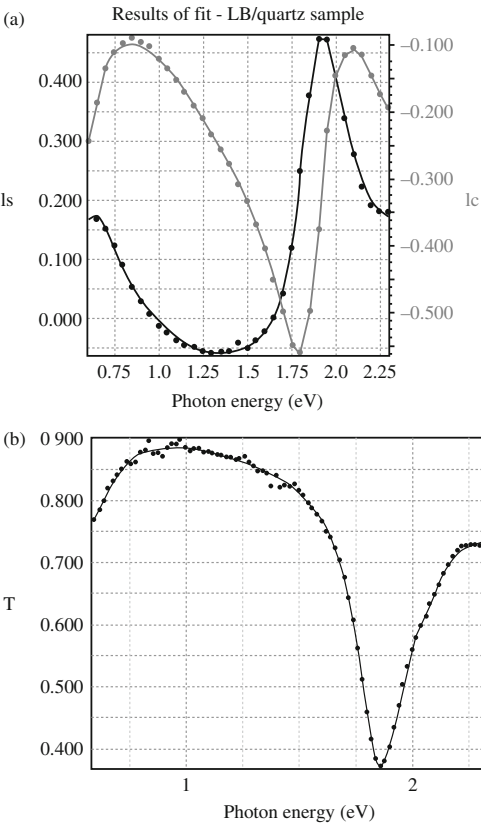


Fig. 9.34. Fit results using (a) ellipsometric measurements (I_S , I_C) and (b) transmission measurements (T) as a function of the wavelength

the angstrom range, kinetic monitoring, and applications in ambient media or liquids. The main limitation at the present time comes from the initial values of the refractive indices of biological media. These indices are rarely to be found in the literature, because experience with this technique has only recently begun to accumulate. It is advisable to fit only the thicknesses and to hold the index fixed, with an average value that is often somewhere near $n = 1.5$.

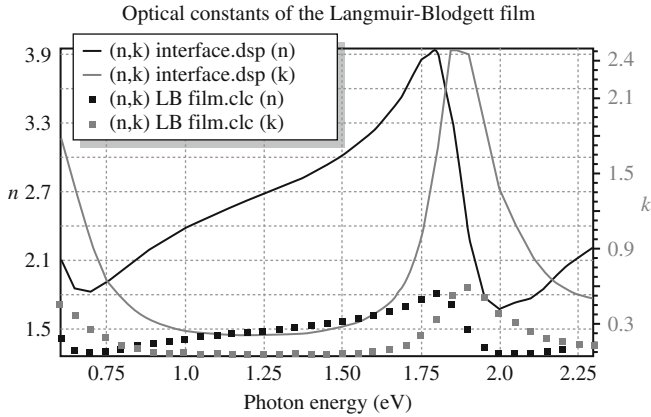


Fig. 9.35. Optical response of the Langmuir–Blodgett film. The figure clearly shows the inhomogeneity of the film depthwise, with much higher optical constants at the interface than at the top of the film

9.3 Optical Spectroscopy Using Waveguides

9.3.1 General Features of Optical Biosensors

Optical waveguide spectroscopy is a method of surface analysis based on the propagation of an evanescent wave along a solid–liquid interface when the refractive index of the solid phase is higher than that of the liquid phase, and when the angle of incidence of the wave on the surface is greater than the limiting angle for total internal reflection. This method was developed in the mid-1980s at the Ecole Polytechnique Fédérale in Zurich [99]. The underlying principle of the method makes it suitable for use as a biosensor [100], since many interactions relevant to biology occur at a solid–liquid interface.

There are several major types of biosensor, exploiting different means of analysis [101]:

- Biosensors based on electrical detection. This type of detection can be used when the product of an enzyme reaction is oxidised or reduced and the current or change in interfacial capacitance can be registered.
- Biosensors based on heat detection. The reaction heat (either positive or negative) produced during a biological recognition phenomenon is recorded using a thermocouple.
- Mechanical biosensors. In this case, binding effects between a ligand and a receptor manifest themselves through a change in natural frequency of a piezoelectric crystal onto which one of the partners in the bond has been grafted. Detectors using a quartz crystal microbalance belong to this family (see Sect. 9.6).
- Interferometric biosensors, especially those using Mach–Zender interferometers.

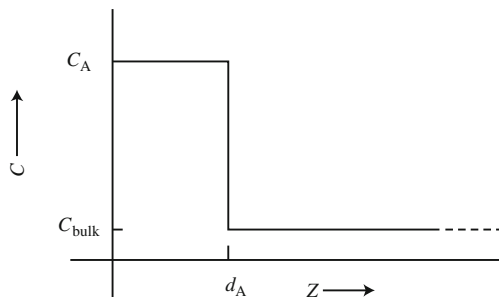


Fig. 9.36. Idealised concentration profile (along the z axis, normal to the surface) of a film of biomolecules deposited at a solid–liquid interface. C_A is the concentration in the film, C_{bulk} is the concentration in the solution, and d_A is the film thickness

- Optical biosensors. These exploit the fact that a change in composition of an interface leads to a modification of the reflectivity. This family of biosensors includes surface plasmon spectroscopy, attenuated total internal reflection near infrared or fluorescence spectroscopy, scanning angle reflectometry, and the various techniques of ellipsometry and spectroscopy of normal modes coupled in a waveguide.

In this section we shall discuss in detail the spectroscopy of normal modes coupled in a waveguide with integrated optics. However, it will be useful to begin with a brief overview of some features common to all optical biosensors:

- Most optical biosensors can be used to study molecules that have not been modified by fluorescent or radioactive labels.
- One can monitor not only the binding kinetics of a soluble ligand on its receptor immobilised at the surface of a solid substrate, but also the desorption kinetics of the ligand. This feature of waveguides with integrated optics is shared by in situ ellipsometers and surface plasmon resonance devices.

The relevant phenomena take place at a solid–liquid interface. For this reason, the apparent affinity constant obtained is not directly comparable with the value that would be found in a homogeneous medium, e.g., by equilibrium dialysis experiments or various types of spectroscopy. Indeed, the concentration of ligand in contact with the surface is not necessarily the same as the concentration in the homogeneous solution far from the interface. The ligand concentration in the thin liquid film in the vicinity of the interface drops during the binding process, but subsequently increases due to diffusion from the bulk of the solution toward this film. It is essential to take into account this diffusion and also convection effects (e.g., due to stirring in the measurement flask, flow of solution over the biosensor surface) in order to give a detailed description of the binding processes. These constraints are common to all biosensors.

Note that in the case of surface plasmon spectroscopy, where a gel is often immobilised on a gold surface, receptor molecules can be grafted onto the active groups of the gel, which then plays the role of a 3D reservoir for these molecules. This has the effect of bringing the measurement conditions closer to the effective conditions to be found within a biological medium, and also of enhancing the sensitivity of the analysis (because the number of receptor molecules is increased relative to the situation where these molecules are grafted onto a planar substrate). On the other hand, it considerably increases the difficulty in analysing the signal, because it modifies the transport of analyte molecules through the pores in the gel as compared with an aqueous solution [102].

9.3.2 Optical Spectroscopy of Normal Modes Coupled in a Waveguide

Optical Characteristics of a Film of Biomolecules Bound to an Interface

The amount of biomolecules bound to an interface is given by the Gibbs surface excess:

$$\Gamma = \int_0^\infty [C_A(z) - C_{\text{bulk}}] dz, \quad (9.18)$$

where the z axis is normal to the average plane of the interface (see Fig. 9.36). In order to determine Γ , which represents the number or mass of bound molecules per unit area, one needs to know the interfacial concentration profile $C_A(z)$. In (9.18), C_{bulk} is the concentration of the relevant molecules in the bulk of the solution, i.e., far from the surface under investigation. Usually, when one studies the binding of biomolecules to a solid–liquid interface, this concentration profile is taken as a step function. In fact, the concentration of deposited biomolecules is assumed to be constant and equal to C_A for z in the interval from 0 to d_A , where d_A is the optical thickness of the deposited film. In most cases, d_A is close to one of the geometrical dimensions of the biomolecule, e.g., the hydrodynamic radius of a protein, the length of its major or minor axis if the protein is treated as an ellipsoid, the length of a strand of DNA immobilised at the interface, etc. (see Fig. 9.36). In general, this step-shaped concentration profile is justified by the compact nature of biomolecules like proteins or DNA.

If it is also assumed that $C_A \gg C_{\text{bulk}}$ in (9.18), the amount of molecules deposited per unit area is given by

$$\Gamma = C_A d_A. \quad (9.19)$$

However, the concentration of solute in the adsorbed thin film is directly proportional to the refractive index of the solution contained in this film [103, 104]:

$$n_A = n_C + C_A \frac{dn}{dc}. \quad (9.20)$$

In (9.20), n_C is the refractive index of the analyte solution and dn/dc is the rate of change of the refractive index of the solute with the concentration in the solvent. The validity of this relation has been demonstrated experimentally [105]. Substituting (9.20) into (9.19), we obtain

$$\Gamma = \frac{(n_A - n_C)d_A}{dn/dc}. \quad (9.21)$$

This relation implies that one must measure n_A , d_A , n_C , and dn/dc in order to determine the number of molecules deposited per unit area during the molecular recognition process between an immobilised receptor and a soluble ligand. The refractive index of the pure solvent and the value of dn/dc are measured independently using a refractometer. (It is advisable to carry out these measurements at the same wavelength as the light delivered by the source of the optical biosensor in order to avoid problems due to dispersion.) The optical measurement consists therefore in determining the values of the index n_A and the optical thickness d_A of the adsorbed film. This means that one must determine *two* unknowns. Note that, for an optically anisotropic deposited film, one must determine two refractive indices for a biaxial film, e.g., a lipid bilayer, or three refractive indices for a triaxial medium.

One should question the validity of (9.21), which is based on the concentration profile shown in Fig. 9.36. However, it has been shown that the product $(n_A - n_C)d_A$ is an optical invariant, i.e., a physical quantity that does not depend on the refractive index profile chosen to model the interface [106]. Although one may doubt the separate values calculated for n_A and d_A , which are effectively based on an interface model that does not necessarily correspond to reality, the value of the product $(n_A - n_C)d_A$ remains robust. Let us now see how the values of n_A and d_A can be determined by waveguide spectroscopy.

Principles of Waveguide Spectroscopy

In optical spectroscopy using waveguides, a monochromatic light beam is directed onto a diffraction grating incorporated in a waveguide F. The waveguide, deposited on a silica substrate S, is made from a high-index medium, typically a mixture of silica and titanium dioxide [108]. It is in contact with a solution C (an aqueous solution in our case) containing the solute whose binding to the solid–liquid interface one hopes to investigate. The refractive index n_F of the waveguide is higher than the refractive index n_S of the silica substrate, which is itself higher than the refractive index n_C of the solution. Let d_F be the thickness of the waveguide film. A further layer A of thickness d_A and refractive index n_A is usually deposited on the waveguide. This layer is the subject of study, e.g., in our research, a multilayer of polyelectrolytes or polypeptides. This setup is shown schematically in Fig. 9.37.

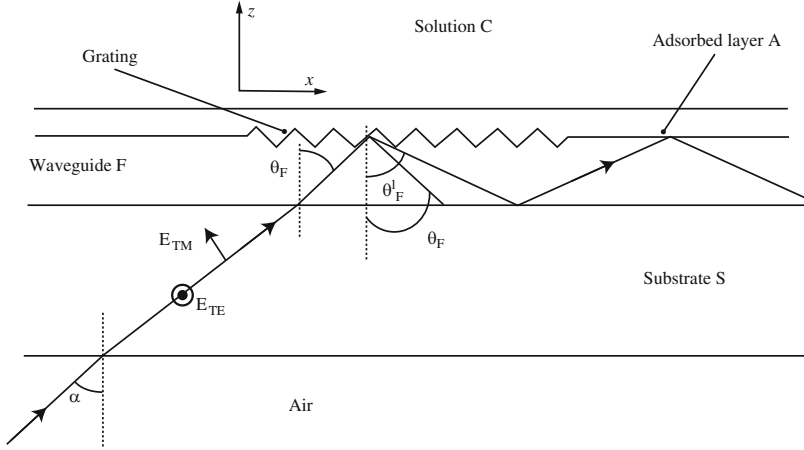


Fig. 9.37. Spectrometer used in waveguide spectroscopy. Directions z and x correspond to the direction normal to the plane of the interface and the direction of propagation along the waveguide, respectively. The two components of the electric field corresponding to the two polarisations TE (transverse electric) and TM (transverse magnetic) are indicated. α and θ_F are the angles of incidence at the air–glass and glass–waveguide film interfaces, respectively

The laser beam arrives at the silica–air interface at an angle of incidence α . Let z be the axis normal to the waveguide and x the axis perpendicular to z and lying in the plane of incidence (see Fig. 9.37). The laser beam is refracted at the air–silica and silica–waveguide interfaces. The incident beam is characterised by a wave vector $\mathbf{k}$, whose component k_x is given by

$$k_x = \frac{2\pi}{\lambda} n_{\text{air}} \sin \alpha . \quad (9.22)$$

The wave vector and the other quantities characterising an electromagnetic wave are defined in Appendix A at the end of this section.

At each refraction, the component k_x remains unchanged. Consequently, the laser beam arrives at the diffraction grating of the waveguide at an angle θ_F (see Fig. 9.37) satisfying

$$n_F \sin \theta_F = n_{\text{air}} \sin \alpha . \quad (9.23)$$

The diffraction grating in the waveguide film diffracts the light, not only in the direction $\theta = \theta_F$, but also in other directions θ_F^l such that

$$N = n_F \sin \theta_F^l = n_{\text{air}} \sin \alpha + \frac{l\lambda}{\Lambda} , \quad (9.24)$$

where l is a whole number called the order of diffraction, λ is the wavelength of the incident beam, and Λ is the periodicity of the diffraction grating integrated into the waveguide film.

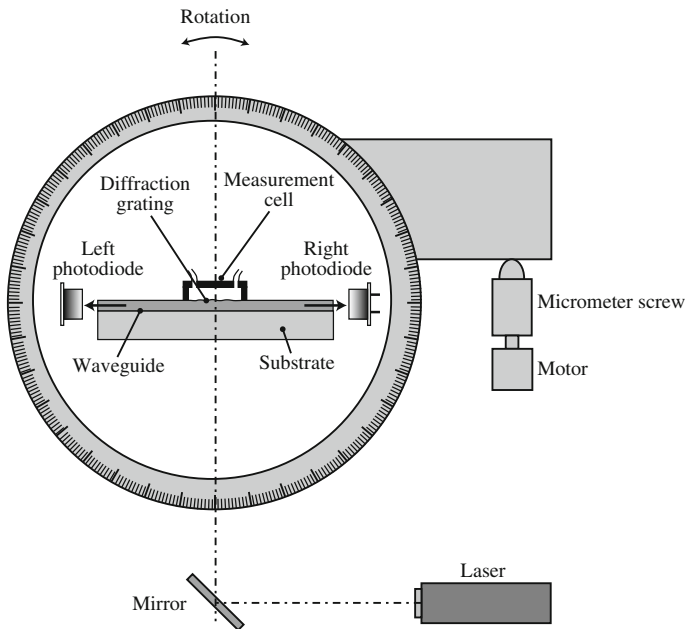


Fig. 9.38. Experimental setup for coupling TE and TM waves in a plane waveguide. A computer controls the motor rotating the measurement cell and photodiodes. When coupling conditions are satisfied, the beam diffracted by the grating propagates through the waveguide to the photodiodes. To determine the coupling angles, the device scans from -8 to $+8$ degrees on either side of an arbitrary reference angle and measures the four coupling angles (two for the TE polarisation and two for the TM polarisation). The average absolute values of the TE or TM angles correspond to the two coupling angles

When one of the diffracted beams propagates in the waveguide, there is a coupling between the incident beam and the waveguide. The component k_x of the wave vector of the diffracted wave is then equal to

$$k_{x,F} = \frac{2\pi}{\lambda} N = k_0 N . \quad (9.25)$$

Propagation of the wave in the waveguide can be treated as a succession of reflections on the two interfaces of the waveguide (see Fig. 9.37). At each reflection, there is a change in the phase of the reflected wave. Let $\Phi_{S,F}$ and $\Phi_{F,C}$ be the phase shifts introduced during reflections at the silica–waveguide and waveguide–solution (or waveguide–adsorbed layer in the case where a film of biomolecules is deposited on the waveguide) interfaces, respectively. To obtain propagation of the diffracted wave in the waveguide, there must be constructive interference between the initial beam and the beam reflected at the S/F and F/C interfaces. This imposes the relation

$$2k_{z,F}d_F + \Phi_{S,F} + \Phi_{F,C} = 2\pi m , \quad (9.26)$$

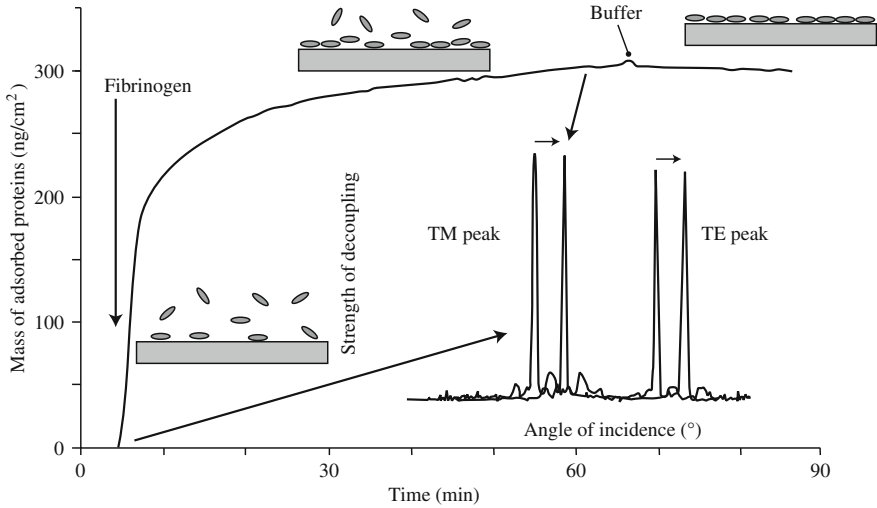


Fig. 9.39. Time dependence of the positions of the TE and TM peaks (*insert*) and of the surface concentration when the waveguide surface is set in contact with a solution of proteins. When the peak positions α_{TE} and α_{TM} no longer change with time, the amount of adsorbed molecules has reached a constant value

where m is a whole number called the order of the waveguide and $k_{z,F}$ is the z component of the wave vector of the diffracted wave. Furthermore, the component $k_{z,F}$ satisfies

$$k_{z,F}^2 + k_0^2 N^2 = n_F^2 k_0^2, \quad (9.27)$$

which follows directly from the wave equation for an electromagnetic wave. This equation and the equations of continuity for the electric and magnetic fields are given in Appendix A at the end of this section.

Any electromagnetic wave can be decomposed into a wave polarised perpendicularly to the plane of incidence, called the transverse electric (TE) or s wave, and a wave polarised in the plane of incidence, called the transverse magnetic (TM) or p wave. If a film is deposited on the waveguide surface, the phase shifts induced by the successive reflections will depend not only on the thickness and refractive index of the film, but also on the polarisation of the wave.

In waveguide optical spectroscopy, the helium–neon laser beam (HeNe, $\lambda = 632.8\text{nm}$), incident on the waveguide surface, is linearly polarised, but contains the two components TE and TM (see Fig. 9.37). One then scans over the angle of incidence α . To do this, the waveguide and injection cell, which contains the buffer solution or the solution containing the biomolecules to be adsorbed, are mounted on a goniometer, itself placed on a motorised stage. The general setup is shown schematically in Fig. 9.38.

For most angles of incidence, (9.26) is not satisfied. The diffracted waves cannot therefore propagate through the waveguide. But when the angle of incidence α reaches the value for which this relation is satisfied, there is coupling between the incident wave and the waveguide. A detector (photodiode) positioned at each output of the waveguide measures the intensity of light propagating through the waveguide and thereby determines the coupling angles for the two polarisations. We work with single-mode waveguides, in which only the mode corresponding to $m = 0$ can propagate. Moreover, the gratings printed on the waveguide are such that only the order $l = 1$ can be diffracted. There is a coupling angle for each of the polarisations TE and TM. To each coupling angle, there corresponds an effective refractive index N_{TE} or N_{TM} , depending on whether it is an angle for the TE or TM polarisation, respectively. As the waveguide is symmetrical, by scanning the angle of incidence between two symmetrically placed values $-\alpha_{\text{max}}$ and $+\alpha_{\text{max}}$, two coupling angles are found for each polarisation, one on the left of the waveguide and the other on the right. The average of the absolute value of these two angles provides an accurate value for the coupling angle, without the need to establish the zero for the setup. An example spectrum is shown in the insert of Fig. 9.39.

Signal Processing

The signal must first be processed without the film, in order to determine the characteristics of the waveguide alone, in particular, its refractive index n_{F} and the thickness d_{F} of the film which actually constitutes the waveguide. This is a calibration stage. The molecules under investigation are then adsorbed and the new values of N_{TE} and N_{TM} are determined during adsorption. These values of N_{TE} and N_{TM} are directly calculated from the values of the angles of incidence which allow coupling of a wave in the waveguide. The waveguide is scanned periodically, so that the values of N_{TE} and N_{TM} can be recorded in a regular manner. The periodicity of these measurements determines the temporal resolution of the device. Our own equipment, built at Unit 595 of INSERM (France), has a time resolution of the order of 100 s. The successive values of N_{TE} and N_{TM} are used to determine the index and thickness of the film as a function of time. By considering the silica substrate (index n_{S}), the waveguide (index n_{F}), and the solution (index n_{C}) flowing above the surface, the phase equation for the two polarisations is

$$2k_{z,\text{F}}d_{\text{F}} + \Phi_{\text{F},\text{S}} + \Phi_{\text{F},\text{C}} = 0. \quad (9.28)$$

In the waveguide F, the wave undergoes total internal reflection and the wave vectors have real components. In the two media S and C, the field is evanescent and the wave vector is pure imaginary. The reflection coefficient r of an interface is defined by

$$r = \frac{E_{\text{r}}}{E_{\text{i}}}, \quad (9.29)$$

where E_i and E_r are the magnitudes of the electric field of the incident wave and the wave reflected by the interface, respectively. This coefficient is a complex number given by

$$r = |r| \exp(i\Phi) , \quad (9.30)$$

where Φ is the phase shift between the incident and reflected waves, and $|r| = 1$, since the reflection of the wave is total. The coefficients of reflection at the F/C and S/F are thus given by

$$r_{F,C} = \exp(i\Phi_{F,C}) , \quad r_{F,S} = \exp(i\Phi_{F,S}) . \quad (9.31)$$

The reflection coefficients and the phase shifts are calculated using the matrix method [107, 109].

The components of the wave vectors, perpendicular to the interface, are given by

$$\begin{cases} k_{z,F} = k_0(n_F^2 - N^2)^{1/2} , \\ k_{z,S} = ik_0(N^2 - n_S^2)^{1/2} , \\ k_{z,C} = ik_0(N^2 - n_C^2)^{1/2} , \end{cases} \quad (9.32)$$

where $k_{z,F}$, $k_{z,S}$, and $k_{z,C}$ are the components of the wave vectors normal to the interfaces in the waveguide, in the silica substrate, and in the solution. N is either N_{TE} or N_{TM} , depending on whether one is considering the phase shifts of the s or the p waves.

The reflection coefficients given by (9.31) can then be calculated using a matrix method [107], as presented in Appendix B at the end of this section. In this method, the film is decomposed into homogeneous and isotropic layers, each of which is characterised by its thickness and its refractive index, the quantities one seeks to determine. One can then calculate the amount of adsorbed molecules in this layer using (9.21).

Consequences: Resolution and Sensitivity

The angular resolution is 3×10^{-4} degrees, which leads to a resolution in the effective refractive index ΔN of the order of 10^{-5} , according to (9.24). Calculation shows that this results in a detection limit of the order of a few ng cm^{-2} [101]. Given that a monolayer of proteins corresponds to an adsorbed quantity of the order of 100 ng cm^{-2} , we find that the intrinsic detection limit of this method corresponds to a few percent of a monolayer. This explains why optical waveguide spectroscopy has been so abundantly used to study the dynamics of protein adsorption.

In order to monitor a very fast phenomenon, one could simply follow the evolution of a single coupling peak for the light in the waveguide. This would give qualitative access to high-speed kinetics, but to the detriment of quantitative information. Note that, in order to overcome this problem, a new design of equipment is currently being developed. This is an optical waveguide in which the light enters the guide at grazing incidence, and where the

diffraction grating is used to emit radiation out of the waveguide when it propagates effectively along it. The angle at which this light is emitted out of the waveguide, which corresponds to the value of α_{TE} or α_{TM} , is measured by a row of parallel photodiodes below the surface of the waveguide.

9.3.3 Applications of Optical Waveguide Lightmode Spectroscopy

Antigen–Antibody Reactions and Comparison with Other Techniques

Like the method of surface plasmon resonance, waveguide spectroscopy has been widely used to detect antigens in a biological fluid when specific antibodies are immobilised at the surface. This is possible because the technique is very sensitive and also because the measurement cell has a very small volume (typically 20–50 μL). Note that the volume of the connecting tubes must be added to this volume. These tubes can of course be made very short and, by choosing the ‘right’ measurement cell, it is possible to work with less than 100 μL of solution.

It is interesting to compare the results obtained by waveguide spectroscopy with those obtained by ellipsometry and quartz crystal microbalance (QCM) when different proteins adsorb onto substrates coated with a thin film of TiO_2 , but also when surfaces saturated with proteins are exposed to their corresponding antibodies. Some experimental curves and a summarising diagram are shown in Fig. 9.40. It turns out that the results obtained by the two optical techniques are very close to one another, but significantly lower than those obtained by quartz crystal microbalance. This should come as no surprise, since optical techniques are sensitive to the dry mass of the molecules deposited (which alone contributes to increasing the refractive index of the film), whereas acoustic techniques such as QCM are sensitive not only to the mass of bound molecules, but also to the water bound directly to these molecules or in hydrodynamic interaction with the adsorbed film.

Using Optical Waveguide Lightmode Spectroscopy to Monitor the Construction of Polyelectrolyte Multilayers

Over the past few years, we have been trying to functionalise the surfaces of biomaterials with active substances inserted in polyelectrolyte multilayer films (see Chap. 21 for applications of these polyelectrolyte multilayer films). This type of film is obtained by sequential adsorption of polycations and polyanions on the surface of a charged interface under the physicochemical conditions of the deposit [112]. It is essential to monitor the construction of the polyelectrolyte multilayer film by an in situ method in order to obtain a good understanding of the growth mechanism of the multilayer film [113, 114]. Figure 9.41 shows the results obtained by monitoring the construction of a $\text{PEI}-(\text{PGA}-\text{PAH})_n$ multilayer by OWLS, where PEI is polyethyleneimine,

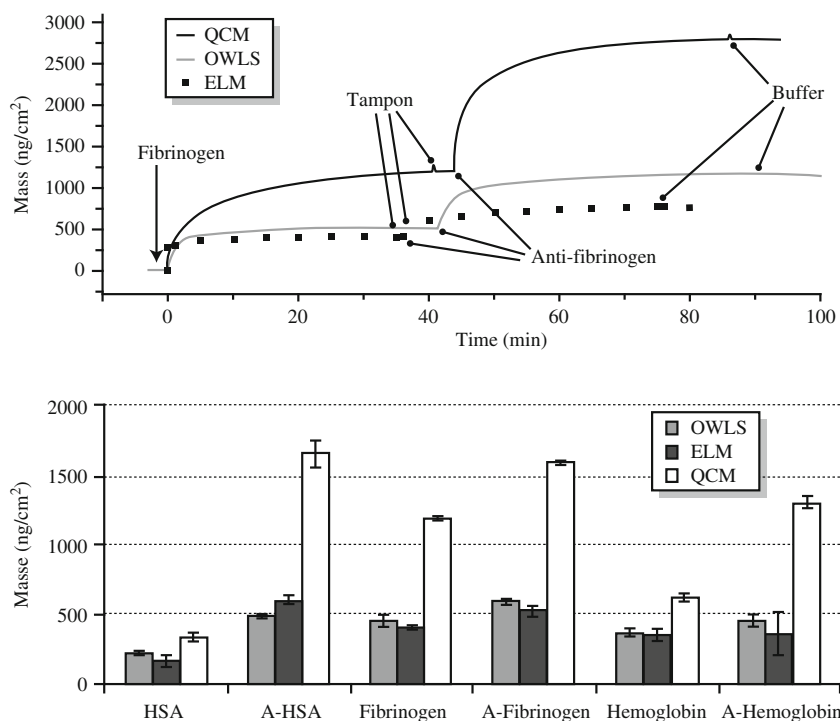


Fig. 9.40. (a) Adsorption kinetics of human fibrinogen dissolved at a concentration of $80\text{ }\mu\text{g mL}^{-1}$ in a HEPES buffer at 10 mM and pH 7.4 on surfaces modified by a 10-nm thick deposit of TiO_2 . The different curves correspond to the adsorbed quantities measured by ellipsometry (ELM), waveguide spectroscopy (OWLS), and quartz crystal microbalance (QCM). After rinsing with the buffer (arrows), the surfaces are set in contact with an antifibrinogen solution (polyclonal antibodies) at $80\text{ }\mu\text{g mL}^{-1}$ before a further rinsing with the buffer (arrows again). (b) Summary of comparative experiments for various proteins: human serum albumin (HSA), an antibody directed against human serum albumin (A-HSA), human fibrinogen, a polyclonal antibody directed against fibrinogen, human hemoglobin, and a polyclonal antibody directed against human hemoglobin. Three experimental techniques are compared: ellipsometry (ELM), optical waveguide lightmode spectroscopy (OWLS), and quartz crystal microbalance (QCM). Data reproduced from [111]

PGA is poly-L-glutamic acid, PAH is polyallylamine, and n denotes the number of pairs of layers deposited. Surprisingly, the raw signal, i.e., the value of N_{TE} (and the value of N_{TM} not shown in Fig. 9.41a), no longer increases continuously after depositing the sixth pair of layers, while experiments carried out with the quartz crystal microbalance show that the amount of matter deposited continues to grow well beyond the seventh pair of layers [114]. On the contrary, in the experiment corresponding to Fig. 9.41, the value of N_{TE} increases when the polyanion is injected and decreases by the same amount

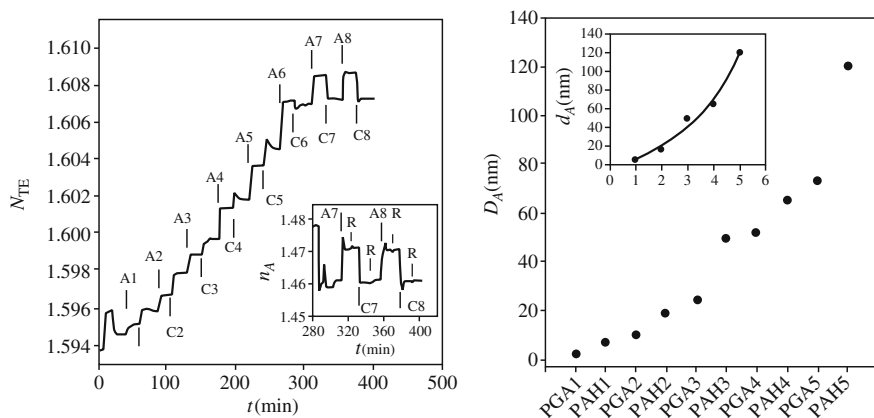


Fig. 9.41. (a) Using waveguide spectroscopy to monitor the construction of multilayer films of polyelectrolytes on a waveguide made from a mixture of silicon oxide and titanium oxide. The polyelectrolytes used here were poly-L-glutamic acid (negatively charged under the conditions of this study) and polyallylamine (positively charged under the conditions of this study), dissolved at 5 mg mL^{-1} in a MES-TRIS buffer at pH 7.4 and in the presence of sodium chloride at 100 mM. Injections of polycations and polyanions are represented by C and A, respectively. After deposition of the seventh layer of polyanion, the evolution of the apparent refractive index for the TE polarised wave becomes cyclic. Identical curves are obtained for the TM polarisation. (b) Evolution of the optical thickness of the multilayer film whose construction is shown in (a) (taken from [114]). *Insert:* Fitting the dependence of the optical thickness on the number of deposited bilayers to an exponential function

when the polycation is injected. This observation helped us to understand a key feature of multilayer films like PGA/PAH (see Fig. 9.41) or HA/PLL [113], where HA stands for hyaluronic acid and PLL for poly-L-lysine.

When the evolution of the optical thickness is calculated assuming a homogeneous and isotropic monolayer model, it transpires that these films are characterised by an exponential increase in the optical thickness (see Fig. 9.41b). When the film continues to grow beyond a characteristic distance, which is of the order of the penetration depth of evanescent waves, the values of N_{TE} and N_{TM} are expected to level out. This would mean that the evanescent wave emerging from the waveguide is no longer sensitive to changes in the optical response of the film close to the interface between the multilayer film and the polyelectrolyte solution. This effect has been observed for multilayer films such as those constructed from poly-4-styrene sodium sulfonate and polyallylamine. However, N_{TE} and N_{TM} undergo cyclic variations during the construction of films like PEI-(PGA-PAH) $_n$ (see Fig. 9.41a). This means that the deposition of polyelectrolytes, well beyond the penetration depth of the evanescent wave in the multilayer film, leads to changes in the refractive index of the film in the region that is effectively probed by the evanescent wave. This led us to

suggest that at least one of the two polyelectrolytes used in the films diffuses during adsorption, followed by diffusion out of the film when it is set in contact with a solution of polyelectrolytes of the opposite sign. Diffusion in the film must then lead to an increase in the average index of the film and hence an increase in N_{TE} and N_{TM} . Diffusion out of the film, on the other hand, must lead to a decrease in the film index, and hence to a decrease in N_{TE} and N_{TM} . Examination of Fig. 9.41a suggests that PGA can diffuse in the film during its adsorption onto the film and diffuse out of the film when the latter is set in contact with a solution of PAH. These diffusion phenomena into and out of the multilayer film, which cause the exponential growth of the film thickness, have been demonstrated by confocal laser scanning microscopy [115].

9.3.4 Conclusions

In this section, we have described the basic physical principles used in waveguide spectroscopy, explaining how the raw signal is used to calculate the optical properties of a film deposited at a solid–liquid interface. We have shown that this technique for in situ characterisation can be used to monitor molecular recognition processes of interest in biology or to understand the growth mechanisms of self-assembled films with very good sensitivity and good time resolution. This technique can also be used in analytical chemistry to monitor water quality and measure the concentration of toxic gases output from a production line.

Appendix A. Reflection of Light by a Perfect (Fresnel) Interface

To understand the optical techniques used to characterise interfaces, including optical waveguide spectroscopy, one must first understand how light reflects from a perfect interface, the so-called Fresnel interface. This kind of interface corresponds to a sudden change in refractive index when the light goes from one medium to the other, e.g., from a solid substrate to a liquid in contact with it. To this end, we begin by recalling Maxwell's equations in a non-magnetic medium [107]:

$$\nabla \wedge \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}, \quad (9.33)$$

$$\nabla \wedge \mathbf{H} = \mathbf{j} + \frac{\partial \mathbf{D}}{\partial t}, \quad (9.34)$$

$$\nabla \cdot \mathbf{B} = 0, \quad (9.35)$$

$$\nabla \cdot \mathbf{D} = \rho, \quad (9.36)$$

where $\mathbf{E}$, $\mathbf{B}$, $\mathbf{H}$, $\mathbf{D}$, $\mathbf{j}$, and ρ are the electric field, magnetic induction, magnetic field, displacement field, current density, and charge density, respectively. In uncharged media, the charge and current density are both zero, i.e., $\mathbf{j} = 0$ and $\rho = 0$. Further, $\mathbf{D}$ is related to $\mathbf{E}$ by

$$\mathbf{D} = \varepsilon \varepsilon_0 \mathbf{E} , \quad (9.37)$$

where ε is the dielectric constant of the medium and ε_0 is the dielectric permittivity of the vacuum. The dielectric constant is related to the refractive index of the medium by

$$\varepsilon = n^2 , \quad (9.38)$$

and it is generally a complex number. Maxwell's equations lead to the wave equation

$$\Delta \mathbf{E} = \frac{\varepsilon}{c^2} \frac{\partial^2 \mathbf{E}}{\partial t^2} , \quad (9.39)$$

where Δ is the Laplacian operator. Plane, monochromatic waves are defined by an equation of the form

$$\mathbf{E} = \mathbf{E}_0 \exp [i(\omega t - \mathbf{k} \cdot \mathbf{r})] , \quad (9.40)$$

where $\mathbf{k}$ is the wave vector, ω the angular frequency, t the time, $\mathbf{r}$ the spatial position, and $\mathbf{E}_0$ a constant vector. Inserting (9.40) into (9.39) and using (9.38), it follows that

$$k^2 = \left(\frac{2\pi}{\lambda} \right)^2 n^2 , \quad (9.41)$$

where λ is the wavelength of the light in vacuum.

Let us now show that an absorbent medium is characterised by a complex refractive medium. The intensity of a light ray is given by $I = E E^*$. Assume that the light wave propagates along the x axis. According to the Beer–Lambert law, the intensity is given by

$$I = I_0 \exp(-\alpha x) , \quad (9.42)$$

where α is the attenuation coefficient. Now if the light propagates along the x axis, (9.40) becomes

$$\mathbf{E} = \mathbf{E}_0 \exp [i(\omega t - kx)] . \quad (9.43)$$

Supposing that $n = n' + in''$, we then also have $\mathbf{k} = \mathbf{k}' + i\mathbf{k}''$, by (9.41). Finally, the light intensity is

$$I = E_0 E_0^* \exp(-2k''x) . \quad (9.44)$$

Hence,

$$\alpha = 2 \frac{2\pi}{\lambda} n'' . \quad (9.45)$$

The Beer–Lambert law thus implies that the refractive index has an imaginary part n'' .

The Maxwell equations also imply that, for a plane wave,

$$\mathbf{B} = \frac{\mathbf{k} \wedge \mathbf{E}}{\omega} . \quad (9.46)$$

Moreover, it can be shown that, at the interface between two dielectric media, we have

$$\begin{aligned} D_{z,1} &= D_{z,2} , \\ E_{x,1} &= E_{x,2} , \\ B_{z,1} &= B_{z,2} , \\ B_{x,1} &= B_{x,2} , \end{aligned} \quad (9.47)$$

where the x axis is parallel and the z axis is perpendicular to the interface. The four relations in (9.47) express the boundary conditions used to calculate the reflection coefficients.

Appendix B. Matrix Method

In this approach, the deposited film is decomposed into ν homogeneous and isotropic strata parallel to the substrate. The j th stratum is characterised by its thickness d_j and its refractive index n_j . The incident and reflected electric fields are then given by

$$\begin{pmatrix} E_i \\ E_r \end{pmatrix} \propto \mathbf{l}_F^{-1} \mathbf{l}_1 \mathbf{L}_1 \mathbf{l}_1^{-1} \cdots \mathbf{l}_\nu \mathbf{L}_\nu \mathbf{l}_\nu^{-1} \mathbf{l}_C \begin{pmatrix} 1 \\ 0 \end{pmatrix} , \quad (9.48)$$

up to an irrelevant multiplicative factor which cancels from the ratio E_r/E_i in (9.29). Each stratum labelled by j is characterised by two matrices $\mathbf{l}_j$ and $\mathbf{L}_j$, with inverses $\mathbf{l}_j^{-1}$ and $\mathbf{L}_j^{-1}$, respectively. These matrices are given by

$$\mathbf{l}_j = \begin{pmatrix} 1 & 1 \\ k_j/n_j^{2\rho} & -k_j/n_j^{2\rho} \end{pmatrix} \quad (9.49)$$

and

$$\mathbf{L}_j = \begin{pmatrix} \exp(ik_j d_j) & 0 \\ 0 & \exp(-ik_j d_j) \end{pmatrix} , \quad (9.50)$$

where k_j is given by (9.32), and $\rho = 0$ for the TE wave and $\rho = 1$ for the TM wave. The matrices $\mathbf{l}_F^{-1}$ and $\mathbf{l}_C$ express the optical properties of the waveguide film and the aqueous solution, respectively. In (9.50), i is such that $i^2 = -1$.

If the film can be treated as a single layer, only one stratum is required here ($\nu = 1$). However, in the case of protein adsorption onto polyelectrolyte multilayer films, we appeal to a bilayer model ($\nu = 2$) to fit (9.48) to the experimental data. In this type of experiment, the homogeneous and isotropic monolayer model is used to calculate the optical parameters of the multilayer film. The latter are introduced in the matrix elements of $\mathbf{l}_1$ and $\mathbf{L}_1$ of (9.48) in the second stage, which aims to determine the optical parameters of the second layer (the layer of adsorbed proteins) from measurements of the values of N_{TE} and N_{TM} obtained during the kinetic adsorption of proteins. Practical considerations concerning analysis of the experimental data can be found in [110].

9.4 Vibrational Spectroscopy

9.4.1 General Features

A molecule can be considered as composed of N atoms of mass m_i , joined together by springs of differing stiffness k_j depending on the binding force. Through this organisation, molecules tend to vibrate in specific ways. For a non-linear molecule comprising N atoms, there will be $3N - 6$ vibrations. These vibrations are quantised and each occurs at a well-defined energy corresponding to the energies of infrared photons. During a vibration, it is not only the masses that move, but also the charges carried by the atoms. In certain cases, one thus observes a polarisability or an oscillating dipole which can interact with the electric field of an electromagnetic wave. With regard to infrared spectroscopy, at certain frequencies, the energy transported by the electromagnetic wave is absorbed when the exciting frequency is equal to the frequency corresponding to the vibrational energy level. In the case of Raman spectroscopy, the exciting wave is scattered with a change in frequency. The change in frequency between the exciting wave and the scattered wave corresponds to the frequency of the vibrational energy level. Not all vibrations of a given molecule will be active in infrared or Raman spectroscopy. They occur in infrared spectroscopy only if they lead to a change in the derivative $\partial\mu/\partial q$ of the dipole moment, and they are only active in Raman spectroscopy if they produce a change in the differential polarisability tensor $\partial\alpha/\partial q$ [116]. The change in the dipole moment is a vector which has a well-defined direction relative to the geometrical axes of the molecule for each vibration. The change in polarisability is a rank three tensor relative to the axes of the molecule.

In infrared spectroscopy, it can be shown that the intensity of a mode is proportional to the square of the scalar product of the electric field vector of the incident wave with the derivative of the dipole moment vector:

$$I \propto \left(\frac{\partial\boldsymbol{\mu}}{\partial q} \cdot \mathbf{E} \right)^2 = \left(\frac{\partial\mu}{\partial q} \right)^2 E^2 (\cos\theta)^2, \quad (9.51)$$

where θ is the angle between the exciting electric field and the direction of the vector $\partial\boldsymbol{\mu}/\partial q$.

A mode will be more intense as the term $\partial\mu/\partial q$ itself increases, but also as the electric field becomes more closely aligned with the direction of the derivative of the dipole moment. Owing to the sensitivity of vibrational spectroscopy to the force constants between atoms, these methods are often used to characterise molecules, but also to observe the changes they undergo when they interact with their surroundings. Moreover, in ordered systems, these methods can be used to determine the orientations of molecular groups.

When nanoscale entities are dispersed in a liquid or solid medium, vibrational spectroscopies in their conventional form can be used to specify the nature of bonding in such systems. In this way Raman spectroscopy can

unambiguously ascertain the state of C–C bonds in molecules or macrostructures like C_{60} or hard carbon [117]. In some cases, owing to the size of the objects, particularly in the intermediate range between the molecular state and a crystal state, there may be some difficulty in interpreting vibrations.

As mentioned previously, vibrational spectroscopy is well suited for the study of nano-objects deposited on a surface. However, there is then a double-edged problem of detecting the bands and understanding the spectra. Owing to the small size of the objects, their characteristic absorption can only be very weak. One must therefore use methods able to observe these systems with great sensitivity. In addition, in the vicinity of an interface, owing to the presence of an incident wave and a reflected wave, anisotropic electric fields will form which depend on the polarisation of the incident wave. From (9.51), the spectra will be sensitive to the orientation of the molecular groups. Moreover, the infrared spectra of thin films are often distorted, and this all the more so as the mode becomes more intense, due to the presence of multiple reflections within the film.

In Sect. 9.4.2, we shall discuss the rules governing electric fields at interfaces, the standard IR transmission and reflection techniques that follow from these rules, and modulation techniques used to enhance the sensitivity of infrared spectroscopy. In Sect. 9.4.3, we describe current possibilities in resonance Raman spectroscopy and surface-enhanced Raman spectroscopy. Finally, in Sect. 9.4.4, we discuss current prospects for vibrational methods in combination with the techniques of near-field microscopy.

9.4.2 Infrared Spectroscopy

External Reflection. Infrared Reflexion Absorption Spectroscopy (IRRAS)

Whatever the type of surface or interface infrared spectroscopy, it is essential to determine the electric fields produced close to the interface in order to understand the infrared spectra that may be produced. To do this, one must first define the polarisation of an electromagnetic wave which encounters the interface.

An electromagnetic wave is said to be parallel polarised (p) when the electric field of the wave is contained in the plane of incidence defined by the normal to the surface and the direction of the incident wave (see Fig. 9.42 top). An electromagnetic wave is said to be perpendicularly polarised (s) when the electric field of the wave is perpendicular to the plane of incidence (see Fig. 9.42 bottom). It is usual to define a coordinate system associated with the interface, in which z is normal to the surface, x lies in the plane of incidence, and y is perpendicular to the plane of incidence. It follows that, in the commonest case, when there is no rotation of the plane of polarisation during reflection, the p polarisation can only produce a field in the z or x directions, whereas the s polarisation can only produce a field in the y direction.

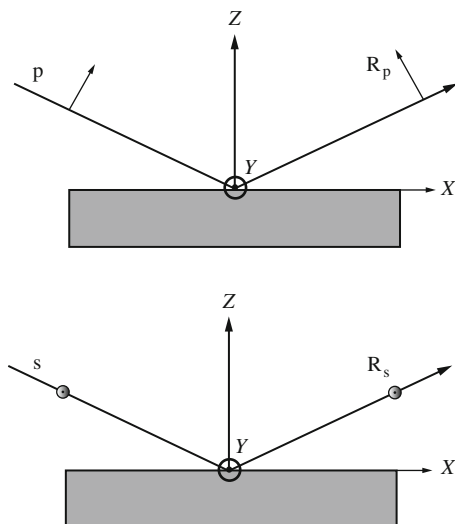


Fig. 9.42. Definition of the p and s polarisations and the X , Y , and Z axes for a plane interface

The distribution of the fields in the vicinity of a substrate depends only on the absorption by the substrate at the frequency ν of the exciting wave. This absorption is directly related to its complex dielectric constant $\hat{\epsilon}(\nu) = \epsilon_r + i\epsilon_i$, and hence to its complex refractive index $\hat{n}(\nu) = n + ik$, since, in the case where the magnetic susceptibility is equal to 1, $\hat{\epsilon}(\nu) = \hat{n}^2$. The extinction coefficient k is directly related to the absorption of the medium, but also to its conduction at the frequency of the exciting wave [118]. Two different types of behaviour may thus be expected close to a surface, depending on whether the substrates are conducting (absorbent), e.g., metals, or insulating (dielectric), e.g., silica. The anisotropic fields near surfaces arise due to the combination of incident and reflected fields, owing to the fact that the reflected wave undergoes a temporal phase shift directly related to the complex index of the surface.

Conducting Substrates in the Infrared

For the main part, this category concerns metals and some semiconductors, together with indium-doped tin oxide (ITO), used as a transparent conducting electrode in most display systems. As can be seen from Fig. 9.43, at the surface of a metal, the fields produced by an electromagnetic wave are highly anisotropic and depend on both the polarisation and the angle of incidence of the wave.

An s polarised wave produces a very weak field, whereas the p polarised wave produces a field only in the z direction. This behaviour is explained by the fact that it is almost impossible to create an electrostatic potential at the

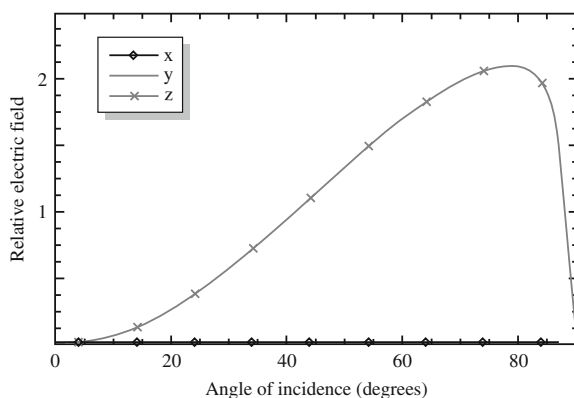


Fig. 9.43. Dependence of the electric field strength in the x , y , and z directions at the surface of a metal on the value of the angle of incidence

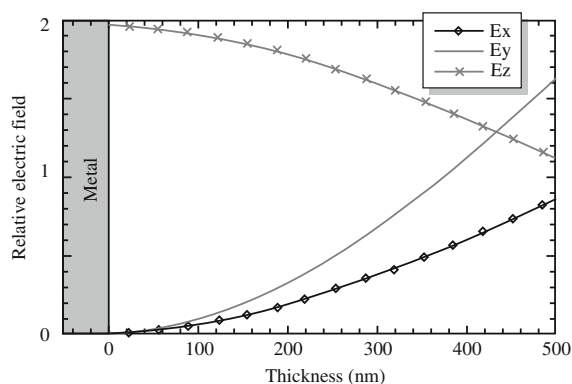


Fig. 9.44. Dependence of the electric fields in the x , y , and z directions on the distance to the surface of a metal ($i = 75^\circ$)

surface of a conductor. Moreover, the field produced by p in the z direction gets stronger and goes through a maximum at grazing incidence.

As can be seen from Fig. 9.44, the anisotropy of the fields extends over a significant distance. It occurs over several thousand angstroms in the infrared. Thin films of this thickness will thus be excited anisotropically in the x , y , and z directions.

The above properties will determine which infrared method should be used to study thin films at the surface of a conductor. Due to this conduction and the associated absorption, the sample cannot be analysed by transmission. Samples are therefore studied only by reflection.

According to (9.51), the intensities of the bands increase with the magnitude of the exciting field. On a conductor, one must therefore use a large angle of incidence to satisfy this condition.

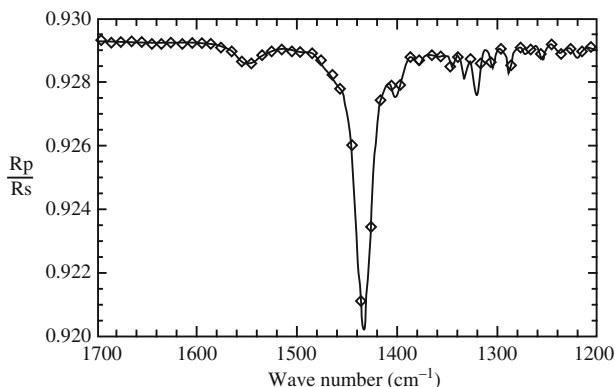


Fig. 9.45. IRRAS spectrum of five layers of cadmium arachidate on gold (thickness 12.5 nm)

Since FT infrared spectrometers operate with a single beam (one records a reference signal, then the sample signal, and it is the ratio of the second signal to the first that gives the sample spectrum), and since the *s* polarised wave produces no field, the latter can be used as the reference signal.

The *p* polarised wave for its part can be used to detect the spectrum of molecules deposited on the surface. The resulting spectrum is found from the ratio of the polarised reflectances, i.e., $S = R_p/R_s$. The IRRAS technique has two advantages when used on conductors:

- At grazing incidence, the field is intensified and the absorption bands are therefore also more intense.
- Since the field is only in the *z* direction, $\mathbf{E}$ can be replaced by $\mathbf{E}_z$ in (9.51). θ is then the angle between the normal to the surface and the change in the dipole moment. There is thus a surface selection rule which indicates that the absorption intensities are proportional to the square of $\cos \theta$.

It follows that an absorption in the plane is not detected, while an absorption in the *z* direction will be intensified. An example of the application of this rule is given in Fig. 9.45.

The IRRAS spectrum shown in Fig. 9.45 was produced by five Langmuir–Blodgett layers of cadmium arachidate on gold. The thickness of the resulting film was of the order of 12.5 nm. The molecule comprises a long carbon chain $(\text{CH}_2)_{18}\text{CH}_3$ and a carboxylate polar tail COO^- . The intense vibration at 1418 cm^{-1} is characteristic of the symmetric vibration of the carboxylate, while the weak band at 1560 cm^{-1} is characteristic of the antisymmetric vibration of the same group. The spectrum of an isotropic phase of this same compound shows that the antisymmetric vibration is twice as intense as the symmetric one. The discrepancy observed in the spectrum of the thin film suggests a preferred orientation of the system with the C_2 axis of the carboxylate close to the normal to the surface. The chains are oriented almost

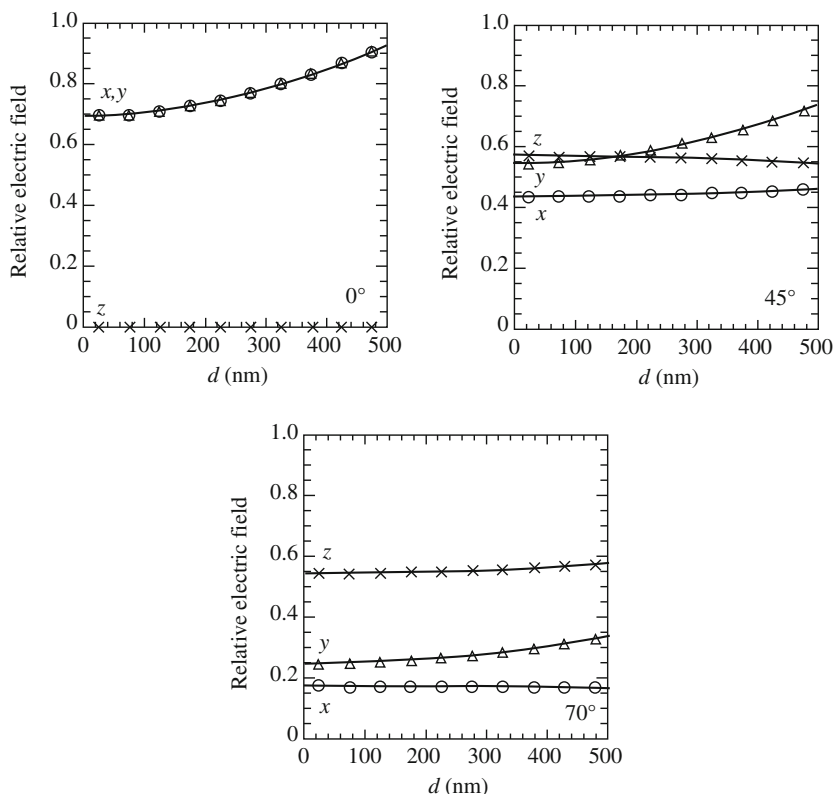


Fig. 9.46. Electric fields in the x , y , and z directions at the surface of a dielectric as a function of the distance d to the surface for different values of the angle of incidence i ($i = 0^\circ, 45^\circ, 70^\circ$)

vertically, because the CH_2 deformations near $1,450\text{ cm}^{-1}$ have disappeared, while bands corresponding to coupling between the wagging and twisting vibrations of the CH_2 groups, normally very weak, appear between $1,350$ and $1,250\text{ cm}^{-1}$, directed along the chain axes.

There is one last point, specific to the excitation of molecules by a field normal to the surface. The electric field has a longitudinal direction relative to the film. This is why it is the longitudinal optical (LO) components of each mode that appear and not the transverse optical (TO) components. The LO components are shifted to higher frequencies than the TO modes in most cases. The shift is greater when the absorptions are more intense. It is of the order of a few tenths of cm^{-1} for weak bands, and a few cm^{-1} for the stronger bands of organic compounds, and can exceed $100\text{--}200\text{ cm}^{-1}$ for the very strong absorptions of inorganic compounds.

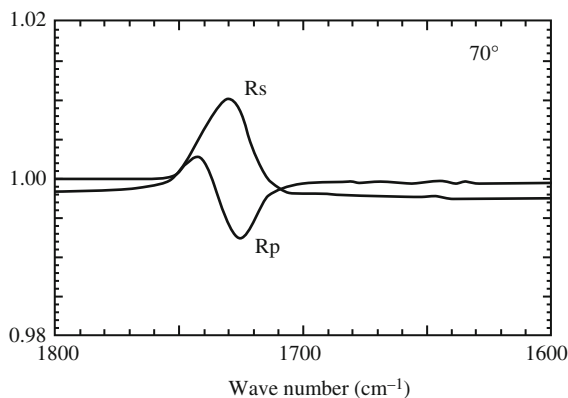


Fig. 9.47. IRRAS spectra along p and s for a thin film (20 nm) of PMMA on glass. Normalisation by the p and s spectra of the glass reference

Dielectric Substrates in the Infrared

It is difficult to predict the anisotropy of the fields at the surface of a dielectric. Only a calculation using the Fresnel relations can obtain their distribution as a function of the polarisation and the angle of incidence. Figure 9.46 shows the values and the extent of the fields for three different angles of incidence on a dielectric of index 1.45. Note that there is no intensification phenomenon in any privileged direction and that the field distribution varies considerably with the angle of incidence. The field in the z direction dominates for large angles, but remains of the same order of magnitude as the field in the y direction produced by the s component.

Due to these cross-overs of the electric fields, the spectra can become complicated, because the TO and LO components will be able to appear simultaneously. In order to simplify analysis of the spectra, it is better to use just the s component, which produces a field only in the y direction. The angle of incidence is then chosen so that the bands are as intense as possible. It can be shown that, for a substrate with index $n = 1.45$, this angle is close to 60° . When the spectrum is recorded, the signals are normalised by the spectrum of the substrate used as reference. Figure 9.47 shows, in the frequency range of the carbonyls, the p and s polarised spectra of a thin film of polymethylmethacrylate (PMMA) on glass.

The spectrum of the p components is more complex than that of the s component, because it contains the TO and LO components in opposite directions. The TO component lies at 1732 cm^{-1} and is directed downwards, while the LO component is at 1740 cm^{-1} and is directed upwards. The component R_s is simpler since it only contains the TO component. However, this component is reversed relative to its observation with the p polarisation. This example shows that it is possible to study thin films on a dielectric, in particular, using the s polarisation. The selection rule in this case is rather simple. In

fact, it is the opposite of the rule for conducting substrates, viz., the intensity of a band is proportional to the square of the sine of the angle θ . The bands will be intense for components $(\partial\mu/\partial q)$ lying in the plane, and extinguished when the directions of the derivatives of the dipole moment are vertical.

Comment. It is relatively easy to obtain an infrared reflection spectrum on a dielectric. However, interpretation of the result can prove very difficult when close to an absorption band of the substrate. Indeed, it is almost impossible to understand the spectral distortion due to mixing between the vibrational modes of the substrate and the film.

To sum up, IRRAS spectroscopy is a relatively simple technique for obtaining information about the nature, organisation, and orientation of molecules near surfaces. However, this method has the disadvantage of requiring two spectra recorded at different times. Fluctuations of the spectrometer and the environment limit the signal-to-noise ratio of the spectra and make it difficult to obtain spectra from monolayers or sub-monolayers. To increase the signal-to-noise ratio, a new, real-time differential technique called polarisation modulation IRRAS has been developed for studying thin films on metals [119].

Polarisation Modulation Infrared Absorption Spectroscopy (PMIRRAS)

The idea here is to modulate the polarisation of the incident beam at a high frequency compared with the frequencies of bands produced by an infrared interferometer. The experimental setup for PMIRRAS, depicted in Fig. 9.48, shows that the infrared beam leaving the spectrometer is first polarised, then goes through a photoelastic modulator which produces a high-speed modulation of the polarisation between states p and s . The beam is then reflected off the sample at grazing incidence, before being focused onto a sensitive detector.

The photoelastic modulator is made from an isotropic crystal which becomes alternately birefringent by means of the strain produced at 45° by piezoelectric crystals modulated at frequency ω_m . The signal output from the detector has the form

$$I_d = CI_0(\omega_i) \left\{ [(R_p + R_s) + (R_p - R_s)J_0] + (R_p - R_s)J_2 \cos 2\omega_m t \right\}, \quad (9.52)$$

where C is a constant depending on the detector response at each frequency, ω_i are the frequencies of intensity modulation produced by the interferometer, usually lying in the range from 100 Hz to 5 kHz, J_2 and J_0 are Bessel functions introduced by the modulator due to the fact that it cannot be half-wave simultaneously for all wavelengths, ω_m is the modulation frequency of the piezoelectric crystals ($\omega_m \propto 30$ kHz), and R_p and R_s are the p and s polarised reflectances, respectively.

The signal I_d contains two terms, one modulated at low frequency that is particularly sensitive to the sum of the polarised reflectances, and the other

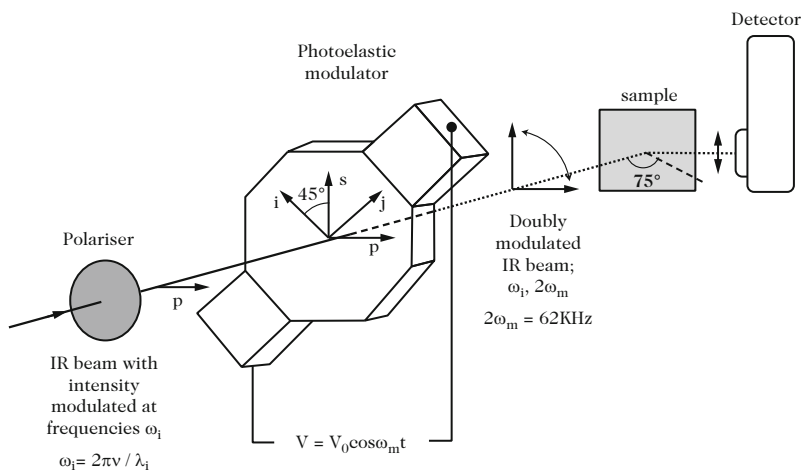


Fig. 9.48. Experimental arrangement for PMIRRAS

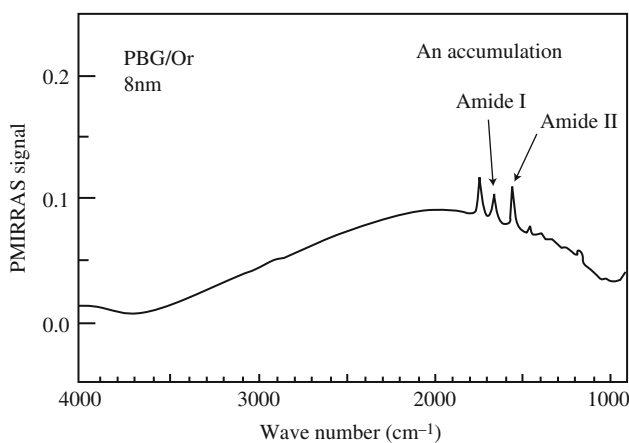


Fig. 9.49. PMIRRAS spectrum of a polybenzylglutamate film on gold, made by scanning with the movable mirror of the interferometer

doubly modulated at low frequency and at high frequency ($2\omega_m$), proportional to the difference between the reflectances. The two components can be separated electronically to yield the spectra of the sum and difference parts. By normalising the difference signal by the sum signal, one obtains a signal

$$S = \frac{(R_p - R_s)J_2}{(R_p + R_s) + (R_p - R_s)J_0} \quad (9.53)$$

Since R_p is close to R_s on a conductor, this signal simplifies to give

$$S = \frac{(R_p - R_s)J_2}{R_p + R_s} \quad (9.54)$$

This signal is only sensitive to absorptions depending on the polarisation. The spectrum will not be sensitive to isotropic absorption of the surroundings, e.g., H_2O , CO_2 , etc. R_s is used as reference signal and R_p as sampling signal. In fact, PMIRRAS spectroscopy can be used to transform the single-beam interferometer into a double beam. Figure 9.49 gives the PMIRRAS spectrum obtained in a single scan (acquisition time around one second) of a thin film (8 nm) of a synthetic polypeptide, viz., polybenzylglutamate (PBG), on gold. The gain of at least one order of magnitude in the signal-to-noise ratio means that a reasonable spectrum can be obtained over a short time span. PMIRRAS spectra on conductors have the same selection rule as IRRAS. From the spectrum shown in Fig. 9.49, it can be deduced that the polypeptide has an α -helical secondary structure and that the axes of the helices lie in the plane of the substrate.

The spectrum has a characteristic shape, arising from the presence of the Bessel function J_2 in the expression for S . The spectrum can be corrected for this function to give the IRRAS signal R_p/R_s [119].

In the dielectric case, (9.53) cannot be simplified. In addition, since the polarised reflectances are very different, the signal contains a significant contribution from the substrate. The spectra of thin films can only be observed by normalising by the PMIRRAS signal of the substrate alone, viz., $S_n = S(\text{sample})/S(\text{substrate})$. PMIRRAS was originally developed only to study films on conducting substrates, but in 1992, Blaudez et al. [120] showed that it could also be used on non-conducting solid or liquid surfaces. The study showed that PMIRRAS could only be applied to dielectrics for a limited range of angles of incidence, which depends on the index of the substrate, e.g., the optimum angle is 75° for glass and water. On this type of substrate, there is a novel selection rule which leads to a situation where bands associated with vibrations in the plane appear positively in the spectra, whereas bands associated with vibrations perpendicular to the substrate appear negatively in the spectra. In addition, there is a ‘magic’ angle for which the bands disappear from the spectra. This angle is equal to 38° for water and 39° for glass. More quantitatively, the intensities of the bands are proportional to $\sin^2 \theta - \sin^2 \theta_m$, where θ_m is the magic angle. This selection rule can be checked for the spectrum of a cadmium arachidate monolayer deposited on glass, as shown in Fig. 9.50.

As observed in Fig. 9.45, the antisymmetric vibration of the carboxylate group at $1,565\text{ cm}^{-1}$ lies in the plane. For this reason, it gives a positive vibration on glass. The symmetric vibration of the carboxylate group at $1,445\text{ cm}^{-1}$ lies along the normal to the surface, and thus appears negatively in the spectrum. One can also make out a positive band at $1,465\text{ cm}^{-1}$, associated with deformational vibrations of CH_2 , because these lie more or less in the plane.

To sum up, using the various infrared transmission and reflection techniques, one can analyse nanoscale systems deposited on most substrates, thereby obtaining precise information about the kind of molecules, and their organisation and orientation. However, infrared spectroscopy is particularly

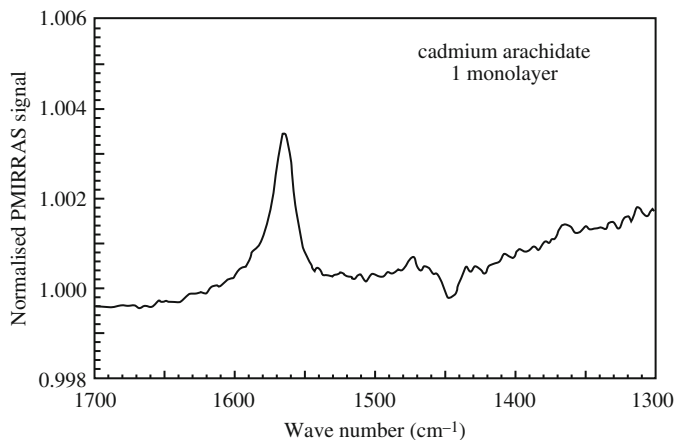


Fig. 9.50. Normalised PMIRRAS spectrum of a cadmium arachidate monolayer of thickness 2.5 nm on glass

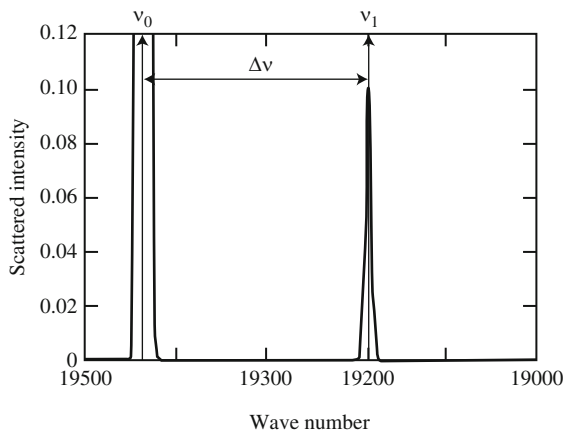


Fig. 9.51. Schematic view of a Raman spectrum. ν_0 is the Rayleigh scattering frequency of the exciter, and ν_1 the Raman scattering frequency. The frequency difference $\nu_0 - \nu_1$ is equal to the vibrational level ν_v

sensitive to polar vibrations, which lead to a significant change in the dipole moment. It is sometimes almost blind to certain vibrations. We shall see in the next section that Raman spectroscopy is then highly complementary since, in contrast to infrared spectroscopy, the polar modes have low intensity in Raman spectroscopy, whereas modes that are only slightly polar, which lead to significant changes in the electron cloud, yield intense bands in Raman spectroscopy. In addition, it is also possible to carry out infrared microscopic analyses in transmission and reflection modes. In this case, one can obtain spectra with lateral resolutions of the order of ten micrometers. It is nevertheless

impossible to use PMIRRAS and gain access with this lateral resolution to films with thicknesses less than 100 nm.

Infrared Transmission

In some cases, e.g., fluorine, germanium, silicon, etc., dielectrics can be transparent enough in the infrared to allow transmission spectra to be taken at normal incidence. In this case, the selection rule for the orientation of the transition moments is simple, because at normal incidence the wave can only produce an electric field lying in the plane. Silicon is particularly interesting for this type of study. For the purposes of transparency, it must be undoped and polished on both sides. However, the resulting transmittance spectra will have low intensities, because for thin films, transmission intensities are inversely proportional to $(2 + n_s)$, where n_s is the substrate index. Since silicon has an index of 3.4 in the infrared, this effect leads to a significant reduction in the signal.

9.4.3 Raman Spectroscopy

Basic Principles

Raman spectroscopy exploits an inelastic scattering effect with a change of frequency. It is not a resonant absorption process. For this reason, it is a low intensity process, only discovered in 1928 [121]. The underlying idea of the experimental arrangement for Raman spectroscopy is rather simple. A sample is excited at a monochromatic frequency that is high (UV, visible, or near IR) compared with frequencies in the mid-infrared, and then the frequencies emitted around the excitation frequency are observed using a dispersive grating. The frequency differences are directly related to the vibrational levels of the sample (see Fig. 9.51).

Since its invention, Raman spectroscopy has undergone a series of improvements which have made it today an extremely effective characterisation method. It became commonplace in the laboratory with the development of lasers and benefited successively from the improvement of detectors and the possibility of combining it with microscopy [122]. More recently, thanks to the invention of the notch filter, very bright spectrometers have appeared, able to obtain the sample spectrum in a very short lapse of time. Moreover, by combining with confocal microscopes as shown in Fig. 9.52, these devices can carry out 2D and 3D mapping with a lateral resolution of the order of $1\text{ }\mu\text{m}$ and a longitudinal resolution of $2\text{--}3\text{ }\mu\text{m}$. The size of the focal point is such that samples with a volume of $1\text{ }\mu\text{m}^3$ can be analysed, corresponding to a mass in the picogram range. As can be seen from Fig. 9.52, the excitation laser is totally reflected by the notch beam splitter M before being focused by the lens L_1 . Owing to the change in frequency, the Raman emissions go through M before being focused on the confocal pinhole P. The pinhole P only

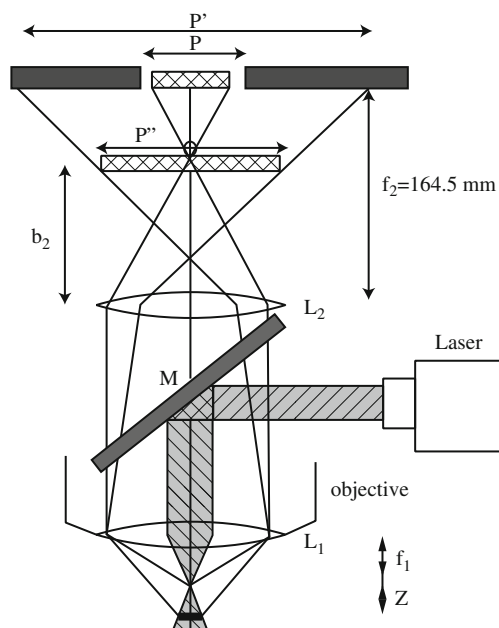


Fig. 9.52. Setup for a Raman confocal microscope

lets through the image of the focal point, because all other points outside the focal plane give images P' and P'' that are much bigger than the confocal hole and are therefore obturated.

After going through the confocal pinhole, the wavelengths are dispersed by a grating before being collected on a plane CCD detector. By confocality, the spectra of buried interfaces or objects under a surface (1–3 mm) can be obtained, since the matter above the sample makes no contribution to the Raman spectrum.

Comments. The Raman intensity is related to the values of the components of the order 3 differential polarisability tensor associated with the relevant vibration:

$$(\alpha') = \begin{pmatrix} \alpha'_{xx} & \alpha'_{xy} & \alpha'_{xz} \\ \alpha'_{yx} & \alpha'_{yy} & \alpha'_{yz} \\ \alpha'_{zx} & \alpha'_{zy} & \alpha'_{zz} \end{pmatrix}.$$

In standard Raman spectroscopy, this tensor is symmetric, and a vibrational mode is active if at least one of its components is nonzero. The fact that the Raman intensities depend on a tensor means that a field produced in one direction can produce emission in some other direction. In principle, this complicates the determination of orientations from interfacial fields.

However, for most vibrations, the diagonal terms, when they are nonzero, are more intense than the off-diagonal terms. It follows that, if a field is

produced in the (x, y) plane, it will be mainly those vibrations for which the terms $\alpha_{xx'}$ and $\alpha_{yy'}$ are nonzero that will appear in the spectrum. Vibrations sensitive only to $\alpha_{zz'}$ will tend to vanish. Consequently, there is a simple selection rule which also allows one to obtain information concerning the orientation of molecular systems.

Raman spectroscopy has two weak points:

- The sample is subjected to a very high power per unit volume, because all the energy of the laser is focused on a very small volume. Some samples cannot withstand this power input and are degraded, e.g., formation of carbon for some organic compounds, or change structure.
- The sample may contain small amounts of fluorescent impurities, and since the fluorescence phenomenon is much more intense than the Raman effect, the spectrum of these impurities can mask or distort the Raman spectrum of the sample.

Methods for Enhancing the Signal

Despite the developments discussed above, conventional Raman spectroscopy depends on an intrinsically weak signal. There is no particular difficulty in obtaining the spectra of pure products, or substances in sufficient concentration in solution, but it becomes problematic when there are not many molecules in the vicinity of the focal point, as happens for very dilute systems or systems located close to an interface. To get around this difficulty, one must appeal, as in infrared spectroscopy, to all those methods that can be used to enhance the Raman signal. Three such methods have been envisaged to this end.

Composition of an Incident and Reflected Wave

In Raman microscopy, excitation occurs along the axis of a microscope. For this reason, the electric field of the laser can only produce a field perpendicular to this axis, i.e., in the plane of the substrate if one is investigating a plane interface. Under such conditions, the rules are the same as for infrared spectroscopy on a surface. For a conducting substrate, the exciting wave produces no field in the plane, whereas this field is maximum for a dielectric. It follows that the same sample in an ultrathin film will not give a spectrum in the first case, whereas it will be easy to obtain one in the second. As in infrared spectroscopy, one can exploit the composition of incident and reflected fields to make combinations of layers that enhance the Raman spectrum of a film deposited on such a combination. Indeed, looking at the effective intensity of the electric field produced at the surface of a metal coated with a layer of silica, as shown in Fig. 9.53, one finds that the resultant field is much stronger at 700 Å from the surface than right up close to it. The ratio of the fields is of the order of 20, and this will lead to similar ratios for the Raman spectra. It is thus judicious to place the film or the system under investigation on an assembly of metal plus 70 nm of silica, rather than directly on the metal. These

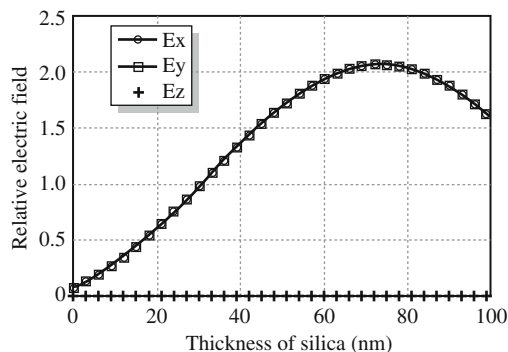


Fig. 9.53. Changes in the electric field produced at normal incidence in a thin film of silica on a metal ($\lambda = 488$ nm)

curves also confirm that, on a metal, the fields in the plane are very weak, so it would not be easy to obtain Raman spectra without designing some form of grazing-incidence focusing.

The curves of the effective fields show that the enhancement is produced by interference. Beyond 70 nm, the field begins to fall off and goes periodically through minima and maxima. A similar effect is obtained for substrates like silicon that are poorly absorbent but highly reflective.

Comment. This phenomenon is also effective in epifluorescence microscopy, where the fluorescence intensity depends on the amplitude of the effective electric field.

Resonance Raman Spectroscopy

Another way to enhance the Raman signal comes from the possibility of obtaining a resonance effect. In conventional Raman spectroscopy, the frequency of the excitation beam does not correspond to a frequency absorbed by the sample. However, it can be shown that, for some molecules, when the laser excitation comes close to an absorption band, there is a very strong intensification of some bands of the sample. In some case, this intensification can reach 10^6 , and the Raman effect is then resonant.

In order to obtain such conditions, several excitation wavelengths must be available in the Raman spectrometer, so that one can select the one giving the greatest enhancement. Resonance Raman spectroscopy involves one further difficulty, which arises from laser absorption by the sample. Without taking any particular steps, the sample will be degraded. In order to avoid this, several arrangements can be used to ensure that the sample is not always excited in the same place. For a solution, a circulation cell is used, whereas for solids, either the laser or the sample is made to vibrate.

Comment. Like fluorescence spectroscopy, resonance Raman spectroscopy has the advantage of being selective. Indeed, if a molecule comprises resonant and non-resonant groups, the behaviour of the various groups making up the molecule can be monitored by varying the excitation wavelength.

Surface Enhanced Raman Scattering

There is another way of intensifying the Raman signal, called surface enhanced Raman scattering (SERS). This effect was first demonstrated in 1975 on traces of pyridine adsorbed onto rough gold or silver films [123]. The enhancement can be considerable, arising essentially from the creation of very strong electric fields at the surface of metal particles in the rough film. Among other things, these fields result from resonant coupling of the fields between neighbouring particles. For this reason, surfaces must be prepared in a perfectly controlled way with regard to particle size and interparticle distances. Gold and silver particles are the best, owing to the indices of these metals in the visible.

Despite the drawback of having to use a particular type of surface, SERS has the double advantage that it does not use any excitation in the absorption band of the sample and that it is highly selective with regard to molecules close to metallic grains. The latter property is due to the fact that the extent of the enhanced fields is confined to the close vicinity of the grain surface.

9.4.4 Prospects for Vibrational Spectroscopy in the Study of Nano-Objects

As shown in the above sections, vibrational spectroscopy provides a very effective tool for studying groups or assemblies of nanosystems. This is particularly true in the study of carbon nanotubes by Raman spectroscopy. Some new approaches are currently being developed to carry out vibrational spectroscopy on single nano-objects. One of these approaches seeks to combine a near-field microscope with a spectrometer.

In this area, Raman spectroscopy is the most advanced (SNOM-Raman), because the problematic is the same as for scanning near-field optical microscopy (SNOM) in fluorescence spectroscopy. This method provides a way of going beyond the diffraction limit for visible light and reaching lateral resolutions of the order of a few hundred angstroms. In 2000, R.M. Stöckle et al. reported the Raman spectrum of a thin film of C₆₀ on glass [124]. They estimate that the probed region measures 55 nm, close to the dimensions of the tip apex used in the experiment (50 nm). The observed volume is thus very small, and it can be estimated that the number of molecules analysed under these conditions is of the order of 10⁴, far fewer than the number of molecules required to carry out optical Raman microscopy.

It is more difficult to adapt near-field microscopy in the infrared, owing to the lower sensitivity of the detectors, longer wavelengths than in the visible, and less intense absorption bands. However, several groups are developing

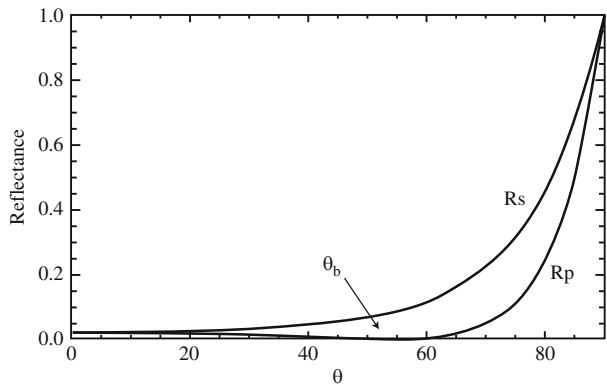


Fig. 9.54. Fresnel diagram of the components R_p and R_s polarised parallel and perpendicularly to the plane of incidence

Table 9.4. Dependence of the Brewster angle on the substrate index

Index	Brewster angle
1.333	$53^{\circ}12$
1.450	$55^{\circ}40$
2.400	$67^{\circ}30$

laboratory apparatus which, for some systems, should be able to demonstrate the feasibility of this type of experiment. In order to solve the problems of detection, luminosity, and coupling with the spectrometer, the light sources used are tunable laser diodes. Under these conditions, a sample is imaged at well-defined frequencies, characteristic of the absorptions. Despite the difficulties involved here, recent work has indeed demonstrated the feasibility of this type of experiment. For example, Taubner et al. [125] have reported quite remarkable results for a study of very small inclusions ($\approx 500\text{ nm}$) in a polymer, in which the field of view has dimensions $2.5 \times 3.5\text{ }\mu\text{m}$ and the lateral resolution is 100 nm . This resolution is much better than can be obtained with an optical microscope ($10\text{ }\mu\text{m}$) and there is some hope that such instruments will become commercially available within a few years.

To sum up, methods of vibrational spectroscopy are making progress with detection and resolution, in parallel with the miniaturisation of the systems to be investigated. There is every hope that instruments capable of systematic study of nanostructures will be available within ten years or so.

9.5 Brewster Angle Microscopy

Brewster angle microscopy was first developed at the beginning of the 1990s [128,129], exploiting the reflection properties of polarised light on a dielectric.

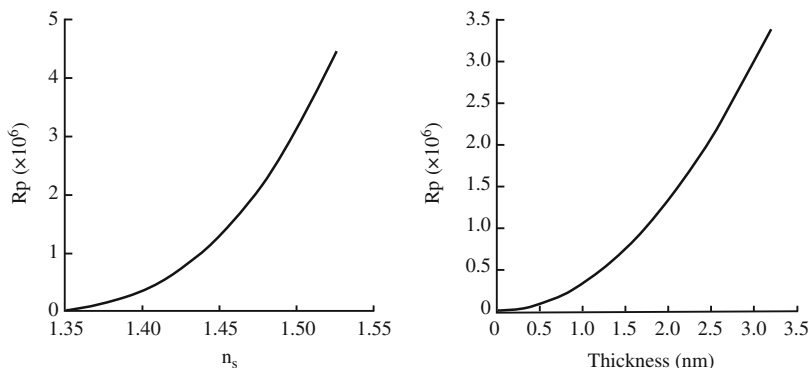


Fig. 9.55. Dependence of R_p at the Brewster angle on (a) the film index for a fixed thickness (2 nm) and (b) the film thickness for a fixed index ($n = 1.45$)

The Fresnel diagram of the polarised reflectances of a dielectric (non-absorbing for the wavelength under consideration) shows that there is an angle θ_b , called the Brewster angle, for which an electromagnetic wave polarised in the plane of incidence (p) is completely refracted and hence not at all reflected (see Fig. 9.54).

In air, the angle θ_b is related to the refractive index n_s of the substrate by $n_s = \tan \theta_b$ (see Table 9.4). At the Brewster angle, when the dielectric surface is coated by a thin film with index n_f that differs from the index of the substrate, part of the wave is reflected. As can be seen from Figs. 9.55a and b, the reflected intensity grows with the difference between n_f and n_s and also with the film thickness.

On a dielectric substrate, it is thus possible to detect the regions where molecules are located with great sensitivity, compared with non-coated regions of the substrate which do not give a reflected signal.

Instruments exploiting this idea have been designed to study domains formed by Langmuir and Gibbs films (monolayers of amphiphilic molecules that are insoluble or only partially soluble in water) on a water surface.

Figure 9.56 shows the setup used for this type of device. The main feature is a high-precision goniometer, providing very good control over the angles of the incident and reflected beams. One arm of the goniometer carries the excitation laser and its polariser, while the other arm carries a sensitive CCD camera. The reflected beam, having passed through the polariser, is focused by a microscope objective on the detector of the camera. A laser with a power of a few mW suffices to detect reflection from the thin film.

The objectives used typically produce magnifications of $\times 5$, $\times 10$, or $\times 20$, which gives fields of view of $900 \times 1,200 \mu\text{m}$, $450 \times 600 \mu\text{m}$, $200 \times 300 \mu\text{m}$, with corresponding lateral resolutions of $5 \mu\text{m}$, $2 \mu\text{m}$, and $1 \mu\text{m}$, respectively. Moreover, owing to the angle of observation, for a given position of the objective, only a narrow band of the image is clear. To overcome this problem,

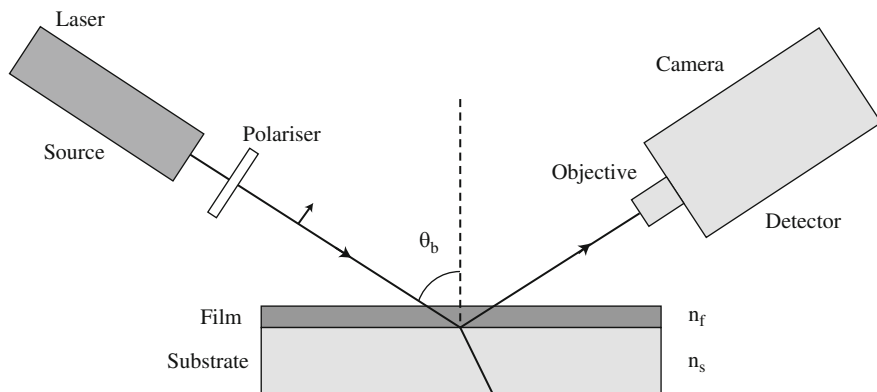


Fig. 9.56. Brewster angle microscope

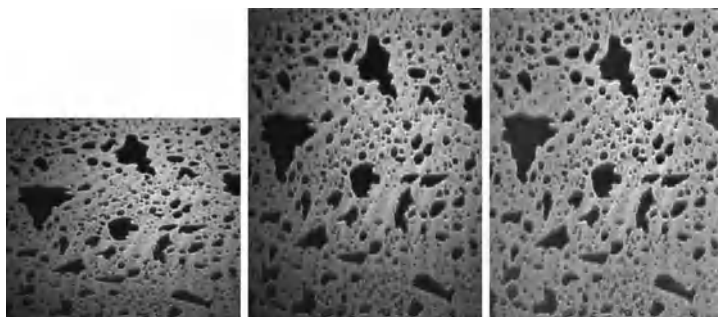


Fig. 9.57. Brewster images of a fluorinated lipid layer at the air/water interface. Each image measures 450 mm across the page and 600 mm in the other direction. *Left:* Original image. *Centre:* Image corrected for distortion of distances. *Right:* Image corrected for vignetting

when an image is recorded, a scanner is used to give the objective a to-and-fro movement, whereby focusing can be achieved successively on different parts of the field of view. The computer driving the device can then reconstitute a clear image by combining all the focused regions. The inclined incidence also compresses the image in the direction of observation. This distortion is easily corrected by an image processing program by multiplying the compressed direction by a factor equal to $1/\cos\theta_b$. For water, this factor is equal to 1.666.

It should be stressed that Brewster angle microscopy will only work in practice if reflection by the rear face is completely eliminated. For water, a black glass bevel is placed at the bottom of the container, and this almost completely absorbs the refracted wave.

Figure 9.57a shows an example of the kind of raw image produced by this technique for a film of fluorinated lipids deposited on water. The dark zones correspond to the water surface (or to zones in which the molecules are highly

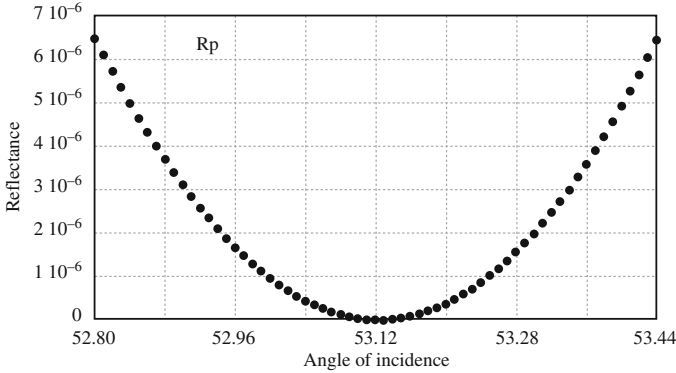


Fig. 9.58. Dependence of R_p on the angle of incidence for values around the Brewster angle in the case of water

diluted), while the bright zones correspond to microdomains in which most of the lipids have gathered. After dimensional correction, one obtains Fig. 9.57b, which better reflects the shape of the domains and their true sizes. The image is still impaired by a strong vignetting effect (the centre is much brighter than the edges), but this too can be corrected (see Fig. 9.57c).

In itself, Brewster angle microscopy is a relatively simple method for obtaining information about the organisation of microdomains on dielectric surfaces. But it has a greater potential, going beyond the simple recording of images. To begin with, the video signal from the camera can be recorded. It is thus possible to obtain a film of the formation of domains on a surface and hence to investigate the kinetic aspects and mechanisms of such structure formation, even though in this case only a single region of the image is focused. In addition, the images obtained express the absolute reflectances through the grey level recorded by the camera. If the grey levels can be calibrated as a function of the shutter speed of the camera, it should then be possible to determine the reflectances in order to evaluate the thickness of the observed films. Indeed, using the Fresnel relations, the reflectance R_p at the Brewster angle can be expressed in terms of the thickness e and index n_f of the thin film [see (9.55)]. Since the refractive index n_f can be estimated for a given film, it is then straightforward to evaluate the film thickness e .

The p polarised reflectance of a thin film of thickness e and refractive index n_f on a dielectric of index n_s observed at its Brewster angle is given by

$$R_p(e) = \left[\frac{n_f \cos \theta_b - \sqrt{(n_f^2 - \sin^2 \theta_b)(1 + n_s^2)}}{n_f \cos \theta_b + \sqrt{(n_f^2 - \sin^2 \theta_b)(1 + n_s^2)}} \right]^2 \frac{n_f^2 - \sin^2 \theta_b}{n_f^2} 16\pi^2 n_f^2 \left(\frac{e}{\lambda} \right)^2. \quad (9.55)$$

During an experiment, the camera must therefore be calibrated at its various shutter speeds. To this end, one can exploit the reflection properties of the substrate near the Brewster angle. If the index of the substrate is perfectly

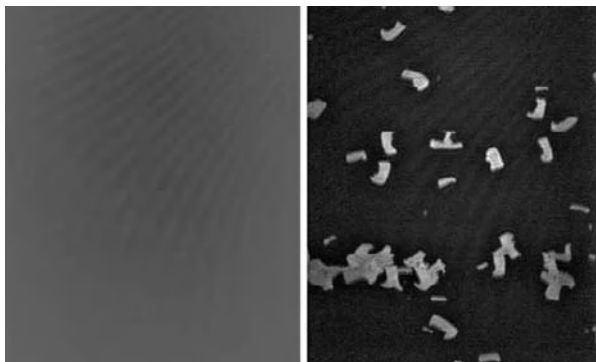


Fig. 9.59. Brewster images of cholesterol layers at the air/water interface. *Left:* Before collapse. Camera shutter speed 1/50 s. *Right:* After collapse. Camera shutter speed 1/500 s

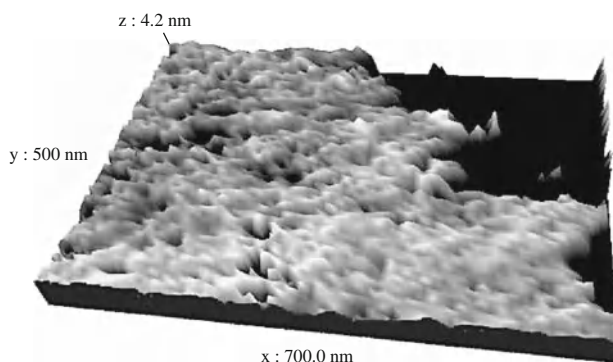


Fig. 9.60. Ellipsometric image of an organic film grafted on silicon ($\lambda = 532$ nm)

known, it is then possible to determine theoretically the curve of the polarised reflectances parallel to the plane of incidence around the Brewster angle. This function is shown for water in Fig. 9.58.

At the beginning of the experiment, the grey levels given by the camera are recorded for the substrate alone, ranging over the angles in the neighbourhood of θ_b . It then suffices to establish the correspondence between the grey levels and the absolute reflectance values. Figure 9.59 shows an example application of this method. Cholesterol forms a stable Langmuir monolayer at the air/water interface in Fig. 9.59a. However, beyond the ‘collapse’, domains begin to form (light grey in Fig. 9.59b), reflecting the 3D growth of the cholesterol. Figure 9.59a was recorded at 1/50 s, while Fig. 9.59b was recorded at 1/500 s, in such a way that the greater reflected intensity of the domains would not saturate the camera.

Using the calibration curves, we determined that the monolayer thickness was close to 1.8 nm, while that of the domains observed in Fig. 9.59b was close

to 5.2 nm, suggesting the formation of trilayers. These results show that the transition between the 2D and 3D cholesterol systems moves directly from a monolayer to a trilayer, without going through a bilayer. Moreover, the shape of the domains suggests that, at room temperature, they are in a crystalline or liquid crystalline state, because in the case of an isotropic fluid phase the domains at the air/water interface are circular.

Furthermore, adding a compensator after the laser polariser in a Brewster angle microscope setup, imaging ellipsometers have been constructed (Beaglehole Instruments, Nanofilm Technologie, Optrel, Sopra). In addition to benefiting from the lateral resolutions and field of view of the Brewster system, these instruments can study any kind of plane gas/liquid, gas/solid, liquid/solid, or liquid/liquid interface, with no restriction due to the kind of substrate. Figure 9.60 shows an example of the use of this type of instrument to study organic monolayers grafted on silicon. The bright region corresponds to the grafting of a layer of thickness 2.5 nm, while the black region corresponds to the silicon surface and hence to a region where grafting has failed.

To sum up, Brewster angle microscopy and ellipsometric imaging are relatively simple imaging techniques to implement and able to obtain topographic information about microdomains forming at plane interfaces. With ellipsometry, it is even possible to study buried interfaces. Although the lateral resolution with these methods is not greater than 1 μm , the longitudinal resolution is of the order of 0.1 nm.

9.6 Quartz Crystal Microbalance with Dissipation Monitoring (QCM-D)

9.6.1 Introduction

The quartz crystal microbalance (QCM) is a device for measuring the mass of a thin film, exploiting the fact that the resonance frequency of a quartz plate will change when its mass is increased by the deposition of the film (see Fig. 9.61).

This device was developed at the end of the 1950s to measure the mass of a rigid deposit, tightly bound to the crystal, e.g., a metal film, either in air or in vacuum. In the 1980s, the technique was extended to measurements in liquid environments, e.g., buffers, culture medium, solvents, etc. [130]. Hence, the QCM apparatus greatly contributed to the study of protein and polymer adsorption on liquid/solid interfaces [131], in particular due to its high sensitivity, reaching 1 ng/cm². More recently, such measurements have been extended to many other types of investigation, e.g., concerning the formation of lipid bilayers from lipid vesicles [132], specific antigen/antibody recognition [133], DNA hybridisation [134], conformational change in proteins after adsorption [135], cell adhesion and spreading [136], enzyme activity on an immobilised substrate [137], the biomineralisation of calcium phosphate [138],

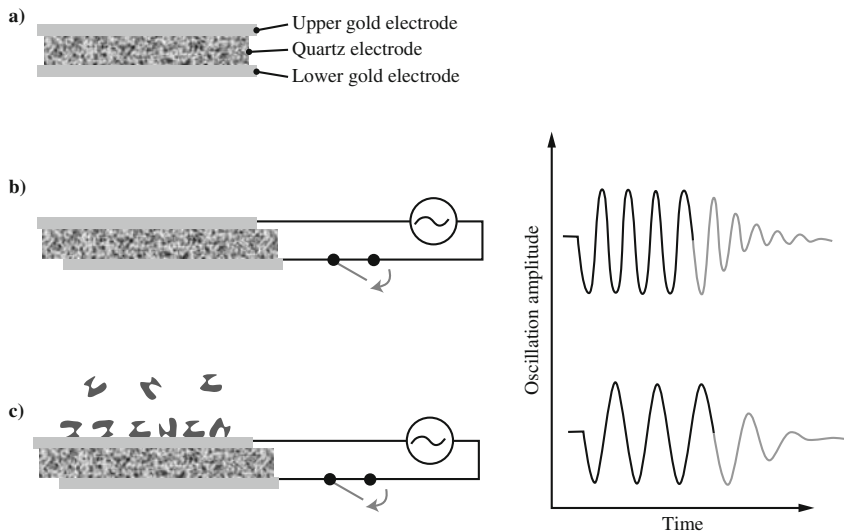


Fig. 9.61. (a) QCM setup showing a quartz crystal sandwiched between two electrodes causing it to oscillate when connected to a power supply. (b) Under a potential difference, the crystal oscillates at its natural frequency (*black curve*). When the current is switched off, the oscillation is observed to fall off (*grey curve*). (c) When material is adsorbed on the crystal in the form of molecules, cells, particles, etc., the oscillation frequency is reduced (*black curve*) and the damping of the oscillation after cutoff depends on the viscoelasticity and thickness of the deposit

bacterial adhesion and the formation of biofilms [131], and others. All these areas of investigation require greater and greater measurement accuracy and information enabling interpretation of the relevant mechanisms.

Under these conditions, energy dissipation in both the film and the liquid in contact with it must be taken into account. The crystal itself is in fact a damped oscillator, and a full theoretical analysis of the whole crystal–film–liquid system must take this into account [139]. In the rest of this section, we shall show how the viscoelastic characteristics of the film and the liquid, together with the damping coefficient of the crystal, are related to the resonance frequency of the whole system and to the damping of its oscillations when the exciting voltage applied to the crystal has been cut off (see Fig. 9.61). We examine successively the case of a crystal in vacuum, a crystal in contact with a viscous liquid, and finally a crystal coated with a homogeneous film in contact with a viscous liquid. We then show how the formalism can be extended to a virtually unlimited number of superposed films, each with its own characteristics (shear modulus, shear viscosity, density, thickness). The scope of this method will be illustrated by simulated and experimental examples.

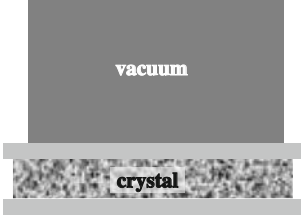


Fig. 9.62. Schematic view of a QCM crystal in vacuum

9.6.2 Vibration of a Damped Harmonic Oscillator Subject to Forces

Near its resonance frequencies (fundamental frequency and odd harmonics), the quartz crystal is treated as a damped harmonic oscillator with mass equal to half that of the crystal [140]. The damping parameter γ' and its stiffness constant k are related to the viscoelastic properties of quartz. Moreover, the viscoelastic properties of the film deposited on the crystal and those of the semi-infinite liquid in contact either with the film, or directly with the crystal, are modelled by Voigt–Kelvin elements, each of which comprises a spring and a piston in parallel [141]. Note that, if the film consists of several layers of different types, each can in principle be represented by its own element.

Let u be the displacement of an arbitrary point on the upper surface of the crystal, along the axis Ox , parallel to the crystal surface, which indicates its direction of vibration. In its most general form, the time dependence of u satisfies

$$\frac{1}{2}m_q \frac{\partial^2 u}{\partial t^2} = -\gamma' \frac{\partial u}{\partial t} - ku + F_p + F_{\text{ext}}, \quad (9.56)$$

where m_q is the mass per unit area of the crystal (in the following, $m_q/2$ will be denoted by M), t is the time, F_p is the force per unit area exerted by an arbitrary medium in contact with the upper face of the crystal, and F_{ext} is the external oscillating force (per unit area) imposed on the crystal by the voltage applied to it.

The description of the mechanical behaviour of the crystal makes heavy use of the Fourier transform (FT) of functions of time. The FT of a function f will be denoted by $\tilde{f}$.

9.6.3 Crystal in Vacuum

When the crystal vibrates in vacuum (Fig. 9.62) ($F_p = 0$), the equation of motion simplifies to

$$M \left(\frac{\partial^2 u}{\partial t^2} + \gamma_a \frac{\partial u}{\partial t} + \omega_1^2 u \right) = F_{\text{ext}}, \quad (9.57)$$

where $\gamma_a = \gamma'/M$ and $\omega_1 = \sqrt{k/M}$. The angular frequency ω_1 corresponds to the fundamental resonance frequency f_1 of the undamped oscillator

($\omega_1 = 2\pi f_1$). If a layer of perfectly rigid material, with mass m per unit area, is deposited on the crystal, this mass oscillates exactly like the upper face of the crystal, assuming that there is no slipping at the material/crystal interface. The force applied by the crystal to the deposited layer is then $m\partial^2 u/\partial t^2$. By the principle of action and reaction, the opposite force will apply to the upper surface of the crystal. This force corresponds to F_p in (9.56), which becomes

$$M \frac{\partial^2 u}{\partial t^2} = -\gamma' \frac{\partial u}{\partial t} - ku - m \frac{\partial^2 u}{\partial t^2} + F_{\text{ext}} . \quad (9.58)$$

The displacement u thus satisfies

$$(M + m) \left(\frac{\partial^2 u}{\partial t^2} + \gamma_b \frac{\partial u}{\partial t} + \Omega_1^2 u \right) = F_{\text{ext}} , \quad (9.59)$$

where

$$\gamma_b = \frac{\gamma'}{M + m} , \quad \Omega_1 = \sqrt{\frac{k}{M + m}} \approx \omega_1 \left(1 - \frac{m}{2M} \right) ,$$

assuming that $m \ll M$. We thus obtain the Sauerbrey relation [142], which gives the change in resonance frequency as a function of m :

$$\Delta f_1 = \frac{\Omega_1 - \omega_1}{2\pi} = -\frac{mf_1}{2M} = -\frac{m}{C_1} , \quad (9.60)$$

where C_1 is the Sauerbrey constant defined by

$$C_1 = \frac{2M}{f_1} . \quad (9.61)$$

This constant is given by the manufacturer of the crystal and is usually determined at the fundamental frequency f_1 . The deposited mass m can also be measured at a frequency that is an odd multiple of the fundamental frequency. For the harmonic of order ν ($f_\nu = \nu f_1$), the frequency shift is Δf_ν and the mass m is obtained from the Sauerbrey relation rewritten in the form

$$m = -\frac{C_1}{\nu} \Delta f_\nu . \quad (9.62)$$

This relation shows that if the deposited film behaves effectively like a perfectly rigid film, rigidly bound to the crystal, the ratios $\Delta f_\nu/\nu$, obtained for $\nu = 1, 3$, etc., are all equal.

9.6.4 Crystal in Contact with a Viscous Medium

Now consider a crystal in contact with a semi-infinite, viscous, and homogeneous medium, e.g., a buffer solution, but with no film deposited on its surface

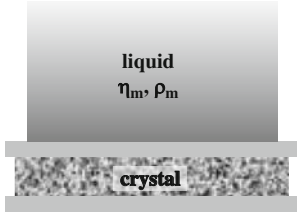


Fig. 9.63. Schematic view of a QCM crystal placed in contact with a viscous liquid characterised by its viscosity η_m and its density ρ_m

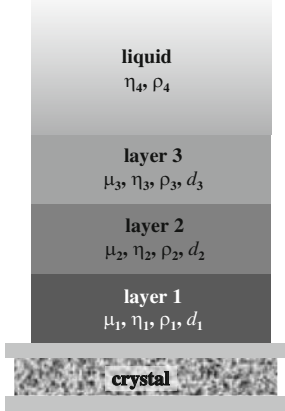


Fig. 9.64. Schematic view of a QCM crystal on which a film has been deposited. In this example, the film comprises three successive layers, each characterised by its shear modulus, shear viscosity, density, and thickness. A viscous liquid is in contact with the film

(Fig. 9.63). In the viscous fluid, characterised by its viscosity η_m and density ρ_m , the Navier–Stokes equation is

$$\frac{\partial \mathbf{v}_m}{\partial t} + (\mathbf{v}_m \cdot \nabla) \mathbf{v}_m = -\frac{1}{\rho_m} \nabla p + \frac{\eta_m}{\rho_m} \Delta \mathbf{v}_m + \mathbf{F}_m, \quad (9.63)$$

where $\mathbf{v}_m$ and $\mathbf{F}_m$ are the velocity of a fluid element and an external force applied to it, respectively, while p is the pressure. We shall neglect pressure effects and also the non-linear term $(\mathbf{v}_m \cdot \nabla) \mathbf{v}_m$. Moreover, we shall consider the case where there is no external force acting on the fluid. In addition, given that the fluid only moves in the direction of vibration of the crystal, the vector equation (9.63) reduces to a scalar equation in one dimension. This equation is satisfied by the unknown function v_m , which depends on the time and a single space variable, viz., the distance z to the upper surface of the crystal. Equation (9.63) now reduces to a scalar diffusion equation:

$$\frac{\partial v_m}{\partial t} = \frac{\eta_m}{\rho_m} \frac{\partial^2 v_m}{\partial z^2}. \quad (9.64)$$

The FT of this equation with respect to time is

$$\frac{\partial^2 \tilde{v}_m}{\partial z^2} = i\omega \frac{\rho_m}{\eta_m} \tilde{v}_m. \quad (9.65)$$

The general solution of this equation is

$$\tilde{v}_m(z) = \tilde{v}_0 \exp(-ik_m z) + \tilde{v}_1 \exp(ik_m z), \quad (9.66)$$

with

$$k_m = \sqrt{-i\omega \frac{\rho_m}{\eta_m}} = (1-i) \sqrt{\frac{\omega \rho_m}{2\eta_m}} = \frac{1-i}{\delta}, \quad (9.67)$$

where δ is the viscous penetration depth [141]. Since $\tilde{v}_m(z)$ must tend to 0 when z tends to infinity, the coefficient $\tilde{v}_1$ must be zero and (9.66) reduces to

$$\tilde{v}_m(z) = \tilde{v}_0 \exp \left[-(1+i) \sqrt{\frac{\omega \rho_m}{2\eta_m}} z \right] \equiv \tilde{v}_0 \exp \left[-(1+i) \frac{z}{\delta} \right]. \quad (9.68)$$

The effect of the viscous medium on the resonance frequency of the crystal will be further discussed in the next section, as a special case of the crystal–film–fluid system described below.

9.6.5 Crystal Covered with a Stratified Viscoelastic Medium in Contact with a Viscous Medium

We now consider the general case of a film made up of successive layers, each being homogeneous and bounded by planes parallel to the crystal surface (see Fig. 9.64). Due to the structure of the film, its properties only vary with the position z relative to the upper surface of the crystal, situated at $z = 0$. The upper face of the film is in contact with a viscous fluid which is considered to extend to $z = \infty$. We thus generalise the result due to Voinova et al. [141] for a film comprising one or two layers to a film with an arbitrary number n of layers. To do this, we adapt the method due to Abélès, developed in optics to calculate the reflectance coefficients for a film with a refractive index gradient.

Each layer is characterised by its shear modulus, shear viscosity, density, and thickness. Hence, the j th layer ($1 \leq j \leq n$) is characterised by the four parameters μ_j , η_j , ρ_j , and d_j . The fluid in contact with the film is treated as the $(n+1)$ th layer of the system and is characterised by its viscosity η_{n+1} and density ρ_{n+1} . These two parameters are denoted by η_m and ρ_m in Sect. 9.6.4, a notation that will be resumed later. Note also that the viscous fluid is semi-infinite, i.e., $d_{n+1} = \infty$, and has no rigidity, i.e., $\mu_{n+1} = 0$. Finally, all the layers are numbered from the crystal surface. The first layer lies between the planes $z = z_0 = 0$ and $z = z_1 = d_1$, and the j th layer between the planes $z = z_{j-1} = d_1 + d_2 + \cdots + d_{j-1}$ and $z = z_j = z_{j-1} + d_j$. The total thickness of the film is then $d \equiv z_n$. The viscous fluid occupies the region between $z = z_n$ and $z = \infty$.

To calculate the response of the crystal, we must find the shear stress σ_{xz} that the film applies to the upper surface of the crystal. Indeed, as we shall see, the force F_p [see (9.56)] is related to σ_{xz} . Since the film is stratified, we proceed layer by layer. To simplify the notation, we shall write σ for σ_{xz} . For the j th layer, the shear stress is defined by

$$\sigma_j = \mu_j \frac{\partial u_j}{\partial z} + \eta_j \frac{\partial v_j}{\partial z}. \quad (9.69)$$

It represents the force, parallel to the direction of vibration of the crystal, which applies to a unit area situated between the planes $z = z_{j-1}$ and $z = z_j$ and parallel to these. In (9.69), u_j and v_j denote the displacement and speed ($v_j = \partial u_j / \partial t$), respectively, of a point within the j th layer. These two quantities are functions of t and z . In the j th viscoelastic layer, the Navier–Stokes equation including elasticity leads to

$$\frac{\partial v_j}{\partial t} = \frac{1}{\rho_j} \frac{\partial \sigma_j}{\partial z} = \frac{\mu_j}{\rho_j} \frac{\partial^2 u_j}{\partial z^2} + \frac{\eta_j}{\rho_j} \frac{\partial^2 v_j}{\partial z^2}. \quad (9.70)$$

The FT of this equation with respect to time gives

$$\frac{\partial^2 \tilde{v}_j}{\partial z^2} = -\frac{\omega^2 \rho_j}{\mu_j^*} \tilde{v}_j, \quad (9.71)$$

where $\mu_j^* = \mu_j + i\omega\eta_j$ is the complex shear modulus of the j th layer. The general solution of (9.71) is

$$\tilde{v}_j(z) = \tilde{v}_j^+(z) + \tilde{v}_j^-(z), \quad (9.72)$$

where

$$\begin{cases} \tilde{v}_j^+(z) = \tilde{v}_{j,0}^+ \exp(ik_j z), \\ \tilde{v}_j^-(z) = \tilde{v}_{j,0}^- \exp(-ik_j z), \end{cases} \quad (9.73)$$

with $\tilde{v}_{j,0}^+$ and $\tilde{v}_{j,0}^-$ constants imposed by the boundary conditions on the two planes defining the layer, and k_j the complex number defined by

$$k_j = \omega \sqrt{\frac{\rho_j}{\mu_j^*}}. \quad (9.74)$$

Replacing μ_j^* by its definition, we obtain

$$k_j = \omega \left(\frac{\rho_j^2}{\mu_j^2 + \omega^2 \eta_j^2} \right)^{1/4} \left\{ \cos \left[\frac{1}{2} \arctan \left(\frac{\omega \eta_j}{\mu_j} \right) \right] - i \sin \left[\frac{1}{2} \arctan \left(\frac{\omega \eta_j}{\mu_j} \right) \right] \right\}. \quad (9.75)$$

The reciprocal of the real part of ik_j defines the penetration depth δ^* for a viscoelastic medium, generalising the definition of δ given by (9.67). This quantity is of great practical importance, because it corresponds to the characteristic distance over which the film can be probed by the microbalance. In the same way as Voinova et al. [141], we define the parameter $\chi = \mu_j / \omega \eta_j$, whence one has the following expression for δ^* :

$$\delta^* = \sqrt{\frac{2\eta_j}{\omega \rho_j}} \sqrt{\frac{1 + \chi^2}{\sqrt{1 + \chi^2} - \chi}} \equiv \delta \sqrt{\frac{1 + \chi^2}{\sqrt{1 + \chi^2} - \chi}}. \quad (9.76)$$

If $\mu_j = 0$, δ^* reduces to δ . Note that, for fixed ω , ρ_j , and η_j , the quantity δ^* increases monotonically with μ_j . However, for fixed ω , ρ_j , and μ_j , this quantity δ^* has a minimum at $\eta_j = \sqrt{3}\mu_j/\omega$, with value (see Fig. 9.65)

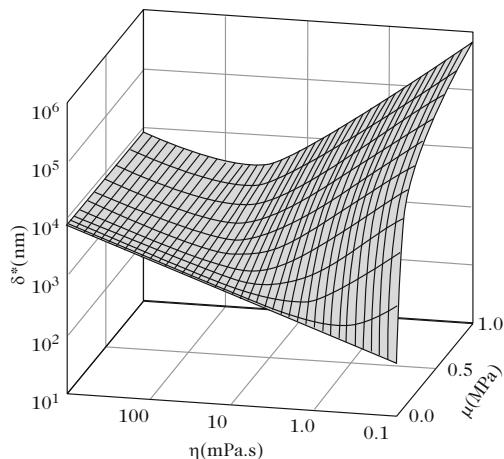


Fig. 9.65. Penetration depth δ^* as a function of the shear modulus μ and shear viscosity η , at $f = 5$ MHz and $\rho = 1$ g/cm³, calculated using (9.76)

$$\frac{2}{\omega} \sqrt{\frac{2\mu_j}{\rho_j}}.$$

Equation (9.72) can be written for the planes located at heights $z = z_{j-1}$ and $z = z_j$, which bound the j th layer. Using (9.73), it is then observed that the components of the velocity on the lower plane are related to those on the upper plane by the matrix relation

$$\begin{pmatrix} \tilde{v}_j^+(z_{j-1}) \\ \tilde{v}_j^-(z_{j-1}) \end{pmatrix} = \begin{pmatrix} \exp(-ik_j d_j) & 0 \\ 0 & \exp(ik_j d_j) \end{pmatrix} \begin{pmatrix} \tilde{v}_j^+(z_j) \\ \tilde{v}_j^-(z_j) \end{pmatrix}, \quad (9.77)$$

where $d_j = z_j - z_{j-1}$. In the following, the 2×2 matrix defined by (9.77) will be denoted $\mathbf{L}_j$. At the interface between layers j and $j+1$, the non-slipping (or continuity) condition implies the equality of the velocities and shear stresses (or their Fourier transforms) on either side of the interface:

$$\begin{cases} \tilde{v}_j(z_j) = \tilde{v}_{j+1}(z_j), \\ \tilde{\sigma}_j(z_j) = \tilde{\sigma}_{j+1}(z_j). \end{cases} \quad (9.78)$$

Combining (9.69), (9.72), (9.73), and $\tilde{u}_j = \tilde{v}_j/i\omega$, the FT of the shear stress can be expressed in the j th layer as

$$\tilde{\sigma}_j = \mu_j \frac{\partial \tilde{u}_j}{\partial z} + \eta_j \frac{\partial \tilde{v}_j}{\partial z} = \frac{\mu_j^*}{i\omega} \frac{\partial \tilde{v}_j}{\partial z} = \frac{k_j \mu_j^*}{\omega} [\tilde{v}_j^+(z) - \tilde{v}_j^-(z)]. \quad (9.79)$$

Then the relations (9.78) associated with (9.72) can be rewritten

$$\begin{cases} \tilde{v}_j^+ + \tilde{v}_j^- = \tilde{v}_{j+1}^+ + \tilde{v}_{j+1}^-, \\ k_j \mu_j^* (\tilde{v}_j^+ - \tilde{v}_j^-) = k_{j+1} \mu_{j+1}^* (\tilde{v}_{j+1}^+ - \tilde{v}_{j+1}^-), \end{cases} \quad (9.80)$$

at the interface, i.e., at $z = z_j$. These two equalities are expressed by the matrix relation

$$\begin{pmatrix} 1 & 1 \\ k_j \mu_j^* & -k_j \mu_j^* \end{pmatrix} \begin{pmatrix} \tilde{v}_j^+ \\ \tilde{v}_j^- \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ k_{j+1} \mu_{j+1}^* & -k_{j+1} \mu_{j+1}^* \end{pmatrix} \begin{pmatrix} \tilde{v}_{j+1}^+ \\ \tilde{v}_{j+1}^- \end{pmatrix}, \quad (9.81)$$

in which the four velocity components are still defined at $z = z_j$. The 2×2 matrix on the left-hand side of (9.81) is denoted by $\mathbf{M}_j$ and the one on the right-hand side by $\mathbf{M}_{j+1}$. Equation (9.81) then implies that

$$\begin{pmatrix} \tilde{v}_j^+ \\ \tilde{v}_j^- \end{pmatrix} = \mathbf{M}_j^{-1} \mathbf{M}_{j+1} \begin{pmatrix} \tilde{v}_{j+1}^+ \\ \tilde{v}_{j+1}^- \end{pmatrix}, \quad (9.82)$$

where the inverse of $\mathbf{M}_j$ is defined by

$$\mathbf{M}_j^{-1} = \frac{1}{2} \begin{pmatrix} 1 & 1/k_j \mu_j^* \\ 1 & -1/k_j \mu_j^* \end{pmatrix}. \quad (9.83)$$

Equations (9.77) and (9.82) can be used to relate the velocity in the first layer, at height $z = 0$, to the velocity in the liquid at height $z = z_n = d$:

$$\begin{pmatrix} \tilde{v}_1^+(z=0) \\ \tilde{v}_1^-(z=0) \end{pmatrix} = \mathbf{L}_1 \mathbf{M}_1^{-1} \mathbf{M}_2 \mathbf{L}_2 \mathbf{M}_2^{-1} \mathbf{M}_3 \cdots \mathbf{L}_n \mathbf{M}_n^{-1} \mathbf{M}_{n+1} \begin{pmatrix} 0 \\ \tilde{v}_{n+1}^-(z_n) \end{pmatrix}. \quad (9.84)$$

Note that $3n$ matrices appear in the matrix product. The first $3n - 1$ matrices contain the characteristics of the n layers making up the film, while the last matrix depends on the characteristics of the viscous medium. Note also that $\tilde{v}_{n+1}^+(z_n) = 0$ is imposed by the boundary condition $\tilde{v}_{n+1}(z \rightarrow \infty) \rightarrow 0$ [as in (9.66), which applies to the case where the liquid is in direct contact with the crystal]. The matrix product leads to a resultant matrix

$$\mathbf{R} = \begin{pmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{pmatrix},$$

and hence to

$$\begin{pmatrix} \tilde{v}_1^+(z=0) \\ \tilde{v}_1^-(z=0) \end{pmatrix} = \tilde{v}_{n+1}^-(z_n) \begin{pmatrix} R_{12} \\ R_{22} \end{pmatrix}. \quad (9.85)$$

Since $F_p = \sigma(z=0)$, it follows from (9.79) and (9.85) that the FT of the force per unit area exerted on the crystal by the film-liquid system is

$$\begin{aligned} \tilde{F}_p = \tilde{\sigma}(z=0) &= \frac{\mu_1^*}{i\omega} \frac{\partial \tilde{v}_1}{\partial z} \Big|_{z=0} = \frac{k_1 \mu_1^*}{\omega} [\tilde{v}_1^+(z=0) - \tilde{v}_1^-(z=0)] \\ &= \frac{k_1 \mu_1^*}{\omega} (R_{12} - R_{22}) \tilde{v}_{n+1}^-(z_n). \end{aligned} \quad (9.86)$$

Equation (9.85), associated with (9.72) and (9.73), leads to the FT of the velocity in the first layer at $z = 0$:

$$\tilde{v}_1(z=0) = \tilde{v}_1^+(z=0) + \tilde{v}_1^-(z=0) = (R_{12} + R_{22})\tilde{v}_{n+1}^-(z_n). \quad (9.87)$$

Given that $\tilde{v}_1 = i\omega\tilde{u}_1$, (9.87) becomes

$$\tilde{v}_{n+1}^-(z_n) = \frac{i\omega\tilde{u}_1}{R_{12} + R_{22}}. \quad (9.88)$$

Substituting this into (9.86), we obtain the expression for $\tilde{F}_p$ as a function of the characteristics of the film and the liquid and the FT of the displacement at $z=0$:

$$\tilde{F}_p = ik_1\mu_1^* \frac{R_{12} - R_{22}}{R_{12} + R_{22}} \tilde{u}_1(z=0). \quad (9.89)$$

By continuity (non-slipping condition), $\tilde{u}_1(z=0)$ in the film is the same as $\tilde{u}$, which is the FT of the displacement of a point in the upper surface of the crystal. Consequently, the FT of the equation of motion (9.56) leads to

$$\tilde{u} = \frac{F_0}{M \left(-\omega^2 + i\gamma_a\omega + \omega_1^2 - \frac{ik_1\mu_1^*}{M} \frac{R_{12} - R_{22}}{R_{12} + R_{22}} \right)}, \quad (9.90)$$

where it has been assumed that F_{ext} is given by $F_{\text{ext}} = F_0 \exp(i\omega t)$. It follows that $\tilde{F}_{\text{ext}}$ has been replaced by F_0 . The frequency for which the real part of the denominator vanishes on the right-hand side of (9.90) provides a good approximation for the resonance frequency, corresponding to the presence of the film deposited on the crystal and the viscous fluid above the film. It follows that the frequency shift is given by

$$\Delta f_{\text{film+fluid}} = \frac{\Delta\omega}{2\pi} \approx -\frac{1}{4\pi M\omega_1} \text{Re} \left(ik_1\mu_1^* \frac{R_{12} - R_{22}}{R_{12} + R_{22}} \right)_{\omega=\omega_1}. \quad (9.91)$$

For a harmonic oscillator governed by (9.57), the dissipation factor (the reciprocal of the quality factor) near the resonance is given by $D = \gamma_a/\omega_1$. When the crystal is coated with a viscoelastic film in contact with a viscous fluid, the dissipation factor becomes [143]

$$D_{\text{film+fluid}} = \frac{1}{\omega_1} \left[\gamma_a - \text{Im} \left(\frac{i\sqrt{\rho_1\mu_1^*}}{M} \frac{R_{12} - R_{22}}{R_{12} + R_{22}} \right)_{\omega=\omega_1} \right]. \quad (9.92)$$

In (9.91) and (9.92), the right-hand sides are evaluated at $\omega = \omega_1$. If the frequency shift and the dissipation factor are needed for a harmonic of order $\nu > 1$, then ω_1 is simply replaced by $\omega_\nu = \nu\omega_1$ in these results.

Equation (9.91) gives the change in the resonance frequency when the crystal is coated by a viscoelastic film in contact with a viscous medium, compared with the resonance frequency of the same crystal in vacuum. Equation (9.91) also gives the dissipation factor of the crystal for the crystal–film–fluid system.

From an experimental standpoint, it is convenient to use the crystal–fluid system (where the fluid can be a buffer solution, for example) as the reference

system. This system can be viewed as a particular form of the general system, where the parameters μ , η , and ρ of all the layers in the film are equal and equal to those of the viscous liquid. The matrix product R then reduces to $R = L_1 L_2 \cdots L_j \cdots L_n$. Since L_j is a diagonal matrix, whatever the value of j , R is also a diagonal matrix. Consequently,

$$\frac{R_{12} - R_{22}}{R_{12} + R_{22}} = -1.$$

It follows that the frequency shift and the dissipation factor become

$$\Delta f_{\text{buffer}} = \frac{\Delta\omega}{2\pi} \approx -\frac{1}{4\pi M} \sqrt{\frac{\rho_m \eta_m \omega_1}{2}} \quad (9.93)$$

and

$$D_{\text{buffer}} = \frac{1}{\omega_1} \left(\gamma_a + \frac{1}{M} \sqrt{\frac{\rho_m \eta_m \omega_1}{2}} \right), \quad (9.94)$$

where the index $n + 1$ has been replaced by m to indicate explicitly that these two equations apply to the particular film-viscous medium system. By subtracting (9.93) and (9.94) from (9.91) and (9.92), respectively, we obtain

$$\begin{aligned} \Delta f &= \Delta f_{\text{film+buffer}} - \Delta f_{\text{buffer}} \\ &\approx -\frac{1}{4\pi M \omega_1} \left[\text{Re} \left(i k_1 \mu_1^* \frac{R_{12} - R_{22}}{R_{12} + R_{22}} \right)_{\omega=\omega_1} - \omega_1 \sqrt{\frac{\rho_m \eta_m \omega_1}{2}} \right], \end{aligned} \quad (9.95)$$

$$\begin{aligned} \Delta D &= D_{\text{film+buffer}} - D_{\text{buffer}} \\ &\approx \frac{1}{M \omega_1} \left[\text{Im} \left(i \sqrt{\rho_1 \mu_1^*} \frac{R_{12} - R_{22}}{R_{12} + R_{22}} \right)_{\omega=\omega_1} + \sqrt{\frac{\rho_m \eta_m \omega_1}{2}} \right]. \end{aligned} \quad (9.96)$$

Equations (9.95) and (9.96) are the basic relations used to analyse experimental data in order to extract the film parameters. Here also, ω_1 should be replaced by $\nu\omega_1$ if the measurements correspond to harmonics of higher order. To avoid any ambiguity, it may be useful to specify that Δf and ΔD , given by (9.95) and (9.96), respectively, reflect the presence of the deposited film, but are not identical to the values of Δf and ΔD that would be measured if the film were in contact with the vacuum.

In the particular case of a film treated as a single layer (medium 1) deposited on the crystal and in contact with a viscous fluid (medium 2), the relevant ratio is given explicitly by

$$\left(\frac{R_{12} - R_{22}}{R_{12} + R_{22}} \right)_{\omega=\omega_1} = - \left[\frac{i \sin(k_1 d_1) + (1 + i) \sqrt{\frac{\rho_2 \eta_2 \omega}{2 \rho_1 \mu_1^*}} \cos(k_1 d_1)}{\cos(k_1 d_1) - (1 - i) \sqrt{\frac{\rho_2 \eta_2 \omega}{2 \rho_1 \mu_1^*}} \sin(k_1 d_1)} \right]_{\omega=\omega_1}. \quad (9.97)$$

According to (9.74), the product $k_1 d_1$ can be written

$$k_1 d_1 = \frac{\rho_1 d_1}{\sqrt{\rho_1 \mu_1^*}}.$$

Moreover, the product $\rho_1 \mu_1^*$ expands to give $\rho_1 \mu_1 + i\omega \rho_1 \eta_1$. As a consequence, medium 1 is characterised by just three independent parameters $\rho_1 \mu_1$, $\rho_1 \eta_1$, and $\rho_1 d_1$. This means that any increase (decrease) in ρ_1 can be exactly balanced by an opposing decrease (increase) in the parameters μ_1 , η_1 , and d_1 . Note also that the product $\rho_1 d_1$ represents the mass per unit area of the deposited film, including the solvent molecules present in the film [134, 144].

9.6.6 Numerical Simulation of the QCM Response

In the following, we consider changes in the resonance frequency and the dissipation factors of crystal–fluid and crystal–film–fluid systems. The film deposited on the crystal will be considered as a single, homogeneous layer ($n = 1$). This data will be simulated using (9.95), (9.96) and (9.97) for different combinations of the film thickness, shear modulus, and shear viscosity, keeping the density of the film fixed at 1 g/cm^3 . In addition, the density of the viscous fluid (solvent) will be fixed at 1.009 g/cm^3 , and its viscosity at $0.91 \text{ mPa}\cdot\text{s}$. These values correspond to an aqueous solution of NaCl with concentration 0.15 M at 25°C . We then examine the inverse problem, in which the quantities Δf and ΔD are measured and the film parameters have to be deduced from them.

In the first example (Fig. 9.66), the shear modulus is fixed at 0. The film is thus treated as a purely viscous medium. Its viscosity is fixed at $5 \text{ mPa}\cdot\text{s}$ (order of magnitude of the viscosities found by Höök et al. for a layer of proteins adsorbed on the crystal (see Table 1 in [134])). Figures 9.66A and B show $\Delta f/\nu$ and ΔD as a function of the film thickness d for the fundamental frequency ($\nu = 1$, $f_1 = 5 \text{ MHz}$) and the three odd harmonics at $\nu = 3, 5$, and 7 . The respective penetration depths are $\delta^* = 564, 326, 252$, and 213 nm . Recall that Δf and ΔD as given by (9.95) and (9.96) represent the changes in f and D caused by the presence of the film ‘inserted’ between the crystal and the solvent, where it is understood that the crystal–solvent system plays the role of reference. We observe that $\Delta f/\nu$ and ΔD become practically insensitive to d beyond a certain value of the thickness. This maximal probing distance is of the order of $2\delta^*$, which decreases with ν . Furthermore, it can be seen from Fig. 9.66A that the four curves are indistinguishable while the thickness remains below about 100 nm . This means that the film behaves like a Sauerbrey film for small enough thicknesses. Figure 9.66C shows ΔD as a function of $\Delta f/\nu$. The end point of each spiral corresponds to many values of the thickness, all of which lead to the same combination $(\Delta f, \Delta D)$.

The second example (Fig. 9.67) illustrates the case of a viscoelastic film which is both rigid (with stiffness $\mu = 0.2 \text{ MPa}$) and viscous (with viscosity

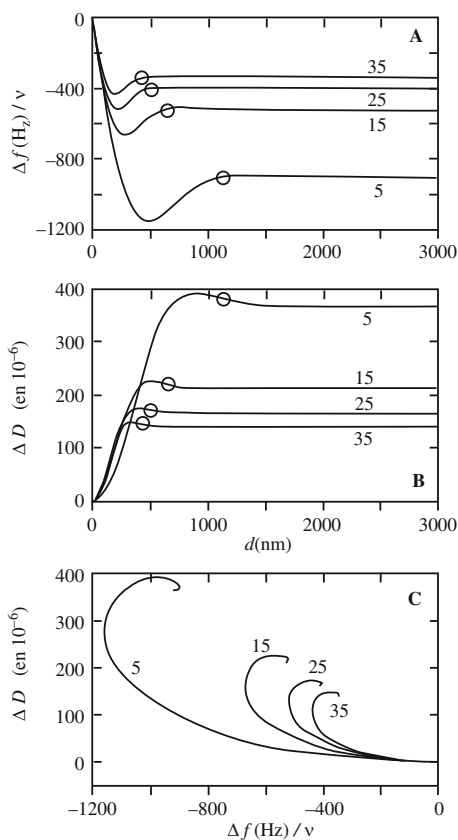


Fig. 9.66. Shift in the resonance frequency (**A**) and the dissipation factor (**B**) for the crystal covered with a single layer in contact with a viscous liquid, as a function of the film thickness d . The film is characterised by $\mu = 0$ and $\eta = 5$ mPa.s, and the fluid by $\eta = 0.91$ mPa.s. (**C**) Change in the dissipation factor plotted as a function of the change in the resonance frequency. Curves are labelled by the resonance frequency of the crystal in MHz. The crystal–fluid system is the reference. The density of the film is arbitrarily fixed at 1 g/cm^3 . The abscissa values of the circles placed on the curves in (**A**) and (**B**) give the value of $2\delta^*$

$\eta = 5$ mPa.s), its thickness varying from 0 nm to 6,000 nm. The numerical values of μ and η lie in the range of orders of magnitude found by Höök et al. for a layer of Mefp-1 (mussel adhesive) protein adsorbed onto the crystal coated with gold and crosslinked (see Table 1 in [134]). Figures 9.67A and B show $\Delta f/\nu$ and ΔD , respectively, as a function of the film thickness, while Fig. 9.67C shows ΔD as a function of $\Delta f/\nu$. Qualitatively, Fig. 9.67 resembles Fig. 9.66. However, there is a significant difference between the results from these two examples. The penetration depths for the four frequencies (5, 15, 25, and 35 MHz) are equal here to 1,553, 435, 295, and 237 nm, respectively. The increase in δ^* compared with the example shown in Fig. 9.66 means that the oscillations in the frequency shift and the dissipation factor undergo less damping, especially for the lowest frequency ($f_1 = 5$ MHz). It is interesting to note that, for the same value of the viscosity, which is the determining factor in the energy dissipation, the penetration depth increases with the elastic modulus, even though the latter is not involved in dissipation.

These two examples show that, for a given film, there is a limiting thickness, beyond which measurement becomes impossible. In other words, the

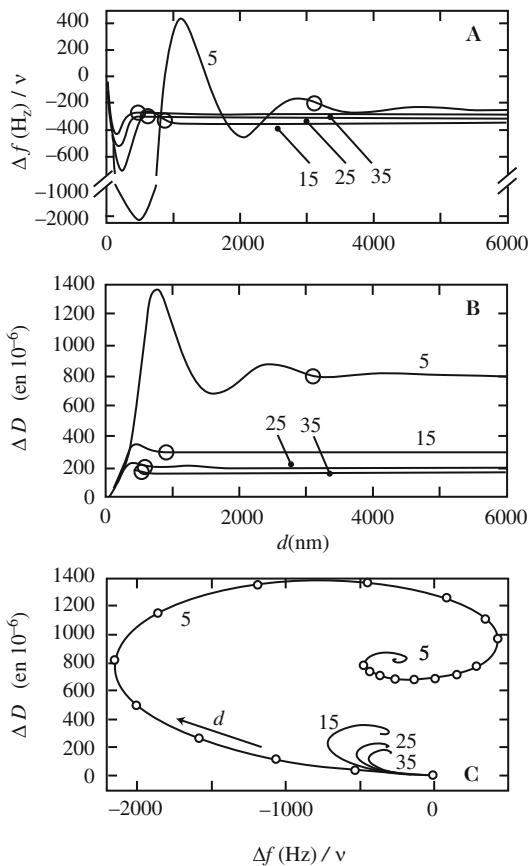


Fig. 9.67. Shift in the resonance frequency (A) and the dissipation factor (B) of a crystal covered with a single film in contact with a viscous liquid, as a function of the film thickness d . The film is characterised by $\mu = 0.2$ MPa and $\eta = 5$ mPas, and the fluid by $\eta = 0.91$ mPas. (C) Change in the dissipation factor plotted against the change in the resonance frequency. Curves are labelled by the resonance frequency of the crystal in MHz. The crystal–fluid system is the reference. The film density is arbitrarily fixed at 1 g/cm^3 . The abscissa values of the circles placed on the curves in (A) and (B) indicate the value of $2\delta^*$. In (C), the arrow indicates the direction of increasing thickness. White disks on curve 5 correspond to thicknesses 0, 100, 200, . . . , 2, 000 nm

Table 9.5. Polyelectrolytes and conditions for their use in the construction of PEI/(HA/PLL)_x multilayers

Polyelectrolyte	Standard abbreviation	pH of polyelectrolyte solution	Ionic strength of polyelectrolyte solution	Charge
Hyaluronic acid	HA	4	0.15 M	Negative
Poly-L-lysine	PLL	4	0.15 M	Positive
Poly(ethyleneimine)	PEI	4	0.15 M	Positive

growth of the film can only be monitored up to a certain thickness. Thus, due to the attenuation of the acoustic wave, the addition of further matter cannot be detected. The detection limit depends on the film parameters μ and η , which will obviously not be known in a real experiment. It is not therefore possible to predict the thickness beyond which the microbalance will no longer be able to inform about the construction of a given film.

9.6.7 Analysis of a Specific Experiment: Construction of a Polyelectrolyte Multilayer Film

Polyelectrolyte multilayer films are obtained by alternately depositing layers of polycations and polyanions on solid substrates. These deposits are accomplished by simply dipping the substrates alternately into polycation and polyanion solutions. The construction is therefore thoroughly straightforward, and the main advantage of these films resides in the accurate control over their thickness achieved by choosing the number of layers to be deposited.

As an example, we examine the construction of a polyelectrolyte multilayer film [145–150]. A precursor layer of poly(ethyleneimine) (PEI) is first adsorbed onto the crystal by simply setting the crystal in contact with a PEI solution (physisorption). This first deposit is followed by alternating deposits of a natural polysaccharide, hyaluronic acid (HA), and a polyamino acid called poly-L-lysine (PLL), dissolved in solutions of NaCl 0.15 M at pH 4 (see Table 9.5). Under these conditions, the HA chains are negatively charged (polyanions), while the PEI and PLL chains are positively charged (polycations).

The resonance frequencies and dissipations in the presence of NaCl alone are used as reference values and measured before depositing the polyelectrolytes. The resonance frequencies and dissipations are then measured after a rinsing phase that follows each polyelectrolyte deposit. Experimental results for the construction of a PEI/(HA/PLL)₈ film are given in Fig. 9.68. Figure 9.68 shows that the curves are not superposed, indicating that the film does not behave as a Sauerbrey film. This is confirmed by the high values of the dissipation (Fig. 9.68B). In addition, note that the frequency shift does not vary monotonically (Fig. 9.68A), which gives rise to the spirals shown in Fig. 9.68C. This example shows that the gradual increase in the mass of the film does not necessarily result in a systematic reduction in the resonance frequencies. Intuitive interpretation of the curves in Fig. 9.68A is therefore not without risk.

The experimental data were analysed using the formalism presented above to extract the film parameters. The film was treated as a single layer, with density arbitrarily fixed at 1 g/cm³. The parameters resulting from this analysis are shown in Fig. 9.69.

Note first of all that the parameters μ and η remain roughly constant in this example. It is worth mentioning that this is not always the case. It is not unusual for the parameter μ to fluctuate wildly, while η often tends to increase during film construction. The physical meaning of these variations remains obscure, so care must be taken when drawing conclusions from such results. On the other hand, the thickness increases in the way predicted by the model (apart from the small apparent reductions in thickness corresponding to deposition of layers HA₇ and HA₈, which may be due to a swelling/shrinking process in the film [149]).

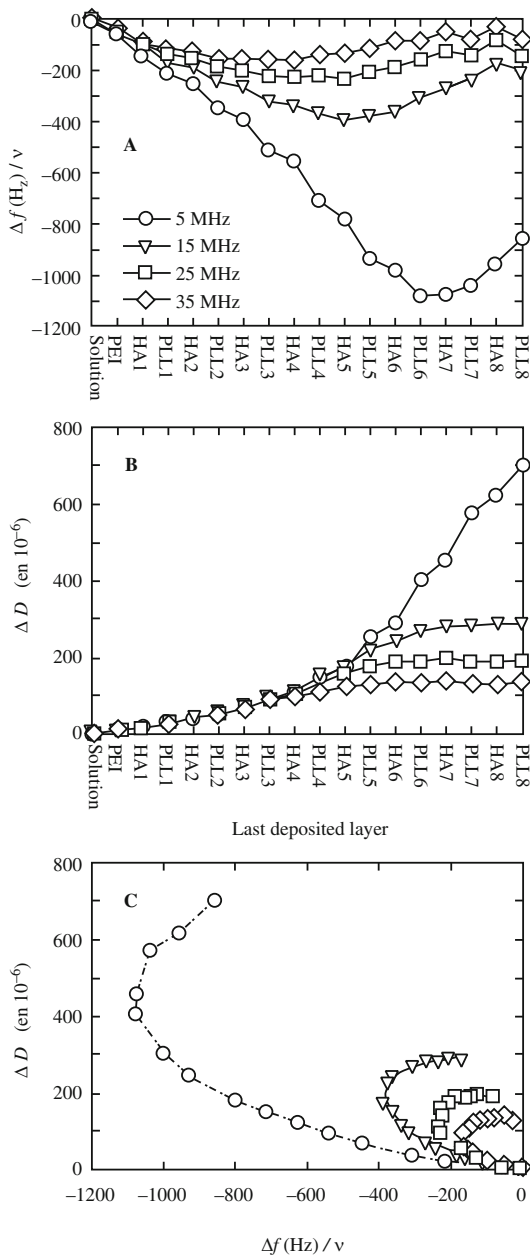


Fig. 9.68. Experimental results for the change in the resonance frequency (A) and the dissipation factor (B) observed during the construction of a polyelectrolyte multilayer film PEI/(HA/PLL)₈. Spirals appear in the graph of ΔD as a function of $\Delta f / \nu$ (C) due to the non-monotonic variation of Δf observed in (A)

In this section, we have generalised the analysis due to Voinova et al. for the processing of experimental data provided by the quartz crystal microbalance with dissipation monitoring (QCM-D). The examples described show that it is generally difficult to interpret these data intuitively and that one must

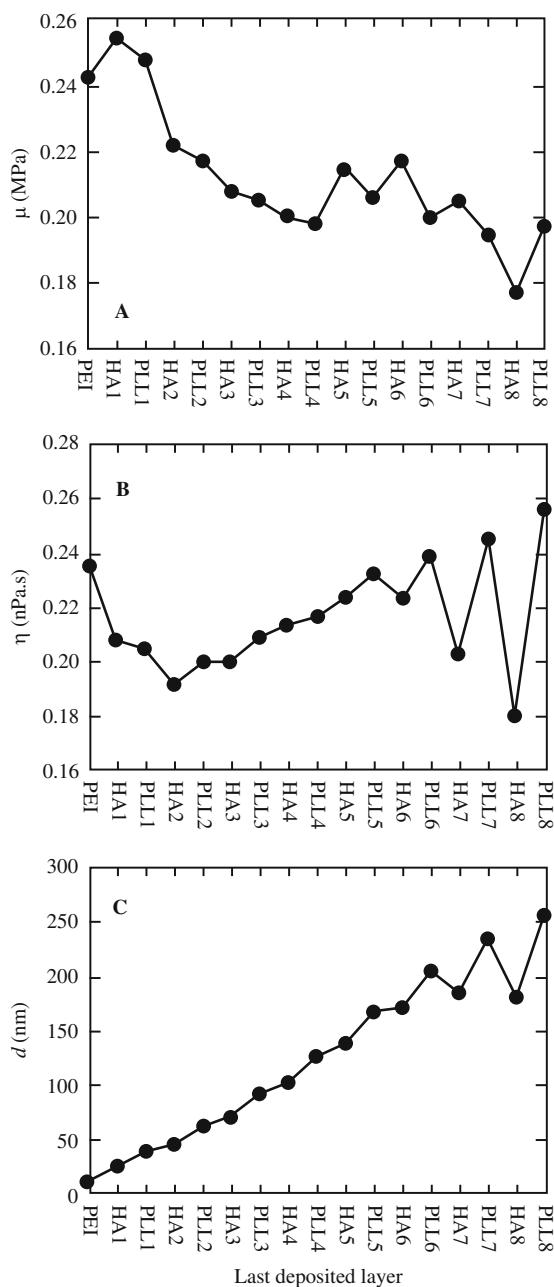


Fig. 9.69. Parameters for the PEI/(HA/PLL)₈ film constructed in the presence of an aqueous solution of NaCl at 0.05 M and pH 4. The parameters were deduced from experimental results assuming the film to have a density of 1 g/cm³. (A) Shear modulus. (B) Shear viscosity. (C) Thickness

therefore have recourse to a model that takes into account the changes in both the resonance frequencies and the dissipation factors, in order to extract at least the film thickness during the construction.

9.7 Grazing Incidence Neutron and X-Ray Reflectometry

X-rays, which have wavelengths in the nanometer range ($\lambda \approx 0.1 \text{ nm}$), were widely used during the last century to determine the structure of crystals, from the atomic scale to the highly complex assemblies involved in proteins and viruses. Today, with the development of synchrotron radiation sources, new approaches based on X-ray reflectometry are used to analyse liquid and solid surfaces and the structure of thin films on the nanoscale [151]. Reflectometry using thermal neutrons (which also have a wavelength of a few tenths of a nanometer) constitutes a highly complementary tool for these structural studies of surfaces and interfaces, thanks particularly to the contrast it provides between hydrogenated and deuterated molecular species [152].

9.7.1 Reflection of X-Rays by a Plane Interface. Critical Angle and Fresnel Law

After the first attempts to deflect or reflect X-rays by Röntgen in 1895, it was long thought that these rays would always propagate in a straight line, no matter what interfaces or materials they were made to cross. These observations led to the conclusion that the refractive index for X-rays always had to be close to $n = 1$. So in order to observe the reflection of X-rays from the plane surface of a solid, very small angles of incidence α would have to be used, close to grazing incidence. Reflection is total for angles of incidence $\alpha < \alpha_c$ (critical angle of total reflection), but then falls off very rapidly as the angle of incidence is increased.

The reflection coefficient $R(\alpha) = I_r(\alpha)/I_0$ is defined as the ratio of the reflected intensity I_r in the specular direction ($\alpha = \alpha_r = \alpha_i$) to the intensity I_0 of the incident beam (see Fig. 9.70).

The existence of an angular interval in which reflection is total arises because, in condensed matter, a refractive index slightly less than unity can be defined for X-rays. More exactly, the so-called Snell–Descartes law of refraction ($\sin i = n \sin r$) provides a relation between the index n of the material medium and the critical angle of total reflection $\alpha_c = \pi/2 - i_c$, which corresponds, with the condition $r = \pi/2$, to

$$n = 1 - \delta = \cos \alpha_c \approx 1 - \alpha_c^2/2. \quad (9.98)$$

It can be shown that the difference of the index from unity depends essentially on the wavelength λ of the X-rays and the electron density ρ_e of the reflecting medium:

$$\delta \approx \alpha_c^2/2 = \lambda^2 \rho_e \frac{r_e}{2\pi}, \quad (9.99)$$

where $r_e = 2.8 \times 10^{-5} \text{ \AA}$, the classical radius of the electron.

For $\lambda = 0.15 \text{ nm}$, we thus obtain:

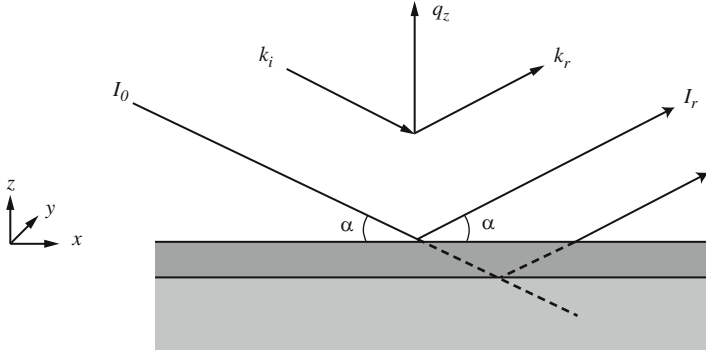


Fig. 9.70. The geometry of specular reflection on a plane surface. Note that the angle of incidence α has been exaggerated for clarity. The incident beam of intensity I_0 is characterised by the wave vector $\mathbf{k}_i$. The reflected beam is characterised by the wave vector $\mathbf{k}_r$. Its intensity I_r results from the combination of partial reflections on the different interfaces. The scattering vector, normal to the interface, is $\mathbf{q} = \mathbf{q}_z = \mathbf{k}_r - \mathbf{k}_i$

- $\alpha_c = 2.9 \text{ mrad}$ and $\delta = 4.5 \times 10^{-6}$ for the free surface of water,
- $\alpha_c = 3.5 \text{ mrad}$ and $\delta = 6 \times 10^{-6}$ for silica, on the surface of silicon,
- $\alpha_c = 5 \text{ mrad}$ and $\delta = 2 \times 10^{-5}$ for the surface of gold.

As shown in Fig. 9.70, the angle of specular reflection can also be characterised by the scattering vector $\mathbf{q}_z$, with magnitude $q_z = 4\pi(\sin \alpha)/\lambda \approx 4\pi\alpha/\lambda$. Using this quantity, one can remove the wavelength dependence of the observed phenomena, defining the critical scattering vector by $\mathbf{q}_c$, with $q_c = 0.038(\rho_e)^{1/2}$.

Note that the values of n , δ , α_c , and q_c can be calculated for a whole set of materials and wavelengths at the website [153].

The laws of classical optics, which describe the phenomenon of total reflection, can also be used to describe the drop in the reflection coefficient beyond the critical angle α_c (or indeed the same for the critical scattering vector $\mathbf{q}_c$). Indeed, when absorption can be neglected, the Fresnel formula

$$\frac{I(q)}{I_0} = \left[\frac{q_z - (q_z^2 - q_c^2)^{1/2}}{q_z + (q_z^2 - q_c^2)^{1/2}} \right]^2 \quad (9.100)$$

predicts, for $q \gg q_c$, an asymptotic behaviour represented by the Fresnel reflection coefficient $R_F(q)$:

$$\frac{I(q)}{I_0} \approx R_F(q) = \frac{q_c^4}{16q^4}. \quad (9.101)$$

For an interface that is flat on average but slight rough, the reflected intensity is partly scattered around the specular direction. The reflection coefficient $R_F(q)$ in the specular direction is then adjusted by a small correction similar to the Debye–Waller factor used in X-ray crystallography, such that

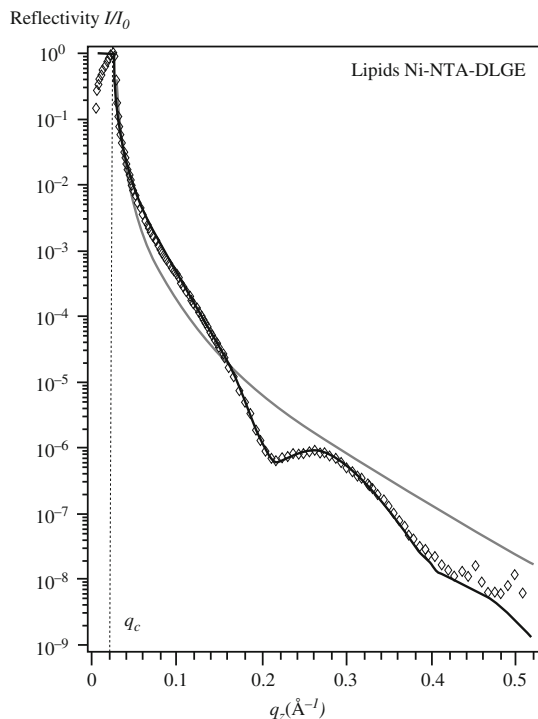


Fig. 9.71. Specular reflectivity curves on a free water surface [156]. The value q_c corresponds to the critical angle of total reflection. In the region of very small angles of incidence ($q_z < q_c$), the horizontal cross-section of the beam is greater than the dimensions of the container and ‘total’ reflection only occurs near $q_z = q_c$. The *grey curve* corresponds to the Fresnel reflectivity at the surface of pure water, affected by a ‘roughness’ effect which is only sensitive beyond $3q_c$. The reflectivity measured after depositing a monolayer of NiNTA-DLGE lipid ligands on the water surface is shown by *diamonds*. The *continuous black curve* was calculated using the density profile in Fig. 9.72 (*grey curve*)

$$R_F(q, \sigma) = R_F(q) \exp(-q^2 \sigma^2 / 2), \quad (9.102)$$

where σ is the root mean square (rms) value of the interface roughness. Given the importance of this effect on the reflected intensity, the flatness of the solid surfaces used must be paid special attention. With regard to the free surfaces of liquids, thermally excited capillary waves lead to a roughness with rms value $\sigma = 0.3 \text{ nm}$ for a water surface at room temperature.

9.7.2 Interference Produced by a Homogeneous Film of Nanometric Thickness

If the interface comprises a homogeneous film of thickness a few nanometers (see Fig. 9.70), the partial reflection of X-rays on the upper and lower surfaces

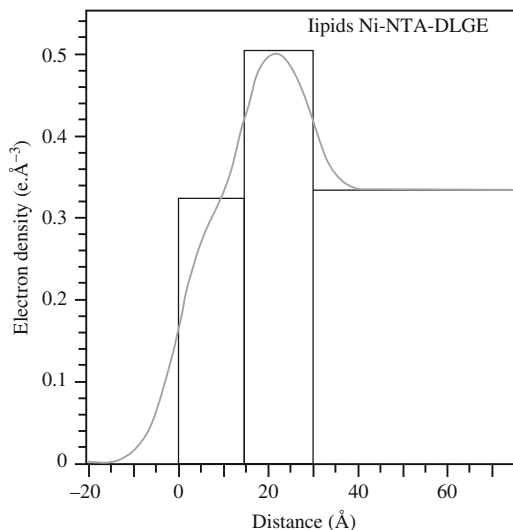


Fig. 9.72. Electron density profiles $\rho(z)$ used to model the reflectivity curve in Fig. 9.71. When there is no roughness (*black lines*), we have, starting from the surface, a first layer of thickness $d_1 = 1.5$ nm and electron density $\rho = 0.32$ electron/ $\text{\AA}^3$, which corresponds to the hydrophobic aliphatic chains of the lipids, than a layer of thickness $d_2 = 1.5$ nm and electron density $\rho = 0.5$ electron/ $\text{\AA}^3$, which corresponds to the polar heads of the lipids, including the NTA group which chelates a nickel ion and serves to immobilise the proteins. Beyond 3 nm, we return to the electron density of the subphase, viz., $\rho = 0.34$ electron/ $\text{\AA}^3$

of the film produces interference leading to a periodic modulation of the reflection coefficient $I(q)/I_0$. The period δq of this modulation is essentially determined by the thickness e of the film:

$$\delta q = 2\pi/e. \quad (9.103)$$

Concerning the amplitude of the modulation, this depends on the reflectivity and hence on the electron density contrast at the surface between the air and the thin film, and also at the surface between the thin film and the substrate. To analyse an experimental reflectivity profile like the one shown in Fig. 9.71, the Fresnel formula can be applied to each surface to calculate a theoretical reflectivity curve for which the parameters, i.e., the average electron densities and thicknesses of the thin film and the substrate, are fitted to describe the experimental curve as closely as possible. The roughnesses of each surface must also be taken into account in this fitting process.

More generally, the reflectivity of a plane interface can be calculated from its electron density profile $\rho(z)$, or more precisely, from the derivative $d\rho(z)/dz$ of this profile, because reflection is produced by variations in the electron density. The Fourier transform of this quantity is used to calculate the reflectivity curve, i.e., the variation of the reflected intensity $I(q)/I_0$ as a function of the

scattering vector $\mathbf{q}$. One then has

$$\frac{I(q)}{I_0} = R_F(q) \left[\int \frac{d\rho(z)}{dz} \exp(iqz) dz \right]^2. \quad (9.104)$$

The example shown in Fig. 9.71 was obtained with a phospholipid monolayer deposited on a water surface. Owing to the difference in electron density between the region of the hydrophilic polar heads and the region of the hydrophobic aliphatic chains, the best fit of the experimental curve is obtained by considering a model with two superposed homogeneous layers on the water surface (see Fig. 9.72). Because of the reduction in surface tension due to the presence of an amphiphilic lipid film, a surface roughness of about 0.5 nm must also be taken into account for each of the three interfaces. Using the above formula, this interface roughness can be described by an average electron density with rounded profile.

9.7.3 Determining the Density Profile of a Stratified Layer. Resolution

When the reflectivity curve of an interface can be measured over a wide range of angles, its electron density profile can be analysed in more detail by using a stratified layer model with different thicknesses, electron densities, and roughnesses for the various layers. In the so-called kinematic description of the reflectivity of each interface, the reflection coefficient between two consecutive layers is calculated as a function of the reflection coefficient of the lower interface. The recurrence relation for the reflection coefficient between layers $n - 1$ and n is

$$R_{n-1,n}(q) = a_{n-1}^2 \frac{R_{n,n+1}(q) + R_F(q) \exp(-q^2\sigma^2/2)}{1 + R_{n,n+1}(q) R_F(q) \exp(-q^2\sigma^2/2)},$$

where $R_F(q)$ is the Fresnel coefficient of the interface between layers n and $n - 1$, σ is the roughness of the interface between layers n and $n - 1$, $a_n = \exp(-iqd_n)$ is the phase shift produced by layer n , and d_n is the thickness of layer n .

The resultant reflectivity of the stratified multilayer is given by the squared modulus of the reflection coefficient $R_{0,1}(q)$ of the first interface. This kind of algorithm is used in several downloadable computation programs [154, 155].

As an illustration, Fig. 9.73 shows a reflectivity curve that is rather rich in detail, recorded for a monolayer of proteins immobilised under a monolayer of lipid ligands on the surface of water [156]. The electron density profile is calculated using a model comprising 30 stratified layers of the same thickness, but with adjustable electron density. The electron density profile giving the best fit with experimental data is shown in Fig. 9.74.

The thickness of the layer corresponds to an orientation that is practically perpendicular to the surface of these long proteins. The electron density profile

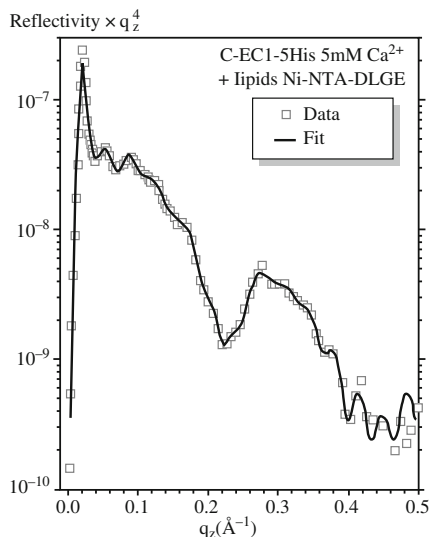


Fig. 9.73. Specular reflectivity curve on a water surface [156]. After depositing a monolayer of NiNTA-DLGE lipid ligands at the surface, proteins endowed with a C-terminal polyhistidine sequence were incubated in the subphase for 10 h. The graph shows the reflectivity $I(q_z)/I_0$ multiplied by q_z^4 , and hence independent of the Fresnel coefficient which tends to mask the modulation. Squares represent the measured values, and the continuous curve the prediction calculated with the density profile shown in Fig. 9.74 (grey curve). The recombinant protein is a cell adhesion protein. It contains the five extracellular domains of the *Xenopus* C-cadherin

can thus be compared with the known structural data for the subunits of the protein (Fig. 9.75).

This choice of example is also intended to elucidate the limits of the method with regard to spatial resolution and sensitivity. Since the measurable range of reflected intensity is limited to $q_{\max} = 5 \text{ nm}^{-1}$, this corresponds in position space to a maximum theoretical resolution given by $d_n = \pi/q_{\max} = 0.6 \text{ nm}$. In reality, fully repeatable results have been obtained by restricting to horizontal slices of thickness $d_n = 0.9 \text{ nm}$. It should also be noted that, in each slice, the electron density is averaged over macroscopic distances parallel to the surface. It is only therefore possible to extract detailed structural information for close-packed layers of molecules with an orientation that is sufficiently well defined relative to the surface. Even when closely packed, the protein layer contains a certain amount of water, of the order of 50%. It thus exhibits an electron density that is only 15–20% greater than that of water, and this implies a rather low level of contrast compared with the subphase. For this reason, the results presented here could only be obtained using a very bright X-ray source, at the European Synchrotron Radiation Facility (ESRF) in Grenoble [157].

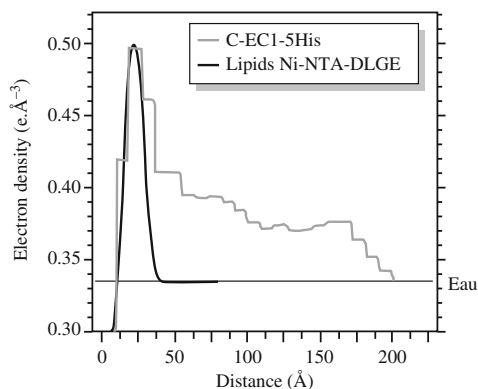


Fig. 9.74. Vertical electron density profile used to model the reflectivity curve in Fig. 9.73. Close to the surface is a layer of thickness 3 nm which corresponds to the lipid ligands, followed by a layer of about 16 nm corresponding to five extracellular domains of C-cadherins which exhibit a certain density modulation

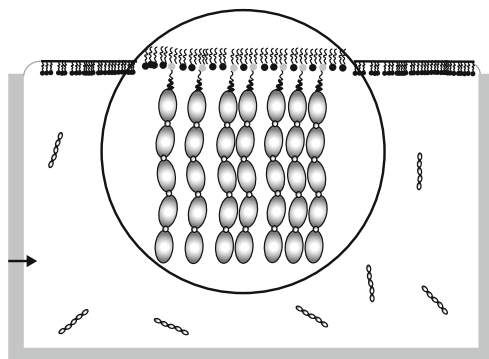


Fig. 9.75. Five extracellular domains of C-cadherins carrying a hexahistidine sequence which immobilises them at the water surface on a monolayer of lipids chelating nickel

9.7.4 Neutron Reflectometry: Contrast Variation

Beams of thermal neutrons with wavelengths in the nanometer range can also be used to analyse the nanoscale structure of plane surfaces and interfaces, using a reflectometry method similar to the one described above for X-rays. The main difference is the interaction of the neutrons with condensed matter. Whereas X-rays interact mainly with the electrons and are thus sensitive to variations in the electron density, neutrons are only scattered by the atomic nuclei. The first consequence is a difference in scattering power between different isotopes of the same atom. For the hydrogen atom, in particular, this difference is very significant, and heavy water or other molecules labelled by deuterium can be used to carry out structural studies of biomolecules by

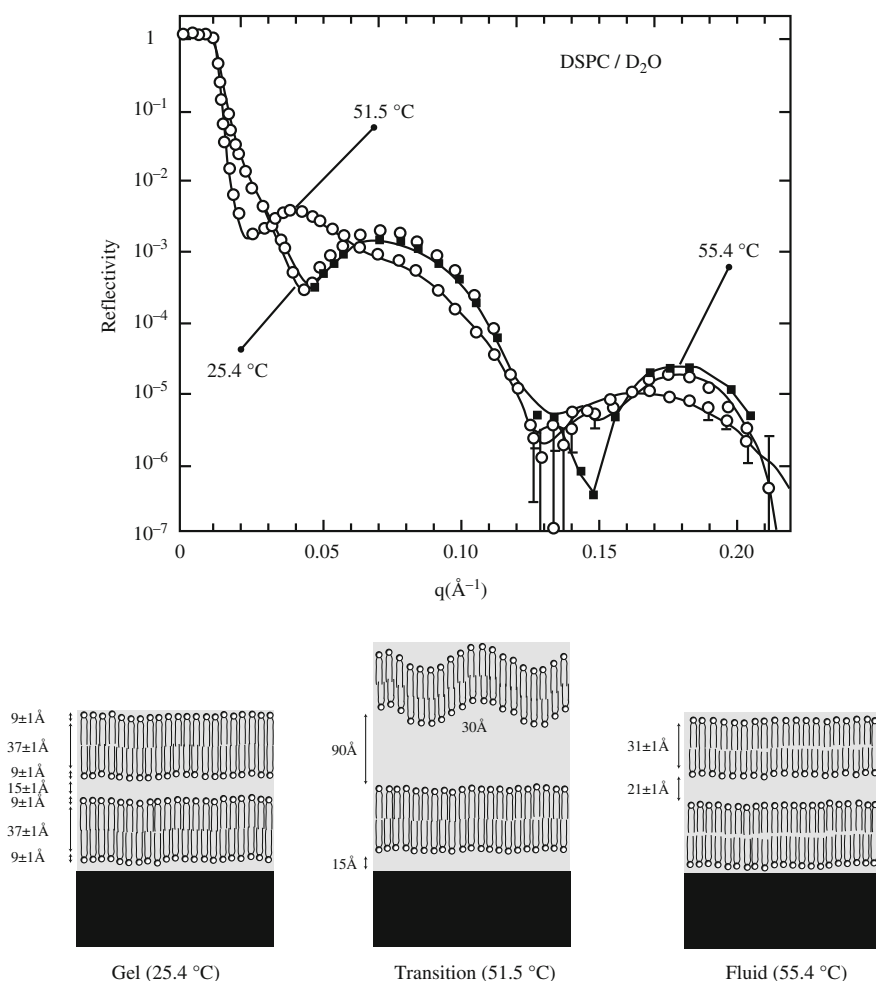


Fig. 9.76. Neutron reflectivity curves on a double bilayer of phospholipids (DSPC) deposited at the interface between a silicon slab and a buffer containing heavy water. At each of three temperatures, a profile of the scattering length density is fitted to the experimental curve, and this is used to determine the thicknesses and roughnesses of the lipid and heavy water layers shown schematically at three different temperatures

neutron scattering [158,159]. Unlike X-rays with wavelengths in the angstrom range, which are quickly absorbed by a few millimeters of condensed matter, thermal neutrons, interacting only with nuclei, can cross several centimeters of some materials. Neutron reflectivity is thus particularly well suited to studying buried solid interfaces or solid/liquid interfaces. However, it should be noted that thermal neutron beams available near a reactor like the one at the Institut Laue-Langevin (Grenoble) have fluxes several orders of magnitude

lower than X-ray fluxes. This means that plane samples with as big an area as possible must be used to increase the signal.

To calculate the reflectivity of a plane interface, the same formulas can be used as for X-rays, replacing the electron density $\rho(z)$ by the scattering length density:

$$\rho_c(z) = \sum_i c_i b_{ci} ,$$

where c_i is the concentration of the isotope i and b_{ci} is the coherent scattering length, which depends on the relevant isotope i .

The example shown in Fig. 9.76 concerns phospholipid bilayers deposited on the plane surface of a silicon slab in contact with a buffer solution [160]. To obtain the best contrast at the lipid/water interfaces, the solution is made with heavy water D₂O. In addition, since the slab of monocrystalline silicon is particularly transparent to neutrons of wavelength greater than 0.63 nm (no diffraction or scattering), the reflectivity is measured on the inner face of the silicon. The reflectivity curve measured up to $q_{\max} = 2.5 \text{ nm}^{-1}$ can be used to determine a density profile with a resolution of the order of 1.2 nm, but by changing the contrast. For example, by repeating the measurement with mixtures of H₂O and D₂O, greater accuracy is obtained for the thickness of the water films on either side of the first bilayer. Under the effect of temperature, the phospholipid bilayers change from a ‘gel’ phase to a ‘fluid’ phase, resulting in a modification of the equilibrium distance and the roughness of the second bilayer.

References

Section One. Biosensors Based on Surface Plasmon Resonance

1. Scheller, F.W., Wollenberger, U., Warsinke, A., Lisdat, F.: Research and development in biosensors, *Curr. Opin. Biotechnol.* **12**, 35–40 (2001)
2. D’Orazio, P.: Biosensors in clinical chemistry, *Clin. Chim. Acta* **334**, 41–69 (2003)
3. Rogers, K.R.: Principles of affinity-based biosensors, *Mol. Biotechnol.* **14**, 109–129 (2000)
4. Rodriguez-Mozaz, S., Marco, M.P., Lopez de Alda, M.J., Barcelo, D.: Biosensors for environmental monitoring of endocrine disruptors: A review article, *Anal. Bioanal. Chem.* **378**, 588–598 (2004)
5. www.biacore.com
6. Cooper, M.A.: Label-free screening of biomolecular interactions, *Anal. Bioanal. Chem.* **377**, 834–842 (2003)
7. Goldstein, B., Coombs, D., He, X., Pineda, A.R., Wofsy, C.: The influence of transport on the kinetics of binding to surface receptors: Application to cells and BIAcore, *J. Mol. Recognit.* **12**, 293–299 (1999)
8. Karlsson, R., Falt, A.: Experimental design for kinetic analysis of protein–protein interactions with surface plasmon resonance biosensors, *J. Immunol. Methods* **200**, 121–133 (1997)

9. Karlsson, R., Roos, H., Fägerstam, L., Persson, B.: Kinetic and concentration analysis using BIA technology, *Methods: A Companion to Methods in Enzymology* **6**, 99–110 (1994)
10. Myszk, D.G., He, X., Dembo, M., Morton, T.A., Goldstein, B.: Extending the range of rate constants available from BIACORE: Interpreting mass transport-influenced binding data, *Biophys. J.* **75**, 583–594 (1998)
11. *BIAevaluation Software Handbook*: Biacore AB, Uppsala, Sweden (2004)
12. *BIAsimulation Software Handbook*: Biacore AB, Uppsala, Sweden (1996)
13. Zanier, K., Charbonnier, S., Baltzinger, M., Nomine, Y., Altschuh, D., Trave, G.: Kinetic analysis of the interactions of human papillomavirus E6 oncoproteins with the ubiquitin ligase E6AP using surface plasmon resonance, *J. Mol. Biol.* **349**, 401–412 (2005)
14. Myszk, D.G.: Improving biosensor analysis, *J. Mol. Recognit.* **12**, 279–284 (1999)
15. Karlsson, R., Stahlberg, R.: Surface plasmon resonance detection and multispot sensing for direct monitoring of interactions involving low-molecular-weight analytes and for determination of low affinities, *Anal. Biochem.* **228**, 274–280 (1995)
16. Frostell-Karlsson, A., Remaeus, A., Roos, H., Andersson, K., Borg, P., Hamalainen, M., Karlsson, R.: Biosensor analysis of the interaction between immobilized human serum albumin and drug compounds for prediction of human serum albumin binding levels, *J. Med. Chem.* **43**, 1986–1992 (2000)
17. Nieba, L., Krebber, A., Pluckthun, A.: Competition BIAcore for measuring true affinities: Large differences from values determined from binding kinetics, *Anal. Biochem.* **234**, 155–165 (1996)
18. Andersson, K., Hamalainen, M., Malmqvist, M.: Identification and optimization of regeneration conditions for affinity-based biosensor assays. A multivariate cocktail approach, *Anal. Chem.* **71**, 2475–2481 (1999)
19. Zeder-Lutz, G., Benito, A., Van Regenmortel, M.H.: Active concentration measurements of recombinant biomolecules using biosensor technology, *J. Mol. Recognit.* **12**, 300–309 (1999)

Applications

20. Myszk, D.G.: Survey of the 1998 optical biosensor literature, *J. Mol. Recognit.* **12**, 390–408 (1999)
21. Rich, R.L., Myszk, D.G.: Survey of the 1999 surface plasmon resonance biosensor literature, *J. Mol. Recognit.* **13**, 388–407 (2000)
22. Rich, R.L., Myszk, D.G.: Survey of the year 2000 commercial optical biosensor literature, *J. Mol. Recognit.* **14**, 273–294 (2001)
23. Rich, R.L., Myszk, D.G.: Survey of the year 2001 commercial optical biosensor literature, *J. Mol. Recognit.* **15**, 352–376 (2002)
24. Rich, R.L., Myszk, D.G.: Survey of the year 2002 commercial optical biosensor literature, *J. Mol. Recognit.* **16**, 351–382 (2003)
25. Rich, R.L., Myszk, D.G.: Survey of the year 2003 commercial optical biosensor literature, *J. Mol. Recognit.* **18**, 1–39 (2005)
26. Van Regenmortel, M.H.: Analysing structure–function relationships with biosensors, *Cell. Mol. Life Sci.* **58**, 794–800 (2001)
27. Homola, J.: Present and future of surface plasmon resonance biosensors, *Anal. Bioanal. Chem.* **377**, 528–539 (2003)

28. Karlsson, R.: SPR for molecular interaction analysis: A review of emerging application areas, *J. Mol. Recognit.* **17**, 151–161 (2004)
29. Timmins, J., Schoehn, G., Ricard-Blum, S., Scianimanico, S., Vernet, T., Ruigrok, R.W., Weissenhorn, W.: Ebola virus matrix protein VP40 interaction with human cellular factors Tsg101 and Nedd4, *J. Mol. Biol.* **326**, 493–502 (2003)
30. Desmaris, F., Lemaire, D., Ricard-Blum, S., Chatrenet, B., Forest, E.: Structural characterization of the feline-immunodeficiency-virus envelope glycoprotein 36 ectodomain for the development of new antivirals, *Biochem. J.* **389**, 559–567 (2005)
31. Bokken, G.C., Corbee, R.J., van Knapen, F., Bergwerff, A.A.: Immunochemical detection of *Salmonella* group B, D and E using an optical surface plasmon resonance biosensor, *FEMS Microbiol. Lett.* **222**, 75–82 (2003)
32. Indyk, H.E., Filonzi, E.L.: Direct optical biosensor analysis of folate-binding protein in milk, *J. Agric. Food Chem.* **52**, 3253–3258 (2004)
33. Nanao, M., Ricard-Blum, S., Di Guilmi, A.M., Lemaire, D., Lascoux, D., Chabert, J., Attree, I., Dessen, A.: Type III secretion proteins PcrV and PcrG from *Pseudomonas aeruginosa* form a 1:1 complex through high affinity interactions, *BMC Microbiol.* **3**, 21 (2003)
34. Mullett, W.M., Lai, E.P., Yeung, J.M.: Surface plasmon resonance-based immunoassays, *Methods* **22**, 77–91 (2000)
35. Lofas, S.: Optimizing the hit-to-lead process using SPR analysis, *Assay Drug. Dev. Technol.* **2**, 407–415 (2004)
36. Natsume, T., Nakayama, H., Isobe, T.: BIA-MS-MS: Biomolecular interaction analysis for functional proteomics, *Trends Biotechnol.* **19** (10 Suppl.), S28–33 (2001)
37. Yuk, J.S., Ha, K.S.: Proteomic applications of surface plasmon resonance biosensors: Analysis of protein arrays, *Exp. Mol. Med.* **37**, 1–10 (2005)
38. Ricard-Blum, S., Bernocco, S., Font, B., Moali, C., Eichenberger, D., Farjanel, J., Burchardt, E.R., van der Rest, M., Kessler, E., Hulmes, D.J.S.: Interaction properties of the procollagen C-proteinase enhancer protein shed light on the mechanism of stimulation of BMP-1, *J. Biol. Chem.* **277**, 33864–33869 (2002)
39. Blanquet-Grossard, F., Thielens, N.M., Vendrely, C., Jamin, M., Arlaud, G.J.: Complement protein C1q recognizes a conformationally modified form of the prion protein, *Biochemistry* **44**, 4349–4356 (2005)
40. Myszka, D.G., Abdiche, Y.N., Arisaka, F., Byron, O., Eisenstein, E., Hensley, P., Thomson, J.A., Lombardo, C.R., Schwarz, F., Stafford, W., Doyle, M.L.: The ABRF-MIRG'02 study: Assembly state, thermodynamic, and kinetic analysis of an enzyme/inhibitor interaction, *J. Biomol. Tech.* **14**, 247–269 (2003)
41. Cseh, S., Vera, L., Matsushita, M., Fujita, T., Arlaud, G.J., Thielens, N.M.: Characterization of the interaction between L-ficolin/p35 and mannan-binding lectin-associated serine proteases-1 and -2, *J. Immunol.* **169**, 5735–5743 (2002)
42. Gregory, L.A., Thielens, N.M., Matsushita, M., Sorensen, R., Arlaud, G.J., Fontecilla-Camps, J.C., Gaboriaud, C.: The X-ray structure of human mannan-binding lectin-associated protein 19 (MAp19) and its interaction site with mannan-binding lectin and L-ficolin, *J. Biol. Chem.* **279**, 29391–29397 (2004)
43. Teillet, F., Dublet, B., Andrieu, J.P., Gaboriaud, C., Arlaud, G.J., Thielens, N.M.: The two major oligomeric forms of human mannan-binding lectin:

- Chemical characterization, carbohydrate-binding properties, and interaction with MBL-associated serine proteases, *J. Immunol.* **174**, 2870–2877 (2005)
44. Myszka, D.G., Wood, S.J., Biere, A.L.: Analysis of fibril elongation using surface plasmon resonance biosensors, *Methods Enzymol.* **309**, 386–402 (1999)
 45. Hasegawa, K., Ono, K., Yamada, M., Naiki, H.: Kinetic modeling and determination of reaction constants of Alzheimer's beta-amyloid fibril extension and dissociation using surface plasmon resonance, *Biochemistry* **41**, 13489–13498 (2002)
 46. Cannon, M.J., Williams, A.D., Wetzel, R., Myszka, D.G.: Kinetic analysis of beta-amyloid fibril elongation, *Anal. Biochem.* **328**, 67–75 (2004)
 47. Cairo, C.W., Strzelec, A., Murphy, R.M., Kiessling, L.L.: Affinity-based inhibition of beta-amyloid toxicity, *Biochemistry* **41**, 8620–8629 (2002)
 48. Cairo, C.W., Kiessling, L.L.: Immobilized amyloid-b experiments using SPR technology, *Biacore Journal* **1**, 4–7 (2003)
 49. Aguilar, M.I., Small, D.H.: Surface plasmon resonance for the analysis of beta-amyloid interactions and fibril formation in Alzheimer's disease research, *Neurotox. Res.* **7**, 17–27 (2005)
 50. Sota, H., Hasegawa, Y., Iwakura, M.: Detection of conformational changes in an immobilized protein using surface plasmon resonance, *Anal. Chem.* **70**, 2019–2024 (1998)
 51. Masuda, T., Yasumoto, K., Kitabatake, N.: Monitoring the irradiation-induced conformational changes of ovalbumin by using monoclonal antibodies and surface plasmon resonance, *Biosci. Biotechnol. Biochem.* **64**, 710–716 (2000)
 52. Yamaguchi, S., Mannen, T., Zako, T., Kamiya, N., Nagamune, T.: Measuring adsorption of a hydrophobic probe with a surface plasmon resonance sensor to monitor conformational changes in immobilized proteins, *Biotechnol. Prog.* **19**, 1348–1354 (2003)
 53. Zako, T., Harada, K., Mannen, T., Yamaguchi, S., Kitayama, A., Ueda, H., Nagamune, T.: Monitoring of the refolding process for immobilized firefly luciferase with a biosensor based on surface plasmon resonance, *J. Biochem. (Tokyo)* **129**, 1–4 (2001)
 54. Mannen, T., Yamaguchi, S., Honda, J., Sugimoto, S., Kitayama, A., Nagamune, T.: Observation of charge state and conformational change in immobilized protein using surface plasmon resonance sensor, *Anal. Biochem.* **293**, 185–193 (2001)
 55. Paynter, S., Russell, D.A.: Surface plasmon resonance measurement of pH-induced responses of immobilized biomolecules: Conformational change or electrostatic interaction effects?, *Anal. Biochem.* **309**, 85–95 (2002)
 56. May, L.M., Russell, D.A.: The characterization of biomolecular secondary structures by surface plasmon resonance, *Analyst* **127**, 1589–1595 (2002)
 57. Stokka, A.J., Flatmark, T.: Substrate-induced conformational transition in human phenylalanine hydroxylase as studied by surface plasmon resonance analyses: The effect of terminal deletions, substrate analogues and phosphorylation, *Biochem. J.* **369**, 509–518 (2003)
 58. Gambari, R.: Biospecific interaction analysis (BIA) as a tool for the design and development of gene transcription modifiers, *Curr. Med. Chem. Anti-Canc. Agents* **1**, 277–291 (2001)
 59. Katsamba, P.S., Park, S., Laird-Offringa, I.A.: Kinetic studies of RNA–protein interactions using surface plasmon resonance, *Methods* **26**, 95–104 (2002)

60. Wang, R., Minunni, M., Tombelli, S., Mascini, M.: A new approach for the detection of DNA sequences in amplified nucleic acids by a surface plasmon resonance biosensor, *Biosens. Bioelectron.* **20**, 598–605 (2004)
61. Duverger, E., Frison, N., Roche, A.C., Monsigny, M.: Carbohydrate–lectin interactions assessed by surface plasmon resonance, *Biochimie* **85**, 167–179 (2003)
62. Sadir, R., Baleux, F., Grosdidier, A., Imberty, A., Lortat-Jacob, H.: Characterization of the stromal cell-derived factor-1 α -heparin complex, *J. Biol. Chem.* **276**, 8288–8296 (2001)
63. Ricard-Blum, S., Feraud, O., Lortat-Jacob, H., Rencurosi, A., Fukai, N., Dkhisli, F., Vittet, D., Imberty, A., Olsen, B.R., van der Rest, M.: Characterization of endostatin binding to heparin and heparan sulfate by surface plasmon resonance and molecular modeling: Role of divalent cations, *J. Biol. Chem.* **279**, 2927–2936 (2004)
64. Shen, B., Shimson, S., Smith, M.M., Ghosh, P.: Biosensor analysis of the molecular interactions of pentosan polysulfate and of sulfated glycosaminoglycans with immobilized elastase, hyaluronidase and lysozyme using surface plasmon resonance (SPR) technology, *J. Pharm. Biomed. Anal.* **31**, 83–93 (2003)
65. MacKenzie, C.R., Hirama, T.: Quantitative analyses of binding affinity and specificity for glycolipid receptors by surface plasmon resonance, *Methods Enzymol.* **312**, 205–216 (2000)
66. Sola, M., Bavro, V.N., Timmins, J., Franz, T., Ricard-Blum, S., Schoehn, G., Ruigrok, R.W., Paarmann, I., Saiyed, T., O'Sullivan, G.A., Schmitt, B., Betz, H., Weissenhorn, W.: Structural basis of dynamic glycine receptor clustering by gephyrin, *EMBO J.* **23**, 2510–2519 (2004)
67. Pattnaik, P.: Surface plasmon resonance: Applications in understanding receptor–ligand interaction, *Appl. Biochem. Biotechnol.* **126**, 79–92 (2005)
68. Abdiche, Y.N., Myszka, D.G.: Probing the mechanism of drug/lipid membrane interactions using Biacore, *Anal. Biochem.* **328**, 233–243 (2004)
69. Frostell-Karlsson, A., Widegren, H., Green, C.E., Hamalainen, M.D., Westerlund, L., Karlsson, R., Fenner, K., van de Waterbeemd, H.: Biosensor analysis of the interaction between drug compounds and liposomes of different properties: A two-dimensional characterization tool for estimation of membrane absorption, *J. Pharm. Sci.* **94**, 25–37 (2005)
70. Nordin, H., Jungnelius, M., Karlsson, R., Karlsson, O.P.: Kinetic studies of small molecule interactions with protein kinases using biosensor technology, *Anal. Biochem.* **340**, 359–368 (2005)
71. Kim, K., Cho, S., Park, J.H., Byun, Y., Chung, H., Kwon, I.C., Jeong, S.Y.: Surface plasmon resonance studies of the direct interaction between a drug/intestinal brush border membrane, *Pharm. Res.* **2**, 1233–1239 (2004)
72. Canziani, G.A., Klakamp, S., Myszka, D.G.: Kinetic screening of antibodies from crude hybridoma samples using Biacore, *Anal. Biochem.* **325**, 301–307 (2004)
73. Morel, G., Ricard-Blum, S., Ardail, D.: Kinetics of internalization and subcellular binding sites for T3 in mouse liver, *Biol. Cell* **86**, 167–174 (1996)
74. Gestwicki, J.E., Cairo, C.W., Mann, D.A., Owen, R.M., Kiessling, L.L.: Selective immobilization of multivalent ligands for surface plasmon resonance and fluorescence microscopy, *Anal. Biochem.* **305**, 149–155 (2002)

75. Lang, K., Hatt, H., Niggemann, B., Zaenker, K.S., Entschladen, F.: A novel function for chemokines: Downregulation of neutrophil migration, *Scand. J. Immunol.* **57**, 350–361 (2003)
76. Hoffmann, J., Pietrzik, C.U., Kummer, M.P., Twiesselmann, C., Bauer, C., Herzog, V.: Binding and selective detection of the secretory N-terminal domain of the alzheimer amyloid precursor protein on cell surfaces, *J. Histochem. Cytochem.* **47**, 373–382 (1999)
77. Quinn, J.G., O’Kennedy, R., Smyth, M., Moulds, J., Frame, T.: Detection of blood group antigens utilising immobilised antibodies and surface plasmon resonance, *J. Immunol. Methods* **206**, 87–96 (1997)
78. Abad, L.W., Neumann, M., Tobias, L., Obenauer-Kutner, L., Jacobs, S., Cullen, C.: Development of a biosensor-based method for detection and iso-typing of antibody responses to adenoviral-based gene therapy vectors, *Anal. Biochem.* **310**, 107–113 (2002)
79. Sz wajkajzer, D., Carey, J.: RaPID plots: Affinity and mechanism at a glance, *Biacore Journal* **1**, 19 (2003)
80. Saerens, D., Kinne, J., Bosmans, E., Wernery, U., Muyldermans, S., Conrath, K.: Single domain antibodies derived from dromedary lymph node and peripheral blood lymphocytes sensing conformational variants of prostate-specific antigen, *J. Biol. Chem.* **279**, 51965–51972 (2004)
81. Genst, E., Handelberg, F., Van Meirhaeghe, A., Vynck, S., Loris, R., Wyns, L., Muyldermans, S.: Chemical basis for the affinity maturation of a camel single domain antibody, *J. Biol. Chem.* **279**, 53593–53601 (2004)
82. Haughey, S.A., O’Kane, A.A., Baxter, G.A., Kalman, A., Trisconi, M.J., Indyk, H.E., Watene, G.A.: Determination of pantothenic acid in foods by optical biosensor immunoassay, *J. AOAC Int.* **88**, 1008–1014 (2005)
83. Kalman, A., Girard, P., Wahlström, L.: A biosensor-based method for the determination of vitamin B12 concentration in food products, *Biacore Journal* **4**, 16–17 (2004)
84. Haasnoot, W., Smits, N.G., Kemmers-Voncken, A.E., Bremer, M.G.: Fast biosensor immunoassays for the detection of cows’ milk in the milk of ewes and goats, *J. Dairy Res.* **71**, 322–329 (2004)
85. Leonard, P., Hearty, S., Quinn, J., O’Kennedy, R.: A generic approach for the detection of whole *Listeria monocytogenes* cells in contaminated samples using surface plasmon resonance, *Biosens. Bioelectron.* **19**, 1331–1335 (2004)
86. Gillis, E.H., Gosling, J.P., Sreenan, J.M., Kane, M.: Development and validation of a biosensor-based immunoassay for progesterone in bovine milk, *J. Immunol. Methods* **267**, 131–138 (2002)
87. Johansson, A.: New Biacore tests for the control of illegal hormones in meat production, *Biacore Journal* **4**, 18 (2004)
88. Gustavsson, E.: Biacore methods for detection of β -lactam antibiotics in milk, *Biacore Journal* **2**, 22–23 (2003)
89. Ferguson, J., Baxter, A.: The measurement of chloramphenicol in food products using Qflex kit chloramphenicol, *Biacore Journal* **2**, 16–17 (2003)
90. Abraham, R., Quimby, M.: Validation of a Biacore C concentration assay, *Biacore Journal* **3**, 12–15 (2003)
91. Yang, C.Y., Brooks, E., Li Y., Denny, P., Ho, C.M., Qi, F., Shi, W., Wolinsky, L., Wu, B., Wong, D.T., Montemagno, C.D.: Detection of picomolar levels of interleukin-8 in human saliva by SPR, *Lab. Chip* **5**, 1017–1023 (2005)

Section Two. Ellipsometry

92. Azzam, R.M.A., Bashara, N.M.: *Ellipsometry and Polarized Light*, Elsevier, Amsterdam (1987)
93. Tompkins, H.G., McGahan, W.A.: *Spectroscopic Ellipsometry and Reflectometry. A User's Guide*, Wiley, New York (1999)
94. Stchakovsky, M.: PhD thesis, University of Paris-Sud (1991)
95. Tompkins, H.G., Irene, E.A.: *Handbook of Ellipsometry*, William Andrew, Norwich, Springer, Berlin Heidelberg New York (2005)
96. Palik, E.D.: *Handbook of Optical Constants of Solids*, Academic Press, Orlando (1985)
97. Agius, B.: *Surfaces, Interfaces et Films Minces: Observation et Analyse*, Dunod, Paris (1993)
98. Fleury, P., Mathieu, J.P.: *Lumière*, Eyrolles, Paris (1961)

Section Three. Optical Spectroscopy Using Waveguides

99. Tiefenthaler, K., Lukozs, W.: J. Opt. Soc. Amer. B **6**, 209 (1989)
100. Ziegler, C., Göpel, W.: Curr. Opin. Chem. Biol. **2**, 585 (1998)
101. Ramsden, J.J.: J. Molec. Recogn. **10**, 109 (1997)
102. Schuck, P.: Biophys. J. **70**, 1230 (1996)
103. De Feijter, J.A., Benjamins, J., Veer, F.A.: Biopolymers **17**, 1759 (1978)
104. Ball, V., Ramsden, J.J.: Biopolymers **46**, 489 (1998)
105. Kekicheff, P., Laughlin, R.G., Munyon, R.L.: Langmuir **17**, 4693 (2001)
106. Mann, E.K., Heinrich, L., Voegel, J.-C., Schaaf, P.: J. Chem. Phys. **105**, 6082 (1996)
107. Born, M., Wolf, E.: *Principles of Optics*, 7th edn., Cambridge University Press (1999)
108. Goutev, N., Nickolov, Z.S., Ramsden, J.J.: J. Raman Spectrosc. **27**, 897 (1996)
109. Picart, C., Ladam, G., Senger, B., Voegel, J.-C., Schaaf, P., Cuisinier, F.J.G., Gergely, C.: J. Chem. Phys. **115**, 1086 (2001)
110. Picart, C., Gergely, C., Arntz, Y., Voegel, J.-C., Schaaf, P., Cuisinier, F.J.G., Senger, B.: Biosensors & Bioelectron. **20**, 553 (2004)
111. Höök, F., Vörös, J., Rodahl, M., Kurrat, R., Böni, P., Ramsden, J.J., Textor, M., Spencer, N.D., Tengvall, P., Gold, J., Kasemo, B.: Colloids and Surfaces B **24**, 155 (2002)
112. Decher, G.: Science **277**, 1232 (1997)
113. Picart, C., Lavalle, Ph., Hubert, P., Cuisinier, F.J.G.: Decher, G., Schaaf, P., Voegel, J.-C.: Langmuir **17**, 7414 (2001)
114. Boulmedais, F., Ball, V., Schwinté, P., Frisch, B., Schaaf, P., Voegel, J.-C.: Langmuir **19**, 440 (2003)
115. Picart, C., Mutterer, J., Richert, L., Luo, Y., Prestwich, G.D., Schaaf, P., Voegel, J.-C., Lavalle, Ph.: Proc. Natl. Acad. Sci. USA **99**, 12531 (2002)

Section Four. Vibrational Spectroscopy

116. Nakamoto, K.: *Infrared and Raman Spectra of Inorganic and Coordination Compounds. Part A: Theory and Applications in Inorganic Chemistry*, Vol. 1, John Wiley & Sons (1997)
117. Hipps, K.W., Crosby, G.A.: J. Phys. Chem. **83**, 555 (1979)

118. Abelès, F. (Ed.): *Optical Properties of Solids*, North-Holland Publishing Company (1972)
119. Buffeteau, T., Desbat, B., Blaudez, D., Turlet, J.M.: **54**, 11, 1641 (2000)
120. Blaudez, D., Buffeteau, T., Cornut, J.C., Desbat, B., Escafre, N., Pezolet, M., Turlet, J.M.: *Appl. Spectrosc.* **47**, 869 (1993)
121. Raman, C.V., Krishnan, K.S.: *Nature* **121** (3048), 501 (1928)
122. Delhay, M., Dhamelincourt, P.: *J. Raman Spectroscopy* **3**, 3342 (1975)
123. Fleischmann, M., Hendra, P.J., McQuillan, A.: *Chem. Phys. Lett.* **26**, 163–166 (1974)
124. Stöckle, R.M., Suh, Y.D., Deckert, V., Zenobi, R.: *Chem. Phys. Lett.* **318**, 131–136 (2000)
125. Taubner, T., Hillenbrand, R., Keilmann, F.: *Applied Physics Letters* **85**, 5064–5066 (2004)

General References

126. Chalmers, J.M., Griffiths, P. (Eds.): *Handbook of Vibrational Spectroscopy*, Vols. 1–5, John Wiley & Sons (2002)
127. Anderson, A.: *The Raman effect*, Vols. 1–2, John Wiley & Sons (1973)

Section Five. Brewster Angle Microscopy

128. Henon, S., Meunier, J.: *Rev. Sci. Instr.* **62**, 936 (1991)
129. Hönig, D., Möbius, D.: *J. Phys. Chem.* **95**, 4590 (1991)

Section Six. Quartz Crystal Microbalance

130. Nomura, T., Okuhara, M.: *Anal. Chim. Acta* **142**, 281–284 (1982)
131. Rodahl, M., Hook F., Fredriksson, C., Keller, C.A., Krozer, A., Brzezinski, P., Voinova, M.V., Kasemo, B.: *Faraday Discuss.* **107**, 229–246 (1997)
132. Keller, C.A., Kasemo, B.: *Biophys. J.* **75**, 1397–1402 (1998)
133. Höök, F., Rodahl, M., Brzezinski, P., Kasemo, B.: *Langmuir* **14**, 729–734 (1998)
134. Höök, F., Kasemo, B., Nylander, T., Fant, C., Sott, K., Elwing, H.: *Anal. Chem.* **73**, 5796–5804 (2001)
135. Höök, F., Ray, A., Krave, U., Norden, B., Kasemo, B.: *Langmuir* **17**, 8305–8312 (2001)
136. Wegener, J., Seebach, J., Janshoff, A., Galla, H.J.: *Biophys. J.* **78**, 2821–2833 (2000)
137. Snabe, T., Petersen, S.B.: *Chem. Phys. Lipids* **125**, 69–82 (2003)
138. Tanahashi, M., Kokubo, T., Matsuda, T.: *J. Biomed. Mater. Res.* **31**, 243–249 (1996)
139. Rodahl, M., Höök, F., Krozer, A., Brzezinski, P., Kasemo, B.: *Rev. Sci. Instrum.* **66**, 3924–3930 (1995)
140. Mecea, V.M.: *Sensors and Actuators A* **40**, 1–27 (1993)
141. Voinova, M.V., Rodahl, M., Jonson, M., Kasemo, B.: *Physica Scripta* **59**, 391–396 (1999)
142. Sauerbrey, G.: *Z. Phys.* **155**, 206–222 (1959)
143. Rodahl, M., Kasemo, B.: *Sensors and Actuators A* **54**, 448–456 (1996)

144. Höök, F., Vörös, J., Rodahl, M., Kurrat, R., Böni, P., Ramsden, J.J., Textor, M., Spencer, N.D., Tengvall, P., Gold, J., Kasemo, B.: *Colloid Surf. B* **24**, 155–170 (2002)
145. Decher, G., Hong, J.D., Schmitt, J.: *Thin Solid Films* **210**, 831–835 (1992)
146. Decher, G.: *Science* **277**, 1232–1237 (1997)
147. Bertrand, P., Jonas, A., Laschewsky, A., Legras, R.: *Macromol. Rapid. Comm.* **21**, 319–348 (2000)
148. Ladam, G., Schaad, P., Voegel, J.-C., Schaaf, P., Decher, G., Cuisinier, F.J.G.: *Langmuir* **16**, 1249–1255 (2000)
149. Picart, C., Lavalle, Ph., Hubert, P., Cuisinier, F.J.G., Decher, G., Schaaf, P., Voegel, J.-C.: *Langmuir* **17**, 7414–7424 (2001)
150. Picart, C., Mutterer, J., Richert, L., Luo, Y., Prestwich, G.D., Schaaf, P., Voegel, J.-C., Lavalle, Ph.: *Proc. Natl. Acad. Sci. USA* **99**, 12531–12535 (2002)

Section Seven. Grazing Incidence Neutron and X-Ray Reflectometry

151. Benattar, J.J.: *La réflectivité des rayons X*, La Recherche **244**, 722–731 (1992)
152. Daillant, J., Gibaud, A.: *X-Ray and Neutron Reflectivity: Principles and Applications*, Lecture Notes in Physics, Springer, Berlin Heidelberg New York (1999)
153. www-cxro.lbl.gov/~optical_constants/
154. Parratt, L.: *The Reflectivity Tool*, Hahn-Meitner-Institut, Berlin
www.hmi.de/bensc/software/refl/parratt/parratt.html
155. Bardon, S., Ober, R., Valignat, M.P., Vandenbrouck, F., Cazabat, A.M., Daillant, J.: *Phys. Rev. E* **59**, 6808–6818 (1999)
156. Martel L., Johnson C., Boutet S., Al-Kurdi R., Konovalov, O., Robinson I., Leckband D., Legrand, J.F.: *J. Phys. IV France C* **12**, 365–371 (2002)
157. Konovalov, O., Myagkov, I., Struth, B., Lohner, K.: *Eur. Biophys. J.* **31**, 1758–1768 (2001)
158. Johnson, C.P., Fragneto, G., Konovalov, O., Dubosclard, V., Legrand, J.F., Leckband, D.E.: *Biochemistry* **44**, 546–554 (2005)
159. Lösche, M.: *Current topics in Membranes* **52** (2002)
160. Fragneto, G., Charitat, T., Graner, F., Mecke, K., Perino-Gallice, L., Bellet-Amalric, E.: *Europhys. Lett.* **53**, 100–106 (2001)

Mass Spectrometry

D. Pflieger, E. Forest, and J. Vinh

For twenty years or so now, mass spectrometry has been used to get exact measurements of the mass of biological molecules such as proteins, nucleic acids, oligosaccharides, and so on. Over the past ten years, this technology has followed the trend toward miniaturisation and the samples required can be much smaller. In particular, the nanoelectrospray source (online or by needle) allow one to work at flow rates of a few tens of nanolitres/min. There are many applications, both in the field of proteomics and in the analysis of protein structure, dynamics, and interactions. Combining this source with nanoHPLC, complex mixtures only available in small quantities can be separated and analysed online. There are also some advantages over conventional HPLC, despite a set of constraints related to the small dimensions and low flow rates. Combining capillary electrophoresis with the electrospray source also gives useful results, with its own set of advantages and constraints. Finally, developments are currently underway to combine this source with chips, providing a means of separation and analysis online.

10.1 Principles and Definitions

Mass spectrometry has become an almost indispensable tool for characterising and studying biomolecules, especially for structural studies of peptides and proteins. A protein can be identified by measuring the molecular masses of peptide mixtures obtained by endoprotease-type enzyme digestion. Differences in molecular mass can be revealed due to the presence of post-translational modifications or disulfide bridges when the primary amino acid sequence is known, by comparing the expected theoretical mass with the measured experimental mass for each peptide. Analysis by mass spectrometry can be combined with enzyme digestion to determine the amino acid sequences with aminopeptidase or carboxypeptidase enzymes. Studies in the field of cell biology have also been developed over the past few years. Mass spectrometric analysis of

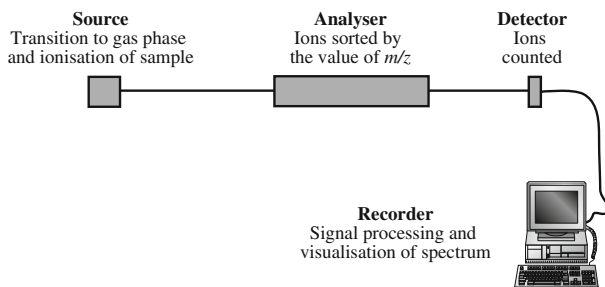


Fig. 10.1. General setup for a mass spectrometer

peptide mixtures can be carried out directly on organs or single cells, e.g., analysis of giant neurons [1] or cells of the lobster secretory gland [2].

10.1.1 What Is Mass Spectrometry?

Mass spectrometry can transform molecules in their natural state into ions in a gaseous state, and thereby obtain their molecular mass m by analysing the mass/charge ratio, denoted by m/z , where m is the mass of the compound and z its charge.

10.1.2 The Mass Spectrometer

A mass spectrometer is a device for measuring the mass/charge ratio m/z of the ions formed from the sample under investigation. It always includes the following elements (see Fig. 10.1):

- An ion source in which the sample is transformed into a gas phase (vaporisation/sublimation/desorption), the molecules are ionised, and the ions are decomposed.
- An analyser able to sort the ions according to the value of the m/z ratio.
- A detector which counts the ions for each value of m/z .
- A recording device for processing the signal and visualising the spectra.
- A calibration system for correlating the actual measured quantity with the m/z ratio.

10.1.3 Terminology

Dalton. Unit of molecular mass, defined as one twelfth of the mass of a ^{12}C carbon atom (1.66×10^{-24} g), approximately equal to the mass of one ^1H atom.

Mass Range. This is the range of values of the mass/charge ratio m/z , specified by a minimum and a maximum, that can be detected with the given mass spectrometer. The mass range is a function of the technological characteristics of the analyser.

Molecular Ion. The molecular ion refers to the intact ionised molecule. However, this term is also used to refer to ionic forms with an odd number of electrons, such as the cationic radical M^+ and the anionic radical M^- . Ions formed in ESI and MALDI are often species with an even number of electrons, such as $[M+H]^+$ or $[M-H]^-$, called protonated or deprotonated ions, respectively, or sometimes pseudo-molecular ions. (This is also the case for FAB, LSIMS, and other ionisation modes.)

Average Mass. The average mass of a molecule is calculated using the average atomic masses of the individual elements making it up ($C = 12.011$, $H = 1.008$, $N = 14.007$, $O = 15.999$, etc.). It is the centroid of the isotopic distribution.

Monoisotopic Mass. The monoisotopic mass of a molecule is calculated using the mass of the most abundant isotope for each individual element making it up ($C = 12.000,00$, $H = 1.007,825$, $O = 15.994,9$, etc.).

Mass Accuracy. A distinction must be made between errors in the actual determination of m/z and ambiguities due to the ionisation process.

Error in the Determination of m/z . Systematic error in the sense of theoretical error, calibration error, or operator error. The accuracy can be assessed from the standard deviation after a series of measurements. Giving a standard deviation does not mean that the true value lies in the resulting interval, but it does indicate the degree of chance in the measurement. Systematic errors must be measured with a compound of known mass. When only one spectrum can be obtained from a biological sample available in small amounts, the standard deviation must be evaluated by determining the mass at each acquisition, e.g., 10 acquisitions, and the average can give the molecular mass with its standard deviation. If the acquisition system gives the average of 10 acquisitions directly, it does not give the standard deviation. Naturally, the averages are the same, but the first method is more informative.

Loss of Accuracy due to the Ionisation Process. Accidental error due to artifacts such as fragmentation, clustering, structural modification, presence of adducts, and so on. In most cases, these ambiguities are revealed experimentally, with measurements on known compounds. An accurate device can give an order of mass accuracy independently of the mass resolution for a sample containing a single compound. A low resolving power is compatible with a high mass accuracy. However, it should be noted that, in MALDI, protein adducts are a source of peak broadening. They depend on the type of protein. Calibration is not always possible, because the adducts cannot always be resolved. This is an example of loss of accuracy due to lack of resolution.

Resolution. A distinction is made between mass resolving power and mass resolution. They both refer to the separating power of the mass spectrometer for two ions with similar masses M and $M + \Delta M_x$, such that their adjacent peaks in a mass spectrum have the same size and profile (Gaussian, Lorentzian, or triangular) with a given overlap (10% valley as measured on the base line, full width at half maximum, etc.), denoted by the index x .

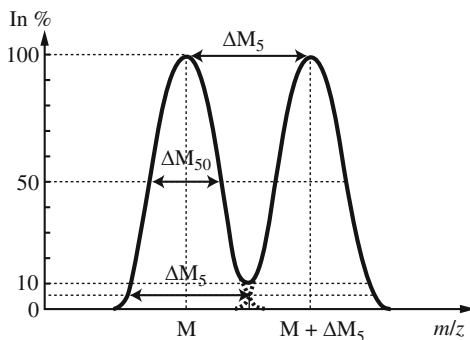


Fig. 10.2. Calculating the resolution

Mass Resolution. This is the difference ΔM_x in the mass (or rather the ratio m/z) which separates two adjacent peaks of masses M and $M + \Delta M_x$.

Mass Resolving Power. This is the dimensionless number $R = M/\Delta M_x$ (or $\Delta M_x/M$ for some authors).

The resolving power at 10% overlap can also be calculated with a single peak, by making the approximation that the width at 5% height of a single peak is equal to the distance between the tops of two peaks separated by a valley at 10% height. This value is called the theoretical resolving power. In this case, the calculation does not take into account problems due to any difference in profile of the two peaks (see Fig. 10.2). Another standard definition uses the width ΔM_{50} at half height, called the full width at half maximum (FWHM).

Sensitivity. This is the minimal amount of product needed to have a signal on a mass spectrum with a given signal-to-noise ratio. The absolute sensitivity of a device is measured in counts/mg of product.

10.2 Ionisation Sources for Biomolecules

10.2.1 Applications in Biology and Biochemistry

During the twentieth century, the techniques of mass spectrometry were applied across a wide range of disciplines, from space research to physico-chemical analysis of materials. One of the first peptide analyses was carried out on fortuitine, a peptide of molecular mass 1,359 Da, in the 1960s using electron bombardment [3]. There then followed a period of around twenty years when applications of mass spectrometry to peptides and proteins stagnated. Genuine biochemical applications to the study of peptides and proteins only began to take off at the beginning of the 1980s, with the discovery of ionisation by fast atom bombardment (FAB), followed closely by other techniques

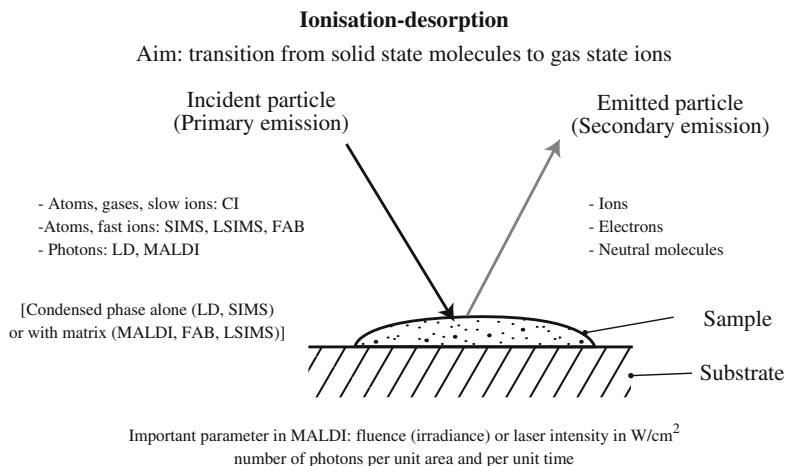


Fig. 10.3. Ionising a solid state sample

such as plasma desorption mass spectrometry (PDMS), liquid secondary ion mass spectrometry (LSIMS), electrospray (ESI), and matrix-assisted laser desorption/ionisation (MALDI). In parallel with the new ionisation techniques, time-of-flight (TOF) or quadrupole filter analysers were being developed. Although sector devices were the norm for high-resolution mass spectrometry up until the beginning of the 1990s, time-of-flight, quadrupole, and ion trap mass spectrometers, which are cheaper and easier to implement, have today come to dominate for applications in biology and/or biochemistry. However, with the falling cost of very high resolution, such devices are coming back into use, e.g., Fourier transform ion cyclotron resonance or orbitrap.

PDMS, FAB, and LSIMS

Applications of mass spectrometry to the study of biomacromolecules are largely based on the discovery in 1980 of new ways of ionising polar and labile molecules. Indeed, these compounds can be ionised by bombarding with beams of high energy primary atoms or ions. These are the ionisation modes for PDMS (used to analyse underivatised polypeptides and proteins), FAB, and LSIMS. The success of these methods led to the creation of a whole new discipline, whose aim was the structural study of biomolecules of fundamental importance in the biomedical sciences. The underlying principle of ionisation is shown in Fig. 10.3.

Plasma desorption mass spectrometry (PDMS) was developed in 1974 par Torgerson et al. [4]. It provided a way of analysing large, non-volatile organic molecules, such as certain underivatised peptides [5] or insulin [6]. This technique uses the nuclear fission products of ^{252}Cf , which pass through a thin

aluminium wall where the sample is deposited. It has a mass range of about 20 kDa.

In 1976, two years after the Torgerson publication on PDMS, Benninghoven presented the idea of secondary ion mass spectrometry (SIMS) [7]. The technique here is to focus a beam of primary ions on the solid sample. This causes a shock wave in the sample, which ejects secondary ions and molecules. However, the signals obtained by SIMS were short-lived, because the impact surface was quickly destroyed by the high-energy flux of primary ions. Barber had the idea of adapting this technique by first using a primary beam of argon atoms at 3–10 keV rather than a beam of primary ions. This avoids surface charge phenomena which were assumed to hinder the focusing of ions in the mass spectrometer [8]. The sample is introduced in the source in the presence of a low-volatility liquid matrix like glycerol, thioglycerol, or 3-nitrobenzylic acid. This ionisation technique was called fast atom bombardment (FAB). Barber then had the idea of using this liquid matrix with a beam of primary cesium ions at 20–30 keV rather than a fast atom beam. The use of primary ions leads to the same phenomenon of secondary ion emission as the use of primary atoms. The charge of the particles facilitates the focusing of the primary beam. This technique was called liquid secondary ion mass spectrometry (LSIMS). With a liquid matrix, the surface is constantly renewed with sample, so that secondary ion spectra can be obtained over longer periods. Soon after the publication of the analysis of insulin by PDMS [6], the LSIMS spectra of this protein were also published [9, 10].

However, all these techniques suffer from a certain number of rather serious limitations. The mass range is limited, with the most favourable cases reaching 25 kDa. The sensitivity is not as good as desired, requiring a few tens of picomoles to analyse a protein, while the analysis of mixtures is subject to significant signal suppression effects, making it much too selective.

ESI and MALDI

In 1988, two new mass spectrometry techniques were described. These enormously extended the field of application of mass spectrometry for the analysis of proteins and peptides, and seemed to correct the weak points of previously developed methods. During the American Society of Mass Spectrometry conference in San Francisco (California, USA) in June 1988, John Fenn of Yale university presented the application of electrospray ionisation (ESI) to protein analysis. This proved to be a genuine innovation for the mass spectrometry of biomolecules with high molecular mass. A few months later, during the International Mass Spectrometry Conference in Bordeaux (France), Franz Hillenkamp presented another ionisation mode called matrix-assisted laser desorption/ionisation (MALDI). In his presentation, he described protein mass measurements up to 117 kDa by time-of-flight mass spectrometry. MALDI looked as promising as ESI. John Fenn and Koichi Tanaka received the Nobel Prize for Chemistry in 2002 for the development of methods for

the identification and structural analysis of biological macromolecules, and for the development of soft desorption and ionisation methods for analysis of biological macromolecules by mass spectrometry.

The ESI and MALDI ionisation modes both provide efficient production of molecular ions from large biopolymers [11]. In 1988, the main question was: which technique would become the choice method for the future in a protein chemistry laboratory? It would seem today that there is no need to opt for one or the other, according to Peter Roepstorff in 1996 [12]:

In fact, at present, I consider the two techniques to be highly complementary. They have both dramatically improved the perspectives for the application of mass spectrometry in protein chemistry to such an extent that protein chemistry laboratories without access to these two techniques or at least one of them cannot be considered up to date.

10.2.2 Electrospray Ionisation (ESI)

The contribution of mass spectrometry to current understanding in biology still remains rather modest. However, electrospray ionisation (ESI) mass spectrometry, in the same way as MALDI, is now considered to be an essential tool in this field. Moreover, new possibilities are always appearing with the rapid development of associated technology (with regard to both instrumentation and methodology). We shall see that the sensitivity of the analysis can be improved by miniaturising the source. This ionisation mode is commonly associated with quadrupole filter analysers, but it has also recently been combined with orthogonal acceleration time-of-flight (oaTOF) analysers.

Description of the Ionisation Process

We shall not consider here the mechanical aspects of ESI. Two reviews provide a detailed description of this ionisation source [13, 14]. The process is simple enough to describe. A solution of the sample is introduced into a capillary tube which is taken to a high electrical potential. The strong electric field applied to the capillary outlet causes a cloud of charged particles to form. These particles simultaneously cross an electric field gradient and a pressure gradient in the direction of the mass spectrometry analyser. During this transport, the droplets get smaller by evaporation of the solvent due to successive Coulomb explosions, i.e., spontaneous division of the charged droplet into smaller droplets, caused by the high surface charge (see Fig. 10.4). Application of a pressure at the capillary input can facilitate nebulisation, depending on the flow rate and composition of the solvent [15]. In addition, a gas flow, usually nitrogen, is applied to the interface to encourage evaporation of the solvent. Some interfaces are heated. The ions formed at atmospheric pressure are then channeled by a set of samplers, i.e., a set of pumped orifices, toward the analyser which is housed in a high vacuum.

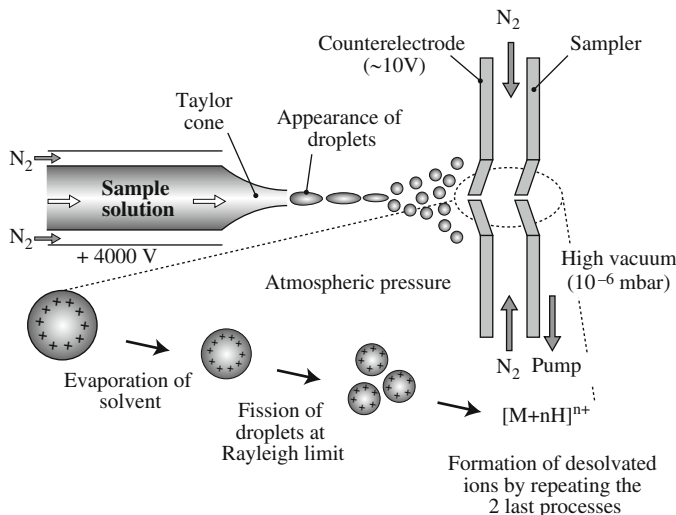


Fig. 10.4. The basics of electrospray ionisation

The electrospray process consists of three main steps:

1. droplet formation,
2. droplet diffusion,
3. gas phase ion formation.

If the end of the capillary is at a positive potential, cations will accumulate at the surface, causing the formation of a Taylor cone at the capillary end, along the axis of the applied field (see Fig.10.4). When this field is strong enough, the cone stretches into a long filament, which breaks up into positively charged droplets when the electrostatic force overcomes the surface tension [16]. The droplets have diameters in the micrometer range, depending on the electric potential, the flow rate of the solution, and the type of solvent. When the solvent evaporates this leads to a reduction in the droplet size. Now, the maximal charge for the stability of a spherical charged droplet of diameter d is proportional to $d^{3/2}$. The Coulomb explosion of the droplet into smaller droplets occurs at the Rayleigh limit. This is the limit at which electrostatic repulsion due to the charges overcomes the surface tension that was holding the droplet together. Various theoretical models have been put forward. The two main ones are the charged residue model and the ion desorption or evaporation model.

The ionic current of the electrospray, generally 0.1–0.3 mA, depends only slightly on the conductivity of the solution, the integrated current of ions actually transmitted (the time average over all ions) being 10–100 pA. This is due to the suppression effect of the buffers or problems of ionisation competition.

Ions formed in the positive mode are multiply protonated species.

Multiply Charged Species

ESI spectra generally contain a set of peaks corresponding to multiply charged species of the form $[M + H]^{n+}$, where M is the molecular mass of the analysed molecule, and n is the number of charges carried by this molecule when ionised. The mass M is very simply determined by deconvolution of the spectrum and transformation of the latter into a mass–intensity curve. For example, for two successive peaks of the m/z ratio measured on the spectrum with m_1 and m_2 (such that $m_1 < m_2$), and charges n_1 and n_2 (such that $n_1 = n_2 + 1 = n + 1$), we know that

$$m_1 = \frac{M + n_1}{n_1} = \frac{M + n + 1}{n + 1}, \quad m_2 = \frac{M + n_2}{n_2} = \frac{M + n}{n}.$$

We can thus calculate the charge state n and the molecular mass M :

$$n = \frac{m_1 - 1}{m_2 - m_1}, \quad M = \frac{(m_2 - 1)(m_1 - 1)}{m_2 - m_1}.$$

Since the electrospray produces multiply charged ions, the analysis of high molecular mass proteins (greater than 100 kDa) can be carried out using analysers with limited mass range. The electrospray has thus often been coupled to a quadrupole analyser, owing to its low cost and simplicity of use. But it has also been combined with magnetic sector mass spectrometers [17, 18], ion trap devices [19–23], Fourier transform ion cyclotron resonance analysers [24–30], and TOF analysers [31–37].

To increase the sensitivity of the electrospray, the source has been gradually miniaturised to become the microelectrospray (microESI) and then the nanoelectrospray (nanoESI). In a time-of-flight analyser, or ion trap, all ions effectively produced in the source can be transmitted to the detector, in contrast to quadrupole or magnetic sector analysers which operate by scanning. This therefore improves the sensitivity.

Preparing the Sample

Sample preparation is in principle rather straightforward. The compound to be analysed must be dissolved at a concentration of 1 fmol/mL to 10 pmol/mL in a solvent like methanol (or acetonitrile)/H₂O 1:1 (v/v) containing 1–5% (v/v) of acetic acid (or 1% formic acid), for the positive mode, or 5–50 mM ammonium acetate (or in the presence of NH₄OH) in the presence of a halogenated solvent such as trifluoroethanol to stabilise the anions, for the negative mode. Many organic solvents are compatible with ESI.

However, surfactants, non-volatile substances, and salts are poorly tolerated. Salts perturb the electrospray process and produce a series of adducts of the form $M + Na^+$, $M + K^+$, $M + H + Na^{2+}$, etc., which complicates the spectrum and reduces sensitivity. Presence of the buffer Tris is transparent. Flow rates in the injection capillary are of the order of mL/min for standard ESI and nL/min for nanoESI.

Limitations

Everyday use of ESI is in fact limited by the sensitivity and resolution of conventional mass spectrometers, especially those using quadrupoles. Large multiply charged molecules can be analysed, but the complexity of the spectra and the need for higher resolution, better mass measurement accuracy, greater sensitivity, and a broader dynamic range (to detect low intensity dissociation products) are still limiting factors when determining the sequences of large biopolymers.

ESI requires very pure samples, and the presence of other compounds or the micro-heterogeneity of many biopolymers can restrict its field of application. There are other limiting factors, such as the very small amount of material available. Since the search for detailed structural information requires MS/MS analyses, or the characterisation of specific non-covalent weak interactions, a particular preparation of biological samples is required (due largely to the incompatibility of ESI with common biological matrices).

For the analysis of small quantities of mixtures of biological products, the electrospray alone can sometimes provide insufficient information about each of the compounds. Combining mass spectrometry with separation techniques is a very useful alternative when the amount of available sample gets smaller, e.g., combining with liquid-phase chromatography (LC), capillary electrophoresis (CE), or capillary electrochromatography (CEC).

Improving Sensitivity in ESI MS: Microspray, Nanospray, Picospray

ESI is a high yield ionisation method. Unfortunately, sampling and transmission of ions in the mass spectrometer is much less efficient, since the yield is estimated at less than 10^{-3} [38]. This is mainly due to space charge effects in the source and the low yield of ion transport at the interface between the source and the analyser. The sensitivity of ESI, in the picomole range, is unfortunately inadequate for much biological analysis.

Some applications, in fact the most interesting but also the most delicate, require analytical techniques on the level of a single cell, in which the amounts of majority compounds are sometimes estimated at a few tens of attomoles. Detection techniques such as laser-induced fluorescence or electrochemical detection are extremely sensitive and can be associated with high resolution separation techniques such as capillary electrophoresis. However, the identification of compounds is solely based on their electrophoretic mobility. Progress with ESI soon showed that such levels of sensitivity would not be unattainable. Ions in spectra obtained by ESI come only from the sample solution. There is no matrix. This is an advantage of ESI over other techniques such as FAB or MALDI which, at a given sensitivity, can be limited by chemical noise due to the matrix. The ESI source was miniaturised during the 1990s in order to achieve more sensitive analysis. This work was done by

several groups. Emmett and Caprioli [39] called it the microspray. Wilm and Mann [40, 41] called it the nanospray. And finally, McLafferty and coworkers presented their picospray, operating with a slightly lower flow rate [42]!

Smith and coworkers had carried out a study on the use of low flow rates in ESI, in particular to combine capillary electrophoresis with ESI MS. They showed that a reduction in flow rate allowed one to increase the sensitivity, improve the stability of the spray, and use aqueous solvents. The end of the capillary is treated with hydrofluoric acid (HF) to reduce the thickness of the wall before coating with silver. Using capillaries with an internal diameter of 5 mm, it was possible to separate proteins and carry out an accurate mass measurement with about 600 amol of each protein [43, 44].

Emmett and Caprioli [39] used a capillary with a bigger diameter (internal diameter 50 mm) and a higher flow rate (about 800 nL/min). The end of the capillary was also treated with HF to obtain a wall thickness of 10–20 mm. The potential was applied simply using the conductivity of the solution. To analyse peptides, the capillary was filled with the stationary phase in order to carry out online concentration and desalting. In this way, they were able to analyse methionine enkephaline from 10 mL at 100 amol/mL. The system was used to study the metabolism of GABA B receptor antagonists *in vivo* in the rat brain after microdialysis [45].

For their nanospray, Wilm and Mann used a pulled glass capillary with inner diameter 1–3 mm. The capillary was coated with metal, and 1 mL of the sample solution was placed directly in the capillary before mounting it in the source. The flow rate, estimated at 20 nL/min, is maintained by the electrospray process itself. It is initiated and stabilised by a low pressure applied at the capillary input [40]. This device was used to sequence tryptic fragments in the femtomole range [46].

Finally, for their picospray, McLafferty and coworkers used a pulled fused silica capillary (inner diameter 5–20 mm throughout its length, and inner diameter 1–5 mm at the end), treated for a wall thickness of 50–80 nm. The flow rate is estimated in this case at about 1.5 nL/min. Combined with an FT-ICR mass spectrometer, it was able to analyse 10 amol of cytochrome c [42], and to measure the mass of carbonic anhydrase in a single red blood cell [47].

These are promising results for the ultrasensitive analysis of peptides and proteins in the attomole (10^{-18} mol) or zeptomole (10^{-21} mol) range. Methodological problems remain, however. A fundamental challenge is to improve the treatment of the inner capillary surface, and hence reduce sample loss and improve separation quality. Improvement of preparation procedures, as well as sample and buffer manipulation, is also useful to avoid obstructing the capillary, a common problem when the inner diameter is small.

The main advantages of miniaturisation are as follows:

- With such flow rates, there is no longer any need for a gas or liquid support (sheath flow).

- The smaller electrical potential needed for a stable spray means that the ionisation needle can be placed inside the gas vortex which enters the mass spectrometer, so that there is better transmission of the sample into the first vacuum region [48].
- The reduction of the infusion flow rate means that smaller charged droplets can be generated. These have a greater area to volume ratio, favouring ion desorption in the gas phase. A stable Taylor cone at low flow then requires a reduction in the dimensions of the ionisation capillary, these being available commercially with different inner diameters for several years now.

The better ionisation performance and greater sampling efficiency have a tremendous effect on the sensitivity of this kind of analysis. For example, Wilm and Mann estimate the proportion of sample actually introduced into the analyser at one molecule per 390 analyses, with the nanoESI source supplying 20 nL/min [40]. Previously, Smith et al. had measured a transfer efficiency to the analyser of only one in 10^4 with a standard ESI source operating at 3–6 μ L/min [49]. As early as 1995, McLafferty and coworkers obtained a detection limit in the attomole range with their miniaturised ESI source and a Fourier transform ion cyclotron resonance (FT-ICR) mass spectrometer [47]. These new ESI interfaces thus improve the detection limit, thereby allowing a smaller sample consumption, and are in fact more tolerant with regard to buffers and salts [40, 50].

10.2.3 MALDI

Historical Review

Ever since Karas and Hillenkamp described matrix-assisted laser desorption/ionisation (MALDI) [51], which can analyse proteins of more than 100 kDa, this technique has become as widespread in mass spectrometry as in biochemistry. MALDI applies to a wide range of biologically relevant molecules. It allows one to work with a broad mass range (from below 1 kDa to several hundred kDa), with a good mass measurement accuracy (especially for peptides and small proteins), while preparation and implementation remain relatively simple. Analysis is fast and tolerance to buffers, salts, and many surfactants is relatively good. Pulsed lasers have been used since 1976 [52] to produce ionised peptides from solid samples in mass spectrometry. This preliminary research and the applications that resulted from it over the following 10 years turned out to be useful only for a few short peptides. The probability of obtaining a useful mass spectrum depended on the specific physical properties of the peptide (volatility, photoabsorption spectra, etc.). As can be seen from Fig. 10.3, laser desorption (LD) ion sources are physically very similar to secondary ion sources (secondary ion mass spectrometry or SIMS), except that the high-energy particles used for irradiation are in fact laser photons. Many lasers have been used and lead to comparable performance. Organic compounds deposited on a surface could be analysed

by laser desorption with production of intact ions up to molecular masses of around 1,000 Da. Beyond this mass, detected species were almost always dissociation products of the original molecule [53]. Furthermore, the wavelength and pulse width of available lasers were limited and could not be easily adapted to different compounds. The use of a matrix as mediator for laser desorption/ionisation experiments had obvious advantages. One only had to look for small, polar compounds, soluble in solvents adapted to the relevant biological macromolecule, with strong absorption at the wavelength of the given laser, capable of codesorbing and transferring their charge to bigger bio-organic compounds. Although the principle was widely accepted in 1984, it took three years to find such a compound and validate the idea.

Two groups tackled the problem, one in Germany and the other in Japan. Tanaka et al. [54] developed a method using a finely ground metal powder in suspension in glycerol as matrix. Bombardment of different samples mixed with this matrix by UV laser photons led to mass spectra for molecules with molecular masses as high as 30 kDa, with high sensitivity and a good signal-to-noise ratio. In Hillenkamp's group, Karas et al. [55] had noticed that, for most sample deposits analysed by LD, the laser light was only poorly, or not at all, absorbed by the deposit. Almost all the energy was absorbed by the substrate, while the samples were virtually transparent to the photons. To improve the ion emission process, they had the idea of increasing the sample/substrate interaction by mixing to obtain a solid composite. The substrate had become a solid matrix holding the analyte molecules and protecting them from the destructive effect of the laser. Nicotinic acid, which absorbs at a wavelength of 266 nm (fourth harmonic of the Nd/YAG laser), was used to desorb proteins of more than 10 kDa [51], then 100 kDa in the following year [56]. The method using a solid matrix had a sensitivity 500 to 1,000 times greater than that using glycerol, and produced a better quality signal. For these reasons, the method using a solid matrix presented by Hillenkamp and Karas was soon to eclipse the method due to Tanaka. It was called matrix-assisted laser desorption/ionisation or MALDI.

Method of Ionisation

The MALDI technique generally uses a pulsed laser beam operating in the UV (although some instruments have been developed in the IR, despite the greater difficulty in constructing IR sources) to desorb and ionise a mixture of matrix and sample co-crystallised on a metal surface. The matrix minimises sample degradation caused by absorption of energy from the incident laser beam. The energy transmitted by the laser is absorbed by the matrix, thereby causing its expansion to a gas phase, carrying the sample molecules with it (see Fig. 10.5).

Laser irradiation thus leads to the ejection of the matrix and sample molecules in the gas phase. The sample is mainly ionised by proton transfer, either before desorption while still in the solid phase, or after desorption by collision with the excited matrix or with other molecules in the plasma, to give

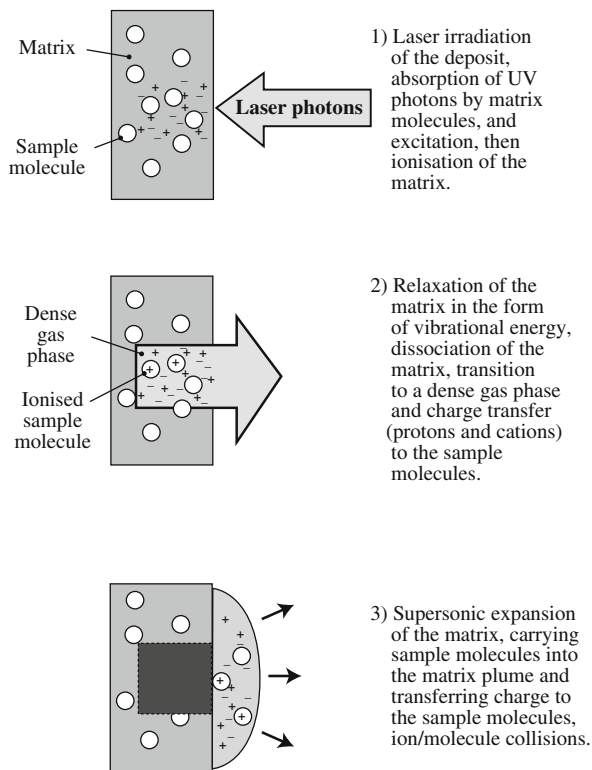


Fig. 10.5. Fundamentals of MALDI. Adapted from [57–60]

singly or multiply charged ions of the form $[M + nH]^{n+}$. Singly charged ions $[M + H]^+$ generally dominate in MALDI spectra. The origin of the protons has been studied recently. Labile protons (carboxylic or hydroxylic groups) from the matrix, the solvent, or even the sample itself, but also non-labile protons from the matrix, all seem to be involved [61]. The exact mechanism of the MALDI process is not yet fully understood. This process can be divided into four main steps (see Fig. 10.5):

1. firing and impact of laser photons,
2. ablation of matrix ions and molecules on the metal surface of the target,
3. expansion of the cloud ejected by ablation in the source and ionisation of neutral elements,
4. extraction of ions toward the analyser by application of an electric field.

Lasers of various wavelengths produce largely the same effect. This suggests that the mechanism exciting the matrix is not significant for the emission of ionised biopolymers. UV lasers probably involve an electronic excitation of the matrix, followed by transfer of the internal energy to higher vibrational energy levels, whereas IR lasers supply vibrational energy directly to certain

degrees of freedom, e.g., the Er/Yag laser acts on the stretching motions of the O–H bond [62]. The excitations induced by the laser are in part rapidly converted to thermal motions of the matrix molecules in a crystal volume located close to the surface [63], and in part transferred to the analyte molecules [62]. These motions appear to be violent enough to break the intermolecular bonds responsible for the cohesion of the crystal, and also to cause ablation of part of the crystal surface. However, the mechanisms for the transfer of electron energy to vibrational and rotational energy depending on the degrees of freedom of the molecules remain totally unknown. Bencsura et al. showed recently that these transfers depend on the amplitude of the sudden increase in internal energy of the matrix molecules, and by extension, on the fluence of the laser pulses used [62].

When the ions are formed in the gas phase, they can be accelerated toward the analyser by applying an electrostatic field. The MALDI source is usually coupled to a time-of-flight (TOF) analyser, which is well suited to pulsed ionisation by laser desorption. However, there is no fundamental reason for restricting MALDI to TOF. MALDI can also be combined with a Fourier transform ion cyclotron resonance analyser [64–68], an ion trap [69–72], or a magnetic sector analyser [73, 74], not to mention combinations of sector/ion trap, sector/TOF, TOF/ion trap, and TOF/TOF [60].

10.2.4 NanoSIMS and Ion Microscopy

For a long time, desorption techniques via secondary emission following irradiation of the solid sample by fast atom bombardment (FAB) or ion bombardment (secondary ion mass spectrometry or SIMS) saw few applications in biology. But today, SIMS is one of the most sophisticated and powerful ionisation methods for biological imaging. Originally introduced in the 1960s by Castaing and Slodzian [75], this technique suffered from limitations due to its poor imaging resolution (0.5–1 μm) and due also to the noise induced by adjoining a matrix for the analysis of purified solutes. There is a parallel with the mechanism in MALDI, where the primary irradiation is carried out using photons rather than fast atoms or ions. When a solid sample is irradiated by primary ions with an energy of a few keV, some of the particles emitted from the plate are ionised (see Fig. 10.3). SIMS involves mass spectrometric analysis of these secondary ions. It thus provides information about the elemental, isotopic, and molecular composition of the upper layers of the sample prepared on the target.

SIMS thus came into its own at the end of the 1990s with the development of SIMS imaging or ion microscopy [76], and with the commercialisation of a first instrument by Cameca, which made this technology accessible to biological imaging laboratories. The approach via analytical chemistry must face a general problem when analysis has to be carried out within subcellular compartments, in order to compare different states of some type of tissue or cell to obtain a better understanding of the involvement of specific molecules

in the organelles. Standard approaches begin with a cell separation and measurements on the purified fractions. However, this approach can suffer from a bias introduced by an artificial redistribution of the relevant analytes from a given physiological location to other intracellular sites with a greater affinity, but no longer relevant to the cell physiology.

The latest developments in analytical techniques tend towards the direct detection of the relevant compounds at the subcellular level in tissues and/or cells. For example, confocal microscopy has been used for cation imaging (Ca^{2+} and Na^{+}) in the cell, with the help of fluorescent molecules that can be induced in the presence of specific cations [77, 78], for immunocytochemical studies of receptors using fluorescent antibodies, or for molecular imaging by fluorescence or multiphoton microscopy [79, 80]. For example, an electron gun was used to study the elemental composition of cellular sub-compartments such as the endoplasmic reticulum [81, 82]. Imaging techniques using SIMS appeared gradually in biology and medicine, and are growing in importance thanks to their sensitivity and the possibilities of detecting isotopes and imaging molecules in single cells or subcellular compartments.

The main problem is the preparation of samples. SIMS is carried out in vacuum, which rules out the analysis of living cells. In order to study cells in their native state and with their physiological biochemical distribution, some way must be found to immobilise them cryogenically.

Instrumentation

Ion microscopes associate the SIMS source with two types of analyser: sector analysers [75] and time-of-flight analysers [83]. Magnetic sector instruments, still the most widespread at the present time, give submicrometer spatial resolution. Many current configurations derive from the original work by Castaing and Slodzian [75]. Other instruments able to achieve higher spatial resolution have been developed [84, 85]. The spatial resolution is obtained by an ion beam, which can be focused on areas of about 50 nm in diameter for Cs^{+} and 150 nm for O_2^{+} . This beam then scans the whole sample. As in many experimental setups, the higher the resolution, the lower the sensitivity.

Typically, the nanoSIMS source includes the primary optics and a secondary optical system coupled with the analyser and detector of the mass spectrometer. The primary source contains the optics required to focus the primary O_2^{+} or Cs^{+} ion beam. These ions are accelerated to the desired energy (keV). The beam is then focused by electrostatic lenses onto the sample at a voltage of 3,000–4,500 V in vacuum (10^{-9} torr). When the primary ion beam bombards the sample surface, characteristic secondary ions are produced. Most of the chemical bonds are broken and atoms and/or fragments of molecules are ejected from the surface layers of the sample over a thickness of 1–2 nm, either as neutrals, or in the form of charged particles, which are then extracted from the source via the immersion lens and focused with several transfer lenses. The ionic image of each irradiated zone is then recorded. To

restitute the image of the whole sample, the latter is scanned and the data gathered point by point for a 3D reconstruction (xy space, m/z ratio). As for the electrospray and MALDI, the details of quantification are still to be fully elucidated.

Preparing the Sample

The preparation depends on the type of sample (tissue, cells). Furthermore, the sample must be a conductor since it is raised to a potential of a few kV. Cells must be fixed to avoid any diffusion of intra- and extracellular solutes, or intraorganelles within the cell itself. All molecules can therefore suffer the effects of artifactual diffusion and redistribution. Small ions like Ca^{2+} , Na^+ , and K^+ , like proteins, can diffuse in fixed tissues [86]. Moreover, binding in aqueous polymer matrices, commonly used for transmission electron microscopy, is not possible in mass spectrometry. Finally, freezing and drying often degrade cells by modifying the morphology of the original sample. On the other hand, a rather brutal freezing method can preserve the surface structure of atomic layers to a thickness of about $10\text{ }\mu\text{m}$ [87–89]. Another problem is to deal with contamination from the extracellular medium, which contains many non-volatile salts that are detrimental to analysis by mass spectrometry. Drastic washing would perturb the equilibrium of the cells and could even cause lysis. For SIMS imaging of tissues, small pieces of tissue are rapidly frozen to fix the diffusing species as well as possible. The frozen tissue is then cut into slices from 0.5 to a few micrometers thick with a microtome, before being flattened on the support at low temperature, thereby achieving good adhesion and good conductivity [90,91]. The slices are then coated with a film of Au/Pd to optimise the conductivity of the whole preparation. Gold clusters $(\text{Au}_3)^+$ have also been used recently as primary ion source. This can considerably increase the emission of secondary ions of lipids and peptides. This source has been tested with success on slices of mouse brain [92] and on mouse paws, to characterise Duchenne muscular dystrophy [93].

10.3 Analysers

10.3.1 General Considerations

The most common conventional analysers associated with ESI and MALDI sources are:

- the quadrupole filter (Q = quadrupole),
- the 2D or 3D ion trap,
- the time-of-flight (TOF) analyser,
- Fourier transform ion cyclotron resonance traps (FT-ICR),
- and any combination of these analysers.

Table 10.1. Main features of standard analysers for MS

Analyser	Resolution ($M/\Delta M$)	Mass range	Accuracy (ppm)
3D ion trap	1 (full scan) at 1,000 (high res.)	4×10^3 Da	100–300
2D ion trap	10 (full scan) at 10,000 (high res.)	4×10^3 Da	100–300
Q	1	4×10^3 Da	100–300
TOF	10^4	3×10^5 Da	<30
Orbitrap	5×10^4	5×10^3 Da	<5
FT-ICR	10^5	5×10^3 Da	<3

Magnetic sector devices, which are still used in ion microscopy, are much less sensitive and should gradually be replaced by TOF.

An analyser is characterised by at least three key parameters:

- resolution,
- accessible mass range,
- sensitivity.

Table 10.1 sums up the main features of the various analysers.

Quadrupole and TOF analysers have characteristics such as resolving power and mass range that are well suited to the analysis of biological samples, combined with relatively simple use. TOF measures the time of flight of sample ions and relates this to their mass/charge ratio. The quadrupole operates as a filter, letting through only one given mass/charge ratio at any given time, then scanning over the relevant mass range. Quadrupole with TOF has become a classic combination, as we shall see later.

The ion trap has some interesting features: very high sensitivity and the possibility of carrying out rather complete structural analysis with a single analyser. However, its mass range is limited ($4\text{--}6 \times 10^3$). The resolution of ion trap devices varies enormously depending on the scan rate. This is why the table only gives broad ranges compatible with fast analysis times in the context of LC–MS combinations, in particular.

Sensitivity is generally estimated globally for a particular association of a type of ion source with the analyser. It varies significantly as instruments are further developed. At the present time, quantities in the femtomole range (10^{-15} mole) can be detected with all these instruments.

10.3.2 Time-of-Flight Analyser

MALDI is generally used in conjunction with the time-of-flight (TOF) analyser, as already mentioned. TOF analysis is well suited to the pulsed nature of MALDI desorption. In addition, since TOF analysis has in theory no upper limit to the mass range, it is compatible with MALDI, with produces ions with high values of m/z .

The ions formed in the source by laser impact are then accelerated by a strong, uniform electric field. All ions with a given value of m/z acquire the same kinetic energy. The ions then enter the free field region (FFR) of the analyser, where their speed remains constant. The time taken by the ions to reach the detector will then be correlated with the value of their ratio m/z . The heavier the ion, the higher the value of m/z and the slower it will be. The smaller and lighter it is, the faster it will be. By using a pulsed laser and by measuring the time of flight of the ions, i.e., the time they require to reach the detector, their mass/charge ratio can be determined. In theory, infinitely high masses can be measured, but in practice, the limit is around 300,000 Da.

Linear Mode

The time-of-flight (TOF) analyser separates the ions according to their speed when they move in the flight tube. A bunch of ions is generated in the source, at the analyser inlet. These ions are accelerated in the flight tube by an electric field applied between the sample holder and the extraction grid (accelerating voltage 10–25 kV). Since the ions are accelerated over the same distance l_{acc} by the same force F , they all acquire the same kinetic energy. A population of ions with a given mass distribution, but all with the same kinetic energy, has a well-defined velocity distribution v_f . The average speed v_f of each ion is inversely proportional to the square root of the ratio m/z for that ion.

Once accelerated, the ions enter the flight tube (length $L = 1\text{--}3\text{ m}$ in general), which is free of field (see Fig. 10.6). The ions separate in this flight tube in a way that depends on the speed acquired in the acceleration zone, before reaching the detector placed at the far end of the flight tube. For a given flight length L and accelerating voltage V_{acc} , the time t_{total} elapsed between extraction of the ions and their arrival at the detector is proportional to the square root of the ratio m/z .

Ions with the smallest values of m/z will arrive first. The spectrum is obtained by measuring the signal recorded by the detector as a time–intensity function for each bunch of ions. Since all the ions are detected, TOF analysers have very high sensitivity.

The resolving power with TOF is limited by various factors, such as initial energy distribution, localisation of the ions before acceleration, time and length of acceleration, and so on. TOF analysers operating in linear mode are often limited by their resolution ($M/\Delta M_{50}$ about 2,000).

Reflectron Mode

Although in theory all ions with the same mass/charge ratio should acquire the same kinetic energy, in practice, this is not quite true, because there is always some initial kinetic energy distribution, before the ions are even accelerated, and this effect limits the mass resolution. Indeed, all ions with the same value of m/z will not take exactly the same time to arrive at the detector, and this

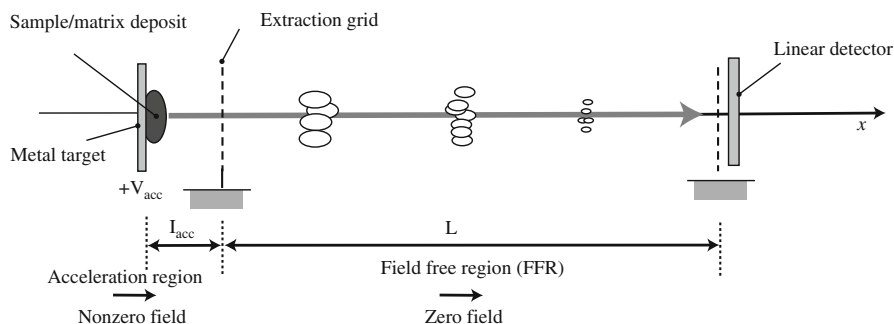


Fig. 10.6. Setup for linear time-of-flight (TOF) analysis

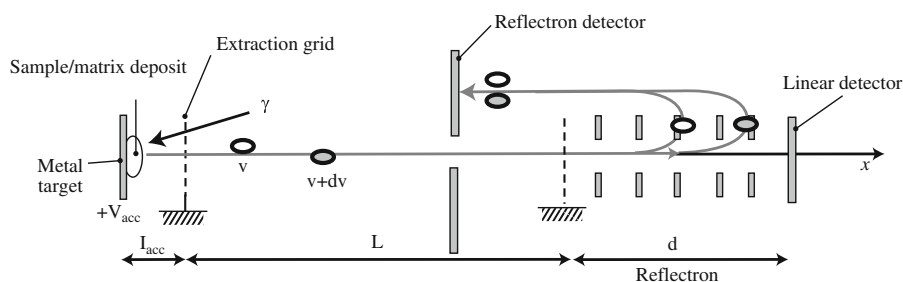


Fig. 10.7. Setup for reflectron time-of-flight (TOF) analysis

leads to an inevitable broadening of the peaks. In order to increase resolution, a reflectron or electrostatic mirror is used. An electric field is applied in the opposite direction, so that the ions are slowed to a halt and go back the other way. This lengthens the time of flight and makes it possible to refocus the ions of the same ratio m/z and different kinetic energy on the reflectron detector. In this way, the resolution is significantly improved for peptides, for example.

The most effective focusing method yet developed is the reflectron. The main feature is an electrostatic mirror which imposes an electric field opposing the advance of the ions. This electric field is in the opposite direction to the one used to accelerate the ions. The ions therefore turn around and leave the reflectron with a longitudinal velocity in the opposite direction to that of their initial velocity. The most energetic ions go further into the reflectron field and hence take longer to be reflected. However, they get there before the less energetic ions. This means that all the ions with the same value of m/z can be focused on a plane. The optics of the instrument can be adjusted in such a way that the ions get slightly deflected toward another detector, off the longitudinal displacement axis of the ions (see Fig. 10.7).

The electrostatic mirror can lengthen the flight distance without increasing the size of the instrument. The reflectron detector is positioned on the focal

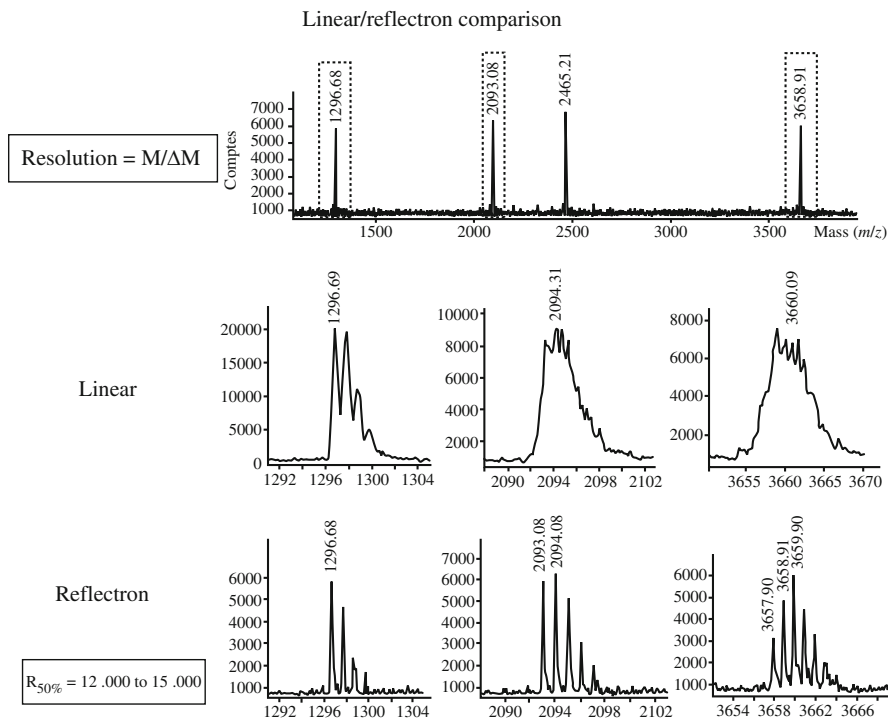


Fig. 10.8. Comparing the resolution obtained on a standard peptide mixture using linear and reflectron modes

plane of the ions with the same value of m/z . The reflectron corrects for the kinetic energy dispersion, but it does not correct for the temporal dispersion.

Comparing Linear and Reflectron Modes

In linear mode, a given peptide gives a single peak corresponding to the average mass of the compound.

The reflectron mode can differentiate between the different isotopes of a given species. For a peptide, the first peak can be considered as the monoisotopic peak, i.e., the peak corresponding to the species made up of the most abundant isotopes with the lowest masses (^1H , ^{12}C , ^{16}O , and ^{14}N), while the second peak corresponds principally to species carrying a ^{13}C , for example. The resolution is clearly improved (see Fig. 10.8).

Orthogonal Acceleration Time-of-Flight Analyser

To couple a time-of-flight analyser with a continuous ion source, the ions must be sampled over time in such a way as to introduce them discontinuously into

the analyser. The idea of off-axis acceleration has been known since the 1960s from work carried out in research and development at Bendix (1964). However, orthogonal acceleration was only described 15 years later.

In an orthogonal acceleration time-of-flight (oaTOF) analyser, the ions are sampled from a parallel incident ion beam produced by a continuous source. This low-energy ion beam is contained in an ion acceleration region, initially held free of field. An electric field is then applied in the form of a rapid pulse to the electrodes bounding this region. This field induces a force that is strictly and exclusively normal to the axis of the incident ion beam. As this beam is parallel, all the ions have zero average velocity and very low velocity dispersion in the direction of the force, before it is applied. The finite width of the beam gives a spatial dispersion which is easily corrected by refocusing. The velocity dispersion is negligible.

There are many advantages: very efficient ion sampling in time, simultaneous correction of velocity dispersion in space, improvement (without reflectron) of the resolving power, maximal efficiency for high masses, and minimal temporal dispersion. The resolution can be further improved by coupling a magnetic field analyser with an orthogonal acceleration time-of-flight analyser, because the ion beam is narrower.

10.3.3 Quadrupole Analyser

Quadrupole analysers comprise four metal bars, held at constant electrical potentials, two positive and two negative, with like pairs being diametrically opposite one another. This setup produces a constant electric field, to which a variable component is added. The resulting electric field is directed perpendicularly to the ideal trajectory of the ions. The ions accelerated along the quadrupole axis have the same speed in this direction, but are subject to a transverse force resulting from the action of the electric field. Only those ions whose trajectory is held between the four bars can be detected. For a given value of the electric field, the quadrupole only lets through the resonant ions, i.e., those of a given m/z ratio. By steadily varying the electric field, a given range of values of m/z can be scanned.

Theory

The quadrupole is probably the most widely used analyser in mass spectrometry. It comprises four bars of hyperbolic or circular cross-section, arranged around the trajectory of the ions through the analyser. It filters the ions according to the value of the ratio m/z . Each opposing pair of rods has an electrical potential of opposite sign to the other pair of rods (see Fig. 10.9). This potential is produced by superposing a continuous component U and a sinusoidal component of the form $V \cos(\omega t)$.

The trajectory of an ion with given value of m/z is governed by the Mathieu equations:

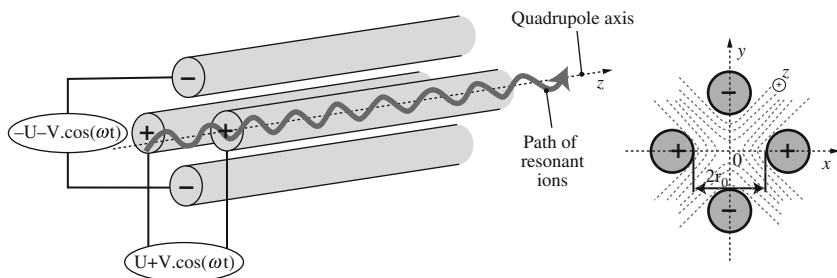


Fig. 10.9. Quadrupole analyser

$$\frac{\partial^2 x}{\partial(\omega t/2)^2} + \left\{ a + 2q \cos [2(\omega t/2)x] \right\} = 0 ,$$

$$\frac{\partial^2 y}{\partial(\omega t/2)^2} + \left\{ a + 2q \cos [2(\omega t/2)y] \right\} = 0 ,$$

where

$$a = \frac{8zeU}{mr_0^2\omega^2} , \quad q = \frac{4zeV}{mr_0^2\omega^2} ,$$

whence

$$\frac{a}{q} = \frac{2U}{V} ,$$

and $2r_0$ is the distance between two opposite bars. Solving these equations gives a domain of stability, i.e., the set of pairs of values of a and q for which those ions with a specified value of m/z will pass through the electric field produced by the four bars. To these pairs of values (a, q) correspond values of the amplitudes U and V , and the frequency ω .

Practice

We shall not attempt to solve the above equations! The operation of the quadrupole can be described qualitatively. The alternating component induces an oscillation in the trajectory of the ions. When the electrical potential applied to the left- and right-hand bars is positive, the resulting force repels the positive ions toward the vertical plane, but when the electrical potential applied to the upper and lower bars becomes positive in its turn, the positive ions are repelled toward the horizontal plane (see Fig. 10.10).

The amplitude of this oscillation depends on the value of m/z for the given ions. Below a minimal threshold value of m/z , the ions have too high an acceleration (and oscillation amplitude) and collide with the bars of the

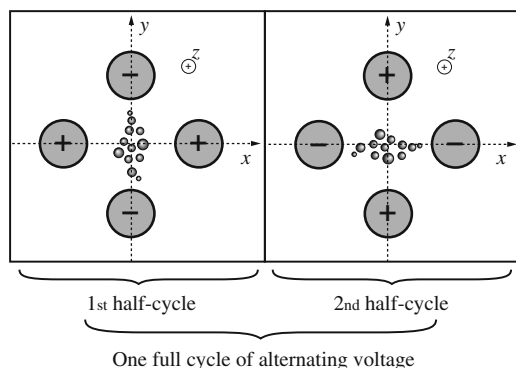


Fig. 10.10. Oscillation of the rod potentials (cross-section)

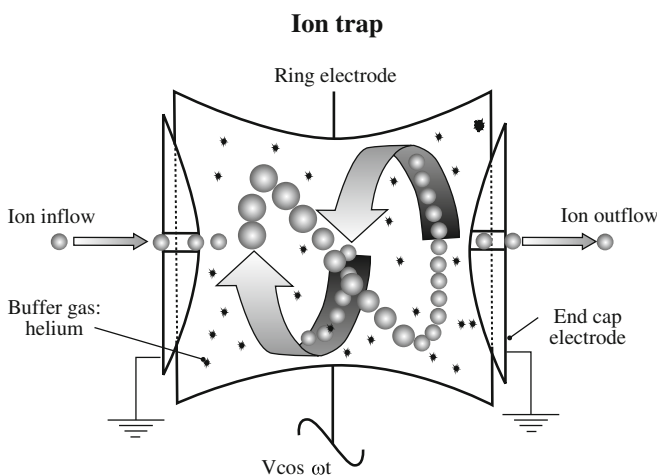


Fig. 10.11. Three-dimensional ion trap

quadrupole, whence their motion is arrested. The quadrupole acts as a high-pass filter. The threshold value depends on the amplitude V and frequency ω of the alternating voltage.

The continuous component acts in combination with the alternating component as a low-pass filter. Ions with higher values of m/z than some maximal threshold value have lower acceleration, and the continuous component will have a proportionally greater effect on their trajectories. They do not refocus along the z axis of the quadrupole during the alternating voltage cycle with frequency ω . They thus gradually drift away from the axis under the effect of this continuous component. At the analyser output, their trajectory has been too far deflected to be able to reach the detector. The threshold value depends on the two amplitudes U and V and the frequency ω of the alternating voltage.

The combination of high- and low-pass filters allows one to select a value of m/z . The values of the frequency and the ratio U/V are determined so that the minimal and maximal threshold values are as close as possible, in order to obtain the best possible resolution. By scanning over the voltage, for fixed U/V and frequency, the device scans over the ratios m/z of the resonant ions. Ions with different ratios m/z will successively be allowed to get through the field, bearing in mind that the ratio m/z of the selected ions varies linearly with the voltages U and V . The other ions will have unstable trajectories and will be filtered, without ever reaching the detector.

The quadrupole is an inexpensive device, and has the advantages of being compact, with a high throughput and able to scan at a high rate. The maximal mass range is $2,000 < m/z < 8,000$. The resolution of these analysers depends on the measured mass M . In practice, $R = 2M$ at 50% valley, i.e., for $M = 7,000$, $R_{50} = 14,000$.

10.3.4 Ion Trap

Three-Dimensional Ion Trap

Paul and Steinwedel described the 3D ion trap in 1960. This analyser comprises an annular hyperbolic electrode in the shape of a spool, covered by two spherical end cap electrodes, connected to one another electrically, as shown in Fig. 10.11. Conceptually, one could view the ion trap as a circular or toroidal quadrupole in which the inner bar has been reduced to a fictional point, the outer bar has become a circle, and the two upper and lower bars have become two end caps.

A sort of 3D quadrupole is obtained by superposing continuous and alternating voltages in which the ions are held captive (trapped) on a 3D figure-of-eight trajectory called a Lissajous curve. The ion trap is a scanning mass analyser which appeals to detection by resonance frequency. However, the scanning here is different. Ions with different values of m/z are simultaneously present in the trap, but they are ejected according to the value of m/z in order to record the spectrum.

For each value of m/z , the ions oscillate in the radial and axial (z) directions at natural frequencies that differ from the radio frequency (RF) applied to the electrodes. By superposing an alternating (a.c.) voltage in the z direction, with the same frequency as the resonance frequency of the ions for a given value of m/z , energy is transferred to these ions. Their trajectory in the z direction (toroidal axis of symmetry) is then destabilised: this is radial ejection. By scanning the frequency of the a.c. voltage, ions with different values of m/z are expelled successively.

The ion trap is a low-cost analyser, but the mass range is somewhat limited. The minimal mass of the acquisition mass range is a function of the maximal mass chosen, e.g., if one wishes to detect ions up to 2,000 Da, the minimal acquisition mass will be around 200 Da.

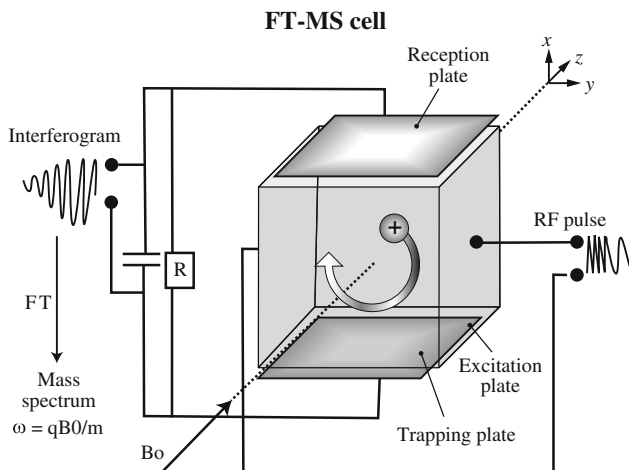


Fig. 10.12. Schematic view of the FT-ICR cell

Ions with ratio $m/z = 8,000$ can be analysed, but if the ratio m/z is too low, the ions are no longer trapped. The sensitivity is comparable to TOF. However, since the dynamic range of the ion trap is low, its resolution can be considerably reduced when a species is highly abundant in the trap.

Linear Ion Trap

The same scanning method can be applied to a multipole system which plays the role of the ring electrode, with supplementary electrodes (multipole or lens) on either side. The latter play the role of the end cap electrodes. Hence, the trap is produced at the centre of the central quadrupole. The size of the device provides for a larger trapping volume, thus minimising ion interactions and space charge effects for a given number of ions present. The sensitivity of the device is thereby improved, all the more so in that the transfer and trapping efficiency is increased.

10.3.5 Fourier Transfer Ion Cyclotron Resonance (FT-ICR) Analyser

New 2D ion traps are currently being developed and commercialised by the manufacturers. They have a different geometry which improves capacity and throughput to the detector (see Fig. 10.12).

The Fourier transform ion cyclotron resonance analyser was developed in 1974 by Comisarow and Marshall [94, 95]. This technique has applications in a variety of fields from analytical chemistry to the physical chemistry of ion/molecule reactions. Recent technological developments with this analyser have allowed it to be adapted to soft ionisation sources such as electrospray

(ESI) or matrix-assisted laser desorption/ionisation (MALDI), making it a very powerful tool for analysing biomacromolecules. By increasing the magnetic field (up to 20 tesla in some devices), it has been possible to analyse large proteins using ESI with high resolution and sensitivity [96]. One should also mention the analysis of hemoglobin contained in a single red blood cell [97]. During the latest mass spectrometry conferences, more and more reported applications have been concerned with environmental issues. For example, at the congress organised by the American Society of Mass Spectrometry in 1998, Alomary et al. presented studies of pesticides and dioxins in water with a sensitivity in the range of a few parts per quadrillion (ppq) [98]!

Fourier transform ion cyclotron resonance mass spectrometry is a technique with excellent linearity between the cyclotron frequency and the mass/charge ratio of the ion, under constant magnetic field conditions. The name is abbreviated to FT-ICR MS or just FT MS.

The Fourier transform principle states that, when the intensity of a signal is measured as a function of time, it will be found to be a superposition of terms corresponding to different frequencies, each with its own intensity. The Fourier transform provides a way of retrieving the component frequencies and their corresponding intensities.

The FT MS ion analysis happens in three stages:

- trapping the ions,
- exciting the ions,
- detecting the ions.

The FT-ICR analyser exploits the ion cyclotron resonance. The ions are produced in the source, then stored in the cell of a cyclotron resonance analyser located in a uniform magnetic field produced by a powerful magnet. The ions are held in a circular orbit produced by the magnetic field with a frequency given by $w = qB/m$, where w is the cyclotron frequency (rad/s), B is the magnetic field (T), and m/q is the mass/charge ratio of the ions.

The cyclotron resonance can be induced by applying a variable electric field, e.g., a sinusoidal field. When the frequency of this sinusoidal component is equal to the cyclotron frequency of the ions, the resonance conditions are fulfilled and the ions are accelerated toward a longer radius of gyration. This is fundamental to analysis in mass spectrometry: the ions that are not associated with this cyclotron frequency are not accelerated. The FT-ICR mass spectrometer detects the image current induced by the motion of ions stored in the analyser cell. This is shown in Fig. 10.12 for a group of positive ions in coherent cyclotron resonance between two electrodes. When the ions move away from electrode 1 and approach electrode 2, the induced electric field causes a displacement of electrons in the outer circuit toward the resistance. They then accumulate on electrode 2. During the other half of the cyclotron orbit, the electrons leave electrode 2 and accumulate on electrode 1 when the ions come near. The flow of electrons in the outer circuit is called the image

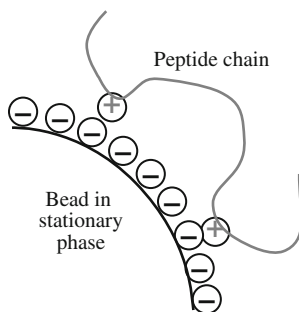


Fig. 10.13. Separation of a peptide by cation exchange chromatography. The solute is positively charged and can interact with the anionic stationary phase

current. This signal can then be amplified and detected. It is thereby possible to detect the ions without collision on the electrodes.

In practice, the ions are injected into a cylindrical cell a few centimeters in diameter placed in a magnetic field of a few tesla, e.g., 3 T. For such a field, the cyclotron frequency is 1.72 MHz at the mass of 28 Th and 12.03 KHz at 4,000 Th. The technique consists in simultaneously exciting all the ions present in the cyclotron by scanning rapidly over a broad range of frequencies, in a time of the order of one microsecond. The result is not only to produce a trajectory passing close to the wall perpendicular to the orbit, but more importantly to set all the ions in phase. The complex wave detected as a function of time can then be transformed into an intensity–frequency function by means of the Fourier transform method.

An FT-ICR device is well-suited to the analysis of complex mixtures. Resolution can reach 8×10^6 .

10.4 Combined Liquid Phase Separation and Mass Spectrometry

10.4.1 Chromatographic Techniques

Ion Exchange Chromatography (IEC)

In ion exchange chromatography, the stationary phase in the column carries charged chemical groups, with which the ionised solutes will be able to set up electrostatic interactions. In this way, cation exchange chromatography (SCXLC for strong cation exchange) uses a stationary phase carrying negative charges, while anion exchange chromatography (SAXLC for strong anion exchange) uses a stationary phase carrying positive charges. Figure 10.13 shows the separation of a polypeptide by SCXLC.

This chromatography can be coupled with reversed-phase chromatography, allowing a 2D separation.

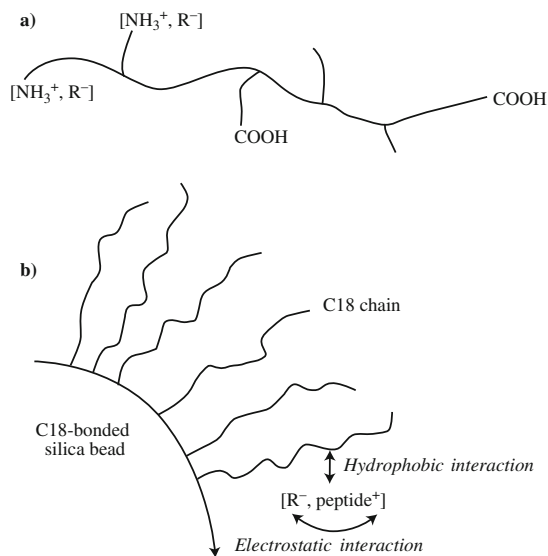


Fig. 10.14. (a) Schematic view of a peptide in the acidic mobile phase of RPLC. Its primary amine functions are positively charged and can form pairs of ions with the counterion of the acid, e.g., trifluoroacetate when trifluoroacetic acid is used. The notation R symbolises the aliphatic, hydrophobic nature of the counterion. The carboxylic acid functions of the peptide are neutral. (b) Representation of the interaction of the ion pair with the octadecyl-bonded phase

Reversed-Phase Liquid Chromatography (RPLC)

Over the decades between 1940 and 1970, ion exchange chromatography, which to a first approximation separates analytes according to their charge, remained the predominant chromatographic technique for analysing amino acids and small peptides. However, progress in high-performance liquid chromatography (HPLC) rekindled interest in non-polar stationary phases. In 1977, Molnar and Horvath tested an octadecyl-bonded column on peptides produced by protein digestion. They obtained a fast separation that was more effective than on an ion exchange column, except for polar amino acids and weakly bonded peptides [99]. At the end of the 1970s, RPLC came to the fore in peptide analysis. In reversed-phase chromatography, a column containing the hydrophobic stationary phase, such as silica particles bonded to aliphatic groups (octyls C8, octadecyls C18), is used. The solutes are separated according to the strength of their interaction with the stationary phase, hence in accordance with their hydrophobicity.

The operating conditions generally used to separate peptides at the end of the 1970s were as follows [100, 101]:

- The columns are filled with silica particles bonded by alkyl chains, usually octadecyls.

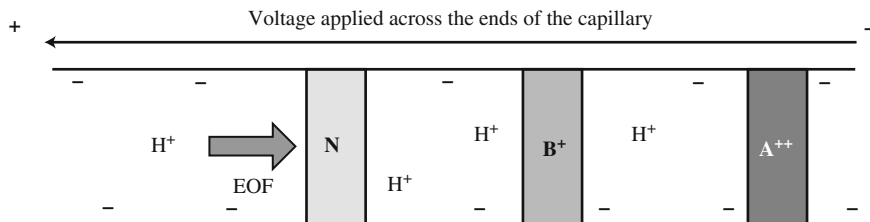


Fig. 10.15. Separation by capillary electrophoresis (CE). Charged analytes like B⁺ and especially A⁺⁺ migrate more quickly than neutrals like N

- The pH of the mobile phases is acidic, between 2 and 4, e.g., imposed by a phosphate buffer.
- Acetonitrile or methanol is used as an apolar modifier, generally in programmed elution. Indeed the latter is required to separate a mixture of peptides with too broad a range of polarities.

The separation of peptides on supports with reversed phase polarity is largely dictated by the hydrophobicity [102] and can involve a mechanism for the formation of ion pairs with the anion of the acidic buffer [103]. With an acidic pH (roughly 2–3), the carboxylic acid functions of the side chains are neutral, and the protonated amine functions can associate with the counterion of the buffer (see Fig. 10.14). This reduction of the peptide charge has the effect of increasing its retention.

10.4.2 Electrophoretic Techniques

Capillary (Zone) Electrophoresis (CE/CZE)

In capillary electrophoresis, species are separated according to their electrophoretic mobility. The sample is introduced into a capillary, across the ends of which a voltage is applied. The silica walls of the capillary are negatively charged (deprotonated silanols). The protons released into the solution give rise to an electro-osmotic flow (EOF) when the voltage is applied. The neutral solutes will migrate at the speed of the EOF, whereas the positively charged analytes will migrate at a speed resulting from the sum of the EOF and their own mobility (see Fig. 10.15). The negatively charged analytes migrate towards the anode and are not detected.

Capillary Electrochromatography (CEC)

Electrochromatography is a hybrid technique combining capillary electrophoresis and chromatography. The silica capillary used here contains the stationary phase, either in the form of a film covering the inner surface of the capillary, or in the form of bonded silica particles, and the analytes interact with this

stationary phase. When a voltage is applied across the ends of the capillary, an EOF is created. The analytes have varying affinities for the stationary phase, and this determines their separation. The main difference with liquid phase chromatography (LC) is that no pressure is applied. The mobile phase is set in motion by applying the voltage.

Capillary Isoelectric Focusing (CIEF)

In capillary isoelectric focusing, the analytes – usually peptides or proteins – are separated according to their pI. A pH gradient is generated in the capillary using ampholytes (or zwitterions), molecules which carry both an acid and a base function. These ampholytes must have pI values covering the range required for separation of the proteins. A base solution is placed at the cathode, and an acid solution at the anode. When a voltage is applied across the ends of the capillary, the analytes and ampholytes migrate until they reach a position in the capillary where they are globally neutral. An equilibrium is then reached. The species separated according to their pI values are then carried to the detector, e.g., by applying a pressure at the end of the capillary.

10.5 Which Mass Spectrometer Should Be Coupled with Separation Techniques: ESI or MALDI?

10.5.1 Combinations with HPLC

RPLC discussed above for the separation of proteolytic peptides can be coupled with mass spectrometry, to identify the analysed peptides. Since the end of the 1980s, the two ionisation techniques ESI and MALDI have become the norm for biomolecular analysis. Here we compare them in the narrower field of proteomic analysis by RPLC–MS/MS. Various analysers can be used downstream of the ionisation source. We shall compare them in terms of sensitivity, resolution, dynamic range, and flow rate, the relevant criteria when studying complex mixtures of small quantities of peptides.

The electrospray source is usually associated with liquid phase separation techniques, because it allows online analysis by mass spectrometry of the outflow from the column or capillary. Microspray interfaces are compatible with flow rates generally less than $\mu\text{L}/\text{min}$, while the nanospray source, which uses capillaries with an orifice of $1\text{--}2\mu\text{m}$, can produce a stable signal at a few nL/min [104].

Bibliographical references using ESI sources subsequent to these miniaturisation developments are not always very precise in the terms employed. Indeed, no distinction is made between ESI, microESI, and nanoESI sources in many cases. However, capillary separation techniques generally work at flow rates of the order of a few nL to a few hundred nL per minute, so the associated ionisation is microESI or nanoESI. The optimal low flow rates for



Fig. 10.16. Photograph of a MALDI sample plate able to take 100 deposits. A few crystal samples are visible

miniaturised sources ($1\text{--}10\mu\text{L}/\text{min}$) and the gain in sensitivity provided by columns with small diameters have stimulated the development of small systems.

The MALDI source was for a long time less favoured in LC-MS/MS combinations, in particular because it uses solid state samples (see Fig. 10.16), so there is no possibility of direct online analysis. Precision automation had to be designed to deposit the eluate from the separation column. MS/MS analysis is achieved by MALDI-IT [105], MALDI-Q-TOF [106–108], or MALDI-TOF-TOF [109], which sometimes require incorporation of a ‘home-made’ source [106, 108]. This type of MALDI-MS/MS interface has only been commercially available for a couple of years now.

Considering a one-hour elution window for an LC separation, about 60 deposits containing 1 min of eluate, i.e., about 200 nL, are formed. A high frequency of laser pulses is required for the LC-MALDI-MS/MS approach to compete with combined LC-ESI-MS/MS in terms of analysis rate. Hence, software had to be (and remains to be) developed to combine the data resulting from each deposit. Despite the particular requirements that have been fulfilled by recent technical developments, the LC-MALDI-MS/MS combination has a certain number of unequalled advantages, as described in a series of studies in 2003 [110–113].

A column with greater diameter can be used, so that more sample can be loaded. The peptide concentration is then obtained on the MALDI plate [111]. Higher mass peptides can be analysed than in ESI-MS/MS [110]. The sample can be further processed on the MALDI plate, e.g., enzymatic digestion to reveal post-translation modifications. The same sample can then be analysed before and after this treatment, to identify the peptides whose mass has changed.

The MALDI source allows one to decouple the MS/MS and MS analysis. This is advantageous for comparative analysis. It is thereby possible to select specifically the peptides detected at fragmentation with different abundances in the two samples being compared [114].

Ions giving weaker signals can be selected, and the fragmentation period can be increased, to the detriment only of the analysis time. In contrast, the LC-ESI-MS/MS combination preferentially fragments ions giving stronger signals. Moreover, the duration of MS/MS is necessarily limited, owing to the continuous elution of the peptides. It should therefore be possible to obtain MS/MS spectra with a better signal-to-noise ratio from weakly detected peptides in MALDI-MS/MS.

With this off-line MALDI-MS/MS combination, it is actually possible to 'stop' the flow of time: a sample giving unsatisfactory results can be analysed again, after further (chemical or enzymatic) treatment of the deposit, or simply by insisting on the fragmentation of certain peptides.

Moseley and coworkers analysed the product of tryptic digestion of bovine mitochondrial ribosomes. They used the LC-MS/MS approach with both an ESI source and a MALDI source. Only 63% of the peptides were characterised in common in the two cases, whence each analytical strategy brings specific information [115]. In addition, different types of ion fragment were observed with the two sources: the MS/MS spectra provided by each combination could be complementary [116]. However, it may be that the singly-charge species generated in the MALDI source are more difficult to fragment than the multiply charged species formed in ESI.

10.5.2 Coupling with Electrophoretic Techniques

The various techniques of electrophoresis have also been combined with ESI or MALDI mass spectrometry [117]. CE-ESI-MS interfaces are mainly based on previously developed LC-MS interfaces, with the further constraint of continuity of the electrical circuit.

The coaxial liquid sheath interface, with three concentric capillaries, is most commonly used [118]. The central capillary, used for the CE separation, is inserted in the region of the ESI source at atmospheric pressure by means of a metal tube. This tube, positioned concentrically around the capillary, serves as the ES needle and delivers the sheath liquid which closes the electrical circuit between the CE buffer and the ES needle. A third concentric tube made of steel carries the spray gas. The sheath liquid, with a flow of a few $\mu\text{L}/\text{min}$, contains a volatile solvent such as methanol to encourage nebulisation. A compromise must be made when choosing the electrolyte concentration, between efficiency of separation (favoured by high ionic strengths) and efficiency of nebulisation (favoured by low salt concentrations). One variant here consists in replacing the sheath liquid by a complementary liquid using a T junction. Nanospray interfaces have also been developed in which this sheath liquid is not necessary.

A complementary liquid can nevertheless be inserted in parallel to stabilise nebulisation + [119].

The complexity of clinical samples has been tackled using combined CE–MS. Human urine and serum have been analysed to identify diagnostic markers of kidney disease [120], diabetes, or hyperplasia of the prostate gland [121].

The three types of interface developed for CE–MS have also been used to combine capillary electrochromatography (CEC) with MS. Owing to the effects of dilution, the sensitivity of the coaxial interface is 20–40 times lower than that of the nanospray [122]. The main applications have been to the analysis of model peptides and lysates of proteins.

Capillary isoelectric focusing (CIEF) coupled online with MS is considered to be a genuine 2D separation system and an alternative to 2D gels. Many articles have been published. In particular, the separating ability of CIEF–ESI–FTICR–MS has been demonstrated by analysing lysates of *Escherichia coli* and *Deinococcus radiodurans* cells [123].

Although they are not online combinations, CE–MALDI associations have also been implemented. Work has been done to automate the deposition of analytes on the MALDI plate after CE separation, with applications to standard proteins, protein and cell lysates, and so on [124].

10.6 Nanotechnology for the MS Interface

Since the best sensitivity of the nanoelectrospray source is obtained at flow rates of a few tens of nL/min [125], the connections between the different elements of the chromatography and MS devices should limit unnecessary volumes as far as possible to avoid broadening the peaks downstream of the LC column. The small volumes used here are sensitive to leakage and blockages. This is why efforts have been made to integrate various stages upstream of the electrospray ionisation into microfluidic systems. The technology of microelectromechanical systems (MEMS) can be used to fabricate microfluidic channels with inner diameters in the range 10–100 μm . These were originally applied to on-chip electrophoretic separations. Of the two types of ionisation (nanoESI and MALDI), nanoESI is the most commonly used interface because it works in solution, making it more easily compatible with a microfluidic system. Many solutions have been proposed for this combination since 1997 [126], with the first work by Ramsey [127] and Karger [128] on glass substrates. However, the electrospray needle was not well suited in terms of profile, and also because the materials used (quartz or glass) were too hydrophilic. A new design then used polycarbonate, better suited to the problem of wetting [129], but less efficient for separation. A hybrid solution was presented by Aebersold and coworkers [130], in which a capillary for the electrospray was stuck on top of glass microfluidic chip. But one of the main problems was the alignment of the capillary with the microfluidic channels. The following solutions gradually

integrated the spray needle into the microfluidic circuit, on silica or polymer substrates, and the result was then commercialised.

10.6.1 Microfluidic Chip Associating Chromatography and Nanospray Tip

Typical systems associating nanoLC with MS generally incorporate a sample enrichment–desalting column able to preconcentrate large volumes of the sample in saline solutions. A microfluidic system associating these different elements together with a nanoelectrospray tip in a single chip have been commercialised by Agilent. They can be used to analyse a peptide mixture with a TOF or ion trap device, with a detection sensitivity better than the femtomole [131].

The central feature of this new technology is a reusable polymer microfluidic chip. Smaller than a credit card, this chip integrates sample enrichment and separation columns of a nanoflow LC system with the various connections and a nanoelectrospray tip. This technology removes the need for half of the connections usually required in a nanoflow LC–MS system, thereby considerably reducing the risk of leakage or dead volume and facilitating the use of this type of analysis.

The other element of the system is the interface which can be mounted on a TOF or ion trap. This interface allows for rapid and accurate insertion of the chip, giving it an optimal position with regard to the mass spectrometer.

This chip is fabricated by laminating polyimide films, with laser-etched channels, outlets, and frits. The enrichment and separation columns are filled with the reversed phase of standard chromatography. The chip is sandwiched in a valve for loading and separating the sample.

20 fmol of lysates from trypsin-digested BSA were analysed, with reproducible separation at flow rates in the range 100–400 nL/min. Two peptides of the BSA were identified by comparison with a database after MS/MS and injection of 600 amol of lysate [122].

The online integration of 2D chromatography has also been achieved by inserting an ion exchange column directly onto the valve. This was used to investigate the complexity and diversity of the plasma proteome [132].

10.6.2 Nanospray Tip Array Chip

Another type of nanoESI infusion system was designed by Zhang et al. [133] and subsequently commercialised by Advion [134]. The support is made by deep reactive ion etching (DRIE) on a silica substrate, with a channel through it and a tip (10 μm ID $\times$ 30 μm OD) connected to one face. A reservoir on the opposite face can contain the liquid to be analysed by direct infusion, and the voltage for electrospray formation is switched on by a liquid junction at the rear of the chip. This system can also be connected to a liquid phase separation with very low dead volume. The device was subsequently adapted

to multichannel chips with 96, 384, or 1,536 parallel nanoelectrospray tips [135]. This solution was commercialised to carry out high flow analysis in a miniaturised format and at low cost, for specialised analytical platforms. It can be adapted to most commercially available mass spectrometers.

References

1. Hsieh, S., Dreisewerd, K., van der Schors, R.C., Jimenez, C.R., Stahl-Zeng, J., Hillenkamp, F., Jorgenson, J.W., Geraerts, W.P., Li, K.W.: Separation and identification of peptides in single neurons by microcolumn liquid chromatography–matrix-assisted laser desorption/ionization time-of-flight mass spectrometry and postsource decay analysis, *Anal. Chem.* **70** (9), 1847–1852 (1998)
2. Redeker, V., Toullec, J.Y., Vinh, J., Rossier, J., Soyez, D.: Combination of peptide profiling by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry and immunodetection on single glands or cells, *Anal. Chem.* **70** (9) 1805–1811 (1998)
3. Barber, M.: *Biochem. Biophys. Res. Commun.* **18**, 469–473 (1965)
4. Torgerson, D.F., Skowronski, R.P., Macfarlane, R.D.: New approach to the mass spectrometry of non-volatile compounds, *Biochem. Biophys. Res. Commun.* **60**, 616–621 (1974)
5. Macfarlane, R.D., Torgerson, D.F.: Californium-252 plasma desorption mass spectrometry, *Science* **191**, 920–925 (1976)
6. Hålekaussön, P., Kamenski, I., Sundquist, B., Fohlman, J., Perterson P., McNeal, C.J., Macfarlane, R.D.: 127-I plasma desorption mass spectrometry of insulin, *J. Am. Chem. Soc.* **104**, 2948–2949 (1982)
7. Benninghoven, A., Jaspers, D., Sichtermann, W.: Secondary-ion emission of amino acids, *Appl. Phys.* **11**, 35–39 (1976)
8. Barber, M.: Fast atom bombardment of solids (FAB): A new ion source for mass spectrometry, *J. Chem. Soc. Chem. Comm.* **1981**, 325–327 (1981)
9. Barber, M., Bordoli, R.S., Elliot, G.J., Sedgwick, R.D., Tyler, A.N., Green, B.N.: Fast atom bombardment mass spectrometry of bovine insulin and other large peptides, *J. Chem. Soc. Chem. Comm.* **1982**, 936–938 (1982)
10. Dell, A., Morris, H.R.: Fast atom bombardment–high field magnetic mass spectrometry of 6000 Da polypeptides, *Biochem. Biophys. Res. Commun.* **106**, 1456–1461 (1982)
11. Chait, B.T., Kent, S.B.: Weighing naked proteins: Practical, high-accuracy mass measurement of peptides and proteins, *Science* **257** (5078), 1885–1894 (1992)
12. Roepstorff, P.: Mass spectrometry in the analysis of peptides and proteins, past and present. In: *Protein and Peptide Analysis by Mass Spectrometry*, ed. by J.R. Chapman, Humana Press, Totowa, NJ (1996) pp. 1–7
13. Cole, R.B.: Some tenets pertaining to electrospray ionization mass spectrometry, *J. Mass Spectrom.* **35** (7), 763–772 (2000)
14. Cech, N.B., Enke, C.G.: Practical implications of some recent studies in electrospray ionization fundamentals, *Mass Spectrom. Rev.* **20** (6), 362–387 (2001)

15. Whitehouse, C.M., Dreyer, R.N., Yamashita, M., Fenn, J.B.: Electrospray interface for liquid chromatographs and mass spectrometers, *Anal. Chem.* **57**, 675 (1985)
16. Gaskell, S.J.: Electrospray: Principles and practices, *J. Mass Spectrom.* **32**, 677–688 (1997)
17. Gallagher, R.T., Chapman, J.R., Mann, M.: Design and performance of an electrospray ion source for a doubly-focusing magnetic sector mass spectrometer, *Rapid Commun. Mass Spectrom.* **4**, 369–372 (1990)
18. Chapman, J.R., Gallagher, R.T., Barton, E.C., Curtis, J.M., Derrick, P.J.: Advantages of high-resolution and high-mass range magnetic-sector mass spectrometry for electrospray ionization, *Org. Mass Spectrom.* **27**, 195–203 (1992)
19. Niessen, W.M.: Advances in instrumentation in liquid chromatography–mass spectrometry and related liquid-introduction techniques, *J. Chromatogr. A* **794** (1–2), 407–435 (1998)
20. Arnott, D., Henzel, W.J., Stults, J.T.: Rapid identification of comigrating gel-isolated proteins by ion trap mass spectrometry, *Electrophoresis* **19** (6), 968–980 (1998)
21. Qin, J., Herring, C.J., Zhang, X.: De novo peptide sequencing in an ion trap mass spectrometer with ^{18}O labeling, *Rapid Commun. Mass Spectrom.* **12** (5), 209–216 (1998)
22. McLuckey, S.A., Van Berkel, G.J., Goeringer, D.E., Glish, G.L.: Ion trap mass spectrometry. Using high-pressure ionization, *Anal. Chem.* **66** (14), 737A–743A (1994)
23. Henion, J., Wachs, T., Mordehai, A.: Recent developments in electrospray mass spectrometry including implementation on an ion trap, *J. Pharm. Biomed. Anal.* **11** (11–12), 1049–1061 (1993)
24. Amster, I.J.: Fourier transform mass spectrometry, *J. Mass Spectrom.* **31**, 1325–1337 (1996)
25. Bruce, J.E., Smith, V.F., Liu, C., Randall, L.L., Smith, R.D.: The observation of chaperone–ligand noncovalent complexes with electrospray ionization mass spectrometry, *Protein Sci.* **7** (5), 1180–1185 (1998)
26. Cassady, C.J., Carr, S.R.: Elucidation of isomeric structures for ubiquitin $[\text{M} + 12\text{H}]^{12+}$ ions produced by electrospray ionization mass spectrometry, *J. Mass Spectrom.* **31** (3), 247–254 (1996)
27. McLafferty, F.W., Senko, M.W., Little, D.P., Wood, T.D., O'Connor, P.B., Speir, J.P., Chorush, R.A., Kelleher, N.L.: Electrospray ionization and Fourier-transform mass spectrometry in biomedical research. In: *Proceedings of the 13th International MS Conference*, Budapest, Hungary, John Wiley (1994)
28. Solouki, T., Reinhold, B.B., Costello, C.E., O'Malley, M., Guan, S., Marshall, A.G.: Electrospray ionization and matrix-assisted laser desorption/ionization Fourier transform ion cyclotron resonance mass spectrometry of permethylated oligosaccharides, *Anal. Chem.* **70** (5), 857–864 (1998)
29. Tolic, L.P., Bruce, J.E., Lei, Q.P., Anderson, G.A., Smith, R.D.: In-trap cleanup of proteins from electrospray ionization using soft sustained off-resonance irradiation with Fourier transform ion cyclotron resonance mass spectrometry, *Anal. Chem.* **70** (2), 405–408 (1998)
30. Yang, L., Lee, C.S., Hofstadler, S.A., Pasa-Tolic, L., Smith, R.D.: Capillary isoelectric focusing electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry for protein characterization, *Anal. Chem.* **70**, 3235–3241 (1998)

31. Verentchikov, A.N., Standing, K.G.: Reflecting time-of-flight mass spectrometry with an electrospray ion source and orthogonal extraction, *Anal. Chem.* **66**, 126–133 (1994)
32. Mirgorodskaya, O.A., Shevchenko, A.A., Chernushevich, I.V., Dodonov, A.F., Miroshnikov, A.I.: Electrospray ionization time-of-flight mass spectrometry in protein chemistry, *Anal. Chem.* **66**, 99–107 (1994)
33. Burdick, D.J., Stults, J.T.: Analysis of peptide synthesis products by electrospray ionization mass spectrometry, *Methods Enzymol.* **289**, 499–519 (1997)
34. Figeys, D., Lock, C., Taylor, L., Aebersold, R.: Microfabricated device coupled with an electrospray ionization quadrupole time-of-flight mass spectrometer: Protein identifications based on enhanced-resolution mass spectrometry and tandem mass spectrometry data, *Rapid Commun. Mass Spectrom.* **12** (20), 1435–1444 (1998)
35. Olling, A., Breimer, M.E., Peltomaa, E., Samuelsson, B.E., Ghardashkhani, S.: Electrospray ionization and collision-induced dissociation time-of-flight mass spectrometry of neutral glycosphingolipids, *Rapid Commun. Mass Spectrom.* **12** (10), 637–645 (1998)
36. Shevchenko, A., Chernushevich, I., Ens, W., Standing, K.G., Thomson, B., Wilm, M., Mann, M.: Rapid de novo peptide sequencing by a combination of nanoelectrospray, isotopic labeling and a quadrupole/time-of-flight mass spectrometer, *Rapid Commun. Mass Spectrom.* **11** (9), 1015–1024 (1997)
37. Morris, H.R., Paxton, T., Dell, A., Langhorne, J., Berg, M., Borboli, R.S., Hoyes, J., Bateman, R.H.: High sensitivity collisionally-activated decomposition tandem mass spectrometry on a novel quadrupole/orthogonal-acceleration time-of-flight mass spectrometer, *Rapid Commun. Mass Spectrom.* **10**, 889–896 (1996)
38. Smith, R.D., Light-Wahl, K.J., Winger, B.E., Goodlett, D.R.: Electrospray ionization. In: *Biological Mass Spectrometry: Present and Future*, ed. by T. Matsuo et al., John Wiley and Sons, Baffinslane, Chichester UK (1995) pp. 41–74
39. Emmett, M.R., Caprioli, R.M.: Micro-electrospray mass spectrometry: Ultra-high-sensitivity analysis of peptides and proteins, *J. Am. Soc. Mass Spectrom.* **5**, 605–613 (1994)
40. Wilm, M., Mann, M.: Analytical properties of the nanoelectrospray ion source, *Anal. Chem.* **68** (1), 1–8 (1996)
41. Wilm, M.S., Mann, M.: Electrospray and Taylor-cone theory, Dole's beam of macromolecules at last?, *Int. J. Mass Spectrom. Ion Processes* **136**, 167–180 (1994)
42. Valaskovic, G.A., Kelleher, N.L., Little, D.P., Aaserud, D.J., McLafferty, F.W.: Attomole-sensitivity electrospray source for large-molecule mass spectrometry, *Anal. Chem.* **67** (20), 3802–3805 (1995)
43. Smith, R.D., Walh, J.H., Goodlett, D.R., Hofstadler, S.A.: *Anal. Chem.* **65**, 574A–584A (1993)
44. Wahl, J.H., Goodlett, D.R., Udseth, H.R., Smith, R.D.: *Anal. Chem.* **64**, 3194–3196 (1992)
45. Andren, P.E., Emmett, M.R., DaGue, B.B., Steulet, A.F., Waldmeier, P., Caprioli, R.M.: Blood–brain barrier penetration of 3-aminopropyl-*n*-butylphosphinic acid (CGP 36742) in rat brain by microdialysis/mass spectrometry, *J. Mass Spectrom.* **33** (3), 281–287 (1998)

46. Wilm, M., Shevchenko, A., Houthaeve, T., Breit, S., Schweigerer, L., Fotsis, T., Mann, M.: Femtomole sequencing of proteins from polyacrylamide gels by nano-electrospray mass spectrometry, *Nature* **379** (6564), 466–469 (1996)
47. Valaskovic, G.A., Kelleher, N.L., McLafferty, F.W.: Attomole protein characterization by capillary electrophoresis mass spectrometry, *Science* **273** (5279), 1199–1202 (1996)
48. Davis, M.T., Stahl, D.C., Hefta, S.A., Lee, T.D.: A microscale electrospray interface for on-line, capillary liquid chromatography/tandem mass spectrometry of complex peptide mixtures, *Anal. Chem.* **67** (24), 4549–4556 (1995)
49. Smith, R.D., Loo, J.A., Edmonds, C.G., Barinaga, C.J., Udseth, H.R.: New developments in biochemical mass spectrometry: Electrospray ionization, *Anal. Chem.* **62** (9), 882–899 (1990)
50. Juraschek, R., Dulcks, T., Karas, M.: Nanoelectrospray: More than just a minimized-flow electrospray ionization source, *J. Am. Soc. Mass Spectrom.* **10** (4), 300–308 (1999)
51. Karas, M., Hillenkamp, F.: Laser desorption ionization of proteins with molecular masses exceeding 10,000 dalton, *Anal. Chem.* **60**, 2299–2301 (1988)
52. Posthumus, M.A., Kistemaker, P.G., Menzelaar, H.L.C.: *Anal. Chem.* **60**, 985 (1978)
53. Hillenkamp, F.: Laser induced ion formation from organic solids. In: *Ion Formation from Organic Solids*, ed. by A. Benninghoven, Springer, Berlin Heidelberg New York (1983) pp.190–205
54. Tanaka, K., Waki, H., Ido, Y., Akita, S., Yoshida, Y., Yoshida, T.: Protein and polymer analysis up to m/z 100,000 by laser ionization time-of-flight mass spectrometry, *Rapid Commun. Mass Spectrom.* **2**, 151–153 (1988)
55. Karas, M., Bachmann, D., Bahr, U., Hillenkamp, F.: Matrix-assisted laser desorption of non-volatile compounds, *Int. J. Mass Spectrom. Ion Processes* **78**, 53–68 (1987)
56. Karas, M., Bahr, U., Hillenkamp, F.: UV laser matrix desorption/ionization mass spectrometry of proteins in the 100,000 dalton range, *Int. J. Mass Spectrom. Ion Processes* **92**, 231–242 (1989)
57. Karas, M., Bahr, U., Ingendoh, A., Nordhoff, E., Stahl, B., Strupat, K., Hillenkamp, F.: Principles and applications of matrix-assisted UV-laser desorption/ionization mass spectrometry, *Anal. Chim. Acta* **241**, 175–185 (1990)
58. Cotter, R.J.: *Time of Flight Mass Spectrometry: Instrumentation and Applications to Biological Research*, A.C. Society, Washington D.C. (1997)
59. Hillenkamp, F., Karas, M., Beavis, R.C., Chait, B.T.: Matrix-assisted laser desorption/ionization mass spectrometry of biopolymers, *Anal. Chem.* **63**, 1193A–1202A (1991)
60. Fenselau, C.: MALDI MS and strategies for protein analysis, *Anal. Chem.* **69** (21), 661A–665A (1997)
61. Wong, C.K.L., So, M.P., Chan, T.-W.D.: Origins of the proton in the generation of protonated polymers and peptides in matrix-assisted laser desorption/ionization, *Eur. J. Mass Spectrom.* **4**, 223–232 (1998)
62. Bencsura, A., Navale, V., Sadeghi, M., Vertes, A.: Matrix-guest energy transfer in matrix-assisted laser desorption, *Rapid Commun. Mass Spectrom.* **11**, 679–682 (1997)
63. Carroll, J.A., Beavis, R.C.: Matrix-assisted laser desorption and ionization. In: *Laser Desorption and Ablation*, ed. by J.C. Miller and R.F. Haglund Jr., *Experimental Methods in Physical Sciences* (1999) Chap. 7

64. Li, Y., Hunter, R.L., McIver Jr., R.T.: High resolution mass spectrometer for protein chemistry, *Nature* **370** (4 August), 393–395 (1994)
65. McIver, R.T., Li, Y., Hunter, R.L.: High resolution laser desorption mass spectrometry of peptides and proteins, *Proc. Natl. Acad. Sci. USA* **91**, 4801–4805 (1994)
66. Solouki, T., Marto, J.A., White, M.W., Guan, S., Marshall, A.G.: Attomole biomolecule mass analysis by matrix-assisted laser desorption/ionization Fourier transform ion cyclotron resonance, *Anal. Chem.* **67**, 4139–4144 (1995)
67. Sheng, L.S., Covey, J.E., Shew, S.L., Winger, B.E., Campana, J.E.: Matrix-assisted laser desorption ionization Fourier-transform mass spectrometry, *Rapid Commun. Mass Spectrom.* **8**, 498–500 (1994)
68. Castoro, J.A., Wilkins, C.L., Woods, A.S., Cotter, R.J.: Peptide amino sequence analysis using matrix-assisted laser desorption/ionization and Fourier transform mass spectrometry, *Biol. Mass Spectrom.* **30**, 94–98 (1995)
69. Jonscher, K.R., Yates, J.R.: Matrix-assisted laser desorption ionization quadrupole ion trap mass spectrometry of peptides, *J. Biol. Chem.* **272** (3 January 1997), 1735–1741 (1997)
70. Schwartz, J.C., Bier, M.E.: Matrix-assisted laser desorption of peptides and proteins using a quadrupole ion trap mass spectrometer, *Rapid Commun. Mass Spectrom.* **7**, 27–32 (1993)
71. Doroshenko, V.M., Cotter, R.J.: A quadrupole ion trap/time-of-flight mass spectrometer with a parabolic reflectron, *J. Mass Spectrom.* **33** (4), 305–318 (1998)
72. Chambers, D.M., Goeringer, D.E., McClukey, S.A., Glish, G.L.: Matrix-assisted laser desorption of biological molecules in the quadrupole ion trap mass spectrometer, *Anal. Chem.* **65**, 14–20 (1993)
73. Bordoli, R.S., Howes, K., Vickers, R.G., Bateman, R.H., Harvey, D.J.: Matrix-assisted laser desorption mass spectrometry on a magnetic sector instrument fitted with an array detector, *Rapid Commun. Mass Spectrom.* **8**, 585–589 (1994)
74. Hill, J.A., Annan, R.S., Biemann, K.: Matrix-assisted laser desorption ionization with a magnetic mass spectrometer, *Rapid Commun. Mass Spectrom.* **5**, 395–399 (1991)
75. Castaing, R., Slodzian, G.: Microanalyse par émission ionique secondaire, *J. Microsc.* **1**, 395–410 (1962)
76. Chandra, S., Smith, D.R., Morrison, G.H.: Subcellular imaging by dynamic SIMS ion microscopy, *Anal. Chem.* **72** (3), 104A–114A (2000)
77. Tsien, R.Y.: Fluorescent probes of cell signaling, *Ann. Rev. Neurosci.* **12**, 227–254 (1989)
78. Miyawaki, A., et al.: Fluorescent indicators for Ca^{2+} based on green fluorescent proteins and calmodulin, *Nature* **388**, 882–887 (1997)
79. Yeung, E.S.: Following single cell dynamics with native fluorescence microscopy, *Anal. Chem.* **71**, 522A–529A (1999)
80. Maiti, S., Shear, J.B., Williams, R.M., Zipfel, W.R., Webb, W.W.: Measuring serotonin distribution in live cells with three-photon excitation, *Science* **275**, 530–532 (1997)
81. Somlyo, A.P., Bond, M., Somlyo, A.V.: Calcium content of mitochondria and endoplasmic reticulum in liver frozen rapidly in vivo, *Nature* **314**, 622–625 (1985)

82. Buchanan, R.A., Leapman, R.D., O'Connell, M.F., Reese, T.S., Andrews, S.B.: Quantitative scanning transmission electron microscopy of ultrathin cryosections: Subcellular organelles in rapidly frozen liver and cerebellar cortex, *J. Struct. Biol.* **110**, 245–255 (1993)
83. McArthur, S.L., Vendettuoli, M.C., Ratner, B.D., Castner, D.G.: Methods for generating protein molecular ions in ToF-SIMS, *Langmuir* **20**, 3704–3709 (2004)
84. Levi-Setti, R., Chabala, J.M., Wang, Y.-L.: Aspects of high resolution imaging with a scanning ion microprobe, *Ultramicroscopy* **24**, 97–114 (1988)
85. Slodzian, G., Daigne, B., Girard, F., Boust, F., Hillion, F.: Scanning secondary ion analytical microscopy with parallel detection, *Biol. Cell.* **74**, 43–50 (1992)
86. Vos, J.W., Hepler, P.K.: Calmodulin is uniformly distributed during cell division in living stamen hair cells of *Tradescantia virginiana*, *Protoplasma* **201**, 158–171 (1998)
87. Chandra, S., Morrison, G.H., Wolcott, C.C.: Imaging intracellular elemental distribution and ion fluxes in cultured cells using ion microscopy: A freeze-fracture methodology, *J. Microsc. (Oxford)* **144**, 15–37 (1986)
88. Ausserer, W.A., Chandra, S., Morrison, G.H.: Morphological and elemental integrity of freeze-fractured, freeze-dried cultured cells during ion microscopic analysis, *J. Microsc. (Oxford)* **154**, 39–57 (1989)
89. Chandra, S., Morrison, G.H.: Sample preparation of animal tissues and cell cultures for secondary ion mass spectrometry (SIMS) microscopy, *Biol. Cell* **74**, 31–42 (1992)
90. Pacholski, M.L., Cannon Jr., D.M., Ewing, A.G., Winograd, N.: Static time-of-flight secondary ion mass spectrometry imaging of freeze-fractured, frozen-hydrated biological membranes, *Rapid Commun. Mass Spectrom.* **12**, 1232–1235 (1998)
91. Sod, E.W., Crooker, A.R., Morrison, G.H.: Biological cryosection preparation and practical ion yield evaluation for ion microscopic analysis, *J. Microsc. (Oxford)* **160**, 55–65 (1990)
92. Touboul, D., Halgand, F., Brunelle, A., Kersting, R., Tallarek, E., Hagenhoff, B., Laprévote, O.: Tissue molecular ion imaging by gold cluster ion bombardment, *Anal. Chem.* **76**, 1550–1559 (2004)
93. Touboul, D., Brunelle, A., Halgand, F., De La Porte, S., Laprevote, O.: Lipid imaging by gold cluster time-of-flight secondary ion mass spectrometry: Application to Duchenne muscular dystrophy, *J. Lipid Res.* **46**, 1388–1395 (2005)
94. Comisarow, M.B., Marshall, A.G.: Fourier transform ion cyclotron resonance spectroscopy, *Chem. Phys. Lett.* **25**, 282–283 (1974)
95. Marshall, A.G., Comisarow, M.B.: Frequency-sweep Fourier transform ion cyclotron resonance spectroscopy, *Chem. Phys. Lett.* **26**, 489–490 (1974)
96. Senko, M.W., et al.: Electrospray ionization Fourier transform ion cyclotron resonance at 9.4 T, *Rapid. Comm. Mass Spectrom.* **10**, 1824–1828 (1996)
97. Hofstadler, S.A., et al.: Analysis of single cells with capillary electrophoresis electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry, *Rapid Comm. Mass Spectrom.* **10**, 919–922 (1996)
98. Alomary, A., Anderson, T., Kahl, S., Solouki, T.: Development of mass spectral methods for direct and low level analysis of environmental pollutants, 363, *46th Amer. Soc. for Mass Spectrom. Annual Conf. on Mass Spectrometry and Allied Topics*, Orlando, FL (May 1998)

99. Molnar, I., Horvath, C.: Separation of amino acids and peptides on non-polar stationary phases by high-performance liquid chromatography, *J. Chromatogr.* **142**, 623–640 (1977)
100. Mönch, W., Dehnen, W.: High-performance liquid chromatography of peptides, *J. Chromatogr.* **140**, 260–262 (1977)
101. O'Hare, M.J., Nice, E.C.: Hydrophobic high-performance liquid chromatography of hormonal polypeptides and proteins on alkylsilane-bonded silica, *J. Chromatogr.* **171**, 209–226 (1979)
102. Mönch, W., Dehnen, W.: High-performance liquid chromatography of polypeptides and proteins on a reversed-phase support, *J. Chromatogr.* **147**, 415–418 (1978)
103. Hancock, W.S., Bishop, C.A., Prestidge, R.L., Harding, D.R., Hearn, M.T.: Reversed-phase, high-pressure liquid chromatography of peptides and proteins with ion-pairing reagents, *Science* **200** (4346), 1168–1170 (1978)
104. Abian, J., Oosterkamp, A.J., Gelpi, E.: Comparison of conventional, narrow-bore and capillary chromatography/mass spectrometry for electrospray ionization mass spectrometry: Practical considerations, *J. Mass Spectrom.* **34**, 244–254 (1999)
105. Krutchinsky, A.N., Kalkum, M., Chait, B.T.: Automatic identification of proteins with a MALDI-quadrupole ion trap mass spectrometer, *Anal. Chem.* **73** (21), 5066–5077 (2001)
106. Krutchinsky, A.N., Zhang, W., Chait, B.T.: Rapidly switchable matrix-assisted laser desorption/ionization and electrospray quadrupole–time-of-flight mass spectrometry for protein identification, *J. Am. Soc. Mass Spectrom.* **11** (6), 493–504 (2000)
107. Griffin, T.J., Gygi, S.P., Rist, B., Aebersold, R., Loboda, A., Jilkine, A., Ens, W., Standing, K.G.: Quantitative proteomic analysis using a MALDI quadrupole time-of-flight mass spectrometer, *Anal. Chem.* **73** (5), 978–986 (2001)
108. Wattenberg, A., Organ, A.J., Schneider, K., Tyldesley, R., Bordoli, R., Bateman, R.H.: Sequence dependent fragmentation of peptides generated by MALDI quadrupole time-of-flight (MALDI Q-TOF) mass spectrometry and its implications for protein identification, *J. Am. Soc. Mass Spectrom.* **13** (7), 772–783 (2002)
109. Bienvenut, W.V., Deon, C., Pasquarello, C., Campbell, J.M., Sanchez, J.C., Vestal, M.L., Hochstrasser, D.F.: Matrix-assisted laser desorption/ionization-tandem mass spectrometry with high resolution and sensitivity for identification and characterization of proteins, *Proteomics* **2** (7), 868–876 (2002)
110. Griffin, T.J., Lock, C.M., Li, X.J., Patel, A., Chervetsova, I., Lee, H., Wright, M.E., Ranish, J.A., Chen, S.S., Aebersold, R.: Abundance ratio-dependent proteomic analysis by mass spectrometry, *Anal. Chem.* **75** (4), 867–874 (2003)
111. Griffin, T.J., Li, X.-J., Lock, C.M., Chervetsova, I., Marelli, M., Aitchison, J.D., Aebersold, R.: Complex protein mixture analysis by LC-MALDI-TOF hybrid mass spectrometry: Evaluating the performance and potential for high throughput proteomic analysis, *51st Conference on Mass Spectrometry and Allied Topics*, Poster WPX487 Montreal, Quebec, Canada (8–12 June 2003)
112. Bodnar, W.M., Blackburn, R.K., Krise, J.M., Moseley, M.A.: Exploiting the complementary nature of LC/MALDI/MS/MS and LC/ESI/MS/MS for increased proteome coverage, *J. Am. Soc. Mass Spectrom.* **14** (9), 971–979 (2003)

113. Hansen, A.M., Nielsen, P.F., Jensen, O.N., Krogh, T.N.: Differences and complementarities of LC-QTOF and LC-MALDI TOFTOF datasets, *51st Conference on Mass Spectrometry and Allied Topics*, Poster WPW446, Montreal, Quebec, Canada (8–12 June 2003)
114. Griffin, T.J., Lock, C.M., Li, X.J., Patel, A., Chervetsova, I., Lee, H., Wright, M.E., Ranish, J.A., Chen, S.S., Aebersold, R.: Abundance ratio-dependent proteomic analysis by mass spectrometry, *Anal. Chem.* **75** (4), 867–874 (2003)
115. Bodnar, W.M., Blackburn, R.K., Krise, J.M., Moseley, M.A.: Exploiting the complementary nature of LC/MALDI/MS/MS and LC/ESI/MS/MS for increased proteome coverage, *J. Am. Soc. Mass Spectrom.* **14** (9), 971–979 (2003)
116. Wattenberg, A., Organ, A.J., Schneider, K., Tyldesley, R., Bordoli, R., Bateman, R.H.: Sequence dependent fragmentation of peptides generated by MALDI quadrupole time-of-flight (MALDI QTOF) mass spectrometry and its implications for protein identification, *J. Am. Soc. Mass Spectrom.* **13** (7), 772–783 (2002)
117. Stutz, H.: Advances in the analysis of proteins and peptides by capillary electrophoresis with matrix-assisted laser desorption/ionization and electrospray mass spectrometry detection, *Electrophoresis* **26**, 1254–1290 (2005)
118. Ding, J., Vouros, P.: Advances in CE/MS, *Anal. Chem.* **71**, 378A–385A (1999)
119. Hsieh, F., Baronas, E., Muir, C., Martin, S.A.: A novel nanospray capillary zone electrophoresis/mass spectrometry interface, *Rapid Commun Mass Spectrom.* **13**, 67–72 (1999)
120. Wittke, S., Fliser, D., Haubitz, M., Bartel, S., Krebs, R., Hausadel, F., Hillman, M., Golovko, I., Koester, P., Haller, H., Kaiser, T., Mischak, H., Weissinger, E.: Determination of peptides and proteins in human urine with capillary electrophoresis mass spectrometry, a suitable tool for the establishment of new diagnostic markers, *J. Chromatogr. A* **1013**, 173–181 (2003)
121. Guzman, N.: Determination of immunoreactive gonadotropin-releasing hormone in serum and urine by on-line immunoaffinity capillary electrophoresis coupled to mass spectrometry, *J. Chromatogr. B* **749**, 197–213 (2000)
122. Barcelo-Barrachina, E., Moyano, E., Galceran, M.T.: State-of-the-art of the hyphenation of capillary electrochromatography with mass spectrometry, *Electrophoresis* **25**, 1927–1948 (2004)
123. Jensen, P.K., Pasa-Tolic, L., Peden, K.K., Martinovic, S., Lipton, M.S., Anderson, G.A., Tolic, N., Wong, K.K., Smith, R.D.: Mass spectrometric detection for capillary isoelectric focusing separations of complex protein mixtures, *Electrophoresis* **21**, 1372–1380 (2000)
124. Page, J.S., Rubakhin, S.S., Sweedler, J.V.: Single-neuron analysis using CE combined with MALDI MS and radionuclide detection, *Anal. Chem.* **74**, 497–503 (2002)
125. Schmidt, A., Karas, M., Dülcks, T.: Effect of different solution flow rates on analyte ion signals in nano-ESI MS, or: When does ESI turn into nano-ESI?, *J. Am. Soc. Mass Spectrom.* **14**, 492–500 (2003)
126. Sung, W.C., Makamba, H., Chen, S.H.: Chip-based microfluidic devices coupled with electrospray ionization mass spectrometry, *Electrophoresis* **26**, 1783–1791 (2005)
127. Ramsey, R.S., Ramsey, J.M.: Generating electrospray from microchip devices using electroosmotic pumping, *Anal. Chem.* **69**, 1174–1178 (1997)

128. Xue, Q., Foret, F., Dunayevskiy, Y.M., Zavracky, P.M., McGruer, N.E., Karger, B.L.: Multichannel microchip electrospray mass spectrometry, *Anal. Chem.* **69** (3), 426–430 (1997)
129. Wen, J., Lin, Y., Xiang, F., Matson, D.W., Udseth, H.R., Smith, R.D.: Microfabricated isoelectric focusing device for direct electrospray ionization mass spectrometry, *Electrophoresis* **21**, 191–197 (2000)
130. Figeys, D., Ning, Y., Aebersold, R.: A microfabricated device for rapid protein identification by microelectrospray ion trap mass spectrometry, *Anal. Chem.* **69** (16), 3153–3160 (1997)
131. Yin, H., Killeen, K., Brennen, R., Sobek, D., Werlich, M., van de Goor, T.: Microfluidic chip for peptide analysis with an integrated HPLC column, sample enrichment column, and nanoelectrospray tip, *Anal. Chem.*, **77**, 527–533 (2005)
132. Fortier, M.H., Bonneil, E., Goodley, P., Thibault, P.: Integrated microfluidic device for mass spectrometry-based proteomics and its application to biomarker discovery programs, *Anal. Chem.* **77**, 1631–1640 (2005)
133. Schultz, G.A., Corso, T.N., Prosser, S.J., Zhang, S.: A fully integrated monolithic microchip electrospray device for mass spectrometry, *Anal. Chem.* **72**, 4058–4063 (2000)
134. Zhang, S., Van Pelt, C.K., Henion, J.D.: Automated chip-based nanoelectrospray mass spectrometry for rapid identification of proteins separated by two-dimensional gel electrophoresis, *Electrophoresis* **24**, 3620–3632 (2003)
135. Dethy, J.M., Ackermann, B.L., Delatour, C., Henion, J.D., Schultz, G.A.: Demonstration of direct bioanalysis of drugs in plasma using nanoelectrospray infusion from a silicon chip coupled with tandem mass spectrometry, *Anal. Chem.* **75** (4), 805–811 (2003)

Electrical Characterisation and Dynamics of Transport

N. Picollet-D'Hahan, C. Amatore, S. Arbault, L. Thouin, A.-L. Biance,
G. Oukhaled, L. Auvray, J. Weber, N. Minc, and J.-L. Viovy

11.1 Ion Channels and the Patch-Clamp Technique

Four molecule channels, three ions, two pumps, and some ATP
are all it takes [...] to make an adjustable biological clock which
runs spontaneously and relentlessly.

J.-P. Changeux [1]

Ion channels lie at the interface between two areas of science: as physiological relays in biology and as picoampere/millivolt conductors in physics. It is precisely the complementarity of these two worlds which led to the emergence of the novel patch-clamp technique. By providing direct access to the activity of the protein channel, the patch clamp brought about a revolution in the molecular scale study of biomembranes.

This section discusses the experimental aspects of this electrophysiological technique, after a brief summary of the basic physical concepts, and illustrates the way ion channels are involved at the very heart of the main cell functions. Finally, we outline their privileged position as molecular targets of therapeutic importance in the treatment of channelopathies (ion channel disorders).

While the patch clamp is considered as the 'standard' for studying the electrical activity of cells, one should not forget the current technological developments that have sprung from it, combining performance and high analysis rates in the discovery of new molecules.

Throughout this section, the reader will be pointed toward the literature concerning channel biology, the patch-clamp method, and the emergence of parallel electrophysiological techniques.

11.1.1 What Is an Ion Channel?

When we remember something, read a book, raise our hand, etc., all these actions are the final result of complex electrical activity governed by billions

of nerve cells. These cells communicate by means of electrical currents which enter and exit them thanks to the essential privileged structures we call ion channels.

A century ago, it was already suspected that such tiny openings must exist! But it was not until about twenty years ago that scientists were able to demonstrate that these orifices or channels were multipurpose entities made from proteins. When ions move along these channels (the word 'ion' comes from the Greek *ienai* which means 'to move'), this ion flow generates an electrical current. Without the help of molecular biology, the channel would have remained a model for a membrane pore characterised by a selective filter and activation or inactivation gates. As we shall see in this section, it was techniques for the extraction of messenger RNA and the heterologous expression of proteins which, combined with the patch-clamp technique, made it possible to determine in particular the amino acid structure and the interaction sites of the channel openers and inhibitors, and to better understand the mechanisms underlying ion selectivity and permeation [2, 3].

To sum up, a question arises if one attempts to define the ion channel: From which standpoint should we view it? For the electrophysiologist, a channel is a resistance through which a current can be dissipated. For the cell biologist, a channel is a vector for ion flow. For the molecular biologist, a channel is a protein complex made up of subunits, themselves made up of domains and segments, fulfilling the role of gates and regulatory zones. It really is the convergence of different methodological approaches which makes it possible to investigate the protein channel from all these different angles and hence to better understand its structure and operation.

How Does an Ion Channel Work?

Every cell extracts the molecules and ions it needs from the extracellular fluid. This flow of molecules and ions takes place through the plasma membrane, e.g., glucose, Na^+ , Ca^{2+} , and, in eukaryotic cells, through the membranes of the intracellular compartments (nucleus, endoplasmic reticulum, mitochondria), e.g., mRNA, Ca^{2+} , ATP. While the lipid bilayer is permeable to water (by osmosis) and some small uncharged molecules such as O_2 and CO_2 which diffuse freely, note that it is not permeable to ions (cations such as K^+ , Na^+ , Ca^{2+} or anions such as Cl^- , HCO_3^-), small hydrophilic molecules (such as glucose), or macromolecules (such as proteins and RNA). In this section, we shall examine only the way that ions are transported through the cell membrane, i.e., facilitated diffusion. This takes place via proteins or assemblies of proteins 'built into' the membrane. These transmembrane proteins form channels filled with water, through which the ion moves in the direction of the electrochemical gradient. In other words, the conductance increases as the electrochemical gradient grows in the direction favouring ion transfer, from the more concentrated solution toward the less concentrated solution. These

channels which facilitate diffusion can be opened or closed, and are said to be gated.

Properties of the Channels

Biological membranes house a wide range of channels with different biophysical, physiological, and pharmacological properties.

Activation Mode

There are several categories of ion channel: ligand-gated, mechanically-gated, voltage-gated, and light-gated channels, and receptor channels coupled with a G protein [2, 4]:

- *Ligand-Gated Channels.* These channels open or close in response to the binding of small signalling molecules or ligands. Whether they be extracellular or intracellular, these ligands are never transported when the channel opens. Extracellular ligands bind to the extracellular face of the channel. An example are neurotransmitters like acetylcholine which open cationic channels and initiate a nerve impulse or muscle contraction. The binding of GABA¹ (gamma-aminobutyric acid) in some synapses inhibits the production of a nerve input. Intracellular ligands bind to the cytosolic face of the protein. Second messengers like cyclic AMP regulate channels involved in the initiation of nerve inputs in response to smells and light. ATP is also needed to open chloride channels and evacuate Cl⁻ or bicarbonate (HCO₃⁻) from the cell. Note that, even if energy released by ATP hydrolysis is needed to open the channel, there is no active transport since the ions diffuse in the direction of their concentration gradient. K_{ATP} channels, potassium channels inhibited by intracellular ATP, have also been described in beta pancreatic cells, cardiac, skeletal, and smooth muscles, and some neurons [5]. The probability of their opening, which is barely sensitive to the potential at all, depends on the ATP concentration on the inner face of the membrane. These channels help to maintain the membrane potential at negative values under pathological conditions (anoxia) in which the intracellular ATP concentration is abnormally low.
- *Mechanically-Gated Channels.* The mechanical deformation of cells, pressure, or stretching of the membrane are mechanical stimuli which can induce some channels to open. An example would be acoustic waves in the cells of the inner ear, which open channels inducing nerve pulses that the brain interprets in terms of sounds. The mechanisms underlying the action of these channels are not fully understood yet. It would seem that a pressure exerted on the membrane is transmitted to the channel either directly

¹ There are two main types of GABAergic receptor in the brain: (1) GABA_A sensitive to muscimol (agonistic) and picrotoxin (antagonistic). When GABA binds to its recognition site, it opens a chlorine channel which, allowing Cl⁻ ions through, causes hyperpolarisation of the target cell. (2) GABA_B, discussed further below.

by the membrane lipids, or indirectly via elements of the cytoskeleton or extracellular elements [2].

- *Voltage-Gated Channels.* In the so-called excitable cells, such as neurons or muscle cells, some channels open or close in response to changes in the charge (measured in volts) across the plasma membrane. Among the channels that are sensitive to the potential in this way, one should mention the outward-rectifying K^+ channels, whose conductance increases with the depolarisation (see the note on sign conventions and definitions on p. 650), and Na^+ and Ca^{2+} channels (T or L channels) in excitable cells. In these channels there are one- or two-gate systems sensitive to the potential, which control channel opening. These gates are in fact electrical charges carried by the amino acids forming the channel [6].
- *Light-Gated Channels.* Micro-organisms like the green alga *Chlamydomonas* often develop by phototaxis. In this case, the photoreceptors are (7-helix) transmembrane proteins, ion channels called channelopsin or channelrhodopsin (ChR1 and ChR2) [7], which are sensitive to light and selectively allow through H^+ ions (ChR1) or cations (ChR2) [8]. Most opsins are actually G proteins, i.e., proteins binding GTP, coupled to receptors which indirectly open the channel. Unlike these opsins, ChR2 is unique because it contains the channel itself. The heterologous expression of this channel can be very useful for manipulating the intracellular pH or the membrane potential, simply by illumination.
- *Receptors Coupled to a G Protein.* In some cases, receptor and channel are two different proteins separated from one another. Coupling occurs via a G protein which modulates the channel activity directly or through a second messenger [2]. An example is the $GABA_B$ receptor. The G protein is itself coupled to a Ca^{2+} channel. In this case, it causes a reduction in the calcium currents and hence a reduction in the release of neurotransmitter in the nerve endings. Coupled to a K^+ channel, the G protein increases the conductance with regard to K^+ ions and thereby induces a hyperpolarisation of the post-synaptic neurons. The binding of GABA to the $GABA_A$ and $GABA_B$ receptors leads to inhibition of neurotransmission.

Note that channels selective with respect to Ca^{2+} , K^+ , and Cl^- ions are also present in the membranes of intracellular organelles such as the sarcoplasmic reticulum, the mitochondria, and the nucleus. These channels are said to be endocellular. Without being exhaustive, one could mention the receptors of inositol 1,4,5-triphosphate (InsP3) and the receptors of ryanodine (RyRs), which play a crucial role in Ca^{2+} -dependent functions such as the excitation–contraction coupling, fertilisation, or secretion.

It is to a large extent patch-clamp measurements that have led to an understanding of the finer properties of ion channels, demonstrating in particular that facilitated diffusion is an all-or-nothing phenomenon, in the sense that the channel is either open or closed.

Ion channels are characterised by their activation mode, as we have just seen. They are also described in terms of their ion specificity, kinetic behaviour, and conductance.

Selectivity

Most channels only let through a single category of ions. One speaks of ion selectivity or specificity. Potassium channels are thus more permeable to potassium ions than to sodium ions. In 1955, Hodgkin and Keynes showed that ions go through these channels in single file. The work of C. Miller in the 1980s demonstrated the existence of 4 ion binding sites in a pore. Recently, work by McKinnon (Nobel Prize for Chemistry in 2003) [9] provided a way of visualising these binding sites in the crystal structure of potassium channels. The selectivity of calcium channels for divalent rather than monovalent ions was demonstrated in the 1980s as a result of work by W. Almers, R. Tsien, and P. Hess. The ion selectivity of channels is partly due to the pore size, estimated at 3.3×3.3 angstroms for voltage-gated K^+ channels, 3.1×5.1 angstroms for voltage-gated Na^+ channels, and 6.5×6.5 angstroms for nicotine receptors. Furthermore, ions in solution are surrounded by a cloud of water molecules (or hydration mantle). When they enter a channel, they drop most of these water molecules, replacing them by interactions with the polar groups of the amino acids forming the walls of the pore. It is thus the ability of inorganic ions to form hydrogen bonds with the oxygen atoms of the pore that will determine their permeation, more than the size of the ions themselves. In other words, an ion that happens to be smaller than another may not actually be more permeative, if it is less inclined to form hydrogen bonds.

Gating

The gating process is that property of channels whereby they modify their permeability with regard to ions in response to electrical or chemical stimuli. In other words, it is a mechanism for the transition from an open to a closed state, or vice versa. This process seems to depend on mechanisms of conformational change in protein channels.

Current research has focused on the molecular mechanisms governing this opening and closing machinery. Progress can be expected from combinations of high performance techniques. Hence, on the one hand, X-ray crystallography gives access to structural data and images of the channels. On the other, recordings of electrical activity are also available. Some research scientists use electron cryomicroscopy to determine the protein structure. This technique has lower resolution than X-ray crystallography, but it is more flexible with regard to identifying the conformation states of the protein. Another approach exploits fluorescence techniques to observe the motions of protein domains or residues. Research is also seeking to combine these different approaches with electrophysiology (see Chap. 19).

11.1.2 Physiological Role of Ion Channels

It was with the invention of the patch-clamp technique that the fundamental involvement of ion channels was clearly established in all the cell functions. The properties of the membrane potential, the action potential, and their variations under environmental effects can now be explained in molecular terms. The ion channels, these sentinel proteins at the heart of the biological membrane, control the flow from one side of this barrier to the other. They allow the exchange of ions between the cell and its surrounding medium, but also the exchanges between the intracellular medium (the cytosol) and the organelles. Moreover, they set up a signalling relationship between the cells and play a communicating role in functions such as secretion, excretion, the sense of smell, contraction, excitability, and so on [2–4].

Consequences of a Change in Channel Activity

Three types of effect can occur in succession following a change in the activity of a channel, rather in the manner of a reaction cascade:

1. The primary effect is a change in the membrane potential, e.g., with an outward flow of K^+ and a net inward flow of Na^+ in the case of a membrane that is permeable to Na^+ and K^+ .
2. For this reason, as a secondary effect, the ion concentrations on either side of the membrane will tend to change. Even if these ion flows are counterbalanced by active transfers (pumps), the transient variations in the ion concentrations nevertheless remain significant and can have short-term physiological effects. These effects can induce the activation of transporters or enzymes sensitive to Ca^{2+} or they may even have consequences for the cell as a whole, e.g., changes in volume following the activation of sodium channels.
3. Finally, as a tertiary effect, the activities of many enzymes, e.g., calmodulin or phospholipase C, are catalysed by a small change in the cytoplasmic ions. Note that calcium is an ion that plays a major physiological role in organisms, especially in muscle contraction or synaptic transmission, either by activating potassium channels that are sensitive to calcium, or by inhibiting certain channels in the endocellular membranes or by modulating the activity of voltage-gated channels.

Main Cell Functions

A range of the main cell functions involve ion channels: excitation (action potential), synaptic transmission (pre- and post-synaptic channels), transduction (receptors coupled to G proteins, receptor channels), contraction, secretion (insulin, exocrine), cell division, immune response, etc. Other functions

are now suspected of being controlled by channels, such as cell migration and adhesion (phosphorylation of K^+ channels), apoptosis² (K^+ and Ca^{2+} channels), cancerisation (K^+ channels comparable with hERG,³ Na_v channels, and Cl^- channels), and multidrug resistance (Na^+ , K^+ , and Cl^- channels). Note, however, that it is often difficult to know whether a change in channel functioning or expression is actually the cause or the consequence of a cell disorder.

11.1.3 Pharmacological Dysfunction

Like physiology, pathology is understood in terms of cell and tissue dysfunction. Some pathologies are related to hyperexcitation or hyperexcitability of nerve channels, or alternatively, a change in the nerve conduction. We shall discuss here two examples of pathologies involving ion channels, which stand out as possible therapeutic targets: pain and multiple sclerosis (or demyelinating polyneuropathy). The channel is thus one element in which pathologies may take their source. Around thirty pathologies have been identified that are related to ion channel dysfunction. These pathologies, also called channelopathies, are caused by spontaneous mutations affecting the genes coding for ion channels [10]. Once again, these disorders will be illustrated by various examples, although the list is not exhaustive: cystic fibrosis, myotonia, cardiopathy, and cancer [4, 11].

Pain Channels

Nociception is due to the activation of specific channels in sensory neurons. Let us mention two recently cloned channels involved in the transduction of nociceptive messages: the acid-sensing ion channel (ASIC) is activated during acidosis associated with an inflammation or ischemia, and the heat-activated channel (increase from $22^\circ C$ to $48^\circ C$) (vanilloid receptor subtype 1 or VR1) is also activated by capsaicin, the hot and irritating constituent of certain types of chili pepper. Local anesthetics block most ion channels, in particular, voltage-gated Na^+ channels, by interacting with a channel receptor. Likewise, the opioids are known to inhibit transient Ca^{2+} channels and activate inward-rectifying K^+ channels (see the note on sign conventions and definitions on p. 650). These two actions work to hyperpolarise the neurons, reduce the frequency of action potentials, and inhibit the release of neurotransmitters.

Multiple Sclerosis

Multiple sclerosis is a progressive deterioration of the central nervous system, in which scattered patches of myelin (the protective covering of the

² Apoptosis or programmed cell death, the natural process of elimination of excess, damaged, or aged cells.

³ hERG (human eag related gene) channels.

nerve fibres) are destroyed in the encephalon and the spinal chord. One of the therapies envisaged for multiple sclerosis and other forms of demyelinating polyneuropathy would be to use K^+ channel blockers, e.g., 4-AP, or 4-aminopyridine. The effect of this blockage would be to increase the duration of action potentials, i.e., increase local currents and thereby improve conduction in demyelinated axons. According to traditional Chilean medicine, the beneficial effects of a herbaceous plant called *Ruta graveolens* are known to improve the symptoms of multiple sclerosis by a blocking effect on the K^+ currents in the Ranvier node (a gap in the myelin sheath along an axon, allowing conduction by jumps in the electrical signals).

Cystic Fibrosis

Cystic fibrosis is a lethal hereditary disease that is most common in European and North American populations, affecting one birth in 2000. The gene responsible for this disease was discovered and cloned in 1989. It codes for the protein CFTR, which a Cl^- channel activated by the protein kinase A. On this gene, the deletion of three nucleotides coding for a phenylalanine at position 508 represents 70% of the mutant alleles. This mutation $\Delta F508$ affects delivery of the protein to the membrane and this means that it remains blocked and degrades in the endoplasmic reticulum. Other mutations affect the regulation or conductance of CFTR. The result is a deficiency in the secretion of Cl^- , accompanied by a reduction in the secretion of Na^+ and water into the mucous compartment. In the lungs, this has the effect of thickening the mucus and favouring chronic inflammation due to the development of pathogenic bacteria. Hence, the activation of CFTR would stimulate the outflow of chlorine by epithelial cells and inhibit the inflow of Na^+ . Pharmacotherapeutic methods and gene therapy are being used to develop ways of correcting for deficiencies in the transport of Cl^- ions.

Myotonia

Various pathologies are linked to the dysfunction of Na^+ channels (hyperkalemic paralysis), Cl^- channels (Thomsen's disease, Becker's disease), and Ca^{2+} channels (hypokalemic paralysis) in the skeletal muscle. Generally, it is through experiments with directed mutations and heterologous expression of mutated proteins that we now have a better understanding of the way mutations affect the working of these channels, clarifying, at least in part, the relation between the mutation and the clinical symptoms of these disorders.

Cardiopathies. The Cardiac Syndrome of Prolonged QT Interval

Various cardiac pathologies are characterised by an abnormally long QT interval, visible on the electrocardiogram, together with a distorted T wave. The

QT interval corresponds to the period of the ventricular systole, which lasts from the beginning of the excitation of the ventricles until the end of their relaxation. The common origin of various syndromes (arrhythmia, tachycardia, torsades de pointes, ventricular fibrillation) is a failing in the repolarisation of the ventricular action potential, resulting from various mutations of fast sodium or potassium channels. Three genes are particularly involved here: they code for the KvLQT1 channel, hERG, and SCN5A. Their various mutations lead to the forms LQT1, LQT2, and LQT3 of the syndrome.

It has recently been shown that certain drugs used in therapy induce cardiac arrhythmia. Some people acquire the prolonged QT syndrome owing to the fact that these drugs block the hERG channels. If this QT interval of the action potential becomes too long, the ventricular muscle can no longer respond uniformly to the beats and there is a desynchronisation of the cells toward chaotic electrical activity and sudden death. Some drugs had to be withdrawn from the market because of these side-effects on the heart [12]. The hERG channels are now used in the pharmaceutical development process as a probe for the side-effects of new drugs [13].

Cancers

The cell cycle depends on the translocation of ions through the membrane. Inhibition of cell proliferation has been observed in several types of cell, e.g., in the lungs, bladder, prostate, and breasts, when antagonists of potassium channels are applied. In breast cancer, it has been shown that control over cell proliferation and progression of the cell cycle depends on the activity of potassium channels.

Take the example of the IK(Ca) channels, i.e., potassium channels activated by calcium. These channels are involved in cell proliferation in prostate cancer. Decker et al. [14] have shown that the potassium channels are involved in the oncogenic transformation. By using selective blockers for the IK(Ca) channels, measuring the intracellular calcium, and introducing the human gene coding for the IK(Ca) channel into the Jurkat T cells, it has been shown that the IK(Ca) channels regulate the proliferation of T cells by controlling the inflow of calcium [15]. The inward flow of calcium is regulated by hyperpolarisation (see the note on sign conventions and definitions on p. 650) of the plasma membrane by IK(Ca). Note that the effects of IK(Ca) are the same if the cells are slightly or highly metastatic.

As another example, consider the BK(Ca) channels, which are large conductance calcium-activated potassium channels. These channels are involved in various cell mechanisms, such as control of the arterial tonus [16], or the generation and maintenance of hypertension [17]. These channels are sensitive to charybdotoxin (ChTX) and iberiotoxin (IbTX). Studies have shown that the expression of BK(Ca) channels drops off during development and reappears whenever cell proliferation occurs as a result of a lesion or illness [18].

The BK(Ca) channels are expressed in a cell line from a breast cancer (MCF-7) [19]. It has also been shown that the BK(Ca) channels play a role in the activation of B lymphocytes [20] and in the regulation of calcium involved in the activation of human T lymphocytes [21]. Ouadid-Ahidouch et al. have shown that the expression of BK(Ca) channels depends on the cell cycle and that these channels seem more relevant during the S phase than during the G1 phase of the cell cycle. The BK(Ca) channels thus help to sustain the synthesis of DNA by increasing the induction of mitogenic agents when the calcium concentration increases [22]. Experiments inhibiting the BK(Ca) channels by specific antagonists (ChTX or IbTX) have shown that these channels have no effect either on the growth of meningiomas [23], or on the development of prostate cancer [24]. On the other hand, they are involved in the growth of astrocytoma cells [25] and in the proliferation of retinopathy [26].

Still in the family of potassium channels, the hERG³ channels are voltage-gated channels deriving from the ether a-go-go (eag) channels. The role of these channels in the repolarisation of the cardiac action potential is well known [27]. They are expressed in several tumour lines, and also in primary cancer cells such as the adenocarcinoma of the endometrium [28], myeloid leukemia [29], and lymphoid leukemia [30]. The expression of hERG allows the growth of tumour cells and hence the development of cancers when there is nothing to induce apoptosis [31]. Tumour growth is linked to a change in the resting electrical potential of the cancer cell membrane. The hERG channels can modulate this resting potential [32], by depolarising the membrane [33]. The inhibition of these channels induces a reduction in cell proliferation [32]. Overexpression of eag channels induces transformations which give cells the characteristics of tumour cells. In leukemias, the hERG channels modulate the advance of the cancer through the mitotic cycle. Indeed, it has been shown that the gene coding for the hERG protein is mainly expressed during the S phase of the cell cycle [34]. The cells in a 'cancerous' colon express the hERG protein, whereas the same cells in a 'healthy' colon do not express hERG [35]. The role of hERG channels in primary solid tumours such as colorectal cancer can determine the acquisition of characteristics which will subsequently lead to an invasive, metastatic cancer. In the future, the immunohistochemical detection of hERG may be useful for medical prognosis. The hERG channels may eventually provide new molecular targets for therapy in the treatment of cancer.

Voltage-gated potassium channels play a role in cell proliferation. If these channels are inhibited, cell proliferation is observed to be inhibited in both slightly and highly metastatic prostate cancer cells [36–38]. The level of expression of voltage-gated potassium channels seems to be associated with the advance of metastasis. In other words, the more metastases there are, the more the channels are expressed [24].

The examples mentioned here are not intended to be exhaustive, but they do illustrate the involvement of potassium channels in the cancerisation phenomenon, and they show that there is a great deal of research. Other kinds of

channel, such as the calcium channels, are also involved in cell proliferation and apoptosis [39].

11.1.4 Direct Ways of Studying Ion Channels

Basic Concepts

The idea of animal electricity or bioelectricity was born in the eighteenth century with the work of Galvani (1737–1798), Volta (1745–1827), and Du Bois-Reymond (1818–1896), who demonstrated the existence of an electrical potential difference between the interior and exterior of cells, in frog nerves and muscles. The membrane potential hypothesis, which marked the true beginning of our understanding of channels, was put forward in 1902 by Bernstein. According to this hypothesis, the membrane of resting cells is selectively permeable to K^+ ions, and the membrane potential produced by passive transmembrane diffusion of these ions is equal to their equilibrium potential or Nernst potential, described in 1888. It then took another seventy years to be able to see these channels operating on an individual scale and to understand their structure [3].

The Hodgkin–Huxley Model

The model due to Hodgkin and Huxley, established for the giant axon of the squid, goes further into the details, stipulating that the resting potential should effectively be close to the equilibrium potential E_K for the potassium ion and that, in addition, the action potential results from a sudden and transient increase in the selective permeability to sodium. This increase is followed, after a time delay, by an increase in the permeability to potassium. Hodgkin and Huxley gave an experimental demonstration of this theory of a self-propagating action potential and showed for the first time that the sodium and potassium conductances were independent. To establish this model, they used the so-called voltage-clamp technique. In the case of giant axons, two electrodes are inserted into the body of the axon. One imposes the membrane potential V_m at a chosen value, while the other uses a feedback amplifier to deliver a current I equal and opposite to the transmembrane ion current, tending to cancel the difference $V_i - V_m$, where V_i is the imposed voltage.

Formulating the Ion Theory. Nernst Potential and Ohm's Law

The motion of an ion, e.g., a positive ion X , depends on its membrane permeability and its concentration gradient across the membrane. The result is a net flow of the ion from the compartment C_1 where its concentration is highest to the compartment C_2 where its concentration is lowest. This difference of concentration is called the concentration gradient.

The displacement of positive charges induces an excess of negative charges at the membrane surface, and this tends to retain the X ions in C_1 , slowing down their displacement toward C_2 . An electrical potential gradient thus arises.

The sum of these two chemical and electrical forces is called the transmembrane electrochemical potential. This potential is zero when the two forces are equal and opposite. The system is then said to be at equilibrium and the net flow of X is zero. This equilibrium potential, called the diffusion potential of ion X and denoted by E_X , is given by the Nernst equation:

$$E_X = \frac{RT}{zF} \ln \frac{[X]_{C_2}}{[X]_{C_1}} = 58.2 \log \frac{[X]_{C_2}}{[X]_{C_1}} ,$$

where $R = 8.31 \text{ V K}^{-1} \text{ mol}^{-1}$ is the perfect gas constant, T is the temperature in kelvin, z is the valence of the ion X, $F = 9.65 \times 10^4 \text{ C mol}^{-1}$ is the Faraday constant, and E_X is given in mV.

In a given physiological situation, when the transmembrane potential changes, there is a net flow of ions which thus generates an electrical current in the direction determined by the difference V_m minus the equilibrium potential for the ion X. Its value is given by

$$I_X = G_X(V_m - E_X) ,$$

where G is the electrical conductance in siemens and $V_m - E_X$ is the electrochemical gradient in volts. Since G is the reciprocal of the resistance R of the membrane to the transfer of X ions, the equation becomes

$$V_m - E_X = R_X I_X .$$

If $V_m - E_X > 0$, the current is outgoing, directed from the inner to the outer face of the membrane. If $V_m - E_X < 0$, the current is ingoing.

Sign Conventions and Definitions

Incoming currents correspond to an inflow of cations or an outflow of anions. They are shown directed downward and their value is attributed a minus sign. Inward-rectifying channels favour an incoming ion flow. Outgoing currents correspond to an outflow of cations or an inflow of anions. They are shown directed upward and their value is attributed a plus sign. Outward-rectifying channels favour an outgoing ion flow.

When resting, the membrane potential is negative with respect to the extracellular medium. Depolarisation is the gradual trend of the membrane potential to move toward positive values. Conversely, repolarisation is the gradual trend of the membrane potential to move toward negative values. Hyperpolarisation refers to the trend toward a more negative transmembrane potential.

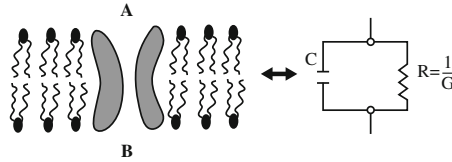


Fig. 11.1. Lipid bilayer (the cell membrane). *Left:* Transmembrane protein channel embedded in the bilayer and separating a conducting extracellular medium A from a conducting intracellular medium B. *Right:* Equivalent circuit diagram with a capacitor and a resistor in parallel

Generation of the Membrane Potential

In cells, the ions on either side of the membrane are unequally distributed. The K^+ ions are concentrated inside, while the Na^+ , Ca^{2+} , and Cl^- ions are concentrated outside. Since most cells are more permeable to K^+ ions, the latter tend to come out under the effect of their concentration gradient. There is then a deficit of positive charges on the inner face and an excess on the outer face. This charge separation forms a virtual capacitor with a capacitance of around $1 \mu F/cm^2$. The difference in the charge distribution produces an electric field which opposes the departure of K^+ ions. An equilibrium is reached when the electric field gradient exactly balances the concentration gradient, i.e., when

$$V \approx V_K = \frac{RT}{F} \log \frac{[K]_e}{[K]_i} \approx -80 \text{ mV} \quad (\text{at } 20^\circ\text{C}).$$

It is the free membrane permeability of potassium (and chlorine) which explains why the resting potential is negative, at a value of around -80 mV . In reality, the membrane is permeable to differing degrees with respect to several ionic species. Their potential is therefore situated between the equilibrium potentials of the different permeating ions. The value of the membrane potential thus depends on the relative conductances for the different ions. This explains why any relative change in some specific conductance will modify the membrane potential, which will then tend toward the equilibrium potential of the most permeating ion.

Capacitive and Ion Currents

This change in the membrane potential will result in a net transfer of charge such that $Q = CV$, where C is the capacitance measured in farads. This charge transfer induces the so-called capacitive current. Indeed, owing to the thinness of its membrane (about 10 nm), the cell behaves like a capacitor made from two saline conducting media, the intracellular and extracellular media, separated from one another by an insulator, the lipid bilayer (Fig. 11.1). The capacitance measures the ability to accumulate or store charges when a potential difference

is applied across this membrane. It is proportional to the area and inversely proportional to the distance between the two conducting layers or conducting saline media.

If a current is applied to the membrane, either by ion channels elsewhere in the cell, or by an electrode, this current will first recharge the membrane capacitance, and then it will change the membrane potential. This is illustrated by representing the membrane as an impedance of resistance R in parallel with a capacitance C (see Fig. 11.1).

Some Orders of Magnitude

- Thickness of a biological membrane: less than 10 nm.
- A transmembrane potential of 100 mV produces an electric field of 10^5 V/cm.
- Capacitance of a biomembrane: $1 \mu\text{F}/\text{cm}^2$ ($0.01 \text{ pF}/\mu\text{m}^2$).
- Number of K^+ ions transferred via the channel per second: 10^6 – 10^7 .

History of the Patch Clamp

The history and study of ion channels was marked by two key periods. The first, as already mentioned, happened at the beginning of the 1950s, when Hodgkin and Huxley, at the same time as Fatt and Katz, discovered the voltage-gated sodium and potassium channels. The second was the development of the patch-clamp technique by Neher and Sakmann at the beginning of the 1980s [40–42].

Indeed, although the Hodgkin–Huxley model was applicable to all excitable cells, physiologists came up against two difficulties: for one thing, it was impossible to see a single channel, and for another, the idea of impaling cells with two electrodes to apply the voltage-clamp technique could not be adapted to the problem of recording currents in small cells measuring only about ten μm across. Hence in 1976, the patch clamp was invented to deal with these obstacles. These two revolutions worked synergistically to extend electrophysiology toward novel biological syntheses from the animal or plant worlds, and a genuine marriage between biophysics and the approaches used in molecular and cell biology. Since the advent of molecular biology, electrophysiology has become molecular.

However revolutionary the patch-clamp technique may be as it is currently practised, it is nevertheless the result of successive improvements, as described by Erwin Neher in his Nobel Prize speech in 1992 [43]:

We made many systematic attempts to overcome the seal problem manipulating and cleaning cell surfaces, coating pipette surfaces, reversing charges on the glass surface, etc.) with little success. [...] By about 1980, we had almost given up on attempts to improve the seal, when we noticed by chance that the seal suddenly increased by more than two orders of magnitude when slight suction was applied to the pipette.

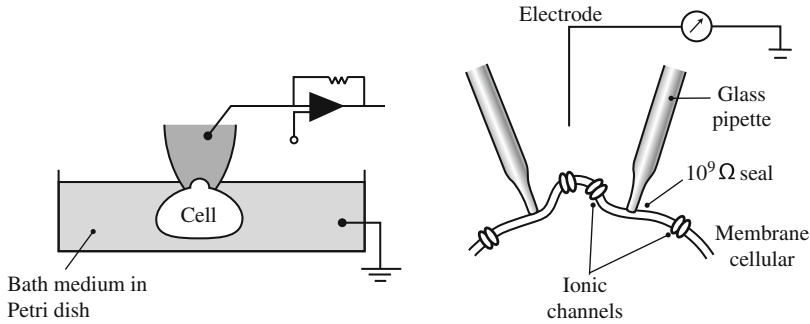


Fig. 11.2. Setting up a patch clamp. *Left:* Glass pipette sealed to the cell surface, itself stuck to the bottom of the Petri dish. *Right:* Magnification of the cell–pipette contact area, which defines the gigaseal, a high-resistance seal

The resulting seal between the cell membrane and the pipette, in the gigohm range ($1 \text{ G}\Omega = 10^9 \Omega$) was called a gigaseal. Improving the seal greatly reduced the background noise.

So the patch clamp is a voltage clamp adapted to a fragment of biological membrane, able to record the currents arising in individual channels (single-channel recording). Note that the patch-clamp technique is also used, by abuse of terminology, to refer to a voltage clamp applied to a whole cell (whole-cell recording). In the latter case, the macroscopic current recorded is the current arising from channels uniformly distributed over the whole surface of a cell [3, 44].

Experimental Implementation

The patch-clamp technique is thus defined as clamping a potential to all or part (a patch) of the cell membrane. The membrane is said to be clamped to the control potential, i.e., the potential fixed by the experimenter. The current crossing the membrane, or just a patch of the membrane, will then be recorded at this given potential. In general, the potential at the membrane is increased in steps and the current through all the channels opened by this change in potential is then measured. Patch-clamp experiments are set up according to the following sequence [3, 45]: the cell is connected via a patch micropipette to an amplifier head stage (or preamplifier, or current–voltage converter), itself connected to an amplifier, itself connected to a computer (see Fig. 11.2). We shall see that each of these steps requires a very specific preparation with correspondingly specific precautionary measures.

Biological Preparation

Primary Cultures. Between cells that are generally organised in tissues (organotypic slices) and cell suspensions accessible to patch-clamp measurements,

there is a whole series of preparatory steps that will affect the final current measurements. The primary culture phase includes enzymatic dissociation stages, generally combined with mechanical dissociations, and purification by sorting to obtain homogeneous suspensions, without clusters, aggregates, cell debris, or dead cells. To this end, gradient (Percoll) and flow cytometry techniques are used to enrich the populations to 85–95%. Selective cell culture, based on criteria such as adhesion, proliferation, or resistance to toxins, can also be used as a cell selection tool [3].

Heterologous Expression. Today, molecular biology provides ways of expressing artificially the genes coding for the relevant ion channels, using so-called heterologous systems. This expression is indeed artificial, because the channels are overexpressed, and hence over-represented in relation to native channels, and disconnected from the other ion channels. Such manipulations are of course envisaged with cloned ion channels, on which one hopes to carry out biochemical studies of the molecular structure and gain information about the intrinsic properties of the channel, e.g., permeation, activation and inactivation, rectification, pharmacological regulation, and so on. The heterologous systems used are, for example, large cells, such as the xenopus oocyte (between 300 μm and 1 mm in diameter), microinjected with mRNA or cDNA in the cytoplasm or the nucleus, respectively. The patch-clamp analysis requires a special preparation, involving removal of the vitelline membrane using tweezers or enzyme digestion [46]. An asymmetry is found to arise in the expression of the channels. In fact they are more expressed in the animal pole (pigmented zone) than the vegetal pole (bright zone). In general, the experimenter arranges to patch the pigmented side (dark zone) of the egg. Note that the xenopus oocyte has been widely used as an electrophysiological tool since 1971, when its utility as a support for expressing structure proteins or metabolic enzymes was first established. Today, it is used as support for expression of transporters, pumps, and ion channels.

Mammalian cells of more typical sizes and better suited to the whole-cell recording configuration, e.g., CHO, HeLa, HEK, Jurkat, etc., with diameters in the range 15–20 μm , are also used for transfection. Plasmids containing the gene coding for the protein channel also tend to carry the gene coding for a fluorescent protein, such as the green fluorescent protein (GFP). This allows optical control of the transfection. The efficiency of these transient transfections remains relatively low (varying from 30–60%). By molecular engineering and successive cultures with selection pressure, it is also possible to create cell lines with stable expression which incorporate the relevant gene at 100%. In this case, the plasmids containing the gene coding for the protein channel generally also carry a gene for resistance to certain molecules such as antibiotics, dissolved in the culture medium. For example, Chinese hamster ovary (CHO) cells are sensitive to methotrexate and die when incubated with this molecule. On the other hand, these same CHO cells, cotransfected by the gene coding for CFTR (which is the defective chloride channel in the case of cystic fibrosis), and by a mutated form of the gene coding for the enzyme insensitive to

methotrexate, resists this molecule. This artificial selection pressure thus provides a way of selecting those cells that have received the CFTR gene. Some groups have built up molecular and cellular tools that are also relevant here, such as inducible expression systems. So for example, cell lines transfected by the gene coding for the $K_v1.2$ channel will express different amounts of protein depending on the incubation time in tetracycline [47]. Such a tailor-made form of expression alone would provide a way of producing a graded electrical signal, by varying the incubation time. Indeed, for an incubation time close to zero, the level of protein expressed is very low and the model would allow single-channel recording. On the other hand, for longer incubation times, in the range 4–15 h, macroscopic currents could be obtained in whole-cell recording. These approaches remain artificial insofar as the relevant channel is not necessarily in its normal, or native, environment. The cell plays the role of host to the channel here, allowing the patch-clamp study to be carried out.

Microtransplantation. Some rather recent approaches can be used to reconstitute the natural diversity of endogenous ion channels and receptors in vitro, whence it is possible to reproduce and quantify the electrophysiological response of native membranes. The microtransplantation technique was developed by R. Miledi and coworkers. The idea is to extract native membranes from neuronal cell lines or human brain samples, and produce vesicles [48]. These vesicles are collected and microinjected into the xenopus oocyte. By membrane fusion, the native channels are delivered to the oocyte membrane and ready to be patched after a latency of 12 h (compared with three days for expression in the case of RNA microinjection).

Environment for Biological Preparation

Other factors affect the seal and recording quality: the primary culture protocol (asepsis, glucose-containing and buffered media, absence of calcium to avoid intracellular connections, choice of digesting enzyme, oxygenation, temperature, and so on) and the conditions whereby the ready-to-patch cell is immobilised at the bottom of the Petri dish. Indeed, the material chosen (plastic, glass, etc.), the way of fixing the cells (collagen, polyamines, Matrigel, carbon fibres, etc.), the type of support (box, filter membrane, etc.), and the cell density (monolayer, single cells, etc.) will strongly influence the cell morphology, the distribution of channels in the membrane, adhesion, and hence certain ion conduction, and it will in some cases cause the cell physiology to evolve (but note that the latter does not evolve in culture in the case of cell lines) [3].

Micropipettes

Micropipettes are obtained from glass tubes. To do this, the tube is held at both ends, in such a way that the middle of the tube is placed close to a hot tungsten filament, and separated into two parts by drawing the ends apart.

Two temperature cycles are programmed so that the tube can be tapered: the first, high temperature heating phase is very short, used to draw the capillary down to a diameter in the range 200–300 μm ; the second, lower temperature heating phase aims to continue melting the glass segment over a longer period, until the tapered tube breaks cleanly. This process involves a careful preparation of the temperature ramps, depending on the geometry and diameter chosen for the pipette. (In general, an ultrafine tip is sought in the range 2–4 μm .) This diameter is checked by measuring the resistance after filling the capillary with an electrophysiological (conducting saline) solution. This resistance is deduced by measuring the current due to a bias of a few mV. Many parameters of these pipettes will determine the success or otherwise of the patch clamp: the choice of glass (different types of hard or soft glass) [49], which depends on the type of cell one aims to study, the diameter of the aperture at the end, which will determine the type of recording configuration, and the state of the surface of the end of the pipette and the microroughness, which will determine the quality of the seal with the cell membrane. With this in mind, some experimenters have sought to improve the pipette fabrication process by ‘polishing’ the ends using a red-hot microforge under the microscope. Finally, the degree of cleanliness of the pipette will affect the quality of the seal. Particular care is taken over storage conditions for these pipettes. For this reason, before drawing, the capillaries are generally placed in a beaker filled with degassed ethanol. The capillaries are then dried in an oven. The micropipettes are drawn extemporaneously, i.e., just prior to the experiment.

From the electrical point of view, the pipette produces electrical interference noise. Indeed, by forming a capacitor of capacitance C (pF) (the glass wall separating two conducting media), it produces a high-frequency noise when it combines with the resistances of the solutions (bath and pipette). To reduce this capacitance and the associated noise, the tip of the pipette can be coated with an electrical insulator such as Sylgard, a resin which polymerises at a temperature of around 80° [49].

Micromanipulation

Mechanical Features: Penetration of the Pipette into the Cell. Once prepared in this way, the micropipette is filled 1/3 full with electrophysiological solution, using a syringe equipped with a thin, flexible plastic tube. Air bubbles are carefully removed from within the pipette by gently tapping with the fingers. Such bubbles might introduce a bias into current measurements by artificially increasing the resistance of the pipette, owing to a break in the electrical conduction. When the pipette has been filled in this way, it is threaded onto a chlorided electrode filament (see the note on Ag/AgCl electrodes below) and screwed into a holder, itself fixed onto the end of the amplifier head stage as shown in Fig. 11.3. The micropipette is pushed in until it comes into contact with the rubber seal. It is through this thread, bathed in the intra-pipette

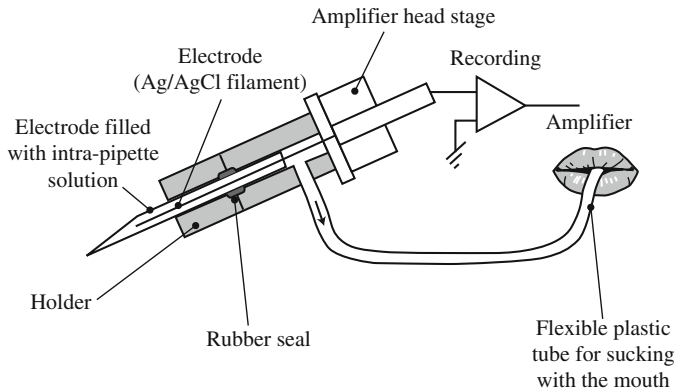


Fig. 11.3. Assembly of the glass micropipette, containing the recording electrode, on the holder and the amplifier head stage

medium, that the currents crossing the membrane will be conducted to the recording circuit.

The amplifier head stage is itself positioned on a support fixed on an antivibration table, allowing motion in the x , y , and z directions. By means of a macrometer screw, the experimenter carefully lowers the pipette toward the cell to be patched. Then using micrometer screws and under visual control via a microscope connected to a video camera, the tip of the pipette is manoeuvred into contact with the cell. To begin with, the experimenter exerts a positive pressure by blowing gently into a flexible plastic tube connected to the support (see Fig. 11.3). This flow of air avoids any possible obstruction of the electrode by particles, as well as cleaning very locally the region of the membrane destined to receive the pipette. The tip of the pipette then comes into contact with the membrane which is invaginated, i.e., it is pulled inside the tip of the pipette. The experimenter visualises and checks this invagination on the screen (clearly visible on the xenopus oocyte). At this point, the experimenter releases the positive pressure and sucks gently through the plastic tube. A very close contact then forms between the membrane molecules and the glass molecules. This is referred to as the seal.

The formation of a seal is the objective shared by all patchers. As we have seen, its success depends on the geometry and state of the pipette surface, but also on the cell morphology, the cell–pipette setup, and environmental factors such as the absence of vibrations and air movement. This is why an antivibration table is used, mounted on an air cushion, which determines the measurement space. The high electrical resistance of the seal will allow the measurement of low-noise currents and requires the use of a Faraday cage to protect the measurement from any source of interference, such as electromagnetic radiation. In order to do away with the manual, or rather oral aspect of the sealing process, some authors have suggested automating the pressure

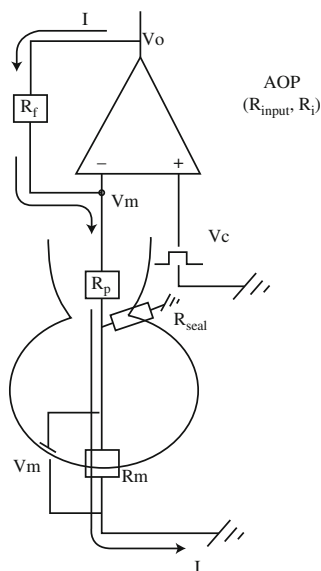
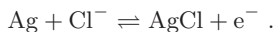


Fig. 11.4. Amplifier head

control inside the pipette using solenoid valves, as a way of standardising the procedure [50].

Ag/AgCl Electrodes

Unlike gold or titanium electrodes, chlorided silver electrodes are not polarisable. The Ag/AgCl electrode is selective to Cl^- ions and its potential is a function of the activity of the Cl^- ions in the solution. It converts ion currents in the solutions into a flow of electrons in the metal and takes part in the reversible chemical reaction with the ions in solution:



Charges do not therefore accumulate on the electrode during the reaction and its potential is stable. Regular regeneration of the electrodes by chlorination (by bleach) avoids any voltage drift.

Electrical Features. Apart from its mechanical function of holding the micropipette, the amplifier head stage has two electrical functions: it records the current which crosses the membrane, or the membrane patch, and it transfers the stimulation signal to the membrane.

The amplifier head contains two elements: the connector of the electrode holder and the operational amplifier (integrated circuit). Note that some commercially available amplifier heads also contain cooling systems for the electronic components to reduce unwanted electronic noise, and systems to compensate for the capacitance or stimulation artefacts.

- The connector between the amplifier head and the patch pipette is made from a dielectric material with a central conducting wire connected to the input of the operational amplifier. This considerably decreases the background noise from the system.
- The operational amplifier and the resistances represent the central part of the amplifier head. The amplifier is characterised by a high input resistance (greater than 10^{12} ohms) and low output noise. This input resistance is very high to ensure that the resistance of the pipette remains negligible. The output voltage of the amplifier is proportional to the recorded current I , which explains why the amplifier is called a current–voltage converter. It converts the high-resistance input signal into a low-resistance output signal, accessible to any recording device. Because it converts very low currents into voltages, the preamplifier is placed in a Faraday cage. It includes a set of electronic units able to calibrate and correct the signal output by the amplifier head, as well as the stimulation signal which enter the amplifier head. In general, amplifiers include a system for balancing the existing potential between the intra-pipette solution and the bath solution, a system compensating for the capacitance of the electrode, a system compensating for the series resistances and capacitance of the membrane, a control system for the holding voltage, low- and high-pass filters, and a system for subtracting the leakage current. The series resistances are located at the tip of the pipette and are due to the presence of small pieces of membrane which always block the pipette to some extent. These resistances can vary greatly from one experiment to another. A voltage in the mV range is sent into the amplifier and comes out in the form of a voltage in the range 0–5 V transmitted through the PC to the acquisition card. Signals from the amplifier are handled by dedicated software which also sets up the stimulation protocol (voltage steps). For further reading, see [3, 45].

Let us illustrate the working of the amplifier head by an example. Recall to begin with that a cell membrane can be reduced to its electrical equivalent circuit (see Fig. 11.1): a membrane resistance R_m of about 10^8 ohm and a current generator. A membrane potential V_m set up across the membrane results from the current I generated by ions passing through open ion channels. The value of V_m is fixed by the experimenter at V_c , the command voltage (see Fig. 11.4). Also known are the value V_o of the output voltage and the value of the resistance R_f , fixed between 50 M Ω and 10 G Ω depending on the amplifier head stage (to adjust the gain).

In our example, we seek to record the current crossing the membrane as a result of a voltage step of -80 mV (holding voltage) to -20 mV. V_c is adjusted to the level of the stimulator and the amplifier requires V_m to be equal to V_c at all times.

Some Orders of Magnitude for the Resistances in Whole-Cell Recording

The intracellular medium is continuously connected to the intra-pipette medium via an overall resistance of $R_p + R_m + R_{\text{seal}}$ such that:

- Pipette resistance R_p : 10^6 ohm.
- Input resistance R_i : 10^{12} ohm.
- Stimulator resistance R_s : 10^4 ohm.
- Membrane resistance R_m : 10^8 ohm.
- Feedback resistance R_f : 10^{12} ohm.
- Seal resistance R_{seal} : 10^9 ohm.

The overall resistance is R : 10^{12} ohm.

To obtain $V_m = V_c$, the operational amplifier generates a voltage V_o that will induce a current I in the circuit $R_f \Rightarrow R_{\text{pipette}} \Rightarrow R_m \Rightarrow R_{\text{bath}} \Rightarrow \text{earth}$, until the voltage V_m becomes equal to the voltage V_c , i.e., -20 mV. Following a voltage step, the channels through the membrane will open and as a result the resistance R_m will fall, the current I will change, and so will V_o . This change in the current only reflects the change in the membrane resistance, given that the values of R_f , R_p , and R_{bath} are constant. Applying Ohm's law $I = V/R$, the current I can be calculated automatically from V_o using the relation

$$I = \frac{V_o - V_m}{R_f} = \frac{V_o - V_c}{R_f}.$$

Now consider the 'anomalous' case in which the input resistance of the amplifier is not very high, e.g., $R_i = 10^7$ ohm rather than 10^{12} ohm. There are two consequences: the membrane voltage will be short-circuited and the measurement of I will be incorrect. The overall resistance becomes 1.1×10^7 ohm, i.e., ten times smaller than R_m . The current I will thus pass through the low resistance electronic circuit, i.e., the input resistance of the amplifier, rather than passing through the membrane.

Consider also the case where the cell-pipette seal is insufficient, with resistance less than the gigaohm, and/or there are impurities in the tip of the pipette, for example. As a consequence, there will be a lot of noise in the recording of the transmembrane current in the case of a poor seal (then R_{seal} decreases and R_m is inaccessible) and/or the current will be biased by the presence of impurities (in this case, R_p increases and R_m is inaccessible once again).

Current Recording

Seal Formation and Recording Configurations

The seal involves physicochemical processes that are still poorly understood. In principle, it involves molecular processes, although these have not been

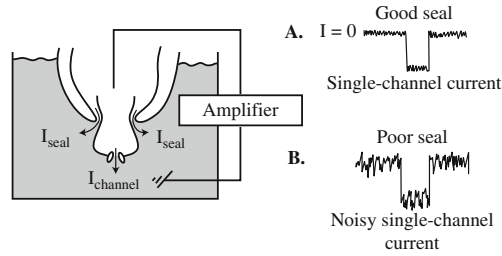


Fig. 11.5. Cell-pipette seal and the consequences of a poor seal, i.e., a seal with resistance less than the gigaohm (gigaseal), on single-channel recording. (A). If the current is 10 times greater through the seal [leakage due to a poor seal (B)] than through the channel, then the stochastic fluctuations in the current will be $\sqrt{10}$ times or 316% higher than in the case of a perfect seal. From M. Joffre [3]

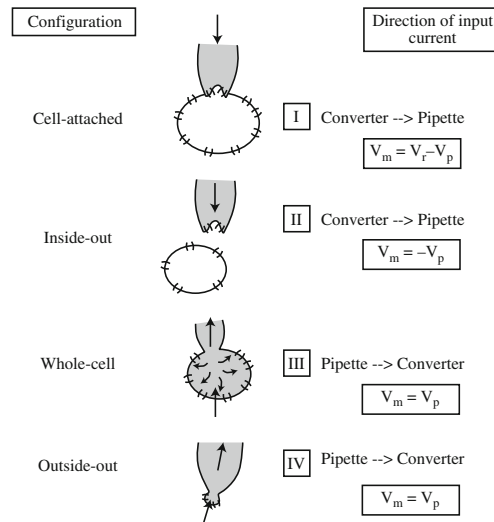


Fig. 11.6. Different ion-current recording configurations in the patch-clamp approach

established at the present time (see Fig. 11.5). Several hypotheses have been put forward that could explain the nature of the interactions, such as electrostatic charges, hydrogen bonds between lipids and glass, calcium bridges, van der Waals forces, and so on. In any case, it would seem that the seal occurs between the membrane and the tip of the glass pipette, but especially between the membrane invaginated inside the pipette and the glass walls. As we shall see in Chap. 19, this lack of exact knowledge concerning the membrane-pipette interactions is currently a major obstacle in the choice of material for developing patch-clamp chips, or indeed choosing the geometry and size of

the recording site on the cell. Indeed it is a real problem not to be able to imitate or reproduce the finer properties of the membrane–pipette coupling in a micro-orifice in a plane substrate (patch-clamp chip).

The seal has electrical and mechanical stability. Once obtained, the seal stabilises the interface between the cell and the pipette, while at the same time inducing a strong barrier against the diffusion of ions, and delimiting two independent compartments in which it is relatively easy to change the ion composition. Once this stability has been achieved, there are four possible recording configurations for the ion currents using mechanical micromanipulation (see Fig. 11.6):

1. Once the seal has been made, the pipette is always held in contact with the cell and imprisons in its tip a small number of channels (usually between 1 and 100). This is the cell-attached configuration [51].
2. After this configuration, when the piece of membrane is torn off by simply withdrawing the pipette, the inside-out configuration is obtained, i.e., the piece of membrane remains fixed to the pipette and it is now its inner face which is exposed to the extracellular medium. In this configuration, it then becomes possible to modulate the inner medium of the cell.
3. To obtain the whole-cell configuration, we return to the initial seal configuration, but this time the membrane fixed to the pipette is ruptured and the pipette and the cytosol enter into contact. This rupture of the membrane can be obtained by several processes: either by a large pressure difference under suction, or by applying an electrical pulse of 1 V for a few microseconds or a few milliseconds. The currents resulting from N channels distributed over the membrane can then be recorded. In this configuration, the physiological conditions are biased insofar as diffusion from the intra-pipette medium modifies the content of the intracellular medium. A less invasive trick provides a way around this problem: a polyene antibiotic permeabilising agent such as amphotericin or nystatin can be applied. These molecules form channels in the membranes (also called membrane partitioning) containing cholesterol or ergosterol. These agents which tend to perforate the membrane, once introduced into the micropipette, form channels which are permeable to monovalent cations and Cl^- , but exclude multivalent ions such as Ca^{2+} or Mg^{2+} . Note that the conductance of channels formed by amphotericin is twice the conductance of channels formed by nystatin. In addition, the presence of these perforating agents will affect the quality of the seal, so special experimental precautions have to be taken. As we have seen, some systems providing heterologous expression of the channels, such as fibroblasts, HEK cells, or CHO cells, lend themselves well to patch-clamp analysis in this configuration. On the other hand, bulky cells like the xenopus oocyte are not so suitable, because the potential cannot be imposed and distributed in a uniform way over the whole of the membrane. In this case, one uses the double intracellular

electrode technique, one electrode imposing the potential and the other measuring the current.

4. Once the whole-cell configuration has been obtained, the pipette can be withdrawn. A piece of the membrane then detaches from the cell and forms a half-vesicle at the tip of the pipette. The extracellular face is this time exposed to the bath medium. This is the outside-out configuration.

Choosing a Configuration

The cell-attached, inside-out, and outside-out configurations are used to study ion channels on an individual level. The current is recorded for the imposed voltages in the presence of various ion solutions, whence the ion selectivity and channel conductance can be determined. The effects of regulatory agents or blockers are studied by measuring the fractional opening time, or opening probability, of the channels. In these configurations, one generally works in the steady state.

In the whole-cell configuration, single-channel currents are no longer distinguished and one records the currents produced by sequential application of voltage steps of fixed duration and variable value. This configuration is particularly well-suited to excitable tissue currents, because it informs as to the electrical properties of the currents under investigation.

Recording and Analysing Ion Currents

Consider first the single-channel currents. A channel can only exist in an open state or a closed state. The activity of the channel induced by an extremely rapid change in molecular configuration shows up through rectangular steps in the current, with amplitude determined by the electrochemical gradient of the surrounding ions and by the conductance of the channel itself. Their duration reflects the presence of the open or closed state of the channel [52]. These are stochastic quantities depending on the gating (opening and closing) mechanisms of the channel. As well as the fast intramolecular rearrangements which occur in the transition from one state to the other, a sufficient energy step is also required to get over the free energy barrier separating the two states. This random crossing of the energy barrier is due to rapid vibrations of the protein channel, related to its thermal energy. These vibrations occur at a frequency of 10^{-11} – 10^{-12} Hz. The random nature of the transitions between states should be borne in mind.

It is essential to limit noise when recording single-channel currents smaller than 1 pA. Experimental conditions are sought with a signal-to-noise ratio greater than 3. The predominant noise is thermal noise from the seal, which one seeks to reduce. The noise can be significantly reduced by choosing a low-noise pipette glass, e.g., a hard glass such as Pyrex or quartz, but there are other tricks, such as restricting immersion of the tip to 2 or 3 mm.

The analysis of a single-channel current (in general, the current in a single channel is around 2.5 pA) serves to understand the operating kinetics of the

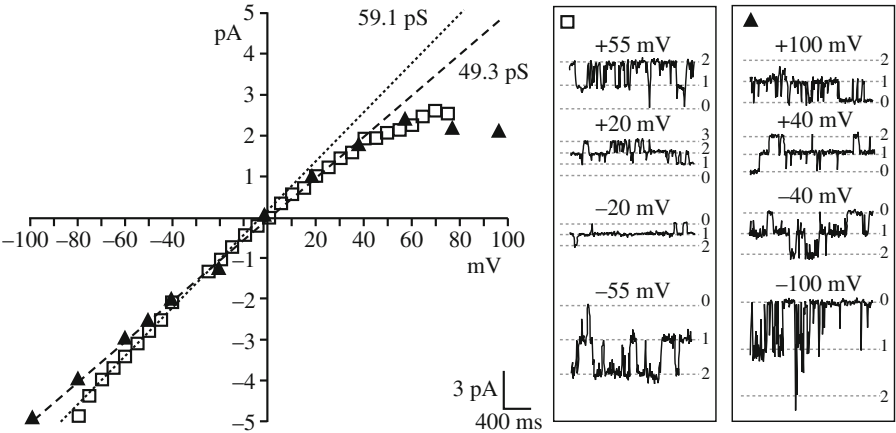


Fig. 11.7. Single-channel recording for K_{ATP} channels of the skeletal muscle for different voltages (*right*) and the corresponding current–voltage characteristic (*left*). With the kind permission of C. Arnould [52]

channel, and identify its different conductance states, the transition rates between states, and so on. The recording thus looks like a sequence of opening and closing events (see Fig. 11.7). This chain of events is very closely analysed, and if possible modelled and represented specifically in the form of amplitude histograms, occupation times of the states, etc. This detailed analysis of a single-channel current leads to the determination of individual kinetics [3,53].

Consider now macroscopic current recording. Applying a given voltage to the whole cell membrane and adjusting the compositions of the intracellular and extracellular media, macroscopic currents can be obtained (in the nanoampere or even the microampere range), resulting from the sum over N channels distributed almost uniformly over the membrane. This time, in contrast to the single-channel analysis, the aim is to characterise a channel by its average behaviour. Since one has a population of N channels, it is easy to see that the global probability for these N channels to make a transition is equal to the probability for one channel to make N transitions. Hence the macroscopic currents account for the behaviour of N stochastic events.

From the experimental standpoint, we find here the same constraint of establishing a high-resistance seal in order to reduce background noise in the signal. In addition, in the whole-cell configuration there are two new phenomena related to rupture of the membrane: a sudden change in the clamp voltage imposed on the membrane and a new ionic and chemical equilibrium between the pipette and the cytoplasm (dialysis). The pipette resistance is then in series with the access resistance which characterises the junction with the cytosol.

Variants of the Patch-Clamp Technique

As already mentioned, the patch-clamp technique cannot be adapted to the study of large cells like the xenopus oocyte. In this case, the double electrode technique is preferable. The flexibility of the patch-clamp technique, combined with its potential for evolution and extension, and its flexibility of use, are what make it so successful, and experimenters have taken full advantage of this versatility to adapt the patch-clamp technique or platform [54] to specific problem situations.

Returning to the case of the oocyte, for which it is difficult to impose a uniform voltage over the whole membrane by means of a single electrode, if one seeks nevertheless to work in the cell-attached configuration, one solution is to use wider bore pipettes, with diameters around 10 μm rather than the 2 μm conventionally used. Since the seals here do not generally exceed a hundred or so megaohms, this has been called the loose patch-clamp technique. It is still suitable for measurements on muscles, and neurons such as the squid giant axon. One should also mention the technique of giant-membrane patch recording. Using pipettes with diameters in the range 10–40 μm , this can examine membrane macrofragments and hence higher channel densities.

On small cells, when moving to the whole-cell configuration, the rupture of the membrane is traumatic for the cell, which finds its cytoplasmic medium replaced by a pipette medium of known and controlled ionic composition. Washing of metabolites and other vital cytosolic factors in the cell then occurs and some cell functions may be lost. This phenomenon, known as run-down, reflects the inactivation of certain channels. To remedy this problem, one has recourse to a membrane perforation or dialysis technique using antibiotic agents (amphotericin B, nystatin, see p.662). The piece of membrane is no longer ruptured, but permeabilised by the formation of pores measuring a few angstroms in diameter, which provide electrical connection between the cell and the pipette, while preserving the vital contents of the cell. This technique is used in particular to study signalling mechanisms, but also exocytosis in secretory cells, and the cell response to metabolic changes or osmolarity.

Recording has also been carried out on tissue sections, in order to investigate the cell response within its original tissue and cell architecture. In particular, such experiments have been done on freshly prepared organotypic brain slices from young rats. The technique can be used to study fast, low amplitude processes thanks to the high recording quality, mainly due to a high signal-to-noise ratio. It can also investigate interactions between biochemical and electrical events. For example, combining this approach with calcium (fluorescence) imaging, one can monitor calcium movements due to exocytosis of neurotransmitters and the neuronal electrical response.

Yet other variants have been introduced to study for example the dynamics of exocytosis, in which the final stage of membrane fusion leads to changes in the membrane capacitances of femtofarad order, which can be measured. In other situations, for example, when studying cell endomembrane channels,

there is a technique using artificial membranes, enriched in purified ion channels (see the detailed review in [3]).

Indirect Techniques for Studying Ion Channels

As we have just seen, the patch-clamp technique is the only one allowing direct access to the electrical activity of the channels, insofar as it either reveals the electrical current, or quantifies the transmembrane ion flow. It is certainly useful to situate this technique among the indirect methods that we are not discussing in detail here: the binding technique, used to detect the interaction of a radioactively labelled compound with an ion channel, the flow technique, which uses radio-isotopes (Rb for potassium channels, Li for sodium channels) to follow the flow of specific ions through the cell, and finally, fluorescence techniques, which measure the membrane potential or changes in the ion concentrations resulting from ion flow. These indirect measurements are made using fluorescent probes, e.g., Fura-2 or Indo-1 as calcium probes, whose spectrum changes as a result of modifications in the membrane potential or the ion concentration (see the review in [59]).

The membrane potential can also be measured using an extracellular approach. Although less informative than an intracellular measurement, the extracellular measurement can nevertheless provide information about the activity of the action potential in a cell population, e.g., in small networks of neurons, brain slices, or cardiac tissue. A highly innovative bioelectronic interfacing approach has been designed and implemented by Fromherz and coworkers [55, 56]. The membrane potential producing a strong electric field can be coupled to a field-effect semiconductor which amplifies extracellular currents (see Chap. 19).

11.1.5 Conclusion: Prospects for the Patch-Clamp Technique and the High-Throughput Revolution in Electrophysiology

The very wide field of applications of the patch-clamp technique, including the prospect of new therapeutic methods, reflects the desire to push the performance and potential of this approach to its limits. Ever more sophisticated analytical strategies emerge, exploiting these innovative ideas and principles to refine our knowledge and understanding of cell physiology, physiopathology, and molecular pharmacology.

As these scientific and technical advances have been made, the ion channel has been transformed from a mere concept into a major actor in cell signalling processes, with a molecular identity as a protein inserted into a membrane, the scene of a complex structure–function relation, at the very heart of a range of pathological conditions known as channelopathies. Ion channels have thus become increasingly relevant molecules in medicine as potential therapeutic targets. The functional analysis of these protein channels represents a genuine bottleneck in the process of discovering new active pharmacological

compounds and a major challenge for the pharmaceutical industry, which is trying to find and validate innovative molecules, specifically targeting these channels and other membrane transporters [57, 58]. Finally, owing to their bioselectivity, these channels are now ideal candidates for the development of biosensors used to detect pathogens and toxins.

As we have seen in this section, the patch-clamp technique, although extremely effective, is highly sensitive to interference and requires a considerable infrastructure to run properly, including a Faraday cage, anti-vibration table, etc. Experimental implementation demands an expert touch to position the glass pipette by micromanipulation, or to make the seal under microscope observation. Being an electrophysiologist is no part-time occupation! In general, it takes a good first year of a doctoral thesis to master the technique with all its subtleties, and an uncommon patience! What is more, even a senior patcher will only run out 20 measurements on a good day.

In contrast to the realities of the research laboratory, there are the requirements of the pharmaceutical industry: the ever increasing number of candidate molecules and mutant proteins, generated by new techniques of combination chemistry and genetic engineering, respectively, and the urgent need for new, reliable and efficient high-throughput screening tools, able to obtain results on small samples [59–61]. In its present configuration (one pipette yields one measurement), it is clear that the patch-clamp technique will not be able to meet the demand. And yet it is essential to exploit its high performance and potential in terms of resolution. This is the technological challenge of the 2000s: how to reconcile patch-clamp methods with high throughput. How can the patch clamp be exploited on a large scale and hence made accessible to non-specialists? How can it become a commonplace screening platform without a whole complement of restrictive infrastructure? How can it be adapted to analyse millions of cells each week? All these questions reveal the need to literally revolutionise the very concept of the patch-clamp micropipette [62]. We shall see in Chap. 19 how cell chips can bring a part answer to such questions, opening the way to a new era in science and technology.

11.2 Amperometry

Using the techniques of modern genomics, local imaging, near-field microscopy, and combinations of these, our structural understanding of fundamental biological phenomena on the cell level has gone ahead in leaps and bounds.

In many cases, and more and more regularly, the architecture of cell organelles or cell assemblies in living tissues is coming within the reach of our understanding, even on the molecular or supramolecular level. However, despite the ever more detailed description of the functional landscape of the cell, the molecular mechanisms underlying intra- and intercellular communication often remain confused and approximative. This dichotomy may seem surprising, but it is easily explained when one realises that the chemical and

biological molecules which define cell communication channels are present in extremely small amounts (between a thousand and a million neurotransmission molecules, for example), emitted for very short time lapses or with fast temporal modulation. For a given molecule, the information actually exchanged between a source and a receptor consists essentially in a different variation in its concentration flow. Only this transient infinitesimal information has any real meaning and this explains why it is so difficult to obtain.

It may also seem surprising that nature would select a means of communication based upon this system, i.e., a very small number of molecules emitted and received in volumes that are generally in the femtolitre range. However, the response of a biological molecular detector is based partly on selective detection and partly on chemical or biochemical reaction rates. Now in a cell, for a given substrate–detector combination, the molecule–substrate reaction rate depends on the probability of an encounter between the molecule and the substrate, i.e., the concentration of the molecule in a nanoscale region centered on the substrate. For example, if one thinks of a neuronal synapse, one bit of information comprises a mere ten thousand molecules or so. This seems extraordinarily few on the macroscopic scale. Moreover, given the chemical nature of neurotransmitters and their spectroscopic properties, it is impossible to observe the corresponding flow, even with the best spectroscopic methods currently available. However, these molecules or ions are emitted in a gap (the synapse) which is itself infinitely small, in such a way that, in this synapse, the concentration varies significantly (from micro- to millimolar). The detectors implanted on the surface of the membrane of the receptor neuron thus have no difficulty in detecting the bit of information kinetically.

This example of natural nanotechnology perfectly illustrates the biological problem situation discussed in this chapter and the basic principles of the faradaic electrochemical methods used to tackle it. These methods use ultramicroelectrodes, i.e., micro- or nanoscale electrodes, whose main properties will be described here. To avoid confusion with other phenomena such as ion currents, capacitive currents, etc., or the corresponding analytical methods described elsewhere in this chapter, we begin with a brief summary of the basic principles underpinning faradaic electrochemical measurements.

11.2.1 Basics of Faradaic Electrochemistry

Any experiment in electrochemistry necessary encompasses two large families of phenomena, which need to be considered separately owing to their different physicochemical nature and the different energies involved.

In the strict sense of the term, electrochemistry refers to a context in which electrons are transferred from the electrode to a chemical substrate, or conversely. There is a genuine chemical reaction between a reagent, an electron or a hole, and some chemical structure. This reaction involves energies of the order of the electronvolt and an activation stage, i.e., a chemical activation

barrier. With reference to the pioneering work by Faraday, this class is referred to as faradaic electrochemistry.

At the same time, the structuring of the electrode–solution interface, the resistance of the medium, etc., involve electrostatic and electrokinetic physical phenomena with energies of order kT . Any variation in the potential of the interface, or the passage of a current in the cell, will cause changes in the relevant physical equilibria. The result is usually a current due to the displacement of ions, but not involving any chemical modification of the substrates. One then speaks of ion currents (i.e., capacitive currents, as explained below), electroosmotic currents, or electrophoretic currents, and more generally of non-faradaic processes.

These different types of phenomenon are often unthinkingly identified, given that they are obviously very closely related. This introduces a harmful confusion in the understanding of electrochemical phenomena. For this reason, we begin by examining each class of process separately depending on whether it is faradaic (chemical, eV) or non-faradaic (physical, kT).

Non-Faradaic Processes

The Electrode–Solution Interface

The presence of immobile charge carriers at the electrode–electrolyte interface and the existence of an electrical potential difference across this interface results in an accumulation of charge on either side of the interface. The charge carried by the electrode represents either an excess or a deficit of electrons. On the electrolyte side, the charge on the electrode is balanced by an accumulation of ions of opposite charge and by a deficit of ions of the same charge as that carried by the electrode. An electrochemical double layer then forms at the interface, within which there is a potential difference. As in any junction, this potential difference arises in the same way as in a charged capacitor to remove the discontinuity in the potential. This situation can be more complex if there are specific adsorption processes, whether they involve ions or neutral molecules.

Depending on the nature of the electrode and the electrolyte, there is one value of the potential for which the interface carries no charge. This is the zero charge potential E_{eq} . When the electrode potential is imposed or modified, it necessarily becomes greater or less than E_{eq} . For a negative difference, the electrode surface is then negatively charged. The cations contained in the solution and the dipoles present, e.g., solvent molecules, are attracted to the interface. Although the region in which the charge accumulates in the electrode can be considered extremely thin, this is not so for the layer of solution in which the ion distribution is no longer electrically neutral. Indeed, a potential difference at the interface imposes a complex structuring of the solution, due to the different nature and behaviour of ions or polar molecules, and the low density of charge carriers.

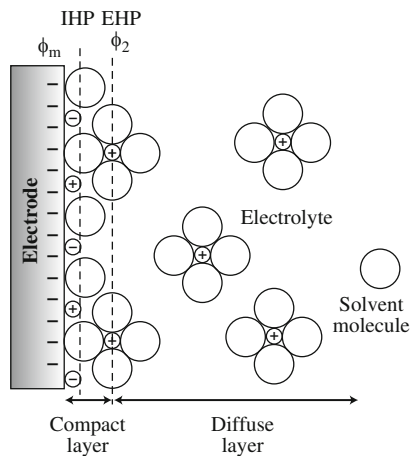


Fig. 11.8. Electrochemical double layer. ϕ_m is the potential of the electrode and ϕ_2 the potential in solution in the outer Helmholtz plane (OHP). The inner Helmholtz plane is abbreviated to IHP

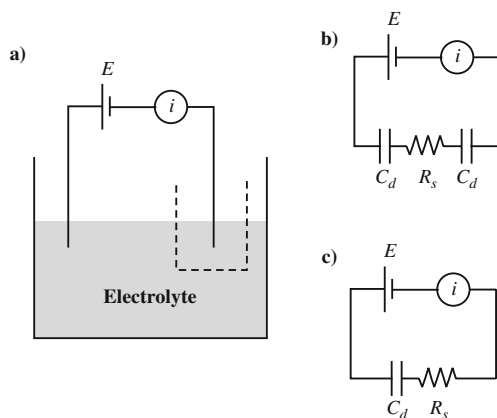


Fig. 11.9. Two-electrode setup. (a) Electrochemical cell. (b) and (c) Equivalent electrical circuits in the absence of faradaic processes when the second electrode is polarisable (b) or non-polarisable (c)

The electrical double layer thereby produced comprises a highly ordered compact layer or Helmholtz layer, made up of ions and dipoles that are strongly adsorbed on the interface (see Fig. 11.8). It corresponds to the minimal approach distance (in the angstrom range) of the excess ions that come to balance the excess charge on the electrode. A distinction is nevertheless possible between the inner Helmholtz plane (IHP), i.e., the plane in which the ions are specifically adsorbed onto the electrode, and the outer Helmholtz plane (OHP), i.e., the plane in which solvated ions are held (see Fig. 11.8). The

thickness of this region is determined by the solvation sphere of these ions and the intercalation of a layer of adsorbed solvent molecules between the ions and the electrode. Beyond this, the solvated (not specifically adsorbed) ions are subject to long-range electrostatic forces, independent of their chemical properties, and thermal excitation within the solution. This sets up a 3D space charge region called the diffuse layer, or Gouy–Chapman layer, within which these ions distribute themselves in such a way as to balance the two constraints.

The potential distribution at the interface thus occurs over a region with thickness depending on the total ion concentration in solution. The latter generally reaches a few tens of angstroms (the Debye length). Under typical conditions in which the electrolyte concentration is high enough, the potential difference in the diffuse layer is very low, i.e., $\phi_2 - \phi_s$ is of the order of ten millivolts. It is then mainly within the compact layer that the potential difference between the electrode and the solution occurs (see Fig. 11.8).

Before exchanging electrons with the electrode, an electroactive species that is not specifically adsorbed will thus be subjected to the potential difference $\phi_m - \phi_2$ at its minimal approach distance, i.e., at the outer Helmholtz plane.

Interface Capacitance

The electrical behaviour of the electrochemical double layer is responsible for the observed capacitive processes, even in situations where there is no electrochemical reaction. The currents measured are due to non-faradaic processes, because they now involve electron transfer at the electrode–electrolyte interface.

Hence, any change in the potential of one electrode will necessarily lead to a reorganisation of the ions and dipoles in the electrochemical double layer and a modification in the charge at the interface. This will result in a flow of electrons in the outer circuit, which will be measured as a charge current from the interface capacitance. In electrochemistry, charge currents or capacitive currents have to be minimised when carrying out experiments. Indeed, they add to the faradaic processes under investigation, i.e., to the processes involving electron transfer at the interface coupled to a transformation of chemical species (see below).

The equivalent circuit of a two-electrode setup (see Fig. 11.9a) comprises two capacitors C_d in series, these representing the electrochemical double layer at the two electrodes, and a resistor R_s representing the resistance of the solution (see Fig. 11.9b). If one of the two electrodes is not ideally polarisable and its potential is fixed, its capacitance in the circuit can be neglected, so that applying a potential difference between the two electrodes will amount to treating a standard RC circuit (see Fig. 11.9c), in which the only capacitance to be taken into account is that of the recording electrode. The change in current at the interface that results from a potential step is then given by

$$i = \frac{\Delta E}{R_s} e^{-t/R_s C_d} . \quad (11.1)$$

The current decreases exponentially with time and tends to zero with a time constant equal to $T = R_s C_d$. When the electrochemical double layer has fully reordered itself after applying a new voltage to the electrode, the capacitive current will become sufficiently negligible compared with the faradaic component. The charge current of the electrochemical double layer will then only constitute a problem for experiments on a short time scale. Note, however, that the term $R_s C_d$ decreases with the size of the electrode.

Faradaic Processes

Electrochemical reactions involve oxidation and reduction via electrodes which can be made from metal or some other electron conducting material:



These reactions thus result from charge transfer or electron transfer at the electrode–electrolyte interface between an electron conductor and an ion conductor. Species that can take part in an electrochemical reaction at the surface of an electrode are said to be electroactive. They can be organic, inorganic, neutral, or charged.

An electrochemical reaction corresponds to local electrolysis of a solution in the immediate vicinity of an electrode and gives rise to the detection of a faradaic current in the outer circuit:

$$\frac{i}{nF} = \frac{dN}{dt} , \quad (11.3)$$

where $F = 96,485^\circ\text{C mol}^{-1}$ is the Faraday constant and N is the number of moles of electroactive species consumed at each electrode in the circuit for n moles of electrons exchanged. Indeed, the overall reaction of the electrochemical cell results from coupling between an oxidation reaction at the anode and a reduction reaction at the cathode. These faradaic processes take place in parallel with non-faradaic processes. The equivalent circuit at an interface then comprises a faradaic impedance Z , added to the previous RC circuit (see Fig. 11.10).

In order for an electrochemical reaction to occur at one of the electrodes, the solution outside the two double layers must be kept electrically neutral and the charge transferred to the second electrode must be the same but of the opposite sign. Furthermore, this reaction will not take place unless the potential difference E between the two electrodes is big enough to exceed the free energy ΔG of the overall reaction of the cell. The voltage E_{appl} applied across the terminals of the cell is

$$E_{\text{appl}} = E + iR_s , \quad (11.4)$$

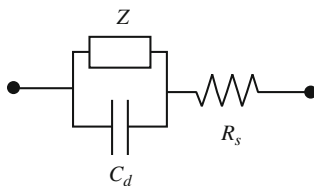


Fig. 11.10. Equivalent circuit for faradaic and non-faradaic processes at each interface

where iR_s is the Ohmic drop due to the resistance of the solution. Its contribution can be minimised by good design of the cell and instrumentation, together with an appropriate electrolyte composition.

In general, electrochemical reactions are only studied on a single electrode called the indicator electrode. The other, auxiliary electrode serves only to carry the current into the electrolyte. Moreover, in order to measure the potential of the indicator electrode, a third, reference electrode with known potential and not polarised is introduced into the setup. The auxiliary electrode can be of any kind provided that it does not affect the electrochemical reaction studied by the indicator electrode. For this reason it is as far away as possible, or even placed in a separate compartment (see Fig. 11.9a).

Although conditions may be suitable to make the reaction thermodynamically favourable, i.e., the applied voltage is sufficiently positive or negative, the reaction rate and hence the recorded current will depend on kinetic processes intrinsic to the electrochemical reaction, such as

- the rate of electron transfer to the interface,
- the mass transfer of electroactive species from the solution to the electrode,
- the chemical reactions preceding or following electron transfer.

Electron Transfer

By its chemical nature, the reaction is an elementary act and hence first order with respect to the two associated electroactive species [see (11.2)]. For one electrode, its rate is given by

$$v = \frac{i}{nFS} = k_{\text{an}}C_{\text{ox}}^{\text{el}} - k_{\text{cat}}C_{\text{red}}^{\text{el}}, \quad (11.5)$$

where k_{an} and k_{cat} are the oxidation and reduction rate constant, respectively, for the relevant electrochemical reaction, $C_{\text{ox}}^{\text{el}}$ and $C_{\text{red}}^{\text{el}}$ are the concentrations of the oxidising and reducing electroactive species at the interface [see (11.2)], and S is the surface area of the electrode. As in any elementary chemical process, the rate constants obey the Arrhenius activation law and depend on the voltage applied to the indicator electrode through the changes in the electrostatic charge:

$$k_{\text{cat}} = k_0 \exp \left[-\frac{\alpha n F}{RT} (E - E^0) \right], \quad (11.6)$$

$$k_{\text{an}} = k_0 \exp \left[\frac{(1 - \alpha) n F}{RT} (E - E^0) \right], \quad (11.7)$$

where k_0 is the standard reaction rate, E_0 is the standard potential, and α is the charge transfer coefficient for the relevant electrochemical system, i.e., the fraction $nF(E - E^0)$ of the total electrical energy that can be used to get over the activation barrier. Putting together (11.5)–(11.7), the current is obtained as a function of the voltage:

$$i = nFSk_0 \left\{ C_{\text{ox}}^{\text{el}} \exp \left[-\frac{\alpha n F}{RT} (E - E^0) \right] - C_{\text{red}}^{\text{el}} \exp \left[\frac{(1 - \alpha) n F}{RT} (E - E^0) \right] \right\}. \quad (11.8)$$

When the applied voltage is high enough, i.e., when $|E - E^0| > 0$, one of the two terms dominates over the other. Depending on the case, the electrochemical reaction gives rise to a reduction process for $E - E^0 < 0$ or an oxidation process for $E - E^0 > 0$. When the reaction rate becomes large, the concentrations $C_{\text{ox}}^{\text{el}}$ and $C_{\text{ed}}^{\text{el}}$ at the electrode vary significantly compared with the values in solution. The current is then soon limited by the mass transport of electroactive species from the solution to the electrode.

Mass Transport

Mass transport occurs in solution because of the electrochemical potential gradient imposed by each electrode, but it also results from natural or forced convection in the solution. The current is proportional to the flow rate J of each electroactive species, given by

$$i = \pm nFSJ. \quad (11.9)$$

For a species of charge z and diffusion coefficient D , the flow rate J contains three terms [see (11.10)], each representing the contribution of one transport mode:

- diffusion, caused by a local concentration gradient ∇C ,
- migration, which is the displacement of charged species due to a local gradient $\nabla \Phi$ in the electrical potential,
- convection, resulting in a local movement of the solution at speed v .

This leads to the expression

$$J = -D\nabla C - \frac{zF}{RT} DC\nabla \Phi + Cv. \quad (11.10)$$

The relative contribution from migration to the transport of an electroactive species varies from one point to another in the solution, but it may be as important as the two other modes in the immediate vicinity of the electrode.

However, for practical reasons, wherein experiments are carried out under conditions in which mass transport can be perfectly controlled by chemical properties, this mode of transport is often rendered experimentally negligible by adding a large amount of supporting electrolyte to the solution. The latter, with a much higher concentration than the electroactive species, takes charge of the current transport by migration between the two electrodes. This leaves experiments implementing a well-defined hydrodynamic regime in the solution, and those in which there is no convective regime, i.e., when the solution is not stirred and the electrode is motionless. In the latter case, diffusion is the main transport mode, apart from natural convection of the solution which occurs some hundred or so micrometers from the electrode surface.

Amperometric Detection

Amperometric detection consists in applying voltage steps to the indicator electrode and measuring the resulting faradaic current. The voltage can be constant or vary continuously, and the current can be measured as a function of time or voltage [65, 66]. In the following, we shall only discuss methods in which the transport of electroactive species is limiting, i.e., methods in which the voltage is subjected to sudden, large amplitude variations and the transient current is then recorded. These techniques, known as chronoamperometry, are among the most widely used in electrochemistry.

As mentioned earlier, most experiments are carried out under conditions in which migration transport has been rendered negligible. If there is no convection in the solution, the current recorded is governed only by the classical laws of diffusion.

Single Potential Step

The simplest approach is to apply a single potential step at the interface, as shown in Fig. 11.11a. The initial potential E_i is chosen so that no electrochemical reaction is possible at the beginning. For example, for an oxidisable species, the idea is to fix a sufficiently negative potential compared with the standard potential of the redox couple, i.e., $n(E_i - E^0) \ll 0$, and then impose a potential step of amplitude $\Delta E = E_f - E_i$.

The current recorded is a transient current (see Fig. 11.11b), which may or may not be a function of the amplitude ΔE . When the final potential E_f is close to the standard potential E^0 of the electrochemical reaction, the current response does depend on the amplitude ΔE . Electron transfer is in this case the limiting process, both kinetically and thermodynamically. On the other hand, when the difference $n(E_f - E^0)$ increases, the current response tends to a limiting behaviour that is independent of ΔE , in which mass transport becomes the limiting process, with the electrochemical reaction entirely shifted to the interface.

By varying ΔE , for a fixed sampling time θ , one can obtain a voltamogram representing the transient current as a function of the potential applied

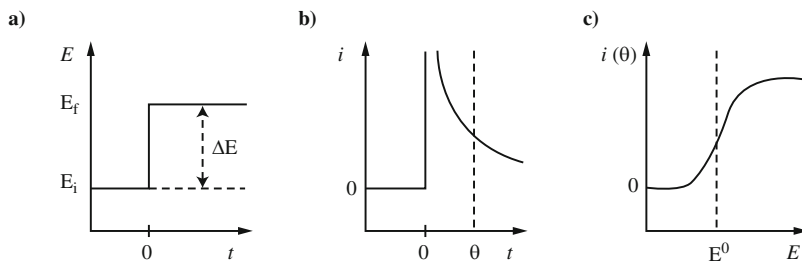


Fig. 11.11. Chronoamperometry by single potential step. (a) Potential step imposed on the electrode. (b) Current response at the electrode. (c) Current-sample voltammogram

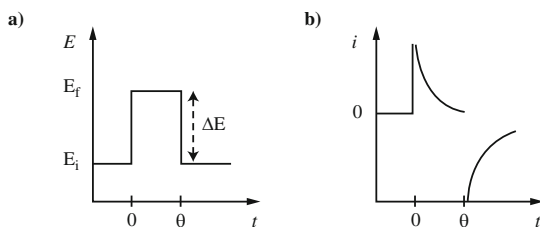


Fig. 11.12. Chronoamperometry by double potential step. (a) Typical waveform of potentials imposed on the electrode. (b) Current response at the electrode

to the electrode (see Fig.11.11c). The current-voltage curve then takes the form of a wave with the characteristics of current limitation by the two types of processes coming into play, viz., electron transfer and mass transport.

When experimental conditions are such that diffusion is the only mode of transport for $n(E_f - E^0) \gg 0$, the time dependence of the theoretical current response is found by applying Fick's second law:

$$\frac{\partial C}{\partial t} = D \nabla^2 C. \quad (11.11)$$

In the linear diffusion regime, i.e., near an infinite plane electrode, the equation becomes

$$\frac{\partial C}{\partial t} = D \frac{\partial^2 C}{\partial x^2}, \quad (11.12)$$

where x is the distance from the plane of the electrode. Solving this with suitable boundary conditions for this type of experiment, one arrives at the Cottrell relation:

$$i = \frac{nFSDC^0}{(\pi Dt)^{1/2}}. \quad (11.13)$$

The depletion of electroactive species in the vicinity of the electrode surface thus goes as $t^{-1/2}$, characteristic of a purely diffusion-limited process, with zero current in the large time limit, i.e., zero concentration gradient at large

times. Under practical conditions, this relationship is satisfied for large electrodes, i.e., millimetric or more. However, there are experimental limitations due partly to non-faradaic processes occurring on short time scales (capacitive current associated with the charge of the electrochemical double layer discussed on p.669) and partly to the long term effects of natural convection in the solution, whose influence on the mass transport of species can no longer be neglected. In the latter case, the current at large times approaches a nonzero limit (see the experimental examples given below, on p.680).

Considering a spherical diffusion field, as observed for smaller electrodes with spherical symmetry (micrometric or submicrometric), Fick's second law (11.11) is expressed differently. For example, for a hemispherical electrode of radius r_0 , it becomes

$$\frac{\partial C}{\partial t} = D \left(\frac{\partial^2 C}{\partial r^2} + \frac{2}{r} \frac{\partial C}{\partial r} \right), \quad (11.14)$$

where r is the distance from the centre of the electrode. Solving this equation, one arrives at the diffusion current

$$i = nFSDC^0 \left[\frac{1}{(\pi Dt)^{1/2}} + \frac{1}{r_0} \right], \quad (11.15)$$

in which the first term, observable on short time scales, corresponds to the Cottrell equation, and the second term is a constant satisfying

$$\lim_{t \rightarrow \infty} i = \frac{nFSDC^0}{r_0}. \quad (11.16)$$

This large time limit arises because the concentration gradient of the species does not vanish in the vicinity of the electrode as it does for a plane electrode. Indeed, the convergent spherical diffusion which dominates at large times guarantees a constant supply of species, with flow rate partly determined by the geometry and the small size of the electrode.

Double Potential Step

This is one of the basic techniques of electrochemistry, involving a current reversal. To begin with a species is electrically generated near the electrode, before reversing the direction of electrolysis to study the properties and/or chemical reactivity of this species. The potential applied to the electrode is a box or double-step potential of fixed width θ , where ΔE is such that $n|E_f - E^0| \gg 0$ (see Fig. 11.12).

Over the time lapse θ , a species is generated locally. Then, after reversal of the potential, this species is consumed at the electrode. During the two phases, the currents recorded are diffusion currents. For a plane electrode, when $t < \theta$, the current obeys the Cottrell relation (11.13), while for $t > \theta$, it is given by

$$i = \frac{nFSD^{1/2}C^0}{\pi^{1/2}} \left[\frac{1}{(t-\theta)^{1/2}} - \frac{1}{t^{1/2}} \right]. \quad (11.17)$$

Calculating the ratio of currents obtained with sampling times θ [given by (11.13)] and 2θ [given by (11.17)], it follows that

$$\frac{i(2\theta)}{i(\theta)} = 1 - \frac{1}{\sqrt{2}} = 0.293. \quad (11.18)$$

Experimentally, this test is often carried out to assess the chemical stability of electroactive species involved in electron transfer. Indeed, any discrepancy with this theoretical value will reflect the occurrence of certain complications in the kinetics or the reaction at the electrode. For example, if the species generated is not stable over a time scale of θ , and if it decomposes to yield non-electroactive products at the detection potential E_i , the ratio $i(2\theta)/i(\theta)$ will clearly be less than 0.293 [see (11.18)]. A quantitative treatment of these data is theoretically feasible for many complex processes, provided that the chemical nature of these processes can be identified experimentally.

11.2.2 Concentration Profiles

Diffusion Layer

The region of the solution in contact with the electrode, where the concentrations of electroactive species differ from those in the solution, is called the diffusion layer. It results from local depletion or enrichment in species due to the electrochemical reaction taking place at this electrode. From a theoretical standpoint, the solution of diffusion equations, whether analytic or numerical, perfectly describes the diffusion transport of the species, for different electrode sizes and geometries.

In the linear diffusion regime with a single potential step, solution of (11.12) leads to the following concentration profiles:

$$C = C^0 \operatorname{erf} \left[\frac{x}{2(Dt)^{1/2}} \right], \quad (11.19)$$

where erf is the error function. This relation describes the theoretical variations of the concentrations at the interface for an electroactive species that is entirely consumed at the electrode (see Fig. 11.13a). It should be remembered that the concentration gradient at the interface is a representation of the flow of this species, and hence an indication of the recorded current [see (11.13)]. An estimate of the thickness of the region where $C \neq C^0$ is given by the approximate Nernst model by considering a constant flow rate, equal to its value at the electrode surface, i.e., at $x = 0$:

$$J = -D \left(\frac{\partial C}{\partial x} \right)_{x=0} \approx -D \frac{C^0 - C^{\text{el}}}{\delta}, \quad (11.20)$$

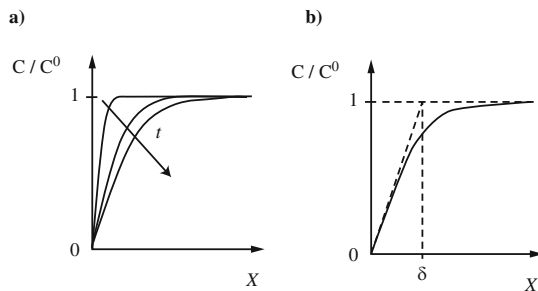


Fig. 11.13. Concentration profile at an interface of a species consumed in the linear diffusion regime. (a) Time dependence according to (11.19). (b) Approximate Nernst model, where δ is the thickness of the diffusion layer

where δ is the equivalent thickness of the diffusion layer (see Fig. 11.13b). In the linear diffusion regime, combining (11.9), (11.13), and (11.20) yields

$$\delta = (\pi Dt)^{1/2}. \quad (11.21)$$

For electrodes with large dimensions, i.e., millimetric or more, the thickness of the diffusion layer goes as $t^{1/2}$ and tends theoretically to infinity. However, in practice, the influence of natural convection in the solution at large times restricts its expansion to a limiting value δ_{conv} . Experimentally, this value does not exceed a few hundred micrometers (see below). It is imposed by the local hydrodynamics of the solution and is thus related to experimental conditions.

For smaller electrodes, i.e., micrometric or submicrometric, when the diffusion regime can be treated as a spherical diffusion regime, solution of (11.14) gives the concentration profile

$$C = C^0 \left\{ 1 - \frac{r_0}{r} \operatorname{erfc} \left[\frac{r - r_0}{2(Dt)^{1/2}} \right] \right\}, \quad (11.22)$$

where erfc is the complementary error function. Since $r - r_0$ is in this case the distance from the electrode surface, these concentration profiles are very similar to those obtained in the case of linear diffusion [see (11.19)]. The difference here lies in the term r_0/r . Combining (11.9), (11.15), and (11.20) shows that at small times, i.e., for thicknesses such that $(\pi Dt)^{1/2} \ll r_0$, δ varies in the same way as in the linear diffusion regime, given by (11.21). On the other hand, for $(\pi Dt)^{1/2} \gg r_0$, the diffusion layer tends toward a fixed thickness, comparable with the size of the electrode, i.e., $\delta = r_0$.

This last condition, viz., $(\pi Dt)^{1/2} \gg r_0$, is one of the fundamental characteristics [67] of ultramicroelectrodes [68–70]. These are electrodes with micrometric or submicrometric dimensions for which a pure diffusion steady state is quickly set up at their surface. It should be stressed that this class of electrodes is not defined solely in terms of the intrinsic electrode dimensions, but also relative to the experimental conditions used. Indeed, a steady state is only

purely diffusional if there is no significant influence of natural convection on the establishment of diffusion layers. The limiting values of δ in the vicinity of these electrodes must then be small enough compared with the limiting thickness δ_{conv} imposed by the hydrodynamics of the solution. In other words, to be immune from the effects of convection, these electrodes must have dimensions much smaller than δ_{conv} . Under typical conditions of electrochemistry, this implies electrodes with dimensions smaller than or equal to about ten micrometers (see the experimental examples discussed below).

Measurements Using Ultramicroelectrodes

Owing to their small dimensions, ultramicroelectrodes have fundamental properties that offer many advantages in electrochemistry, allowing them to be used in what are usually described as extreme conditions. One of the main advantages concerns the spatial resolution of measurements, explained by the very thin diffusion layers. The current recorded is only sensitive to physicochemical processes operating on this scale, directly or indirectly coupled to the electrochemical reaction. This intrinsic feature means that they can be used as local probes or sensors for all electroactive species present in the medium under investigation, and in particular, as we shall see below, for species relevant in biology (see Sect. 11.2.4). The spatial resolution of the measurements is imposed by the size of the indicator electrode used as a probe, and it will improve as the size is made negligible compared with the region of the solution being explored. This property opens the way to many analytical applications on the scale of living cells.

Two other advantages follow directly from the small dimensions of the electrodes [67–70]. To begin with, consider the temporal resolution of these recordings, i.e., the time constant $R_s C_d$ of the electrochemical cells [see (11.1)]. The product $R_s C_d$ depends on the size of the microelectrode and in fact decreases as the electrode gets smaller ($R_s C_d$ is proportional to r_0 for a hemispherical electrode as soon as r_0 is bigger than a few nanometers). At short times, the capacitive current corresponding to the charge of the electrochemical double layer is then minimised with respect to the faradaic current, and this property can be exploited to monitor fast biological or chemical processes associated with electron transfer, e.g., the release of neurotransmitters by a living cell (see p. 684). The second advantage results from a significant reduction in the Ohmic drop $R_s i$ [see (11.4)] in the transient or steady state regime ($R_s i$ is proportional to r_0 for a hemispherical electrode in the transient regime and decreases toward a constant limit at large times). This feature is extremely interesting for the study of resistive media including little supporting electrolyte, or concentrated, even pure media, leading to the recording of high current densities.

The experimental examples discussed below are intended to illustrate the potential of ultramicroelectrodes when they are used as local concentration probes [63, 67–70] during physicochemical processes. Indeed, amperometric

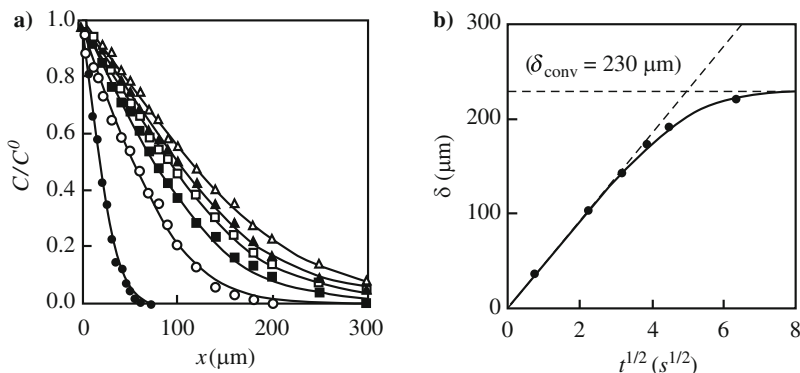


Fig. 11.14. Oxidation of $\text{Fe}(\text{CN})_6^{4-}$ 10 mM in 1 M KCl at a platinum electrode of radius 0.6 mm. (a) Experimental concentration profiles of $\text{Fe}(\text{CN})_6^{3-}$ recorded with a microelectrode of radius about $2 \mu\text{m}$ (data represented by *symbols*) and theoretical concentration profiles (*continuous curves*) calculated for $D = (5.7 \pm 0.5) \times 10^{-6} \text{ cm}^2 \text{ s}^{-1}$ and $\delta_{\text{conv}} = 230 \mu\text{m}$. Concentrations were recorded for different polarisation times (*from left to right* 0.5, 5, 10, 15, 20, and 40 s). (b) Change in the thickness δ of the diffusion layer estimated from the data in (a). The *continuous curve* in (b) represents the theoretical behaviour as predicted by the natural convection model [63]

detection provides an easy way of relating current variations to local concentration gradients. The technique used to establish concentration profiles consists in placing an ultramicroelectrode inside the diffusion layer generated in the vicinity of an active surface, then moving the ultramicroelectrode through the layer in order to map the values at different points. In the examples discussed, the active surface under examination is a second electrode, the working electrode, which is bigger and held at a fixed voltage.

Figure 11.14 shows a first example of amperometric detection carried out in the vicinity of a millimetric electrode in the linear diffusion regime, under conditions in which diffusion is the only limiting process at short times [63, 71]. These measurements represent the time dependence of the change in concentration profile of the electrogenerated species in a direction orthogonal to the plane of the working electrode (see Fig. 11.14a). In agreement with theoretical predictions, the thickness of the diffusion layer varies over short times in a way that accords perfectly with (11.19) (see Fig. 11.14b). However, this is not the case for times longer than a few seconds, where a steady, hydrodynamic or convective diffusion regime is gradually set up, controlled by the natural convection of the solution. In the case of Fig. 11.14, the thickness of the diffusion layer tends toward a limit whose value turns out to be fixed by hydrodynamic and operational conditions (see Fig. 11.14b). The transition between the two regimes can be accounted for by a suitable model, integrating the effects of natural convection in the solution [63].

The resolution of the measurements can be greatly improved by using sub-micrometric probe electrodes. Under these conditions, the currents detected correspond to a sampling volume of a few femtolitres or less. The second example chosen here (see Fig. 11.15) represents a series of concentration profiles established in the vicinity of a micrometric electrode. The probe is a nanometric electrode, displaced in the three space directions by a scanning electrochemical microscope (SECM) stage [76]. Mapping consists in a series of profiles obtained by scanning planes parallel to the surface of the working electrode. These profiles correspond to the steady diffusion regime, rapidly attained in the vicinity of an ultramicroelectrode (see below). The thicknesses of the resulting diffusion layers are indeed of the same order of magnitude as the characteristic dimensions of the working electrode (see Fig. 11.15b). In the present case, natural convection in the solution has a moderate influence, as expected for the size of the working electrode.

11.2.3 Conclusion Regarding Faradaic Electrochemical Detection

We have just seen in the last section that an electrode held at a sufficient potential can detect the existence of a concentration of some electroactive species in the immediate neighbourhood of its surface. In the case of an ultramicroelectrode, this will only detect molecules in a volume of radius comparable with the electrode radius, whereupon it plays the role of a genuine confocal microscope. Information obtained about the chemical composition of the micro- or nanoscopic volume under investigation is entirely faradaic, since it corresponds only to the strength of the electrochemical reactions occurring at the interface [see (11.5)].

A change in the detected current informs directly about a corresponding change in the concentration of the electroactive species in the detected volume, as illustrated in Fig. 11.14a. As a consequence, the time dependence of the steady state current exactly reflects the external flows entering or leaving the micro- or nanoscopic sampling volume, i.e., the flows which control the concentration of the target electroactive species.

These flows may have a biological origin, as we shall see in the following. Local analytical methods based on currents are thus essentially different from those based on electrophysiological methods (see Sect. 11.1) or electrophoretic methods (see Sect. 11.4), which for their part reflect local ion displacements without associated electrochemical reaction. Moreover, due to their kinetic nature, faradaic currents are only affected by the local concentration and not by the amount of electroactive species [see (11.5) and Fig. 11.14a], which is an extraordinary advantage when studying living systems, since biological phenomena generally lead to the emission of very small amounts of species (femtomoles 10^{-15} to zeptomoles 10^{-21}), but locally present in very high concentrations.

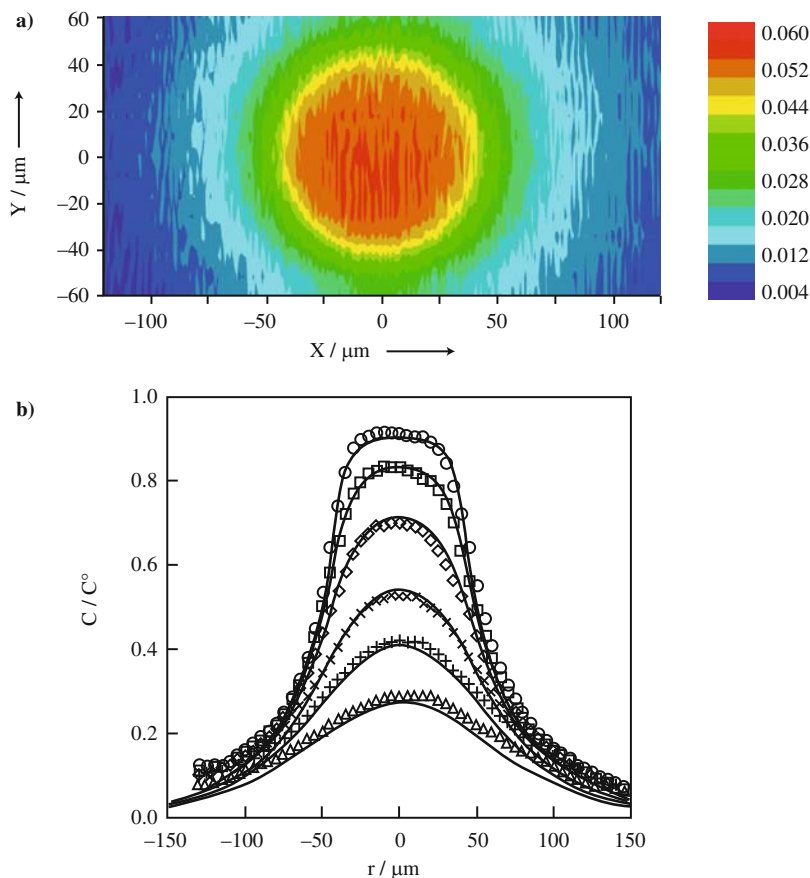


Fig. 11.15. Reduction of $\text{Ru}(\text{NH}_3)_6\text{Cl}_3$, 5 mM in 0.1 M KF at a platinum microelectrode with radius $40\text{ }\mu\text{m}$. **(a)** 2D map of currents measured by a nanoelectrode of radius 80 nm near the microelectrode in the steady state regime. The nanoelectrode scans a plane parallel to the surface of the microelectrode, at a vertical distance $z = 5\text{ }\mu\text{m}$ and scan rate $8\text{ }\mu\text{m s}^{-1}$. **(b)** Experimental concentration profiles (*symbols*) established at different vertical distances z above the surface of the microelectrode: $z = 5$ ($\circ$), 9 ($\square$), 16 ($\diamond$), 28 ($\times$), 40 ($+$), and $57\text{ }\mu\text{m}$ ($\triangle$). *Curves* show theoretical profiles calculated using the natural convection model with $\delta_{\text{conv}} = 135\text{ }\mu\text{m}$ [64]

In the rest of this section, we illustrate these principles by presenting two studies of fundamental biological processes on the scale of a living cell, carried out using ultramicroelectrodes.

11.2.4 Artificial Synapses: Biological Applications to Single Cells

Vesicular Exocytosis of Neurotransmitters

Historically, the first field of application for electrochemical measurements on single cells was the analysis of vesicular exocytosis of neurotransmitters [77–79]. This biological mechanism is what allows the transmission of information between neurons in the so-called chemical synapses (contrasted with electrical synapses or ion channels transmitting a signal by successive membrane depolarisation), but also the release of hormones into the blood. Literally, exocytosis means ‘to empty the contents’. Indeed, neurotransmitter molecules such as dopamine, serotonin, adrenaline, GABA, and so on, are packed into vesicles contained within the emitting cell. These vesicles, generated by the endoplasmic reticulum, are bounded by a phospholipid membrane analogous to the cell membrane and often contain an anionic polyelectrolytic protein matrix, in which the cationic neurotransmitter is trapped in high concentrations, typically several hundred millimoles [80,81]. The vesicles are carried toward the cell membrane by the cytoskeleton and, following a suitable stimulus leading very often to an arrival of calcium ions in the intracellular medium, they subsequently fuse (totally or partially) with the membrane, rapidly delivering their contents as messengers into the extracellular medium [82–85].

Most experimental work on exocytosis has been carried out on the model of the chromaffin cells from the adrenal glands, rather than on neurons. Indeed, neurons are always particularly difficult to isolate and cultivate. Furthermore, the released neurotransmitters can only be detected outside the most important emission region, i.e., outside the synaptic bud. In contrast, the chromaffin cells, normally involved in controlling the flow of adrenaline (and other hormones) in the blood, are easily isolated and cultivated to study their secretion of catecholamines by vesicular exocytosis. Indeed, these cells, measuring around 10 μm in diameter, contain in their cytoplasm a large number (several thousand) vesicles with average diameter 0.3 μm , and each vesicle contains in its turn some 3 million catecholamine molecules (mainly adrenaline, but also noradrenaline and dopamine).

The activation of the opening of membrane calcium channels, by local depolarisation (injection of a solution of K^+ ions at 55 mM) or by a flow of divalent ions (a solution of Ba^{2+} ions at 2 mM), causes a transient increase in the cytoplasmic calcium concentration. This instantaneously triggers ($t \ll$ millisecond) the formation of supramolecular connections of protein type, which set up between the cell membrane and the membrane of the underlying vesicle (see Fig. 11.16, phases I and II). This connection leads to the formation of a transmembrane fusion pore which establishes a contact between the extracellular medium and the medium inside the vesicle, thereby initiating the release of neurotransmitters stored in the vesicle. Although there is still some debate in the literature about the exact nature of the fusion pore (constituted

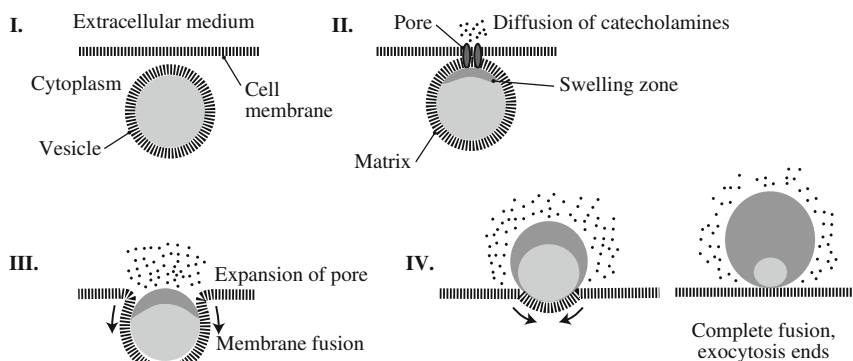


Fig. 11.16. Schematic view of the different stages in exocytosis of a catecholamine-secreting vesicle (containing a dense matrix)

purely by proteins or lipids, or a combination) [86–89], it is known to behave like a transient channel with initial diameter in the nanometer range (value determined by electrophysiological measurements, as described in Sect. 11.4, or by the method described here [90]). However, during exocytosis, this channel does not behave like the majority of ion protein channels, since it can very quickly become unstable (on a millisecond time scale), destructuring and causing the fusion of the cell and vesicle membranes. There are certainly several reasons for this instability, as one can see by considering the fact that the pore and vesicle structures can vary between the different types of cells releasing neurotransmitters and the fact that several biological regulatory mechanisms can be involved. In the specific case of the chromaffin cells, recent work based on the physicochemistry of polyelectrolytic matrices and membranes has confirmed the hypothesis that the vesicular matrix destructures and swells under the effect of the exchange of outward catecholamine cations replaced by monovalent cations (Na^+ , K^+) and water molecules entering the vesicle [80, 90, 91]. The pressure on the vesicular membrane and the mechanical tension on the pore structure due to swelling of the matrix eventually lead to the rupture of the fusion pore and fusion of the membranes (see Fig. 11.16, phase III). This fusion is then pursued irreversibly until the vesicular membrane has been completely exposed to the external medium (see Fig. 11.16, phase IV), at the surface of the cell membrane. The catecholamine molecules can then diffuse out of the swollen matrix and rapidly reach their target in the cell environment (or be transported in the blood from the adrenal gland).

The different phases of the exocytosis were first revealed by electron microscopy and electrophysiological methods, including measurement of the membrane electrical capacitance by the patch-clamp technique with a glass microelectrode (as described in Sect. 11.1) [92, 93], then more recently by evanescent wave fluorescence spectroscopy (see Chap. 7) [94, 95]. But while the descriptive aspects of these methods has become truly exceptional, their

kinetic resolution is not sufficient for observing the dynamics of phase II and even less so for identifying the processes causing the transition between fusion pore (phase II) and membrane fusion (phase III). The development of ultramicroelectrodes and demonstration of their particular properties over the last two decades (see p. 680) have led to their use in an artificial synapse configuration, whereupon it has been possible to measure by amperometry the flow of neurotransmitters emitted by a single cell during a series of vesicular exocytosis processes. At the present time, the resolution of amperometric measurements of these flow rates for released molecules is of the order of a thousand molecules per millisecond. The current detected at the ultramicroelectrode then very faithfully reproduces the vesicular emission rate in real time, and more particularly, the release rate by the nanoscale pore before its expansion and the large-scale release of messenger molecules. This phase can be identified by the presence of a small but steady increase called a foot, just prior to the exocytosis peak (see Fig. 11.17) [93]. Moreover, amperometric measurements have shown that there are several shapes for these foot events, corresponding to different release dynamics via the fusion pore [96,97].

The great accuracy of these measurements recently led to quantitative tests of theoretical models describing vesicular exocytosis. Indeed, it has become possible to exploit each phase of the process (phases II, III, and IV) in a quantitative manner and deduce the dynamical behaviour of the vesicle membrane [90,98]. The release rate of catecholamines, observed in real time by virtue of the short response time of these ultramicroelectrodes (see p. 680 for theoretical explanation), is treated as the convolution product of two kinetic processes, viz., the rate at which the pore opens and the matrix is uncovered on the one hand, and the diffusion kinetics of the catecholamines in the matrix on the other. The hypothesis that changes in the state of the vesicular matrix control these processes has been confirmed, more particularly for exocytosis from chromaffin cells. The exchange between vesicular catecholamine cations and monovalent cations from the external medium necessarily leads to the destructuring and swelling of the matrix. As long as the vesicle membrane imprisons the matrix, swelling is prevented geometrically and therefore has the effect of increasing the internal pressure on the membrane. This pressure has direct consequences on the structure of the pore connecting the vesicle and cell membranes. This situation persists until the surface tension energy of the pore balances its edge energy. When this moment is reached, the pore becomes unstable and must burst, with irreversible fusion of the membranes. The expansion rate of the pore is then controlled for the main part by viscous dissipation of its surface tension energy. The pore diameter thus increases almost exponentially, exposing more and more surface area of the matrix to the solution, which in turn means that there is a large increase in the release rate of catecholamines (phase III).

The process continues until the matrix no longer exerts any pressure on the inner face of the vesicular membrane, which therefore relaxes its surface tension. The transfer of the vesicular membrane into the cell membrane then

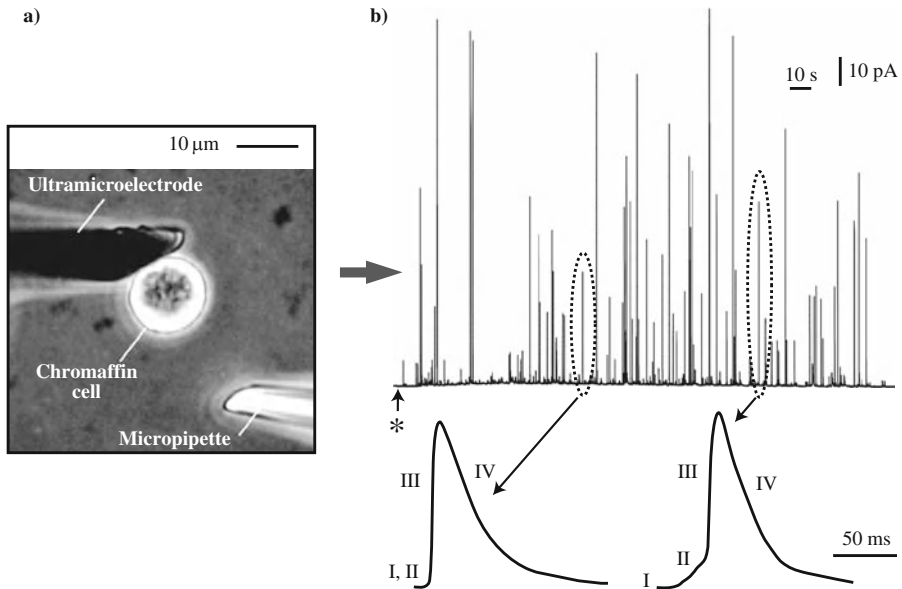


Fig. 11.17. Ultramicroelectrode detection of vesicular exocytosis of catecholamines by a bovine chromaffin cell. (a) Optical microscope observation of the artificial synapse. The ultramicroelectrode is held at a distance $d < 1 \mu\text{m}$ from the cell membrane and a microcapillary is placed in the neighbourhood of the cell to inject a stimulus solution (Ba^{2+} 2 mM, at the time marked by an *asterisk* on the plot). (b) Amperometric plot ($E = +650 \text{ mV}$ vs. Ag/AgCl) recorded during cell secretion. Each peak corresponds to exocytosis of a vesicle, describing its dynamics perfectly. The transient phase of the fusion pore can be observed directly in about 30% of cases by the current foot just prior to the peak (*right-hand peak*, compared with the *left-hand peak* which has no foot)

occurs with almost no change in the tension energy. At this point in the process, it remains only to take into account the lower energy of the toroidal structure of the membrane at the join with the vesicle. This energy, proportional to the perimeter of the torus, can be treated as an edge energy and therefore tends to dissipate when the radius is reduced. The viscous dissipation of the energy thus causes a linear variation of the radius of the junction between the membranes. In conclusion, the matrix plays a decisive role, first by causing the rupture of the initial pore, then by maintaining the membrane under tension at the beginning of its fusion with the cell membrane. At the end of phase III, the matrix has usually only released 20–30% of its contents and is completely exposed to the extracellular fluid. The remainder of the catecholamine molecules are therefore emitted by simple spherical diffusion from the matrix.

On the basis of these ideas, models have been made of the amperometric exocytosis peaks for chromaffin cells [90,99]. They have been used to extract

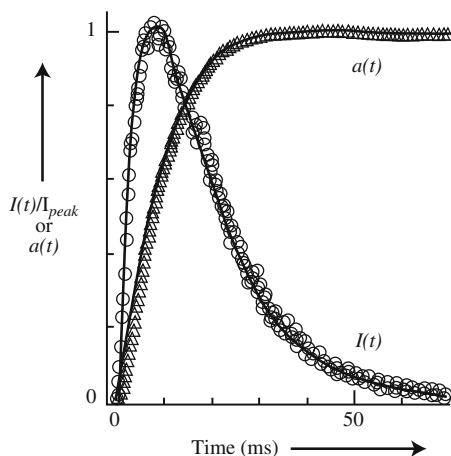


Fig. 11.18. Amperometric exocytosis peak detected during the secretion of catecholamines by a chromaffin cell (*continuous curve*) and theoretical signal (*circles*) calculated from the curve $a(t)$ (sigmoidal shape, *triangles*), representing the change in the surface fraction of the vesicular membrane integrated into the cell membrane. $a(t)$ is calculated at each instant of time from the experimental current by deconvolution of the diffusion processes [90]

the kinetics of pore opening and exposure of the matrix, and also the diffusion kinetics of the catecholamines in the matrix (see Fig. 11.18). Comparing the experimental peaks with model peaks created from the convolution of these kinetic and diffusional phenomena demonstrates the general validity of these hypotheses concerning exocytosis for vesicles containing a dense-core matrix, such as can be observed in the chromaffin cells of the adrenal glands, some neurons, or the mast cells of the immune system. Future developments in electrochemical analysis, focusing on each exocytosis event, and associated theoretical modelling should open new ways of understanding pathological processes perturbing the transmission of nerve or hormone information by exocytosis.

Detecting the Active Species of Oxidative Stress

Over the last ten years, ultramicroelectrode analysis of the living cell has found a particularly important field of application in the study of the release of active species leading to cell oxidative stress. Oxidative stress is usually defined as any process causing change, destructuring, degeneracy, or mortality in the cell, following the uncontrolled production of oxidising, peroxidising, nitrating, nitrosing, and similar species [100–103]. Such species may act at various sites, including DNA (at different points), membrane lipids, certain amino acids in proteins, and enzymes [103–105]. The modifications of these macromolecules are normally repaired by specialised enzymes, e.g., DNA

repair complexes, or prevented by direct elimination of the harmful species by means of antioxidants and specific enzymes, e.g., catalase, SOD, peroxidases, etc. However, when these modifications do occur and repeat sufficiently often, they can eventually lead to a loss of functions in an organ, tissue, or even a whole organism. It has now been shown that oxidative stress may be at the origin of, or contribute to, a large number of human pathologies, such as some cancers (skin, liver, lungs, etc.), viral infections, neurodegenerative diseases (Alzheimer), cardiovascular disease, and others [103,106–108].

The analytical methods used in this field of biological and medical investigation were until recently of an indirect nature and based on the detection of long-term degradation products from oxidative stress. The main benefit of single-cell analysis has thus been the possibility of detecting the initiating species of oxidative stress and characterising their release mechanisms. These species resulting from the oxygen metabolism form a family of compounds derived from enzymatic production, by different sources, of two species of low molecular weight and carrying an unpaired electron: the superoxide ion $\text{O}_2^{\bullet-}$ and nitrogen monoxide $\text{NO}^\bullet$. These primary species are in themselves only weakly oxidising (chemically, they are reducing). Their role is rather as a messenger, regulating certain metabolic activities. Hence, the first biological function of $\text{NO}^\bullet$ to have been demonstrated is control over dilation of vessels and arteries by release from smooth muscle cells surrounding the endothelial cells (which produce $\text{NO}^\bullet$) that form the vascular duct [109]. The superoxide ion for its part mainly targets proteins regulating the development and proliferation of cells. However, $\text{O}_2^{\bullet-}$ dismutates quickly into hydrogen peroxide (H_2O_2) and leads via the so-called Fenton and Haber–Weiss reactions to a highly reactive species, the hydroxyl radical ($\text{OH}^\bullet$), which initiates lipid peroxidation and changes in the DNA bases [105,110]. In addition, the superoxide anion reacts very quickly with nitrogen monoxide to form peroxyntirites (ONOO^- and its acidic form ONOOH), leading to the nitration of amino acids, but also giving rise to a great many other, even more reactive species, e.g., $\text{NO}_2^\bullet$, NO_2^+ , ONOOCO_2^- , and so on, with a strong nitration and nitrosing tendency. It is thus easy to understand the relevance of such analysis for biomedical research on oxidative stress, or the pharmacological control of its effects, which aims to quantify and identify in a single cell and in real time the primary species leading eventually to metabolic or pathological modifications.

The production of $\text{O}_2^{\bullet-}$ and $\text{NO}^\bullet$ or their derivatives by a living cell has been analysed using the artificial synapse method for different cell types, in particular skin cells (fibroblasts and keratinocytes) [111–113] and cells in the immune system (lymphocytes, monocytes, macrophages) [112,114,115]. We shall exemplify the possibilities offered by electrochemical analysis by discussing the study of the release by human skin fibroblasts of the superoxide ion and nitrogen monoxide. Fibroblasts are the majority cells in the dermis. Until recently they were considered as simple connective cells, a more or less passive support for the keratinocytes making up the epidermis. However, recent studies have shown that UVA radiation in sunlight penetrates the skin

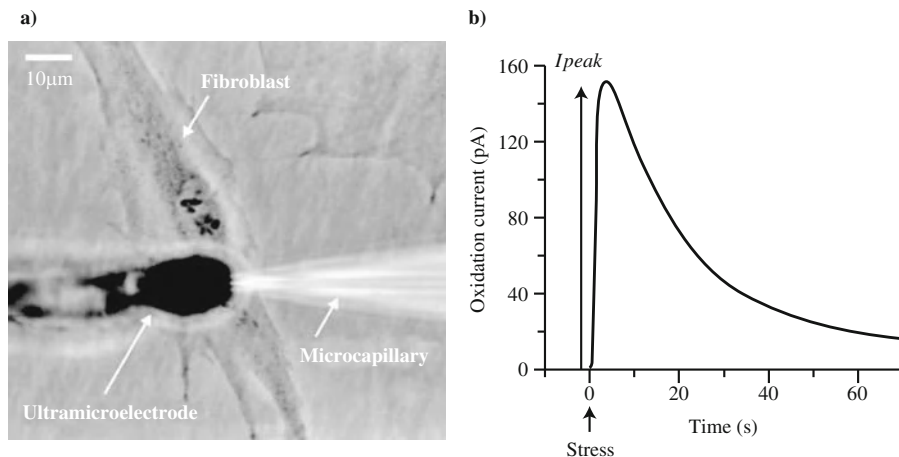


Fig. 11.19. Ultramicroelectrode detection of species produced during oxidative stress of a human fibroblast. (a) Optical microscope observation of the artificial synapse made from a fibroblast (from human skin, in culture at the bottom of a Petri dish) and a platinum-coated carbon ultramicroelectrode placed at a distance $d = 5\ \mu\text{m}$ from the cell membrane. The tip of a glass microcapillary placed between the electrode surface and the cell causes a short stress in the cell membrane by local depolarisation and thereby activates the enzymes producing $\text{O}_2^{\bullet-}$ and $\text{NO}^\bullet$. (b) Amperometric plot ($E = +650\ \text{mV}$ vs. ECSS) recorded during the fibroblast response [113]. Reproduced with the kind permission of the Royal Society of Chemistry

as far as the dermis and has important effects on the oxidative metabolism of the fibroblasts, whose subsequent modification may initiate cancers [116].

Studies of oxidative stress on the scale of a single fibroblast have used modified ultramicroelectrodes, as often happens for the analysis of biologically relevant compounds where one seeks high selectivity with regard to many possible interfering species. Carbon ultramicroelectrodes developed for amperometric measurements of exocytosis, described earlier, have been modified by a micrometric or submicrometric deposit of platinum (platinum black), which provides a large sensitive area with high catalytic activity, especially with regard to several derivatives of $\text{O}_2^{\bullet-}$ and $\text{NO}^\bullet$ [113, 117]. The detection selectivity between these different derivatives is achieved by the choice of potential applied to the electrode in amperometry, since the electrochemical responses of these species (see Fig. 11.11c) on the platinum-coated electrodes are quite distinct (differences in E_0 and electron transfer constants). The first studies on the scale of a single fibroblast showed that, during a brief activation by physical stress on the cell, caused by local depolarisation of its membrane by means of a microcapillary, rapid, high intensity bursts of species flow were detected at the ultramicroelectrode, positioned a few micrometers from the cell surface to avoid distortion by back-reaction of oxide compounds on the cell [111, 118] (see Fig. 11.9). It was shown that this cell secretion corresponded

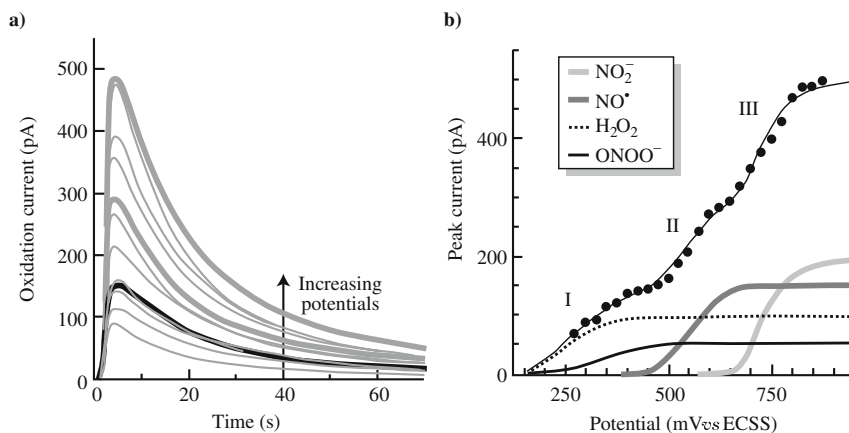


Fig. 11.20. Determining the species produced during oxidative stress of cultured human fibroblasts. (a) Study for different ultramicroelectrode potentials ($300 \leq E \leq 850$ mV vs. ECSS) of the oxidation response recorded by amperometry. (b) Variation of the maximal current of the amperometric spikes in (a) as a function of the detection potential (*bullets*). This curve reveals three distinct waves labelled I, II, and III, which have been compared with oxidation waves of solutions of H_2O_2 , ONOO^- , NO_2^- , and $\text{NO}^\bullet$ obtained in vitro with the same ultramicroelectrodes. The amplitude of the waves of the four compounds has been adjusted to correspond to the amplitude of the cell response, allowing construction of a global voltammogram (*thin black curve*) that agrees extremely well with the points from the ex vivo measurements [113]

locally, i.e., in the membrane, to high concentrations of species (in the millimolar range). However, the kinetics of the cell response does not correspond to release by simple diffusion of derivatives of $\text{O}_2^{\bullet-}$ and $\text{NO}^\bullet$ from some initial receptacle such as a vesicle, but is rather the result of an active enzymatic production process of these species simultaneous with their release. To see this, one only has to compare the widths of the exocytosis spikes, limited by a diffusion process (see Fig. 11.17), with those of the peaks detected here: there is a difference of 3 orders of magnitude.

The kind of species released during cell oxidative stress has been determined on the basis of amperometric recordings at different potentials and a large cell sample (see Fig. 11.20a) [113]. In this way it was possible to represent the variation of the signal amplitude with the measurement potential and produce an experimental curve equivalent to a steady-state voltammogram (see Fig. 11.11c). This type of current–voltage curve gives an electrochemical signature of the electroactive species detected at the ultramicroelectrode surface and can thus be compared with the signature, in this same mode of electrochemical analysis, of solutions of species assumed to be produced in the early stages of oxidative stress. The curve obtained from cell measurements exhibits three identifiable waves of oxidation (numbered I, II, and III in Fig. 11.20b),

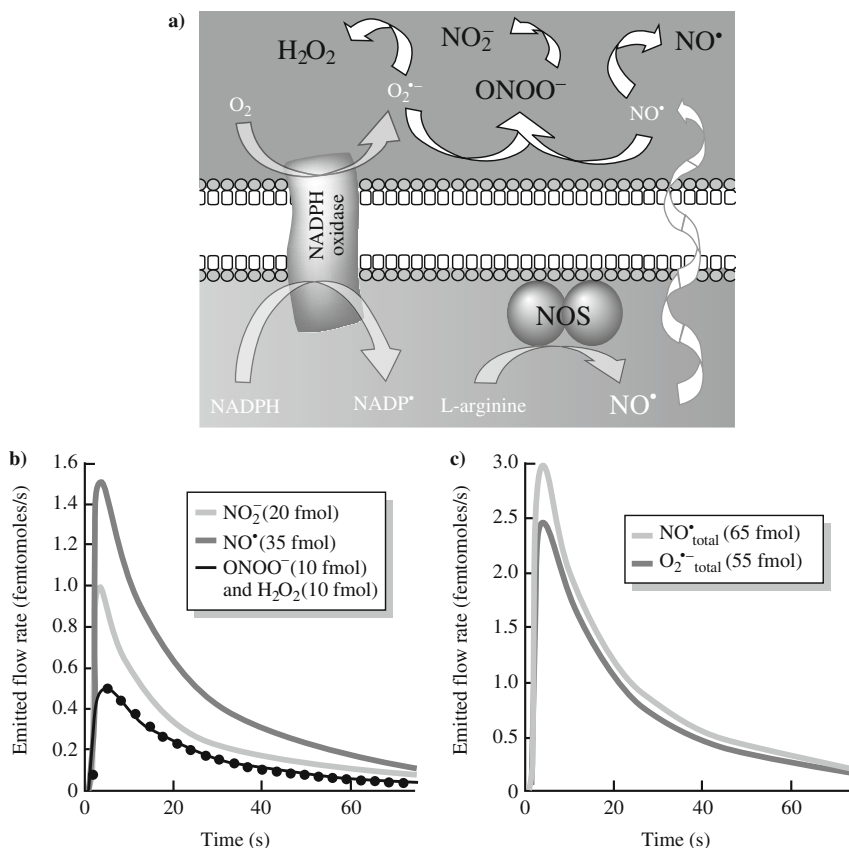


Fig. 11.21. (a) Proposed reaction scheme for the production of oxidative stress species by fibroblasts via the activation of two enzymatic systems, NADPH oxidase and NO synthase, located in the cell membrane. (b) Characterising the flow of each species (H_2O_2 , ONOO^- , NO_2^- , and $\text{NO}^\bullet$) released by a fibroblast during its response to oxidative stress. The flows have been calculated using Faraday's law from changes in the amperometric current (see Figs. 11.20a and b) corresponding to each compound. (c) On the basis of the reaction diagram in (a), an estimate of the initial production flows of the primary species $\text{O}_2^{\bullet-}$ and $\text{NO}^\bullet$ has been made. The total amounts of species produced are given in brackets in (b) and (c), found by integrating each flow curve with respect to time [117, 119]

corresponding therefore to the detection of at least three products. Comparison with in vitro responses of the derivatives of $\text{O}_2^{\bullet-}$ and $\text{NO}^\bullet$ revealed that waves II and III came from the detection of nitrogen monoxide and nitrites, respectively. Wave I does not show the characteristics of a simple oxidation wave due to a single type of compound, but is the result of a mixture, in a close range of potentials, of two oxidation waves due to hydrogen peroxide and peroxynitrite. The sum of these in vitro responses, for concentrations adjusted

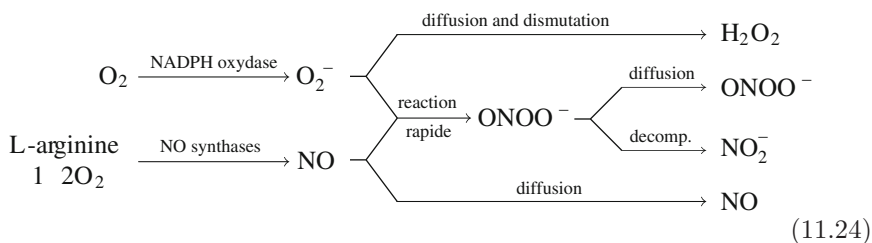
with respect to currents detected on cells, leads to the voltammogram shown by the continuous black curve in Fig. 11.20b, which corresponds well with the *ex vivo* measurement points. These results thus show that the activation of oxidative stress in an isolated fibroblast leads to simultaneous release of a cocktail of compounds that can be selectively determined and quantified by amperometry [113, 119].

Various biochemical tests (specific inhibitors and enzymes) have corroborated the hypotheses put forward concerning the chemical nature of the species and it has been possible to identify with certainty two enzyme systems as production sources, viz., NADPH oxidases and NO synthases, synthesising the two primary species $\text{O}_2^{\bullet-}$ and $\text{NO}^\bullet$, respectively (see Fig. 11.21a) [117]. The process induced by membrane depolarisation and variations in calcium ion concentration between the intra- and extracellular media thus leads to a joint activation of the two enzyme systems and a simultaneous production (on the time scale of the measurements, i.e., $t_{\text{response}} < 100 \text{ ms}$) of the two primary species. This leads to the release of a flow containing all the derivatives observed at the ultramicroelectrode as a current peak, which reaches its maximum in less than one second (see Fig. 11.20a). In addition, the artificial synapse configuration ensures maximal collection of the flow of released species since the sensor is positioned very close to the biological source. As explained in the introduction, this minimises the diffusion time of the electroactive species (see p. 684) and thereby limits any distortion in the kinetics of the observed processes. On the basis of these principles, the current peaks are directly converted by Faraday's law into the flow rate Φ of species released by the cell (see Fig. 11.21b):

$$\Phi_j = \frac{I_j}{n_j F} . \quad (11.23)$$

For each species (labelled by j), this conversion takes into account the number n_j of electrons exchanged during its electrochemical oxidation at the ultramicroelectrode surface, and in the end gives a more representative value than the current measurement of the amplitude of exchanges between the cell and its environment during its reaction to stress.

From the flow curves of the species released by the fibroblasts (see Fig. 11.21b), it has been possible to reconstruct the initial production flows of the two primary species $\text{O}_2^{\bullet-}$ and $\text{NO}^\bullet$ (see Fig. 11.21c). Indeed, all the detected species were produced directly from the superoxide anion and nitrogen monoxide, or from their combination product the peroxynitrite anion:



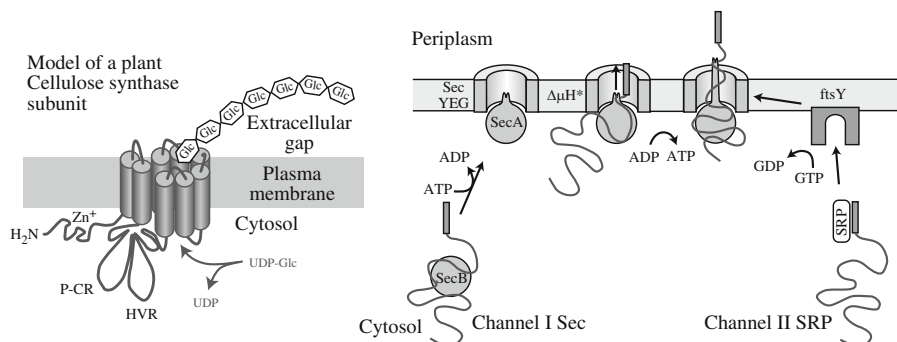


Fig. 11.22. Biological examples of molecular transport through nanopores. *Left:* Synthesis of cellulose. *Right:* Secretion of proteins through a bacterial membrane

The primary flow rates can thus be constructed by the following expressions:

$$(\Phi_{\text{O}_2\bullet-})^{\text{prod}} = 2(\Phi_{\text{H}_2\text{O}_2})^{\text{meas}} + (\Phi_{\text{ONOO}^-})^{\text{meas}} + (\Phi_{\text{NO}_2^-})^{\text{meas}}, \quad (11.25)$$

$$(\Phi_{\text{NO}\bullet})^{\text{prod}} = (\Phi_{\text{NO}\bullet})^{\text{meas}} + (\Phi_{\text{ONOO}^-})^{\text{meas}} + (\Phi_{\text{NO}_2^-})^{\text{meas}}. \quad (11.26)$$

Moreover, each flow rate curve can be integrated to quantify the total release or production of each species by the cell (values in brackets in Figs. 11.21b and c). The values determined here, between ten and a few tens of femtomoles, are considered to be infinitesimal on the macroscopic scale and would be undetectable by any conventional spectroscopic or analytic technique. However, under the conditions provided by the artificial synapse, detection occurs in a volume equivalent to the volume of the immediate environment of the cell, a space in which the cell will carry out most of its exchanges with other cells or molecular targets. This ‘living space’ of the cell, a few picolitres or less, thus receives an amount of messengers, or defense molecules in the present case of oxidative stress, corresponding to very considerable changes in concentration (micro- to millimolar). Insofar as faradaic electrochemical currents reflect chemical kinetics [see (11.5)], they are sensitive to the concentration, i.e., the probability of contact between the electroactive species and the electrode. This explains why the artificial synapse method can detect and resolve processes involving such tiny amounts of messengers, provided that their concentration in the artificial synapse gap is high. In fact, this is the same strategy as the one selected by nature in neuronal or neuromuscular synapses, for example.

Conclusion

These two examples of analysis exploiting the artificial synapse created between a solid ultramicroelectrode and a living cell illustrate the tremendous potential of this type of sensor and demonstrate the relevance of micro- and

nanometric electrochemical techniques for the study of a great many biological processes involving dynamical exchanges between electroactive messenger molecules. The current resolution limit for amperometric measurements, of the order of 1,000 molecules/millisecond, allows a very detailed analysis of the flows exchanged by biological entities, whether they be of micrometric dimensions like the cell, or nanometric dimensions like the fusion pore during vesicular exocytosis of neurotransmitters. The future development of these methods will certainly lead to improvements in this resolution and the possibility eventually to detect the activity of a single biological molecule such as a membrane enzyme.

11.3 Macromolecular Transport Through Natural and Artificial Nanopores. Electrical Detection

11.3.1 Introduction

Many natural or technological processes involve the transport of colloidal objects or macromolecules through channels or pores of very small radius. In chemistry, the first example is provided by filtration, joined by steric exclusion chromatography and capillary electrophoresis. In biology, one could mention the examples of biopolymer synthesis (DNA, RNA, proteins, and polysaccharides), extranuclear transport of messenger RNA [120], the translocation and secretion of proteins [121], and the infection of a cell by certain viruses [122] (see Fig. 11.22). The biological examples are particularly fascinating due to their complexity, their accuracy, and the selective and often active nature of the transport.

It is difficult to study the above processes on the molecular scale [123, 124], and it was a genuine revolution in 1996 when John Kasianowicz, Eric Brandin, Daniel Branton, and David Deamer [125], together with George Church and Richard Baldarelli [126], showed how to observe in a direct manner the passage of a single DNA molecule or single-strand RNA molecule through a nanometric protein pore inserted in a lipid bilayer, using an extremely simple electrical technique (see Fig. 11.23). When the lipid membrane is subjected to an electrical potential difference of 100 mV, the blocking effect of a polymer chain passing through the pore induces a drop in the electric current of the order of 100 pA, which depends on the chemical nature of the chain, its length, the width of the channel, and the ionic strength. This change in the current is easy to measure using modern electrophysiological techniques, provided that it lasts for more than a few microseconds.

The electrical detection method is the same as the one used on a more macroscopic scale in the Coulter counter [127]. In the biological context of membrane proteins, it is related to the patch-clamp technique [128] and the electrophysiological study of single ion channels [128, 129], reconstituted in

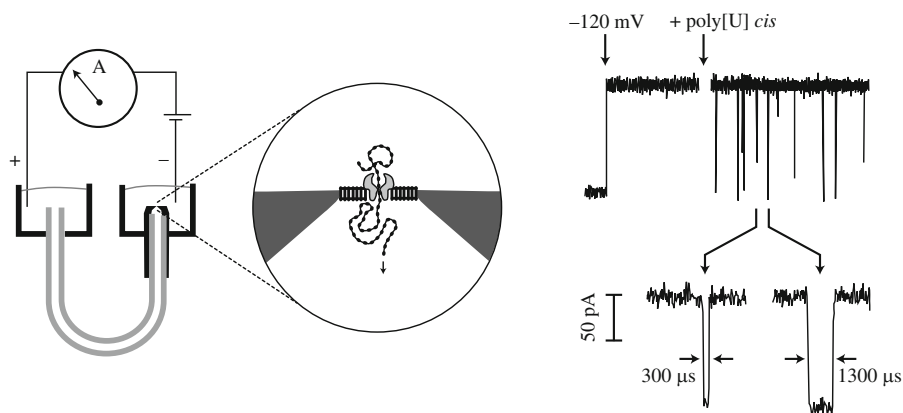


Fig. 11.23. *Left:* Setup for electrical detection of the passage of a macromolecule through a nanopore. *Right:* Electrical plot for the passage of poly-uracil (poly-U) through an α -hemolysin pore. Adapted from [125]

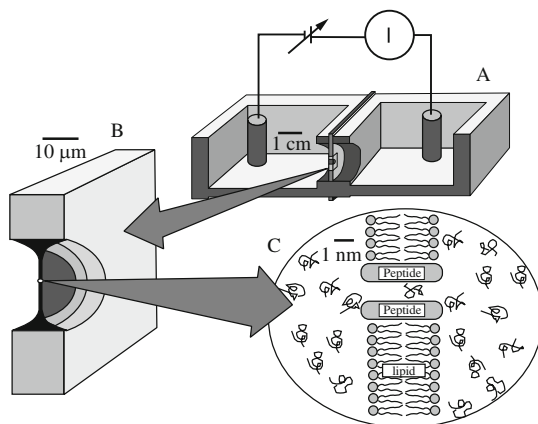


Fig. 11.24. Experiment proposed by Zimmerberg and Parsegian to investigate the passage of polymers through protein channels inserted into a planar lipid bilayer. Adapted from [132]

so-called planar lipid bilayers, deposited on an orifice separating two compartments filled with an electrolyte solution (see Fig. 11.24) and subjected to an electrical potential difference [130]. The idea of passing a polymer chain through a protein pore is also associated with a much earlier suggestion by J. Zimmerberg and A. Parsegian [131] to measure the internal diameter of protein channels by passing chains of increasing size through the channel until it gets blocked [132] (see Fig. 11.24), and also the hypothesis that the lambda phage can inject its DNA through a receptor in the outer membrane of *E. Coli*, namely, LamB porin (maltoporin) [133].

Since then, several groups around the world have been investigating the many applications of macromolecular transport through nanopores, both theoretically [134] and experimentally. These applications concern the ultrafast sequencing of DNA and RNA [135], the manipulation of biological macromolecules [136], the development of chemical and biological nanosensors [137], studies of the fundamental properties of confined polymer chains [138], and the search for new channels, either natural (protein [139]) or artificial, synthetic (based on cyclic molecules [140], carbon nanotubes [141], or tracks left by heavy ion etching [142]) or produced by nanoscale ion beam lithography [143].

It soon became apparent that one way of extending the work on protein channels and getting around certain limitations in their use, such as the fragility of lipid membranes and proteins and the lack of variability in the pore sizes, was to use artificial nanochannels. At the present time there is no simple standard way of fabricating nanometric channels, but focused ion beam lithography [143,144] and electron beam lithography [145], followed if need be by controlled, partial blockage of the channels, has already led to interesting results, for both the fabrication and the translocation of molecules [143]. One of the technical advantages of microelectronics is that it can integrate the channels, electrodes, and the whole of the required electrical, chemical and microfluidic environment into a single device [146].

Historically, it was Charles Bean and Ralph de Blois who carried out the first experiments on particle transport through single artificial pores in 1974 [147]. Their channels, using Nucleopore membranes invented in the laboratories of General Electric [148], were chemically etched by developing high energy heavy ion radiation tracks, and the smallest radius was close to 30 nm. These Nucleopore channels have also been used as a (controversial) model for ion channels in pioneering patch-clamp experiments on artificial porous membranes [149]. Irradiation can now be achieved with a single ion [142], giving rise to a conical pore with radius as small as 2 nm [150].

There are two detailed reviews describing the investigation of molecular transport through pores using electrical methods. The main one, drawn up by Hagan Bailey and Charles Martin [151], describes all results obtained in this field up to the year 2000 and so does not include recent experiments on artificial nanopores. The second [152] is mainly devoted to the study of the dynamics of single-strand DNA molecules in protein pores. The contributions to a conference on the subject in 1999 have been gathered together in one volume [153]. We shall attempt here to give a complete and didactic description of the subject, emphasising the basic ideas. We begin with the principles underlying electrical detection of the passage of a particle through a pore, with some of the necessary mathematics. We then review our physical understanding of the transport of neutral or charged macromolecules through a pore. Finally, we outline the different natural or artificial systems studied to date, and the experiments carried out on them.

One important omission concerns the manipulation of biological structures and force studies by translocation through nanopores used as nanotweezers

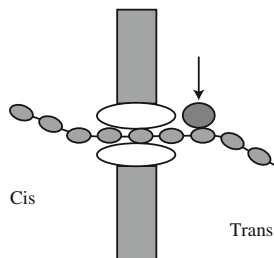


Fig. 11.25. Nanopore as nanotweezers and force machine. A suitably oriented electric field pulls on a macromolecular complex blocked in the pore. By increasing the field, one can measure the force required to break the complex. One can monitor events and exercise feedback control on the applied voltage by measuring the current

(see Fig. 11.25). This difficult subject would require a whole chapter on its own to be treated in a sufficiently didactic way. The first experiments studied the separation of the strands in double-strand DNA [136, 154]. Current projects concern the manipulation of RNA secondary structures, the mechanical unfolding of proteins, and the study of DNA–protein and protein–protein interactions.

11.3.2 Electrical Detection of Particle Transport in a Pore

Electrical Resistance of a Pore

We consider a straight, cylindrical pore of diameter D , passing perpendicularly through an electrically insulating membrane of thickness L , separating two compartments filled with an electrolyte of electrical conductivity σ . A voltage V is applied between the right-hand (*cis*) compartment and the left-hand (*trans*) compartment. There are two cases:

- If the membrane is thick, i.e., $L \gg D$, edge effects are negligible and the electrical resistance R of the pore is given by the usual expression $R = L/\sigma S$, where S is the cross-sectional area of the pore, viz., $S = \pi D^2/4$. Taking $\sigma = 11 \Omega^{-1}\text{m}^{-1}$, the conductivity of a molar solution of KCl, $L = 5 \text{ nm}$, the thickness of a lipid bilayer, and $D = 1 \text{ nm}$, the diameter of a rather big ion channel such as α -hemolysin, this leads to $R = 0.6 \text{ G}\Omega$. The current predicted for a voltage of $V = 100 \text{ mV}$ is $I = 170 \text{ pA}$.
- If on the other hand the membrane is thin, i.e., $L \ll D$, the so-called access resistance R_a is determined by the shape of the electric field lines in the electrolyte at the pore entrance and exit. Assuming that the current lines are uniformly distributed over the entrance and exit half-spaces of the pore, one estimates simply that $R_a = 2/\pi\sigma D$. The solution of Laplace's equation [155] (cited by [129]) gives the exact relation $R_a = 1/\sigma D$.

Dielectric Constant and Surface Charge Effects

Most nanometric channels obtained by chemical etching are electrically charged by spontaneous dissociation of surface groups or adsorption of ions when they are brought into contact with an aqueous solution. The charge of the surface ions, for example negative, is balanced in the solution by a layer of mobile positive ions. This layer is localised over a microscopic thickness in the vicinity of the surface, determined by thermal motions and the total ion concentration. The charged surface attracts ions of opposite sign (called counterions) and repels ions of the same sign. The system constitutes an electrostatic double layer of thickness given by the Debye screening length. The structure of this double layer was analysed qualitatively by Helmholtz in the nineteenth century and quantitatively by Gouy and Chapman at the beginning of the twentieth century, anticipating the Debye–Hückel electrolyte theory. It plays an important role in electrochemistry and in the physics of colloids [156]. The existence of a diffuse ion layer leads to many effects of interest today with the advent of nanotechnology, in particular, the electroosmosis effect. This can be used to produce uniform flows by application of an electric field. Electrostatic effects are also important in protein pores, especially ion channels, because the sign and distribution of charges in the channels are what give them their selectivity. There is an abundant literature on this subject [129], and many numerical simulations [157–159].

The ion channels are inserted into lipid bilayers, which are media of low dielectric constant ϵ_r . This is also true of certain membranes pierced by nanolithographic pores. Typically, $\epsilon_r = 2$ for a lipid bilayer or an organic membrane and $\epsilon_r = 4$ for silica. An ion located within a narrow ion channel is as if placed in a medium with dielectric constant 40 times smaller than water ($\epsilon_w = 80$), and its energy is almost 40 times greater than its electrical energy in water. A high dielectric potential barrier thus opposes the passage of ions through a narrow channel.

The general treatment of surface charge [160] and dielectric constant effects involves lengthy calculations or sophisticated numerical simulations. We shall only give an approximate description here, separating the two problems and restricting to the simple case of channels in which the pore diameter is less than the Debye length. The studied effects then predominate.

Surface Charge and Conductance of a Channel

The ions contributing to electrical conduction in a channel are subjected to an applied electric field, steric forces, and electrostatic forces exerted by the wall and thermal motions. Their motion is described by the Poisson–Nernst–Planck equation, which combines the Poisson equation from electrostatics with the Nernst–Planck (or Smoluchowski) equation for Brownian diffusion motion of ions in an applied field. If the effects of electroosmotic flow are significant, one must also adjoin the Navier–Stokes equation. If the electrolyte concentration is high enough to mean that the Debye length is less than the channel diameter,

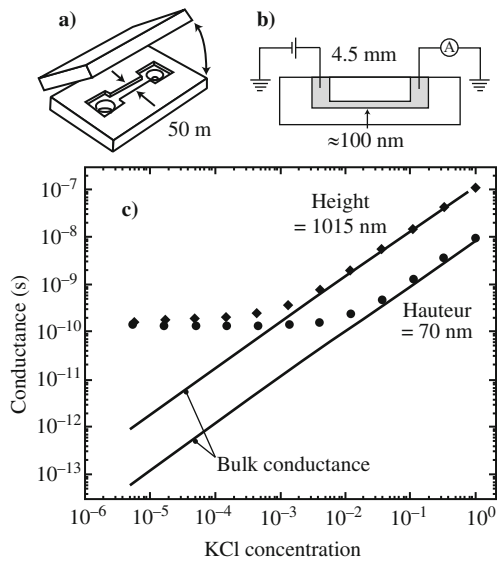


Fig. 11.26. (a) Thin nanochannels. (b) Conductance measurement. The channel thickness is shown in the figure. (c) Conductance of channels as a function of the salt concentration in the solution. Adapted from [161]

the channel resistance is given by the macroscopic laws described previously. The new case arises for low electrolyte concentrations, when the majority charge carriers are the counterions on the channel surface. The conductance G of the channel then tends to a constant proportional to the surface charge of the channel. This effect has recently been observed in channels [161] and very thin silica tubes [162] (see Fig. 11.26).

Dielectric Constant Effect

A charge q placed in water (dielectric constant $\varepsilon_w = 80$), in the vicinity of an external medium of low dielectric constant ε_e , induces polarisation charges of like sign in this medium, which therefore repel it. This effect leads to an increase in the surface energy of an electrolyte solution (investigated by Onsager and Saramas). The calculation of the electric field of an ion placed in the center of an infinitely long cylindrical channel passing through a membrane of dielectric constant ε_e and filled with water can be found in the standard textbooks [163]. It was extended by A. Parsegian and others to the case of a finite length pore [164]. A simplified, semi-quantitative description of this situation is possible [165]. The original idea was due to P.-G. de Gennes.

The external dielectric medium is less permeable than water to the electric field and channels the flux of the electrostatic field created by the charge q along the pore. Due to polarisation charges, the normal component of the electric field at the surface is practically zero when the ratio $\varepsilon_w/\varepsilon_e$ tends to

infinity. To a first approximation, the field $E(x)$ at a large distance x from the charge compared with the pore diameter D is parallel to the pore axis and constant. The equation for the conservation of the flux determines $E(x)$:

$$2\frac{\pi}{4}D^2\varepsilon_w\varepsilon_0|E(x)| = q .$$

It follows that

$$|E| = \frac{2q}{\pi\varepsilon_w\varepsilon_0D^2} .$$

However, at large distances r , the charge appears completely inserted in the external medium and one expects the electric field to be given by the Coulomb law in the external matrix:

$$E(r) = \frac{q}{4\pi\varepsilon_e\varepsilon_0r^2} .$$

Matching the two fields defines a leakage length s :

$$s = \frac{1}{2\sqrt{2}} \left(\frac{\varepsilon_w}{\varepsilon_e} \right)^{1/2} D ,$$

also arising in the full and direct solution of Poisson's equation using Fourier-Bessel series. The length s is the dimension of the distribution of polarisation charges along the pore axis. With $\varepsilon_w = 80$ and $\varepsilon_e = 2$, we obtain $s = 2.24D$, or 4.5 nm for a channel like α -hemolysin in *Staphylococcus aureus*, which is 2 nm across and 5 nm long. When the ionic strength is low, the field lines are almost completely channeled through the channel.

By matching the electrostatic potentials in the different regions, we obtain the expression for the electric potential near the charge q for $r \ll D$. In this region, the potential is given up to a constant by Coulomb's law in water:

$$V(r) = \frac{q}{4\pi\varepsilon_w\varepsilon_0r} + \frac{q}{2\pi\varepsilon_e\varepsilon_0} \left[\frac{4}{\sqrt{2}} \left(\frac{\varepsilon_e}{\varepsilon_w} \right)^{1/2} - 3 \left(\frac{\varepsilon_e}{\varepsilon_w} \right) \right] .$$

The constant, denoted by V_0 , represents the potential created by the polarisation charges at the point where the charge q is located. The dielectric polarisation energy of the charge q is then qV_0 . This is the height of the energy barrier felt by the ion when it goes across the membrane.

Resistance of a Conducting Cylindrical Pore Containing an Insulating Sphere

The discussion in this section is based on [146, 147]. J.C. Maxwell was the first to calculate the change in electrical resistance due to the presence of spherical beads in dilute suspension in a conducting fluid [166]. If ρ_{eff} is the

effective resistivity of an infinitely dilute solution of insulating beads and ρ is the resistivity of the pure electrolyte, Maxwell's result can be expressed by the relation $\rho_{\text{eff}} = \rho(1 + 3f/2 + \dots)$, where f is the volume fraction of the solution occupied by the beads ($f \ll 1$). One can thus deduce a first estimate of the change in resistance of a cylindrical pore of length L and diameter D , when the pore diameter is very big compared with the bead radius. Confinement effects are neglected and the last formula is applied directly. The volume fraction f of a bead of diameter d in the pore is $f = 2d^3/3D^2$. The resistance R^* of the pore containing the bead is then

$$R^* = \frac{4\rho L}{\pi D^2} \left(1 + \frac{d^3}{D^2 L} \right)$$

Introducing the relative change $\Delta R/R = (R^* - R)/R$ in the resistance, this implies that

$$\frac{\Delta R}{R} = \frac{d^3}{D^2 L}.$$

The last relation does not depend on the conductivity of the electrolyte.

The case of beads with almost the same width as the pore ($d \leq D$) can be treated just as simply using the Rayleigh method cited by Maxwell. According to Rayleigh, the dominant contribution to the total resistance is given by the usual expression for the resistance of a cylindrical pore with slowly varying cross-section, used locally:

$$R = \rho \int_{-L/2}^{L/2} \frac{dz}{A(z)},$$

where $A(z)$ is the cross-sectional area of the ring between the sphere and the cylinder at coordinate z . If the origin is taken at the sphere center, assuming the sphere to be centered on the pore axis, then $A(z) = \pi/4(D^2 - d^2 + 4z^2)$ for $z \leq d/2$. After a short calculation, it turns out that the relative change in the resistance is given by

$$\frac{\Delta R}{R} = \frac{D}{L} \left\{ \frac{\arcsin(d/D)}{[1 - (d/D)^2]^{1/2}} - \frac{d}{D} \right\}.$$

There are expressions covering almost the whole range $0 \leq d \leq D$, e.g., the following semi-empirical formula given by C.P. Bean [147]:

$$\frac{\Delta R}{R} = \frac{d^3}{D^2 L} \left[\frac{1}{1 - 0.8(d/D)^3} \right].$$

More sophisticated calculations can treat the case of particles that do not lie on the pore axis [167].

The above calculations and those cited can be used to interpret the relative current variations when a spherical particle passes through the channel, but then do not provide a way of interpreting the full shape of the signal, which depends on the dwell time of the particle in the pore, as well as its more or less random motion and the forces exerted on it. The description of this motion is complex. In particular, it involves the solvent flows induced by the motion of the particle and the associated friction forces. Several classic works have been devoted to this issue [168, 169].

The first experiments were carried out on single Nuclepore channels by Deblois and Bean [147]. Figure 11.27 shows a more recent and particularly striking example.

11.3.3 Polymers Confined in Pores. Statics and Dynamics

The first theoretical work devoted to confined polymers [165, 170, 171] was carried out in the context of a renewed interest in polymer physics, with the development of new chromatographic techniques, but also the effects of the petrol crisis, which made it economically worthwhile to use the enhanced recovery technique to extract crude oil by injecting polymer solutions into porous oil-bearing rocks.

We shall make a distinction between long pores, where the pore is much longer than the polymer chain, and short pores, where the pore is much shorter than the chain. The first case is relevant to chromatography and separation and filtration techniques, while the second holds in recent experiments on protein channels or very thin nanolithographic membranes. It corresponds to the hypothetical situation of a sequencing experiment. Confinement effects predominate in the first case and are negligible in the second. We must also distinguish the very different cases of neutral polymers and electrically charged polymers, also called polyelectrolytes.

A neutral, flexible polymer chain is described in the simplest model by a random walk of N steps, each of length a , where N is the number of monomers in the chain and a is the size of a monomer. A random chain is a statistically scale invariant object with fractal dimension equal to 2. When there are no external perturbations, the root mean squared value of the distance between the ends of the chain is $R = N^{1/2}a$. Real chains in a good solvent – the typical case – have a different structure to ideal chains because of the so-called excluded volume interaction between monomers. A monomer prefers to have contact with molecules in the solvent rather than with other monomers. This produces an effective short-range repulsion between monomers. An excluded volume chain is modelled by a self-avoiding random walk. Like the random walk, it is a scale invariant object, fluctuating and critical in the sense of phase transitions. The root mean squared value of the distance between the ends of the chain now depends on the dimension of the space: $R \approx aN^\nu$, where ν is a critical exponent, equal to 1 in one dimension and about 0.6 ($\approx 3/5$) in three dimensions [172].

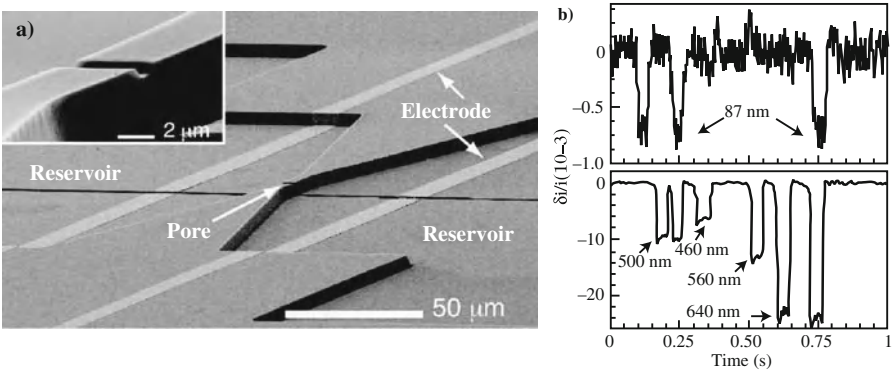


Fig. 11.27. (a) Photograph of the device for electrical detection of beads through microchannels. (b) Electrical signature of the passage of beads through nanochannels. Adapted from [146]

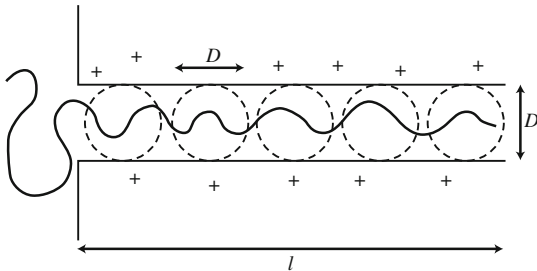


Fig. 11.28. Structure of a neutral polymer chain entering a channel. Adapted from [178]

Neutral Polymers

Structure of a Chain Confined in a Pore

The structure of a neutral, flexible polymer chain in a cylindrical pore is shown in Fig. 11.28 as it was predicted by Mohammed Daoud and Pierre-Gilles de Gennes in 1977 [171]. The chain has a 3D structure at distances less than the pore diameter D , because the monomers ignore the presence of the pore walls. However, the chain is one-dimensional on a large scale. The result is an image rather like a row of blobs, or more poetically, a pearl necklace. Each blob of diameter D contains a 3D piece of chain, comprising $g = (D/a)^{5/3}$ monomers. The length of the chain is the total length of all the blobs placed side by side:

$$R = \frac{N}{g} D = N a \left(\frac{a}{D} \right)^{2/3} . \quad (11.27)$$

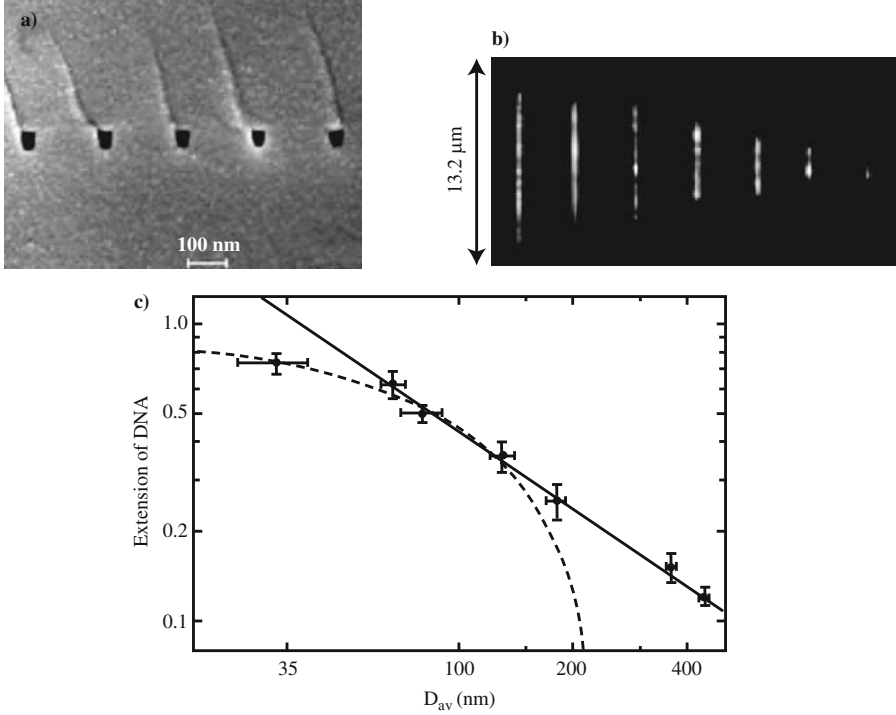


Fig. 11.29. (a) Photograph of nanometric channels (cross-section $30 \times 40 \text{ nm}^2$) printed in a plexiglas matrix coated with silica. Taken from [175]. (b) Single molecule of fluorescent viral DNA (λ phage) confined in nanometric channels of decreasing size (from $30 \times 40 \text{ nm}^2$ to $440 \times 440 \text{ nm}^2$ from *left to right*), identical to those in Fig. 11.22 centre. (c) The experiment shows that confinement stretches the DNA chains

The last relation has long been confirmed by numerical simulations [173]. Chains tethered onto a surface have a 1D structure analogous to that of a chain in a pore of diameter equivalent to the distance between anchoring points. It has been possible to check experimentally, by neutron scattering, that the thickness of a grafted layer is indeed given by (11.27) [174]. A direct observation of fluorescent DNA confined in silica nanochannels has recently given rise to a very elegant test of this law [175] (see Fig. 11.29).

Partition Coefficient

In the spirit of a scaling law description analogous to those occurring in the study of critical phenomena, it is natural to attribute the thermal energy $k_B T$ (k_B is Boltzmann's constant and T the temperature) to a correlated domain or blob, of size D . The free energy of a chain is then $F \approx k_B T(N/g) = Nk_B T(a/D)^{5/3}$. Comparing this free energy with that of a chain in a non-confined dilute solution, which is of order $k_B T$, one calculates the partition

coefficient q , the ratio of the monomer concentrations at equilibrium in the pore and in the bulk:

$$q = \frac{c_p}{c_b} = \exp(-\alpha N) \left(\frac{a}{D} \right)^{5/3}, \quad (11.28)$$

where α is a constant of order 1. This coefficient is exponentially small for long chains, implying that a long isolated chain will almost never enter spontaneously into a small pore.

The first observations of macromolecular permeation through Nuclepores agreed qualitatively with this law [123]. Recently, a quantitative experimental check of this theoretical prediction of the partition constant was achieved by observing the frequency with which short neutral chains of polyethylene glycol pass through an α -hemolysin protein nanopore [176].

Osmotic Forcing

One way of getting a long chain into a pore is to exert a force which pulls or pushes it in. The first option is achieved by flow, and the second by forces of osmotic origin. By increasing the bulk concentration of a polymer solution, we increase the strength of the excluded volume interactions, which compress the chains and force them to tangle up. This is a semi-dilute regime, which corresponds to a monomer concentration c greater than the overlap concentration c^* , given by the average concentration within an isolated coil: $c^* \approx N/R^3 = 1/N^{4/5}a^3$. In a semi-dilute solution, excluded volume interactions are short range and screened at greater distances, because repulsions between chains balance repulsions within a chain. The correlation or screening length ξ of excluded volume interactions is the average distance to a point of entanglement. It is equal to the radius of an isolated chain at $c = c^*$, and equal to the size a of a monomer when there is no solvent. A power law interpolation, justified by the theory of critical phenomena, leads to the relation $\xi = a(ca^3)^{-3/4}$, which has been confirmed experimentally. A semi-dilute solution of polymers is a compact pile of correlated domains or blobs of size ξ . The osmotic pressure Π is proportional to the free energy per unit volume, given by the Widom scaling law according to the prescription $k_B T$ per blob, whence $\Pi \approx k_B T/\xi^3$. A chain in a semi-dilute solution will enter a pore of size D when its monomers have greater chemical potential than the confined monomers. This happens when the correlation length becomes shorter than the pore diameter, i.e., $\xi \leq D$. The geometric interpretation of the pore penetration condition by a dilute polymer solution is straightforward: the size of the blobs in solution must be less than the pore diameter (see Fig. 11.30). At the threshold, the blobs of a chain in solution and a confined isolated chain are identical. This condition implies the paradoxical conclusion, subject of some controversy, that a polymer melt is able to invade pores of any size whatever. In contrast to a claim that is often made, there is no confinement entropy for chains in a polymer melt [172].

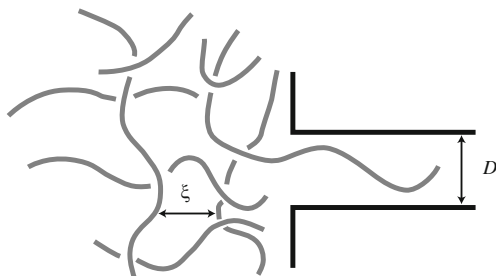


Fig. 11.30. Schematic view of a semi-dilute solution of polymers of correlation length ξ at the entrance to a pore of width D

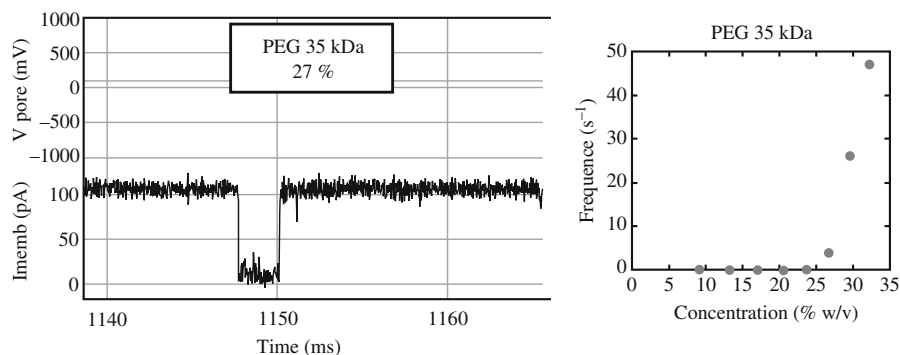


Fig. 11.31. *Left:* Current measured during the passage of a poly(ethylene glycol) chain of mass 35,000 dalton through an α -hemolysin pore. *Right:* Frequency of passage of this polymer at different solution concentrations in g/mL

Observations of threshold effects have shown agreement with the predictions [123, 124], but there has not been any complete verification of the law owing to the difficulty in varying the pore size in a series of experiments. The most spectacular observation of a threshold effect has been carried out on an α -hemolysin protein pore and poly(ethylene glycol) chains of mass 35,000 dalton [177]. It is shown in Fig. 11.31. This example does not strictly correspond to the case of an infinitely long pore, because the dimension of the chain is of the same order of magnitude as the pore length, viz., 5 nm. However, the entry phenomenon is the same in this case as for a long pore.

Chain Dynamics at the Entrance and Within a Pore

We begin by recalling the dynamics of polymer chains [172]. Any particle or molecule in solution is subject to the action of solvent molecules. Their effects are twofold. A friction force opposes the average motion, and a random force due to thermal agitation gives rise to Brownian motion:

- The friction force $\mathbf{f}_v$ is proportional to the average relative velocity $\mathbf{v}$ of the solute molecule and the solvent molecules, i.e., $\mathbf{f}_v = -\zeta_0 \mathbf{v}$, where ζ_0 is the coefficient of friction of the molecule. The latter is given to within an order of magnitude by the Stokes relation $\zeta_0 = 6\pi\eta_s a$, where η_s is the viscosity of the solvent and a the radius of the molecule.
- Thermal motions give rise to a diffusion phenomenon characterised by the diffusion coefficient D_m , related to ζ_0 by the Einstein relation $D_m = k_B T / \zeta_0$.

In this framework, the dynamics of a polymer chain is described by attributing a coefficient of friction ζ_0 to each monomer. Assuming that the total friction force on the chain is the sum of the individual friction forces on each monomer, a friction coefficient $N\zeta_0$ is attributed to the whole chain, where N is the polymerisation index. The corresponding diffusion coefficient is $D_c = k_B T / N\zeta_0$. This description is not complete, however. In the best of cases, it only applies to the overall motion of a chain and not to its internal motions. Furthermore, it neglects hydrodynamic interactions, which arise because the motion of one monomer produces a flow that perturbs the motion of the others. Hydrodynamic interactions are negligible in very open structures, but they become significant for random coils in solution. Because of these interactions, the solvent inside the chain is dragged along at the same velocity as the macromolecule, and the latter behaves from a hydrodynamic point of view like a sphere of radius R , the radius of the chain. As Bruno Zimm first showed in 1956, the diffusion coefficient of a chain is then $D_c = k_B T / 6\pi\eta_s R$, and the relaxation time of a chain is, to within an order of magnitude, R^2 / D_c , so that $\tau \sim \eta_s R^3 / k_B T$.

We now consider the suction of the chain by a fluid flow. A macromolecule can be dragged into a pore by a flow of the solvent. In this case, there is a critical flow rate necessary for the insertion of a linear chain [178]. The flow at the entrance to a channel is convergent and elongational, and the viscous friction forces tend to extend a flexible macromolecule. A chain will enter a pore if it is sufficiently stretched out and its lateral dimension sufficiently reduced. A more precise discussion is delicate because the phenomenon is a dynamic one. The deformation criterion for a chain involves the dimensionless product of the velocity gradient $\dot{\gamma}$ with the relaxation time τ of the macromolecule. The deformation is only large if $\dot{\gamma}\tau \geq 1$. Due to flow conservation, the velocity field is given approximately by $v(r) \approx J/r^2$, where r is the distance to the center of the entry disk and J is the flow rate of the solvent. The velocity gradient is $\dot{\gamma} = \nabla v \approx J/r^3$. In Zimm's model, mentioned briefly above, the relaxation time of a chain is $\tau \sim \eta_s R^3 / k_B T$, where η_s is the viscosity of the solvent [172]. The condition $(\nabla v)\tau = 1$ defines the dimension r_c of the region close to the entrance of the pore where the viscous stress exerted on the chain is high. If the stress is high, each monomer follows the solvent flow independently of the others. The deformation of a chain is 'affine' as long as it is not fully elongated. The deformation of the chain and the deformation of a volume element

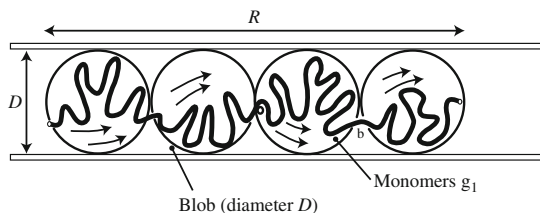


Fig. 11.32. Dynamics of a polymer chain in a channel. Taken from [179]

of the solvent are proportional. The lateral dimension $r_{\perp}$ of the chain at a distance r from the pore is then $r_{\perp}(r) = (r/r_c)R$ and equals $r_{\perp} = (D/r_c)R$ in the pore [172]. A chain will be sucked into the pore if $r_{\perp} < D$ or $R < r_c$. This condition is satisfied if the flow rate through the pore is greater than a critical flow rate J_c given by

$$J_c = \frac{k_B T}{\eta_s}.$$

This value is independent of the pore size and the chain length. In the blob picture described at the beginning of this section, it corresponds to the point at which the Stokes force $\eta_s D v$ on a blob of size D becomes comparable with the entropic force $k_B T/D$ that must be exerted to confine it. (v is the average speed of the solvent in the pore, related to the flow rate by $v \sim J/D^2$.)

We now consider the dynamics of a confined chain. F. Brochard and P.-G. de Gennes [179] studied the Brownian motion of a neutral, flexible polymer chain in a pore. This is a subtle problem. Reasoning by analogy with the 3D case, one might think that a confined chain would drag the solvent along with its motion, but this is impossible. Current conservation in a 1D channel would require that a single chain displace the solvent throughout the channel. The solvent must be at rest outside a moving chain and this requires a recirculation or backflow of the solvent in the chain, as shown schematically in Fig. 11.32.

Boundary conditions require the hydrodynamic interactions in a channel to decrease exponentially beyond a characteristic distance, equal to the width of the channel. The hydrodynamic interactions are said to be screened. It follows that recirculation effects occur over a length scale equal to the pore diameter. Taking this effect into account, the mobility μ of a chain is that of a linear necklace of N/g pearls of diameter D , each being attributed a friction coefficient of order $\eta_s D$. It is given by

$$\zeta = \frac{1}{\mu} \approx \frac{N}{g} \eta_s D = \eta_s R = \eta_s N a \left(\frac{a}{D} \right)^{2/3}.$$

If the passage of a molecule through a pore is purely diffusive, the coefficient of diffusion is $D_f = \mu k_B T$ and the dwell time τ in the pore is given by

$$\tau = \frac{L_{\text{pore}}^2 \eta_s N a}{k_B T} \left(\frac{a}{D} \right)^{2/3},$$

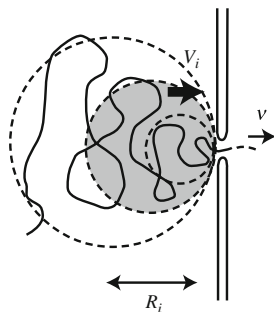


Fig. 11.33. Diagram representing the friction exerted on a polymer chain as it passes through a nanopore. Taken from [180]

where L_{pore} is the pore length. This time varies linearly with the chain mass N .

Consider now passage through short pores. The experiments mentioned in the introduction inspired a great deal of theoretical work and numerical simulations of the dynamics of a polymer chain passing through a very narrow pore [81–84, 180, 185], but there have been few systematic experimental studies.

The simplest case concerns a rigid and electrically neutral linear chain. The motion is purely random and diffusive. The dwell time τ (strictly the average time of first passage) is given to within an order of magnitude by the Brownian law of motion $\tau = L^2/D_f$, where L is the chain length and D_f the diffusion coefficient of the molecule. The value of this diffusion coefficient depends on the relative magnitudes of the various friction forces:

- If the viscous friction of the solvent on the chain outside the pore is much greater than the friction in the pore, the diffusion coefficient is

$$D_f \approx \frac{k_B T}{\eta_s N a},$$

because the friction is proportional to the length of the molecule. The dwell time then varies as the cube of the chain length, i.e., $\tau \propto N^3$.

- In the opposite case, the diffusion coefficient is

$$D_f \approx \frac{k_B T}{\eta_s x},$$

where x is the length over which friction occurs, typically the length of the pore. In this case, the dwell time is proportional to the square of the chain length, i.e., $\tau \propto N^2$.

The case of a neutral and flexible chain is more complex. As Lee and Obukhov were the first to point out [180], the passage of a flexible chain through a short, narrow pore will hardly modify its conformation and the friction due to the solvent is much smaller than that exerted on a rigid linear chain (see Fig. 11.33). Let us imagine a situation where we pull one end of a polymer coil

through a hole at a speed v . The part of the coil upstream of the hole moves at a very low average speed V . This speed V is determined by the condition that the total time $t = L/V$ required to unwind the coil, where $L = Na$ is the total length of the ‘thread’, is the same as the time required to translate the coil over its own length, viz., R/V , where R is the coil radius. Hence, $V \simeq (R/L)v = vN^{-1/2}$.

In this case, the friction force on the coil is the same as the friction force on a sphere of radius R moving at speed V , viz., $f = \eta_s RV$. The mobility of the chain is then given by $\mu \simeq \eta_s R$, the diffusion coefficient by $D_f = k_B T / \mu$, and the dwell time by $\tau \sim N^2$, for an ideal chain. Up to a factor, this is the same result as the one found for localised friction in the pore. It agrees with existing numerical simulations [184].

There are no experimental observations of these laws, because long, flexible and neutral chains do not pass spontaneously through the pores of finite length that are currently in use. Experiments have been carried out on concentrated systems, where passage is facilitated by the screening effects mentioned earlier [177]. In this case, the contribution to the kinetics of reptation effects in solution in the vicinity of the pore no longer seems to be negligible.

Charged Polymers

Polyelectrolytes are ionisable macromolecules, charged in an aqueous solution and very common in nature. Biopolymers like nucleic acids, proteins, and polysaccharides are polyelectrolytes. They often have a very high charge density (which may depend on the pH) and this gives them novel properties. The charged monomers along the chain repel one another and this tends to rigidify these molecules, but the ions in solution screen the electrostatic interactions. The bulk structure and properties of these solutions remain poorly understood, particular in the semi-dilute regime, despite a considerable amount of theoretical and experimental work [186–188].

During a translocation experiment, polyelectrolytes are dragged along by the strong electric field in the pore which is used for electrical detection of the passage of molecules. Recall that a potential difference V of the order of 100 mV applied over a distance of 5 nm, the thickness of a lipid bilayer, leads to an electric field of 20 million V/m. If an elementary charge e is located in the pore, the associated potential energy eV will be 4 times greater than the thermal energy $k_B T$. For single-strand DNA, for example, the effects of thermal motions are small. Experiments in which relatively small DNA molecules pass through hemolysin protein nanopores show that the translocation time is proportional to the length of the molecule, and the frequency of translocations is proportional to the Boltzmann factor of the applied potential [189]. These results are significantly modified for very long double-strand DNA passing through artificial nanopores. The effects of flexibility and friction with the solvent on the part of the macromolecule that has not yet gone through the pore are significant and lead to translocation times varying as $R^2 \sim N^{1.2}$ [190].

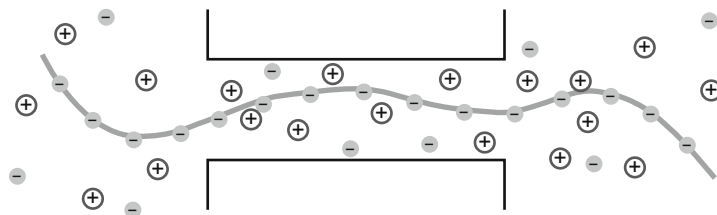


Fig. 11.34. If a pore is long enough, it is better to confine the counterions around a chain to ensure electroneutrality locally, even if the Debye length is greater than the pore diameter [165]

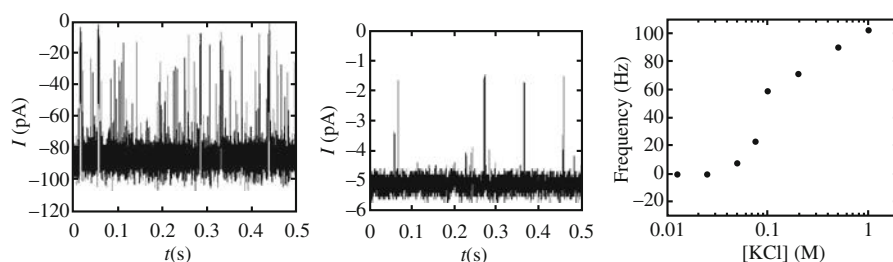


Fig. 11.35. Translocation experiment with dextran sulfate (molar mass 8,500) through a hemolysin protein pore subjected to a potential difference of 100 mV. *Left:* The electric current showing that macromolecules are going through is measured for an electrolyte concentration of 1 M. *Center:* Electrolyte concentration 0.05 M. *Right:* Graph of translocation frequency against electrolyte concentration, showing that translocation is prevented below a certain threshold concentration

The conformation and properties of confined polyelectrolytes in pores have received little theoretical or experimental attention, with the two exceptions described in [165, 191]. The most interesting situation is the one in which the polyelectrolytes are dissolved in the absence of added salts. Electrostatic effects are then the strongest. In the absence of surface charge, two effects prevent a charged linear chain from entering the pore: repulsion by polarisation charges as discussed above and possible compression of the cloud of counterions, required to guarantee electrical neutrality around the chain, when the Debye length κ^{-1} is greater than the pore diameter (see Fig. 11.34). It can be shown that the spontaneous entry of a chain in a long enough pore is unlikely as soon as $\kappa^{-1} \geq D/2$.

Figure 11.35 shows the results of a translocation experiment with a rather short polyelectrolyte, dextran sulfate, through a hemolysin pore, varying the ionic strength of the solution. The frequency of passage of the chains through the pore decreases considerably (for fixed chain concentration) and goes to zero when the electrolyte concentration goes below 0.1 M. This threshold corresponds to a Debye length κ^{-1} of 1 nm, equal to the pore radius, thereby

confirming the theoretical prediction. Naturally, this effect must depend on the strength of the applied field, but this study remains to be done.

11.3.4 Some Natural and Artificial Systems

Experiments on the transport of macromolecules through nanopores have been carried out on several types of pore, i.e., protein pores and artificial synthetic pores. Protein pores are the easiest to obtain, at least in some cases, and they are the most effective. They are less wide and shorter than synthetic pores. We have already discussed several examples of the use of pores. In this section, we describe their fabrication and implementation.

Planar Lipid Membranes and Biological Nanopores

Lipids are amphiphilic molecules with a hydrophilic part, the polar head, and a hydrophobic part comprising one or more carbon chains. In an aqueous solution, the hydrophobic parts are insoluble and, depending on their geometry, lipid molecules assemble spontaneously into different kinds of aggregates, viz., spherical or cylindrical micelles and plane bilayers, with their polar heads at the surface of the aggregate in contact with the water and their hydrophobic tails inside. In a lipid bilayer, the molecules lie head-to-tail and the polar heads are assembled into a plane sheet on either side of a central part itself comprising two sheets of chains. The thickness of the bilayer is approximately equal to twice the length of the lipid chains. The bilayers formed from natural phospholipids, e.g., phosphatidylcholine–lecithin, are the basic constituent of biological cell membranes and the membranes of artificial structures like vesicles [192, 193].

In 1962, Mueller and Rudin discovered [130] that a single macroscopic lipid bilayer could be deposited on a frame or across a macroscopic orifice measuring a few hundred micrometers in diameter and bathed in water, in the same way that soap films can be formed in air. These objects are called black lipid membranes because they are so thin (typically 5 nm), by analogy with ‘black’ soap films, too thin to reflect light. These planar lipid bilayers were soon being used as a support for the insertion of membrane proteins, to carry out *in vitro* studies. A particularly important class of membrane proteins is the ion channels. These are proteins or protein assemblies arranged in such a way as to form a central channel allowing the transport of ions or other molecules through the lipid membrane, which would otherwise be insulating from an electrical point of view [129]. By placing electrodes on either side of the bilayer and imposing a potential difference (see Fig. 11.23), the electric current through the membrane can be measured. In particular one can directly detect the current through a single protein channel inserted into the bilayer, monitoring its fluctuations, although the current in 1 M KCl is typically in the picoampere range for a potential difference of 100 mV. These studies long provided the only means of studying the functioning of ion channels [194] and

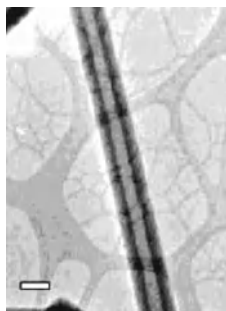


Fig. 11.37. Transmission electron microscope image of a silica nanotube. *Scale bar* 100 nm. Taken from [198]. Copyright 2005, American Chemical Society

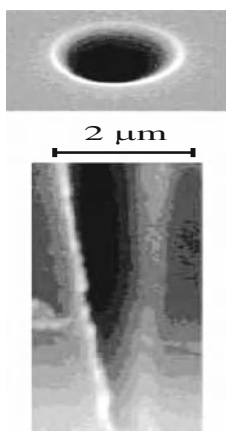


Fig. 11.38. Scanning electron microscope image of a conical pore in an organic membrane. Taken from [150]. Copyright 2004, American Chemical Society

ratio of order 1. Nanochannels often have bigger diameters than nanopores. There are many fabrication methods. The historical reference is irradiation by high-energy heavy ions, which produces Nuclepore channels. One should also mention pulled glass micropipettes. More recently, nanochannel sources have become more diverse.

Improvements in the resolution of conventional microfluidics, using PDMS, have made it possible to fabricate PDMS channels with dimensions between 100 nm and 1 μm . The channels are drawn by electron lithography in a PMMA resist and the motif transferred by nanoimprinting [196]. DNA has been observed to pass through such channels, using electrical methods [146] and fluorescence techniques [197].

Carbon nanotubes are also an interesting source of nanochannels. One advantage is that they have a uniform diameter of 50 nm over almost 100 μm . Carbon nanotubes are mounted under the microscope, then incorporated in a block of resin. The system is subsequently sliced using a microtome, producing nanotube slices of thickness 600 nm. Electrical detection has observed the passage of polystyrene beads through such tubes [141].

More recently, one group has fabricated similar systems from silica [198], using silicon tubes produced by chemical vapour deposition (CVD), the motif

being transferred to the silica by an oxidation and etching technique (see Fig. 11.37). The inner radius of the tubes is then of the order of 50 nm. The translocation of DNA through these tubes has been detected and it was shown how charge on the tube walls affected the measured current.

Nanopore Fabrication

Nanopores are more difficult to fabricate than nanochannels but they have the advantage of increasing the spatial resolution of the confinement potential applied to the macromolecule. A polymer is no longer simply confined in a tube, but held as though by tweezers over a short length of its chain, which has some advantages for sequencing applications. The nanopores can be made in organic or inorganic membranes. Two distinct approaches are being explored at the present time: irradiation of organic membranes by heavy ions, derived from the methods used in the 1970s by C.P. Bean and coworkers [147] to make the original Nuclepores, and irradiation of a ceramic membrane by a focused beam.

Consider first the irradiation of organic membranes by high-energy heavy ions. This method, known as track etching, produces pores of various geometries in organic polymer membranes. Several types of commercial membrane have been used, such as poly(ethyleneterephthalate), polycarbonates, or polyimides. The membrane is perforated by lithography with high-energy heavy ions, e.g., Xe ions at 140 MeV. Each ion breaks chemical bonds as it passes through and leaves a track in the membrane, which can then be revealed by chemical etching, immersing the membrane in a sodium hydroxide solution. Pores of well-controlled geometry and dimensions (cylindrical, conical, or double conical pores with sizes between 10 nm and 5 μm , as illustrated in Fig. 11.38) can be obtained by optimising the development (time, adding or not adding surfactants to the sodium hydroxide solution, differential development on either side of the membrane) but also the way the polymer membrane is produced [150]. In contrast to other lithographic techniques, this method has the advantage of providing very good resolution (better than 10 nm). Indeed, a single heavy ion is sufficient to transform the material, and this limits diffraction and diffusion effects. The size of the track left by the ion varies between 2 and 3 nm. Such perforated membranes are commercially available today under the name of Nuclepores, because they constitute very effective filters, applicable with a wide range of solvents. However, this means of fabrication does require access to sophisticated equipment, such as a particle accelerator, and the number of ions crossing the membrane will depend on the exposure time in the accelerator. It is thus difficult to obtain a low pore density and to isolate a single pore. In the 1970s, Bean had already succeeded in electrically detecting the passage of a bead through such a membrane, and more recently, polymers (DNA or porphyrin) have also been observed to pass through such membranes [150].

Consider now the irradiation of ceramic membranes by focused ion or electron beams. Standard techniques from microelectronics can also be used

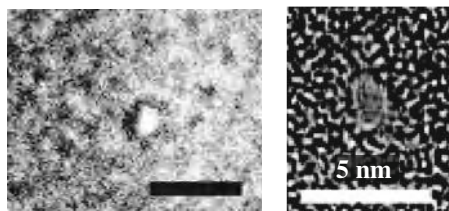


Fig. 11.39. Nanopores in ceramic membranes. *Left:* 3-nm nanopore in a SiC membrane perforated by FIB [144]. *Right:* 3-nm nanopore in a Si₃N₄ membrane perforated by EBL [199]

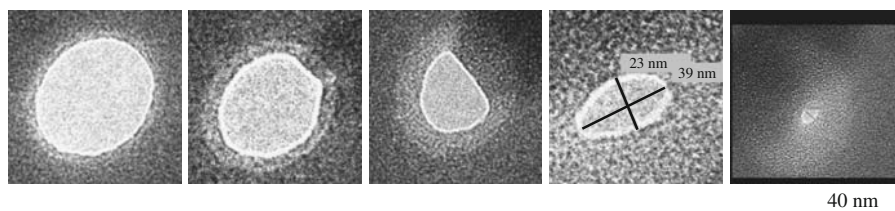


Fig. 11.40. Blocking a hole in a silicon carbide membrane inside a transmission electron microscope. Courtesy of A.L. Biance and J. Gierak

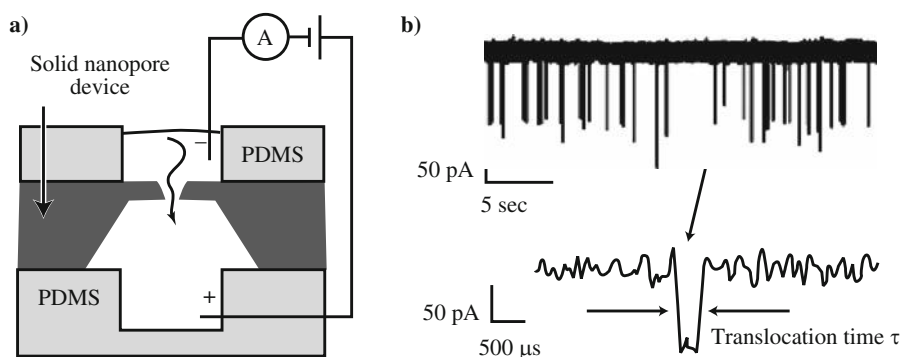


Fig. 11.41. (a) Device made from PDMS (*light grey*) used to integrate a perforated membrane. (b) Current plot obtained when a DNA molecule passed through a silica nanopore. Taken from [145]

to generate single pores on a nanometric scale. At least four groups around the world are exploring the fabrication of such systems [143–145, 199]. Generally speaking, there are four stages in the fabrication of artificial nanopores used to detect the passage of molecules or colloidal objects:

- *Stage One. Fabrication of Ceramic Membranes.* Suspended membranes are fabricated by standard methods: a thin layer of the material making up the membrane is deposited on a substrate which is subsequently etched on its rear face. The materials chosen are silicon nitride [143, 199], silica [145], or

silicon carbide [144], owing to the suitable combination of electrical properties (wide gap semiconductors or insulators) and mechanical properties (high Young's modulus).

- *Stage Two. Perforating the Membrane.* The first technique for perforating membranes is direct action by means of a focused ion beam (FIB) (see Fig. 11.39 left). The ions then have enough energy to remove matter and dig into the membrane, in a phenomenon known as sputtering, even passing right through if it is thin enough. Depending on the dimensions of the ion beam and the thickness of the membrane, resolutions in the range 3–50 nm have been obtained. The second technique is electron or ion beam lithography in a layer of PMMA (see Fig. 11.39 right), yielding pores of diameter 1 nm [199] to 20 nm [145].
- *Stage Three. Partial Blocking of the Hole.* If molecular resolution is not attained during perforation, two methods have been contrived to reduce the hole size. The hole can be partly blocked by depositing an insulating material on the perforated membrane. This method has been tested by the group at Harvard, by depositing aluminium oxide monolayer by monolayer [200] on the perforated membrane. The hole can also be blocked up by favouring the diffusion of species in its vicinity. This diffusion can be speeded up by scanning an electron beam at the edge of the hole (directly under the transmission electron microscope) or by scanning with an ion beam. The electrons heat the material locally, while the ions disturb species at the surface. If the aspect ratio of the hole is favourable, i.e., if the hole is longer than it is wide, this diffusion results in a reduction in the hole diameter (see Fig. 11.40). In the opposite case, the hole will be widened. The physical mechanisms inducing partial blocking of the hole are still not fully understood.
- *Stage Four. Integrating the Device.* The device is generally integrated into a PDMS microfluidic matrix, which allows liquids to circulate without exerting any pressure on the thin, hence fragile, membrane, and also allows integration of electrodes. Such a device is shown schematically in Fig. 11.41, together with the current recorded when a DNA molecule passed through.

Experiments carried out with this type of system have mainly detected the translocation of double-strand DNA and studied the folding of the macromolecule as it passes through the pore. There are still many research groups developing this type of device, and none of the techniques discussed above would appear to stand out in terms of performance. A lot remains to be done.

11.3.5 Conclusion and Prospects

The electrical detection of macromolecular transport through nanopores is a new technique. Its application to biological systems remains limited. The prospect of using this technique as a sequencing method remains realistic, but a great deal of development will be needed first, probably involving a

chemical modification of the pores and various technological artifices such as integration of the electrodes to improve resolution. No one knows whether these efforts will achieve their aim on a reasonable time scale. Other applications nevertheless look very promising in the shorter term. The most important is probably the use of nanopores as tools for micromanipulation of biological macromolecules and force machines. Another use is in the realisation of biomimetic systems for active translocation, such as eukaryotic and prokaryotic translocation apparatus and various protein systems for synthesis of biomacromolecules. Motor molecules and catalysts can be grafted inside natural or artificial pores. Studies are currently under way. From a technical point of view, it looks interesting to combine electrical detection of molecular translocation with optical detection based on various fluorescence techniques, e.g., excitation transfer between a donor molecule grafted onto a confined chain and an acceptor molecule grafted onto the pore wall.

Acknowledgements

Original results presented here come from work financed by *Action Concertée Incitative 'Nanosciences'* in the context of a project entitled *Extrusion moléculaire*, associating Philippe Guégan, Laurent Bacri, and Jérôme Mathé of the *Matériaux Polymères aux Interfaces* (MPI) group at the University of Evry (France), Catherine Amiel and Véronique Wintgens of the *Laboratoire de Recherches sur les Polymères* (LRP) at the University of Paris 12 and the CNRS, Jacques Gierak and Ali Madouri of the *Laboratoire de Photonique et Nanostructures* (LPN) at the CNRS in Marcoussis (France), Yong Chen of the LPN and the *Laboratoire Pasteur de l'Ecole Normale Supérieure*, Elie Raphaël of the *Laboratoire de Physico-Chimie Théorique* at the ESPCI (Paris), and Jean-Louis Sikorav in the biology department of the CEA in Saclay (France). This chapter owes much to them and we offer our warmest thanks.

11.4 Electrophoretic Techniques

11.4.1 Introduction

The considerable recent progress in molecular biology has marked the beginning of a new era in medicine, the pharmaceutical industry, and the field of biotechnology. This progress is largely the result of improved analytical techniques for biological macromolecules. Electrophoresis is without doubt one of the most important of these techniques. The separation of DNA fragments in particular lies at the heart of all sequencing methods, such as those which recently led to a complete reading of the whole human genome and which constitute the foundation of most diagnostic methods for genetic disorders. The fragments to be separated may contain anything from a hundred million bases (the size of a chromosome), through about ten thousand bases (the size

of a gene), to just a hundred or even ten bases (produced by PCR, genes cut by restriction enzymes). These separations must therefore be possible over 8 orders of magnitude. Electrophoretic techniques have thus been developed and adapted to each of these scales.

In this section we discuss the various existing methods and the main physical mechanisms underlying these DNA separation techniques. We then describe more specifically the prospects attributable to miniaturisation.

11.4.2 Migration of a Charged Species in Solution

Electrophoresis is the generic term used to describe the motion of a charged species induced by an electric field in a conducting solution (also called an electrolyte). When this species (molecule, macromolecule, or nanoparticle) is set in solution, a layer of counterions forms around it, called the Stern layer. In the layer of fluid immediately around the molecule, the counterions are considered to be permanently adsorbed (see Fig. 11.8).

There is also a region around the molecule where the counterions are more concentrated than in the rest of the solution, but not immobilised as they are in the Stern layer. This region is called the diffuse layer. The region comprising the two layers is called the electrical double layer. As a consequence, one can define a screening length for the charge of the molecule, corresponding to the dimensions of this double layer. This is the Debye length, denoted here by λ_d . Beyond this length, the effect of the charge of the molecule is no longer felt in the medium.

If an external field is now applied, it exerts a force on the molecule, but also on the ions surrounding it. In particular, the cloud of ions of thickness λ_d surrounding the molecule contains a surplus of counterions which are dragged in the opposite direction to the molecule. They interact hydrodynamically with the latter. The speed of the molecule will thus depend on the thickness of the Debye layer. Since this speed is proportional to the applied field for reasonable fields (applied electric fields are usually less than 400 V/cm), one defines the mobility as the ratio of the speed to the field.

For a small Debye layer compared with the particle size R ($\lambda_d \ll R$), shear is essentially contained in the layer of thickness λ_d and it can be shown that the electrophoretic mobility μ is given by

$$\mu = \frac{\varepsilon_b \varepsilon_0 \zeta}{\eta},$$

where η is the viscosity of the liquid, ε_0 is the permittivity of the vacuum, ε_b is the dielectric constant of the electrolyte, and ζ is the zeta potential on the surface where shear appears in the fluid. Note that, in the latter case (which holds when dealing with DNA molecules, these being uniformly charged), the mobility is independent of the particle size. It is not therefore possible to separate large uniformly charged particles, i.e., large compared with the

Debye length, in a free solution. Since the electric field is not sufficient to achieve the separation, the objects to be separated must be made to migrate in a network of obstacles, with which they interact, in order to separate them according to their size or their conformation. These networks can be formed from polymer gels or nanofabricated.

Main Types of Electrophoresis

To begin with, electrophoresis was developed using gels. A polymer gel such as acrylamide or agarose is used as a separation matrix. The DNA is introduced into pits in the matrix. The gel is bathed in an electrolyte solution in which the electric field is applied. This technique is still commonly used in molecular biology, especially for separating medium-sized DNA fragments containing a thousand or so bases.

Over the last twenty years, gel electrophoresis has been more and more often replaced by capillary electrophoresis. In this case, a fused silica capillary of bore around $50\text{ }\mu\text{m}$ and length 50 cm is filled with a solution of entangled polymers. (Gels are not used in capillaries for practical reasons, e.g., difficulty in getting the gel inside, breaking of the gel, etc.) Each end of the capillary is held in a buffer solution, and an electric field is applied between the two ends (see Fig. 11.42).

Capillary electrophoresis allows a high level of automation and an advantage with regard to sensitivity and analysis rate. Indeed, the high surface to volume ratio of the capillaries results in good thermal dissipation, so much stronger fields can be applied than on a gel (in the range 100–500 V/cm).

Research is turning more and more toward miniaturisation of electrophoresis systems. In this case, the separation channel is etched into a chip, and networks of microfluidic systems can be sculpted on the same chip. The channel can be filled with a solution of entangled polymers or include etched structures (see Sect. 11.4.4). The main interest of this approach lies in the possibility of integrating different operations into the same chip, e.g., DNA extraction, amplification, and separation, but also in the lower consumption of sample material and reduced separation times. We shall describe the use of polymer matrices for electrophoresis in capillaries or on chips, and in particular for DNA separation applications. Finally, we shall discuss applications more specifically devoted to chips, involving etched arrays.

11.4.3 Use of Polymer Matrices

Solutions of Entangled Polymers

In a dilute solution, polymer chains are hydrodynamically isolated from one another. When the solution is more concentrated, up to a limiting concentration C^* called the overlap or entanglement concentration, the chains begin to interact with one another (see Fig. 11.43) and form a network. One then

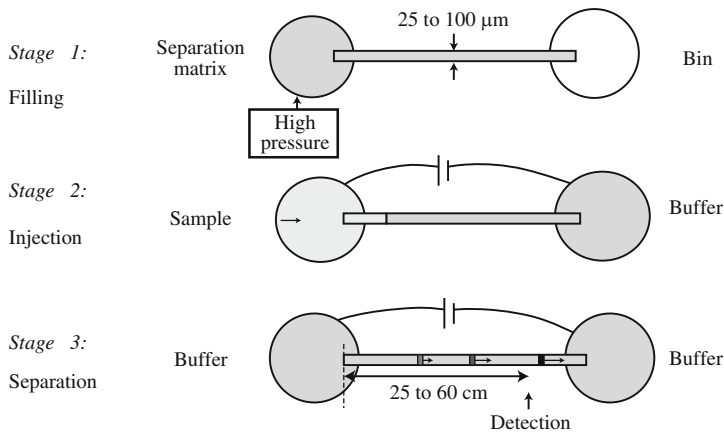


Fig. 11.42. Three stages in a capillary electrophoresis experiment. Stage 1: The polymer solution is injected into the capillary. Stage 2: DNA is injected electrokinetically into the capillary. Stage 3: The DNA molecules are set in motion by applying a strong electric field, separated, then directly detected in the capillary

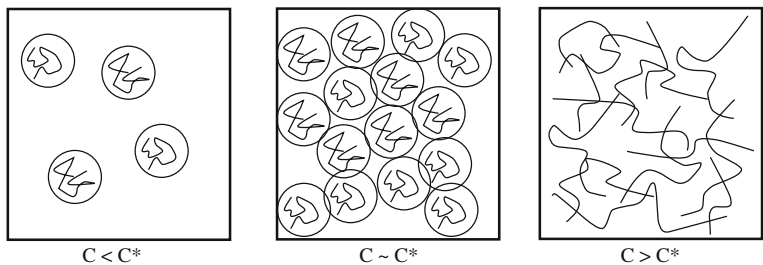


Fig. 11.43. Entanglement concentration. For $C > C^*$, a physical gel forms (non-covalent bonds between the polymers). This case is used to produce a network of obstacles that serves to separate the DNA molecules

speaks of a entangled polymer solution or semi-dilute solution. It is these entanglements that form the network of obstacles which selectively slows down the DNA molecules in accordance with their size.

Different DNA Migration Regimes in a Semi-Dilute Solution

The migration regimes described in this section are, to a first approximation, applicable to both gels and entangled polymer solutions. The migration regime depends on the radius of gyration of the DNA compared with the pore size in the polymer network. These models are summarised in Fig. 11.44. Regimes allowing separation of DNA are the Ogston regime (II) and the reptation regime (III). The pore sizes in entangled polymer solutions typically range from the nanometer to a hundred or so nanometers, leading to a possible

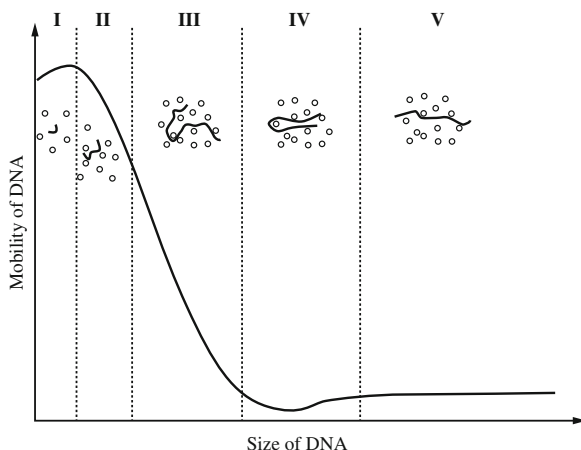


Fig. 11.44. Different DNA migration regimes. The diagram represents the mobility of the DNA as a function of its size in a polymer matrix. Circles represent polymer fibres. (I) The DNA fragments are very small and migrate as in a free solution. (II) The Ogston regime [2], where separation occurs as a filtration. (III) The DNA fragments are bigger than the pore size in the network and migrate by reptation. Their mobility goes as the reciprocal of their size. (IV) An intermediate regime. The molecules may be blocked in a U-shape, which generates a minimum in the mobility curve. (V) Migration with orientation. Chains are oriented by the field and separation is no longer possible

separation range from a few tens to around ten thousand DNA bases. Now, DNA separation requirements range from a few tens to a few million bases.

Surface Coating

Since the surface to volume ratio of capillaries and microchannels is very high, surface effects become more and more significant as separation systems are miniaturised. Hence, the polymer solution used as separation matrix will also play a key role on the walls. Indeed, owing to the presence of ionised silanol groups ($4\text{--}5$ silanols/ nm^2), at the pH at which separations are carried out (around 8), the silica capillary wall is negatively charged (see Fig. 11.45). The counterions in the solution will thus be more concentrated on the walls, over a distance of the order of the Debye length. When the electric field is applied, the positively charged solvent layer will be dragged toward the negative electrode. By viscosity, it will drag all the rest of the solvent with it, creating a flow known as electroosmotic flow. This flow has a profile in which the velocity is uniform over the whole cross-section of the capillary except along the walls, over a distance of roughly the Debye length (see Fig. 11.45 left). Hence, in principle, electroosmosis modifies only migration times, but not peak dispersion.

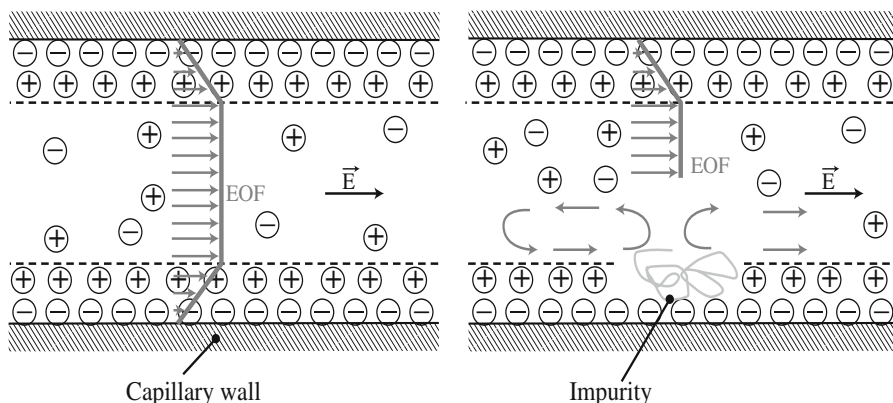


Fig. 11.45. Electroosmotic flow, with uniform charge density (*left*) and recirculation in the presence of heterogeneities (*right*). Negative charges correspond to silanol groups of the capillary wall. Positive charges correspond to the surplus of positive counterions over a distance of the order of the Debye length

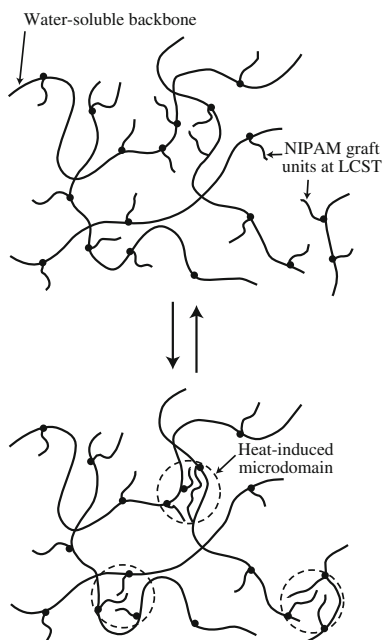


Fig. 11.46. Heat-sensitive polymers. *Top*: At room temperature, there is a little entanglement but the viscosity is relatively low. *Bottom*: At higher temperatures (the transition generally occurs in the range 30–40°C), poly-*N*-isopropylacrylamide micelles form, thus creating the obstacle network. From [202]

However, DNA samples for separation contain impurities, e.g., proteins. The latter adsorb onto the solid surfaces and thereby create charge heterogeneities, with associated electroosmotic flow. These heterogeneities can produce solvent recirculation or backflow (see Fig.11.45 right), leading to broadening of the peaks and non-reproducibility. After 3–4 separations, the

capillary may have become unusable. One of the great challenges of capillary electrophoresis is thus to prevent electroosmosis by decoupling the electrokinetic phenomena of the double layer on the walls from the rest of the fluid, by preventing the adsorption of solutes on the capillary walls. To achieve this, one solution is to coat the capillary wall with a layer of neutral polymers. The separation matrix can sometimes play this role by adsorbing onto the surface of the separation channel. (If not, one must appeal to more demanding strategies such as covalent immobilisation of neutral polymers.)

Examples of Polymer Matrices

Heat-Thickening Polymers

Entangled polymer solutions are rather viscous, which makes it difficult to fill capillaries and microchannels. One solution to this problem consists in synthesising polymers that are not too viscous at room temperature, but which become much more viscous when the temperature is raised [202]. They comprise an acrylamide backbone carrying NIPAM (N-isopropylacrylamide) units at the low critical solution temperature (LCST), which are soluble at low temperature and precipitate out as the temperature is raised. Hence, when the temperature increases, the NIPAM units form micelles measuring a few nanometers, separated by distances of a few tens of nanometers, and connected to one another via the chains of the backbone. A physical gel has thus been constituted by a self-organisation phenomenon related to an equilibrium between entropy and enthalpy. This creates the obstacle network required to carry out separations. In contrast to traditional gels, this gelification is reversible, and the medium becomes liquid once again at room temperature, so that it is easy to get the solution inside the channel (see Fig. 11.46). This self-organisation has been studied by neutron scattering [203]. A Bragg peak clearly appears when the temperature is raised, indicating the ordered nature of the medium, and providing information about its characteristics.

Dynamic Surface-Coating Polymers

We have seen that it is essential to coat the surface to obtain a good separation. In the beginnings of capillary electrophoresis, hydrophilic polymers were grafted covalently on the capillary walls [204]. However, this involves chemical treatment of each capillary, a tedious and costly process. In order to simplify operations, polymer solutions can be used that will play the role of the separation matrix as well as the surface coating. To do this, one can, as before, use a copolymer with a polyacrylamide backbone in order to obtain good separation properties. This time, polydimethylacrylamide (PDMA) graft units are associated with the backbone. The PDMA adsorbs onto the capillary surface and acrylamide loops then form, as shown in Fig. 11.47, preventing the adsorption of solutes on the capillary surface [205, 206]. These copolymers achieve good single-strand and double-strand DNA separations, as shown in Fig. 11.48.

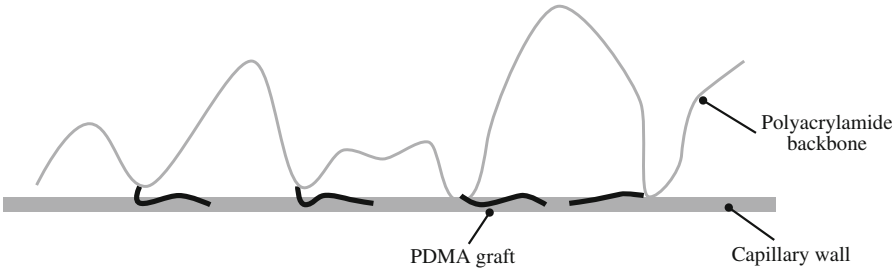


Fig. 11.47. Polymer adsorption at the capillary surface by PDMA grafts and the formation of acrylamide loops

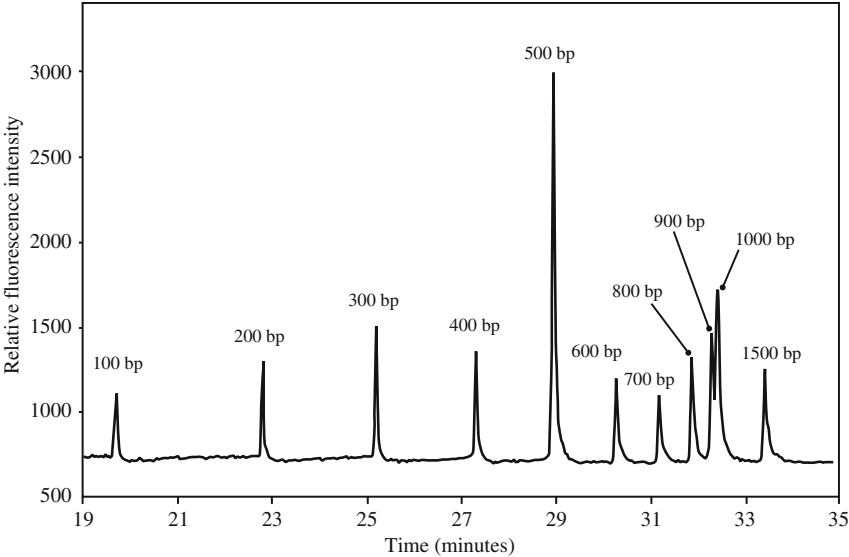


Fig. 11.48. Separation of double-stranded DNA in P(AM-PDMA)

11.4.4 Microfluidic Systems for Separation of Long DNA Fragments

The separation of DNA molecules is important in all size ranges. Gel electrophoresis can no longer discriminate these molecules above a size threshold of around 50,000 base pairs, because these are too big for the pore sizes that can be obtained in a gel (see Fig.11.44, regime V). These separations are nevertheless possible by using a pulsed field, i.e., by periodically changing the direction of the field. However, this pulsed-field technique is extremely slow, one separation typically requiring 24 h [207,208]. The separation of long DNA fragments has not given conclusive results with capillary electrophoresis either, owing to electrohydrodynamic instabilities [209–211].

To solve this problem, microfabrication techniques have provided a way of producing artificial gels, with a range of pore sizes that is better suited to the dimensions of these very long molecules. A variety of structures have been proposed over the last few years, using steric and/or entropic effects, and these have led to time-saving factors as great as 100 compared with conventional pulsed electrophoresis techniques on gels.

Microfabricated and Self-Assembled Obstacle Arrays

It was only a few years after the advent of the first microfluidic devices [212] that it was shown how to separate long DNA molecules in a record time using microlithographic obstacle arrays [213]. The idea is to use microfabrication techniques directly borrowed from microelectronics to design artificial gels with a pore size in the micrometer range. These arrays were originally fabricated in silicon. Today, there is a preference for transparent materials with better insulating properties, such as glasses and plastics. The latter also have the advantage of being somewhat cheaper, which opens interesting possibilities for routine analyses [214]. These arrays (see Fig. 11.49) have been used to show that there is a significant difference in mobility for DNA molecules with sizes above 100 kbp, contradicting the reptation model with orientation (see Fig. 11.44).

Although there can be no doubt that the idea of microlithographic obstacle arrays was inspired by the methods of gel electrophoresis, the possibility of observing DNA migration in the matrix by fluorescence microscopy made use of totally different separation mechanisms. These observations show that it is collisions between the molecules and the obstacles which distort the polyelectrolytes, immobilising them for a time that depends on the chain length [215]. In an array about 1 cm long, the DNA molecules will experience hundreds of collisions on the obstacles in the matrix before reaching the detection point. Each of these collisions can be divided into several stages, as shown in Fig. 11.50c:

1. The molecule collides with the obstacle and deforms from its statistical coil conformation (Fig. 11.50c1).
2. The DNA stretches into a U-shape around the obstacle (Fig. 11.50c2).
3. The molecule escapes with the same motion as a rope around a pulley (Fig. 11.50c3–c5).
4. Free at last, it relaxes into its coil conformation and another collision can begin (Fig. 11.50c6).

The duration of this cycle increases with the length of the molecule and the separation can be likened to a race along an obstacle course, in which the bigger competitors take longer to get past each obstacle. This is what leads to the potential for separation. Recent models have been used to calculate, for regular arrays, the speeds and speed dispersions for each size of DNA as

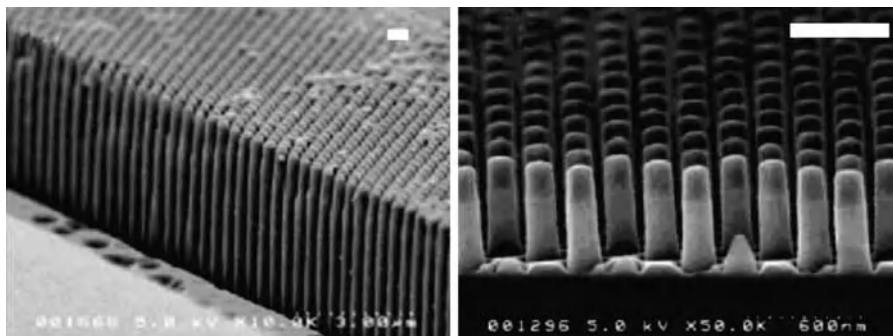


Fig. 11.49. Electron microscope image of obstacle arrays in quartz. Scale bars 1 μm . From [217]. Copyright 2004, American Chemical Society

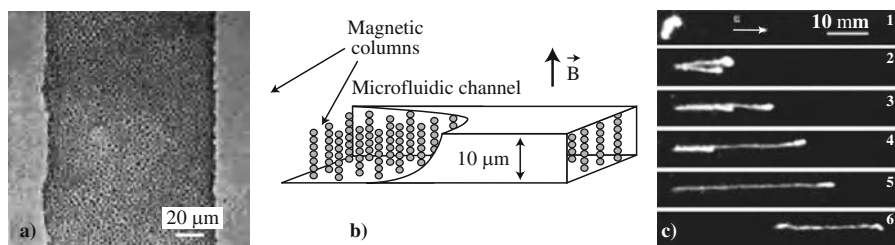


Fig. 11.50. (a) Microphotograph of the self-organised post array inside a separation channel, where a magnetic field is applied perpendicularly to the image plane. (b) Three-dimensional representation of the array in (a). (c) Collision of a single fluorescent DNA molecule with a magnetic column observed by fluorescence videomicroscopy (unpublished data)

a function of the obstacle density and the strength of the applied electric field [216].

More recently, other work [217] has improved the quality of these separations, mainly by increasing the density of the obstacles, this time made from quartz (see Fig. 11.49). These arrays have been used to discriminate DNA molecules in the same size range in a few tens of seconds.

An economical and convenient alternative to these microfabricated arrays consists in using separation matrices made from self-assembled magnetic columns or posts. To do this, one starts with a suspension of nanometric magnetic colloidal particles. When there is no magnetic field, these suspensions are fluid, but when they are subjected to a uniform magnetic field within a channel, they assemble into columns whose diameters, depending on conditions like the concentration of the emulsion and the channel height, vary between the size of a particle (a hundred or so nanometers) and several micrometers (see Fig. 11.50) [218]. This idea has been used [219,220] to separate large DNA molecules (up to about 150,000 base pairs), and separation times of the order

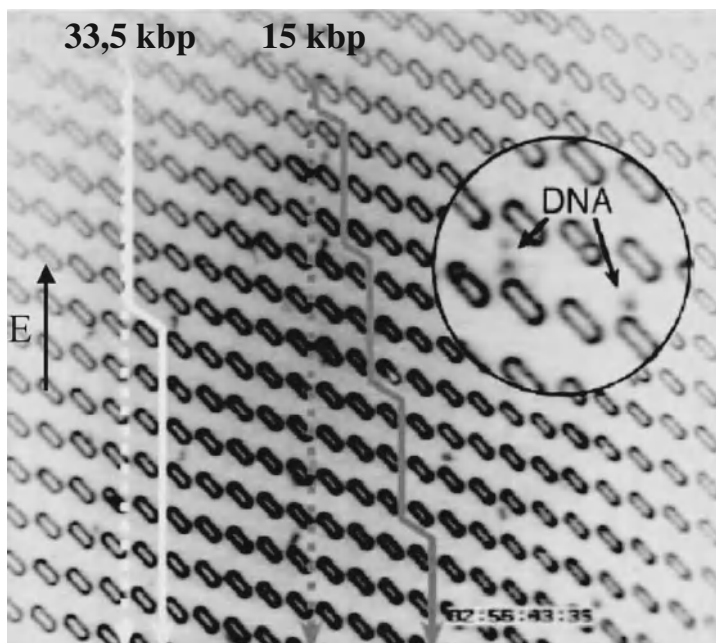


Fig. 11.51. Asymmetric silicon obstacle array and representation of two DNA molecules of different sizes separated by diffusion [221]. Copyright 1999 National Academy of Sciences, USA

of 1 min have been achieved. These magnetic columns offer an extremely economical alternative to systems fabricated by microlithography. Moreover, once the magnetic field has been switched off, a simple excess pressure suffices to rinse out the newly fluidified emulsion and the channels can be used once again.

Separation by Diffusion

The same group that made the first lithographic arrays developed asymmetric arrays a few years later, this time using the different diffusion constants of different-sized DNA molecules to separate them [221]. In Fig. 11.51, it can be seen that the obstacles are tilted relative to the direction of the electrical field (in the vertical direction). This time the molecules migrate without distortion and short DNA fragments, more sensitive to Brownian motion, will diffuse more quickly laterally, so that their trajectories will be much more significantly deflected by the obstacles (right-hand trajectory) than long DNA molecules (left-hand trajectory). The great advantage here is that it is in principle much easier to recover the separated products at the end of the channel than it would be in linear separation systems. This is a continuous separation, well suited to preparative micro (or nano!) applications.

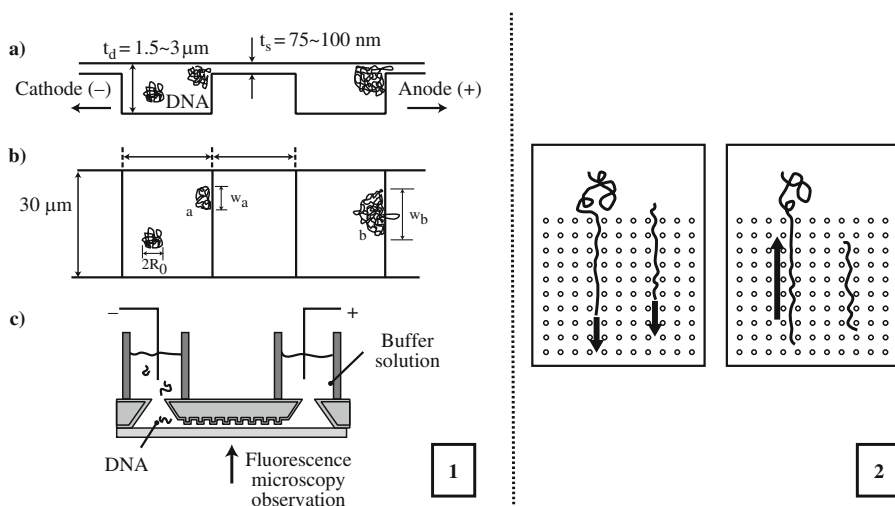


Fig. 11.52. (1) Separation by entropic trap array: (a) Side view of the channel showing the alternating shallow and deep regions. (b) Top view of the channel. The probability of escaping from the trap is proportional to the width w shown in the diagram. (c) Overall view of the device. From [222]. (2) Entropic recoil. Response of a short molecule (*left*) and a long molecule (*right*) to an electric field pulse. From [214]

Entropic Separation

Finally, macromolecules can be separated by exploiting the conformation and hence the entropy of the molecule. A first separation system making use of entropic traps was thus proposed to separate long DNA molecules [222]. This silicon device, shown in Fig. 11.52, consisted of a regular alternation of shallow regions (90 nm), shallower than the radius of gyration of all the molecules to be separated, and deeper regions. When a molecule migrates into a deeper region, it is in the coil conformation. It must then deform when it enters a shallow region. This alternating process produces a difference in mobility between the various DNA sizes and thus provides a way of separating them. Strangely, the authors report a higher mobility for the longer chains and suggest a model which shows that the probability of escaping from a trap is proportional to the contact area between the molecule and the deep region, and thus increases with the size of the molecule. These large molecules escape more quickly because they have more monomers which see the exit slit and which can form the bow head of the polymer.

The same group later suggested a system called entropic recoil [223], which had already been proposed theoretically [224]. The idea, shown in Fig. 11.52 is to force the DNA molecules onto an array of aligned dots with a spacing of a hundred nanometers or so, and then apply an electric field for a very short time, during which the molecules extend into the array. When the field is then

cut off, the small molecules find themselves trapped in the array and cannot coil up again because of the surrounding obstacles. The big molecules for their part will only have extended a part of their chains, the main part of the coil remaining outside the array. Their extended arm can then coil up once more and the center of mass will not have moved. By subsequently applying a series of pulses, the short molecule can be moved without affecting the big one, this leading to excellent separations. However, it would seem that the separation of many species at the same time is more complicated, since the system works in an all-or-nothing way, displacing all molecules below a certain size threshold and leaving the others where they were.

11.4.5 Conclusion

Electrophoretic separation techniques are still at the forefront of a good many analytical processes in biology. Gel electrophoresis devices are commonly found in any biology laboratory. Capillary electrophoresis equipment is indispensable for high-throughput genetic studies such as sequencing or the detection of mutations. These systems are more and more often replaced by miniaturised electrophoresis systems, for the moment mainly in research laboratories, but already for routine applications in the pharmaceutical industry. Such miniature systems bring the prospect of integrating different stages of an analytical sequence into a single chip, which should become a genuine lab-on-a-chip, simultaneously reducing the cost, the analysis time, and the required space. While it has been possible to transpose the separation of relatively short DNA fragments to miniature systems in a rather direct way, the separation of very long DNA molecules has provided a particularly fertile training ground for the imagination and given rise to new separation systems.

References

Section One. Ion Channels and Patch-Clamp Techniques

1. Changeux, J.-P.: *L'homme neuronal*, Fayard, Paris (1993)
2. Dubois, J.-M.: *Les canaux ioniques cellulaires*, Coll. Communication cellulaire (1999)
3. Joffre, M.: *Electrophysiologie moléculaire: la technique du patch-clamp*, Vol. I, Editeurs des Sciences et des Arts, Hermann (2001)
4. Joffre, M.: *Electrophysiologie moléculaire: propriétés, structure moléculaire et rôles physiologiques des canaux ioniques*, Vol. II, Editeurs des Sciences et des Arts, Hermann (2001)
5. Ashcroft, F.M.: Adenosine 5'-triphosphate-sensitive potassium channels, *Ann. Rev. Neurosc.* **11**, 97–118 (1988)
6. Bezanilla, F.: Voltage sensor movements, *J. Gen. Physiol.* **120**, 465–473 (2002)
7. Nagel, G., Ollig, D., Fuhrmann, M., Kateriya, S., Musti, A.M., Bamberg, E., Hegemann, P.: Channelrhodopsin-1: A light-gated proton channel in green algae, *Science* **296**, 2395–2398 (2002)

8. Nagel, G., Szellas, T., Huhn, W., Kateriya, S., Adeishvili, N., Berthold, P., Ollig, D., Hegemann, P., Bamberg, E.: Channelrhodopsin-2: A directly light-gated cation-selective membrane channel, *Proc. Natl. Acad. Sci. USA* **100** (24), 13940–13945 (2003)
9. www.hhmi.org/news/mackinnon.html: Website of R. McKinnon, Nobel Prize for Chemistry 2003
10. Ashcroft, F.: *Ion Channels and Disease: Channelopathies*, Academic Press (2000)
11. Niemeyer, B.A., Mery, L., Zawar, C., Suckow, A., Monje, F., Pardo, L.A., Stühmer, W., Flockerzi, V., Hoth, M.: Ion channels in health and disease, *EMBO Reports* **21** (71), 568–573 (2001)
12. Roden, D.M.: Drug-induced prolongation of the QT interval, *New England J. Med.* **19**, 2231–2238 (2004)
13. Tao, H., Santa, Ana, D., Guia, A., Huang, M., Ligutti, J., Walker, G., Sithipong, K., Chan, F., Guoliang, T., Zozulya, Z., Saya, S., Phimmachack, R., Sie, C., Yuan, J., Wu, L., Xu, J., Ghetti, A.: Automated tight seal electrophysiology for assessing the potential hERG liability of pharmaceutical compounds, *Assay Drug Dev. Technol.* **2** (5), 497–506 (2004)
14. Decker, K., Koschinski, A., Trouliaris, S., Tamura, T., Dreyer, F., Repp, H.: Activation of a Ca^{2+} -dependent K^{+} current by the oncogenic receptor protein tyrosine kinase v-Fms in mouse fibroblasts, *Naunyn Schmiedeberg's Arch. Pharmacol.* **357** (4), 378–384 (1998)
15. Fanger, C.M., Rauer, H., Neben, A.L., Miller, M.J., Rauer, H., Wulff, H., Rosa, J.C., Ganellin, C.R., Chandy, K.G., Cahalan, M.D.: Calcium-activated potassium channels sustain calcium signaling in T lymphocytes. Selective blockers and manipulated channel expression levels, *J. Biol. Chem.* **276** (15), 12249–12256 (2001)
16. Catacuzzeno, L., Pisconti, D.A., Harper, A.A., Petris, A., Franciolini, F.: Characterization of the large-conductance Ca-activated K channel in myocytes of rat saphenous artery, *Pflügers Arch.* **441** (2–3), 208–218 (2000)
17. Liu, J.J., Chao, J.R., Jiang, M.C., Ng, S.Y., Yen, J.J., Yang-Yen, H.F.: Ras transformation results in an elevated level of cyclin D1 and acceleration of G1 progression in NIH 3T3 cells, *Mol. Cell. Biol.* **15** (7), 3654–3663 (1995)
18. Ransom, C.B., Sontheimer, H.: BK channels in human glioma cells, *J. Neurophysiol.* **85** (2), 790–803 (2001). Erratum in: *J. Neurophysiol.* **85** (4): following table of contents (2001). *J. Neurophysiol.* **85** (6), 4 (2001)
19. Ouadid-Ahidouch, H., Roudbaraki, M., Ahidouch, A., Delcourt, P., Prevarsckaya, N.: Cell-cycle-dependent expression of the large Ca^{2+} -activated K^{+} channels in breast cancer cells, *Biochem. Biophys. Res. Commun.* **316** (1), 244–251 (2004)
20. Partiseti, M., Choquet, D., Diu, A., Korn, H.: Differential regulation of voltage- and calcium-activated potassium channels in human B lymphocytes, *J. Immunol.* **1**, **148** (11), 3361–3368 (1992)
21. Lin, C.S., Boltz, R.C., Blake, J.T., Nguyen, M., Talento, A., Fischer, P.A., Springer, M.S., Sigal, N.H., Slaughter, R.S., Garcia, M.L., et al.: Voltage-gated potassium channels regulate calcium-dependent pathways involved in human T lymphocyte activation, *J. Exp. Med.* **1**, **177** (3), 637–645 (1993)
22. Kodal, H., Weick, M., Moll, V., Biedermann, B., Reichenbach, A., Bringmann, A.: Involvement of calcium-activated potassium channels in the regulation of

- DNA synthesis in cultured Muller glial cells, *Invest. Ophthalmol. Vis. Sci.* **41** (13), 4262–4267 (2000)
23. Moll, V., Weick, M., Milenkovic, I., Kodal, H., Reichenbach, A., Bringmann, A.: P2Y receptor-mediated stimulation of Muller glial DNA synthesis, *Invest. Ophthalmol. Vis. Sci.* **43** (3), 766–773 (2002)
 24. Parihar, A.S., Coghlan, M.J., Gopalakrishnan, M., Shieh, C.C.: Effects of intermediate-conductance Ca^{2+} -activated K^{+} channel modulators on human prostate cancer cell proliferation, *Eur. J. Pharmacol.* **27**, **471** (3), 157–164 (2003)
 25. Basrai, D., Kraft, R., Bollensdorff, C., Liebmann, L., Benndorf, K., Patt, S.: BK channel blockers inhibit potassium-induced proliferation of human astrocytoma cells, *Neuroreport*. **25**, **13** (4), 403–407 (2002)
 26. Bringmann, A., Francke, M., Pannicke, T., Biedermann, B., Kodal, H., Faude, F., Reichelt, W., Reichenbach, A.: Role of glial K^{+} channels in ontogeny and gliosis: A hypothesis based upon studies on Muller cells, *Glia* **1**, **29** (1), 35–44 (2000)
 27. Sanguinetti, M.C., Jiang, C., Curran, M.E., Keating, M.T.: A mechanistic link between an inherited and an acquired cardiac arrhythmia: hERG encodes the IKr potassium channel, *Cell* **81** (2), 299–307 (1995)
 28. Cherubini, A., Taddei, G.L., Crociani, O., Paglierani, M., Buccoliero, A.M., Fontana, L., Noci, I., Borri, P., Borrani, E., Giachi, M., Becchetti, A., Rosati, B., Wanke, E., Olivotto, M., Arcangeli, A.: hERG potassium channels are more frequently expressed in human endometrial cancer as compared to non-cancerous endometrium, *Br. J. Cancer*. **83** (12), 1722–1729 (2000)
 29. Pillozzi, S., Brizzi, M.F., Balzi, M., Crociani, O., Cherubini, A., Guasti, L., Bartolozzi, B., Becchetti, A., Wanke, E., Bernabei, P.A., Olivotto, M., Pegoraro, L., Arcangeli, A.: hERG potassium channels are constitutively expressed in primary human acute myeloid leukemias and regulate cell proliferation of normal and leukemic hemopoietic progenitors, *Leukemia* **16** (9), 1791–1798 (2002)
 30. Smith, G.A., Tsui, H.W., Newell, E.W., Jiang, X., Zhu, X.P., Tsui, F.W., Schlichter, L.C.: Functional up-regulation of hERG K^{+} channels in neoplastic hematopoietic cells, *J. Biol. Chem.* **24**, **277** (21), 18528–18534 (2002)
 31. Wang, H., Zhang, Y., Cao, L., Han, H., Wang, J., Yang, B., Nattel, S., Wang, Z.: hERG K^{+} channel, a regulator of tumor cell apoptosis and proliferation, *Cancer Res.* **1**, **62** (17), 4843–4848 (2002)
 32. Bianchi, L., Wible, B., Arcangeli, A., Tagliatela, M., Morra, F., Castaldo, P., Crociani, O., Rosati, B., Faravelli, L., Olivotto, M., Wanke, E.: hERG encodes a K^{+} current highly conserved in tumors of different histogenesis: A selective advantage for cancer cells?, *Cancer Res.* **15**, **58** (4), 815–822 (1998)
 33. Arcangeli, A., Bianchi, L., Becchetti, A., Faravelli, L., Coronello, M., Mini, E., Olivotto, M., Wanke, E.: A novel inward-rectifying K^{+} current with a cell-cycle dependence governs the resting potential of mammalian neuroblastoma cells, *J. Physiol.* **1**, **489** (2), 455–471 (1995)
 34. Crociani, O., Guasti, L., Balzi, M., Becchetti, A., Wanke, E., Olivotto, M., Wymore, R.S., Arcangeli, A.: Cell cycle-dependent expression of hERG1 and hERG1B isoforms in tumor cells, *J. Biol. Chem.* **31**, **278** (5), 2947–2955 (2003)
 35. Lastraioli, E., Guasti, L., Crociani, O., Polvani, S., Hofmann, G., Witchel, H., Bencini, L., Calistri, M., Messerini, L., Scatizzi, M., Moretti, R., Wanke, E.,

- Olivotto, M., Mugnai, G., Arcangeli, A.: hERG1 gene and hERG1 protein are overexpressed in colorectal cancers and regulate cell invasion of tumor cells, *Cancer Res.* 15, **64** (2), 606–611 (2004)
36. Skryma, R.N., Prevarskaya, N.B., Dufy-Barbe, L., Odessa, M.F., Audin, J., Dufy, B.: Potassium conductance in the androgen-sensitive prostate cancer cell line, LNCaP: Involvement in cell proliferation, *Prostate* 1, **33** (2), 112–122 (1997)
 37. Fraser, S.P., Grimes, J.A., Djamgoz, M.B.: Effects of voltage-gated ion channel modulators on rat prostatic cancer cell proliferation: Comparison of strongly and weakly metastatic cell lines, *Prostate* 15, **44** (1), 61–76 (2000)
 38. Laniado, M.E., Fraser, S.P., Djamgoz, M.B.: Voltage-gated K⁺ channel activity in human prostate cancer cell lines of markedly different metastatic potential: Distinguishing characteristics of PC-3 and LNCaP cells, *Prostate* 1, **46** (4), 262–274 (2001)
 39. Carafoli, E.: Calcium signaling: A tale for all seasons, *Proc. Natl. Acad. Sci. USA* 5, **99** (3), 1115–1122 (2002)
 40. Hamill, O.P., Marty, A., Neher, E., Sakmann, B., Sigworth, F.J.: Improved patch-clamp techniques for high-resolution current recording from cells and cell-free membrane patches, *Pflugers Arch.* **391** (2), 85–100 (1981)
 41. Lindau, M., Neher, E.: Patch-clamp techniques for time-resolved capacitance measurements in single cells, *Pflugers Arch.* **411** (2), 137–146 (1988)
 42. Sakmann, B., Neher, E.: Patch-clamp techniques for studying ionic channels in excitable membranes, *Annu. Rev. Physiol.* 46455–46472 (1984)
 43. Neher, E.: Nobel lecture: Ion channels for communication between and within cells, *Embo. J.* **11**, 1679–1692 (1992)
 44. Teulon, J.: Le patch-clamp en bref, *M/S* 5, **20**, 1–2 (2004)
 45. *The Axon Guide for Electrophysiology and Biophysics Laboratory Techniques*: www.moleculardevices.com/transition/ (1993)
 46. Wang, M.H.: A technical consideration concerning the removal of oocyte vitelline membranes for patch-clamp recording, *Biochem. Biophys. Res. Com.* **324**, 971–972 (2004)
 47. Trapani, J.G., Korn, S.J.: Control of ion channel expression for patch-clamp recordings using an inducible expression system in mammalian cell lines, *BMC Neurosci.* 2, **4** (1), 15 (2003)
 48. Palma, E., Torchia, G., Limatola, C., Trettel, F., Arcella, A., Cantore, G., DiGennaro, G., Manfredi, M., Esposito, V., Quarato, P.P., Miledi, R., Eusebi, F.: BDNF modulates GABA_A receptors microtransplanted from the human epileptic brain to *Xenopus* oocytes, *Proc. Natl. Acad. Sci. USA* 102, **5**, 1667–1672 (2005)
 49. Rae, J.L., Levis, R.A.: Glass technology for patch-clamp electrodes, *Methods in Enzym.* **207**, 66–92 (1992)
 50. Heyward, P.M., Shipley, M.T.: A device for automated control of pipette internal pressure for patch-clamp recording, *J. Neurosci. Methods.* 15, **123** (1), 109–115 (2003)
 51. Mason, M.J., Simpson, A.K., Mahaut-Smith, M.P., Robinson, H.P.: The interpretation of current-clamp recordings in the cell-attached patch-clamp configuration, *Biophys. J.* **88** (1), 739–750 (2005)
 52. Vivaudou, M.B., Arnoult, C., Villaz, M.: Skeletal muscle ATP-sensitive K⁺ channels recorded from sarcolemmal blebs of split fibers: ATP inhibition is reduced by magnesium and ADP, *J. Membrane Biol.* **122**, 165–175 (1991)

53. Sackmann, B., Neher, E. (Eds.): *Single-Channel Recording*, 2nd edn., Plenum Press, New York (1995)
54. Jamieson, C.V., MacDonald, A.G.: A method of patch-clamp recording in hyperbaric oxygen, *Undersea Hyperb. Med.* **29** (4), 286–293 (2002)
55. Hutzler, M., Fromherz, P.: Silicon chip with capacitors and transistors for interfacing organotypic brain slice of rat hippocampus, *Eur. J. Neurosc.* **19**, 2231–2238 (2004)
56. Braun, D., Fromherz, P.: Imaging neuronal resistance on silicon chip using fluorescent voltage-sensitive dye, *Biophys. J.* **87**, 1351–1359 (2004)
57. Willumsen, N.J., Bech, M., Olesen, S.P., Jensen, B.S., Korsgaard, M.P., Christophersen, P.: High throughput electrophysiology: New perspectives for ion channel drug discovery, *Receptors and Channels* **9** (1), 3–12 (2003)
58. Owen, D., Silverthorne, A.: Channelling drug discovery, current trends in ion channel discovery research, *Drug Discovery World* 348–361 (2002)
59. Xu, J., Chen, Y., Li, M.: High-throughput technologies for studying potassium channels-progresses and challenges, *Drug Discovery Today* **3** (1), 32–38 (2004)
60. Wood, C., Williams, C., Waldron, G.J.: Patch-clamping by numbers, *Drug Discovery Today* **9** (10), 434–441 (2004)
61. Wang, X., Li, M.: Automated electrophysiology: High throughput of art, *Assay Drug Dev. Technol.* **1** (5), 695–708 (2003)
62. Stett, A., Bucher, V., Burkhardt, C., Weber, U., Nisch, W.: Patch-clamping of primary cardiac cells with micro-openings in polyimide films, *Med. Biol. Eng. Comput.* **41** (2), 233–240 (2003)

Section Two. Amperometry

63. Amatore, C., Szunerits, S., Thouin, L., Warkocz, J.-S.: *J. Electroanal. Chem.* **500**, 62 (2001)
64. Baltes, N., Thouin, L., Amatore, C., Heinze, J.: *Angew. Chem. Int. Ed.* **43**, 1431 (2004)
65. Bard, A.J., Faulkner, L.R.: *Electrochemical Methods*, John Wiley (2001)
66. Southampton Electrochemistry Group: *Instrumental Methods in Electrochemistry*, Ellis Horwood (1990)
67. Amatore, C.: In: *Physical Electrochemistry*, ed. by I. Rubinstein, Marcel Dekker, New York (1995) p. 131
68. Wightman, R.M., Wipf, D.O.: In: *Electroanalytical Chemistry*, Vol. 15, ed. by A.J. Bard, Marcel Dekker, New York (1989) p. 267
69. Montenegro, M.I., Queiros, M.A., Daschbach, J.L.: *Microelectrodes: Theory and Applications*, Kluwer Academic Press, Dordrecht (1991)
70. Amatore, C., Arbault, S., Maisonhaute, E., Szunerits, S., Thouin, L.: In: *New Trends in Molecular Electrochemistry*, Vol. 12, ed. by C. Amatore, A. Pombeiro, FontisMedia, Lausanne (2004) p. 385
71. Amatore, C., Szunerits, S., Thouin, L. et Warkocz, J.-S.: *Electroanalysis* **13**, 646 (2001)
72. Amatore, C., Szunerits, S., Thouin: *Electrochem. Commun.* **2**, 248 (2000)
73. Amatore, C., Szunerits, S., Thouin, L., Warkocz, J.-S.: *Electrochem. Commun.* **2**, 353 (2000)
74. Amatore, C., Pebay, C., Scialdone, O., Szunerits, S., Thouin, L.: *Chem. Eur. J.* **7**, 2933 (2001)

75. Amatore, C., Knobloch, K., Thouin, L.: *Electrochem. Commun.* **6**, 887 (2004)
76. Bard, A.J., Mirkin, M.V.: *Scanning Electrochemical Microscopy*, Marcel Dekker, New York, Basel (2001)
77. Jankowski, J.A., Schroeder, T.J., Holz, R.W., Wightman, R.M.: *J. Biol. Chem.* **267**, 18329 (1992)
78. Finnegan, J.M., Pihel, K., Cahill, P.S., Huang, L., Zerby, S.E., Ewing, A.G., Kennedy, R.T., Wightman, R.M.: *J. Neurochem.* **66**, 1914 (1996)
79. Wightman, R.M., Jankowski, J.A., Kennedy, R.T., Kawagoe, K.T., Schroeder, T.J., Leszczyszyn, D.J., Near, J.A., Diliberto, E.J., Viveros, O.H.: *Proc. Natl. Acad. Sci. USA* **88**, 10754 (1991)
80. Nanavati, C., Fernandez, J.M.: *Science* **259**, 963 (1993)
81. Borges, R., Machado, J.D., Alonso, C., Brioso, M.A., Gomez, J.F.: In: *Chromogranins: Functional and Clinical Aspects*, Vol. 482, Kluwer Academic Publishers, Amsterdam (2000) p. 69
82. Galli, T., Martinez-Arca, S., Paumet, F.: *MS-Médecine Sciences* **18**, 1113 (2002)
83. Burgoyne, R.D., Fisher, R.J., Graham, M.E., Haynes, L.P., Morgan, A.: *Biochem. Soc. Trans.* **29**, 467 (2001)
84. Humeau, Y., Doussau, F., Grant, N.J., Poulain, B.: *Biochimie* **82**, 427 (2000)
85. Rizo, J., Sudhof, T.C.: *Nat. Struct. Biol.* **5**, 839 (1998)
86. Peters, C., Bayer, M.J., Buhler, S., Andersen, J.S., Mann, M., Mayer, A.: *Nature* **409**, 581 (2001)
87. Chen, Y.A., Scheller, R.H.: *Nat. Rev. Mol. Cell Biol.* **2**, 98 (2001)
88. Almers, W.: *Nature* **409**, 567 (2001)
89. Weber, T., Zemelman, B.V., McNew, J.A., Westermann, B., Gmachl, M., Parlati, F., Sollner, T.H., Rothman, J.E.: *Cell* **92**, 759 (1998)
90. Amatore, C., Bouret, Y., Travis, E.R., Wightman, R.M.: *Biochimie* **82**, 481 (2000)
91. Amatore, C., Arbault, S., Bonifas, I., Bouret, Y., Erard, M., Guille, M.: *Chemphyschem.* **4**, 147 (2003)
92. Zimmerberg, J., Curran, M., Brodwick, M., Cohen, F.S.: *Biophys. J.* **51**, A357 (1987)
93. Chow, R.H., Vonruden, L., Neher, E.: *Nature* **356**, 60 (1992)
94. Zenisek, D., Steyer, J.A., Almers, W.: *Nature* **406**, 849 (2000)
95. Oheim, M., Loerke, D., Stuhmer, W., Chow, R.H.: *Eur. Biophys. J. Biophys.* **27**, 83 (1998)
96. Amatore, C., Arbault, S., Bonifas, I., Bouret, Y., Erard, M., Ewing, A.G., Sombers, L.A.: *Biophys. J.* **88**, 4411 (2005)
97. Sombers, L.A., Hanchar, H.J., Colliver, T.L., Wittenberg, N., Cans, A., Arbault, S., Amatore, C., Ewing, A.G.: *J. Neurosci.* **24**, 303 (2004)
98. Amatore, C., Bouret, Y., Midrier, L.: *Chem. Eur. J.* **5**, 2151 (1999)
99. Amatore, C., Bouret, Y., Travis, E.R., Wightman, R.M.: *Angew. Chem. Int. Edit.* **39**, 1952 (2000)
100. Wink, D.A., Mitchell, J.B.: *Free Rad. Biol. Med.* **25**, 434 (1998)
101. Dedon, P.C., Tannenbaum, S.R.: *Arch. Biochem. Biophys.* **423**, 12 (2004)
102. McCord, J.M.: *Am. J. Med.* **108**, 652 (2000)
103. Halliwell, B., Gutteridge, J.M.C.: *Free Radicals in Biology and Medicine*, Oxford University Press, Oxford (1999)
104. Squadrito, G.L., Pryor, W.A.: *Free Rad. Biol. Med.* **25**, 392 (1998)

105. Bergamini, C.M., Gambetti, S., Dondi, A., Cervellati, C.: *Curr. Pharm. Design* **10**, 1611 (2004)
106. Kraemer, K.H., Lee, M.M., Andrews, A.D., Lambert, W.C.: *Arch. Dermatol.* **130**, 1018 (1994)
107. Grisham, M.B., Granger, D.N., Lefer, D.J.: *Free Rad. Biol. Med.* **25**, 404 (1998)
108. Valko, M., Morris, H., Cronin, M.T.D.: *Curr. Med. Chem.* **12**, 1161 (2005)
109. Ignarro, L.J., Buga, G.M., Wood, K.S., Byrns, R.E., Chaudhuri, G.: *Proc. Natl. Acad. Sci. USA* **84**, 9265 (1987)
110. Cadet, J., Sage, E., Douki, T.: *Mutat. Res. Fund. Mol. M.* **571**, 3 (2005)
111. Arbault, S., Pantano, P., Jankowski, J.A., Vuillaume, M., Amatore, C.: *Anal. Chem.* **67**, 3382 (1995)
112. Arbault, S., Pantano, P., Sojic, N., Amatore, C., Best-Belpomme, M., Sarasin, A., Vuillaume, M.: *Carcinogenesis* **18**, 569 (1997)
113. Amatore, C., Arbault, S., Bruce, D., De Oliveira, P., Erard, M., Vuillaume, M.: *Faraday Discuss.* **116**, 319 (2000)
114. Lachgar, A., Sojic, N., Arbault, S., Bruce, D., Sarasin, A., Amatore, C., Bizzini, B., Zagury, D., Vuillaume, M.: *J. Virol.* **73**, 1447 (1999)
115. Porterfield, D.M., Laskin, J.D., Jung, S.K., Malchow, R.P., Billack, B., Smith, P.J.S., Heck, D.E.: *Am. J. Physiol. Lung C.* **281**, L904 (2001)
116. Sary, A., Robert, C., Sarasin, A. *Mutat. Res. Dna Repair* **383**, 1 (1997)
117. Amatore, C., Arbault, S., Bruce, D., De Oliveira, P., Erard, M., Sojic, N., Vuillaume, M.: *Analisis* **28**, 506 (2000)
118. Arbault, S., Edeas, M., Legrand-Poels, S., Sojic, N., Amatore, C., Piette, J., Best-Belpomme, M., Lindenbaum, A., Vuillaume, M.: *Biomed. and Pharmacother.* **51**, 430 (1997)
119. Amatore, C., Arbault, S., Bruce, D., de Oliveira, P., Erard, M., Vuillaume, M.: *Chem. Eur. J.* **7**, 4171 (2001)

Section Three. Macromolecular Transport Through Nanopores. Electrical Detection

120. Citovsky, V., Zambryski, P.: Transport of nucleic acids through membrane channels: Snaking through small holes, *Annu. Rev. Microbiol.* **47**, 167–197 (1993)
121. Matlack, K., Mothes, W., Rapoport, T.: Protein translocation: Tunnel vision, *Cell* **92**, 381–390 (1998)
122. Berrier, C., Bonhivers, M., Letellier, L., Ghazi, A.: High conductance channel induced by the interaction of phage lambda with its receptor maltoporin, *FEBS Letters* **476**, 129–133 (2000); Lambert, O., Letellier, L., Gelbart, W.M., Rigaud, J.L.: DNA delivery by phage as a strategy for encapsulating toroidal condensates of arbitrary size into liposomes, *Proc. Natl. Acad. Sci.* **97**, 7248–7253 (2000)
123. Guillot, G., Léger, L., Rondelez, F.: Diffusion of large flexible chains through model porous membranes, *Macromolecules* **18**, 2531–2537 (1985)
124. Lal, J., Sinha, S.K., Auvray, L.: Structure of polymer chains confined in Vycor, *J. Phys. II* **7** 1597–1615 (1997); Gilbert, E.P., Auvray, L., Lal, J.: Structure of polyelectrolyte chains confined in nanoporous glass, *Macromolecules* **34**, 4942–4948 (2001)

125. Kasianowicz, J.J., Brandin, E., Branton, D., Deamer, D.W.: Characterization of individual polynucleotide molecules using a membrane channel, *Proc. Natl. Acad. Sci.* **93**, 13770–13773 (1996)
126. Church, G.M., Deamer, D.W., Branton, D., Baldarelli, R., Kasianowicz, J.: Characterization of individual polymer molecules based on monomer–interface interactions, US Patent 5,795,782, Aug. 18 (1998)
127. Coulter, W.H.: Means for counting particles suspended in fluid, US Patent 2656508, Oct. 20 (1953)
128. Sakmann, B., Neher, E. (Eds.): *Single Channel Recording*, Plenum Press (1995)
129. Hille, B.: *Ion Channels of Excitable Membranes*, 3rd edn., Sinauer Associates (2001)
130. Mueller, P., Rudin, D.O., Ti Tien, H., Wescott, W.C.: Reconstitution of cell membrane structure in vitro and its transformation into an excitable system, *Nature* **194**, 979–980 (1962); Mueller, P., Rudin, D.O., Ti Tien, H., Wescott, W.C.: Methods for the formation of single bimolecular lipid membranes in aqueous solution, *J. Phys. Chem.* **67**, 534–535 (1963)
131. Zimmerberg, J., Parsegian, V.A.: Polymer inaccessible volume changes during opening and closing of a voltage-dependent ionic channel, *Nature* **323**, 36–39 (1986)
132. Bezrukov, S.M., Vodyanoy, I., Parsegian, V.A.: Counting polymers moving through a single ion channel, *Nature* **370**, 279–281 (1994)
133. Note in [125] and [126]
134. Lubensky, D.K., Nelson, D.R.: *Biophys. J.* **77**, 1824 (1999); Muthukumar, M.: *J. Chem. Phys.* **111**, 10371 (1999); Chuang, J., Kantor, Y., Kardar, M.: *Phys. Rev. E* **65**, 011802 (2002)
135. Akeson, M., Branton, D., Kasianowicz, J.J., Brandin, E., Deamer, D.W.: *Biophys. J.* **77**, 3227 (1999)
136. Bates, M., Burns, M., Meller, A.: *Biophys. J.* **84**, 2366 (2003)
137. Howorka, S., Nam, J., Bayley, H., Kahne, D.: *Angew. Chem. Int. Ed.* **43**, 842 (2004)
138. Movileanu, L., Cheley, S., Bayley, H.: *Biophys. J.* **85**, 897 (2003)
139. Szabo, I., Bathori, G., Tombola, F., Brini, M., Coppola, A., Zoratti, M.: DNA translocation across planar bilayers containing *Bacillus subtilis* ion channels, *J. Biol. Chem.* **272**, 25275–25282 (1997)
140. Bacri, L., Benkhaled, A., Guegan, P., Auvray, L.: Ionic channel behavior of modified cyclodextrins inserted in lipid membranes, *Langmuir* **21** (13), 5842–5846 (2005)
141. Sun, L., Crooks, R.: Single carbon nanotube membranes: A well-defined model for studying mass transport through nanoporous materials, *J. Am. Chem. Soc.* **122** (49), 12340–12345 (2000)
142. Siwy, Z., Apel, P., Dobrev, D., Neumann, R., Spohr, R., Trautmann, C., Voss, K.: Ion transport through asymmetric nanopores prepared by ion track etching, *Nucl. Instr. Meth. Phys. Res. B* **208**, 143–148 (2003)
143. Li, J., Stein, D., McMullan, C., Branton, D., Aziz, M.J., Golovchenko, J.A.: Ion beam sculpting at nanometer length scales, *Nature* **412**, 166–169 (2001)
144. Biance, A.-L., Gierak, J., Bourhis, E., Madouri, A., Lafosse, X., Patriarche, G., Oukhaled, G., Ulysse, C., Galas, J.-C., Chen, Y., Auvray, L.: Focused ion beam sculpted membranes for nanoscience tooling, *Microelectronic Engineering* **83**, 1474–1477 (2006)

145. Storm, A.J., Chen, J.H., Ling, X.S., Zandbergen, H.W., Dekker, C.: Fabrication of solid-state nanopores with single-nanometer precision, *Nature Materials* **2**, 537 (2003)
146. Saleh, O.A.: A novel resistive pulse sensor for biological measurements, PhD thesis, Princeton University (2003)
147. DeBlois, R.W., Bean, C.P.: Counting and sizing of submicron particles by the resistive pulse technique, *Rev. Sci. Instr.* **41** (7), 909 (1970)
148. Fleischer, R.L., Alter, H.W., Furman, S.C., Price, P.B., Walker, R.M.: Particle track etching: Diverse technological uses range from virus identification to uranium exploration, *Science* **178**, 255–263 (1972)
149. Pasternak, C.A., Alder, G.M., Appel, P.Y., Bashford, C.L., Korchev, Y.E., Lev, A.A., Rostovtseva, T.K., Zhitariuk, N.I.: Model pores for biological membranes: The properties of track-etched membranes, *NIM B* **105**, 332–334 (1995)
150. Mara, A., Siwy, Z., Trautmann, C., Wang, J., Kamme, F.: An asymmetric polymer nanopore for single molecule detection, *Nano Letters* **4**, 497–501 (2004); Heins, E.A., Siwy, Z., Baker, L.A., Martin, C.H.: Detecting single porphyrin molecules in a conically shaped synthetic nanopore, *Nano Letters* **5** (9), 1824–1829 (2006)
151. Bayley, H., Martin, C.: Resistive-pulse sensing – From microbes to molecules, *Chem. Rev.* **100**, 2575–2594 (2000)
152. Meller, A.: Dynamics of polynucleotide transport through nanometre-scale pores, *J. Phys.: Condens. Matter* **15**, R581–R607 (2003)
153. Kasianowicz, J.J., Kellermayer, M. (Eds.): *Structure and Dynamics of Confined Polymers*, Kluwer Academic Publisher (2002)
154. Mathé, J., Visram, H., Viasnoff, V., Rabin, Y., Meller, A.: Nanopore force spectroscopy of individual DNA hairpin molecules, *Biophys. J.* **87**, 3205 (2004); Mathé, J., Aksimentiev, A., Nelson, D.R., Schulten, K., Meller, A.: Orientation discrimination of single-stranded DNA inside the α -hemolysin membrane channel, *Proc. Natl. Acad. Sci.* **102**, 12377–12382 (2005)
155. Jeans, J.: *The Mathematical Theory of Electricity and Magnetism*, Cambridge University Press (1925); Hall, J.E.: Access resistance of a small circular pore, *J. Gen. Phys.* **66**, 531–532 (1975)
156. Israelachvili, J.: *Intermolecular and Surface Forces*, Academic Press (1992)
157. Roux, B.: Ion conduction and selectivity in K^+ channels, *Ann. Rev. Biomol. Struc. Dyn.* **34**, 153–171 (2005)
158. Aksimentiev, A., Schulten, K.: Imaging alpha-hemolysin with molecular dynamics: Ionic conductance, osmotic permeability and the electrostatic potential map, *Biophys. J.* **88**, 3745–3761 (2005)
159. Noskov, S.Y., Im, W., Roux, B.: Ion permeation through the alpha-hemolysin channel: Theoretical studies based on Brownian dynamics and Poisson–Nernst–Planck electrodiffusion theory, *Biophys. J.* **87**, 2299–2309 (2004)
160. Levine, S., Marriott, J.R., Robinson, K.: Theory of electrokinetic flow in a narrow parallel-plate channel, *J. Chem. Soc., Faraday Trans.* **2**, 1–11 (1975)
161. Stein, D., Kruithof, M., Dekker, C.: *Phys. Rev. Lett.* **93**, 035901-1–035901-5 (2004)
162. Daiguji, H., Yang, P., Majumdar, A.: *Nanolett.* **4**, 137–142 (2004)
163. Smythe, W.R.: *Static and Dynamic Electricity*, McGraw-Hill (1968) p. 204; Durand, E.: *Electrostatique*, Vol. III, Masson (1966) p. 234

164. Parsegian, A.: Energy of an ion crossing a low dielectric membrane: Solutions to four relevant electrostatic problems, *Nature* **221**, 844–846 (1969); Levitt, D.G.: Electrostatic calculations for an ion channel, *Biophys. J.* **22**, 209–218 (1978); Jordan, P.J.: Electrostatic modelling of ion pores, *Biophys. J.* **39**, 157–164 (1982)
165. Auvray, L.: Sur quelques propriétés de solutions de macromolécules rigides et de polyélectrolytes dans un milieu poreux modèle, MSc thesis, University of Paris-Sud (1982)
166. Maxwell, J.C.: *A Treatise on Electricity and Magnetism*, Vol. III, Clarendon Press, Oxford (1891) p. 440. Reprinted by Dover
167. Berge, L., Jøssang, T., Feder, J.: Off-axis response for particles passing through long apertures in Coulter-type counters, *Meas. Sci. Technol.* **1**, 471–474 (1990)
168. Happel, J., Byrne, B.J.: Motion of a sphere and fluid in a cylindrical tube, *Ind. Eng. Chem.* **46**, 1181–1186 (1954)
169. Anderson, J.L., Quinn, J.A.: Restricted transport in small pores. A model for steric exclusion and hindered particle motion, *Biophys. J.* **14**, 131–150 (1974)
170. Cassasa, E.F.: *J. Polymer Sci. B* **5**, 773 (1967)
171. Daoud, M., de Gennes, P.G.: Statistics of macromolecular solutions trapped in small pores, *J. Physique* **38**, 85–93 (1977)
172. de Gennes, P.G.: *Scaling Concepts in Polymer Physics*, Cornell University Press (1979)
173. Wall, F., Seitz, W.A., Chin, J.C., de Gennes, P.G.: Statistics of self-avoiding walks confined to strips and capillaries, *Proc. Natl. Acad. Sci. USA* **75**, 2069–2070 (1978)
174. Auroy, P., Auvray, L., Léger, L.: Characterization of the brush regime for grafted polymer layers at the solid–liquid interface, *Phys. Rev. Lett.* **66**, 719–722 (1991)
175. Reisner, W., et al.: *Phys. Rev. Lett.* **94**, 196101 (2005)
176. Movileanu, L., Bayley, H.: Partitioning of a polymer into a nanoscopic protein pore obeys a simple scaling law, *Proc. Natl. Acad. Sci. USA* **98** (18), 10137–10141 (2001)
177. Oukhaled, A., Amiel, C., Bacri, L., Raphaël, E., Sikorav, J.L., Auvray, L.: Transport of single neutral and charged macromolecules through proteinic nanopores, *Proceedings of the 40th Intern. IUPAC Symp. on Macromolecules*, Paris 4–9 July 2004, e-Polymers 2005, no. E002, L1188 (2005)
178. de Gennes, P.-G.: Flexible polymers in nanopores, *Adv. Poly. Sci.* **138**, 92–105 (1999)
179. Brochard, F., de Gennes, P.-G.: Dynamics of confined polymer chains, *J. Chem. Phys.* **67**, 52–56 (1977)
180. Lee, N., Obukhov, S.: Diffusion of a polymer chain through a thin membrane, *J. Phys. II France* **6**, 195–204 (1996)
181. Sung, W., Park, P.J.: Polymer translocation through a pore in a membrane, *Phys. Rev. Lett.* **77**, 783 (1996)
182. Lubensky, D.K., Nelson, D.R.: Driven polymer translocation through a narrow pore, *Biophys. J.* **77**, 1824 (1999)
183. Muthukumar, M.: Polymer translocation through a hole, *J. Chem. Phys.* **111**, 10371 (1999)
184. Chuang, J., Kantor, Y., Kardar, M.: Anomalous dynamics of translocation, *Phys. Rev. E* **65**, 011802 (2001)

185. Sakaue T., Raphaël, E.: Dynamics of polymer translocation, *Macromolecules* **39**, 2621–2628 (2006)
186. Oosawa, F.: *Polyelectrolytes*, Marcel Dekker, New York (1971)
187. de Gennes, P.-G., Pincus, P., Velasco, R.M., Brochard, F.: *J. de Phys.* **37**, 1461 (1976)
188. Barrat, J.L., Joanny, J.F.: *Advances in Chemical Physics*, Vol. XCIV (1996)
189. Meller, A., Branton, D.: Single molecule measurements of DNA transport through a nanopore, *Electrophoresis* **23**, 2583–2591 (2002)
190. Storm, A.J., Storm, C., Chen, J., Zandbergen, H., Joanny, J.F., Dekker, C.: Fast DNA translocation through a solid-state nanopore, *Nanolett.* **5**, 1193–1197 (2005)
191. Gilbert, E.P., Auvray, L., Lal, J.: Structure of polyelectrolyte chains confined in nanoporous glass, *Macromolecules* **34**, 4942–4998 (2001)
192. Israelachvili, J.: *Intermolecular and Surface Forces*, Academic Press (1992)
193. Schechter, E.: *Biochimie et biophysique des membranes*, Masson (1997)
194. Miller, C. (Ed.): *Ion Channel Reconstitution*, Plenum Publishing (1985)
195. Song, L., Hobaugh, M.R., Shustak, C., Cheley, S., Bayley, H., Gouaux, J.E.: Structure of staphylococcal alpha-hemolysin, a heptameric transmembrane pore, *Science* **274**, 1859–1865 (1996)
196. Lahmani, M., Dupas, C., Houdy, Ph.: *Nanoscience: Nanotechnologies and Nanophysics*, Springer, Berlin Heidelberg New York (2007)
197. Tegenfeldt, J.O., Austin, R.H., et al.: The dynamics of genomic length DNA in 100 nm channel, *Proc. Natl. Acad. Sci. USA* **101** (30), 10979–10983 (2004)
198. Fan, R., Karnik, R., Yue, M., Li, D., Majumdar, A., Yang, P.: DNA translocation in inorganic nanotubes, *Nanoletters* **5** (9), 1633–1637 (2005)
199. Heng, J.B., Ho, C., Kim, T., Timp, R., Aksimentiev, A., Grinkova, Y.V., Sligar, S., Schulten, K., Timp, G.: Sizing DNA using a nanometer-diameter pore, *Biophys. J.* **87**, 2905–2911 (2004)
200. Chen, P., Mitsui, T., Farmer, D.B., Golovchenko, J., Gordon, R.G., Branton, D.: Atomic layer deposition to fine-tune the surface properties and diameters of fabricated nanopores, *Nano Letters* **4** (7), 1333–1337 (2004)

Section Four. Electrophoretic Techniques

201. Viovy, J.-L.: *Rev. Mod. Phys.* **72**, 813–872 (2000)
202. Sudor, J., Barbier, V., Thiriot, S., Godfrin, D., Hourdet, D., Millequant, M., Blanchard, J., Viovy, J.-L.: New block-copolymer thermoassociating matrices for DNA sequencing: Effect of molecular structure on rheology and resolution, *Electrophoresis* **22** (4), 720–728 (2001)
203. Barbier, V., Hervé, M., Sudor, J., Brulet, A., Hourdet, D., Viovy, J.-L.: *Macromolecules* **37**, 5682–5691 (2004)
204. Hjerten, J.: *Chromatogr.* **347**, 191–198 (1985)
205. Barbier, V., Buchholz, B.A., Barron, A.E., Viovy, J.L.: Comb-like copolymers as self-coating, low-viscosity and high-resolution matrices for DNA sequencing, *Electrophoresis* **23** (10), 1441–1449 (2002)
206. Weber, J., Barbier, V., Pages-Berhouet, S., Caux-Moncoutier, V., Stoppa-Lyonnet, D., Viovy, J.-L.: A high-throughput mutation detection method based on heteroduplex analysis using graft copolymer matrixes: Application to brca1 and brca2 analysis, *Anal. Chem.* **76** (16), 4839–4848 (2004)

207. Schwartz, D.C., Cantor, C.R.: *Cell* **37**, 67–75 (1984)
208. Carle, G.F., Olson, M.V.: *Nucleic Acids Res.* **12**, 5647–5666 (1984)
209. Mitnik, L., Heller, C., Prost, J., Viovy, J.-L.: Segregation in DNA solutions induced by electric fields, *Science* **267**, 219–222 (1995)
210. Magnusdottir, S., Isambert, H., Heller, C., Viovy, J.-L.: Electrohydrodynamically induced aggregation during constant and pulsed field capillary electrophoresis of DNA, *Biopolymers* **49** (5), 385–401 (1999)
211. Isambert, H., Prost, J., Ajdari, A., Viovy, J.-L.: Electrohydrodynamic patterns in charged colloidal solutions, *Phys. Rev. Lett.* **78**, 5, 971–974 (1997)
212. Manz, A., Graber, N., Widmer, H.M.: Miniaturised total chemical analysis systems: A novel concept for chemical sensing, *Sens. Actuators B* **1**, 244 (1990)
213. Volkmuth, W.D., Austin, R.H.: DNA electrophoresis in microlithographic arrays, *Nature* **358**, 600–602 (1992)
214. Slater, G.W., Guillouzie, S., Gauthier, M.G., Mercier, J.-F., Kenward, M., McCormick, L.C., Tessier, F.: Theory of DNA electrophoresis [approximately 1999–2002 (1/2)], *Electrophoresis* **23**, 3791–3816 (2002)
215. Volkmuth, W.D., Duke, T., Wu, M.C., Austin, R.H., Szabo, A.: DNA electrodiffusion in a 2D array of posts, *Phys. Rev. Lett.* **72**, 2117–2120 (1994)
216. Minc, N., Viovy, J.-L., Dorfman, K.D.: Non-Markovian transport of DNA in microfluidic post arrays, *Phys. Rev. Lett.* **94**, 198105 (2005)
217. Kaji, N., Tezuka, Y., Takamura, Y., Ueda, M., Nishimoto, T., Nakanishi, H., Horiike, Y., Baba, Y.: Separation of long DNA molecules by quartz nanopillar chips under a direct current electric field, *Anal. Chem.* **76**, 15–22 (2004)
218. Lawrence, E.M., Ivery, M.L., Flores, G.A., Liu, J., Bibette, J., Richard, J.: Field-induced structure of confined ferrofluid emulsion, *J. Intl. J. Mod. Phys. B* **8**, 2765–2777 (1994)
219. Doyle, P., Bibette, J., Bancaud, A., Viovy, J.-L.: Self-assembled magnetic matrices for DNA separation chips, *Science* **295**, 2237 (2002)
220. Minc, N., Futterer, C., Dorfman, K.D., Bancaud, A., Gosse, C., Goubault, C., Viovy, J.-L.: Quantitative microfluidic separation of DNA in self-assembled magnetic matrices, *Anal. Chem.* **76**, 3770–3776 (2004)
221. Chou, C.F., Bakajin, O., Turner, S.W.P., Duke, T., Chan, S.S., Cox, E.C., Craighead, H.G., Darnton, N., Han, J., Austin, R.H.: Sorting by diffusion: An asymmetric obstacle course for continuous molecular separation, *Proc. Natl. Acad. Sci. USA* **96**, 13762–13765 (1999)
222. Han, J., Craighead, H.G.: Separation of long DNA molecules in a microfabricated entropic trap array, *Science* **288**, 1026–1029 (2000)
223. Cabodi, M., Turner, S.W.P., Craighead, H.G.: Entropic recoil separation of long DNA molecules, *Anal. Chem.* **74**, 5169–5174 (2002)
224. Long, D., Harden, J.L., Viovy, J.-L.: Séparation de polyélectrolytes par perméation de membranes poreuses, *C.R. Acad. Sci. Series II* **321**, 239–246 (1995)

Microfluidics: Concepts and Applications to the Life Sciences

A. Buguin, Y. Chen, and P. Silberzan

12.1 Introduction

With the appearance of the micro total analysis system (μ TAS) [2] or the lab-on-a-chip, the way was opened to novel applications of microfabricated devices, particularly for chemical or biological analysis [3, 4]. There are many advantages with this type of approach, and of such importance that this field has moved forward in leaps and bounds over the past few years. Apart from the issues of mass production, integration, cost, and so on, which are treated elsewhere in this book, the possibility of working on minute amounts of sample means that a good many types of analysis can be viewed from a new angle. However, this undisputed breakthrough itself requires the design and fabrication of tools able to manipulate such small amounts of liquid in a chip. This is the challenge that microfluidics hopes to meet.

Comment. To get some idea of the volumes involved here, the volume of a droplet measuring $100\mu\text{m}$ in diameter is of the order of 1 nL . Standard micropipettes manipulate volumes greater than $1\mu\text{L}$.

To regard this as some kind of microplumbing would nevertheless be unfair. The manipulation of species and the motion of fluids are often intimately connected, and at the end of this chapter we shall describe integrated devices in which the sample is brought to the chip, then manipulated so as to obtain the desired effect, e.g., separation, crystallisation, etc. Another feature of this discipline is that it lies at the interface between several traditional areas, viz., physics, chemistry, biology, and engineering, and at the frontier of nanoscience and nanotechnology.

With regard to biological applications, the extreme complexity of phenomena in living beings means that they can only be analysed by more and more sensitive and selective, hence local, methods if quantitative measurements are to be made in vitro and in real time [5]. These microlaboratories on a chip are today able to accomplish some quite sophisticated tasks, including cell sorting and detection of rare cells, extraction and purification of genomic DNA



Fig. 12.1. Anatomical plate by Antonio Scarpa (1752–1832), showing a complex microfluidic system [9]

from a single cell, crystallisation of proteins, etc. [6]. Some devices are already commercialised [7] and, given the extraordinary capacity for integration and miniaturisation in microelectronics, it is reasonable to expect a considerable surge in research and industrial development in the years to come.

As mentioned above, one of the current features of microfluidics consists in circulating a liquid in a channel with bore between a few micrometers and a few tens of micrometers. In this respect, it is not a new discipline as such, since this kind of problem has already been solved, e.g., in inkjet printers, diesel injectors, and capillary electrophoresis [8]. So microfluidics has not revolutionised the ideas and methods of classical hydrodynamics. On the other hand, nature provides us with examples of systems that are much more elaborate than a simple microchannel. This is exemplified by circulation systems in plants and animals (see Fig. 12.1). Apart from guiding the relevant fluids (vessels), such systems integrate the means for setting them in motion (heart or osmotic pressure), for allowing exchanges with other fluids (lungs), and for restricting their flow (valves).

Current microfabrication techniques, mainly developed from methods used in microelectronics, provide ways of imagining systems as complex as these using synthetic materials. Without claiming to reach the level of complexity found in life, technological efforts today aim to offer rudimentary solutions for all these functions, with a view to integrating them in a later stage. Naturally, and it is also an important prospect, fields of application extend towards

combinatorial chemistry. In this area, reducing the volumes of solvent required accord perfectly with the current awareness of environmental issues.

After a brief review of the implications of size reduction for the tenets of fluid mechanics, we shall describe some of the solutions put forward to achieve the rudimentary functions mentioned above, pausing briefly to describe the fabrication methods themselves. Finally, we discuss some illustrative examples of the way microfluidics has been used in the life sciences.

12.2 Physics of Microfluidic Flows

12.2.1 Fluid Mechanics on Microscopic Scales

The aim in this section is to review some basic notions of fluid mechanics through the modifications required when the classic conservation laws are applied to the mechanics of a fluid microparticle (liquid or gas) [10].

Notion of Fluid Particle

The fluid particle is a basic concept in fluid mechanics. The idea is to introduce a length scale over which the system can be treated as continuous. This means that it is well above the size of a molecule, but much less than the dimensions of the fluid container (the channel). In numerical simulations using the finite-element method, it is the mesh size. This particle must be much smaller than the size of the system one hopes to describe. Hence, for a meteorologist, a particle of size less than 10 km would already be very small, whereas for someone interested in microfluidics and working with channels only 10 μm across, a length of 1 μm would doubtless be too big. On very small length scales, e.g., molecular scales, the problem belongs to statistical mechanics. The fluid particle must contain a number N of molecules that is big enough to justify neglecting fluctuations in the associated quantities. In microfluidics, the first problem is therefore to identify whether there exists a suitable intermediate length scale between the size of a molecule and the size of the channel.

Comment. This description in terms of fluid particles, although acceptable in most cases (in particular for liquids), is no longer valid when N becomes too small. This will be the case, for instance, in rarefied gases, or when the fluid contains particles of mesoscopic dimensions, as happens in colloidal suspensions, for example, where the continuum approach is no longer valid [11].

Fundamental Equation of Motion

The fluid particle is an imaginary entity and serves only to define a length scale (and hence a volume element) to which one can apply the fundamental equation of motion at any point of space:

$$\rho \frac{d\mathbf{v}}{dt} = \sum \mathbf{f}_V, \quad (12.1)$$

assuming that the fluid is incompressible, i.e., $\text{div}(\mathbf{v}) = 0$, which expresses the conservation of matter, where $\mathbf{v}$ is the fluid velocity at the point in question, ρ is its density, and $\mathbf{f}_V$ is a force per unit volume applying to the volume element. This means that $\mathbf{v}$ is a velocity field, a function of the time and space coordinates. The derivative operator can thus be rewritten and the relation takes the form

$$\rho \left[\frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v} \cdot \nabla) \mathbf{v} \right] = \sum \mathbf{f}_V. \quad (12.2)$$

Comment. One can of course include inertial forces in the case of accelerating frames, but we shall see later that these can be neglected for very small volume elements [10].

Forces per Unit Volume

When there is no external force $\mathbf{f}_{V \text{ ext}}$ (such as a gravitational, electric, or magnetic field), it is standard practice in fluid mechanics to separate the forces per unit volume (divergence of the generalised stress tensor $\overline{\overline{\sigma}}_g$) into two contributions: pressure forces p and viscous forces ($\overline{\overline{\sigma}}_g = -p\overline{\overline{\mathbf{I}}} + \overline{\overline{\sigma}}$), where $\overline{\overline{\mathbf{I}}}$ is the identity tensor and $\overline{\overline{\sigma}}$ is the stress tensor (viscous contribution). The divergence of the latter is proportional to the shear under the hypothesis of a Newtonian fluid, where it is assumed that the force per unit volume required for the fluid layers of viscosity η to slide over one another is given by

$$\text{div}(\overline{\overline{\sigma}}) = \eta \Delta \mathbf{v}.$$

Equation (12.2) then becomes the Navier–Stokes equation:

$$\rho \left[\frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v} \cdot \nabla) \mathbf{v} \right] = \mathbf{f}_{V \text{ ext}} - \nabla p + \eta \Delta \mathbf{v}. \quad (12.3)$$

Comment. The Newtonian fluid hypothesis simplifies the problem enormously and the study of the relation $\overline{\overline{\sigma}} = f(\nabla \mathbf{v})$ forms a discipline in its own right, for which microfluidics also provides solutions [12].

Dimensionless Numbers

In order to estimate the nonlinear contributions to (12.3), under the steady-state assumption ($\partial \mathbf{v} / \partial t = 0$), when there is no precise specification of the forces per unit volume, one compares the terms in $\rho(\mathbf{v} \cdot \nabla) \mathbf{v}$ and $\eta \Delta \mathbf{v}$. This leads naturally to the introduction of the Reynolds number, one of the well-known dimensionless numbers in this field:

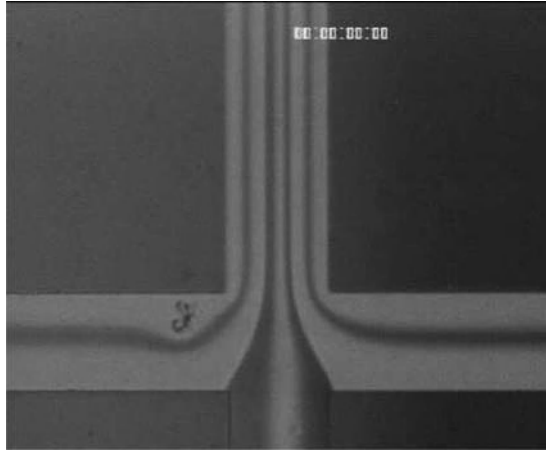


Fig. 12.2. Laminar flow in a microchannel (100 μm), observed by fluorescence microscopy. The channel has 9 inputs fed alternately by fluorescent and colourless liquids

$$\text{Re} = \frac{|\rho(\mathbf{v} \cdot \nabla)\mathbf{v}|}{|\eta\Delta\mathbf{v}|} \approx \frac{\rho UL}{\eta} \approx \frac{UL}{\nu}, \quad (12.4)$$

where L is a characteristic length scale of the system, U the typical flow velocity, and $\nu = \eta/\rho$ the viscous diffusion coefficient (kinematic viscosity) of the fluid.

The Reynolds number decreases with the size of the system. To get an idea of the order of magnitude, in a channel of bore $L \approx 10 \mu\text{m}$, for a characteristic fluid velocity in the cm/s range, we have $\text{Re} \approx 10^{-1}$ for water ($\nu = 10^{-6} \text{ m}^2/\text{s}$). This low value of the Reynolds number shows that, at these length scales, the convective term is practically always negligible compared with the viscous term, whence one obtains the linear form of (12.3). There is no convection (turbulence) in microfluidic systems. Flows are laminar and mixing processes (governed solely by viscous diffusion) are much slower (see Fig. 12.2).

Comment. The Reynolds number can also be viewed as the ratio of a diffusion time and a convection time. In this context, it is analogous to the Péclet numbers associated with diffusion of a particle (Fick's equation UL/D) and heat diffusion (Fourier's equation UL/κ).

Boundary Conditions

Full solution of (12.3) (impossible analytically in most cases) requires specification of the boundary conditions for the system. Apart from the case of rarefied fluids or particular interfaces [13], these conditions are robust and require [10]:

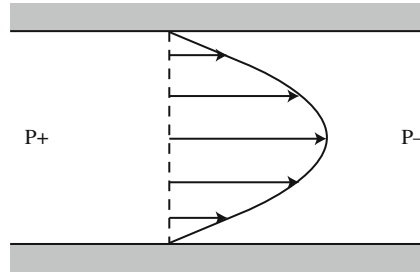


Fig. 12.3. Velocity profile (parabolic) observed at low Reynolds number in a cylindrical channel filled with fluid and subjected to a pressure difference $\Delta P = P_+ - P_-$

- Zero velocity at a solid interface for a viscous fluid (no slipping).
- Continuity of tangential and normal stresses at an interface between two fluids 1 and 2:

$$\sigma_{t1} = \sigma_{t2}, \quad (12.5)$$

$$\sigma_{n1} - \sigma_{n2} = \gamma \left(\frac{1}{R} + \frac{1}{R'} \right), \quad (12.6)$$

where R and R' are the principal radii of curvature of the interface between 1 and 2.

Comment 1. Many experiments show that the zero speed assumption at the solid interface is not justified on the nanoscale [13]. For nanochannels, the slipping phenomenon cannot be neglected.

Comment 2. Equation (12.6), which is none other than the Laplace law, is of particular importance in microfluidics. It is mainly responsible for the filling problems encountered when channels have very small dimensions (trapping of air bubbles).

12.2.2 Setting the Fluid in Motion

Although it is relatively easy to make a microchannel from one of the materials mentioned in the last section, it is quite a different matter to set the fluid, or the particles it contains, in motion. Let us therefore discuss some of the methods available for breathing life into the system, as it were. Several strategies are possible depending on whether one needs to displace the fluid itself, or particles, droplets, or molecules dispersed in this fluid.

Setting in Motion by Pressure Difference: Poiseuille Flow

At low Reynolds number, in the case of a cylindrical channel with axis z , diameter R , and length l , subjected to a pressure difference ΔP between its two openings, equation (12.3) becomes

$$\Delta v_z = -\frac{\Delta P}{\eta l}. \quad (12.7)$$

In this geometry, there is no difficulty in solving this. With the boundary condition $v_z(r = R) = 0$, this implies a flow in which the velocity profile is parabolic, as shown in Fig. 12.3. Such a system is called a Poiseuille flow:

$$v_z(r) = \frac{\Delta P}{4\eta l}(R^2 - r^2), \quad (12.8)$$

where r is the distance to the center of the channel.

Comment 1. Most microfabrication methods produce channels with a rectangular cross-section which makes it more difficult to find the exact solution to (12.3) (for channels with aspect ratio close to 1), but which does not change the general form of the solution.

Comment 2. The velocity varies across the channel, and this can have a significant effect on the resolving power of separation methods [14]. The average velocity over the cross-section of the channel is nevertheless of the order of $\bar{v}_z \approx (\Delta P/\eta l)R^2$.

In order to transport an aqueous solution over distances of around one centimeter, and in a time of about one second, through a channel of diameter $10\text{ }\mu\text{m}$, a pressure of order $\Delta P = 1\text{ bar}$ would be required (corresponding to a water column 10 m high). So a device exploiting hydrostatic pressure would need a macroscopic system, whose dimensions are obviously incompatible with the trend toward miniaturisation, to set these fluids in motion at the relevant speeds. Along the same lines, similar macroscopic systems using a gas as pressure reservoir or a system imposing the flow rate, such as a syringe, can be envisaged. However, this type of solution shows its limits in the need to make an interface between the microscopic device (the microchannel network) and the system used to impose the flow. The experimenter is soon faced with the problem of a highly complex device occupying a very limited area ($\approx 1\text{ cm}^2$), but for which the control system imposing pressures or flow rates occupies an enormous volume (not forgetting the problems due to leakage owing to a large number of connections with the macroscopic world).

This type of system can thus generate a Poiseuille flow within the fluid, and this can be used to transport solutes such as molecules, proteins, or particles in short times. By short times, we understand that the speeds are big enough to be able to neglect solute diffusion, i.e., high Péclet number. However, for Poiseuille flow, the solute will not have the same displacement velocity at different points across the channel [see (12.8)]. This well-known phenomenon (Taylor dispersion), which tends to enhance dispersion, can seriously affect the resolving power of separation methods. Although effectively used in some such methods (field-flow fractionation FFF), this phenomenon disperses suspensions while not being efficient enough to be suitable for mixing problems.

Comment. Apart from electroosmosis, discussed further in the next section, several strategies can be implemented to limit the Taylor dispersion problem. The simplest consists in displacing the solute in a solvent droplet within an immiscible fluid carrier. All diffusion and dispersion is then confined to within the volume of this droplet [40].

Finally, note that there is a formal analogy between fluid mechanics and electrical circuits which, although rather tricky to apply in practice, is worth exploring from the microfluidic standpoint [15].

Setting in Motion by an Electric Field: Electroosmosis

An alternative to inducing liquid displacement by hydrostatic pressure difference is to apply an electric field along the channel. The interaction of this field with the charged walls of the channel sets the fluid in motion by an effect known as electroosmosis [16].

A solid surface in contact with an aqueous solution generally carries an electric charge which depends on the pH conditions. However, this surface charge excess is always balanced by an excess of counterions near the surface, with a concentration profile that depends on the ionic strength of the solvent. The action of an electric field on this accumulation of mobile charges produces (by electrohydrodynamic coupling) a flow called electroosmotic flow.

We begin by calculating this charge distribution. Poisson's equation relates the electric potential ψ to the charge density ρ according to

$$\Delta\psi = -\frac{\rho}{\varepsilon_0\varepsilon_r}, \quad (12.9)$$

where ε_0 is the permittivity of the vacuum ($\approx 8.85 \times 10^{-12} \text{ J}^{-1}\text{C}^2\text{m}^{-1}$), and ε_r is the relative permittivity of the medium (≈ 80 for water).

At equilibrium, the number of ions of each type is given by a Boltzmann distribution, which implies for a symmetric electrolyte:

$$\rho = n_0 \left[-ze \exp\left(\frac{-ze\psi}{k_B T}\right) + ze \exp\left(\frac{ze\psi}{k_B T}\right) \right], \quad (12.10)$$

where n_0 is the number of ions far from the surface, z the valence of these ions, e the elementary charge, T the temperature, and k_B the Boltzmann constant.

The Debye–Hückel approximation considers the limiting case where $e\psi \ll k_B T$. Combining (12.9) and (12.10), one then obtains

$$\psi = \psi_0 \exp(-\kappa x), \quad (12.11)$$

where

$$\kappa^2 = 2 \frac{z^2 e^2 n_0}{\varepsilon_0 \varepsilon_r k_B T}, \quad (12.12)$$

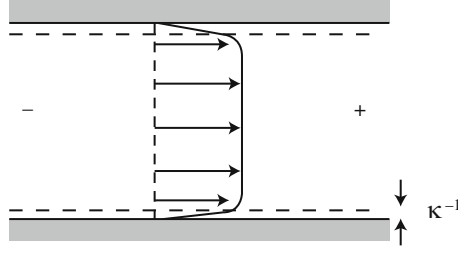


Fig. 12.4. Velocity profile for electroosmotic flow (with a positively charged surface). The fluid velocity is constant across the section of the channel (outside the Debye layer, which has been grossly exaggerated in the diagram). It is instructive to compare this profile with the Poiseuille profile in Fig. 12.3

with κ^{-1} the Debye length, which describes the range of electrostatic interactions within the medium. For a monovalent ion at a concentration of 0.1 M, $\kappa^{-1} \sim 1$ nm. If the concentration drops to 10^{-7} M, as it does in pure water, then $\kappa^{-1} \sim 1$ μ m.

Applying an electric field parallel to this surface sets the counterions in motion over a thickness κ^{-1} and the velocity profile $v(x)$ can be deduced from the Navier–Stokes equation:

$$\rho E + \eta \frac{\partial^2 v}{\partial x^2} = 0. \quad (12.13)$$

Combining this equation with (12.9), we obtain

$$E \varepsilon_0 \varepsilon_r \frac{\partial^2 \psi}{\partial x^2} = \eta \frac{\partial^2 v}{\partial x^2}. \quad (12.14)$$

Equation (12.14) is solved with the usual boundary conditions: the velocity is zero on the surface, and the potential on the surface is defined as the ζ potential. Far from the surface,

$$\frac{\partial \psi}{\partial x} = \frac{\partial v}{\partial x} = 0,$$

which implies that

$$v(x) = \frac{\varepsilon_0 \varepsilon_r E}{\eta} [\psi(x) - \zeta]. \quad (12.15)$$

If the Debye layer of thickness κ^{-1} is neglected compared with the channel diameter, this leads to a profile in which the velocity is almost constant across the whole section of the channel, with value V_{eo} given by

$$V_{eo} = -\frac{\varepsilon_0 \varepsilon_r E \zeta}{\eta}. \quad (12.16)$$

Comment 1. Electroosmosis is the direct counterpart of particle electrophoresis. Indeed, a charged particle moves under the effect of an electric field by the same mechanism as presented above to describe the fluid motion induced by immobile charged surfaces.

Comment 2. Electroosmosis is an extremely effective way of setting fluids in motion. Moreover, because it generates a flow with almost uniform velocity profile in the central part of the channel, this method considerably reduces the problems of Taylor dispersion so detrimental to separation methods.

Comment 3. A strong current enforced in the channel will not only have a transport effect on mobile charges. It will also cause Joule heating, which can become significant. Microsystems, which dissipate heat more efficiently owing to their dimensions, represent a step forward compared with existing techniques.

Alternative Solutions

It is still a delicate matter today to address a large number of channels by pressure. It is much easier to address voltages or electric currents by a judicious arrangement of electrodes, and this is something that the technology developed in microelectronics can handle in a routine manner. The idea of implanting a large number of small heating resistors or electrodes with their addressing system on the substrate of a network of channels would therefore appear to offer an interesting solution [17]. Some research groups have thus examined various alternatives offered by physicochemistry to set fluids in motion. Often the approach chosen involves manipulating droplets placed on a substrate and using capillarity properties near the triple line to displace them. As for two-phase flows, the problems of diffusion and Taylor dispersion can be eliminated by confining the solute within this microdroplet.

The first possibility is to use highly localised heating (and hence a temperature gradient) applied at the edge of a droplet. This phenomenon uses the temperature-dependence of interface tension, known as the thermal Marangoni effect. Applying a thermal gradient breaks the equilibrium of the capillary forces on the triple line contour and thus sets the drop in motion [18, 19]. However, this method has two major drawbacks. Firstly, a temperature gradient has to be applied precisely where it is sometimes necessary (especially in biology) to maintain a constant temperature. Secondly, even if only very small thermal gradients are required theoretically to set a drop in motion, a hysteresis phenomenon affects the contact angle, and this evolves in time due to various sources of pollution [20], making the method extremely difficult to implement in practice.

It is also possible to use electric fields to set the drops in motion. An electric field applied to a water drop placed on a solid substrate will reduce the angle of contact. This is known as electrowetting [21]. By a simple addressing

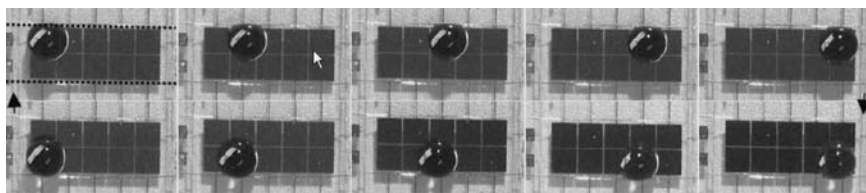


Fig. 12.5. Transport of a water droplet by successive addressing of underlying electrodes (*squares* visible under the droplet). The counterelectrode, not visible in the photo, is a fine gold wire used as a catenary. Each electrode has side $800\text{ }\mu\text{m}$. From [25]

system using electrodes, one can thus vary the wettability of the substrate locally and, by creating wetting gradients between the front and the back of the drop, set it in motion, split it into two parts, or force two drops to mix [22,23]. A series of electrodes suitably placed and addressed will therefore be able to impose an easily controllable motion on these drops. Other strategies have used surface corrugation, which introduces polarity and allows one to displace drops macroscopically by a succession of spreading and retraction effects [24].

Comment. It is in the vicinity of the triple line that the electric field is strongest by the spike effect. These very strong fields tend to alter the surface of the insulating film, which often leads to irreversible hysteresis effects (ageing).

However, the integration of devices using electrowetting remains a delicate matter and requires more work to make its use more flexible.

To end this section, it is worth mentioning some less widely used technologies. Etched chips can be used on a compact disk support, generating flows by the centrifugal force induced when the disk rotates [26]. Other work shows that ultrasonic surface waves, generated at the surface of a piezoelectric crystal, can be effectively used to manipulate liquid drops [27].

These examples, which illustrate the extraordinary level of activity in this field, also show that no solution yet stands out for its superiority.

12.3 Fabrication, Materials, Functions

Several lines of technology have been developed over the past few years. One point specific to microfluidics is still the fabrication of closed channels, i.e., after etching, the device has to be closed by sticking on a top plate. This fact requires specific developments, such as anodic bonding, to assemble the fourth side of the etched structure.

Silicon is the material allowing the best definition by direct printing. Silicon technology, a direct spinoff from the microelectronics industry, has naturally been the source of the most significant technological developments. Glass

can lead to considerable economic advantages. It is often chosen for the qualities of the material. Indeed, glass is a very good electrical insulator, as well as being chemically inert and optically transparent. The third possibility is to use polymers for the advantages of cost and mass production. This technology derives essentially from miniaturisation of techniques in the plastics industry (injection moulding, thermoforming, etc.). In research centers, soft lithography, in particular using the elastomer polydimethylsiloxane (PDMS), holds an important place, since this material is extremely simple to implement and provides a fast way of making prototypes. Apart from the obvious problems of chemical compatibility with the various solvents that may be used in different applications, polymers have properties such as permeability to gases or liquids which can be problematic, but which can also be used to advantage.

Before going into the details of the different types of fabrication, we should mention other techniques such as stereolithography, laser ablation, etc., and bear in mind that hybrid methods, combining two or more fabrication techniques, are commonplace today.

Note that the fabrication of a microfluidic device may involve technical difficulties quite different from those encountered in microelectronics. Device miniaturisation is not simply a question of microfabrication. The problem is to build new concepts and find novel solutions by implementing a range of knowhow: materials science, microfabrication techniques, hydrodynamics, micromechanics, microelectronics, optics, electromagnetism, analytical chemistry, biochemistry, and so on. From this point of view, microfluidics is intrinsically cross-disciplinary, and in this context, the integration of different functionalities on the same chip (which remains the principal aim) is a genuine technological challenge. Finally, since this field is still in its early stages, and since the field of applications is so broad, it is hard to define the basic building blocks that will form the backbone of this discipline, as one can now for microelectronics, for example. The very wide range of applications, needs, and tools to be implemented suggest a slow rate of development in which it is unlikely that a single technological channel will be able to satisfy the many applications envisaged.

12.3.1 Lithography

The first stage, common to all the relevant forms of technology, is to transfer a pattern onto a substrate. This transfer is achieved via a thin polymer film. One usually proceeds by means of proven techniques from photolithography [28]. The first thing is to deposit a thin film of photosensitive resist (photoresist) on a substrate, and then to expose to ultraviolet radiation through an optical mask on which the pattern for the required microfluidic circuit has been drawn. After exposing the resist layer, the pattern is developed in a chemical solution, whence the motifs on the mask are reproduced in negative or positive form depending on the type of resist. This resist layer then serves in its turn as a protective mask when the substrate is wet-etched by chemical

etching or dry-etched by means of reactive ions (reactive ion etching RIE). It then remains only to clean the residual resist from the wafer.

A recent technique for reproducing surface patterns with very high resolution is lithography by nanoimprinting [22, 23]. This technique, proposed as a substitute for conventional photolithography, has the advantage of being able to replicate the motifs on a mould in a resist which then plays the role of mask during the RIE dry-etching of the substrate. Replication is so faithful that even the smallest motifs defined by electron beam lithography can be reproduced. There are two stages in this technique. The first is to print the pattern in a polymer film deposited on a substrate using a patterned mould obtained by conventional techniques (electron beam lithography followed by RIE). Once the polymer film has been structured, the substrate itself is then etched by RIE.

To facilitate separation after moulding, the mould surface must be treated with an anti-adhesive. Note that, although the fabrication of the mould requires the use of a costly conventional high resolution lithographic technique, this mould can then be reused many times for successive replication. Thermoplastic polymers such as polymethylmethacrylate (PMMA) or polycarbonate (PC) are generally used for moulding. Imprinting is achieved at a temperature above the glass transition temperature T_g of the polymer and at a high enough pressure (a few tens of bars for PMMA) for a period of a few minutes. The system is then cooled while maintaining the pressure. When the temperature goes below T_g , the pressure is released and the mould separated from the sample. The relief imprinted in the polymer layer is then transferred to the substrate by RIE.

Nanoimprinting has a potentially very high efficiency since it is a parallel process. However, owing to the heating and cooling stages, the total time of imprinting can grow too long for certain more delicate applications. This technique also raises problems of high-accuracy alignment. For these reasons, an alternative method has been proposed, which consists in first moulding a liquid mixture of a monomer with a catalyst, then solidifying it by UV irradiation through a transparent mould. The advantage with this technique lies in the fact that it can be done at room temperature and low pressure, combined with the fact that it is faster and can achieve high accuracy alignment [33].

12.3.2 Different Technologies

Silicon

This technology was pursued in parallel with the development of microsystem production for other types of applications such as microactuators, accelerometers, microinjection heads, and so on [29–31]. In this case, after etching the material by RIE as discussed above, the top plate is sealed onto the microstructures with a pyrex film using anodic bonding. One thus obtains a ready-to-use microfluidic device. Other micromachining techniques such as

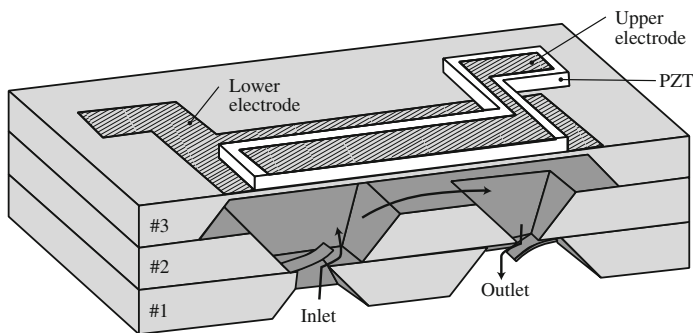


Fig. 12.6. Membrane micropump. Two pump chambers are integrated into this device, arranged in opposite directions, each with two parts separated by a valve. When the membrane, equipped with a piezoelectric actuator, is deformed, the pressure of the upper part of the two chambers varies in such a way that one of the two gates closes and the other opens alternately. A flow is thus set up from the left chamber to the right chamber. From [32]

laser ablation and electroerosion can also be employed to make relatively large microstructures.

In order to make devices with more sophisticated functionalities, one seeks to integrate activation and detection elements, capable of administering and analysing complex flows in situ, on the same chip. Figure 12.6 illustrates this possibility on an elementary component to be integrated. This is a membrane micropump comprising two microvalves assembled in opposite directions and a piezoelectric activation membrane. More generally, actuation can be achieved by electrostatic, electromagnetic, thermopneumatic, or other forces [31]. Fabrication of this type of element is thus a complex operation involving a whole series of delicate operations. Although there is no fundamental problem in integrating several types of element on the same chip, implementation can in practice be much more difficult, whence it will only be envisaged for key applications.

Having said this, silicon micromachining technology remains an interesting choice of strategy in the long term, since it can integrate all types of micro-electronic, microoptical, micromechanical, and microfluidic silicon component on the same chip.

Glass

Glass, or rather, glasses, since the details of the chemical composition differ so much from one sample to another, is traditionally the choice material for most experiments in chemistry and biology. Microstructures are conventionally made in glass by chemical etching with hydrofluoric acid (HF) after the photolithography stage. Once it has been micromachined and connection holes made in it, the microstructured glass plate is finally closed by means

of another glass plate by thermal bonding or chemical bonding with dilute hydrofluoric acid. Other methods such as laser ablation and RIE can also be used. With RIE, specific machining systems must be used. Optical or electron lithography can also be used directly to expose microstructures locally on particular photosensitive glasses. After annealing, the exposed part evolves from an amorphous to a polycrystalline state, much more soluble in the acid used for development.

Plastic

Injection moulding, in which the polymer melt is injected into the shape to be reproduced, is the most widespread method for forming plastic objects. In order to make the micro/nanostructures of an all-plastic microfluidic device, it is possible, and simpler, to mould a plastic sheet directly on a previously machined mould [33] using the techniques described above. This mould can be made from silicon or some other material, e.g., a metal. One of the advantages of moulding techniques is that deep microchannels and shallow nanostructures can be made in a single step. Several technical solutions have been devised:

- Direct imprinting, which means pressing the mould onto the surface of a plastic plate, e.g., PMMA, at a temperature slightly below the glass transition temperature T_g , but at a relatively high pressure. This is known as hot embossing.
- Thermal imprinting in a thin layer of thermoplastic polymer deposited on a rigid substrate. Here, the temperature is well above T_g so that the polymer melts, and a pressure is applied to the whole mould–polymer–substrate system. Once the temperature has gone down, the now solid polymer is easily detached from the mould. This is thermal nanoimprinting.
- UV-assisted imprinting, carried out at room temperature using a transparent mould and a liquid mixture of prepolymer and a crosslinking agent sensitive to ultraviolet radiation. Here the mould fills by capillary forces without the need to apply pressure. After exposure, the now solid polymer is easily detached from the mould. This is known as cold nanoimprinting.
- The compression of thermoplastic polymer granules at a temperature above T_g (hence in the liquid state) to form, after cooling, a single plastic wafer. By directly compressing polymer granules such as PMMA, PC, or cyclic olefin copolymers (COC) between two silicon wafers at high temperature, one of the wafers carrying etched micro- and nanostructures, all the surface structures of the mould can be reproduced with great accuracy. This is called compression moulding.

The bonding together of two plastic plates with nanopatterned surfaces is a critical stage. The seal must be perfect without blocking or deforming the nanostructures. It has been shown that molecular thermal bonding, where two plastic plates are set in contact at a temperature close to the glass transition temperature of the polymer and under a slight pressure, is a reliable and

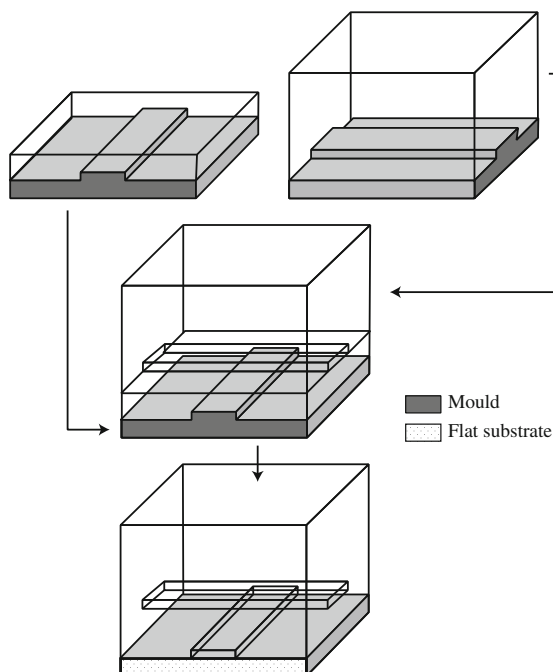


Fig. 12.7. Fabrication of a microvalve by two-layer soft lithography. To begin with, two channels are made in PDMS, a control channel and a working channel. After separation, the two channels are superposed on a substrate and an excess pressure is applied in the control channel. The membrane separating the two channels is deformed and the working channel closes. From [35]

efficient technique from all points of view [33].

Comment. For some applications, the choice of material is also dictated by biocompatibility. Neither silicon, nor glass, nor the majority of plastics are a priori biocompatible. To use them in practice in a biological medium, they must receive appropriate surface treatment, the subject of intense development over the past few years.

12.3.3 Silicone Elastomers

Polydimethylsiloxane (PDMS) is a transparent silicone elastomer (with low intrinsic fluorescence), which is biocompatible and permeable to gases. These properties make it very attractive for many applications.

The structuring method used is soft lithography, which is very easy to implement [34]. The technique is to reproduce a negative of the pattern on a mould obtained by other microfabrication techniques. One begins with a silicon or glass substrate prepared using the methods described above. One

can also start directly with structures microfabricated in a thick resist by photolithography and thereby remove the need for the etch stages. The method is thus to mix the precursor (non-crosslinked, hence liquid PDMS) and the crosslinking agent (also containing the catalyst), then pour this mixture onto the mould. After several hours of crosslinking at 60–80°C, the PDMS has become a flexible elastomer which faithfully replicates the mould geometry. It is then easy to extract it from the mould. To facilitate this step, the mould surface must be modified by a preliminary anti-adhesive treatment. Finally, access is provided for injecting liquids into the channels by piercing the rubber using a needle, and the device is then sealed against a glass plate to minimise leakage. Because it is so easy to implement, soft lithography provides an interesting solution, accessible to all those who wish to carry out microfluidics with a minimum of investment.

This method can also be used to build more complex microfluidic circuits on several levels. For example, two perpendicular channels can be superposed, separated by a thin membrane of the same elastomer (see Fig. 12.7). When one of the channels is under pressure (e.g., using compressed air), the membrane is deformed and thus closes the adjacent channel (pneumatic valve). By placing three of these valves across the same main channel and activating them in sequence, one can alternately generate a pressure difference across each valve. This forces the liquid to move in a controlled way (peristaltic micropump) [35]. Starting from this basic architecture, quite sophisticated microfluidic devices can be produced. For this reason, soft lithography is considered as a high-potential technology, not only for academic research, but also for the development of industrial prototypes.

Comment. PDMS is a hydrophobic rubber, and this may lead to problems, e.g., when filling the channels. However, its surface is easily modified chemically by a low power oxygen plasma. With this treatment, the surface of the material becomes not only hydrophilic, but also sufficiently reactive chemically to be used as a substrate for grafting molecular monolayers. It thus becomes possible to apply treatments minimising non-specific adsorption on structures or walls, or alternatively to graft suitable molecules for some chosen application.

In conclusion, several fabrication methods can be used to make microfluidic devices. The choice of method depends on the particular strategy of the user. A combination of these methods is often needed when fabrication involves integration of elements with different functionalities. For the production of disposable devices, plastic represents the best choice (see Fig. 12.8). In research, PDMS and multilayer soft lithography provide a flexible and efficient approach for many applications. Finally, silicon and glass constitute a line of research and development for high-performance fabrication and complex integration.

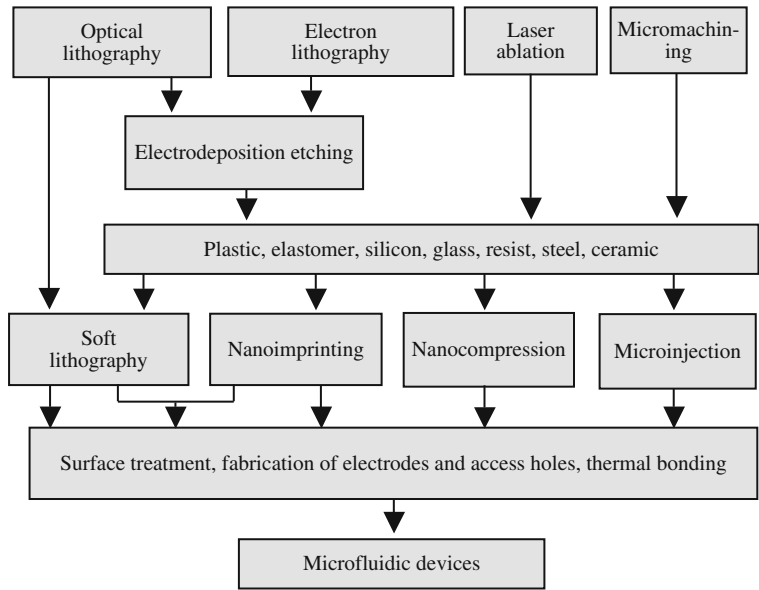


Fig. 12.8. Fabrication of plastic microfluidic devices for mass production

**12.3.4 Elementary Components:
Pumping, Mixing, and Separating in Microvolumes**

In the above, we discussed the two main ways of setting a fluid in motion in a microchannel: application of a pressure difference and the electroosmosis effect. In practice, local acceleration of a liquid can also be controlled by microvalves and membrane or peristaltic micropumps. Moreover, the flow can also be activated and controlled by capillary forces, surface acoustic waves, and so on. Finally, another interesting method is to use a laser beam, either to move a cluster of microbeads around a microfluidic channel using optical tweezers, or to change the state of the polymer placed inside the channel [36].

One of the key tasks in microfluidics is mixing. We have seen that one of the characteristics of microfluidic flows is that they are laminar. In a passive microchannel, the mixing of two liquids cannot occur by diffusion. Even for systems measuring only a few tens of micrometers, the process is too slow to be effective in practice. However, by etching microstructures on the channel walls, one can effectively change the paths of the two liquids and thereby force them to mix by alternating very thin layers of the two fluids. On these very small length scales, diffusion becomes a fast and efficient mixing mechanism [37]. Another mixing method involves injecting the two liquids alternately into a microchannel by means of peristaltic micropumps or by exploiting the electrokinetic effect.

Comment. The two-phase flows discussed above allow fast mixing within the drops due to the flows generated inside them when they move along the channel. The effect is amplified if the drops are forced into bends or baffles. This has interesting consequences: since mixing occurs quickly in each drop, they can be considered as independent reactors.

Another important point concerns temperature control in these integrated devices. Microelectrodes are often used to heat some given point in the device, whence it is possible to adjust the chemical or biological reactivity of the various species present. When rapid response times are required, it is advantageous to use a focused laser pulse for local heating.

Other basic functions, such as concentration, filtering, extraction, and purification of chemical or biological substances, have also been given a great deal of attention over the last few years [38]. Results are often published in conference proceedings (in particular μ TAS) and journals like *Analytical Chemistry*, *Lab on a Chip*, *Electrophoresis*, and so on.

Regarding detection, conventional measurement and analysis methods apply to microfluidics provided that one takes the length scale into account. On the microscale, the most sensitive methods are often those of electrochemistry and optics. A particularly sensitive method is the detection of biological species by surface plasmon resonance. Surface plasmons are excited by an evanescent wave in a thin metal film, usually gold. For a given metal, the resonance is observed with polarised light at a well-defined angle of incidence. This angle changes significantly depending on the close environment of the surface, whence adsorbed chemical or biological species can be detected even in very small amounts. Combining this detection method with molecular anchoring techniques, the walls of microfluidic elements can be functionalised and this type of measurement can be integrated directly into the device.

12.4 Applications

12.4.1 Crystallisation of Proteins

Protein functions are closely related to their structure. X-ray or electron diffraction are the main methods used to determine protein structures, and these require crystals of a reasonable size and quality, with few impurities or grain boundaries. This means that the crystallisation of purified proteins is often one of the limiting stages. It is also a stage that requires screening of many operating conditions and which can therefore benefit directly from the parallel approach offered by microfluidics. We begin with a brief review of the general principles governing crystallisation processes, then outline some of the solutions and advantages offered by microfluidics to achieve these aims, illustrating with three concrete examples.

To obtain a crystal from a solution of soluble (or solubilised) proteins, the physicochemical principle exploited is to add crystallisation factors such as organic solvents, salts, polymers, and so on. The solution is then enriched by vapour diffusion, dialysis, etc., until it exceeds the crystallisation threshold. Transmembrane proteins, which are usually insoluble, pose more problems, because they can only be purified in very small quantities.

Crystallisation agents are used to inhibit or enhance the various interactions (electrostatic, hydrogen bond, van der Waals) which exist naturally between proteins and water. Although it is possible to obtain general constitutive laws, the relative proportions of some agent or other required to obtain the best crystal depends on the protein to be crystallised in a way that is very difficult to predict. One is soon faced with an optimisation problem (to find the best possible crystal) in a phase diagram with several parameters. The solution to this problem is therefore rather time-consuming (scanning over different combinations of the parameters), and the results are often limited by the difficulty in obtaining large amounts of the purified protein.

At the present time, the most effective solution for dealing with this type of problem is to use automated platforms to carry out mixtures. However, such installations tend to be bulky and costly, and require a lot of maintenance. They are able to manipulate volumes of a few tens of nL. This is therefore a propitious area for microfluidics, because it offers the prospect of manipulating smaller volumes, and hence smaller amounts of matter, in a highly parallel way, at low cost, and in a very limited space.

The reduction in the volume of proteins required to test a single point on the phase diagram (corresponding to a mixture) is of course a step forward since it means that more parameter values can be explored. However, if the volume is reduced too far, there are some problems for this type of application:

- The increased surface to volume ratio. If the proteins or crystallisation factors adsorb onto the walls, it becomes difficult to control concentrations accurately.
- The small size of the resulting crystals makes them difficult to manipulate and observe.

The solutions offered today by microfluidics use characteristic volumes in the nanoliter range.

The main advantage actually derives from the fact that microfluidics can juxtapose a large number of mixing chambers on a reduced surface area (several hundred per cm^2). But then, and this is where the difficulty really lies, one must be able to adjust the concentrations of the various solutes extremely accurately in each of these chambers.

To transport the solutes, the main difficulty comes from diffusion and Taylor dispersion (see p. 748). The concentration is not uniform throughout the vector fluid and evolves with the distance travelled. It would thus seem difficult, using a simple Poiseuille flow, to impose precise concentration conditions within each chamber.

To this one must add the problems involved in mixing the solutions. We have only discussed the thermodynamic aspect (phase diagram) of crystal formation, but of course, one must also take into accounts the kinetics of this formation. If crystallisation is fast (reaction time faster than the diffusion time of the solutes in the mixing chamber), one must find some way of accelerating the mixing of the solutions.

Finally, the physicochemical properties of the different solutions, such as viscosity, surface tension, ionic strength, etc., will vary with the type and concentration of the crystallisation agent. Depending on the technique used to displace the liquids, the speed may depend on one of these parameters and each solution will have its own displacement speed for identical external conditions.

There are several ways of getting round these problems:

- Confine the solutions in microdroplets and transport them by means of an immiscible vector fluid.
- Trap the fluid using microvalves.

The soft lithography technique described above lends itself well to this type of fabrication and, apart from a few special cases [39], almost all the solutions put forward use this method.

Series Approach

The idea here is to fabricate series of droplets of different composition [40,41] (each represents one point on the phase diagram) and to transport them in an immiscible liquid (an oil). This type of transport was described earlier in this chapter. To produce these drops (with volume of the order of 10 nL), an injection nozzle with three inlets is made (although the number of inlets can vary): one for the protein solution, one for the medium, and one for the crystallisation agent. Depending on the relative pressures of each of these inlets, drops of variable composition are generated in the nozzle and carried along by an oil. By directing this series of drops into a glass capillary tube, it is then possible to analyse their contents by an X-ray diffraction image, which establishes the quality of the resulting crystals as a function of their composition. Used under these conditions, this method is somewhat brutal since the crystallisation conditions are imposed as soon as the droplets form. However, the phase diagram can be explored under gentler conditions by using an oil that is permeable to water. It is enough to simply juxtapose two drops with different concentrations and the less concentrated drop will empty its contents into its neighbour by simple vapour diffusion, until crystallisation conditions are achieved. This approach is not restricted to this field and can also be used in combinatorial chemistry (for which other fabrication procedures must be used).

Permeability and Soft Lithography

Microfabrication with silicone elastomers soon became widespread owing to its low cost and ease of implementation [42]. This type of material nevertheless possesses certain particular properties: it is permeable to gases and water vapour. The diffusion properties of gases in silicone are put to use in some separation techniques (pervaporation). These properties, often undesirable in other types of application, can be exploited (by adapting the channel geometry and external conditions) to enrich the solute concentration of some part of the fluidic device and thereby achieve conditions under which crystallisation will take place in a gentle and steady manner.

Parallel Approach

This approach is by far the most promising, but it is also the most difficult to implement [43]. It requires a system of valves and pumps to be integrated into the device. The first solution put forward consisted of an array of 144 reaction chambers (25 nL for each mixture), themselves supplied by 48 manually filled pits. In this example, the liquids mix by diffusion when a peristaltic valve is opened. This system has the advantage that the best crystallisation conditions can be read directly and quickly (<1 h).

12.4.2 Separation of DNA Molecules

The chapters in this book devoted to electrophoresis (see in particular Chap. 11) describe what is involved and stress the difficulties inherent in this approach, which remains the reference for DNA size separation. The first attempts using microfluidic systems consisted in a straightforward transposition of capillary electrophoresis to microchannels. Rather than an improvement in performance, which would in any case be limited by the physics of polyelectrolyte electrophoresis, the aim here was to investigate the practical possibilities of the miniaturisation strategy (reduction of sample volumes, increased analysis rate, improved portability, etc.). Another effect of scale reduction is improved heat dissipation, which means that stronger electric fields can be used. For this reason, significant results were reported, using strategies that consisted mainly in miniaturising and integrating well-understood technologies on the same chip. A good example is the work by Burns et al. [19], in which the different stages of reagent transport, PCR, and analysis are combined in the same chip.

Other techniques, using electric fields or otherwise, have also been tested, and we shall review some of them in this short section. The choice of material depends on the technique adopted. When electric fields displace the particles to be separated, even poorly conducting materials like silicon cannot be used. Then systems are made from insulators such as glass or plastics, which is not without creating its own difficulties, in particular with regard to the control

of electroosmosis. When the particles are displaced by a pressure difference, the material must be able to withstand high pressures, and plastics and other elastomers can then be problematic.

Finally, one should note the intrinsic limits of using high fields in electrophoresis. Not only will increasing the field only marginally improve resolution, beyond a certain value, but high fields bring with them instabilities that can cause DNA to form aggregates. The latter then separate out physically from the rest of the solution [44].

Artificial Gels

Another, more original approach is to get rid of the gel completely and reproduce the sieving effect of the pores by means of arrays of physical obstacles. As early as 1992, the pioneering work by R. Austin and coworkers in the USA showed that the possibilities of microfabrication could be fully exploited by replacing the gels conventionally used in electrophoresis by arrays of microfabricated posts [45]. The lithographic techniques discussed above can indeed be used to fabricate tiny posts whose diameter and spacing are in the micrometer range or smaller, i.e., of the same order of magnitude as the radius of gyration of the molecules themselves. These experiments, combined with the possibility of visualising single DNA molecules, have led to rapid progress in understanding the mechanisms involved in the migration of these long molecules in complex geometries possessing some of the features of a real gel.

To push the analogy a step further, pulsed-field electrophoresis (PFC) has been reported in the literature. Once again, the idea is straightforward enough in principle, since it simply involves transposing gel PCF, which is the standard method for separating long DNA molecules, but which remains rather slow, even though it offers unequalled performance in terms of resolution. The molecules are carried into a hexagonal array of posts by an electric field and subjected to a series of fields whose orientation varies in time. This technique exploits the transient regimes. At a given instant, the molecules stretch into the array of posts in the direction of the field. But when the orientation changes, the molecules need a certain time to line up with the new direction. The time required depends on the size of the molecule, and this provides a way of sorting the molecules with respect to this parameter. Here, not only does microfabrication reproduce the conditions in gels, but by creating an extremely well-defined array, it can considerably improve the efficiency of the device. Some versions of this device can separate continuously in extremely short time spans (see Fig. 12.9) [46, 47]. Finally, the first attempts have already been made to adapt this type of device to nanometric scales, and the results are extremely promising [48].

Nanostructures

Chapter 11 describes experiments in which DNA molecules are made to pass through a protein pore. Along the same lines, it has been shown that sealed

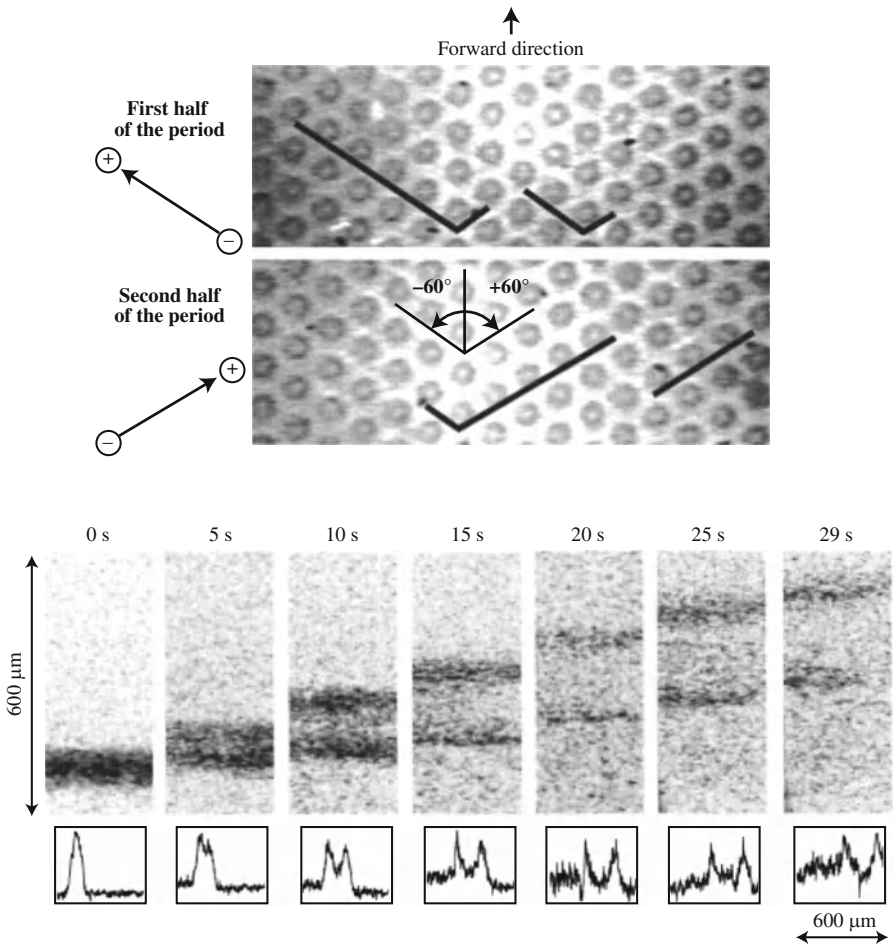


Fig. 12.9. Schematic view of the evolution of two DNA chains in a hexagonal microstructure. The posts are $2\mu\text{m}$ in diameter. The angle between the two successive directions of the field is 120° . The shortest molecules are on average faster, because they have the time to reorient themselves at each field change. *Bottom:* A very fast separation is obtained between λ DNA molecules ($\sim 50\text{ kbp}$) and T4 DNA molecules ($\sim 170\text{ kbp}$), using a pulsed field with period 1 s. From [47]

nanochannels can be fabricated and DNA molecules forced into them by applying an electric field. Today, channels can be fabricated with diameter close to the persistence length of the molecules ($\sim 50\text{ nm}$), which means that they extend to a length close to their end-to-end length. The favoured detection mode is fluorescence videomicroscopy. The length of the chains is measured by taking images of individual molecules, whence their mass can also be ascertained. An electric field can also be used to draw the chains one by one

in front of a micro-, or better, a nano-aperture. The transit time then gives the chain length [49]. If proteins are bound on the chain, they can then be precisely located.

When the chains enter these nanostructures, their entropy is expected to change significantly due to a reduction in the number of available configurations. It has been proposed to exploit this effect by subjecting the molecules to a sequence of constrictions and expansions, once again using electric fields. Indeed, it turns out that the molecules respond differently to these obstacles depending on their size, and such devices provide efficient separations, although not yet fully understood theoretically [50]. It would appear that the effects of strong interactions with the surfaces are superposed on the initially predicted entropic effects, and this underscores a common feature of devices operating at the nanoscale, viz., surface/volume ratios become large, whence a tight control over the physicochemical characteristics of these surfaces is crucial.

Microdielectrophoresis

We have seen how continuous electric fields can be used to achieve separations in electrophoresis. The gradients of these electric fields also play an important role. Indeed, when the field is not spatially homogeneous, a particle with a different complex permittivity to the surrounding medium feels a force $\mathbf{f}$ given by

$$\mathbf{f} \propto a^3 \alpha \nabla |E|^2, \quad (12.17)$$

where a is the radius of the particle and α its polarisability. This is called the dielectrophoretic effect. The sign of this force generally depends on the frequency. When it tends to attract particles toward high-field regions, one speaks of positive dielectrophoresis (positive DEP), and in the other case, negative dielectrophoresis (negative DEP).

Two features distinguish DEP from electrophoresis:

- It is the gradient of the square of the electric field magnitude that arises in the force expression. This effect can thus be observed just as well with a continuous field (although it is then superposed on the electrophoresis effect, which can be problematic) as with an alternating field.
- As a consequence of the first point, by using alternating fields, one can work close to the electrodes and produce gradients directly in the medium. Indeed, continuous fields would lead to electrochemical phenomena, the most obvious of which would be electrolysis.

This effect, which has been known for a long time, has been exploited in a new way by integrating it into microsystems where heat exchange with the environment is optimised and thermal effects minimised. In fact DEP is now commonly used to manipulate and/or characterise particles, from DNA molecules to cells [51]. However, since the effect depends on a gradient, in

order to displace particles over macroscopic distances, one must either set up a large-scale gradient, or combine this phenomenon with a macroscopic flow. Another solution is to generate successive microgradients and move the particles along step by step.

Such systems have been proposed and tested on long DNA molecules or model systems such as latex beads, and it has been shown that this series of elementary displacements does indeed lead to macroscopic displacements which depend heavily on the particle characteristics, and in particular their size or mobility [52].

Continuous Separations with Fixed Field

We saw earlier that some microfabricated devices can be used to reproduce the conditions of pulsed-field electrophoresis, making it possible to envisage continuous separations. Other devices have been proposed, using a single field direction in the same spirit as for DEP, where a total displacement is achieved as a series of individual steps. The first of these systems uses left–right symmetry-breaking obtained by forcing the molecules onto tilted physical obstacles (rectangular posts). As predicted by theory, the big molecules are deflected with a greater angle than the small ones. But, apart from certain limitations due to the physics of the problem, the values obtained for the angles are very small and analysis times long [53]. However, the efficiency of this device can be increased by approaching from a nonzero angle. It is even possible to get round the details of the post inclinations by approaching from a suitably chosen angle. Indeed, there are two principal axes in this problem: one is the axis of the macroscopic field that injects the molecules (hydrodynamic flow, or most often, an electric field), and the other is one of the crystallographic axes of the post array. When these two axes are close, the molecules may prefer the easier crystallographic axis to the field axis. Since this effect is size dependent, it does indeed provide a means of sorting the particles. This kind of sorting has been very clearly demonstrated on latex beads by imposing an array of optical traps [54] or microposts [55].

12.4.3 Cell Sorting

The standard approach in cell biology is to work on a large number of cells and obtain the relevant parameters from a global measurement, averaged over a large number of cells. In some cases, this approach proves to be inadequate. It is then necessary to carry out measurements on single cells, which better reflect the intrinsic diversity and variability of living entities. This is one of the motivations of cell sorting: to pick out each cell individually and analyse it, e.g., from a genetic standpoint. However, practical usage gives other meanings to the term. In particular, it may be necessary to separate distinct cell populations in terms of clear differences in physicochemical characteristics, e.g., adhesion on functionalised surfaces, elasticity, etc., and in as efficient a

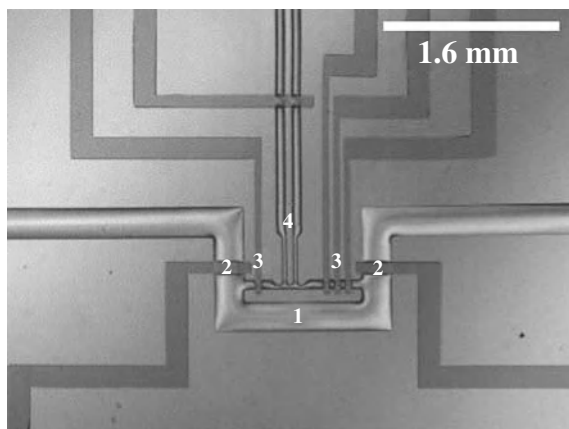


Fig. 12.10. Microphotograph of a two-layer cell sorting device. The control channel (peristaltic pumps and valves) appears *dark grey*, and the fluidic channels *light grey*. When a relevant cell is detected in the central part of the main channel (1), it is closed in by means of valves (2). Four small valves are then activated (3) to displace and extract it via the narrow channels (4). From [60]. Copyright 2004, with the kind permission of Elsevier

manner as possible. A final application concerns the detection of rare cells, down to one in several million. The latter case is relevant in particular to the early detection of some diseases, such as breast cancer.

Over the past few years, our understanding of the processes of cell biology has considerably improved, particularly on the subcellular and molecular level. However, it is still difficult to sort and manipulate single cells in a systematic way. The size of a cell is typically in the range 10–100 μm . The biochemical reaction time for its part varies over a very broad range (from the millisecond to the hour). In addition, cells are fragile, living entities. Methods of manipulation and analysis must therefore be adapted to these scales and these requirements, bearing in mind that the sensitivity of detection with regard to cellular or subcellular constituents must be extremely high. Microfluidics offers solutions that meet some aspects of the challenge, as a result of the many advantages discussed above.

Traditionally, in the macroscopic world, cell sorting is often done by flow cytometry, which involves optical analysis (by fluorescence or scattering) of the characteristics of cells that pass one by one before the detector. Each cell is subsequently enclosed in a drop whose trajectory is controlled by an electric field, in such a way as to sort and collect cells according to certain chosen properties. This is the method known as fluorescence-activated cell sorting (FACS), which is very widely used in cell biology. Up to now microfluidic devices for the purposes of cell sorting (μFACS) are in fact hybrid systems: the detection and analysis part remains macroscopic, while cell injection and displacement are carried out in microchannels, e.g., using the electroosmotic

effect [56], the dielectrophoretic effect [57], optical tweezers [58], etc. Here microfluidics offers the advantage of controlling the flows in channels of comparable size to the cell itself. This confinement effect means in particular that one can probe intercellular interactions or the interaction between the cells and the channel walls (previously grafted with specifically selected molecules). One is taking advantage here of the considerable increase in surface to volume ratio, forcing the cells to interact with a surface [59].

Many solutions remain to be explored with regard to the level of detection sensitivity in these microsystems. Analysis can be carried out on the basis of optical properties (absorption, diffraction, fluorescence, etc.), electrical properties (conductance, impedance, etc.), mechanical properties (Young's modulus via ultrasound), chemical or electrochemical reactivity, or other physicochemical properties.

Certain cancers are diagnosed by detecting pathological cells in very small amounts in biological samples which are necessarily of very limited volume, e.g., blood or bone marrow samples. Typically, the problem is to identify a single specific cell among 10 million on a reasonable time scale, of the order of 1 h. In this case, a two-stage sort can be profitable [60]. In Fig. 12.10, the cells are first injected into a wide-bore channel (lighter grey) with sufficient speed to ensure that several million cells go through in less than 1 h. When a rare cell is identified in the central U-shaped zone, the two valves (darker bands) of the main channel are closed. By means of four control channels, the cells in the central zone move one by one through the narrow part of the channel, whence the cell in question can be located and extracted via the two narrower microchannels.

Apart from the other advantages of size reduction, microfluidic systems have integration potential: the cell sorting operation can thus be integrated into more complex devices with other functionalities. For example, one can envisage a device that would carry out the following series of operations: cell sorting, cell lysis, extraction of DNA from a single cell, amplification by PCR on this DNA, and analysis by capillary electrophoresis with artificial gels as discussed earlier. Hence most analytical operations on single cells can be built sequentially into a single microfluidic chip, with the promise of considerable time-saving and cost reduction for many applications.

Microfluidic cell sorting provides a particularly good illustration of the cross-disciplinary nature of this field. It explicitly involves the knowhow from many different areas: microfabrication, micromechanics, hydrodynamics, optics, electronics, computing, chemistry, biochemistry, and immunology. Research on this theme still offers considerable opportunities and challenges, e.g., with regard to the speed and sensitivity of detection, the integration of excitation sources and high-performance sensors, etc. By developing appropriate applications in the fields of diagnosis and biomedical research, cell sorting devices may be among the first to provide a market for microfluidics.

12.5 Conclusion

Recent advances in the ability to administer liquids in very small amounts and manipulate objects in microfluidic devices in a controlled and accurate way have already led to significant progress in chemistry, biology, and physics. As a consequence, microfluidic systems have become a basic tool for fundamental and applied research, and this in a record time. Furthermore, as we have seen, this is a field of investigation which remains wide open to the inventiveness of research scientists.

From the industrial point of view, we are only at the beginning, but research effort by the major companies shows the interest raised by this technology. Current mergers and alliances show also that, on the one hand, the technology has not yet converged upon one or more solutions likely to become standards, and on the other hand, that microfluidics for the life sciences can only have meaning when combined with real biological challenges and questions.

In a wider perspective, while remaining in the field of biology, it is doubtless in its ability to communicate with the living world that this discipline will find its main expansion. Indeed, nature has been producing and using microscopic vessels for millions of years to transport chemical substances such as nutriments, hormones, gases, waste, etc. One sees here the two possible approaches that open up before us:

- To use living examples to inspire the construction of biomimetic synthetic analogs, using these to approach classic problems in a novel way. One can even imagine hybrid devices including functional proteins, which could fulfill a certain number of functions more effectively than fully synthetic devices.
- To interface ‘intelligent’ systems with living organisms. This brings to mind in particular the field of medicine, with probes, sensors, etc., able to carry out measurements and transmit them in real time, possibly involving communication with other active components. This approach would allow us to implement the various desired functions while improving the quality of health treatment or analysis (by making it less invasive or traumatic), and maintaining low fabrication costs compared with existing systems.

It seems likely that these two strategies, although they still look rather like science fiction, will lead to important achievements within a few years.

References

1. Tabeling, P.: *Introduction à la microfluidique*, Belin (2003)
2. Manz, A., Graber, N., Widmer, H.M.: *Sens. Actuators B* **1**, 244 (1990)
3. Beebe, D.J., Mensing, G.A., Walker, G.M.: *Annu. Rev. Biomed. Eng.* **4**, 261 (2002)

4. Reyes, D.R., Iossifidis, D., Auroux, P.-A., Manz, A.: *Anal. Chem.* **74**, 2623–2636 (2002)
5. Schwarz, M.A., Hauser, P.C.: Recent developments in detection methods for microfabricated analytical devices, *Lab on a Chip* **1**, 1–6 (2001)
6. Andersson, H., Van den Berg, A.: Microfluidic devices for cellomics: A review, *Sens. Actuators B* **92**, 315–325 (2003)
7. See Caliper at www.caliperls.com and Agilent at www.chem.agilent.com
8. Lacher, N.A., Garrison, K.E., Martin, R.S., et al.: Microchip capillary electrophoresis/electrochemistry, *Electrophoresis* **22**, 2526–2536 (2001)
9. Scarpa, A.: In: *Tabulae Neurologicae*, Pavia (1794)
10. Guyon, E., Hulin, J.P., Petit, L.: In: *Hydrodynamique physique*, EDP Science (CNRS), Paris (1991)
11. Jones, D., Leary, B., Boger, D.V.: The rheology of a concentrated colloidal suspension of hard spheres, *J. Coll. Interf. Sci.* **147** (2), 479–495 (1991)
12. Srivastava, N., Davenport, R.D., Burns, M.A.: Nanoliter viscometer for analyzing blood plasma and other liquid samples, *Anal. Chem.* **77** (2), 383–392 (2005)
13. Pit, R., Hervet, H., Leger, L.: Direct experimental evidence of slip in hexadecane at solid interfaces, *Phys. Rev. Lett.* **85** (5), 980–983 (2000)
14. Taylor, G.: Dispersion of soluble matter in solvent flowing slowly through a tube, *Proc. Roy. Soc. Ser. A* **219** (1137), 186–203 (1953)
15. Ajdari, A.: Steady flows in networks of microfluidic channels: Building on the analogy with electrical circuits, *Compt. Rend. Phys.* **5**, 539 (2004)
16. Levich, V.G.: In: *Physico-Chemical Hydrodynamics*, Prentice Hall, Englewood (1962)
17. Hunt, T.P., Lee, H., Westervelt, R.M.: Addressable micropost array for the dielectrophoretic manipulation of particles in fluid, *Appl. Phys. Lett.* **85** (26), 6421–6423 (2004)
18. Brzoska, J.B., Brochard-Wyart, F., Rondelez, F.: Motions of droplets on hydrophobic model surfaces induced by thermal gradients, *Langmuir* **9** (8), 2220–2224 (1993)
19. Burns, M.A., Mastrangelo, C.H., Sammarco, T.S., Man, F.P., Webster, J.R., Johnsons, B.N., Foerster, B., Jones, D., Fields, Y., Kaiser, A.R., Burkes, D.T.: Microfabricated structures for integrated DNA analysis, *Proc. Natl. Acad. Sci. USA* **93**, 5556–5561 (1996)
20. de Gennes, P.-G., Brochard-Wyart, F., Quéré, D.: In: *Gouttes, bulles, perles et ondes*, Belin, Paris (2002); 2nd edn. with CDROM (2005)
21. Vallet, M., Berge, B., Vovelle, L.: Electrowetting of water and aqueous solutions on poly(ethylene terephthalate) insulating films, *Polymer* **37** (12), 2465–2470 (1996)
22. Cho, S.K., Moon, H.J., Kim, C.J.: Creating, transporting, cutting, and merging liquid droplets by electrowetting-based actuation for digital microfluidic circuits, *J. Microel. Syst.* **12** (1), 70–80 (2003)
23. Washizu, M.: Electrostatic actuation of liquid droplets for microreactor applications, *IEEE Trans. Indust. Appl.* **34**, 732–737 (1998)
24. Buguin, A., Talini, L., Silberzan, P.: Ratchet-like topological structures for the control of micro-drops, *Appl. Phys. A* **75**, 207 (2002)
25. Fouillet, Y., Jeanson, H., Chartier, I., Buguin, A., Silberzan, P.: Moving droplets with a micro catenary, *Houille Blanche* **4**, 37 (2003)

26. Hirschberg, D., Palmberg, C., Jörnvall, H., Bergman, T., Gustafsson, M.: Integrated sample preparation and MALDI mass spectrometry on a microfluidic compact disk, *Anal. Chem.* **76**, 253–502 (2004)
27. Wixforth, A.: Acoustically driven planar microfluidics, Superlattices and Microstructures **33**, 389–396 (2003)
28. Madou, M.: In: *Fundamentals of Microfabrication*, CRC Press, Boca Raton (1997)
29. Rai-Choudhury, P. (Ed.): *Handbook of Microlithography, Microfabrication and Microsystems*, Vols. 1 and 2, SPIE Press (1997)
30. Chen, Y., Pépin, A.: Nanofabrication: Conventional and non-conventional methods, *Electrophoresis* **22**, 187 (2001)
31. Koch, M., Evans, A., Brunnschweiler, A.: In: *Microfluidic Technology and Applications*, Research Studies Press Ltd., England (2000)
32. Koch, M., Harris, N., Evans, A.G.R., White, N.L., Brunnschweiler, A.: *Proceedings Transducers '97*, Chicago (1997)
33. Chen, Y., Pépin, A.: Emerging nanolithographic methods. In: *Nanosciences. Nanotechnologies and Nanophysics*, ed. by P. Houdy, C. Dupas, M. Lahmani, Springer, Berlin Heidelberg New York (2006)
34. Xia, Y., Whitesides, G.M.: Soft lithography, *Angew. Chem. Int. Ed.* **37**, 550 (1998)
35. Unger M.A., Chou H.-P., Thorsen, T., Scherer, A., Quake, S.R.: Monolithic microfabricated valves and pumps by multilayer soft lithography, *Science* **288**, 113 (2000)
36. Terray, A., Oakey, J., Marr, D.: *Science* **296**, 1843 (2002)
37. Stroock, A.D., Dertinger, S.K.W., Ajdari, A., Mezic, I., Stone, H.A., Whitesides, G.M.: Chaotic mixer for microchannels, *Science* **295**, 647–651 (2002)
38. See *Proceedings of the International μ TAS Conference* (Micro Total Analysis Systems) published every year by Kluwer, Dordrecht, Netherlands
39. Sanjoh, A., Tsukihara, T.: *J. Cryst. Growth* **196**, 691 (1999)
40. Zheng, B., Tice, J.D., Ismagilov, R.F.: *Adv. Mater.* **16**, 1365 (2004)
41. Zheng, B., Tice, J.D., Roach, L.S., Ismagilov, R.F.: *Angew. Chem. Int. Ed.* **43**, 2508 (2004)
42. Verneuil, E., Buguin, A., Silberzan, P.: *Europhys. Lett.* **68**, 412 (2004)
43. Hansen, C.L., Skordalakes, E., Berger, J.M., Quake, S.R.: *Proc. Nat. Acad. Sci.* **99**, 16531 (2002)
44. Isambert, H., Ajdari, A., Viovy, J.L., Prost, J.: *Phys. Rev. Lett.* **78**, 971–974 (1997)
45. Volkmuth, W., Austin, R.H.: DNA electrophoresis in microlithographic arrays, *Nature* **358**, 600–602 (1992)
46. Huang, L.R., et al.: A DNA prism for high-speed continuous fractionation of large DNA molecules, *Nature Biotech.* **20**, 1048 (2002)
47. Bakajin, O., Duke, T.A.J., Tegenfeldt, J., Chou, C.-F., Chan, S.S., Austin, R.H., Cox, E.C.: Separation of 100-kilobase DNA molecules in 10 seconds, *Anal. Chem.* **73**, 6053–6056 (2001)
48. Kaji, N., Tezuka, Y., Takamura, Y., Ueda, M., Nishimoto, T., Nakanishi, H., Horiike, Y., Baba, Y.: *Anal. Chem.* **76**, 15–22 (2004)
49. Tegenfeldt, J., Prinz, C., Cao, H., Chou, S., Reisner, W., Riehn, R., Wang, Y.M., Cox, E., Sturm, J., Silberzan, P., Austin, R.: The dynamics of genomic-length DNA molecules in 100 nanometer channels, *Proc. Natl. Acad. Sci. USA* **101**, 10979 (2004)

50. Han, J., Craighead, H.G.: Separation of long DNA molecules in a microfabricated entropic trap array, *Science* **288**, 1026 (2000)
51. Schnelle, T., Müller, T., Hagedorn, R., Voigt, A., Fuhr, G.: Single microelectrode dielectrophoretic tweezers for manipulation of suspended cells and particles, *Biochim. Biophys. Acta* **1428**, 99–105 (1999)
52. Gorre-Talini, L., Spatz, J., Silberzan, P., Dielectrophoretic ratchets, *Chaos* **8**, 650 (1998)
53. Huang, L.R., Silberzan, P., Tegenfeldt, J.O., Cox, E.C., Sturm, J.C., Austin, R.H., Craighead, H.: The role of molecular size in ratchet fractionation, *Phys. Rev. Lett.* **89**, 178301 (2002)
54. Korda, P.T., Taylor, M.B., Grier, D.G.: *Phys. Rev. Lett.* **89**, 128301 (2002)
55. Huang, L.R., Cox, E.C., Austin, R.H., Sturm, J.C.: A microfluidic tango: Separation without dispersion, *Science* **304**, 987 (2004)
56. Fu, A.Y., Chou, H.P., Spence, C., Arnold, F.H., Quake, S.R.: *Anal. Chem.* **74**, 2451–2457 (2002)
57. Fiedler, S., Shirley, S.G., Schnelle, T., Fuhr, G.: *Anal. Chem.* **70**, 1909–1915 (1998)
58. Enger, J., Goksör, M., Ramser, K., Hagberg, P., Hanstorp, D.: *Lab on a Chip* **4**, 196 (2004)
59. McCloskey, K.E., Chalmers, J.J., Zborowski, M.: *Anal. Chem.* **75**, 6868–6874 (2003)
60. Studer, V., Jameson, R., Pellereau, E., Pépin, A., Chen, Y.: *Microelectronic Engineering* **73–74**, 852–857 (2004)

Data Processing

P. Grangeat

13.1 Nanobiotechnology and Data Systems

13.1.1 Nanobiotechnology

A new area of biology has been opened up by nanoscale exploration of the living world. This has been made possible by technological progress, which has provided the tools needed to make devices that can measure things on such length and time scales.

In a sense, this is a new window upon the living world, so rich and so diverse. Many of the investigative methods described in this book seek to obtain complementary physical, chemical, and biological data to understand the way it works and the way it is organised. At these length and time scales, only dedicated instrumentation could apprehend the relevant phenomena. There is no way for our senses to observe these things directly.

One important field of application is molecular medicine, which aims to explain the mechanisms of life and disease by the presence and quantification of specific molecular entities. This involves combining information about genes, proteins, cells, and organs. This in turn requires the association of instruments for molecular diagnosis, either *in vitro*, e.g., the microarray or the lab-on-a-chip, or *in vivo*, e.g., probes for molecular biopsy, and tools for molecular imaging, used to localise molecular information in living organisms in a non-invasive way. These considerations concern both preclinical research for drug design and human medical applications. With the development of DNA and RNA chips [1], genomics has revolutionised investigative methods for cells and cell processes [2,3]. By sequencing the human genome, new ways have been found for understanding the fundamental mechanisms of life [4]. A revolution is currently under way with the analysis of the proteome [5–8], i.e., the complete set of proteins that can be found in some given biological medium, such as the blood plasma. The goal is to characterise certain diseases by recognisable signatures in the proteomic profile, as determined from

a blood sample or a biopsy, for example [9–13]. What is at stake is the early detection of disease and personalisation of health care [14].

Another important area in nanomedicine and nanobiology is the control of electrical or mechanical activity using microelectrode arrays or micro electromechanical systems (MEMS), or even nano electromechanical systems (NEMS). These devices are well suited to provide functional support or substitutes for treating certain deficiencies (heart disease, brain disorders, etc.), e.g., for helping handicapped or elderly people. They are also used to study ion transport through cell membranes [15].

13.1.2 Data Systems

In the age of the information society, the aim of observational instruments is to provide a digital representation of the explored medium. They thus constitute an interface between the real world and the digital world. The observed object can be represented by a list of digital numbers. This numerical representation is then a source of information, which can be analysed, visualised, recorded, and disseminated by suitable digital processing systems. This is the view one finds in the convergence of nanotechnology, biology, information technology, and cognitive science, known as NBIC, which brings together these various fields of research. Information then forms the link between technology and knowledge, applied to a particular field: biology.

In this chapter, we shall be mainly concerned with digital data processing close to the sensor, used to make the necessary corrections and extract the desired numerical data. We shall then describe some of the main principles of data analysis. However, we shall not go into more detail concerning the huge field of bioinformatics [16–19], whose main aim is to integrate, exploit, and model data. We thus limit the discussion to that part of data processing that is related to data production.

This field of data processing applied to nanobiotechnology is nevertheless very extensive owing to the widely different means available. So the idea here will be to introduce the concepts required to define these processes. We shall give several examples. However, no attempt will be made to provide an exhaustive overview of all these techniques. The scheme considered here is outlined in Fig. 13.1. Through interaction with the observed medium, the measurement nanosystem provides a set of digital electronic signals. One must then carry out some form of digital processing to obtain the desired representation of the data. The electronic signals are thus processed to reconstruct the relevant numerical representation.

This includes a preprocessing stage to correct for defects in the sensors and improve the data, and a stage to extract the data. This numerical representation supplies the input data for data analysis processes, which seek in particular to extract a parametric or symbolic description, to help the user to interpret the measurements and take appropriate decisions.

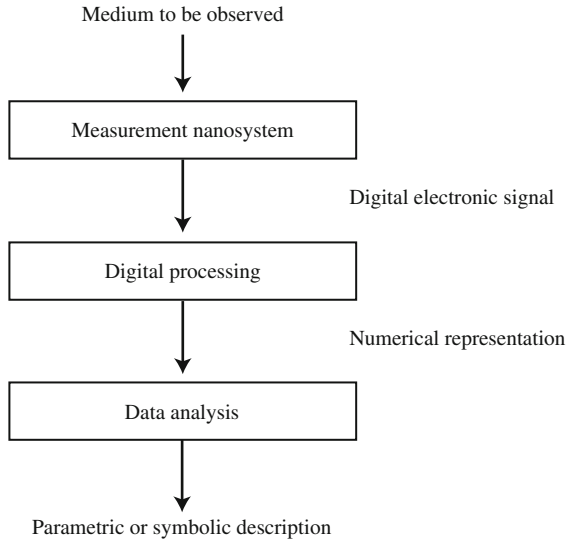


Fig. 13.1. The main steps in a data-processing system

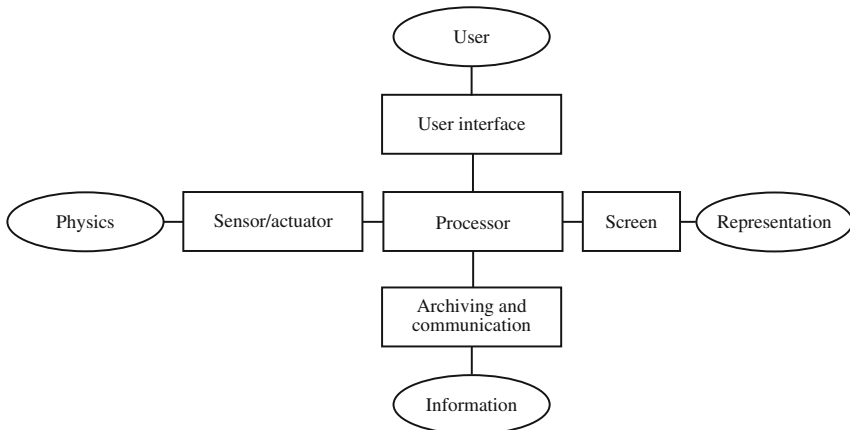


Fig. 13.2. Architecture of a data system

Figure 13.2 sums up the architecture of a data system. Data is processed by a processing unit or processor. The interaction with the physical medium is achieved by sensors or actuators. The user controls the system through an interface such as a keyboard or a mouse. The screen displays the calculated representation. The latter can be archived on a suitable device, such as a mass memory, or communicated to other remote users via a communication network. Conversely, the processor accesses, via its mass memory or communication network, the knowledge required to construct or analyse data, as happens when one accesses a remote data base.

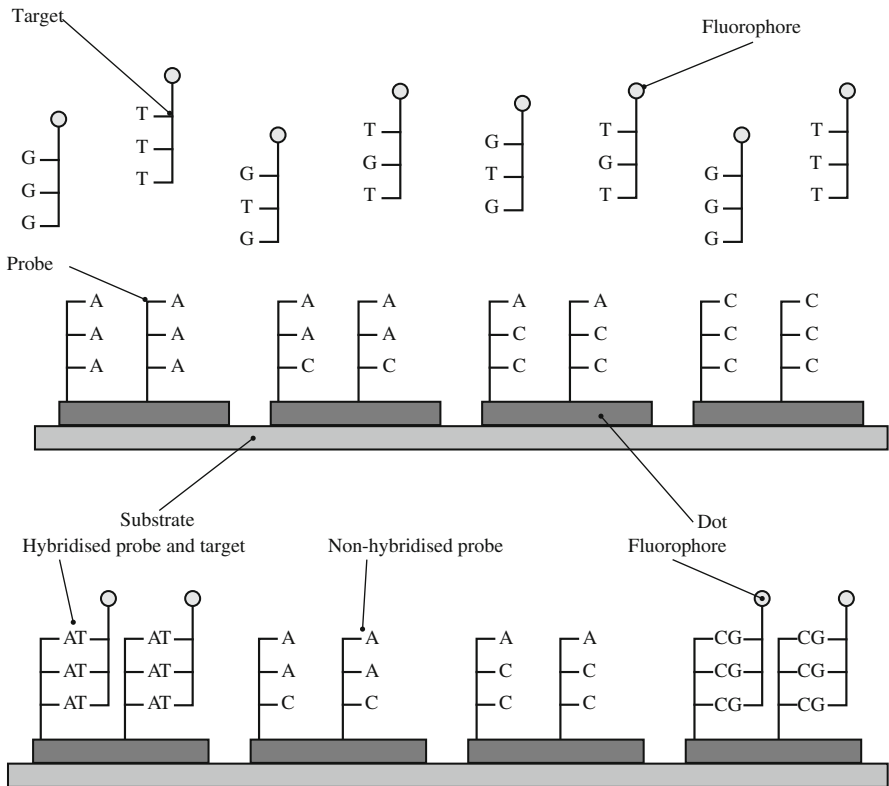


Fig. 13.3. Target/probe hybridisation array. *Upper:* Microarray functionalised for hybridisation reactions by contact with the solution. *Lower:* Microarray after hybridisation, ready for fluorescence recognition

13.1.3 Three Examples

We now present three examples of data-processing problems encountered in nanobiotechnology.

Target/Probe Hybridisation Arrays

A hybridisation array [1, 20] is a component that can recognise, in a mixture containing target molecules, the presence of molecular probes arranged selectively at a set of points on the microarray (see Fig. 13.3). The mixture of targets is characterised by a vector $c(j)$ specifying the number of targets of type j present in the mixture, and which are to be recognised.

This mixture interacts with an array of dots on which the probes are grafted (see Fig. 13.3 upper). After interaction, let $p(i)$ be the number of targets hybridised with the associated probe on dot i (see Fig. 13.3 lower). In the

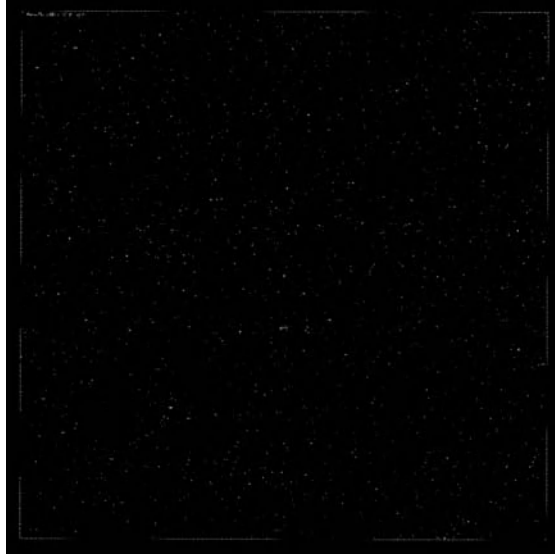


Fig. 13.4. Fluorescence image of an Affymetrix chip (bioMérieux–LETI joint group)

example of Fig. 13.3, hybridisation occurs by the formation of specific bonds between bases A and T on the one hand, and bases C and G on the other. Only those strands carrying A, C, G, T sequences that are complementary term by term will give rise to hybridisation. In this case, the process is interpreted by saying that the probe has recognised the target. In Fig. 13.3, the idea is illustrated for 3-base strands. In reality, oligonucleotides with 15–80 bases are used.

In an optical measurement using fluorescence, the targets are endowed with a fluorescent label. After rinsing the microarray, only those dots on which hybridisation has occurred will carry fluorophores (see Fig. 13.3 lower). To read off the results, the dot is excited by laser radiation and the intensity $m(i)$ of the fluorescence signal from dot i is recorded. Figure 13.4 shows the fluorescence image of an Affymetrix chip made by the joint bioMérieux–LETI team.

The first problem solved by processing is to calculate a restored signal $r(i)$ from the measurements $m(i)$, taking into account any defects in the sensor and the measurement method. In a quantification approach, the aim of processing is then to calculate, for each probe j , the number $c(j)$ of associated targets present in the mixture. This requires a functional model for each of the hybridisation and read stages. In a detection approach, the aim of analysis is to decide for each dot whether the probe j was present in the solution or not. This decision is made by thresholding or classification processes, for example.

Peptide Chromatography and Mass Spectrometry

Consider a mixture of peptides characterised by a vector $c(j)$ specifying the number of molecules of each peptide j present in a mixture to be analysed. This mixture is injected in a chromatography column. Let $p(j, L, t)$ be the number of peptides of type j present at the column outlet at time t , where L parametrises the length of the column.

At each time t , the mixture leaving the chromatography column is ionised, then injected into a mass spectrometer. If $m(v, t)$ is the mass spectrum at time t , with v the time of flight of the ions in the mass spectrometer or the associated mass-to-charge ratio.

A restored signal $r(v, t)$ is calculated, which compensates for any defects in the acquisition system. Then, in a quantification approach, the idea is to calculate, for each peptide j , the number of molecules $c(j)$ that were present in the mixture at the column inlet. To do this, a functional model is required for each step in the sequence chromatography, ionisation, spectrometry. Finally, in a detection approach, one must decide by data analysis whether or not the peptide j was present in the solution. This decision can be made, for example, by thresholding or classification. By data analysis, we also seek to recognise which proteins correspond to the detected peptides.

Imaging

An imaging system produces a map of a physical parameter. In morphological imaging, where one studies the architecture or topology of a scene or object, this parameter may be an absorption density, a refractive index, a depth, and so on. In functional imaging, where one studies the functioning or metabolism of an organism via natural or artificial markers, this parameter will correspond to a density of markers, or a density of emitters, if these markers are labelled by radioactive, luminescent, or fluorescent elements. Let f be the function describing the spatial distribution of this parameter. The parameter is observed using an imaging system which supplies an image m . Different imaging systems can be used, such as near-field microscopes, electron microscopes, confocal microscopes, imaging systems exploiting surface plasmon resonance (SPR) [21, 22] or digital holography, or spectroscopic imaging [23].

The first aim of processing is to calculate a restored signal r . For example, we attempt to reconstruct the initial function f which served to generate the measured image m [24, 25]. This problem becomes still more delicate when reconstructing a 3D image f from a set of 2D projections m recorded from different angles of incidence. This is the problem in tomography [26, 27]. By image analysis, we then seek to recognise the characteristic shapes of target objects or to differentiate pathological images from the others by classification.

In each of these examples, one has to determine minute concentrations of molecules or make nanometric images. Many of these techniques thus involve

nanoscale characterisation. The data parameters and vectors depend on the physical phenomena taking place.

Looking at the problem quite generally, there are three main types of function for these systems:

- The first is detection. The problem is to decide whether or not a given molecule is present in the medium, or to diagnose the presence of a disease, a genetic mutation, or some active element. The resulting data is binary: either the element is present or it is not. From a statistical standpoint, one speaks of a statistical test. One often defines a threshold associated with a decision function to handle this kind of data. Such a process is characterised by its sensitivity, i.e., its ability to minimise false negatives, in which it is decided that the element is not present, whereas it is in fact physically present, and also by its specificity, i.e., its ability to reject false positives, in which it is decided that the element is present, whereas in fact it is not physically present.
- The second is observation. When a mixture is analysed, one seeks to establish a list of its constituents. When a cell is imaged, the idea is to see and identify the various organs. This function is characterised by the resolution limits of the instruments.
- The third is quantification. One seeks to accurately determine the concentration of molecules in a mixture, or the density of a label in a functional image in order to assess the level of activity. This function is characterised by the accuracy, fidelity, and dynamic range of the measurements. An important parameter is the signal-to-noise ratio.

13.1.4 Technological Bottlenecks

Size reduction requires nanobiotechnology to overcome certain technological obstacles which may be summarised as follows:

- *Scale Effects.* The transition to nanometric dimensions changes the orders of magnitude for the relevant physical phenomena [28]. This means that certain phenomena tend to dominate over others, e.g., in microfluidic flows or interactions between molecules and substrates, whence the functional models describing the measurement principle must be suitably adapted. This happens, for example, in impedance spectroscopy. The standard model here is Randles' model, which describes the dynamical behaviour of a redox reaction in the case of semi-infinite diffusion [29]. It includes a resistance, a Warburg diffusion impedance, and an interfacial double-layer capacitor. This model is no longer appropriate when the redox phenomena occur on functionalised microelectrodes, and more sophisticated models are required, such as non-integer derivative models to describe the relevant diffusion-convection effects [30]. Measurement noise also tends to increase when sensor sizes are reduced [31].

- *Effects of Parallel Processing and High Throughput.* High throughput analysis on microarrays involves multiplexing, i.e., carrying out measurements simultaneously. There is a transition here from single input/single output (SISO) measurements to multiple input/multiple output (MIMO) measurements [32]. One simple idea is matrix addressing, in which inputs are made along the rows and outputs along the columns of the device [33]. With such systems, suitable processing must be used to identify the behaviour and parameters of each elementary cell from these input–output relations. Moreover, the transition to high throughput measurements involves a compromise between the measurement rate and the resolution and sensitivity of such devices. This is the case for example in analytical processes where the kinetics of chemical reactions and transport phenomena must be taken into account in the physicochemical separation stage.
- *Reliability.* Several sources of error can affect measurements. The first concerns the effects of dispersion on components either during fabrication or as they evolve over time. The size reduction of detection components is inevitably accompanied by a high statistical scatter in characteristics such as gain, linearity, dynamic range, or functional parameters. The second source of error concerns noise, due in particular to the fact that the interactions producing the signal are random phenomena. Finally, the third source of error is interference effects such as non-specific hybridisation, contamination, or diffusion effects. These error sources require corrective preprocessing and regularisation processes to make such systems more robust. When possible, another solution is to set up servosystems using automated feedback techniques to control measurement quality.
- *Specific Features of the Interface with Biology.* The in vivo exploration of living organisms, e.g., cells, micro-organisms, plants, organs, animals, patients, imposes its own restrictions, such as
 - Biocompatibility, to ensure that measurement conditions are compatible with living functions.
 - The temporal evolution of observations, owing to the fact that living organisms are evolving.
 - The complexity of observed phenomena, discussed further below.
 Dedicated processing is then essential.
- *Measurement Complexity.* Biological media are often highly complex. A liquid like blood or serum contains a wide range of different molecules and cells with a very broad range of concentrations. A single human cell contains some 35,000 genes, several thousand messenger RNA, and several million possible proteins. In addition, many molecules such as DNA or proteins interact with one another via functional networks [34]. This means that the exploration of living beings requires a multiparameter approach, combining several physical means of exploration and also associating a multiscale description ranging from the nanoscale to the macroscopic. The 3D structure of objects such as molecules or cells also carries information regarding the spatial organisation of phenomena. Furthermore, biological

phenomena evolve in time, and it is essential to investigate their kinetics in functional studies. Faced with this complexity, the key notion is then the integration of data in order to combine these investigations. The problem of processing such a huge amount of data demands considerable computational resources, such as processors integrated in components, processor farms, or computational grids, not to mention access to databases via suitable means of telecommunications.

13.1.5 Automated Measurements

From a functional point of view, biochemical analysis generally involves manipulation stages in which one seeks to separate components by physicochemical processes, read stages, and data-processing stages. For complex analytical tasks, these stages must be carried out by an automatic system which controls and coordinates. An example of such automation is provided by robots which deposit drops on microarrays. These robots must coordinate matter transfer, physicochemical processes, read processes, and data handling.

In the context of nanobiotechnology, an important line of research concerns the so-called lab-on-a-chip, where all the different steps must be miniaturised and integrated into a single microsystem [35–40]. The advantages here are that one can work with smaller amounts of sample or reagent, and improve the yield of reactions by increasing the surface-to-volume ratio. Furthermore, it will become possible to combine on the same component both the physicochemical process and the on-board data processing. In molecular medicine, this kind of lab-on-a-chip opens the way to a new type of use, rather like a portable health unit, moving toward decentralised autonomous analytical systems that can be used at the bedside, at the doctor's surgery, in laboratories for biological analysis, or even at the patient's own home. This will contribute to the development of personalised medical treatment. However, other areas will benefit from the possibility of such a portable measurement unit, including environmental control and the fight against bioterrorism and biological weapons.

13.1.6 Layout of this Chapter

This chapter is set out according to the characteristic layout of a data system, described below and illustrated in Fig. 13.1. Section 13.2 describes the different forms of numerical representation and the basic rules for avoiding error. Section 13.3 discusses preprocessing and Sect. 13.4 data extraction. Finally, Sect. 13.5 introduces data analysis.

13.2 Representing Data

13.2.1 Data Structures

A wide range of representations are available for data obtained in nanobiotechnology. The computer code is always a discrete list of numbers. However, this list may correspond to different structures, depending on what the data represent and depending on their dimensions. Let us now describe the main data structures encountered in practice.

For a molecular profile, one has a list of values associated with the different molecules under consideration. The values represent the amounts of molecules in solution. These quantities can be specified by concentrations or by numbers of molecules. The list of values is grouped together in a data vector.

The interactions between molecules in a metabolic network are represented by relational graphs describing the molecules which interact together and the associated reaction rate constants. These graphs are similar to those used in a formal way to describe finite automata networks.

The current measured at an electrode in an electrochemical measurement, the voltage in a potential measurement, and the fluorescence intensity produced by a given laser excitation are 1D signals. In the electronic circuits close to sensors, they often exist analogically and are represented by a function. An analog-to-digital converter transforms these continuous signals into a discrete series of numerical values called samples. In systems where several measurements are made in parallel, as in electrode arrays, the values at a given instant are represented by a vector, in which each index indicates the element carrying out the measurement. Sampled time signals will then specify a vectorial numerical series.

Spectral measurements also correspond to a discrete series of values associated with each of the spectral components under investigation. The latter may be electrical or mechanical frequencies, or nuclear or electromagnetic energies, or mass-to-charge ratios in mass spectrometry. These spectra are also represented by vectors of numerical values. In chromatography, one studies time series of spectra as a function of the exit time from the chromatography column. This then defines a vectorial sequence of numerical values that can be represented in the form of an image or spectrogram.

Images constitute a multidimensional form of signal. They represent the spatial distribution of a characteristic physical parameter. When this spatial distribution is studied on a two-index surface, usually a plane, one speaks of a 2D image. This is the case in particular for microscopes studying the state of a surface. It is also the case for cross-sections acquired by tomographic setups. However, living entities such as cells, organs, animals, or human patients have a bulk structure. It is then desirable to produce 3D images in order to locate the relevant data in an appropriate way. To study a time evolution, a sequence of images is obtained. Each image is described continuously by a function and in a discrete way, after sampling, by a vector. This vector is specified by

concatenating the rows of the image. Image sequences are associated with a vectorial numerical series.

Some devices study the spatial distribution of a spectral measurement, e.g., systems carrying out time-of-flight secondary ion mass spectrometry (TOF-SIMS) [41] or imaging systems using impedance spectroscopy. The results can then be treated as 3D images, in which two of the dimensions correspond to spatial variation and the third to spectral variation.

From the computing point of view, these lists of numerical values are stored in a file. They constitute an elementary piece of data for an examination. To carry out measurement campaigns or routine examinations, one must use the appropriate archiving facilities for this kind of data, also called databases. The data relating to a given patient, or a given experiment, are gathered together in the same directory. These directories can combine data coming from different types of measurement. One then speaks of multimodality. When a large number of cases must be explored systematically, as for exploration of the genome or the proteome, or when a study concerns a large number of patients, a data bank is set up.

13.2.2 Sampling and Quantification

The transformation of a continuous signal into a discrete signal is called sampling. The inverse transformation is referred to as reconstruction. It employs suitable interpolation techniques. For a sampled signal to be able to give back the original continuous signal in an exact manner, the sampling must satisfy Shannon's theorem, which requires the sampling frequency to be equal to at least twice the cutoff frequency of the signal. This condition must be taken into account when studying objects that are very small compared with the size of the sample. Otherwise, one obtains errors known as aliasing.

The sampled data are then quantified using an analog-digital converter. This converter is characterised by its dynamic range and quantification interval. After quantification, measurement results are generally coded on 1, 2, or 4 bytes, where 1 byte corresponds to 8 bits (binary digits).

13.2.3 Measurement Noise

When an electrical signal is generated by some physical phenomenon, this involves the production and transport of electrons or holes in materials. These interactions are governed by statistical laws. The recorded signal will thus always involve random fluctuations referred to as measurement noise [31]. Furthermore, this measurement noise will grow in importance as one approaches nanometric dimensions, close to the dimensions of the atoms and molecules generating these signals. Such fluctuations must therefore be taken into account when describing measurements. The noise is characterised by its probability density, or first order moment, generally described by its standard deviation, and by its autocorrelation function, or second order moment,

describing the mutual dependence of neighbouring samples [24]. The auto-correlation function is often specified by the power spectral density (PSD) of the noise, which is in fact its Fourier transform. At a sensor, the interaction phenomenon generating the signal is often distributed over a certain volume or surface. This means that sample size reduction may be accompanied by an increase in the mutual dependence of neighbouring samples.

13.2.4 Direct or Indirect Measurement

In some cases, data analysis can be carried out directly from a recorded signal. This is the case in the first example described on p. 778, where one wishes to determine whether or not hybridisation has occurred at each given dot. In the second example (see p. 780), the spectrum is used to detect directly whether a peptide of given mass/charge ratio is present or not. But in the other cases, this direct observation is not sufficient. The measurement tool is characterised by a direct transformation relating the desired physical parameter to the measurements. To retrieve the physical parameter from these indirect measurements, one has to solve the inverse problem, i.e., calculate the physical parameter using a system of equations specified by the direct problem and the measurement data. This is generally the situation when one has to make a quantitative measurement of a concentration. In example 1, the relevant parameter is the concentration of molecules present in the solution, whereas the measured quantity is the fluorescence intensity. In example 2, the desired parameter is the concentration of peptides present in the solution, whereas the measured quantity is the mass spectrogram. In example 3 (see p. 780), starting from an impedance spectroscopy image, one seeks to map the concentration of molecular deposits on the surface under investigation. Under an atomic force microscope, the depth of each point on the surface of the object will be deduced from the interaction force between the sensor and the object. This problem is also encountered in tomographic devices such as electron microscopes, where recorded images correspond to a projection of the attenuation density in the direction of the measurement radiation. To reconstruct a 3D map from this attenuation density, projections are carried out from different measurement angles. One then reconstructs the object by inverting the system of equations specified by this set of projections [26].

13.3 Correcting for Sensor Defects and Improving the Data

The aim at this stage of data processing, also called preprocessing, is to prepare the data for direct use by the user, or for subsequent numerical processing. In particular, one must correct for shortcomings due to the detectors and conditions of use, so that measurement results can then be handled in a

standardised way. This requires correcting for systematic errors in the detection method or technologies being used. The aim in preprocessing data is to provide ideal measurement results according to the following criteria:

- Linearity, in the sense that, on a complex object for which the characteristic parameter is a linear combination of the characteristic parameters of the elementary objects making it up, the measurement result for the complex object will just be the corresponding linear combination of the measurements associated with the elementary objects.
- Independence from other measurements, dispersions in the components, and experimental conditions.
- Reproducibility and robustness.
- No aberrations (outliers).
- Good geometric, i.e., spatial or temporal, localisation, when necessary.

To achieve such goals, numerical processing is necessary, as we shall now describe. From a mathematical standpoint, the problem is to define a transformation between a list of raw measurement results and a list of ready-to-use measurement data. A general discussion of preprocessing methods in image handling can be found in the works by Jahne [24] and Maitre [25], and a presentation of preprocessing and normalisation in the context of microarrays is given in the book by Draghici [1].

13.3.1 Linearity and Calibration

This stage involves converting the raw data into usable data. The idea is to specify a conversion rule, using either an explicit formula or a conversion table.

The first example is calibration of the detector response [32, 42–44]. A set of measurements is made on calibrated values. These measurements can be represented on a graph, with the physical quantity on the horizontal axis and the measured value on the vertical axis (see Fig. 13.5). The idea is then to find a calibration curve that passes through the measurement points. One usually adopts a model conversion rule, such as a linear or polynomial model. One then estimates the curve parameters by regression. Statistical analysis of these measurements defines the margin of error associated with the bounding curves of the regression band. Then, whenever a measurement is made, one seeks the associated physical quantity on this calibration curve. Calibration measurements located outside the regression band are known as outliers.

This stage is also an opportunity to introduce a conversion of the measurement results such as a logarithmic conversion. This is necessary when handling results that obey exponential laws, such as the attenuation of radiation in matter, or geometric progressions appropriate for the multiplication of biological species. This stage can also convert the measured electrical quantity into the desired physical quantity.

Finally, this conversion can be used to correct for dispersion in the gain and dynamic range of the measurements. For example, in an imaging system,

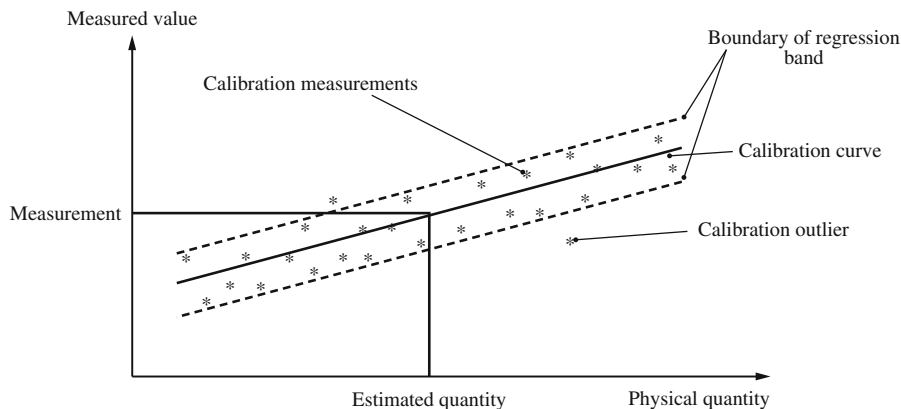


Fig. 13.5. Estimating a physical quantity from a measured value by calibration and regression

one seeks to adapt the dynamic range so as to make full use of the display range of the screen, by means of conversion tables called look-up tables (LUT). One can also arrange for the histogram of displayed values to be as uniform as possible by histogram equalization [45], to adapt the image to the analytical capacities of our own vision.

13.3.2 Independence and Normalisation

In the previous case, we considered an elementary measurement. If we must now carry out several measurements at the same time, the above conversion rule may no longer be adequate. This is the case for example in imaging systems, where the measurement on one pixel is affected by those on other pixels. From a physical point of view, this happens if physical interference effects such as scattered radiation perturb the measurement. In preprocessing, one must take into account cases where interference creates a difference of level, or offset, that must be corrected. There are two different families of approach here.

The first family seeks to estimate the value of the background level on a pixel from neighbouring pixels. The values are analysed in a relative way. One then calculates the difference between the measured value and the estimated background value. This estimate can be made by calculating an average value or by calculating an interpolated value by passing a surface through the values of the neighbouring pixels.

The second family seeks to use reference or control points corresponding to measurements in which one knows that only the offset is present. For example, one can use measurement points outside the zone of the objects, or measurement points where the measured value is forced to zero (blank

measurement). In this case, the offset is estimated by calculating an interpolated value from these reference points.

Similar approaches can be developed to correct for the temporal variations of a signal. Offsets can appear here too, e.g., due to remanence phenomena. These are corrected either by subtracting a sliding average value, or by subtraction on the basis of a reference time series.

13.3.3 Noise and Filtering

Due to the statistical nature of the phenomena producing the electronic signal, every measurement is necessarily random. In data processing, one models this by characterising the measurement by the sum of a deterministic signal and random noise. Sensors are usually integrating devices which average the measured quantity over the active surface of the sensor or the time window of the measurement. The more one tries to improve the spatial or temporal resolution, the more the signal-to-noise ratio between the average value of the signal and the standard deviation of the noise will be reduced. Measurement noise is thus a limiting factor for nanotechnological devices.

The main technique for reducing noise is to calculate an average over several measurements. These measurements can come from repeated acquisitions or from several sensors. For signal or image acquisition systems, one solution is to calculate this average value by combining spatially or temporally neighbouring values. One then speaks of spatial or temporal filtering [24, 25]. In linear filtering, the value of a sample is replaced by a linear combination of the values of neighbouring samples. More sophisticated approaches have been devised which use a wavelet decomposition of the image [25, 46]. Wavelets are bases of functions ensuring good space and frequency localisation, allowing joint space–frequency analysis, spatially adaptive filtering, and multiscale processing. This spatial localisation cannot be achieved with the trigonometric functions used for Fourier analysis. The disadvantage with linear filters is that one has to smooth the image or signal. When this becomes a serious limitation, nonlinear filtering is preferable, with a rank filter or a filter attributing different weights to samples according to some criterion depending on the level of variation of the signal or image [47].

In nanotechnology, techniques from microelectronics provide a way of replacing a single measurement on one sensor by a large number of measurements on several sensors. One advantage here is to be able to introduce these digital smoothing processes to obtain an equivalent noise level. But since these are numerical processes, one has more degrees of freedom to adapt the processing to the desired quality criteria, in particular by introducing nonlinear processes.

13.3.4 Outliers

The introduction of mass production using the technologies of microelectronics brings with it many advantages, but at the same time it creates constraints.

In particular, some measurement points may not work, and hence not deliver a measurement signal, or produce an aberrant signal. This can happen, for example, in image sensors or microarrays. It is then important to be able to identify these sensors and process the measurements in a suitable way.

To decide whether a measurement is an outlier, one needs to be able to detect aberrant behaviour. Such behaviour may for example correspond to a calibration measurement in which the point defined by the calibration value and the measured value lie away from the average calibration curve by a distance greater than the distance that can be attributed to measurement noise (see Fig. 13.5). Another idea is to compare the value with neighbouring values in space or time. If the discrepancy is too large as compared with the measurement noise, the value can be declared an outlier.

Once identified, these outliers can either be rejected, or replaced by a value calculated from neighbouring values by filtering or interpolation. In imaging, when the error only involves a few pixels distributed randomly over the image, one can use a median filter, where the outlier is replaced by the median of the neighbouring values.

13.3.5 Distortion and Geometric Corrections

With an imaging system, one needs to determine not only the level of each pixel, but also its coordinates in the image plane. This is done either by construction of the sensor, or by measurement for scanning systems. When localisation has to be very accurate, it is often necessary to carry out a geometrical rectification of the image. A simple idea is to produce the image of a test card comprising a set of points or a grid and to identify the geometrical transformation required to transform the measured positions into rectified positions. This then specifies a distortion formula. The task then is to calculate the values to be attributed to each point of the rectified image by interpolation on the raw image as dictated by this distortion formula.

The problem of geometrical adjustment of data also arises when comparing images taken over a population of specimens in the case where one hopes to find the correspondence between the relevant regions for the purposes of comparison. The geometrical transformations provide a way of normalising data when there are morphological differences between specimens.

The same problem also occurs in spectroscopic data when aligning different spectra for the purposes of comparison. One approach is to use reference spikes and distort the spectra until their positions correspond [48].

13.4 Data Extraction

13.4.1 Extracting Physical Quantities

Acquisition systems in nanobiotechnology provide an interface between the living world and the world of numerical data. This interface involves

physicochemical phenomena which transform the physical quantities one seeks to characterise into electronic data. In the preprocessing stages, one aims to format this electronic data. In this processing stage, the objective is to extract information, i.e., to retrieve the sought physical quantities from the electronic data.

Applications are extremely varied. The first simple case is the extraction of a signal from noisy data to reduce the statistical fluctuations and remove the other components [31, 49]. A more complex case is the mixing of the desired signals [32, 50]. This happens when one seeks to multiplex measurements. Multiplexing occurs for example when one carries out spectroscopic measurements on a mixture [51] or electrical measurements over a parallel electrode array. The problem then is to recover the elementary signals or the associated physical parameters from the measurement data.

Another example concerns microscope imaging systems, e.g., confocal imaging, where, owing to the limitations of the measurement devices, the observed image is fuzzy compared with the ideal image. The aim of processing in this case will be to restore the image and calculate an image close to the desired exact image [52].

Another case is tomography, where measurements provide a set of projections of the map of the relevant parameter along different measurement axes [26, 27]. This happens in attenuation imaging or emission imaging. Each projection corresponds to the accumulation of the absorber or emitter density along the measurement axis. Image reconstruction then involves calculating the spatial distribution of this density from the projection data. This kind of problem generalises to nanocharacterisation processes, using different physical means to explore the surface or the interior of a biological medium such as a cell or a functionalised surface after chemical deposition (SPR imaging, near-field microscopy, etc.).

Biochemical measurements often involve microfluidic manipulation and transport. This kind of transport often induces convection–diffusion effects which affect the composition of mixtures and the accuracy of such analysis. The problem then is to recover the initial composition of the mixture from measurements made at the outlet of the analytical system.

A final example is the estimation of an amount of hybrid products from electrochemical data obtained on immunological or DNA sensors [29, 53–56].

13.4.2 The Systems Approach

To implement data processing, the systems approach illustrated in Fig. 13.6 describes the measurement process by a system with input quantities X corresponding to the excitation signals, physical parameters P characterising the state of the device, and output quantities specified by the measurement data M [53, 57–60]. This system is modelled by a set of relations G of the form $G(X, P, M) = 0$ between these variables.

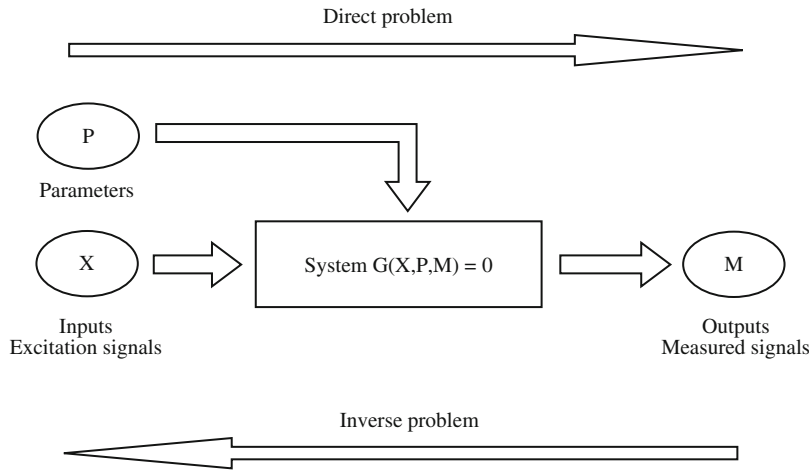


Fig. 13.6. The systems approach and inverse problems

Depending on the application, the required information may correspond either to the input quantities X , e.g., the image of a sample at the microscope input, or to the physical parameters P , e.g., the concentration of fluorescent deposits on a fluorescence measurement setup excited by laser.

These relations assume different forms depending on the application. In continuous methods, the input quantities X and output quantities M are functions. One then distinguishes integral methods, in which the measurement data M are related to the inputs by an integral transformation of the first kind associated with a characteristic integral kernel for the system under investigation [57], also called a Fredholm equation, and differential methods, in which the quantities are related by partial differential equations (PDE) [61]. In discrete methods, the input and output quantities are vectors related to one another by matrix relations, for example. One then distinguishes deterministic approaches and statistical approaches, the latter being used to describe the stochastic nature of the quantities and measurement interactions.

These relations result from physical models implying suitable interaction equations. However, in many cases, the physical models may be too sophisticated or too detailed, introducing too many parameters which it may prove difficult to ascertain. This is often the case in biology and medicine. One then tries to simplify the model to obtain a functional model that only takes into account the relevant variables and reproduces the behaviour, i.e., the relations of cause and effect between the input and output variables and the parameters, as faithfully as possible. Uncontrolled variables will be included in a systematic error term attributed to the approximate nature of the model. The choice of an appropriate model is one of the special skills of the data processor.

13.4.3 Inverse Problems

Once the measurement process has been modelled, the next task is to calculate the required data from the measurement data and knowledge either of the parameters, when one seeks to find the inputs, or of the input data, when one seeks to determine the parameters. The problem is therefore to solve the equation $G(X, P, M) = 0$, using the measurement data and known quantities.

This is an instance of the general field of inverse problems [51,57,62]. To define this solution, several key ideas have been defined by mathematicians and data processors. The direct problem involves calculating measurement data when inputs and parameters are known (see Fig. 13.6). In a certain number of cases, as in tomographic imaging or image restoration, this transformation is given explicitly. One then uses a description of the instrument, e.g., by a convolution equation with the point source response of the device, or by a transformation like the Fourier transform, Radon transform, or X-ray transform [26]. The Radon transform of a function is defined as the set of integrals over the hyperplanes of a space, and the X-ray transform as the set of integrals over the straight lines of the space. In two dimensions, these two transformations are equivalent. The X-ray transformation models measurements carried out with devices such as X-ray tomographs or electron microscopes, where radiation is attenuated between a point source and a detector. It is also used to describe measurements of emitter density along measurement axes, as happens in emission tomography if one neglects interference effects such as self-attenuation of the radiation. In other cases, this transformation is given implicitly by solving a partial differential equation. If one wishes for example to describe the transport of a solution in a microfluidic device, one must solve the fluid dynamics equations, given the initial conditions defined by the input mixture and the physical parameters of the transport equations.

Inverse problems aim to recover the desired information from measurements (see Fig. 13.6). To simplify the discussion here, we consider the case where the unknowns are the input quantities and the physical parameters are assumed to be given. The problem is thus to solve the equation $G(X, P, M) = 0$ given P and M .

Several difficulties may arise when implementing this program. These difficulties follow from the fact that the direct problems describing the measurement principles are not well posed, in the sense that one or more of the following criteria is satisfied:

- Non-existence of the solution. Owing to errors in the measurements or models, it may be that no input can transform exactly to the outputs corresponding to the measurement data.
- Non-uniqueness of the solution. Owing to an insufficient number of measurement results, there is an ambiguity allowing several solutions.
- Instability. A small perturbation of the measurement data leads to a significant perturbation of the solution.

Linear algebra provides general algorithms for discrete measurement systems described by systems of linear equations. Such algorithms will generally be available when considering small variations about some operating point.

A general problem is calibration of instrumental parameters. It is usually carried out by using measurement standards as input and solving the system of equations $G(X, P, M) = 0$ with respect to the variables P for these standards.

Under certain conditions, it may be that the physical parameters are unknown. One strategy then is to carry out a series of solutions, alternately estimating X given P and P given X .

13.4.4 Regularised Solutions

In order to calculate a solution, one must implement a method of numerical computation that gives an acceptable result. One strategy is to smooth the measurement noise, reduce the complexity of the direct problem, and, where necessary, smooth the computed solution. Another strategy is to define a generalised inverse solution. To do this, one defines a norm in the solution space and a norm in the data space. These norms are mathematical expressions characterising the distance between two elements of a Euclidean space [61]. One then defines the generalised inverse solution as the input X of minimum norm for which the associated measurement values specified by the direct problem applied to the solution best approximate the measurement data M actually obtained, i.e., they minimise the norm of the separation between the associated measurement value and the actual measurement value.

Since this solution is generally highly perturbed, one then defines a regularised generalised inverse solution, by adding a penalty term (to the norm to be minimised) that takes into account the regularity of the solution. One then seeks a solution achieving a compromise between regularity and correspondence of the measurement values.

A statistical approach is needed when perturbations or measurement noise become significant. In this case, solutions and measurements are described as stochastic processes [51]. One then seeks the most probable solution in the sense of some given criterion. The most widely used criteria are maximum likelihood, which seeks the solution for which the measurement result corresponds to the most likely output, and Bayesian methods, where the solution chosen is the one that maximises the probability of obtaining a solution conditional on the obtained measurement results [63]. These ideas show in particular that a calculated solution depends on the measurement values, as one would expect, but also on the method used to calculate it. In particular, this calculation allows one to include a priori knowledge of the solution when necessary.

All these ideas are now well known in the fields of physical instrumentation and imaging. Their use in nanobiotechnology is still rather limited. However, it would seem essential in order to obtain reliable quantitative measurement data.

13.5 Data Analysis

The aim of data analysis is to provide the user with tools for describing the measurement data, for interpreting results, and for making decisions [17]. With sophisticated devices such as parallel measurement systems, spectrometers, or imaging systems, the data provided is very rich and hence very complex to analyse. There are then several steps in the analysis:

- selection of relevant measurements,
- identification of statistical parameters in the stochastic laws governing the measurement process,
- identification of geometrical parameters specifying the distribution of the relevant measurements in space or time,
- classification of the measurement values with respect to these parameters to aid in decision-making.

In this section, we describe the main approaches used in each step.

13.5.1 Selecting the Relevant Measurements

The problem here is to select either characteristic points, such as peaks on a spectrum [64] or the maxima and minima on a cyclic voltamperometry curve, or relevant regions, such as a spot on a microarray image [1]. One then distinguishes:

- Manual approaches, in which the user selects these data on a screen via a suitable graphic interface.
- Semi-automated approaches, in which the user first positions points close to the data to be selected, in order to guide the algorithm, which will then identify them automatically.
- Automated approaches, where the relevant data are recognised without the user's intervention.

With the advent of high-throughput measurement devices, the approach must be as fully automated as possible.

A first example is the detection of spectral peaks by looking for the Gaussian curve that best approximates the data. This is achieved, for example, by identifying the points that maximise the correlation with primitives like the Gaussian forms [32].

A second example in imaging is given by recognition algorithms for certain objects [24, 25]. This is relevant when the region to be analysed is specified by the given object. One then proceeds either by a contour approach, in which the algorithm detects the points of the image on the boundary between neighbouring regions, or by a region approach, in which the algorithm divides the image into regions depending on whether or not they belong to the studied objects. These two approaches are complementary. One widely used method for detecting the relevant regions is mathematical morphology [65].

13.5.2 Statistical Analysis

Parameters are calculated from numerical data from the appropriate region.

A first type of parameter describes the probability distribution of an elementary measurement associated with the histogram of values. However, one often accepts simple values such as the mean or variance. A more detailed approach is to use regression to compute the independent factors giving the best description of the measurement data [66].

A second type of parameter describes the distribution function of the measurement data in space or time. In this category, one has in particular the correlation coefficients with the given forms, e.g., Gaussian forms, adaptive filtering with signals of given form, and computation of moments like the center of gravity and moments of inertia if the measurement points are treated as a set of points with mass equal to the measured value. One also has the texture parameters describing the correlations between neighbouring values.

13.5.3 Geometrical Analysis

The problem here is to parametrise the spatial or temporal support of the relevant data.

In imaging, one uses the surface area or perimeter to describe a shape. Other more elaborate parameters characterise the circularity (ratio of the square of the perimeter to the area), or describe boundaries (contours) by Fourier descriptors or fractal dimensions [24].

But one can also parametrise the transformations which deform a known shape into a measured shape. For example, one determines the parameters of an affine transformation, which describes a combination of translations, rotations, and dilations. Another approach is to determine the positions of control points on deformable grids. This is used in medicine and biology when reference images or atlases are available.

As far as the temporal aspect is concerned, one determines the duration of an event, for example.

On a sequence of images, one may for example analyse the speed of displacement of a living cell, either globally, or in a parametrised form, or in the form of a continuous field of displacement vectors.

13.5.4 Classification Methods

The next task is to decide, on the basis of either the selected measurement data, or more often the associated parameters, whether a signal is there or not, or quite generally to classify the data, i.e., attribute them to characteristic classes [17, 67–71]. The first case is a special case with two classes, one determined by the presence of a signal and the other by its absence.

To carry out this classification, one needs class models and classification rules. Such rules can be deterministic. One then characterises each class by

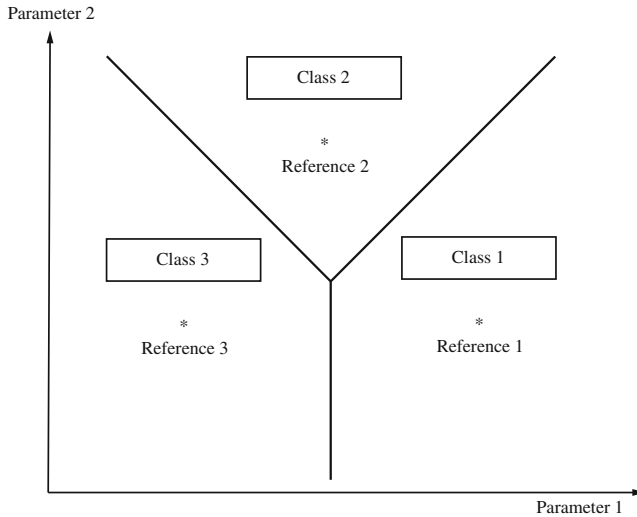


Fig. 13.7. Example of classification. Structure of a two-parameter space with three classes specified by three reference vectors, for a classification criterion based on the minimum distance to the reference vectors

a vector of reference parameters, for example. One defines a distance in the parameter space, e.g., the standard Euclidean distance. The class associated with a vector of reference parameters is then defined as the set of parameters for which the distance to this reference vector is minimum compared with the distances to reference vectors specifying other classes (see Fig. 13.7).

The determination rules can also be statistical. In this case, one defines the probability distribution of a set of parameters when the desired event corresponds to the different vectors of reference parameters. For example, one can use normal distributions. When a measurement is made, one calculates for each class the probability of obtaining the estimated parameters. One then attributes the measurement datum to the class which obtains the maximum probability, i.e., the one maximising the likelihood of this measurement result.

These classification processes are also used to identify outliers. When the distance to the reference vector representing a class is too great, or when the probability of obtaining the result is too low, one concludes that this datum is anomalous. Such a result is then rejected from the analysis process.

To apply this classification process, the classes must be known. To this end, one adopts a learning process. To define this learning process, one must begin with a set of representative trial data. One then distinguishes supervised and unsupervised approaches. In the supervised approach, one knows for each trial datum the name of the class it belongs to. The problem is then to position the boundaries between the parameters, so as to separate these measurement data in the best possible way. In the approaches described above, each class is specified by a vector of reference parameters that must be identified. In

statistical approaches, one also seeks to identify the parameters specifying the probability distributions. In unsupervised approaches, the classes must be inferred. One then applies precise construction rules characterising the parameter grouping. The algorithm seeks to distribute the data in classes that best satisfy these construction criteria. The point about this kind of approach is that it can be fully automated. The disadvantage is that the identified classes do not necessarily correspond to classes of physical objects.

In the connectionist approach [67], one defines a neural network connecting the input parameters and decision-making processes. This is done by a highly parallel automaton. On a formal neuron, the process consists in outputting the result of a threshold function applied to a weighted sum of inputs. This is therefore a complex linear discriminator.

In the structural approach [67], an object is analysed as a set of components and connections. A first approach is to describe the relations between each of these components using a graph structure. The decision-making and classification problem is then solved by graph comparison algorithms. The second approach is to describe complex objects in terms of their components, using a syntax structure defined by the construction rules. These structural approaches provide a way of moving toward a functional analysis of data, or even defining ontologies. The results of data analysis then help us to understand the biological functioning of the observed phenomena.

Acknowledgements

The author would like to thank F. Perraut from the bioMérieux–CEA-LETI joint group for the image of a microarray shown in Fig. 13.4.

References

1. Draghici, S.: *Data Analysis Tools for DNA Microarrays*, Chapman & Hall/CRC Editions (2003)
2. Lockhart, D.J., Winzeler, E.A.: Genomics, gene expression and DNA arrays, *Nature* **405** (15 June), 827–836 (2000)
3. Shih, I.-M., Wang, T.-L.: Apply innovative technologies to explore cancer genome, *Current Opinion in Oncology* **17** (1), 33–38 (2005)
4. Fitch, J.P., Sokhansanj, B.: Genomic engineering: Moving beyond DNA sequence to function, *Proceedings of the IEEE* **88** (12), 1949–1948 (2000)
5. Anderson, N.L., Anderson, N.G.: The human plasma proteome, *Molecular and Cellular Proteomics* **1.11**, 845–867 (2002)
6. Rabilloud, T.: Progrès récents et évolutions en analyse protéomique, *Spectra Analyse* **31** (225), 26–31 (2002)
7. Petricoin, E., Wulfkuhle, J., et al.: Clinical proteomics: Revolutionizing disease detection and patient tailoring therapy, *Journal of Proteome Research* **3**, 209–217 (2004)
8. Petricoin, E., Liotta, L.A.: Proteomic approaches in cancer risk and response assessment, *Trends in Molecular Medicine* **10** (2), 59–64 (2004)

9. Veenstra, T.D., Prieto, D., et al.: Proteomic patterns for early cancer detection, *Drug Discovery Today* **9**, 889–897 (2004)
10. Wiesner, A.: Detection of tumor markers with ProteinChip (R) technology, *Current Pharmaceutical Biotechnology* **5** (1), 45–67 (2004)
11. Steel, L.F., Haab, B.B., et al.: Methods of comparative proteomic profiling for disease diagnostics, *Journal of Chromatography B: Analytical Technologies in Biomedical and Life Sciences* **815** (1–2), 275–284 (2005)
12. Vlahou, A., Fountoulakis, M.: Proteomic approaches in the search for disease biomarkers, *Journal of Chromatography B* **814**, 11–19 (2005)
13. Xiao, Z., Prieto, D., et al.: Proteomic patterns: Their potential for disease diagnosis, *Molecular and Cellular Endocrinology* **230**, 95–106 (2005)
14. Wulfkühle, J., Espina, V., et al.: Genomic and proteomic technologies for individualisation and improvement of cancer treatment, *European Journal of Cancer* **40**, 2623–2632 (2004)
15. Sigworth, F.J., Klemic, K.G.: Microchip technology in ion-channel research, *IEEE Transactions on Nanobioscience* **4** (1), 121–127 (2005)
16. Boguski, M.S., McIntosh, M.W.: Biomedical informatics for proteomics, *Nature* **422** (13 March), 233–237 (2003)
17. Li, L., Tang, H., et al.: Data mining techniques for cancer detection using serum proteomic profiling, *Artificial Intelligence in Medicine* **32**, 71–83 (2004)
18. Penkett, C.J., Bahler, J.: Navigating public microarray databases, *Comparative and Functional Genomics* **5** (6–7), 471–479 (2004)
19. Rajappa, M., Sharma, A., et al.: Bioinformatics and its implications in clinical medicine: A review, *International Medical Journal* **11** (2), 125–129 (2004)
20. Liotta, L.A., Espina, V., et al.: Protein microarrays: Meeting analytical challenges for clinical applications, *Cancer Cell* **3** (April), 317–325 (2003)
21. Rella, R., Spadavecchia, J., et al.: Liquid phase SPR imaging experiments for biosensor applications, *Biosensors and Bioelectronics* **20**, 1140–1148 (2004)
22. Yuk, J.S., Ha, K.S.: Proteomic applications of surface plasmon resonance biosensors: Analysis of protein arrays, *Experimental and Molecular Medicine* **37** (1), 1–10 (2005)
23. Chaurand, P., Schwartz, S.A., et al.: Imaging mass spectrometry: Principles and potentials, *Toxicologic Pathology* **33** (1), 92–101 (2005)
24. Jahne, B.: *Digital Image Processing: Concepts, Algorithms, and Scientific Applications*, Springer, Berlin Heidelberg New York (1997)
25. Maitre, H.: *Le Traitement des Images*, Hermès Science Publications (2003)
26. Grangeat, P.: *La tomographie: Fondements mathématiques, imagerie microscopique et imagerie industrielle*, Hermès Science Publications (2002)
27. Grangeat, P.: *La tomographie médicale: Imagerie morphologique et imagerie fonctionnelle*, Hermès Science Publications (2002)
28. Madou, M.J., Cubicciotti, R.: Scaling issues in chemical and biological sensors, *Proceedings of the IEEE* **91** (6), 830–838 (2003)
29. Diard, J.-P., Le Gorrec, B., et al.: *Cinétique Electrochimique*, Hermann (1996)
30. Chibane, A., Grangeat, P., et al.: Application de modèles à dérivée non entière à la détection électrochimique sur biopuce, 20th Colloque GRETSI sur le Traitement du Signal et des Images, Louvain-la-Neuve (2005)
31. Bergermann, T., Quiaioit, F., et al.: Statistical issues in signal extraction from microarrays, *Proceedings of the SPIE* **4266**, 24–34 (2001)
32. Sharaf, M.A., Illman, D.L., et al.: *Chemometrics*, John Wiley (1986)

33. Zhong, Z., Wang, D., et al.: Nanowire crossbar arrays as address decoders for integrated nanosystems, *Science* **302** (21 November), 1377–1379 (2003)
34. Tavazoie, S., Hughes, J.D., et al.: Systematic determination of genetic network architecture, *Nature Genetics* **22** (July), 281–285 (1999)
35. Brivio, M., Fokkens, R.H., et al.: Integrated micro-fluidic system enabling (bio)chemical reactions with on-line MALDI-TOF mass spectrometry, *Anal. Chem.* **74**, 3972–3976 (2002)
36. Hong, J.W., Quake, S.R.: Integrated nanoliter systems, *Nature Biotechnology* **21** (10), 1179–1183 (2003)
37. Jin, L.J., Ferrance, J., et al.: A microchip-based proteolytic digestion system driven by electroosmotic pumping, *Lab-on-a-Chip* **3**, 11–18 (2003)
38. Puget, P.: Les laboratoires sur puce: Une revue des technologies mises en œuvre, *Spectra Analyse* **32** (233), 29–32 (2003)
39. Sarrut, N., Bouffet, S., et al.: Enzymatic digestion and liquid chromatography in micro-pillar reactors – Hydrodynamic versus electroosmotic flow, *MicroTAS 2004 Conference* (2004)
40. Sauer, S., Lange, B.M.H., et al.: Miniaturization in functional genomics and proteomics, *Nature Reviews: Genetics* **6** (June), 465–476 (2005)
41. Belu, A.M., Graham, D.J., et al.: Time-of-flight secondary ion mass spectrometry: Techniques and applications for the characterization of biomedical surfaces, *Biomaterials* **24**, 3635–3653 (2003)
42. Beebe, K.R., Kowalski, B.R.: An introduction to multivariate calibration and analysis, *Anal. Chem.* **59** (17), 1007A–1016A (1987)
43. Flaten, G.R., Walmsley, A.D.: A design of experiment approach incorporating layered designs for choosing the right calibration model, *Chemometrics and Intelligent Laboratory Systems* **73**, 55–66 (2004)
44. François, N., Govaerts, B., et al.: Optimal designs for inverse prediction in univariate nonlinear calibration models, *Chemometrics and Intelligent Laboratory Systems* **74**, 283–292 (2004)
45. Gonzales, R.C., Wintz, P.: *Digital Image Processing*, Addison-Wesley (1987)
46. Nikolov, S.G., Huter, H., et al.: De-noising of SIMS images via wavelet shrinkage, *Chemometrics and Intelligent Laboratory Systems* **34**, 263–273 (1996)
47. Lukac, R., Plataniotis, K.N., et al.: A multichannel order-statistic technique for cDNA microarray image processing, *IEEE Transactions on Nanobioscience* **3** (4), 272–285 (2004)
48. Walczak, B., Wu, W.: Fuzzy warping of chromatograms, *Chemometrics and Intelligent Laboratory Systems* **77**, 173–180 (2005)
49. Feulade, R.N., Brown, S.D.: An inverse model for target detection, *Chemometrics and Intelligent Laboratory Systems* **77**, 75–84 (2005)
50. Giddings, J.C.: Concepts and comparisons in multi-dimensional separation, *J. of High Resolution Chromatography & Chromatography Communications* **10**, 319–323 (1987)
51. Mohammad-Djafari, A., Giovannelli, J.-F., et al.: Regularization, maximum entropy and probabilistic methods in mass spectrometry data processing problems, *Intl. J. Mass Spectrometry* **215** (1–3), 175–193 (2002)
52. Vandeginste, B.G.M., Kowalski, B.R.: Spatial enhancement and restoration of chemical images from secondary ion mass spectrometry and ion scattering spectrometry, *Anal. Chem.* **55**, 557–564 (1983)
53. Girault, H.H.: *Electrochimie physique et analytique*, Presses Polytechniques et Universitaires Romandes (2001)

54. Drummond, T.G., Hill, M.G., et al.: Electrochemical DNA sensors, *Nature Biotechnology* **21** (10), 1192–1198 (2003)
55. Fabry, P., Fouletier, J.: *Microcapteurs Chimiques et Biologiques – Applications en Milieu Liquide*, Editions Hermès (2003)
56. Mirsky, V.M.: *Ultrathin Electrochemical Chemo- and Biosensors – Technology and Performance*, Springer, Berlin Heidelberg New York (2004)
57. Bonnet, M.: *Problèmes inverses: de l'expérimentation à la modélisation*, Editions TEC & DOC (1999)
58. Ljung, L.: *System Identification – Theory for the User*, Prentice Hall PTR (1999)
59. Landau, I.D., Besançon-Voda, A.: *Identification des Systèmes*, Hermès Science Publications (2001)
60. Dewe, W., Marini, R.D., et al.: Development of response models for optimising HPLC methods, *Chemometrics and Intelligent Laboratory Systems* **74**, 263–268 (2004)
61. Andrews, L.C., Phillips, R.L.: *Mathematical Techniques for Engineers and Scientists*, SPIE Press (2003)
62. Demoment, G., Idier, J., et al.: Problèmes inverses en traitement du signal et de l'image. *Techniques de l'Ingénieur*, Traité TELECOMS, TE 5235-1, TE 5235-25 (2001)
63. Idier, J.: *Approche bayésienne pour les problèmes inverses*, Hermès Science Publications (2001)
64. Yasui, Y., McLerran, D., et al.: An automated peak identification/calibration procedure for high-dimensional protein measures from mass spectrometers, *Journal of Biomedicine & Biotechnology* **4**, 242–248 (2003)
65. Angulo, J., Serra, J.: Automatic analysis of DNA microarray images using mathematical morphology, *Bioinformatics* **19** (5), 553–562 (2003)
66. Tyler, B.: Interpretation of TOF–SIMS images: Multivariate and univariate approaches to image denoising, image segmentation and compound identification, *Applied Surface Science* **203–204**, 825–831 (2003)
67. Belaid, A., Belaid, Y.: *Reconnaissance des Formes: Méthodes et Applications*, InterEditions (1992)
68. Chen, Y., Dougherty, E.R., et al.: Ratio-based decisions and the quantitative analysis of cDNA microarray images, *Journal of Biomedical Optics* **2** (4), 364–374 (1997)
69. Ben-Dor, A., Shamir, R., et al.: Clustering gene expression, *Journal of Computational Biology* **6** (3/4), 281–297 (1999)
70. Brown, M.P.S., Grundy, W.N., et al.: Knowledge-based analysis of microarray gene expression data using support vector machines, *PNAS* **97** (1), 262–267 (2000)
71. Fu, L.M., Youn, E.S.: Improving reliability of gene selection from microarray functional genomics data, *IEEE Transactions on Information Technology in Biomedicine* **7** (3), 191–196 (2003)

Molecular Dynamics. Observing Matter in Motion

C. Chipot

14.1 Introduction

It is particularly important to obtain insights into the structural and dynamical aspects of ordered systems on the atomic level in order to understand the functions of such complex molecular constructions. In many cases, it is impossible to obtain microscopic detail using conventional experimental techniques. However, the genuine explosion of computer resources over the past ten years, together with the development of more effective algorithms, have made it possible to study nanomolecular assemblies of increasing complexity through the methods of theoretical chemistry.

The purpose of this chapter is to examine one aspect of theoretical chemistry, namely, statistical simulations of molecular mechanics. The aim of such simulations is to gain access to the atomic detail of condensed matter through computer experiments. There are many techniques available today to do this, including molecular dynamics, stochastic dynamics and its special cases – e.g., Brownian dynamics or Langevin dynamics – or again Monte Carlo simulations. These different theoretical approaches can be viewed in many ways as a bridge between macroscopic experimental observation and its microscopic counterpart. In the following, we shall be mainly concerned with molecular dynamics [1].

14.1.1 Relating the Microscopic to the Meso- and Macroscopic

In order to correlate the properties of a microscopic system with those of the macroscopic phase, it is essential to remove the problem of edge effects. In practice, periodic boundary conditions are used. This involves replicating the finite set of particles distributed over a box, usually a parallelepiped, in the three space directions (see Fig. 14.1). It turns out that this approach is justified with hindsight by the reliable reproduction of thermodynamic quantities from very small samples. The pseudo-infinite nature of the system generated

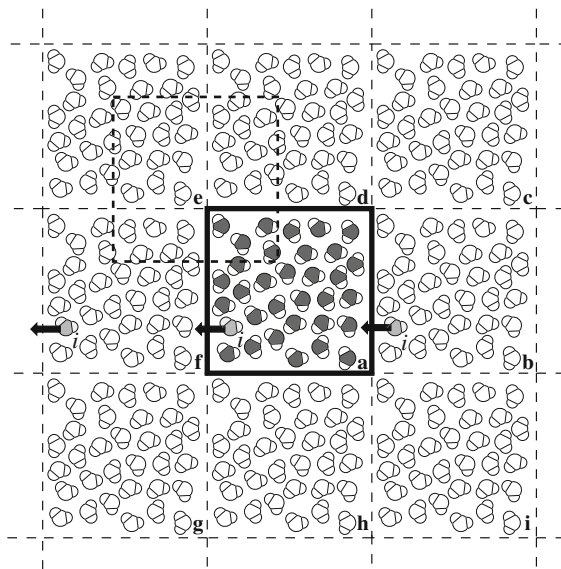


Fig. 14.1. Two-dimensional view of a simulation cell replicated in the three space dimensions. Using periodic boundary conditions, when molecule i leaves the central box (a), its images in the neighbouring ghost cells move in the same way. The dashed box, overlapping cells (a), (d), (e), and (f), symbolises the minimal image convention

here compels one to carry out certain approximations when treating the interactions between molecules [2]. In particular, the so-called minimal image approximation assumes that each particle i in the center cell interacts with the nearest image of all other particles j .

Furthermore, by introducing a cutoff sphere, one can eliminate the interactions beyond some arbitrary distance, defined as being less than or equal to half the smallest side of the simulation cell (see Fig. 14.2). The validity of these approximations will obviously improve as the range of the interactions decreases. Although dispersion and repulsion interactions, which have limited range, do not generally cause any difficulties here, the same cannot be said for certain electrostatic interactions. Since the size of the system generally goes as r^3 , one must assume that the treatment of interactions going as $1/r^n$, for $n < 3$, is likely to lead to error in a model with a cutoff distance. To deal with this situation, it is better to use a lattice summation method, such as the Ewald–Kornfeld summation method [3] or the Ladd method [4], which evaluate interactions of a molecule with all others situated in the central cell, as well as in all image cells. Although these approaches considerably increase the overall cost of the calculation, they are essential for an adequate description of very long range interactions.

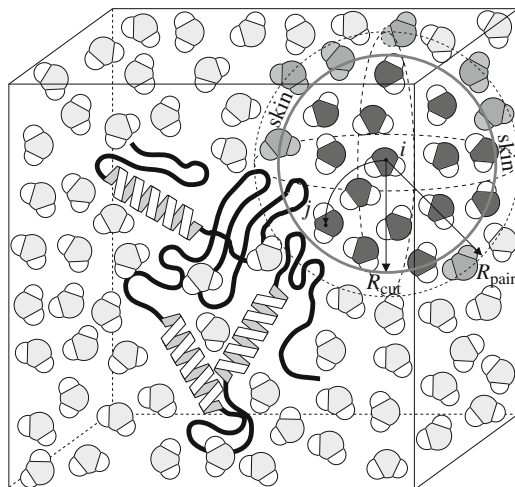


Fig. 14.2. Use of a cutoff sphere of radius R_{cut} to limit the calculation of interactions of particle i with its neighbours in the minimal image convention. A sphere of radius R_{pair} , greater than R_{cut} , is used to establish a list of all the neighbours of i . This list of particle pairs $\{i, j\}$ is periodically updated. The space between the spheres of radii R_{cut} and R_{pair} is called the skin

14.1.2 Legitimacy of Molecular Dynamics Simulations

Can one justify the use of molecular simulations to model condensed matter? Strictly speaking, a complete study of a system as complex as a molecular liquid would require solution of the time-dependent Schrödinger equation for a very large set of electrons and nuclei. Such an idea remains unrealistic, despite recent progress in the field of linear growth calculations, and one is therefore limited to a classical description of the system dynamics. Even in this context, if only for reasons of computation time, quantum molecular simulations are generally limited to particle numbers between a hundred and a few thousand.

The basic idea of molecular dynamics is particularly simple. It consists in generating the trajectories of a finite set of particles by numerical integration of the classical equations of motion. This approach, which appears debatable at the outset, turns out to be justified by two quite remarkable facts:

- By the Born–Oppenheimer approximation, the motion of the electrons can be dissociated from the motion of the nuclei.
- In most cases, the de Broglie wavelength of a particle is significantly shorter than the intermolecular distance, whence quantum effects are globally negligible.

The trajectories determined in this way are used to evaluate static and dynamic properties by time averages, which coincide with statistical averages for the so-called ergodic systems:

$$\lim_{t \rightarrow \infty} \overline{\mathcal{A}}_t = \langle \mathcal{A} \rangle. \quad (14.1)$$

Here, $\mathcal{A}$ refers to some arbitrary observable property, $\overline{\mathcal{A}}_t$ is its time average, and $\langle \mathcal{A} \rangle$ is its statistical ensemble average. In practice, the ergodicity postulate, which assumes that all microstates of the system are visited during a single trajectory, would appear to be satisfied, at least in all simple liquids.

14.2 Basic Principles of Molecular Dynamics

In classical molecular dynamics [1, 2, 5], the trajectories of the various components of the molecular system are generated by integrating Newton's equations of motion. For particle i , these give

$$\begin{cases} m_i \frac{d^2 \mathbf{x}_i(t)}{dt^2} = \mathbf{f}_i(t), \\ \mathbf{f}_i(t) = -\frac{\partial \mathcal{V}(\mathbf{x})}{\partial \mathbf{x}_i(t)}, \end{cases} \quad (14.2)$$

where $\mathcal{V}(\mathbf{x})$ is the potential energy function of the N -particle system, which depends only on the Cartesian coordinates $\{\mathbf{x}_i\}$. Equations (14.2) are integrated numerically with an infinitesimal time step δt , ensuring the conservation of energy of the system, typically $1\text{--}2 \times 10^{-15}$ s (see Fig. 14.3). The choice of δt is closely related to the vibrational frequencies of covalent bonds in the system. Chemical bonds involving hydrogen atoms vibrate with higher frequencies which can only be described with smaller integration intervals.

We shall see in the rest of this chapter that longer integration steps can be used, either by freezing the degrees of freedom associated with high vibrational frequencies by means of constrained molecular dynamics algorithms, or by eliminating these degrees of freedom in the framework of a coarse-grained model.

14.2.1 Validity of Molecular Dynamics Simulations

It would, however, be perfectly unrealistic to expect to generate an exact trajectory over a long time span when Newton's equations of motion are solved numerically, with a finite integration step. But it is not as crucial as one might think at first glance to have an exact solution for (14.2). What matters is that the statistical behaviour of the trajectory should be correct, in order to ensure a sufficiently accurate reproduction of the dynamical and thermodynamical properties of the system. This condition is only satisfied if the integrator used to propagate the motion has the property known as symplecticity [8–10]. A propagator is said to be symplectic if it preserves the metric in phase space.

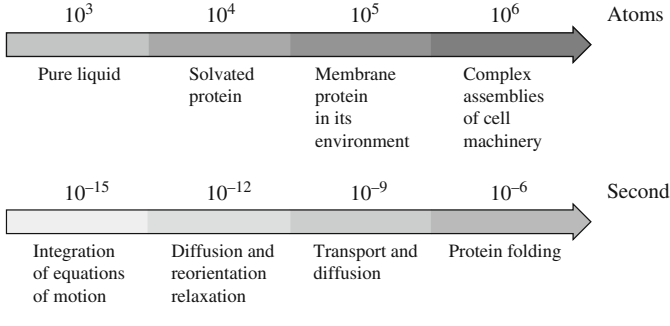


Fig. 14.3. (a) Length and (b) time scales accessible to molecular dynamics simulation. The largest molecular system yet modelled is the movement of transfer RNA into the ribosome, involving 2.64×10^6 atoms [6]. The longest simulation yet achieved in an explicit environment is the folding of a fragment of 47 villin residues, for a duration of 10^{-6} s [7]

It follows that the error associated with this propagator is then necessarily bounded:

$$\lim_{n_{\text{step}} \rightarrow \infty} \frac{1}{n_{\text{step}}} \sum_{k=1}^{n_{\text{step}}} \left| \frac{\mathcal{E}(k\delta t) - \mathcal{E}(0)}{\mathcal{E}(0)} \right| \leq \varepsilon_{\text{MD}}. \quad (14.3)$$

Here, n_{step} is the number of steps in the simulation, $\mathcal{E}(0) \equiv \mathcal{H}(\mathbf{x}, \mathbf{p}_x; 0)$ the initial total energy of the system, and ε_{MD} the upper bound of the energy conservation – 10^{-4} is an acceptable value. Assuming that the integration step is limited, integration of the equations of motion does not give rise to erratic growth in the error for energy conservation, which could significantly affect the statistical behaviour of the molecular dynamics over long times. It is interesting to note that, for a Hamiltonian system, the property of symplecticity implies that the Jacobian

$$\mathbf{J}(\mathbf{\Gamma}_{\delta t}, \mathbf{\Gamma}_0) = \frac{\partial(\mathbf{\Gamma}_{\delta t}^1, \dots, \mathbf{\Gamma}_{\delta t}^N)}{\partial(\mathbf{\Gamma}_0^1, \dots, \mathbf{\Gamma}_0^N)} \quad (14.4)$$

has unit determinant. $\mathbf{\Gamma}_0$ is the initial vector in the N -dimensional phase space, which contains all the position variables $\mathbf{x}$ and momentum variables $\mathbf{p}_x$ describing the system.

14.2.2 Multistep Integration of the Equations of Motion

As mentioned earlier, the long-range nature of charge–dipole interactions, which vary as $1/r^2$, and a fortiori the charge–charge interactions, which go as $1/r$, means that suitable algorithms must be implemented to take such

contributions into account, and this may seriously increase the cost of the simulation. A more formal version of the equations of motion is

$$\boldsymbol{\Gamma}_t = e^{i\mathcal{L}t} \boldsymbol{\Gamma}_0, \quad (14.5)$$

where $\mathcal{L}$ is the Liouville operator which generates a distribution $\varrho(\boldsymbol{\Gamma}, t)$ for a given thermodynamic ensemble according to

$$\frac{\partial \varrho(\boldsymbol{\Gamma}, t)}{\partial t} = -i\mathcal{L}\varrho(\boldsymbol{\Gamma}, t). \quad (14.6)$$

Applying the Trotter formula,

$$e^{i\mathcal{L}\Delta t} = e^{i\mathcal{L}_1\Delta t/2} e^{i\mathcal{L}_2\Delta t} e^{i\mathcal{L}_1\Delta t/2} + \mathcal{O}(\Delta t^3), \quad (14.7)$$

in which $i\mathcal{L} = i\mathcal{L}_1 + i\mathcal{L}_2$, the deconvolution of the short- and long-range contributions is made explicit. In this way, integration steps of different lengths can be used depending on the nature of the relevant interaction. For example, by splitting the total Hamiltonian $\mathcal{H}(\mathbf{x}, \mathbf{p}_x)$ governing the behaviour of the system into a kinetic contribution $\mathcal{T}(\mathbf{p}_x)$ of valence $\mathcal{V}_{\text{valence}}(\mathbf{x})$, a short-range electrostatic contribution $\mathcal{V}_{\text{short}}(\mathbf{x})$, and a long-range electrostatic contribution $\mathcal{V}_{\text{long}}(\mathbf{x})$, it follows that

$$\begin{aligned} e^{i\mathcal{H}(\mathbf{x}, \mathbf{p}_x)\Delta t} &= e^{i\mathcal{V}_{\text{long}}(\mathbf{x})\Delta t/2} \left\{ e^{i\mathcal{V}_{\text{short}}(\mathbf{x})\Delta t/2n} \right. \\ &\quad \times \left[e^{i\mathcal{V}_{\text{valence}}(\mathbf{x})\Delta t/2pn} e^{i\mathcal{T}(\mathbf{p}_x)\Delta t/pn} e^{i\mathcal{V}_{\text{valence}}(\mathbf{x})\Delta t/2pn} \right]^p \\ &\quad \left. \times e^{i\mathcal{V}_{\text{short}}(\mathbf{x})\Delta t/2n} \right\}^n e^{i\mathcal{V}_{\text{long}}(\mathbf{x})\Delta t/2}. \end{aligned} \quad (14.8)$$

This partition of the various contributions to $\mathcal{H}(\mathbf{x}, \mathbf{p}_x)$ is the central idea of so-called multistep integration methods such as *r*-RESPA (reversible reference system propagator algorithm) [11]. It clearly brings out the use of distinct steps to update these contributions, and in this way significantly reduces the computational cost of the simulation.

14.3 Potential Energy Function

This function is the cornerstone of molecular dynamics calculations, because it has the role of reproducing the intra- and intermolecular interactions of the system as faithfully as possible. In principle, this functional can be written as a sum of N terms, viz.,

$$\mathcal{V}(\mathbf{x}) = \sum_i v_1(\mathbf{x}_i) + \sum_i \sum_{j>i} v_2(\mathbf{x}_i, \mathbf{x}_j) + \sum_i \sum_{j>i} \sum_{k>j>i} v_3(\mathbf{x}_i, \mathbf{x}_j, \mathbf{x}_k) + \cdots, \quad (14.9)$$

in which $v_1(\mathbf{x}_i)$, $v_2(\mathbf{x}_i, \mathbf{x}_j)$, and so on, represent the intramolecular potential, the pair interaction potential, etc., respectively. Hence, $\mathcal{V}(\mathbf{x})$ characterises an N -body problem, even though it can be argued that $v_2(\mathbf{x}_i, \mathbf{x}_j)$ is without doubt the most significant term in the intermolecular contribution [2]. This point of view is in fact the starting point for the pair approximation, in which effects of higher order are partially included in an effective potential:

$$\mathcal{V}(\mathbf{x}) \simeq \sum_i v_1(\mathbf{x}_i) + \sum_i \sum_{j>i} v_2^{\text{effective}}(\mathbf{x}_{ij}). \quad (14.10)$$

This approximation is used for most typical force fields, in particular those used to study macromolecular systems, for which calculation time is closely linked to the complexity of $\mathcal{V}(\mathbf{x})$.

14.3.1 Meaning of Different Terms in the Force Field

Among the potential energy functions implementing an effective, pairwise additive description, we have often had recourse to the AMBER series of programs [12, 13]:

$$\begin{aligned} \mathcal{V}(\mathbf{x}) = & \sum_{\text{bonds}} k_r (r - r_0)^2 + \sum_{\text{angles}} k_\theta (\theta - \theta_0)^2 \\ & + \sum_{\substack{\text{dihedral} \\ \text{angles}}} \sum_n \frac{V_n}{2} [1 + \cos(n\phi - \gamma)] \\ & + \frac{1}{k_{\text{vdW}}^{1-4}} \sum_{\substack{i < j \\ \{i, j\} \in 1-4}} \varepsilon_{ij} \left[\left(\frac{R_{ij}^*}{r_{ij}} \right)^{12} - 2 \left(\frac{R_{ij}^*}{r_{ij}} \right)^6 \right] \\ & + \frac{1}{k_{\text{Coulomb}}^{1-4}} \sum_{\substack{i < j \\ \{i, j\} \in 1-4}} \frac{q_i q_j}{4\pi\epsilon_0\epsilon_1 r_{ij}} \\ & + \sum_{\substack{i < j \\ \{i, j\} \in 1-4}} \varepsilon_{ij} \left[\left(\frac{R_{ij}^*}{r_{ij}} \right)^{12} - 2 \left(\frac{R_{ij}^*}{r_{ij}} \right)^6 \right] + \sum_{\substack{i < j \\ \{i, j\} \in 1-4}} \frac{q_i q_j}{4\pi\epsilon_0\epsilon_1 r_{ij}}, \end{aligned} \quad (14.11)$$

in which k_r and r_0 represent the binding force constant and the equilibrium bond length, respectively, k_θ and θ_0 are the angle force constant and the equilibrium valence angle, and $V_n/2$, n , and γ are the torsion barrier, its periodicity, and its phase, respectively. ϵ_0 and ϵ_1 are the permittivity of the vacuum and the relative permittivity, respectively. q_i is the partial charge

carried by the atom i . R_{ij}^* and ε_{ij} are the van der Waals parameters for the pair of atoms $\{i, j\}$, obtained by the Lorentz–Berthelot mixing rules:

$$\begin{cases} \varepsilon_{ij} = \sqrt{\varepsilon_i \varepsilon_j}, \\ R_{ij}^* = R_i^* + R_j^*. \end{cases} \quad (14.12)$$

Given that, during their parametrisation, generally on the basis of sophisticated quantum chemistry calculations, the dihedral terms already contain electrostatic and van der Waals components, most typical force fields make a distinction between interactions of atoms separated by exactly three chemical bonds (these are the so-called 1–4 terms) and all the others, whenever $\{i, j\}$ are not separated by one or two bonds (see Fig. 14.4). The 1–4 contributions are weighted by factors $1/k_{\text{vdW}}^{1-4}$ and $1/k_{\text{Coulomb}}^{1-4}$ as shown in (14.11).

The description of the dihedral angles is a particularly important aspect of potential energy functions. The behaviour of the torsion potential $\mathcal{V}(\phi)$ is often too complex to be faithfully described by including a single term of the Fourier series in (14.11). The case of the phospholipids provides a good illustration of this difficulty. Indeed, the key to accurate reproduction during a statistical simulation of the order parameters S_{CD} of aliphatic chains lies in a good description of the dihedral angles of these chains by an appropriate potential. Only a potential with several terms will be able to account for the subtle *trans-gauche* equilibrium along the chains. The Ryckaert–Bellemans potential [14],

$$\mathcal{V}(\phi) = \sum_{i=1}^6 a_i \cos^i \phi, \quad (14.13)$$

in which the coefficients a_i have been optimised for the internal rotation of n -butane, is usually used to simulate phospholipid bilayers.

14.3.2 Parametrisation of Unbound Atom Terms

The equilibrium between the various terms of the force field is just as subtle. The torsion component is only one aspect of this equilibrium. The choice of the Lennard-Jones parameters R_{ij}^* and ε_{ij} , and the partial charges q_i , is no less critical, to the extent that a judicious choice here determines the accuracy of the calculated thermodynamical quantities. A potential energy function is a construction whose constitutive elements have been calibrated for global reproduction of the key physicochemical properties, although those elements do not necessarily have a physical or chemical meaning.

One popular approach for setting up point charge models is to adjust the electrostatic potential, a genuine fingerprint of the molecule. Considering only a monopole expansion, the optimal set of N_{atoms} net atomic charges $\{q_k\}$ is obtained by minimising the functional

$$f(\{q_k\}) = \sum_{i=1}^{N_{\text{points}}} \left[V^{\text{reference}}(\mathbf{x}_i) - \sum_{j=1}^{N_{\text{atoms}}} \frac{q_j}{r_{ij}} \right]^2, \quad (14.14)$$

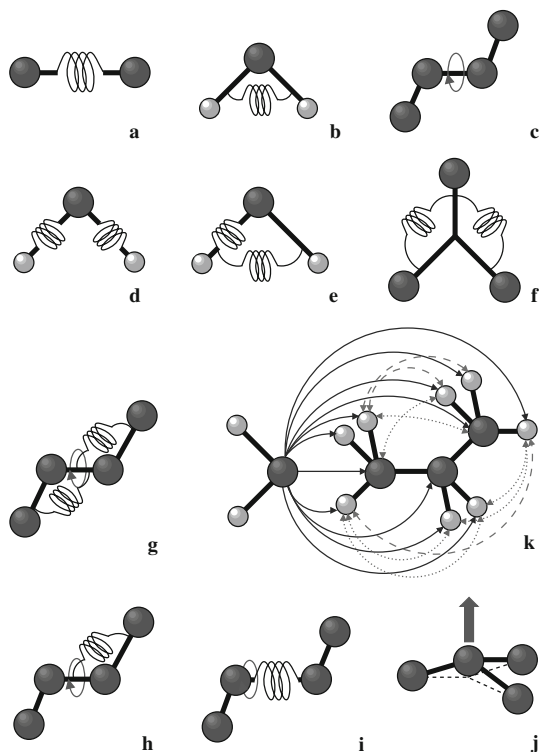


Fig. 14.4. Different terms in an empirical potential energy function. Terms (a)–(j) represent the valence force field, among which (d)–(j) are the so-called cross-terms. (j) is called the out-of-plane term, ensuring that the central atom remains in the plane of the three neighbours to which it is chemically bound. (k) characterises the Coulomb and van der Waals interactions of atoms that are not chemically bound: intermolecular interactions (*continuous curves*), 1–4 intramolecular interactions (*dotted curves*), and > 1–4 intramolecular interactions (*dashed curves*)

where $V^{\text{reference}}(\mathbf{x}_i)$ is the electrostatic potential evaluated at the point $\mathbf{x}_i$ on a grid of N_{points} around the molecule. $V^{\text{reference}}(\mathbf{x}_i)$ is obtained by quantum chemical calculations, from an object of the form $\langle \Psi | 1/|\mathbf{x}_i| | \Psi \rangle$, usually with a high level of sophistication.

Determining the Lennard-Jones parameters often turns out to be a delicate problem. One approach is to fit the repulsion and dispersion contributions from a large number of very precise quantum chemistry calculations carried out for different configurations. This approach is generally applicable to small systems, e.g., the formamide–water heterodimer, for which one seeks to specify the atom–atom van der Waals interaction potential. As an alternative to quantum chemistry calculations, a more heuristic approach uses statistical simulations. Starting from a given set of Lennard-Jones parameters $\{R_{ij}^*, \varepsilon_{ij}\}$,

can one quantitatively reproduce the fundamental thermodynamical quantities associated with a molecular liquid, such as its density ρ , vaporisation enthalpy ΔH_{vap} , or even its diffusion coefficient D ?

14.3.3 Beyond the Usual Force Fields

In many cases, the minimalist description imposed by the functional (14.11) may be inadequate. As with most typical force fields, this functional has a multipurpose aspect, even though it was originally designed for the study of biopolymers and more specifically the study of proteins and nucleic acids.

Classical Description of the Chemical Bond

A more detailed investigation of small organic molecules, often required in the emerging area of de novo drug design, involves a level of calculation that goes beyond the simplistic hypotheses of (14.11). One example is the replacement of the harmonic bond stretching term (see Fig. 14.4a) by a Morse dissociation potential:

$$\mathcal{V}(r) = D_0 \left[e^{-\alpha(r-r_0)} - 1 \right]^2, \quad (14.15)$$

in which D_0 is precisely the dissociation energy and r_0 the equilibrium bond length. The harmonic description itself may appear unsuitable as soon as anharmonic effects are no longer negligible. Cubic and quartic corrections may also prove necessary for both bond stretching and valence angle bending:

$$\mathcal{V}(r) = k_r(r-r_0)^2 \left[1 - k'_r(r-r_0) + k''_r(r-r_0)^2 \right], \quad (14.16)$$

where k'_r and k''_r are the force constants for the cubic and quartic contributions.

Coupling Between Chemical Bond and Valence Angle

When the valence angle widens, bond lengths are reduced, something which is clearly not represented in (14.11). To correct for this, the potential energy function can be enriched with so-called coupling terms, e.g.,

$$\mathcal{V}(r-\theta) = k_{r\theta}(r-r_0)(\theta-\theta_0). \quad (14.17)$$

Apart from the stretch-bend effect characterised here, other coupling terms can be introduced in $\mathcal{V}(\mathbf{x})$, as shown in Figs. 14.4e-i. Apart from complicating the calculation of the potential energy function at each step δt in the integration of (14.2), the introduction of coupling terms also complicates parametrisation.

Beyond a Simple Set of Point Charges

Electrostatic effects are the most problematic, and for several reasons. To begin with, the monopole approximation of the relation (14.14) is not necessarily sufficient for a given molecule. It may then be desirable to include permanent dipoles in the simple model of charges centered on the atoms. More crucial are the induction effects, obviously absent from the pairwise additive approximation (14.10). The number of available solutions to get round this difficulty remains limited. The least costly approach, still very widely used to simulate macromolecular systems, is to increase the net atomic charges artificially, in such a way that these reproduce a characteristic permanent dipole moment, not of the gaseous phase, but of a polar environment. This leads to an ‘average’ representation of the polarisation effects. A more rigorous approach and indeed the choice solution, although much more time-consuming, is to introduce polarisability contributions explicitly in the force field [15]. In this case, the total electrostatic contribution is then

$$\mathcal{V}_{\text{elec}}(\mathbf{x}) = \frac{1}{2} \sum_i q_i V_i. \quad (14.18)$$

Here, the electrostatic potential V_i arises from two sources: on the one hand, the point charges in the system, and on the other, the multipole moments induced at site i . Restricting here to the induced dipole moment $\boldsymbol{\mu}$, linearly related to the electric field $\mathbf{E}_i$, created at point i by all the other polarisable sites $j \neq i$, viz., $\boldsymbol{\mu} = \alpha_i \mathbf{E}_i$, the potential can be written in the form

$$V_i = \sum_{j \neq i} \left(\frac{q_j}{4\pi\epsilon_0 r_{ij}} + \frac{\mathbf{r}_{ij} \cdot \boldsymbol{\mu}_j}{4\pi\epsilon_0 r_{ij}^3} \right). \quad (14.19)$$

Even using the induced moments $\boldsymbol{\mu}_i$ obtained at time t as the starting point at time $t + \delta t$, the convergence of these moments during the molecular dynamics simulation of a polarisable liquid increases the calculation time by a factor of at best two, compared with a simulation using the pairwise additive approximation [16].

All-Atom Versus Coarse-Grained Models

At the opposite extreme to all-atom simulations with explicit allowance for induction effects, cruder approaches have been developed to push back the conventional limits of molecular dynamics, in terms of both system size and time scale (see Fig. 14.3). So-called coarse-grained descriptions take steps in both these directions by significantly reducing the number of particles in the system and allowing the use of a much bigger integration step, typically $\delta t = 40$ fs, by eliminating the harder degrees of freedom.

In the particular case shown in Fig. 14.5, the ratio between the number of atoms in the all-atom and unified-atom or coarse-grained descriptions is

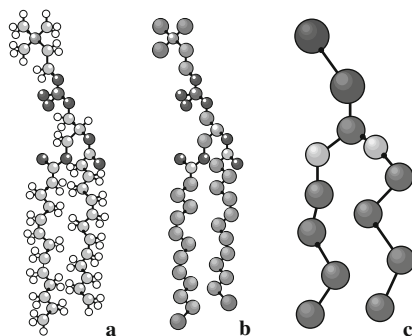


Fig. 14.5. An all-atom description (a), compared with a unified-atom description (b) and a coarse-grained description (c) of a lipid unit. Whereas in the unified atom model, the methylene groups $-\text{CH}_2-$ and methyl groups $-\text{CH}_3$ are treated as van der Waals spheres, several methylene and methyl groups comprise a van der Waals sphere in the coarse-grained description

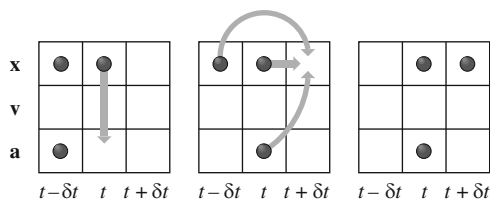


Fig. 14.6. Using the Verlet algorithm to integrate the equations of motion. Note that this scheme does not explicitly involve the velocities of the particles in the system. Starting with the triplet $\{\mathbf{x}_i(t), \mathbf{x}_i(t - \delta t), \mathbf{a}_i(t)\}$ on the left, the acceleration at t is found and used to obtain the position at $t + \delta t$ on the right

of the order of 2 or 9, respectively. The interaction potentials used are often significantly different from those in the all-atom or unified-atom models. Although the electrostatic term is generally Coulomb, or a point dipole interaction, the van der Waals term can be specified by modified Lennard-Jones potentials, e.g., 6–9, or by a Gay–Berne potential [17]. It is interesting to note that, despite their very crude nature, coarse-grained models are able to reproduce certain structural physical properties of the system in a semi-quantitative way. This is the case, for example, for the atomic density profiles of lipid bilayers [18, 19].

14.4 Integrating the Equations of Motion

14.4.1 Molecular Dynamics Integrators

There are several methods for numerical integration of Newton’s equations of motion (14.2). We shall consider three such methods here. The simplest is

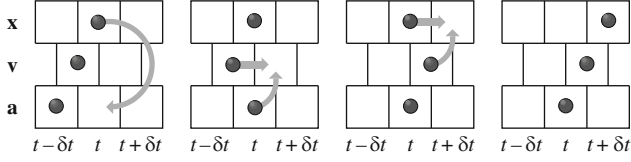


Fig. 14.7. Using the leapfrog algorithm to integrate the equations of motion. Starting from the triplet $\{\mathbf{x}_i(t), \mathbf{v}_i(t - \delta t/2), \mathbf{a}_i(t)\}$ on the left, the acceleration at t and the velocity at $t + \delta t/2$ are deduced and used to find the position at $t + \delta t$ on the right

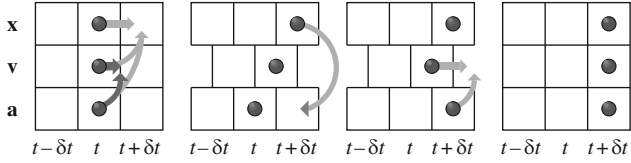


Fig. 14.8. Integrating the equations of motion using the velocity-Verlet algorithm. Generalising the scheme described in Fig. 14.6 to the velocities of the particles, one can obtain the position, velocity, and acceleration at $t + \delta t$

without doubt the Verlet algorithm, which uses the triplet

$$\{\mathbf{x}_i(t), \mathbf{x}_i(t - \delta t), \mathbf{a}_i(t)\},$$

where

$$\mathbf{a}_i(t) = \ddot{\mathbf{x}}_i(t) = \frac{d^2 \mathbf{x}_i(t)}{dt^2} = \frac{\mathbf{f}_i(t)}{m_i}$$

is the acceleration of particle i (see Fig. 14.6) [20]. The change in the positions of the particles is obtained by a Taylor expansion of the position in $t - \delta t$ and $t + \delta t$, leading to

$$\mathbf{x}_i(t + \delta t) = 2\mathbf{x}_i(t) - \mathbf{x}_i(t - \delta t) + \mathbf{a}_i(t)\delta t^2, \quad (14.20)$$

which involves a possible error of $\mathcal{O}(\delta t^4)$. Note that the velocities

$$\mathbf{v}_i(t) = \dot{\mathbf{x}}_i(t) = \frac{d\mathbf{x}_i(t)}{dt},$$

do not appear explicitly in this formula. They are eliminated during the Taylor expansion of $\mathbf{x}_i(t + \delta t)$ and $\mathbf{x}_i(t - \delta t)$. Although not needed for the description of the trajectories, it is essential to calculate them in order to evaluate the kinetic energy $\mathcal{T}(\mathbf{p}_x)$, which depends only on the momentum $\mathbf{p}_x$, and hence the total energy $\mathcal{E} \equiv \mathcal{H}(\mathbf{x}, \mathbf{p}_x)$ of the system. One uses

$$\mathbf{v}_i(t) = \frac{\mathbf{x}_i(t + \delta t) - \mathbf{x}_i(t - \delta t)}{2\delta t}, \quad (14.21)$$

leading to errors of $\mathcal{O}(\delta t^2)$ at each integration step.

Exercise. From the Taylor expansion of the position $\mathbf{x}$ at $t - \delta t$ and $t + \delta t$, obtain the result (14.20). Deduce that the associated error is $\mathcal{O}(\delta t^4)$.

The leapfrog algorithm (see Fig. 14.7), derived from the last, uses the triplet $\{\mathbf{x}_i(t), \mathbf{v}_i(t - \delta t/2), \mathbf{a}_i(t)\}$. The name ‘leapfrog’ is fairly clear from the following expression for the algorithm:

$$\begin{cases} \mathbf{x}_i(t + \delta t) = \mathbf{x}_i(t) + \mathbf{v}_i\left(t + \frac{\delta t}{2}\right) \delta t, \\ \mathbf{v}_i\left(t + \frac{\delta t}{2}\right) = \mathbf{v}_i\left(t - \frac{\delta t}{2}\right) + \mathbf{a}_i(t) \delta t. \end{cases} \quad (14.22)$$

In practice, the first step is to calculate $\mathbf{v}_i(t + \delta t/2)$, from which one can deduce $\mathbf{v}_i(t)$ using

$$\mathbf{v}_i(t) = \frac{\mathbf{v}_i(t + \delta t/2) + \mathbf{v}_i(t - \delta t/2)}{2}, \quad (14.23)$$

this being needed to evaluate the kinetic term $\mathcal{T}(\mathbf{p}_x)$.

Finally, the velocity-Verlet algorithm (see Fig. 14.8) corrects for the main weak point of the Verlet and leapfrog algorithms, viz., the definition of the velocities, with associated error of $\mathcal{O}(\delta t^2)$. The velocities are explicitly included in the Verlet algorithm by writing

$$\begin{cases} \mathbf{x}_i(t + \delta t) = \mathbf{x}_i(t) + \mathbf{v}_i(t) \delta t + \frac{1}{2} \mathbf{a}_i(t) \delta t^2, \\ \mathbf{v}_i(t + \delta t) = \mathbf{v}_i(t) + \frac{\mathbf{a}_i(t) + \mathbf{a}_i(t + \delta t)}{2} \delta t. \end{cases} \quad (14.24)$$

This scheme involves the following two steps:

$$\mathbf{v}_i\left(t + \frac{\delta t}{2}\right) = \mathbf{v}_i(t) + \frac{1}{2} \mathbf{a}_i(t) \delta t, \quad (14.25)$$

from which the thermodynamic forces $\mathbf{f}_i$, and also the accelerations $\mathbf{a}_i$, can be evaluated at time $t + \delta t$. Then

$$\mathbf{v}_i(t + \delta t) = \mathbf{v}_i\left(t + \frac{\delta t}{2}\right) + \frac{1}{2} \mathbf{a}_i(t + \delta t) \delta t, \quad (14.26)$$

from which the kinetic energy can be deduced at time $t + \delta t$, while the potential energy $\mathcal{V}(\mathbf{x})$ at the same instant is calculated in the force loop.

14.4.2 Integration with Constraints

In some circumstances, it may be useful to freeze certain degrees of freedom during the molecular dynamics simulation by introducing holonomic constraints. These eliminate the hard degrees of freedom of the system, corresponding to the highest vibration frequencies, in particular the vibrations of covalent bonds involving hydrogen atoms. By holding the length of these bonds at their nominal value, the integration step δt for the equations of motion can be significantly increased without jeopardising conservation of the total energy of the system.

Freezing degrees of freedom in the molecular dynamics context amounts to solving the equations of motion with constraints. When the length of a chemical bond is frozen, the constraint can be written in the form

$$\chi_{ij}(t) = |\mathbf{x}_j(t) - \mathbf{x}_i(t)|^2 - d_{ij}^2, \quad (14.27)$$

where d_{ij} is the equilibrium length of the chemical bond. Apart from the force $\mathbf{f}_i$ due to intra- and intermolecular interactions, a constraint force $\mathbf{g}_i$ then appears in the equations of motion:

$$m_i \frac{d^2 \mathbf{x}_i(t)}{dt^2} = \mathbf{f}_i + \mathbf{g}_i. \quad (14.28)$$

This constraint force is defined by

$$\mathbf{g}_i = - \sum_j \lambda_{ij}(t) \nabla_i \chi_{ij}(t) = -2 \sum_j \lambda_{ij}(t) \mathbf{x}_{ij}(t), \quad (14.29)$$

where $\lambda_{ij}(t)$ is the Lagrange multiplier associated with the constraint along the chemical bond between atoms i and j . Combined with the Verlet integrator (14.20), the equations of motion take the form

$$\mathbf{x}_i(t + \delta t) = \mathbf{x}_i(t) + \delta t \mathbf{v}_i(t) + \frac{\delta t^2}{2m_i} [\mathbf{f}_i(t) + \mathbf{g}_i(t)]. \quad (14.30)$$

The constrained equations of motion are usually solved by following an iterative Gauss–Seidel scheme, i.e., solving the equations of a linear system one by one, until each holonomic constraint is satisfied.

Exercise. Establish the constrained equations of motion for a three-atom molecule, in which the bond lengths d_{21} and d_{23} are frozen and the valence angle $\theta(1, 2, 3)$ is allowed to vary through the intramolecular potential.

During the simultaneous integration of Newton's equations of motion, the total energy of the system is conserved, and if the volume is held constant, the simulation will generate a microcanonical statistical ensemble, i.e., $(N, V, \mathcal{E})$. However, this situation is not always appropriate, and it may be desirable to carry out simulations in which the temperature or the pressure are independent quantities, rather than derived properties (see Table 14.1).

Table 14.1. Examples of thermodynamic ensembles accessible to molecular dynamics simulations. U is the internal energy of the system, S its entropy, and A its Helmholtz free energy. $P_\perp$ and $P_\parallel$ denote the normal and lateral pressures, respectively, applied to the simulation cell, and $\mathbf{p}_i^2 = p_{x_i}^2 + p_{y_i}^2 + p_{z_i}^2$

Ensemble	Free energy	Conserved Hamiltonian $\mathcal{H}(\mathbf{x}, \mathbf{p}_x)$
$(N, V, \mathcal{E})$		$\mathcal{H} = U$
(N, V, T)	$A = U - TS$	$\mathcal{H} \simeq \sum_i \frac{\mathbf{p}_i^2}{2m_i} + \mathcal{V}(\mathbf{x})$
$(N, P_\perp, \mathcal{A}, T)$	$A = U - TS + P_\perp V$	$\mathcal{H} \simeq \sum_i \frac{\mathbf{p}_i^2}{2m_i} + \mathcal{V}(\mathbf{x}) + P_\perp V$
$(N, P_\parallel, \mathcal{A}, T)$	$A = U - TS + P_\parallel V$	$\mathcal{H} \simeq \sum_i \frac{\mathbf{p}_i^2}{2m_i} + \mathcal{V}(\mathbf{x}) + P_\parallel V$
(N, V, γ, T)	$A = U - TS - \gamma \mathcal{A}$	$\mathcal{H} \simeq \sum_i \frac{\mathbf{p}_i^2}{2m_i} + \mathcal{V}(\mathbf{x}) - \gamma \mathcal{A}$
$(N, P_\perp, \gamma, T)$	$A = U - TS + P_\perp V - \gamma \mathcal{A}$	$\mathcal{H} \simeq \sum_i \frac{\mathbf{p}_i^2}{2m_i} + \mathcal{V}(\mathbf{x}) + P_\perp V - \gamma \mathcal{A}$

14.4.3 Molecular Dynamics at Constant Temperature

Several methods have been proposed for carrying out isothermal molecular dynamics simulations, with varying degrees of sophistication. The simplest is undoubtedly the periodic recalibration of the velocities by a factor of $\sqrt{T/T_T}$, where T_T is the instantaneous kinetic temperature, i.e., $2T(\mathbf{p}_x)/3Nk_B$, with k_B Boltzmann's constant, and T the desired temperature. However, applying this factor at each integration step does not strictly speaking lead to Newtonian molecular dynamics. Newtonian mechanics implies conservation of energy and momentum. Molecular dynamics at constant kinetic temperature involves solving the constrained equations of motion [2]

$$\begin{cases} \dot{\mathbf{x}}_i = \frac{\mathbf{p}_{x,i}}{m_i}, \\ \dot{\mathbf{p}}_{x,i} = \mathbf{f}_i - \xi(\mathbf{x}; \mathbf{p}_x) \mathbf{p}_x, \end{cases} \quad (14.31)$$

in which $\xi(\mathbf{x}; \mathbf{p}_x)$ can be treated as a coefficient of friction ensuring that $\dot{T}_T = 0$. This constraint is chosen in such a way as to perturb the Newtonian trajectory as little as possible:

$$\xi(\mathbf{x}; \mathbf{p}_x) = \frac{\sum_i \mathbf{p}_{x,i} \cdot \mathbf{f}_i}{\sum_i |\mathbf{p}_{x,i}|^2}. \quad (14.32)$$

A second, more rigorous approach consists in introducing an extra degree of freedom s into the equations of motion. The velocity of particle i is then $\mathbf{v}_i = s\dot{\mathbf{x}}_i = s\mathbf{p}_{x,i}/m_i$. Potential and kinetic terms are associated with the degree of freedom s , which is thought of as a thermostat:

$$\begin{cases} \mathcal{V}_s = \frac{(f+1) \ln s}{\beta}, \\ \mathcal{T}_s = \frac{1}{2} \mathcal{Q} \dot{s}^2, \end{cases} \quad (14.33)$$

where $\mathcal{Q}$ is the thermal inertia parameter, which controls temperature fluctuations, f is the number of degrees of freedom of the system ($3N - 3$ if the total momentum $\mathbf{p}$ is constant), and $\beta \equiv 1/k_B T$. This kind of approach is known generically as an extended Lagrangian method, in the sense that the Lagrangian now has the form

$$\mathcal{L}_s(\mathbf{x}; \mathbf{p}_x) = \mathcal{T}(\mathbf{p}_x) + \mathcal{T}_s(\mathbf{p}_x) - \mathcal{V}(\mathbf{x}) - \mathcal{V}_s(\mathbf{x}).$$

The equations of motion can now be re-established in the form

$$\begin{cases} \ddot{\mathbf{x}}_i = \frac{\mathbf{f}_i}{m_i s^2} - 2 \frac{\dot{s} \dot{\mathbf{x}}}{s}, \\ \mathcal{Q} \ddot{s} = \sum_i m_i \dot{\mathbf{x}}_i^2 s - \frac{f+1}{\beta s}. \end{cases} \quad (14.34)$$

This formalism, devised by Nosé [22], was revisited by Hoover [23], who got rid of the time-dependent parameter s . In the constrained equations of motion (14.36), the friction term is now given by a first order differential equation:

$$\dot{\xi} = \frac{f}{\mathcal{Q}} k_B (T_{\mathcal{T}} - T). \quad (14.35)$$

The conserved quantity here is the Hamiltonian

$$\mathcal{H}_s(\mathbf{x}; \mathbf{p}_x) = \mathcal{T}(\mathbf{p}_x) + \mathcal{T}_s(\mathbf{p}_x) + \mathcal{V}(\mathbf{x}) + \mathcal{V}_s(\mathbf{x}).$$

The resulting Nosé–Hoover equations can be written in the form

$$\begin{cases} \dot{\mathbf{x}}_i = \frac{\mathbf{p}_{x,i}}{m_i}, \\ \dot{\mathbf{p}}_{x,i} = \mathbf{f}_i - \xi(\mathbf{x}; \mathbf{p}_x) \mathbf{p}_x, \\ \dot{\xi} = \frac{f}{\mathcal{Q}} k_B (T_{\mathcal{T}} - T). \end{cases} \quad (14.36)$$

As suggested by Fig. 14.9, the chaos generated by the Nosé–Hoover algorithm is inadequate to describe the canonical distribution of small systems involving few degrees of freedom. An alternative approach, including several coupled thermostats, like the one proposed by Martyna et al. [21], is preferable in this kind of situation. However, in most cases, the increase in the number of degrees of freedom produces enough chaos to ensure that the system behaves ergodically, or quasi-ergodically.

The last approach is the weak coupling method [24], which consists in relaxing the instantaneous kinetic temperature $T_{\mathcal{T}}(t)$ to a reference value T :

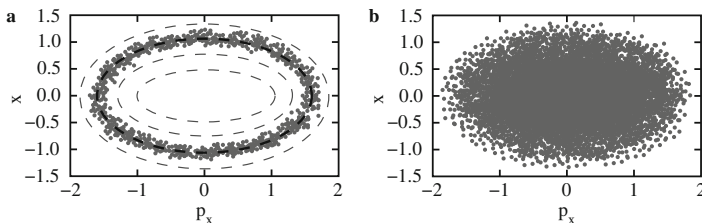


Fig. 14.9. Molecular dynamics at constant temperature in the simplistic case of a 1D harmonic oscillator. In position and momentum space $\{x, p_x\}$, the system evolves on constant-energy orbits. When the number of degrees of freedom is very small, the sampling generated by the Nosé–Hoover algorithm is high non-ergodic (a). The chaotic aspect of the simulation can be recovered, at least in part, by the thermostat chain variant of the Nosé–Hoover algorithm (b) [21]

$$\frac{dT_T(t)}{dt} = \frac{T - T_T(t)}{\tau_T}, \quad (14.37)$$

where τ_T is precisely the relaxation time associated with temperature fluctuations. The kinetic energy is modified by an amount $\Delta\mathcal{T}$ given by

$$\Delta\mathcal{T} = \frac{1}{2}(\chi^2 - 1)Nk_B T_T(t), \quad (14.38)$$

during an integration step δt , by readjusting the velocities by a factor χ given by

$$\chi = \left\{ 1 + \frac{\delta t}{\tau_T} \left[\frac{T}{T_T(t)} - 1 \right] \right\}^{1/2}. \quad (14.39)$$

This aperiodic coupling to a ‘heat reservoir’ by virtue of a first order process does not lead to an oscillating response to temperature changes. On the other hand, it has been shown that this algorithm does not strictly lead to a canonical distribution, in contrast to the Nosé–Hoover method.

Although somewhat on the fringe of the isothermal molecular dynamics algorithms, Langevin dynamics also provides a way of controlling the temperature in statistical simulations. The basic idea here is to include a frictional force, proportional to the velocity, along with the conservative forces. The friction cancels the kinetic energy of the system. A random force $f(t)$ is then introduced to allow the system to diffuse:

$$m_i \frac{d^2 \mathbf{x}_i}{dt^2} = -\frac{d\mathcal{V}(\mathbf{x})}{d\mathbf{x}} - \gamma m_i \frac{d\mathbf{x}_i}{dt} + f(t), \quad (14.40)$$

where γ is a scalar parameter controlling the friction, $f(t)$ white Gaussian noise with zero average, i.e., $\langle f(t) \rangle = 0$, obeying the fluctuation–dissipation theorem [25], i.e.,

$$\langle f(t)f(t') \rangle = \frac{2m_i\gamma\delta(t-t')}{\beta}, \quad (14.41)$$

which is the condition for generating a canonical distribution.

14.4.4 Molecular Dynamics at Constant Pressure

Once again, there are several methods for holding the simulation cell at constant pressure, involving various levels of sophistication. Indeed, it may be desirable in some situations to generate trajectories in the isobaric–isothermal ensemble (N, P, T) .

Just as for the problem of holding the temperature at a constant value, the extended Lagrangian formalism can be applied to the pressure. Initially proposed by Andersen [26], this method involves coupling the system to an external variable V characterising the volume of the simulation cell. This coupling represents the action that would be exerted on the system by a piston. Kinetic and potential terms are associated with it:

$$\begin{cases} \mathcal{V}_V = \frac{1}{2} m_P \dot{V}^2, \\ \mathcal{T}_V = PV, \end{cases} \quad (14.42)$$

where m_P can be thought of as the mass of the piston, and P represents the required pressure. By rescaling the position and velocity variables, $\mathbf{x}$ and $\mathbf{v}$, in the form $\mathbf{s} = \mathbf{x}/V^{1/3}$ and $\dot{\mathbf{s}} = \mathbf{v}/V^{1/3}$, respectively, the kinetic and potential energies take the form

$$\mathcal{V}(\mathbf{x}) \equiv \mathcal{V}(V^{1/3} \mathbf{s}), \quad \mathcal{T}(\mathbf{p}_x) = \frac{1}{2} m V^{2/3} \sum_i \dot{\mathbf{s}}_i^2. \quad (14.43)$$

From the Lagrangian

$$\mathcal{L}_V(\mathbf{x}; \mathbf{p}_x) = \mathcal{T}(\mathbf{p}_x) + \mathcal{T}_V(\mathbf{p}_x) - \mathcal{V}(\mathbf{x}) - \mathcal{V}_V(\mathbf{x}),$$

one can now establish the new equations of motion due to Andersen [26]:

$$\begin{cases} \ddot{\mathbf{s}}_i = \frac{\mathbf{f}_i}{m_i V^{1/3}} - \frac{2}{3} \frac{\dot{\mathbf{s}}_i \dot{V}}{V}, \\ \ddot{V} = \frac{P_P - P}{m_P}, \end{cases} \quad (14.44)$$

in which the force $\mathbf{f}_i$, and the instantaneous pressure derived from the virial expansion, viz.,

$$P_P = \frac{1}{V} \left(\frac{N}{\beta} - \frac{1}{2} \sum_i \mathbf{x}_i \cdot \mathbf{f}_i \right), \quad (14.45)$$

are evaluated from the unscaled coordinates and momenta. The quantity that is conserved during the molecular dynamics simulation is the Hamiltonian of the extended system:

$$\mathcal{H}_V(\mathbf{x}; \mathbf{p}_x) = \mathcal{T}(\mathbf{p}_x) + \mathcal{T}_V(\mathbf{p}_x) + \mathcal{V}(\mathbf{x}) + \mathcal{V}_V(\mathbf{x}),$$

i.e., its enthalpy, plus a kinetic contribution $k_B T/2$ due to fluctuations in the volume of the cell. It should be stressed that, formally, this algorithm generates an isobaric-isenthalpic distribution (N, P, H) . One can obtain an isobaric-isothermal ensemble by coupling to a thermostat, such as the one specified by (14.34).

It has been observed that this approach leads to oscillations in $P_{\mathcal{P}}$, depending on the mass $m_{\mathcal{P}}$ of the piston. Feller et al. [27] suggested a scheme for eliminating this unwanted effect by damping the degree of freedom of the piston via the Langevin equation. Going back to (14.44), one has

$$\begin{cases} \ddot{\mathbf{s}}_i = \frac{\mathbf{f}_i}{m_i V^{1/3}} - \frac{2}{3} \frac{\dot{\mathbf{s}}_i \dot{V}}{V}, \\ \ddot{V} = \frac{P_{\mathcal{P}} - P}{m_{\mathcal{P}}} - \gamma \dot{V} + R(t), \end{cases} \quad (14.46)$$

where γ is the collision frequency, and $R(t)$ is a random force, drawn from a Gaussian distribution with zero average. It is interesting to note that $R(t)$ satisfies the fluctuation-dissipation theorem:

$$\langle R(t_1) R(t_2) \rangle = \frac{1}{m_{\mathcal{P}}} \frac{2}{\beta} \kappa(t_1 - t_2), \quad (14.47)$$

in which $\kappa(t)$ is a damping factor.

A second approach put forward by Berendsen et al. [24], is an extension of the weak-coupling method (described above) to the constant pressure simulation. Just as for the constant temperature algorithm, the equations of motion are modified after relaxing the instantaneous pressure $P_{\mathcal{P}}(t)$ to a reference value P , according to

$$\frac{dP_{\mathcal{P}}(t)}{dt} = \frac{P - P_{\mathcal{P}}(t)}{\tau_{\mathcal{P}}}, \quad (14.48)$$

where $\tau_{\mathcal{P}}$ is the relaxation time associated with pressure fluctuations. By readjusting the atomic coordinates and the size of the periodic cell by a factor ς , the total volume is changed by $\Delta V = (\varsigma^3 - 1)V$, and this of course leads to a change in pressure, given by

$$\Delta P = \frac{\Delta V}{\beta_T V}, \quad (14.49)$$

in which β_T is the isothermal compressibility. By solving (14.48) and (14.49) for a given value of ς , it follows that

$$\varsigma = \left[1 - \beta_T \delta t \frac{P - P_{\mathcal{P}}(t)}{\tau_{\mathcal{P}}} \right]^{1/3}. \quad (14.50)$$

As for the temperature, this algorithm does not lead to a clearly defined thermodynamic ensemble.

14.5 Rigorous Treatment of Electrostatic Interactions

A particularly crucial feature of molecular dynamics simulations is an appropriate treatment of electrostatic interactions. For obvious reasons of cost, the spherical cutoff method discussed earlier is still very widely used, especially when the time scales explored exceed the nanosecond range, as happens for example when simulating phospholipid bilayers. Although the long-range aspect of the dipole–dipole interactions, going as $1/r^3$, is sufficiently limited to ensure a satisfactory reproduction of the structural properties and statistical ensemble averages of the system, the influence of these interactions on complex processes like protein folding requires more careful consideration.

The presence of ionic species is distinctly more problematic, because the cutoff sphere induces a number of artifacts that significantly distort the results of simulations. It should also be stressed that, even considering only dipolar species, the spherical cutoff produces a singularity in the derivatives of the potential energy at the boundary of the cutoff region. The harmful effects of this spherical cutoff can be lessened by including a so-called switching function [2], replacing the discontinuous behaviour of the Heaviside step function by a less abrupt decrease in the vicinity of the sphere. However, this approach does nothing to solve the problems introduced by the presence of ions in the system.

A more rigorous and in principle less costly method for handling charge–charge and charge–dipole interactions exploits the Debye–Hückel theory and solves the linearised Poisson–Boltzmann equation [28]. As shown in Fig. 14.11, this method, known as the generalised reaction field, does not fully eliminate the artifacts when $r \rightarrow R_{\text{cutoff}}$, the cutoff radius.

The most rigorous method is undoubtedly the one proposed by Ewald [3]. Starting from the observation that the Coulomb sum, viz.,

$$\mathcal{V}_{\text{Coulomb}}(\mathbf{x}) = \sum_{i < j} \frac{q_i q_j}{4\pi\epsilon_0\epsilon_1 r_{ij}}, \quad (14.51)$$

extended to the central cell and all its neighbours, does not formally converge, the key idea of the Ewald summation method (Fig. 14.10) is to decompose (14.51) into a sum in position space and a sum in momentum space:

$$\sum_{\mathbf{n}} \frac{1}{|\mathbf{n}|} \mathcal{F}(\mathbf{n}) + \sum_{\mathbf{m}} \frac{1}{|\mathbf{m}|} [1 - \mathcal{F}(\mathbf{n})]. \quad (14.52)$$

By surrounding each point charge in the system by a Gaussian charge distribution

$$\varrho_i(\mathbf{x}) = q_i \alpha^3 \frac{\exp(-\alpha^2 r^2)}{\sqrt{\pi^3}}, \quad (14.53)$$

where α is a positive parameter characterising the width of the Gaussian distribution, the first sum converges rapidly when $\mathbf{n} \rightarrow \infty$, because $\mathcal{F}(\mathbf{n})$

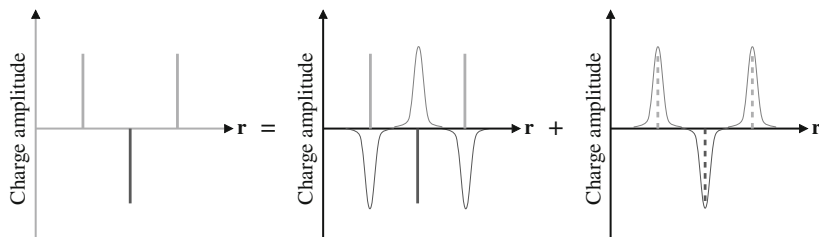


Fig. 14.10. Terms in an Ewald sum for a 1D system of point charges. In position space, each charge is surrounded by a Gaussian charge distribution $\varrho_i(\mathbf{x})$, of equal amplitude and opposite sign. This contribution is counterbalanced in momentum space by a Gaussian distribution $\varrho_j(\mathbf{x})$ of opposite sign

decreases very fast. The contribution in position space is short range (see Fig. 14.10). The second sum evaluated in momentum space uses a Fourier transform to solve Poisson's equation, i.e.,

$$\nabla^2 V_i(\mathbf{x}) = -4\pi\varrho_i(\mathbf{x}).$$

The transform decreases rapidly and the sum converges in the same way [29].

The incorporation of lattice sums according to the scheme proposed by Ewald in a molecular dynamics code, using a macromolecular force field like the one described in (14.11), can be summed up as follows:

$$\begin{aligned} \mathcal{V}_{\text{Ewald}}(\mathbf{x}) &= \frac{1}{2V\epsilon_0} \sum_{\mathbf{k} \neq 0} \frac{\exp(-k^2/4\alpha^2)}{k^2} \left[\sum_j q_j \exp(-i\mathbf{k} \cdot \mathbf{r}_j) \right] \left[\sum_j q_j \exp(i\mathbf{k} \cdot \mathbf{r}_j) \right] \\ &+ \frac{1}{4\pi\epsilon_0} \sum_i \sum_{j>i} \frac{q_i q_j}{r_{ij}} \text{erfc}(\alpha r_{ij}) - \frac{\alpha}{4\pi^{3/2}\epsilon_0} \sum_i q_i^2 \\ &+ \frac{1}{4\pi\epsilon_0} \sum_i \sum_{\substack{j \text{ bound to } i \\ j>i}} \frac{q_i q_j}{r_{ij}}. \end{aligned}$$

The first term is a sum over all vectors $\mathbf{k}$ in momentum space, and α is a positive parameter characterising the width of the Gaussian charge distribution surrounding each point charge in the system. The second term is a sum in position space, and $\text{erfc}(x)$ is the complementary error function, i.e., $1 - \text{erf}(x)$. The third and fourth contributions are corrections, given that the sum in momentum space is over all atomic pairs $\{i, j\}$, including therefore self-, 1-2, and 1-3 terms.

Formally, the computing investment involved in the classic approach using Ewald lattice summation is $\mathcal{O}(N^2)$, where N is the number of particles making up the system. As has been shown by Perram et al. [31] and Fincham [32],

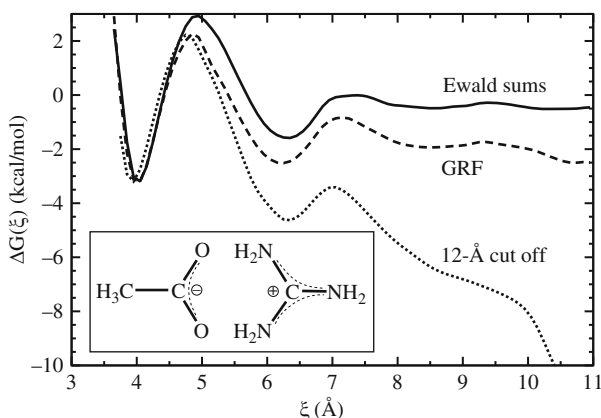


Fig. 14.11. Free energy profile characterising the approach between a guanidinium cation and an acetate anion in the C_{2v} geometry, in aqueous solution [30]. *Continuous curve*: Ewald sum (EW) simulation, $\alpha = 0.3 \text{ \AA}^{-1}$. *Dashed curve*: Generalised reaction field (GRF) simulation, $R_{\text{cutoff}} = 12 \text{ \AA}$. *Dotted curve*: Spherical cutoff (SC) simulation, $R_{\text{cutoff}} = 12 \text{ \AA}$

this cost can be reduced to $\mathcal{O}(N^{3/2})$ by a judicious choice of the width α of the Gaussian distributions, the number of vectors $\mathbf{k}$, and the cutoff of the pairwise interactions in position space. In general, one must balance the CPU time invested in evaluating position space and momentum space summations in order to achieve an $N^{3/2}$ dependence.

Less costly alternatives than conventional Ewald summation are based on an evaluation of the sum in momentum space using the fast Fourier transform (FFT) [33]. A 3D grid is constructed to fill the space in which the molecular dynamics simulation is carried out. The charges carried by the particles making up the system are interpolated on this grid, and the corresponding charge distribution, denoted by $\varrho(\mathbf{x})$, is calculated. Using the FFT method, the transform $\hat{\varrho}(\mathbf{k})$ of the charge distribution is determined on the basis of vectors $\mathbf{k}$ in momentum space. The long-range contribution of the electrostatic potential is then evaluated from

$$\hat{V}_{\text{long}}(\mathbf{k}) = \hat{\mathcal{G}}(\mathbf{k})\hat{\varrho}(\mathbf{k}),$$

where $\hat{\mathcal{G}}(\mathbf{k})$ is the influence function defined by

$$\hat{\mathcal{G}}(\mathbf{k}) = \frac{\hat{\lambda}(\mathbf{k})}{\epsilon_0 k^2},$$

in which $\lambda(\mathbf{x})$ is a distribution depending only on the geometrical characteristics of the simulation cell. By the inverse transform, the contribution $V_{\text{long}}(\mathbf{x})$ is estimated at different points of the 3D grid. Forces of electrostatic origin are then determined by numerical differentiation of the potential. Finally, the

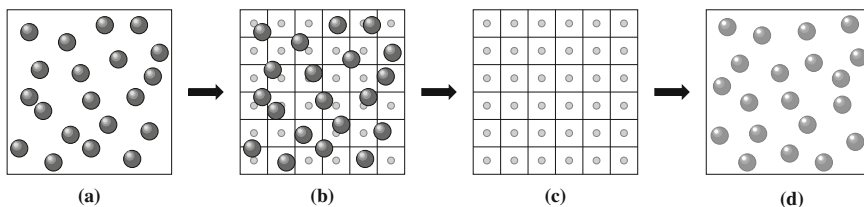


Fig. 14.12. Particle mesh method on a 2D lattice. (a) System of charged particles. (b) The charges are interpolated on a 2D grid. (c) Using a fast Fourier transform (FFT) method [33], the potential and forces are evaluated at each point of the grid. (d) The forces on the particles are then interpolated, and the particle positions updated

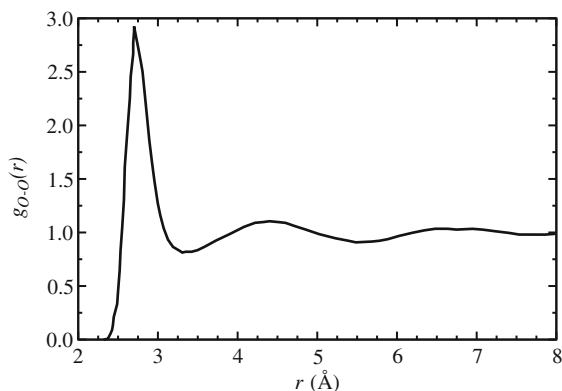


Fig. 14.13. Oxygen–oxygen radial distribution function for TIP4P water [38], obtained from a molecular dynamics simulation at 300 K

electric field and potential are interpolated from the grid to the particle positions (see Fig. 14.12). This scheme represents the central idea of particle mesh algorithms [34, 35], such as particle mesh Ewald (PME) or particle–particle particle mesh [36] (P³M), which has a formal cost going as $\mathcal{O}(N \ln N)$.

14.6 Some Properties Accessible to Simulation

14.6.1 Structural Properties from Simulations

Among the static properties that can be extracted from the trajectories generated by molecular dynamics simulations, structural quantities are particularly interesting, insofar as they describe the local order of the molecular system. Radial distribution functions $g(x)$ give the probability of finding a pair of atoms separated by a distance x , relative to the expected probability for a totally random distribution with the same density (see Fig. 14.13) [2, 25, 37]. The function $g(x)$ is defined by integrating the configurational distribution

function over all atomic positions except for the positions of the two relevant atoms:

$$g(\mathbf{x}_1; \mathbf{x}_2) = \frac{N(N-1)}{\rho^2 \int e^{-\beta V(\mathbf{x}_1, \dots, \mathbf{x}_N)} d\mathbf{x}_1 \dots d\mathbf{x}_N} \int e^{-\beta V(\mathbf{x}_1, \dots, \mathbf{x}_N)} d\mathbf{x}_3 \dots d\mathbf{x}_N, \quad (14.54)$$

where ρ is the density of the liquid. For a system in which all the atoms are identical, the radial distribution function simplifies to an ensemble average over pairs of atoms:

$$g(\mathbf{x}) = \frac{V}{N^2} \left\langle \sum_i \sum_{j \neq i} \delta(\mathbf{x} - \mathbf{x}_{ij}) \right\rangle. \quad (14.55)$$

For a pair of atoms $\{i, j\}$, the radial distribution function is easily evaluated from:

$$g_{ij}(r) = \frac{\langle n_j(r + \delta r) \rangle}{4\pi \rho_j \int r^2 dr}, \quad (14.56)$$

in which $\langle n_j(r + \delta r) \rangle$ is the average number of sites j whose distance from i lies in the range r to $r + \delta r$, and ρ_j is the average density of sites j in the sample.

Exercise. Find the expression for the radial distribution function using (14.54).

Another quantity that is just as relevant for understanding the ordering of the molecular liquid is the Kirkwood factor $G_K(R)$, which depends on the distance [39]. This quantity characterises the correlation between the dipole moment $\boldsymbol{\mu}_i$ of a reference molecule i and the dipole moment of its neighbouring molecules j contained within a sphere of radius R centered on i :

$$G_K(R) = \frac{\left\langle \sum_{i,j; r_{ij} < R} \boldsymbol{\mu}_i \cdot \boldsymbol{\mu}_j \right\rangle}{N\mu^2}. \quad (14.57)$$

The Kirkwood factor is particularly sensitive to the way electrostatic effects are handled in the simulation. A spherical cutoff leads to many artifacts in the dipole–dipole correlation at the sphere boundary.

14.6.2 Dynamical Properties from Simulations

As its name suggests, molecular dynamics can also reveal the dynamical properties of a system. Among these, temporal correlation functions are very informative regarding the relaxation times of different degrees of freedom in the

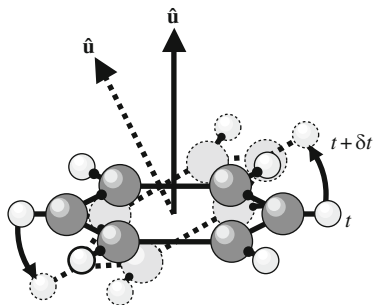


Fig. 14.14. Reorientation of a benzene molecule over a period of time. $\hat{\mathbf{u}}$ is the unit vector along the C_6 axis of the molecule

system [1,2]. Quite generally, the correlation between two quantities $\mathcal{B}$ and $\mathcal{B}'$ is given by the coefficient

$$c_{\mathcal{B}\mathcal{B}'} = \frac{\langle \delta \mathcal{B} \delta \mathcal{B}' \rangle}{\sigma(\mathcal{B})\sigma(\mathcal{B}')}, \quad (14.58)$$

where $\delta \mathcal{B} = \mathcal{B} - \langle \mathcal{B} \rangle$, with $\langle \mathcal{B} \rangle$ the statistical ensemble average of the quantity $\mathcal{B}$, while

$$\sigma(\mathcal{B}) = \sqrt{\langle \mathcal{B}^2 \rangle - \langle \mathcal{B} \rangle^2}.$$

The coefficient $c_{\mathcal{B}\mathcal{B}'}$ varies between 0 and 1, where 0 indicates that there is no correlation. This formulation can be generalised to the case where $\mathcal{B}$ and $\mathcal{B}'$ are evaluated at different times. The correlation function is then

$$c_{\mathcal{B}\mathcal{B}'}(t) = \frac{\langle \delta \mathcal{B}(t) \delta \mathcal{B}'(0) \rangle}{\sigma(\mathcal{B})\sigma(\mathcal{B}')}. \quad (14.59)$$

The time average of the numerator is taken over all time origins. If it happens that $\mathcal{B}' \equiv \mathcal{B}$, the resulting function is called an autocorrelation function, defined by

$$c_{\mathcal{B}\mathcal{B}}(t) = \frac{\langle \delta \mathcal{B}(t) \delta \mathcal{B}(0) \rangle}{\langle \delta \mathcal{B}(0) \delta \mathcal{B}(0) \rangle}. \quad (14.60)$$

Integrating the correlation function from $t = 0$ to $t = \infty$ yields the correlation time. As an example, consider the reorientational correlation time for benzene in water, as illustrated in Figs. 14.14 and 14.15. This is necessarily a longer simulation than for pure water, given that the statistical average refers to a single molecule.

The quantity evaluated here is the autocorrelation coefficient $c_{\hat{\mathbf{u}}\hat{\mathbf{u}}}(t) = \langle \hat{\mathbf{u}}(t) \hat{\mathbf{u}}(0) \rangle$, where $\hat{\mathbf{u}}$ is the unit vector along the C_6 or C'_2 axis of the benzene molecule. The inertia tensor of the molecule does suggest, however, that the

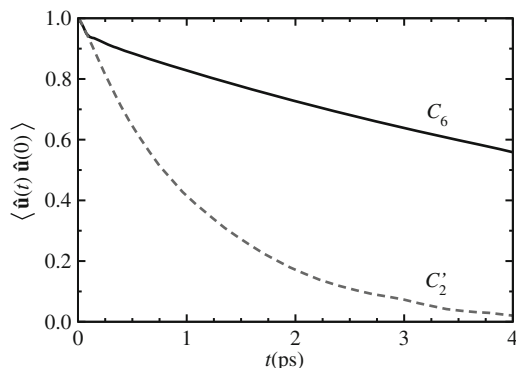


Fig. 14.15. Reorientational autocorrelation functions for benzene in liquid water at 300 K. Integrating the profiles yields a correlation time $\tau(C_6) = 2.4$ ps and $\tau(C'_2) = 0.5$ ps

rotational motion about the C'_2 axis will decorrelate more quickly than that about the C_6 axis.

Given the correlation functions of the derivative $\dot{\mathcal{B}}$ of $\mathcal{B}$, rather than $\mathcal{B}$ itself, numerical simulations can estimate transport coefficients. For example, equilibrium molecular dynamics simulations can estimate the diffusion coefficient D via the integral [2]

$$D = \frac{1}{3} \int_0^\infty \langle \dot{\mathbf{x}}_i(t) \cdot \dot{\mathbf{x}}_i(0) \rangle dt, \quad (14.61)$$

in which $\dot{\mathbf{x}}_i(t) \equiv \mathbf{v}_i(t)$ is the velocity of the center of mass of molecule i . For sufficiently long times, D can be obtained from Einstein's relation

$$D = \frac{1}{3} \frac{\langle |\mathbf{x}_i(t) - \mathbf{x}_i(0)|^2 \rangle}{2t}. \quad (14.62)$$

14.6.3 Molecular Dynamics and Free Energy

So far we have seen that molecular dynamics can inform about both structural and dynamical properties of systems. We shall now explain how an ensemble of configurations generated by molecular dynamics can be used to estimate fundamental thermodynamic quantities such as the free energy, which is used to predict the propensity of chemical species to associate or react.

In many respects, the foundations of free energy calculations were laid by Kirkwood [40], Zwanzig [41], and Bennett [42]. In 1954, Zwanzig put forward a strategy based on simple principles of statistical mechanics, which could be used to calculate differences of free energy ΔA between two thermodynamic states a and b of a system. This was called free energy perturbation (FEP):

$$\Delta A = -\frac{1}{\beta} \ln \left\langle \exp \left\{ -\beta [\mathcal{H}_b(\mathbf{x}, \mathbf{p}_x) - \mathcal{H}_a(\mathbf{x}, \mathbf{p}_x)] \right\} \right\rangle_a. \quad (14.63)$$

Here, $\mathcal{H}_a(\mathbf{x}, \mathbf{p}_x)$ is the Hamiltonian of the N -particle system in the thermodynamic state a , while $\langle \cdots \rangle_a$ denotes an average taken over an ensemble of configurations that are representative of the state. In most cases, the states a and b are sufficiently orthogonal to disallow direct application of (14.63). In practice, the reaction path between the two states is split into a series of unphysical intermediate substates, connected via a coupling parameter. Equation (14.63) can be expressed formally as a continuous integral, which leads to the so-called thermodynamic integration (TI) method:

$$\Delta A = \int_0^1 \left\langle \frac{\partial \mathcal{H}(\mathbf{x}, \mathbf{p}_x; \lambda)}{\partial \lambda} \right\rangle_\lambda d\lambda, \quad (14.64)$$

where λ is the coupling parameter relating state a , with $\lambda = 0$, to state b , with $\lambda = 1$.

Exercise. Starting with the original expression $A = -(1/\beta) \ln Q$ for the Helmholtz free energy, where Q is the canonical partition function, show that the difference in free energy between two states a and b can be expressed as the ensemble average (14.63).

The umbrella sampling (US) method [43] provides another way, related in some ways to the FEP and TI methods, for obtaining results that can be directly compared with experiment. In the framework of this approach, sampling along an order parameter ξ , i.e., a genuine reaction coordinate, can be restricted to a finite portion of the configuration space by means of judiciously chosen external bias potentials $\mathcal{V}_{\text{ext}}(\xi)$. Torrie and Valleau have shown that the unbiased ensemble average of a quantity $\mathcal{B}$ can be recovered from a biased statistical ensemble:

$$\langle \mathcal{B} \rangle = \frac{\langle \mathcal{B} \exp(-\beta \mathcal{V}_{\text{ext}}) \rangle_{\text{bias}}}{\langle \exp[-\beta \mathcal{V}_{\text{ext}}(\xi)] \rangle_{\text{bias}}}. \quad (14.65)$$

It follows that the free energy profile along ξ can be directly determined from the probability $\mathcal{P}_{\text{bias}}(\xi)$ of finding the system at different values of the order parameter calculated in the biased configuration space:

$$A(\xi) = -\frac{1}{\beta} \ln \mathcal{P}_{\text{bias}}(\xi) - \mathcal{V}_{\text{ext}}(\xi) + A_0, \quad (14.66)$$

where A_0 is a constant. One limitation of the US method lies in the need to define, without prior knowledge, the form of the external potentials $\mathcal{V}_{\text{ext}}(\xi)$. This proves to be a difficult task when dealing with qualitatively novel problems. An attractive solution is provided by the adaptive bias force (ABF) method [46], in which one integrates the average force exerted along ξ and

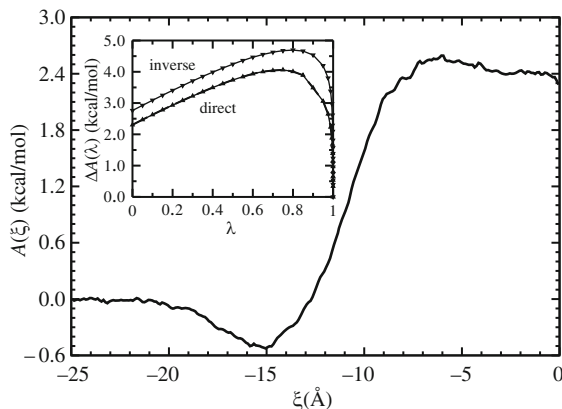


Fig. 14.16. Estimate of the free energy of hydration of methane by transfer between air and aqueous phases, and by annihilation (direct simulation) and creation (reverse simulation) in an aqueous medium (*insert*). In the first case, a methane molecule moves freely along the normal to the interface formed by a sheet of water in equilibrium with its vapour phase, using the ABF method. In the second case, the interactions between the methane and the water are gradually cancelled or created depending on the value of λ . These two methods produce the same result quantitatively, viz., 2.4 ± 0.4 kcal/mol [44], in good agreement with the experimental value of 2.0 kcal/mol [45]

obtained from an unconstrained molecular dynamics simulation [44]. During the simulation, a bias force is estimated so that, when it is applied to the system, one obtains a Hamiltonian for which no residual average force acts along ξ . Then all values available to the order parameter are sampled with the same probability, thereby improving the accuracy of the free energy estimates.

As can be seen from Fig. 14.16, the fact that the two estimates of the free energy of hydration of methane via two different approaches (ABF and FEP) should happen to coincide suggests that, independently of the intrinsic quality of the force field, the methods underlying these calculations has been perfectly mastered today. This example also illustrates the tangible link that free energy calculations can provide between theory and experiment.

The present section makes no claim to be exhaustive. The interest reader is referred to the specialised literature, e.g., [47], for more detail.

14.7 Molecular Dynamics and Parallelisation

Among the intrinsic limitations of molecular dynamics, the most serious is undoubtedly the computer time required to explore complex molecular systems over significant lapses of time. The complexity increases with the number of atoms, which can easily exceed 10,000 in the case of biological systems. For such assemblies, it is often difficult to achieve simulation times compatible

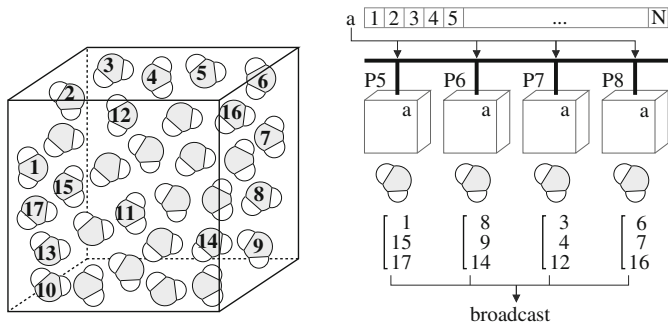


Fig. 14.17. Decomposing a task into groups of atoms. The work done by the central loop in calculating forces is split over the various available processors, each one handling a set of atoms that are reasonably close to one another in space

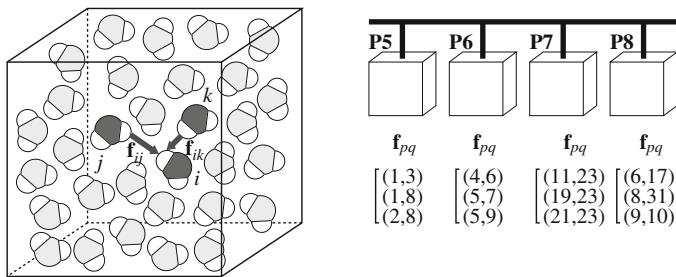


Fig. 14.18. Parallelisation by dividing up the force calculation. The central computation loop for the forces is split over the various available processors and the contributions to the force exerted on the relevant atom i are evaluated concomitantly

with the physical reality of the modelled phenomenon, which range from nanoseconds to microseconds. This difficulty comes from the step used to integrate the equations of motion, which is incommensurable with the simulation times envisaged. A good illustration is provided by models of biological membranes, in which the collective motions of lipid chains relax on time scales between 10^{-11} and 10^{-6} s. This time scale should be compared with the typical integration step of around 2×10^{-12} s for molecular dynamics, which ensures energy conservation in the system. With coarse-grained models, freed from the harder degrees of freedom of the system, the equations of motion can be integrated with time steps of the order of 40×10^{-12} . This improves the prospects for investigating complex phenomena on time scales of microsecond order. However, such models are rather crude and supply only a qualitative image of the process. An atomistic description of molecular systems and the limitations associated with them are the price to pay for quantitative results.

Over the past fifteen years, the development of parallel architectures has considerably increased the available computer resources, pushing back the

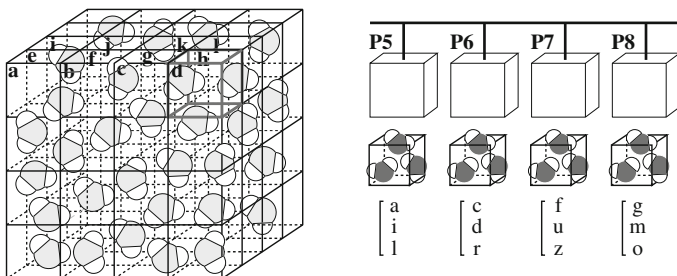


Fig. 14.19. Decomposition into spatial domains. The simulation cell is cut up into several subcells, each one being handled by a processor in the parallel architecture. In order to balance the work load, the cell is split into more subcells than there are processors

usual limits of molecular simulations. In particular, massively parallel multiple instruction multiple data (MIMD) computers have made it possible to sample phase space over periods of time never before achieved for complex molecular assemblies. With the advent of these new architectures, various parallelisation strategies have been devised. The simplest of these decomposes the calculation task into groups of atoms, as shown in Fig. 14.17. The full set of N particles is shared over the n_{proc} available processors. A table defined in the central memory contains all data concerning the positions of the particles. This data is necessarily shared over all the processes to compute the forces involving particles that do not belong to the same group.

It is immediately clear that this division into groups of atoms is limited to systems of moderate size and involves a considerable amount of memory. Parallelisation by splitting the force calculation over n_{proc} processors is without doubt the most efficient in memory terms (see Fig. 14.18). In this approach, pairwise interactions are evaluated in units distributed over the different processors. This scheme has the clear advantage that a scalar molecular dynamics code can be used, by parallelising the computer-costly routines. Compilation directives like OPENMP lend themselves particularly well to this approach in the framework of shared memory architectures [48]. The difficulty in implementing such a strategy arises from the possibility that an atom j may interact simultaneously with atom j and atom k , when these contributions are evaluated in different units. In this case, the contributions of j and k to the force $\mathbf{f}_i$ exerted on i cannot be updated at the same time on distinct threads.

The performance of these strategies which divide up the force calculation tends rapidly to an asymptotic limit with increasing n_{proc} . This solution is attractive whenever it can be applied to systems of reasonable size, using architectures with a limited number of processors, typically $n_{\text{proc}} < 16$. The most promising and most efficient alternative to the last two approaches is to decompose into spatial domains, as depicted in Fig. 14.19. The main idea

here is to divide up the simulation cell into n_{cell} subcells handled, in the best of situations, quasi-independently by the n_{proc} processors in the architecture. Complete independence is impossible, even in the case of atomic fluids, because some communication, even very minimal, is required to update the forces exerted on the various particles making up the system. The situation is substantially complicated when a single molecule, e.g., a protein, extends over several contiguous subcells. This explains among other things why a parallel molecular dynamics program based on spatial decomposition can only be developed from scratch, following a quite different line of thought to methods that handle groups of atoms or divide up the forces, whence there is no way of recycling a standard scalar code.

This idea has seen a certain number of improvements, but the details go beyond the scope of this chapter. However, among the most important of these, it is worth mentioning the idea of load balancing, which shares the $n_{\text{cell}} \gg n_{\text{proc}}$ over the available processors and moves them around during the first steps of the simulation, in such a way that the computational effort is similar for all the processors. This strategy is available in many molecular dynamics codes, e.g., NAMD [49], one of the most widely used today to model complex biological systems.

14.8 Conclusion

The examples discussed in this chapter give an idea of the field of investigation accessible to numerical simulations for handling the kind of problems faced by the modeller on an everyday basis. This field of investigation is gradually broadening as the performance-to-cost ratio improves for intensive computer calculation. Furthermore, the availability of massively parallel architectures, although still somewhat unfairly distributed around the world, opens the way to particularly ambitious applications of molecular dynamics. The study of macromolecular systems with relevance to physicochemistry and biology, over significant time scales, is a good illustration. Returning to Fig. 14.3 (see p. 807) and considering the literature over the past few years, one finds that molecular dynamics can rather easily tackle systems of up to 10^5 atoms over periods of a few nanoseconds. For instance, the fluctuations of a membrane protein in its natural environment fall within this range of sizes and times. Combining with free energy calculations, which allow one to escape the limitations of Boltzmann sampling, molecular dynamics also provides a way of modelling slow processes arising in such systems, such as assisted transport of small molecules by a dedicated membrane protein [50].

However, two comments can be made on the basis of the progress made over the past few years. Firstly, the quality of a calculation depends closely on the simulation strategy. Depending on the problem at hand, some methods must be ruled out in favour of more rigorous, and often more costly approaches. This is the case when we consider electrostatic effects. In the least

costly approach, these are handled by a pairwise additive approximation and a spherical cutoff for long-range forces, but in a more sophisticated treatment, lattice sums or even distributed polarisability models may be included. While the general methodology is now well established and robust, not everything can yet be simulated by molecular dynamics, given that Newton's equations of motion are integrated numerically in infinitesimal steps of 1–2 fs, to be compared with the time scales under investigation and the size of the system.

Moreover, great care must be taken with regard to the nature of the system when parametrising the potential energy function. Although as a result of many years of development the standard force fields have reached the maturity required to be able to tackle a wide range of macromolecular assemblies with some confidence, it is true to say that many applications need considerable further development. This is the case, for example, with lipid membranes, for which multipurpose functionals do not provide a satisfactory solution. Finally, although the generalisation of molecular modelling to all aspects of chemistry may encourage an ever greater use of numerical simulation, in most situations the codes used should not be considered as 'black boxes'. In any case, only a close examination of the results, a detailed analysis of the trajectories generated, and a careful investigation of the derived thermodynamic properties can confirm the correctness and the physicochemical meaning of such calculations [51].

Acknowledgements

I would like to thank my colleagues François Dehez, Jérôme Delhommelle, and Mounir Tarek for accepting to check and comment on the text, with suggestions for improving the content.

References

1. Frenkel, D., Smit, B.: *Understanding Molecular Simulations: From Algorithms to Applications*, Academic Press, San Diego (1996)
2. Allen, M.P., Tildesley, D.J.: *Computer Simulation of Liquids*, Clarendon Press, Oxford (1987)
3. Ewald, P.: Die Berechnung optischer und elektrostatischer Gitterpotentiale, *Ann. Phys.* **64**, 253–287 (1921)
4. Ladd, A.J.C.: Long-range dipolar interactions in computer simulations of polar liquids, *Mol. Phys.* **36**, 463–474 (1978)
5. van Gunsteren, W.F., Berendsen, H.J.C.: Computer simulation of molecular dynamics: Methodology, applications, and perspectives in chemistry, *Angew. Chem. Int. Ed. Engl.* **29**, 992–1023 (1990)
6. Sanbonmatsu, K.Y., Simpson, J., Tung, C.S.: Simulating movement of tRNA into the ribosome during decoding, *Proc. Natl. Acad. Sci. USA* **102**, 15854–15859 (2005)

7. Duan, Y., Kollman, P.A.: Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution, *Science* **282**, 740–744 (1998)
8. Tuckerman, M.E., Martyna, G.J.: Understanding modern molecular dynamics: Techniques and applications, *J. Phys. Chem. B* **104**, 159–178 (2000)
9. Leimkuhler, B., Reich, S.: *Simulating Hamiltonian Dynamics*, Cambridge University Press (2005)
10. Hairer, E., Lubich, C., Wanner, G.: Geometric numerical integration. Structure-preserving algorithms for ordinary differential equations. In: *Springer Series in Computational Mathematics*, Vol. 31, 2nd edn. Springer, Berlin Heidelberg New York (2006)
11. Martyna, G.J., Tuckerman, M.E., Tobias, D.J., Klein, M.L.: Explicit reversible integrators for extended systems dynamics, *Mol. Phys.* **87**, 1117–1128 (1996)
12. Cornell, W.D., Cieplak, P., Bayly, C.I., Gould, I.R., Merz Jr., K.M., Ferguson, D.M., Spellmeyer, D.C., Fox, T., Caldwell, J.C., Kollman, P.A.: A second generation force field for the simulation of proteins, nucleic acids, and organic molecules, *J. Am. Chem. Soc.* **117**, 5179–5197 (1995)
13. Kollman, P., Dixon, R., Cornell, W., Fox, T., Chipot, C., Pohorille, A.: The development/application of a ‘minimalist’ force field using a combination of ab initio and experimental data. In: *Computer Simulation of Biomolecular Systems: Theoretical and Experimental Applications*, ed. by W.F. Van Gunsteren, P.K. Weiner, Escom, The Netherlands (1997) pp. 83–96
14. Ryckaert, J., Bellemans, A.: Molecular dynamics of liquid alkanes, *Chem. Soc. Faraday Discuss.* **66**, 95–106 (1978)
15. Chipot, C., Ángyán, J.G.: Continuing challenges in the parametrization of intermolecular force fields. Towards an accurate description of electrostatic and induction terms, *New J. Chem.* **29**, 411–420 (2005)
16. Wang, W., Skeel, R.D.: Fast evaluation of polarizable forces, *J. Chem. Phys.* **123**, 164107 (2005)
17. Gay, J.G., Berne, B.J.: Modification of the overlap potential to mimic a linear site–site potential, *J. Chem. Phys.* **74**, 3316–3319 (1981)
18. Shelley, J.C., Shelley, M.Y., Reeder, R.C., Bandyopadhyay, S., Klein, M.L.: A coarse grain model for phospholipid simulations, *J. Phys. Chem. B* **105**, 4464–4470 (2001)
19. Nielsen, S.O., Lopez, C.F., Srinivas, G., Klein, M.L.: Coarse grain models and the computer simulation of soft materials, *J. Phys.: Condens. Matter* **16**, R481–R512 (2004)
20. Verlet, L.: Computer ‘experiments’ on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules, *Phys. Rev.* **159**, 98–103 (1967)
21. Martyna, G.J., Klein, M.L., Tuckerman, M.E.: Nosé–Hoover chains: The canonical distribution via continuous dynamics, *J. Chem. Phys.* **97**, 2635–2645 (1992)
22. Nosé, S.: A molecular dynamics method for simulations in the canonical ensemble, *Mol. Phys.* **52**, 255–268 (1984)
23. Hoover, W.G.: Canonical dynamics: Equilibrium phase-space distributions, *Phys. Rev. A* **31**, 1695–1697 (1985)
24. Berendsen, H.J.C., Postma, J.P.M., Van Gunsteren, W.F., DiNola, A., Haak, J.R.: Molecular dynamics with coupling to an external bath, *J. Chem. Phys.* **81**, 3684–3690 (1984)
25. McQuarrie, D.A.: *Statistical Mechanics*, Harper and Row, New York (1976)
26. Andersen, H.C.: Molecular dynamics simulations at constant pressure and/or temperature, *J. Chem. Phys.* **72**, 2384–2393 (1980)

27. Feller, S.E., Zhang, Y.H., Pastor, R.W., Brooks, B.R.: Constant pressure molecular dynamics simulations – The Langevin piston method, *J. Chem. Phys.* **103**, 4613–4621 (1995)
28. Barker, J.A.: Reaction field, screening, and long-range interactions in simulations of ionic and dipolar systems, *Mol. Phys.* **83**, 1057–1064 (1994)
29. Toukmaji, A.Y., Board Jr., J.A.: Ewald summation techniques in perspective: A survey, *Comput. Phys. Comm.* **95**, 73–92 (1996)
30. Rozanska, X., Chipot, C.: Modeling ion–ion interaction in proteins: A molecular dynamics free energy calculation of the guanidinium–acetate association, *J. Chem. Phys.* **112**, 9691–9694 (2000)
31. Perram, J.W., Petersen, H.G., de Leeuw, S.W.: An algorithm for the simulation of condensed matter which grows as the $3/2$ power of the number of particles, *Mol. Phys.* **65**, 875–889 (1988)
32. Fincham, D.: Optimisation of the Ewald sum for large systems, *Mol. Sim.* **13**, 1–9 (1994)
33. Cooley, J.W., Tukey, J.W.: An algorithm for the machine calculation of complex Fourier series, *Math. Comput.* **19**, 297–301 (1965)
34. Darden, T.A., York, D.M., Pedersen, L.G.: Particle mesh Ewald: An $N \log N$ method for Ewald sums in large systems, *J. Chem. Phys.* **98**, 10089–10092 (1993)
35. Essman, U., Perera, L., Berkowitz, M., Darden, T., Lee, H., Pedersen, L.G.: A smooth particle mesh Ewald method, *J. Chem. Phys.* **103**, 8577–8593 (1995)
36. Hockney, R.W., Eastwood, J.W.: *Computer Simulation Using Particles*, IOP Publishing, Bristol, England (1988)
37. Chandler, D.: *Introduction to Modern Statistical Mechanics*, Oxford University Press (1987)
38. Jorgensen, W.L., Chandrasekhar, J., Madura, J.D., Impey, R.W., Klein, M.L.: Comparison of simple potential functions for simulating liquid water, *J. Chem. Phys.* **79**, 926–935 (1983)
39. Neumann, M., Steinhauser, O.: The influence of boundary conditions used in machine simulations on the structure of polar systems, *Mol. Phys.* **39**, 437–454 (1980)
40. Kirkwood, J.G.: Statistical mechanics of fluid mixtures, *J. Chem. Phys.* **3**, 300–313 (1935)
41. Zwanzig, R.W.: High-temperature equation of state by a perturbation method. I. Nonpolar gases, *J. Chem. Phys.* **22**, 1420–1426 (1954)
42. Bennett, C.H.: Efficient estimation of free energy differences from Monte Carlo data, *J. Comp. Phys.* **22**, 245–268 (1976)
43. Torrie, G.M., Valleau, J.P.: Nonphysical sampling distributions in Monte Carlo free energy estimation: Umbrella sampling, *J. Comput. Phys.* **23**, 187–199 (1977)
44. Hénin, J., Chipot, C.: Overcoming free energy barriers using unconstrained molecular dynamics simulations, *J. Chem. Phys.* **121**, 2904–2914 (2004)
45. Ben-Naim, A., Marcus, Y.: Solvation thermodynamics of nonionic solutes, *J. Chem. Phys.* **81**, 2016–2027 (1984)
46. Darve, E., Pohorille, A.: Calculating free energies using average force, *J. Chem. Phys.* **115**, 9169–9183 (2001)
47. Chipot, C.: Calculating free energy differences from perturbation theory. In: *Free Energy Calculations. Theory and Applications in Chemistry and Biology*, ed. by C. Chipot, A. Pohorille, Springer, Berlin Heidelberg New York (2006)

48. Couturier, R., Chipot, C.: Parallel molecular dynamics using OPENMP on a shared memory machine, *Comp. Phys. Comm.* **124**, 49–59 (2000)
49. Phillips, J.C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, L., Schulten, K.: Scalable molecular dynamics with NAMD, *J. Comput. Chem.* **26**, 1781–1802 (2005)
50. Tajkhorshid, E., Nollert, P., Jensen, M.Ø., Miercke, L.J.W., O’Connell, J., Stroud, R.M., Schulten, K.: Control of the selectivity of the aquaporin water channel family by global orientational tuning, *Science* **296**, 525–530 (2002)
51. van Gunsteren, W.F., Mark, A.E.: Validation of molecular dynamics simulation, *J. Chem. Phys.* **108**, 6109–6116 (1998)

Real-Time PCR

A. Evrard, N. Boulle, and G.s Lutfalla

Over the past few years there has been a considerable development of DNA amplification by polymerase chain reaction (PCR), and real-time PCR has now superseded conventional PCR techniques in many areas, e.g., the quantification of nucleic acids and genotyping. This new approach is based on the detection and quantification of a fluorescent signal proportional to the amount of amplicons generated by PCR. Real-time detection is achieved by coupling a thermocycler with a fluorimeter. This chapter discusses the general principles of quantitative real-time PCR, the different steps involved in implementing the technique, and some examples of applications in medicine.

15.1 Real-Time PCR

15.1.1 Polymerase Chain Reaction

The polymerase chain reaction (PCR) provides a way of obtaining a large number of copies of a double-stranded DNA fragment of known sequence. This DNA amplification technique, developed in 1985 by K. Mullis (Cetus Corporation), saw a spectacular development over the space of a few years, revolutionising the methods used up to then in molecular biology. Indeed, PCR has many applications, such as the detection of small amounts of DNA, cloning, and quantitative analysis (assaying), each of which will be discussed further below.

Basics of the Chain Reaction

PCR exploits a heat-stable enzyme, DNA polymerase, which is able to copy a strand of DNA in the form of a complementary DNA strand (making use of the complementarity of the bases A, G, C, and T in the structure of DNA). During PCR, two synthetic primers are used, generally comprising around twenty bases. These each attach themselves specifically to one of the two

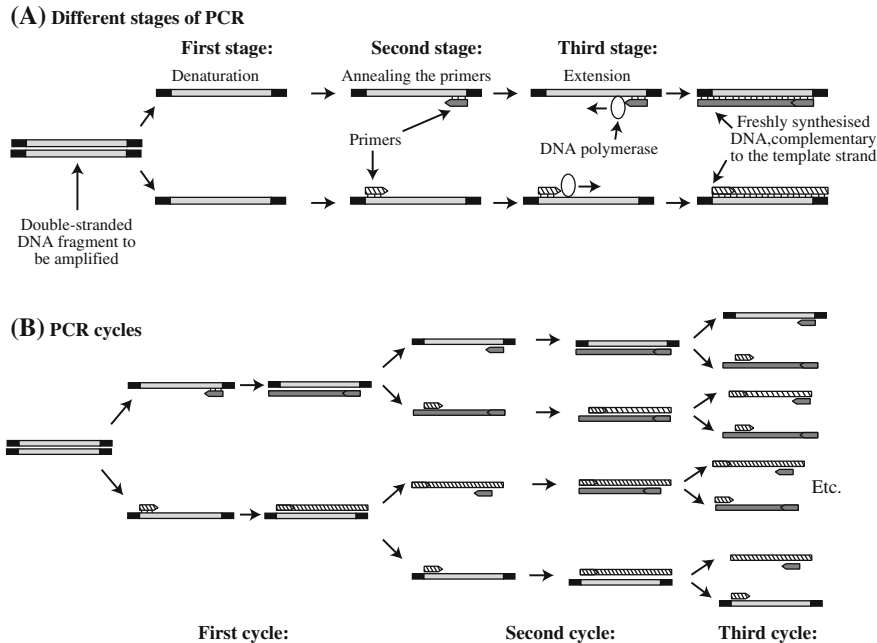


Fig. 15.1. Principle of the chain reaction. (a) The stages in a PCR cycle. (b) A sequence of PCR cycles

DNA strands, at the ends of the fragment to be amplified, serving as starting point for the DNA polymerase (see Fig. 15.1a).

PCR amplifies DNA by a sequence of cycles, each cycle comprising three steps: thermal denaturation, annealing (or hybridisation), and extension (or elongation). The first stage, denaturation, separates the two strands of the DNA to be amplified by a brief rise in temperature. Once the DNA has been denatured, the temperature is reduced to allow specific annealing of the two primers, each onto one of the two DNA strands. In the third stage, known as extension, DNA polymerase, in the presence of the four deoxyribonucleotide triphosphates (dATP, dGTP, dCTP, dTTP), synthesises the complementary DNA strands of the dissociated DNA strands, starting the synthesis from the primers. These PCR cycles of denaturation, annealing, and extension, are then repeated many times. Each amplicon, i.e., the DNA fragment synthesised at the end of the previous cycle, serves as DNA template for the subsequent cycle (see Fig. 15.1b). Hence, at the end of a PCR process, usually comprising 20–40 cycles, the initial DNA fragment will have been considerably amplified.

Conventional PCR, of the kind just described, is carried out in a device known as a thermocycler, able to adjust the temperature of the reaction medium in such a way that the different stages of the PCR cycle can follow on continuously. New-generation thermocyclers use the Peltier effect (see

below) to achieve extremely fast temperature variations, whence all the cycles and the amplification can be finished in about 2 h.

Peltier Effect

The Peltier effect is a thermoelectric effect that can be used to heat or cool a given medium. If we consider a conductor at constant temperature, the heat energy J_q received or released by the conductor is proportional to the current J , i.e., $J_q = aJ$, where a is the Peltier constant. If two conductors with different Peltier constants a_1 and a_2 are set in contact, part of the heat is transferred to the surroundings. The heat transferred is given by

$$\frac{dJ_q}{dt} = (a_1 - a_2)J.$$

Reversing the current has the effect of cooling the surroundings. In a thermocycler, several conductors with different Peltier constants a are set side by side, so that the temperature of the reaction medium can be rapidly varied and adjusted.

PCR Kinetics

The PCR obeys an exponential law as long as the substrate, i.e., the DNA to be amplified, remains the limiting factor. If the PCR efficiency is optimal (efficiency = 1), the number of DNA fragments generated doubles each cycle.¹ One then obtains the following equation:

$$Q_n = q_0 \times 2^n \quad (\text{in the } n \text{ th cycle}),$$

where Q_n is the amount of DNA amplified after n cycles, and q_0 is the initial amount of DNA to be amplified.

As the PCR proceeds, the reagents become the limiting factor (rather than the DNA to be amplified) and the amplification rate gradually decreases. The PCR rate then levels out and the number of generated amplicons drops off.

The PCR kinetics can be divided into three phases (see Fig. 15.2A):

- A first phase in which the number of amplified DNA fragments (amplicons) is below the detection threshold (due to background noise).
- An exponential phase.
- A plateau.

The DNA amplified after n cycles can be assessed once the PCR has stopped, by marking with a double-stranded DNA intercalator, such as ethidium bromide. When the DNA is labelled in this way, it can be visualised after migrating on an agarose gel (see Fig. 15.2B).

From this kinetics, it is theoretically possible to determine the initial amount of DNA q_0 in a sample (Fig. 15.3). Hence, if Q_{An} and Q_{Bn} are the

¹ In practice, the PCR efficiency is often less than unity (see Sect. 15.2.3).

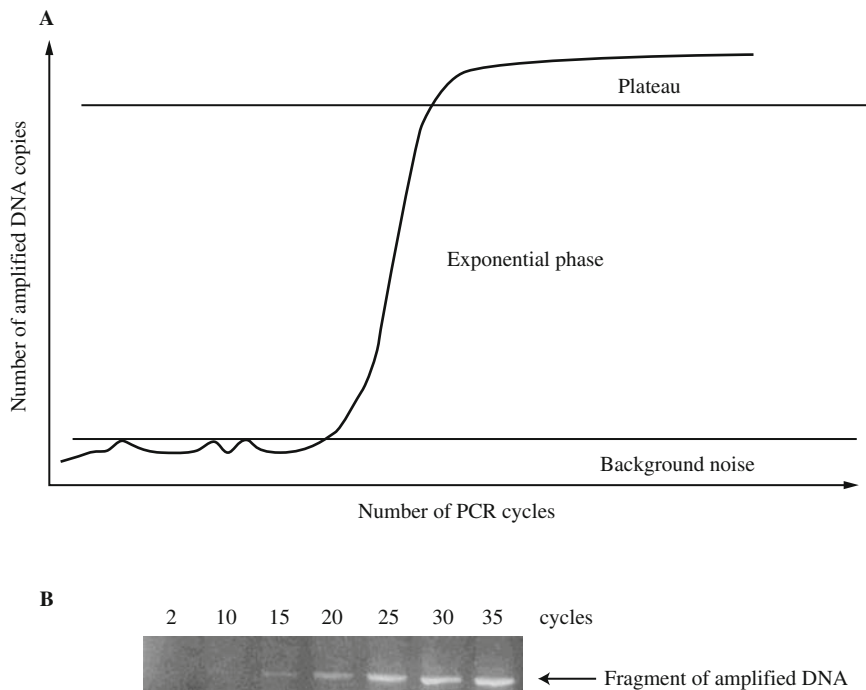


Fig. 15.2. (A) PCR kinetics. (B) Visualising amplified DNA fragments on agarose gel after marking with ethidium bromide, a double-stranded DNA intercalating agent

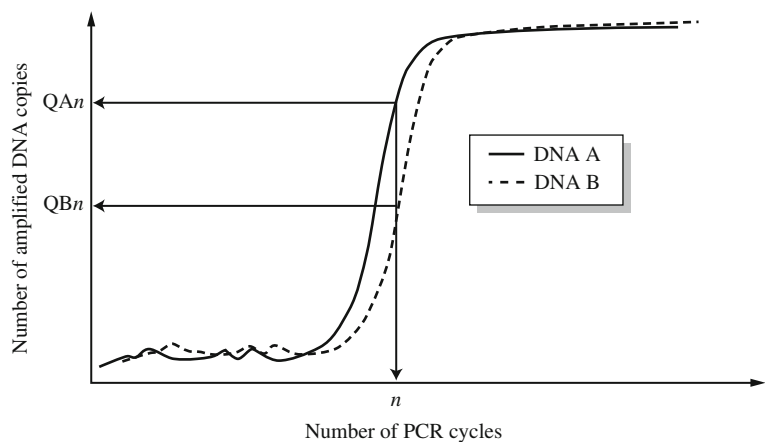


Fig. 15.3. Conventional PCR and DNA quantification by the end-point method

amounts of DNA from samples A and B amplified after n cycles, the ratio of the initial amounts of DNA in these two samples A and B, viz., q_{A0} and q_{B0} , is given by

$$\frac{Q_{An}}{Q_{Bn}} = \frac{q_{A0} \times 2^n}{q_{B0} \times 2^n}.$$

This is a comparative measurement, comparing one sample with another, or comparing a sample with some calibration curve.

This kind of assay uses the end-point method, i.e., the PCR is stopped for all samples (here, samples A and B) after n cycles (see Fig. 15.3). The amplified amounts of samples A and B after n cycles, denoted Q_{An} and Q_{Bn} , can be evaluated by marking the DNA with ethidium bromide, for example, and carrying out agarose gel electrophoresis (see Fig. 15.2B). More sensitive detection methods have been devised, such as using a fluorescent primer during the PCR, which both marks and quantifies the amplified DNA.

However, such measurements remain semi-quantitative and lack sensitivity. Moreover, experimentally, they are sometimes difficult to apply, because they require all the samples to be detectable at cycle n , and all PCRs to be in the exponential phase of the amplification, rather than in the plateau phase.

Basics and Utility of Quantitative Real-Time PCR

Quantitative real-time PCR is a recent approach based on ‘live’ monitoring of the PCR by detecting and quantifying a fluorophore incorporated into the amplified DNA products. During the PCR, emissions by the fluorophore, measured cycle by cycle, are directly proportional to the amount of amplicons that has been generated. This is achieved by combining a thermocycler with a fluorescence detector. Various fluorescence formats are currently available and will be discussed in the following.

The kinetics of real-time PCR is the same as described on p. 843. Quantification by real-time PCR involves determining the cycle threshold C_t or crossing point C_p , which is the number of cycles beyond which the amplification reaction becomes detectable, i.e., the signal becomes significantly stronger than the background noise (see Fig. 15.4). The more DNA there was initially in the reaction medium, the more quickly the exponential phase of the PCR will get started. Hence, the number of PCR cycles needed to detect the amplification products (and hence the C_t) will be lower as the initial DNA concentration q_0 is higher. So the underlying idea of quantitative real-time PCR is fundamentally different from conventional end-point PCR, since one is concerned with the point at which the PCR gets going, corresponding to the very beginning of the exponential phase of the reaction. Real-time PCR quantification is not therefore affected by the depletion of the reagents that characterises the plateau phase. Furthermore, since amplification and detection of amplified products are carried out in the same tube, one avoids post-PCR manipulation, a source of contamination.

The cycle-by-cycle detection of amplified products and the use of the beginning of the exponential PCR phase for quantification makes real-time PCR a much more sensitive and reproducible DNA quantification tool than the

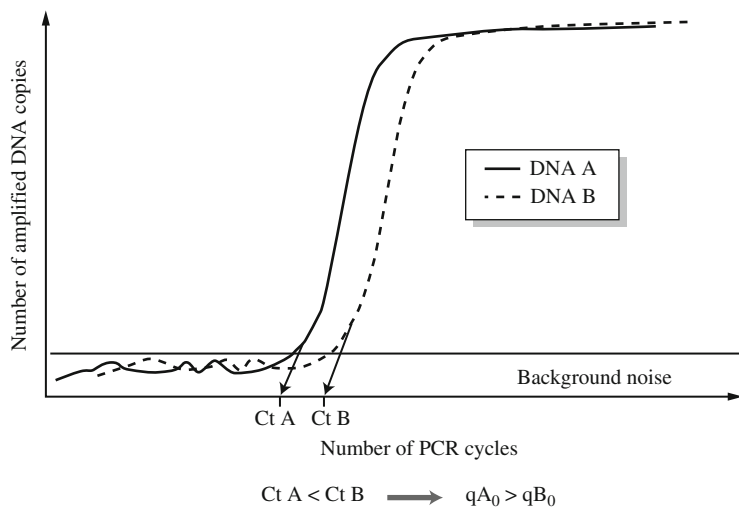


Fig. 15.4. Real-time PCR and DNA quantification

semi-quantitative methods used previously. Using this technique, an amount as small as 10 copies of DNA can be detected and measured in a reaction medium [1]. Moreover, the quantification range can be extended, up to 10 log for some parameters.

In addition to assaying, other applications of real-time PCR based on the fluorescence detection of amplified DNA have been considerably developed over the past few years, e.g., the search for mutations and genotyping. These will be discussed in the following.

15.1.2 Equipment Used for Quantitative Real-Time PCR

Even recently, there were only two or three quantitative real-time PCR systems available on the market [1], whereas today there are more than a dozen [2,3]. This equipment differs in the type of thermocycler, the number of optical filters available to analyse the various fluorophores in the reaction medium, the reaction substrates, and the number of samples that can be analysed, not to mention the time required for the PCR. The software supplied to handle quantitative PCR results may also differ in some ways.

Most quantitative real-time PCR equipment on offer today is based on use of the Peltier effect, as described above. These systems can bring about fast temperature changes (of the order of 3°C per second). One of the first systems of this type was the GeneAmp SDS 5700, commercialised by Applied Biosystem. A quite different idea is implemented by the LightCycler, produced by Roche Diagnostic, which was also one of the first real-time PCR devices to be commercialised. In this case, temperature changes are obtained by an air flow system through the device. Here, the PCR is carried out in a narrow

glass capillary, allowing extremely fast temperature changes in the reaction medium (20°C per second).

The number of optical filters available on the real-time PCR equipment determines the number of different fluorophores that can be detected in any given PCR process (multiplex PCR). The number of filters in quantitative PCR systems varies from 2 in the DNA Engine Opticon (manufactured by Biorad) to 5 or 6 filters in more recent systems such as ABI 7300 and 7500 Real-Time PCR Systems made by Applied Biosystem, or LightCycler 2.0 Instrument made by Roche Diagnostic. The fluorophores themselves can be excited by a laser or a halogen lamp. Many quantitative PCR systems use plates of 96 pits to carry out PCR. The capacity of more recent ones has risen to 384 pits, e.g., ABI Prism SDS 7900 HT made by Applied Biosystem. Other instruments have lower capacity, e.g., 32 glass capillaries for Roche Diagnostic's LightCycler. The volume of the reaction medium also varies from one instrument to another, ranging from 10 to 100 μ L. Another variable is the time required for real-time PCR, ranging from 30 min for the fastest systems, e.g., Smartcycler made by Cepheid, to 1 h for the LightCycler, and 2 h for most other instruments. Note also that Cepheid's Smartcycler is designed to carry out several quantitative PCR analyses at the same time.

Recently, an improved system has been proposed by Applied Biosystem, using microarrays able to carry out high throughput real-time PCR. This is the TaqMan Low Density Array. These microarrays contain up to 384 pits in which the reagents required for the real-time PCR are deposited in lyophilised form (reaction medium, DNA polymerase and specific primers for the genes to be quantified). To launch the PCR, the user simply places the desired DNA template in soluble form on the microarray, before putting the microarray in the real-time PCR instrument. These microarrays can be fabricated by selecting a panel of relevant genes whose specific primers are deposited by the manufacturer in each of the pits, whereupon a large number of quantitative PCRs can be carried out in real time.

15.1.3 Fluorescence Formats

Fluorescent DNA Markers

The detection of nucleic acids using fluorescent markers like ethidium bromide, SybrGreen I (SG), or PicoGreen (PG) has become a major application in molecular biology, both in experiment and diagnostics. For example, since it was first commercialised at the beginning of the 1990s, SybrGreen has been used successively to detect nucleic acids in gel or solution, for fluorescence imaging techniques, for flow cytometry, and most recently, for real-time PCR and capillary electrophoresis. The basic chemical structure of SybrGreen and PicoGreen is a monomeric cyanine (see Fig. 15.5). The first real-time PCR experiments were carried out with ethidium bromide, which intercalates in the DNA double helix. These intercalation properties cause breaks in the

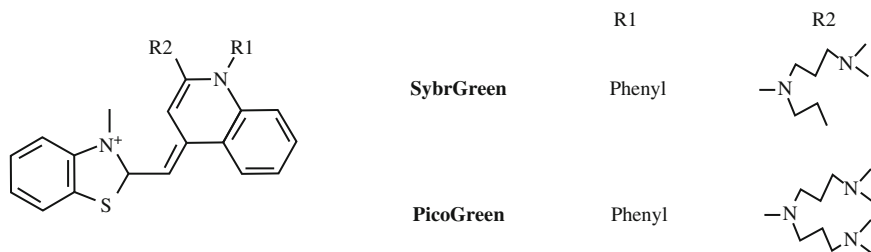


Fig. 15.5. Chemical structures of fluorescent markers based on monomeric cyanine

DNA and hence lead to non-optimal PCR. For this reason, ethidium bromide was soon replaced by SybrGreen I and its analogs.

SybrGreen I binds in the minor groove of the DNA double helix, and fluoresces at a wavelength of 530 nm, while the fluorescence of the dye in solution, which is not bound to the DNA, is negligible. This property is exploited during PCR, where the fluorescent signal grows as the amount of newly synthesised DNA increases. The SybrGreen I molecules bind at regular intervals onto the DNA, at a rate of about one molecule every 10 base pairs, and the fluorescence emitted is thus proportional to the length of the amplified fragments. Maximal fluorescence is obtained at the end of the extension stage (see Fig. 15.6A).

The fluorescent markers can simply be added to the conventional PCR mixtures containing water, buffer, MgCl_2 , nucleotides, and a thermostable DNA polymerase. They are easy and cheap to use, and provide a way of avoiding the often tedious requirement of designing specific fluorescent probes and new sets of primers. The main drawback with these markers is their lack of specificity, given that they will bind onto any double-stranded DNA sequence. This means that non-specific amplifications and primer dimers will cause unwanted fluorescence which adds to the specific signal and perturbs the quantification of the target sequence. The use of such markers therefore involves rigorous implementation and maximal optimisation of the PCR conditions. The melting temperature T_m of the amplified products can be checked using DNA melting curves at the end of the PCR to guarantee the specificity of the signal (see Sect. 15.2).

In real-time PCR, fluorescent markers are mainly used to quantify nucleic acids. However, they can also be used in some genotyping applications with analysis of melting curves (see Sects. 15.2 and 15.3). For example, if a few base pairs are deleted in the relevant gene, this will modify T_m sufficiently to allow detection by SybrGreen. Recently, Idaho has commercialised a new marker called LCGreen, which can be used to produce high-resolution melting curves for detection of small sequence variations. The particular DNA binding properties of this new marker should even allow the detection of single nucleotide polymorphisms (SNP), concerning a single base pair, something that can only otherwise be done using fluorescent probes at the present time.

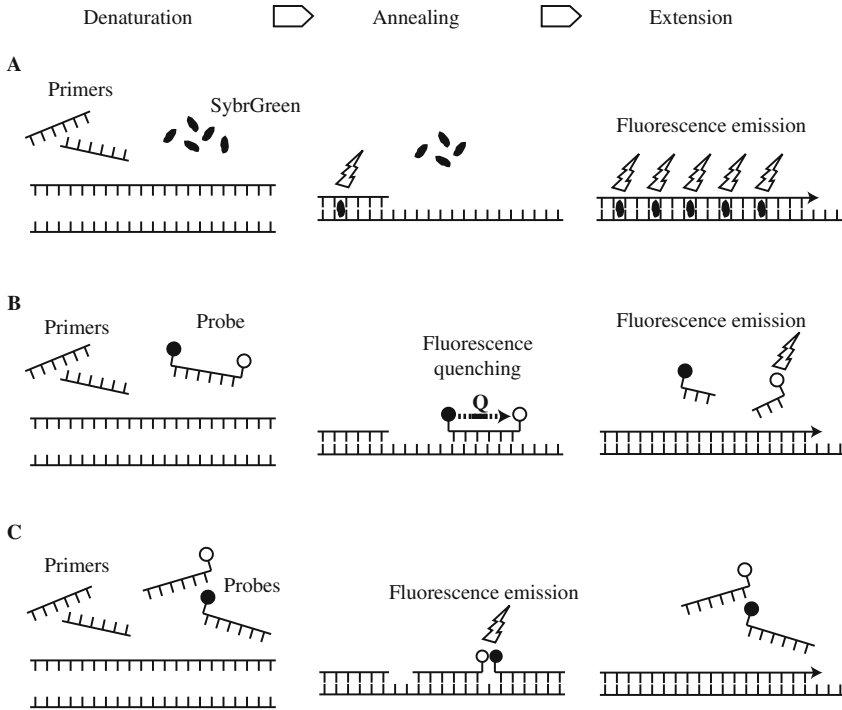


Fig. 15.6. Fluorescence formats most commonly used in real-time PCR. **(A)** SybrGreen. During denaturation, the fluorophore is free in solution and no fluorescence can be detected. During annealing, the SybrGreen begins to bind, giving maximal fluorescence at the end of extension. **(B)** Hydrolysis probes. The fluorescence of the fluorophore is inhibited by the nearby quencher. After annealing and during extension, the probe is hydrolysed by the DNA polymerase, thereby releasing the fluorophore which emits at its specific wavelength. **(C)** Hybridisation probes. The two probes hybridise near to one another on the target sequence and energy is transferred by resonance between the donor and acceptor fluorophores

Fluorescent Nucleic Acid Probes

The probes used for real-time PCR are oligonucleotides of different sizes coupled to one or more fluorophores. Depending on the situation, one or two probes are needed to detect fluorescence. Whatever the type of probe, the fluorescence emission is conditioned by the way the probe hybridises with the target sequence. Probes are more specific than the intercalators discussed above. On the other hand, they are less sensitive since a single fluorophore molecule binds to each strand of newly synthesised DNA. Probes cannot replace the conventional primers needed for amplification and must be able to hybridise with the target sequence within the region bounded by the primers (except in the case of scorpion probes). A great deal of work is being done on the chemistry of these probes and many different systems are currently on

Table 15.1. Excitation and emission wavelengths of different fluorophores and quenchers used with TaqMan probes

	Absorption [nm]	Emission [nm]
Fluorophore		
FAM	492	515
TET	521	536
JOE	527	548
HEX	535	556
TAMRA	555	580
ROX	575	602
Cy3	552	565
Cy5	651	674
Cy7	743	767
Quencher		
Methyl Red	410	–
ElleQuencher	650, 600	–
TAMRA	555	580

offer. We shall only be concerned here by the most representative systems on the market.

Hydrolysis Probes

The so-called TaqMan hydrolysis probes are historically the most widely used. The idea is to use a single probe carrying a fluorophore at 5' and a fluorescence absorber or quencher at 3'. The quencher can absorb the fluorescence emitted by the fluorophore and dissipate it in the form of heat. During the annealing stage, the probe hybridises in a specific way at the same time as the PCR primers (which requires a similar annealing temperature for both primers and probe). During extension, the exonuclease activity of the DNA polymerase hydrolyses the probe and releases the fluorophore at 5'. The latter, once it has moved far enough away from its quencher, will emit a characteristic fluorescence signal that can be measured at the end of extension (see Fig. 15.6B). In this case, it is therefore the number of fluorophore molecules released from the probe that is proportional to the amount of amplified target DNA.

Several types of fluorophore and quencher are available on the market and one is free to choose suitable excitation and emission wavelengths (see Table 15.1). It should be noted that some fluorophores such as TAMRA can also be used as a quencher. Multiplex PCR is achieved by using two different fluorophores on probes specific to two different targets (see Sect. 15.2).

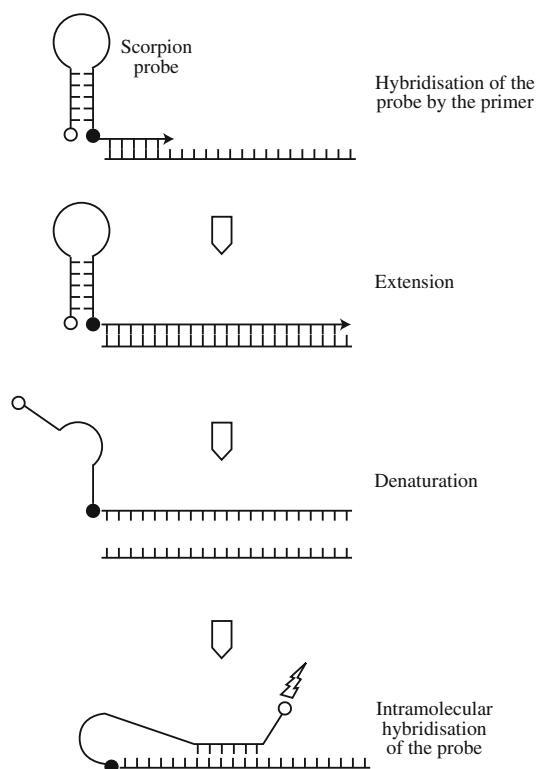


Fig. 15.8. Using the scorpion probe. The probe hybridises covalently with the quencher via the forward primer. After extension, the new DNA strand is bound to the probe, which folds over onto the neighbouring target sequence, thereby releasing the fluorophore and its quencher

and the TaqMan probes, but at the end of hybridisation, where fluorescence is maximal.

Hybridisation probes can be used both to quantify nucleic acids and to search for point mutations. The approach used for genotyping is to combine probes with post-PCR melting curves. In this case, one is not interested in the denaturation of the target itself, but rather in the dehybridisation of the probes from the target. The probes must anneal with the sequence carrying the relevant mutation. If the mutation is present, the probe anneals and creates a mismatch, which leads in turn to a drop in the melting temperature. This change in T_m can be detected by producing the post-PCR melting curve. This method is unable to detect new mutations, but provides extremely fast and reliable genotyping of known mutations (see Sect. 15.3). The fact that there is no post-PCR manipulation minimises the risk of contamination and makes this technique particularly attractive for diagnostics, especially when used with high-throughput instrumentation.

Molecular Beacons

Molecular beacons (MB probes) are probes with hairpin structure which, like hydrolysis probes, carry a fluorophore at 5' and a fluorescence absorber or quencher at 3' (see Fig. 15.7). In the non-hybridised state, the probe is folded owing to the partial complementarity of the sequence. In this configuration, the fluorophore and quencher are close to one another and no fluorescence is detected. During the annealing stage, the nucleotide sequence forming the loop anneals onto the target sequence, thereby moving the fluorophore and its quencher further apart. Molecular beacons are conventionally used to quantify nucleic acids, but post-PCR melting curves can also be made for genotyping applications.

Scorpion Probes

Despite their similar structure to molecular beacons, these probes differ from those described so far, because they play the double role of probe and PCR primer, thus serving as intramolecular markers. They have a hairpin structure, in which the fluorophore and quencher are therefore juxtaposed in the unhybridised state. The probe is bound covalently via the quencher to the 5' end of the forward primer (see Fig. 15.8). After extension, the newly synthesised strand is permanently bound to the probe. After denaturation, there is an annealing stage at a temperature which favours intramolecular hybridisation of the probe onto the neighbouring target sequence. During this phase, the probe acquires the so-called scorpion conformation, wherein the fluorophore and its quencher are moved apart. This type of probe can be used to quantify nucleic acids, but also to detect single-base mutations.

15.2 Implementing Quantitative Real-Time PCR

In this section, we discuss a basic application of quantitative PCR, i.e., one in which an amount of template is measured relative to reference samples. This quantification can be made absolute provided that the reference samples are themselves absolute. At the end of the chapter, we shall explain how the ideas discussed below can be applied to other areas of real-time PCR, in particular genotyping (see Sect. 15.3).

The key point when setting up a quantitative real-time PCR experiment is the specificity of the amplified product. The experiment cannot strictly speaking be described as quantitative unless a single type of DNA fragment is amplified in each tube. One must therefore use the hot start method described below. When setting up such experiments, whatever the fluorescence format to be used for subsequent experiments (hybridisation probe, hydrolysis probe, etc.), one must begin by a preparation using a fluorescence marker that labels all amplified DNA molecules, such as SYBR Green (see Sect. 15.1.3).

Setting up the experiments with sequence-specific fluorescence probes would be a mistake, because it would then be impossible to ascertain all the polymerisation reactions occurring in each tube. The preparation discussed here is thus described using SYBR Green to monitor the amounts of amplified DNA.

Hot Start PCR

In PCR, during the first temperature rise, since the sample has not yet been denatured, it is not unusual for the primers to be hybridised either on top of one another or together. They may then constitute primed templates on which the DNA polymerase will act to alter their specificity, with the consequence that unwanted amplification products will turn up later on. For this reason, PCR users realised that the specificity of PCR could be improved by adding the DNA polymerase after the initial denaturation. This is the basis of the hot start method. Today, there are two main variants. The first uses a DNA polymerase complexed with a specific antibody which blocks its activity. It thus remains inactive during the first temperature ramp. The initial denaturation stage lasting 3 min is sufficient to completely denature the antibody and release the active enzyme. The second variant exploits the fact that the pH of all biological buffers depends on the temperature. The polymerase is chemically modified by adding chemical groups that deactivate the enzyme. These groups are chosen in such a way that their bond with the enzyme is stable at the pH of the buffer at low temperature, but unstable at the pH it has at 95°C. The enzyme is then activated during the denaturation stages which, for this reason, must be prolonged. By abuse of language, one speaks of enzyme self-activation.

15.2.1 Denaturation and Amplification Curves

A real-time PCR experiment comprises three successive stages (see Fig. 15.9): an initial stage to denature the samples (95°C for hot start, with activation of the DNA polymerase), an amplification stage with repeated denaturation/annealing/extension cycles at temperatures 98°C, 65°C, and 72°C, respectively, and a final phase of gradual denaturation which serves to ascertain the quality of the amplified DNA molecules.

At the end of each extension cycle, the fluorescence of each sample is measured, which amounts to measuring the amount of amplified DNA. Figure 15.10 shows the amplification results with an initial range of concentrations of the template to be amplified: a sample and three successive ten-fold dilutions. For each sample, four stages are observed. In a first stage, the fluorescence signal cannot be distinguished from the background noise. Then in a second stage lasting a few cycles, the signal grows exponentially, before increasing linearly in the third stage, and reaching a plateau in the fourth. The curves shown in Fig. 15.10 are in fact corrected curves: at each measured value, one subtracts the value measured in the first cycle. This is known as arithmetic correction.

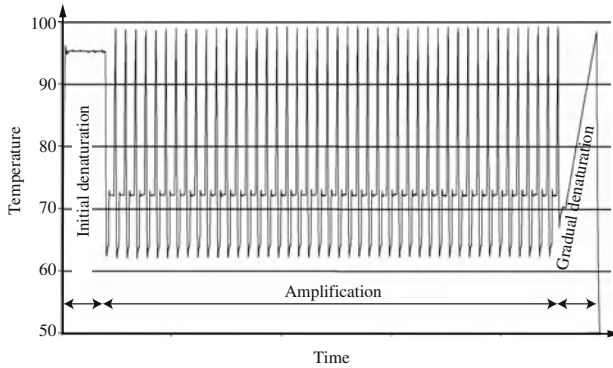


Fig. 15.9. Time-dependence of the temperature during a real-time PCR experiment

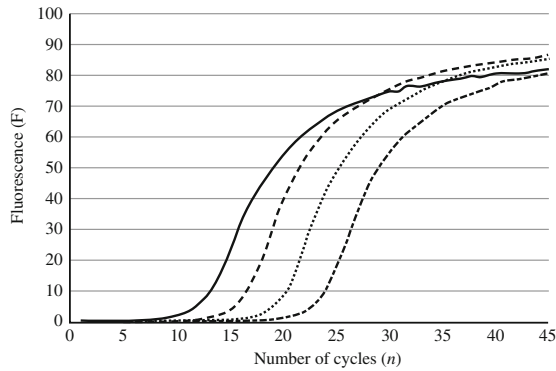


Fig. 15.10. Amplification phase. Variation of the fluorescence with the number of cycles during a real-time PCR experiment

Selecting the Annealing and Denaturation Temperatures

The annealing temperature is a parameter that must be adjusted to suit each pair of primers (see Sect. 15.2.2). In the experiment shown in Fig. 15.1, the annealing temperature is 65°C.

The denaturation temperature generally used is 95°C, but for some templates that are particularly rich in GC, a temperature of 98°C must be used. The incubation time in the different stages depends both on the chosen enzyme system and on the type of equipment used. If the chosen hot start system consists of a polymerase coupled to an antibody, an initial denaturation lasting 3 min is enough to fully activate the DNA polymerase, but if the chosen system uses self-activation, the initial denaturation stage must be longer, and longer denaturation times must be added in each cycle. The times must also be adapted to suit the instrument, because temperature transfer times can vary significantly from one device to another (see Sect. 15.1.2)

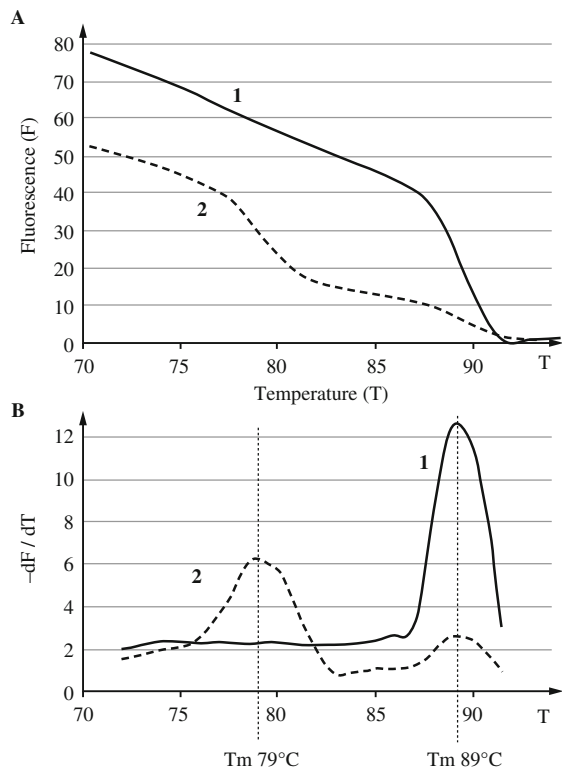


Fig. 15.11. Denaturation stage. Variation of the fluorescence with temperature T during a real-time PCR experiment

During the gradual denaturation stage (see Fig. 15.9), double-stranded DNA samples are steadily heated and the fluorescence is monitored continuously. Figure 15.11A shows the temperature dependence of the fluorescence. In the simple case where a single DNA fragment has been amplified (see Fig. 15.11A, curve 1), following a slow drop in fluorescence due to the change in pH caused by the temperature variation, a sudden drop in fluorescence is observed, reflecting the denaturation of the amplified double-stranded DNA molecules. For this reason, one speaks of melting curves. The point of inflection of the curve during this denaturation, where the derivative has a maximum (see Fig. 15.11B), is taken to indicate the value of T_m for the amplified molecules. This T_m measurement for the amplified DNA molecules is the counterpart in real-time PCR of a measurement of the size of the amplified molecules in conventional PCR. If several types of product have been amplified (see Fig. 15.11, curve 2), it is easy to observe the two denaturation stages, and examination of the derivative reveals as many peaks as there are different amplified products, provided that they have different values of T_m (two different products are amplified in Fig. 15.11, curve 2).

Influence of pH on Fluorescence

Like many fluorochromes, the fluorescence intensity of SYBR Green complexed with DNA depends on the pH. The drop in pH, which is linear during the temperature ramp, induces a linear reduction in the fluorescence intensity of SYBR Green.

15.2.2 Optimising the Annealing Temperature: Specificity

When setting up a quantitative PCR experiment, one must first choose a pair of oligonucleotide primers for the PCR. Once these oligonucleotides have been synthesised, one must then test whether these primers are compatible with a robust quantification experiment. To do this, the first thing is to check that conditions can be found in which the primer pair amplifies only a single DNA fragment, and that the amplified DNA fragment is indeed the one required. One thus begins by testing the ability of this pair of oligonucleotides to prime the DNA amplification using different annealing temperatures (see p. 855). Typically, with 20-base oligonucleotides, one tests primer hybridisation temperatures in the range 60–70°C. These amplifications are carried out using these different hybridisation temperatures on DNA samples containing (positive controls) or not containing (negative controls) the relevant template. At this stage, a pair of primers is considered suitable if one can identify an annealing temperature for which amplification is only observed for the positive controls, and if the amplification generates only one specific product. In the case where several products are amplified, the denaturation curve looks like the dotted curve 2 in Fig. 15.11. It is commonly necessary to test several pairs of primers before finding a suitable one. If the pair is suitable, the amplified product is purified and one can then either analyse it on agarose gel, or sequence it, in order to find out whether it is indeed the required DNA fragment. If the sequence confirms that the amplified fragment is indeed the desired product, the preparation continues by determining the efficiency of the real-time PCR.

15.2.3 Determining the Amplification Efficiency

Figure 15.10 shows amplification curves in which the sample fluorescence is represented as a function of the number of PCR cycles. In such a representation, it can be seen that the period over which the amplification is exponential is relatively short, and that certain factors soon become limiting, with a slowing effect on the amplification. The latter becomes linear before finally levelling out. It is clear that, in order to carry out the quantification, the amplification reaction must be considered before any factor other than the initial template concentration can intervene. It is thus essential to work in the exponential stage of the PCR. In order to get a better picture of the exponential amplification stage, one must change the representation to a logarithmic graph of the fluorescence as a function of the number of cycles (see Fig. 15.12A). In

this figure, the ordinate is the amount of amplified DNA, which is directly proportional to the measured fluorescence, specified in arbitrary units. It can be seen that most of the curves are chaotic below the fluorescence level 1. This is due to the sensitivity threshold for the systems used to measure the fluorescence intensity. One is thus only concerned with the linear part of the amplification curves in this logarithmic representation. In the example shown, the curves are linear for 4 successive cycles, with fluorescence values in the range 1–10. For each sample, Fig. 15.12B shows the four corresponding points on the amplification curves, and a linear regression line for the four points.

Let us see how this representation is used to carry out the quantification. Let q_0 be the initial amount of template in a given sample. After n amplification cycles, one expects to find an amount Q of amplified product, where $Q = q_0 2^n$ (see p. 843). Now, in practice, no PCR experiment has efficiency 1. One cannot therefore assume that, in each amplification cycle, the amount of DNA will have exactly doubled. The relation between Q and q_0 thus has the form $Q = q_0(1 + \rho)^n$, where ρ is the PCR efficiency. As the fluorescence intensity is directly proportional to the amount of amplified DNA, this equation should describe the curves relating the fluorescence intensity to the number of amplification cycles. Taking the logarithm of both sides to accord with the representation in Fig. 15.12B, one thus obtains

$$\log Q = \log q_0 + n \log(1 + \rho) .$$

This shows that the slopes of the regression lines in Fig. 15.12B are directly related to the efficiency of the PCR. In fact, these lines are steeper for greater PCR efficiency, i.e., for ρ closer to unity. Quantification is only possible if, for the same experiment, all the regression lines turn out to be parallel, i.e., only if the reaction efficiency is the same for all the samples. As the efficiency is never exactly the same in all the samples of a calibration curve, it is more interesting to use another method to calculate the efficiency.

Consider the number of cycles required to reach a given fluorescence level in samples with different initial amounts of template. These points of intersection or crossing points are denoted by Cp and expressed in terms of the number of cycles. One starts once again with the relation $Q = q_0(1 + \rho)^n$, from which one can deduce the relation between n and q_0 for given Q :

$$n \log(1 + \rho) = \log Q - \log q_0 , \quad \text{whence } n = \frac{\log Q}{\log(1 + \rho)} - \frac{\log q_0}{\log(1 + \rho)} .$$

On the graph of Fig. 15.12C, we can then identify the four crossing points between the regression lines of Fig. 15.12B and the horizontal line with unit ordinate. This gives the graph of n as a function of q_0 shown in Fig. 15.12C. One then plots the linear regression curve for this graph, and its slope is a measure of the efficiency, viz.,

$$-\frac{1}{\log(1 + \rho)} .$$

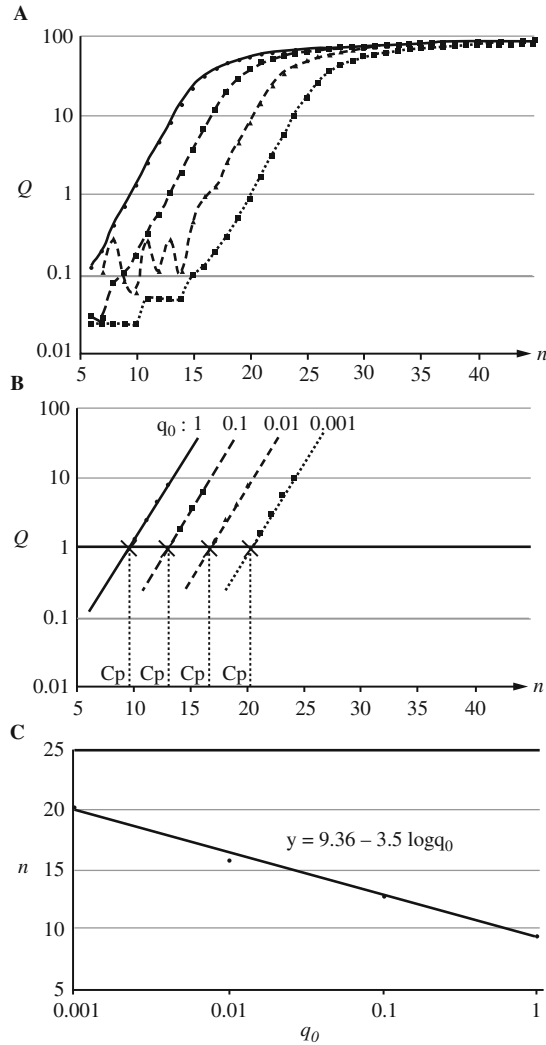


Fig. 15.12. Determining the PCR efficiency. (A) Amount of amplified product as a function of the number n of cycles. (B) Determining the crossing points C_p . (C) Correlation between the crossing points expressed relative to the number of cycles and initial amounts of DNA

With unit theoretical efficiency, one would obtain a slope of -3.32 . In the case shown, the slope is in fact -3.5 , corresponding to a PCR efficiency of 0.93, which is a good level. Experience shows that results are all the more reproducible as the PCR efficiency is closer to unity.

Pairs of primers for which experimental conditions cannot be found that give an efficiency greater than 0.8 cannot be used for delicate quantification

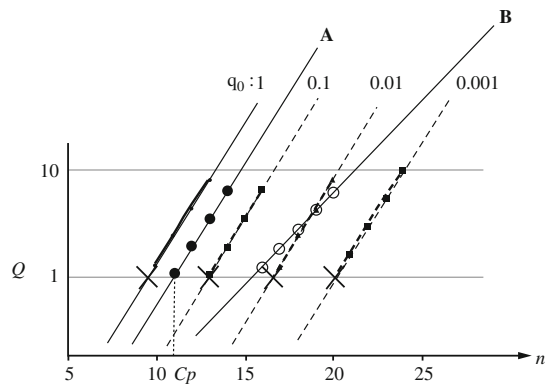


Fig. 15.13. Relative quantification

experiments. In general, the PCR efficiency can be modified and improved by altering the hybridisation temperature of the primers during the amplification cycles (but take care over the specificity!).

15.2.4 Relative Quantification

In quantitative PCR, quantification is not absolute but relative. Indeed, the amplification obtained on a given sample is compared with amplifications obtained for a calibration series.

Consider an experiment of the kind shown in Fig. 15.12, but in which, apart from the samples of the 10-fold dilution series treated here as a calibration system, there are two samples A and B with unknown initial template concentrations. After amplification, the first thing is to check that, in all the samples, the nature of the amplified DNA is the same (see Sect. 15.2.1). To do this, one first checks that the melting curves (like those shown in Fig. 15.11) are the same. In the experiment illustrated in Fig. 15.11, the two samples contained an equivalent initial amount of a DNA template that was supposed to give an amplified product with a T_m of 89°C measured by the melting curve. Sample 2 (curve 2 of Fig. 15.11) had a second template that could be amplified by the same pair of primers, giving an amplified product with lower T_m , at 79°C. It can be seen that amplification of the DNA fragment with $T_m = 79^\circ\text{C}$ significantly perturbed the amplification of the DNA fragment with $T_m = 89^\circ\text{C}$. Owing to this interference, quantitative PCR cannot be achieved if one amplifies several fragments simultaneously in the same sample. We shall see below that there are some exceptions to this rule.

For the quantification itself, one uses the kind of representation shown in Fig. 15.12B, including all the samples. Figure 15.13 shows the results. It can be seen that, for sample A, the amplification curve in this logarithmic representation is indeed parallel to the calibration curves. For sample B, the amplification curve is not parallel to those of the calibration system. The

PCR efficiency was not therefore the same for this sample as for the points of the calibration system, and quantification cannot be done on this point. Such behaviour generally reflects a problem with the quality of the sample. The problem can be solved either by purifying the sample again, or by diluting it sufficiently to ensure that the impurities no longer interfere with the PCR.

To determine the initial amount q_0 of the template in sample A, one determines the abscissa (in terms of the number of cycles) of the crossing point (Cp) between the corresponding regression line and the straight line of unit ordinate. This number is plotted in Fig. 15.12C and used to calculate the corresponding value of q_0 . Like any measurement, it will involve errors. No result can be taken seriously unless at least three repeats have been done and the results presented in the form of a confidence interval.

15.2.5 Multiplex PCR

The idea of multiplex PCR is all the more attractive in that, if fluorescence formats like hybridisation probes, hydrolysis probes, or other sequence-specific systems are used, each of the relevant molecules can be labelled by a different colour and the number of measurements per sample thereby increased (see Sect. 15.1.3).

As discussed above, it is difficult to do quantitative PCR if one hopes to simultaneously amplify several molecules in the same sample. Indeed, since the PCR quickly reaches limiting conditions (see below), the most abundant molecules in the initial sample will amplify correctly, but the less abundant molecules will not be amplified in sufficient amounts to be detected. The commonest use of multiplex PCR involves simultaneous amplification of several loci from genomic DNA in the context of genotyping experiments (see Sect. 15.3.2). By definition, all loci are originally present in similar amounts. If the pairs of primers used to amplify the different loci have similar efficiencies, these loci will be amplified to a similar degree and, with appropriate probes, several loci can be genotyped simultaneously.

However, conditions can be found in which multiplex PCR is possible with samples containing highly variable amounts of template to be quantified. The idea is then to arrange for conditions in which the amplification of the most abundant molecules ceases before conditions become limiting for the less abundant molecules. To do this, a limiting amount of primers is used to amplify the most abundant fragment. Due to its abundance, and despite the limiting amount of primers, the quantification on this fragment is accurate. Since the amplification of this first fragment has not exhausted the reaction medium, the amplification of the second fragment can also be correctly achieved.

Non-Specific Amplifications and Limiting Conditions

The reaction media for quantitative PCR are generally prepared so that the first component to become limiting in the reaction will be the nucleotides. They are prepared in such a way as to amplify an amount of material that will give a fluorescence

signal 50–100 times greater than the level above which one obtains a good concentration/response curve. Typically, for the examples discussed in this chapter, a level in the range 50–100. Under such conditions, once the relevant molecule has been amplified, it is no longer possible to amplify the other molecules in the sample. This is relatively important, because the PCR is so powerful that it is commonplace for a primer pair to end up, after a large number of cycles, by amplifying something other than the specific template. There are two ways to avoid the problems related to such non-specific amplifications: either find the perfect primer pair, or restrict the amount of nucleotides to stop the amplifications as soon as enough material has been amplified in good conditions for carrying out the assay. Since perfect primer pairs are hard to find, one generally opts to limit the amount of nucleotides in the reaction medium. In order to get around this problem when carrying out multiplex PCR, one can on the other hand increase the amount of nucleotides, but limit the amount of primer, so that the amplification of a given fragment is limited by consumption of the specific primers before the other components of the reaction medium become limiting. In this way, and without knowing in advance the relative abundance of the different templates, several fragments can be amplified. The disadvantage with this method is that limiting the primer concentration can perturb the quantification of small amounts of template, whence the repeatability of the results will leave something to be desired.

15.3 Applications of Real-Time PCR

15.3.1 Real-Time PCR for the Quantification of Viral Genomes

The quantification of DNA molecules by quantitative real-time PCR is useful in many different areas of medicine, such as virology, bacteriology, or cancerology. Depending on the situation, this quantification can be used for diagnosis, to establish a prognostic marker, or to monitor the effects of treatment. The contribution of real-time PCR in pharmacogenetics will be discussed in Sect. 15.3.2. In this section, we shall describe an application of real-time PCR in virology, concerning the diagnosis and monitoring of an infection by the hepatitis C virus.

This virus was first identified in 1989. Infection by this virus is a genuine public health problem, both in terms of the number of people contaminated (600,000–800,000 people in France) and also due to the strong tendency of this virus to cause chronic liver disease (cirrhosis), which can lead to liver cancer (hepatocellular carcinoma). The virus is transmitted parenterally (transfusions prior to 1990, drug use), sexually, and nosocomially (medical equipment contaminated during hemodialysis, endoscopy with biopsies, acupuncture, etc).

The hepatitis C virus (HCV) is a flavivirus with a genome comprising a single-strand RNA of about 10,000 nucleotides. It is characterised by high genetic variability, which depends on the part of the genome considered (see Fig. 15.14). For example, the 5' region of the viral RNA (5' NC) is conserved

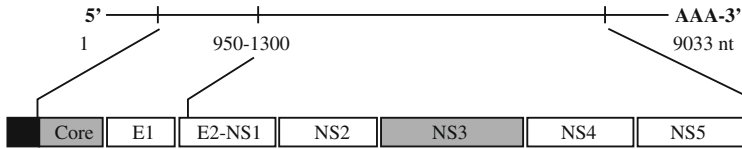


Fig. 15.14. Structure of the genome of the hepatitis C virus. With the kind permission of Dr J. Ducos, CHU, Montpellier, France

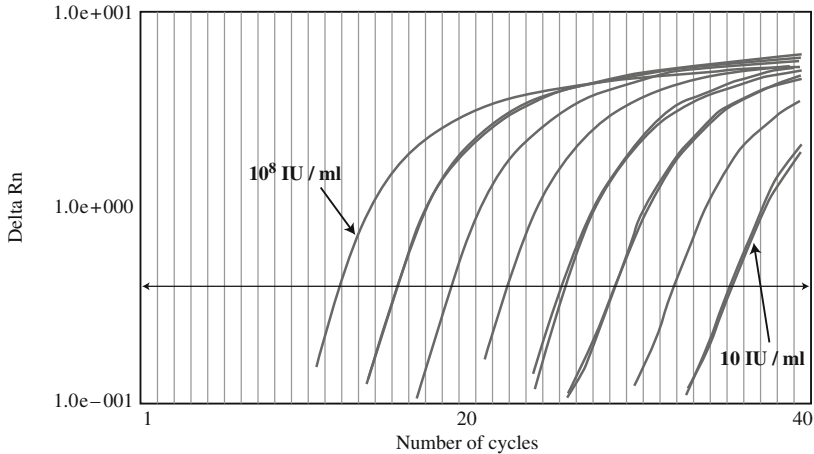


Fig. 15.15. Example of international standard calibration curves used in quantitative real-time PCR to assay the hepatitis C virus. With the kind permission of Dr J. Ducos, CHU, Montpellier, France

among the variants, while the E1 and E2/NS1 regions exhibit the highest level of variability. This genetic variability has led to the classification of HCV into 6 main genotypes, with very different response to the various antiviral treatments.

Acute HCV infection is clinically indiscernible in most cases. The RNA of the virus is the first detectable marker in the serum of the contaminated person, 7 to 21 days after infection. The anti-HCV antibodies produced by immune reaction appear somewhat later, 20 to 150 days after contamination. Recovery from acute infection is defined by absence of detectable HCV RNA in the patient's serum. So diagnosis of viral hepatitis C is based on serology, in particular, detection of anti-HCV antibodies in the serum, and molecular biology, i.e., detection of the viral genome by PCR and determination of the viral genotype by sequencing.

Quantitative real-time PCR has allowed the development of a sensitive method for assaying the HCV genome in the serum. Quantification of this viral genome requires complete extraction of the RNA in the serum of the tested person. This RNA must then be transcribed into DNA, because PCR cannot be used with an RNA template. To do this, there is an intermediate

stage called reverse transcription, in which a complementary DNA copy is synthesised from the RNA, using an enzyme of viral origin called reverse transcriptase. This complementary DNA is then used as DNA template for the quantitative PCR process. The primers chosen to amplify the genome of the hepatitis C virus are positioned on the most highly conserved portion (or core) of the viral genome, in order to recognise and assay all 6 viral genotypes (see Fig. 15.14). Quantification of the viral genome in an unknown serum sample is measured from the crossing point C_p (see Sect. 15.2) obtained by quantitative real-time PCR. This crossing point is compared with a calibration curve established by international standards, extending from 10 to 10^8 IU (international units) per mL (see Fig. 15.15).

The amount of viral genome (or viral load) in the serum is thus measured from the calibration curve and expressed in IU/mL of serum. Quantitative real-time PCR significantly improves the sensitivity for detecting the genome of the hepatitis C virus, currently of order 20 IU/mL, which corresponds to about 100 copies of the viral genome per mL, as compared with preexisting quantification techniques. The initial quantification of the viral genome has a diagnostic value, but it also affects the choice of therapy. Indeed, a viral load greater than 800,000 IU/mL is considered high and the patient will be treated with antiviral agents such as interferon or ribavirin. PCR quantitative assays of the viral load also play a role in monitoring the response to therapy. Hence, during antiviral treatment, which may last for different lengths of time depending on the genotype of the HCV, the effects of the therapy are assessed by measuring the viral load in the serum after one month of treatment, at the end of the treatment, and then again six months after the end of the treatment. Quantitative PCR detection of the viral genome after 6 months signals a relapse. On the other hand, if the viral genome is not detected by quantitative PCR two years after the end of the treatment, the patient is considered to be permanently cured.

Measurement of the HCV viral load by quantitative real-time PCR is now a routine component of medical examination in virology, supplying essential information for diagnosis and therapy. Furthermore, the use of international standards as calibration curves for these assays has made it possible to standardise and compare results obtained in different medical laboratories. The routine use of this technique in hospitals depends on strict control of the various stages in the quantitative PCR process, between extracting the RNA from the patient's serum to carrying out the quantitative real-time PCR. This has recently been facilitated by automation of these steps, thereby avoiding contamination during PCR and improving the reproducibility of the results.

Other applications of real-time quantitative PCR for DNA assays have been devised in virology and other areas of medicine, and this quantitative approach will certainly be further developed in medical analysis in the years to come.

15.3.2 Real-Time PCR in Pharmacogenetics

Introduction to Pharmacogenetics

Pharmacogenetics is the study of the genetic mechanisms responsible for the way different individuals respond to a given medical treatment. The variability of this response, from the point of view of the efficiency or toxicity of the treatment, is a major public health problem. Indeed, it is estimated that 100,000 deaths per year in the United States are due to the side effects of medicines, representing the fourth largest cause of mortality, and that the cost of this iatrogeny is something like 100 billion dollars. In France, the *Agence Française de Sécurité Sanitaire des Produits de Santé* estimates that some 128,000 hospital admissions (with a cost of around 320 million euros) can be attributed to ill-suited prescription of medicines. Pharmacogenetics is concerned more precisely with medicines that are said to have a narrow therapeutic margin, i.e., overdose leads to particularly harmful consequences, and medicines for which patient response exhibits a high level of variability from one individual to another. This is the case for anticancer chemotherapy and drugs used in psychiatry.

Differences in sensitivity to drugs are often due to genetic variations in metabolism and transport, as well as pharmacological targets. For example, the quantity and functionality of enzymes responsible for biotransformations, as for other proteins in the organism, largely depend on the genetic information that codes them. Now these genes can exhibit a certain level of variability in their sequence from one individual to another. This is referred to as genetic polymorphism. These variations in the sequence may be point mutations, i.e., single nucleotide polymorphisms (SNP), partial or total deletions, or gene amplifications (several copies of the same gene). Such variations will lead to the quantitative or qualitative phenotypic changes (see Fig. 15.16) that underlie the interindividual differences in response.

Up until recently, there were only a few phenotyping methods, such as measurement of enzyme activity or the use of test substrates, for predicting the therapeutic or toxic response to a medicine. Advances in molecular biology and the identification of genetic polymorphisms involving phenotypic variability have led to the development of new genetic tests that are faster, easier to reproduce, and above all well suited to high-throughput screening using modern techniques like DNA chips and real-time PCR. The prerequisite for using such tests is to ensure that the chosen genotype test can predict a given phenotype with certainty. This requires preparatory studies of the correlation between genotype and phenotype, which remains a major field of investigation in medical biology.

Applications in Pharmacogenetics

The use of real-time PCR in pharmacogenetics can be exemplified by discussing the genes for dihydropyrimidine dehydrogenase (DPD) and cytochrome

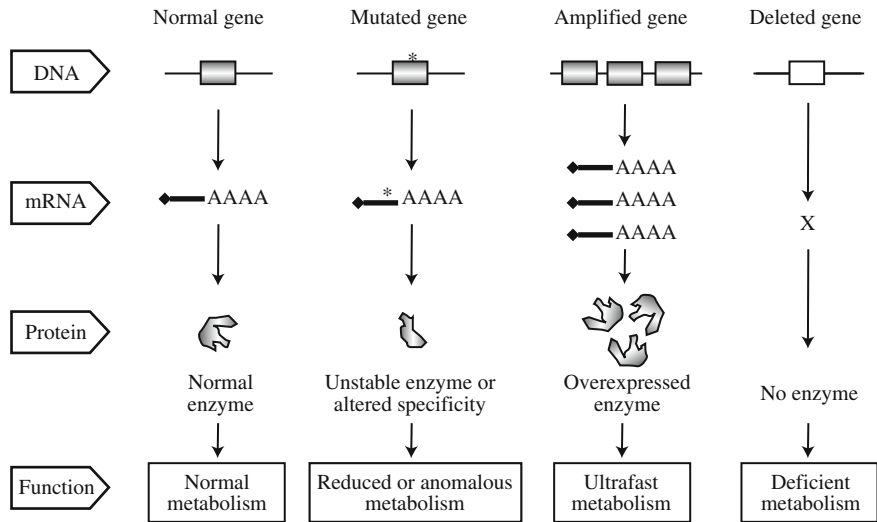


Fig. 15.16. Consequences of different genetic polymorphisms on the expression of enzymes responsible for the biotransformation of drugs and the functional impact on the metabolism

P450 2D6. These two examples will illustrate the relevance of this technique for the two main applications, viz., the search for point mutations and the quantification of DNA. A prerequisite for any genotype analysis is to obtain easily purifiable genomic DNA from any cell in the organism. It is traditional to use the lymphocytes in a blood sample or saliva from the mouth.

Dihydropyrimidine Dehydrogenase (DPD)

Dihydropyrimidine dehydrogenase is an enzyme found essentially in the liver and lymphocytes. Its physiological role is to transform uracil, a pyrimidine base DNA precursor, into dihydrouracil. It also transforms 5-fluorouracil into its inactive metabolite, 5-dihydrofluorouracil. 5-fluorouracil is one of the most widely used anticancer agents in the world. After administering it to a cancer patient, 80% of the dose will be degraded by DPD. Now it turns out that some people with a DPD deficiency develop very severe, even fatal toxic reactions after administration of 5-fluorouracil. It is thus particularly important to identify patients with a DPD deficiency before treatment, to avoid a major toxic reaction. Pharmacogenetic studies have shown that a DPD deficiency is often related to a point mutation located at the splice site of exon 14. (The exons constitute the coding part of the gene, and splicing is the biological mechanism whereby the exons are assembled to form messenger RNA.) The presence of this mutation (a guanine is replaced by a cytosine) perturbs the splicing process and leads to an mRNA which lacks exon 14 (referred to as exon skipping) and a truncated enzyme that cannot function correctly (see Fig. 15.17A).

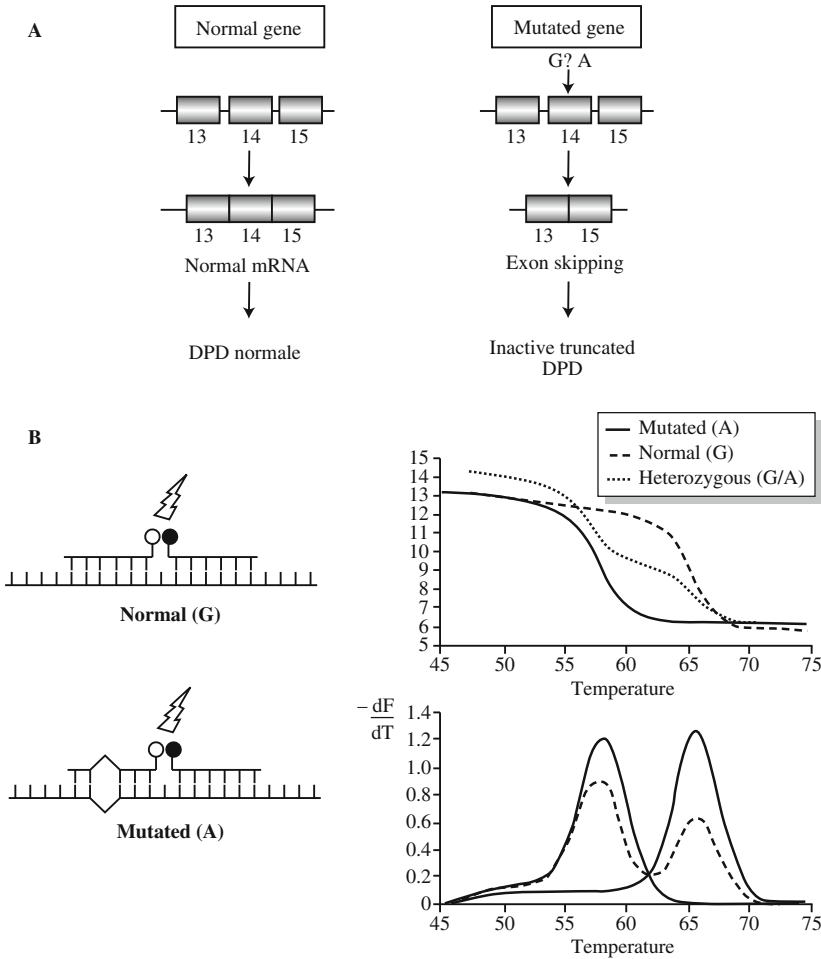


Fig. 15.17. (A) Consequence of the mutation $G \rightarrow A$ of exon 14 for dihydropyrimidine dehydrogenase (DPD). The mutation suppresses a splice site and leads to a non-functional truncated protein. (B) Genotyping the mutation by real-time PCR. One of the two marked probes hybridises with the mutation to be detected. The presence of the mutation causes a mismatch which reduces the melting temperature. Comparing the values of T_m for the mutated and wild-type sequence, a homozygous or heterozygous mutation can be detected. (The values of T_m are more easily calculated from the curves for the derivatives $-dF/dT$)

Today one can envisage systematic genotyping of DPD in all patients treated with 5-fluorouracil using real-time PCR, by making a post-PCR melting curve with allele-specific fluorescent probes. Figure 15.17B shows an example of genotyping with hybridisation probes using FRET (see p. 849). One of the two probes overlaps the sought mutation. The presence of the mutation leads to a mismatch with the probe and a reduction of the melting

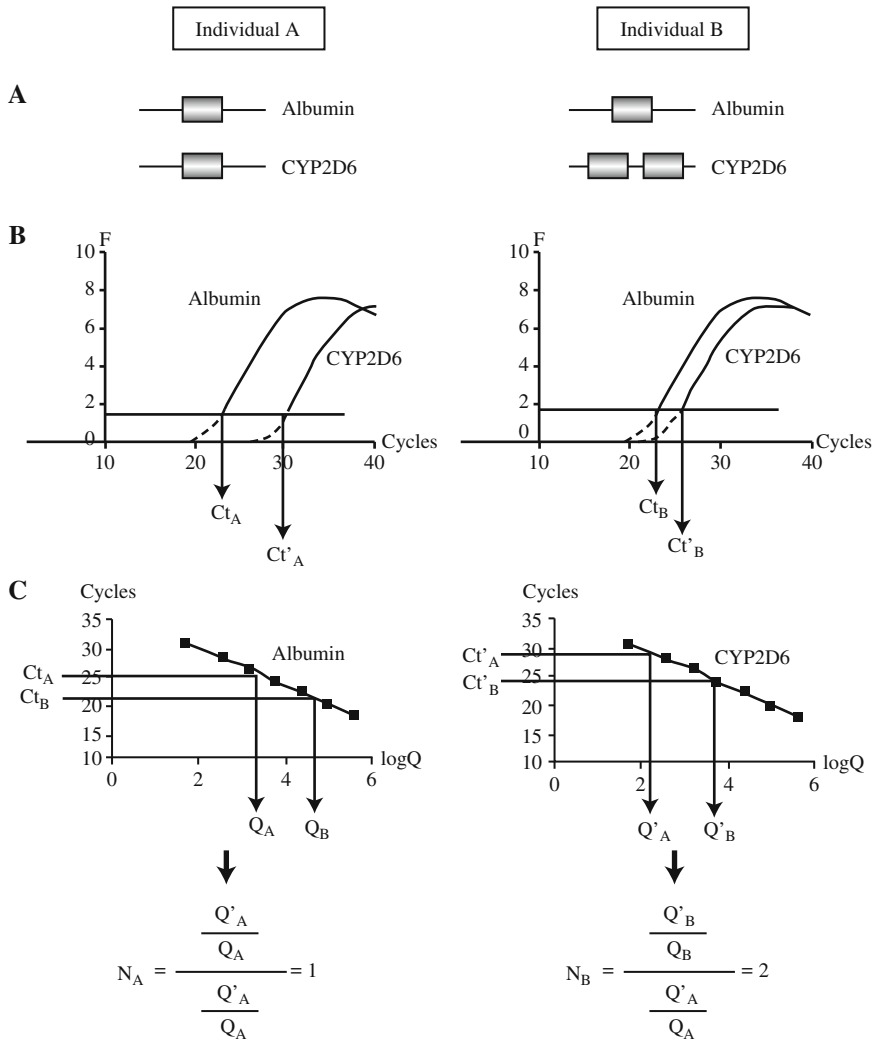


Fig. 15.18. Relative quantification of the gene for cytochrome P450 2D6 using real-time PCR. **(A)** The individual B to be tested is compared with a reference individual A in whom the gene is not duplicated. The gene for albumin serves as an internal standard for normalising the initial amounts of DNA. **(B)** The amplification curves are used to obtain the values of Ct. **(C)** The corresponding amounts of DNA are obtained from the calibration curves for the two genes

temperature. Indeed, the probe detaches more easily when the temperature increases and the detected fluorescence falls off rapidly. This fast and reliable technique can detect homozygous mutations (two mutated alleles) and heterozygous mutations (a single mutated allele).

Cytochrome P450 2D6 (CYP450 2D6)

The P450 cytochromes (CYP450) form a family of phase I hepatic enzymes responsible for the biotransformation of many substances foreign to the organism (xenobiotics). CYP450 2D6 is responsible for metabolising many drugs, such as antidepressants, morphines, and antiarrhythmics, and its level of expression thus influences the therapeutic or toxic response. The gene coding for CYP450 2D6 has many polymorphisms, and individuals can be classified into slow, intermediate, and ultrafast metabolisers. One commonly observed genotypic variation is a duplication of the gene, leading to an overexpression of the enzyme.

Figure 15.18 shows an assay for the gene for CYP450 2D6 by real-time PCR. This is a relative quantification like the one described in Sect. 15.2. The aim is to compare an individual B of unknown status with a normal individual A, for whom the relevant gene is not duplicated. To do this, an internal standard is needed to account for any difference in the initial amount of DNA between the two samples. One therefore uses a second, non-polymorphic gene, always present as a single copy in all individuals, e.g., the gene for albumin (see Fig. 15.18A). The amplification curves for these two genes and for the two individuals are used to obtain the values of C_t , the crossing threshold, discussed in Sect. 15.2 (see Fig. 15.18B). These values are then compared with the previously established calibration curves, from which the corresponding amounts of DNA can be deduced (see Fig. 15.18C). The number N of genes for CYP 2D6 is then expressed relative to the reference individual, for whom $N = 1$. This technique for assaying the gene can identify duplications ($N = 2$), but also homozygous deletions ($N = 0$), or heterozygous deletions ($N = 0.5$).

References

1. Bustin, S.: Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays, *J. Mol. Endocrinol.* **25**, 169–193 (2000)
2. Bustin, S.: Quantification of mRNA using real-time reverse transcription PCR (RT-PCR): Trends and problems, *J. Mol. Endocrinol.* **29**, 23–39 (2002)
3. Eurogentec: www.eurogentec.com. Technical resources, documentation, your one-stop-shop real-time PCR supplier

Biosensors. From the Glucose Electrode to the Biochip

L. Blum and C. Marquette

The biosensor was born over forty years ago, when Clark and Lyons [1] had the idea of carrying out specific glucose concentration measurements by detecting the oxygen consumed during the enzymatic oxidation of this metabolite, catalysed by glucose oxidase, using an electrochemical sensor. The enzyme was used in solution, confined to the end of the sensor. In parallel, during the 1960s, more and more studies were being carried out on the properties of immobilised enzymes and their use. Then in 1967 an enzyme electrode for specific glucose assays was described by Updike and Hicks [2]. In this work, glucose oxidase was incorporated in a polyacrylamide gel. In 1969, by immobilising urease in the same type of matrix, Guilbault and Montalvo [3] described the first

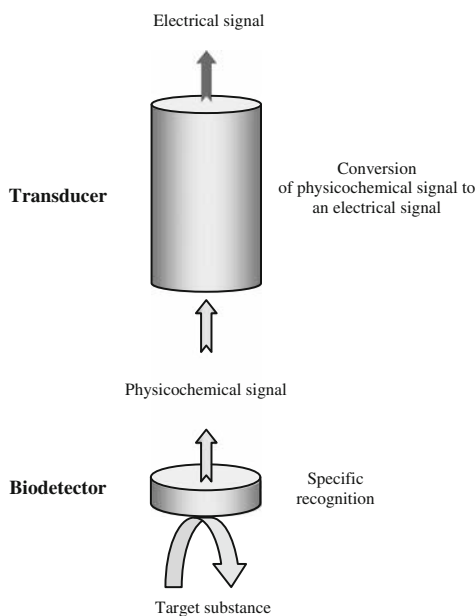


Fig. 16.1. General arrangement of a biosensor

enzyme electrode associated with a potentiometric measurement for assaying urea by detecting ammonium ions (NH_4^+) produced during the reaction.

This founding research opened the way to many developments associating not only enzymes and electrochemical detection, but also other types of biomolecules and other detection methods [4–6]. A biosensor can thus be defined more precisely as any measurement device in which the sensitive component is of a biological nature. The sensitive biological element, or bioreceptor, is able to specifically recognise a target substance present in a complex medium. This specific recognition generates a physicochemical signal which is then transformed by a transducer into a measurable and interpretable electrical signal (see Fig. 16.1).

The bioreceptor is usually immobilised and in contact with the transducer, thereby constituting the sensitive layer of the biosensor. It acts as an extremely sensitive filter as compared with a chemical sensor. Biosensors are thus a powerful tool for highly specific analysis. On the other hand, the intrinsic fragility of the biological element imposes mild conditions of use, especially with regard to the temperature, pH, and ionic strength.

Depending on the kind of physicochemical signal generated by the molecular recognition, various forms of transducer can be used:

- electrochemical transducer (electrodes for amperometric or potentiometric detection, field-effect transistors),
- mass transducer (piezoelectric quartz),
- thermal transducer (thermistor),
- optical transducer (optical fibres and optoelectronic systems).

16.1 Bioreceptors

In theory, any biochemical or biological structure capable of specific recognition can potentially be used to form the sensitive layers in the biosensor. Historically, it was mainly from protein structures like enzymes and antibodies that the first biosensors were constructed. Subsequently, other molecular systems were used, in particular, nucleic acids (DNA or RNA), the latter being used for the design of biochips (see Sect. 16.7).

Although some devices incorporating membrane detectors, subcellular organelles, or finely cut pieces of biological tissue have been reported, this work remains marginal and many such projects have never gone beyond academic status. Indeed, when a biological entity is isolated from its natural environment, it becomes extremely fragile, hence unstable, and conditions for manipulation are sometimes incompatible with any kind of routine use. But note that microorganisms, mainly bacteria and yeasts, can play the role of the sensitive element in a biosensor.

Under a wider definition of the biosensor, artificial structures might be included, which would be qualified as biomimetic, in the sense that they are

designed to integrate the recognition function of certain biomolecules. These are mainly catalytic antibodies, molecularly imprinted polymers, and artificial receptors.

16.1.1 Natural Bioreceptors

Protein Structures: Enzymes and Antibodies

Proteins are natural macromolecules consisting of a chain of amino acids bound together by amide (peptide) bonds. There are twenty main amino acids from which proteins can be built up. They carry a carboxyl group and an amine group on the same carbon atom. The two other substituents of this carbon atom are a hydrogen and a side chain whose nature determines a particular amino acid. The molecular weights of amino acids vary between 75 for glycine and 204 for tryptophan. Since the number of amino acids in a protein can vary between about fifty and several thousand, the molecular weight of a protein can be anything from a few thousand to a few hundred thousand.

The nature and ordering of the amino acid chain in a protein constitute its primary structure. Some amino acids may only occur once in a given protein, or not at all, while others may appear many times. The possibilities for the primary structure are almost infinite. As an example, most bacteria produce some 3,000 different proteins, while the human organism produces several tens of thousands. In reality, each protein has a 3D structure determined by the spatial arrangement of the polypeptide chain. This organisation constitutes the secondary and tertiary structures, which are stabilised by a certain number of covalent bonds (disulfide bridges) and non-covalent bonds (hydrogen bonds, hydrophobic interactions, ionic interactions, and van der Waals interactions). Finally, some proteins have several polypeptide chains, each with their own primary, secondary, and tertiary structure, but associating together to form a quaternary structure. The spatial arrangement of a protein plays a dominant role in its function, and in particular in its specific recognition properties, if it has any.

Enzymes have catalytic activity. Almost all the reactions occurring in living organisms are catalysed by enzymes. Several thousand different enzymes have been found in different organisms. The reagents in a reaction catalysed by an enzyme are called substrates. During such a reaction, the substrates bind temporarily with the enzyme to form a complex. The 3D structure of these biological catalysts is characteristic of the given enzyme. The conformation is such that the enzyme encloses a pocket called the active site, in which the substrates can enter to interact with the enzyme. There is then a change in the conformation of the enzyme, allowing other interactions to set up with the substrate and facilitating its transformation into a transition state. Once in the transition state, the substrate is transformed into a product which detaches from the enzyme. The latter then resumes its initial conformation, absolutely intact and ready to engage in a new catalytic cycle. The formation

of the specific enzyme–substrate complex distinguishes enzymes from other catalysts. An enzyme only catalyses one type of reaction (oxidation, hydrolysis, etc.), and its action is usually limited to a single compound (substrate) or a single family of compounds. It is indeed this specificity which explains why there are so many different enzymes to be found in living organisms. Of course, many different reactions are necessary for life and hundreds of different compounds are involved in these reactions. Furthermore, these proteins have a much more efficient catalytic activity than chemical catalysts.

Antibodies (Ab) are specialised proteins which specifically recognise an antigen (Ag), i.e., a protein or other macromolecule that is foreign to the organism, and form an antigen–antibody complex. Antibodies are secreted by specialised cells in vertebrates and constitute the front line of the immune defense system. Antibodies have a molecular weight of around 150,000 and comprise 4 polypeptide chains, arranged in a Y-shape. They have two binding sites for the antigen, one on each branch of the Y. Antibodies are highly specific for the compounds that incite their formation, which explains why they are so useful as reagents for analysis.

Whole Cells

Whole-cell biosensors exploit the fact that certain strains of microorganism preferentially metabolise some particular compound (usually organic), and that this assimilation is accompanied by the production and excretion of a substance that can be detected by a transduction system. Microbial biosensors generally lack sensitivity compared with enzyme biosensors, due to the existence of several enzyme systems within each cell, and their response time is usually quite long (several tens of minutes). However, for some applications, a global measurement is more useful than a specific measurement, and in this case, whole-cell biosensors are particularly well suited. This is especially true for environmental applications, where it may be better to assess the overall toxic effect of a sample rather than trying to detect the presence of some specific toxin.

16.1.2 Artificial Bioreceptors

Catalytic Antibodies

Although antibodies with a natural catalytic activity have been identified in certain pathologies, most catalytic antibodies (abzymes) used to design biosensors have been produced artificially [7]. The main approach for obtaining catalytic antibodies is to produce antibodies directed against an analogue compound of the transition state for a given reaction. Among the different antibodies obtained, some will have both the specific recognition capabilities of conventional antibodies and a catalytic activity that allows them to transform the target substance (the antigen) by a process similar to the enzyme

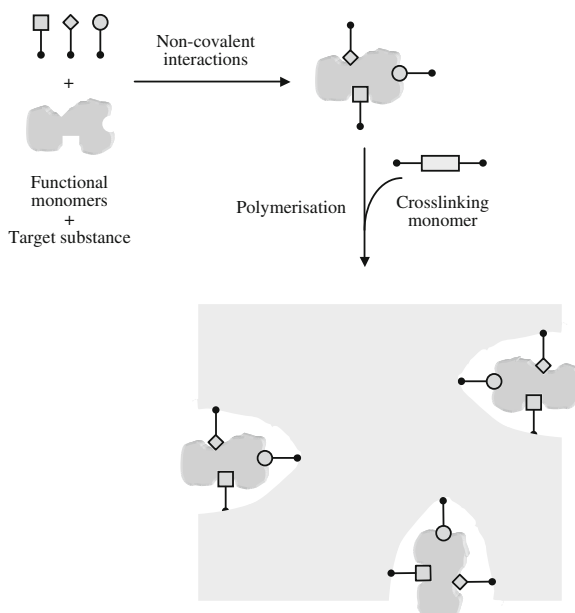


Fig. 16.2. The main steps in the preparation of a molecularly imprinted polymer

reaction and with comparable efficiency. The idea is simple enough, but there are certain problems with implementing it, especially regarding the design of structures mimicking the transition state.

Molecularly Imprinted Polymers

Molecular imprinting consists in creating receptor sites specific to some target substance in a synthetic polymer [8]. Molecular imprints, complementary images of the chosen target, are produced by setting the target or some analogue of it in contact with a mixture of monomers carrying functional groups able to bind to the target by non-covalent interactions, e.g., hydrogen bonds, hydrophobic interactions, ionic interactions. The target substance thus acts as a template around which the polymer will form.

Polymerisation is achieved in the presence of a crosslinking monomer in order to obtain a rigid 3D structure incorporating sites complementary to the target substance. After eliminating the target, a molecularly imprinted polymer is obtained that can specifically recognise the target substance (see Fig. 16.2).

Such systems are often made to design sensors and analytical systems for which it is difficult to find or obtain natural bioreceptors in the form of antibodies or enzymes. This is an attractive approach, but the results are sometimes disappointing. The main difficulty lies in a judicious choice of monomer for polymerisation.

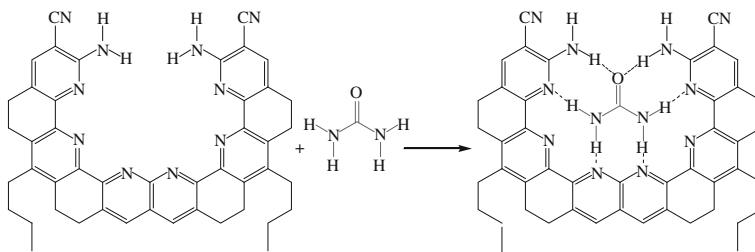


Fig. 16.3. Synthetic receptor for urea ($\text{H}_2\text{N}-\text{CO}-\text{NH}_2$). The urea binds specifically through hydrogen bonds onto the receptor, which has a complementary structure [9]

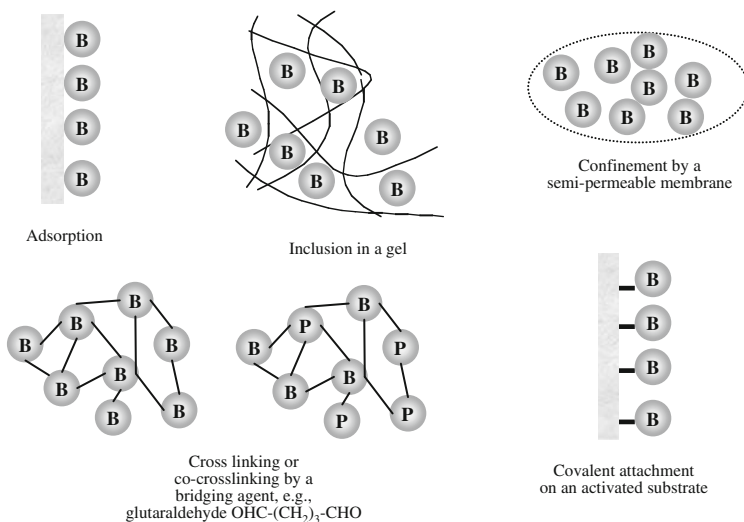


Fig. 16.4. Different ways of immobilising a bioreceptor B. P is an inactive protein like bovine serum albumin

Artificial Receptors

Artificial receptors are organic compounds obtained by synthesis in such a way that their structure contains a site complementary to the target substance. Target–receptor interactions are generally hydrogen bonds, and in the ideal case only the target substance will be able to bind. As an example, a synthetic receptor for urea is shown in Fig. 16.3.

16.1.3 Using Ligand–Receptor Systems

Molecularly imprinted polymers, artificial receptors, or antibodies form a ligand–receptor complex with their target substances, without producing a

physicochemical signal, in contrast to enzymes. The specific formation of such complexes can nevertheless be detected qualitatively and quantitatively, by either direct or indirect methods.

Indirect methods involve a second molecular system, biological or otherwise, able to emit a physicochemical signal which specifically indicates the formation of the ligand–receptor complex. For example, in the case of conventional immunoassays, the formation of the Ag–Ab complex is revealed via a bioconjugate, either made catalytic by coupling to an enzyme, or labelled by a light-emitting probe or a radioactive element. In the same way, in an immunosensor, i.e., a biosensor designed to carry out immunoassays, it is the bioconjugate itself that emits the physicochemical signal translated by the transducer into an electrical signal. Such systems, mainly developed for immunoassaying, can be transposed to most ligand–receptor systems.

Direct methods are ones that do not require labelling. These include measurements of mass variations, optical properties (changes in absorbance, fluorescence, or refractive index), or electrical properties, due solely to the specific formation of the ligand–receptor complex.

Conventional antibodies, molecularly imprinted polymers, and artificial receptors have no catalytic activity and do not transform their targets, which remain complexed with the bioreceptor. Before the biosensor can be reused, the complex must therefore be dissociated and the device must be washed.

16.2 Immobilisation Methods

The bioreceptor must be in intimate contact with the transducer and may be directly immobilised at the transducer surface. However, the sensitive element is usually immobilised on or in some artificial substrate. Immobilisation techniques were originally developed for enzymes, but most can be transposed to different types of bioreceptor [10–12]. Immobilisation methods are usually broken down into five categories: adsorption, inclusion, confinement, crosslinking, and covalent coupling on an active substrate (see Fig. 16.4). Substrates of different kinds have been used (see Table 16.1 for some examples). Membrane substrates are generally the best suited for making biosensors.

16.2.1 Adsorption

Adsorption is simply due to interactions between the bioreceptor and its support. If the latter is not charged, low energy bonds (hydrogen bonds, van der Waals forces, or hydrophobic effects) are established between bioreceptor and support. If it is charged, ionic bonds will also occur. However, bioreceptors immobilised in this way can easily desorb through a change in pH or ionic strength, and this method is not widely used for bioreceptors.

Table 16.1. Examples of immobilisation substrates

Immobilisation method	Substrate
Adsorption	Agarose, alumina, starch, carbon, cellulose, collagen, collodion, silica gel, ion exchange resins, porous glass, polyamide, polystyrene, polyurethane
Inclusion	Gels (alginate, starch, carrageenan, gelatine, polyacrylamide, polysiloxane, polystyrene, polyvinyl alcohol, silica)
Confinement	Dialysis and ultrafiltration membranes
Covalent coupling	Polyamino acids and proteins (polyglutamic acid, L-alanine and L-glutamic acid copolymer, collagen, fibroin) Polysaccharides (polygalacturonic acid, agarose, starch, cellulose, chitin, dextrose) Synthetic polymers (polyacrylamide, polyamide, polymethacrylate, polystyrene, polyvinyl alcohol) Inorganic substrates (silica gel, gold, quartz, porous glass)

16.2.2 Inclusion

The idea here is to incorporate the bioreceptor in a polymer which is usually in the form of a gel. The bioreceptor is not therefore directly bound to the substrate. This technique, also easy to implement, allows a homogeneous distribution of the bioreceptor throughout the gel, but it may cause damage to biological structures, which may be affected by the reagents used in the polymerisation process. Inclusion is mainly used to immobilise whole cells or subcellular fractions.

16.2.3 Confinement

In the confinement technique, the bioreceptor remains in solution inside a compartment bounded by a semi-permeable membrane which only allows through small molecules. The bioreceptor thus remains in solution within a microcompartment of volume in the μL range or less, bounded by the membrane.

16.2.4 Crosslinking

The use of bifunctional agents such as glutaraldehyde $\text{CHO}-(\text{CH}_2)_3-\text{CHO}$ provides a way of crosslinking enzyme molecules one to another or co-crosslinking them in the presence of an inactive protein (serum albumin or gelatine). Crosslinking is also used to increase the stability of enzyme-substrate complexes obtained by adsorption or inclusion.

16.2.5 Covalent Bonding on an Activated Substrate

The covalent bonding of bioreceptors, mainly proteins, on activated substrates is achieved by setting up covalent bonds between functional groups of the substrate and functional groups of the bioreceptor that are not involved in the molecular recognition process. The most widespread methods use $-\text{COOH}$, $-\text{NH}_2$, $-\text{OH}$, and $-\text{SH}$ groups on the substrate. These groups are chemically rather inactive and they must first be transformed into activated functions in order to be able to react under mild conditions with the functional groups of the bioreceptor. For proteins, binding occurs through functional groups carried by the side chains of the amino acids, viz., amine, imine, amide, hydroxyl, carboxyl, and thiol, but also via the *N*-terminal or *C*-terminal group, or again via the carbohydrate part in the case of glycoproteins. Some of the main activation methods are described briefly below.

Activation of Carboxylic Groups

There are two techniques here:

- The acyl nitride method. Here the carboxylic groups are first esterified by methanol in an acidic medium, then transformed into hydrazides by treating with hydrazine, and finally into acyl nitride (treating with NaNO_2). By reacting with the amine groups of the bioreceptor, the latter gives an amide bond.
- The carbodiimide method. The carbodiimides ($\text{R}_1-\text{N}=\text{CN}-\text{R}_2$) react with $-\text{COOH}$ groups in an acid medium to give an *O*-acylisourea derivative able to form an amide bond with an $-\text{NH}_2$ group of the bioreceptor.

Activation of Amine Groups

The technique here uses diazonium salts. Aromatic amine functions can be treated by NaNO_2 to give a diazonium salt, which can react with an imidazole nucleus (histidine residue) or a phenol nucleus (tyrosine residue).

Activation of Hydroxyl Groups

The method here uses cyanogen bromide. When the substrate carries α -glycol groups, it can be treated with cyanogen bromide (CNBr) to obtain imido-carbonates that can react with amine groups on a bioreceptor to give *N*-substituted isoureas.

Activation of Sulfhydryl Groups

The formation of disulfide bridges ($-\text{S}-\text{S}-$) between a thiol group of the support and a thiol group of the bioreceptor is made possible after activating

the support by means of 2,2'-dipyridyldisulfide. This method produces a reversible covalent immobilisation of the bioreceptor. Theoretically, the support and bioreceptor can be regenerated by treating the former with a reducing agent for the sulfhydryl groups (dithiothreitol, cysteine, β -mercaptoethanol).

These methods for activating the substrate are all relatively simple in principle, but often complicated to actually implement. Today, ready-to-use pre-activated supports are available commercially. Immobilisation occurs within a few minutes by simply setting the support in contact with a solution containing the bioreceptor.

16.3 Biosensors with Electrochemical Detection

16.3.1 Enzyme Electrodes

An enzyme electrode comprises an immobilised enzyme preparation (the bioreceptor) and an electrochemical sensor for potentiometric or amperometric detection (the transducer), capable of specifically detecting some chemical species involved in the enzyme reaction.

Electrochemical Sensors

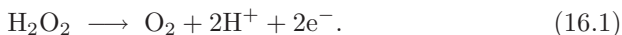
Several types of selective electrochemical sensor can be used to make an enzyme electrode. These can be divided into two types, viz., potentiometric and amperometric.

Potentiometric detection operates at constant or zero current. One then measures the voltage variation, which is proportional to the logarithm of the concentration of the reactive species. The commonest example of this type of sensor is the pH electrode. The sensitivity to H_3O^+ ions is due to the particular composition of the glass membrane of the electrode. By suitably modifying this composition, electrodes can be obtained that are sensitive to other ions, such as Na^+ , K^+ , and NH_4^+ . Gas diffusion potentiometric electrodes can measure the partial pressure of a dissolved gas in a solution. They comprise a pH electrode and a hydrophobic membrane permeable to the gas. A filler solution is placed between the glass electrode and the hydrophobic membrane. The composition of this solution depends on the kind of gas whose concentration is to be measured. The partial pressure of CO_2 and NH_3 can be measured with such electrodes.

With sensors using amperometric detection, the voltage is held constant and the current variation is measured. The latter is directly proportional to the concentration of the reactive species. The electrodes are metal (Pt, Au, Ag), and a bias is imposed by a polarograph, which can also measure the current variation consequent to electrochemical reduction or oxidation of the reactive species. Detectors using amperometric detection are mainly used to

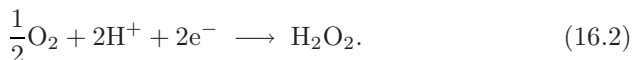
measure the concentration of hydrogen peroxide and the partial pressure of oxygen (p_{O_2}).

To measure the concentration of H_2O_2 , the metal electrode (usually platinum) is held at a constant positive potential between +600 and +700 mV higher than an $\text{Ag}/\text{AgCl}/\text{Cl}^-$ reference electrode. The oxygenated water oxidises on contact with the biased electrode according to the reaction



The resulting change in current is then directly proportional to the amount of oxidised hydrogen peroxide.

The oxygen partial pressure p_{O_2} is measured using a Clark electrode. This sensor has a metal electrode on which a hydrophobic membrane has been deposited. The latter is usually teflon, permeable only to gases. By imposing a constant potential of around -700 mV relative to an $\text{Ag}/\text{AgCl}/\text{Cl}^-$ reference electrode, the dissolved oxygen that comes into contact with the electrode is reduced and the resulting current variation is directly proportional to the partial pressure of oxygen in the solution:



Glucose Electrode

The glucose oxidation reaction, catalysed by glucose oxidase (GOD), can be taken to illustrate the various kinds of electrochemical sensors. The reaction occurs in two stages and involves a non-protein (prosthetic) organic structure, tightly bound to the enzyme, viz., flavin adenine dinucleotide (FAD). FAD is a redox coenzyme which alternately accepts and releases electrons. In the first stage, the glucose is oxidised by the GOD–FAD complex to give gluconolactone and the reduced complex GOD–FADH₂ (see Fig. 16.5). Gluconolactone undergoes a fast reversible hydrolysis into gluconic acid, while in a second stage the GOD–FADH₂ complex is oxidised by molecular oxygen to give back the GOD–FAD complex and hydrogen peroxide.

The appearance of gluconic acid can in principle be detected by a pH electrode, but in practice, since the enzyme reactions occur in a buffered medium, the pH variations are very slight. The glucose can also be assayed by detecting either the oxygen consumption (oxygen electrode) or the appearance of hydrogen peroxide (metal electrode). These enzyme electrodes, with which a species involved in the reaction is directly detected by an electrochemical sensor, are referred to as first generation. The principle of an enzyme electrode for assaying glucose, with amperometric detection of the resulting hydrogen peroxide, is illustrated in Fig. 16.6.

Other oxidases work in the same way as glucose oxidase, specifically catalysing the oxidation of certain substrates with consumption of oxygen and

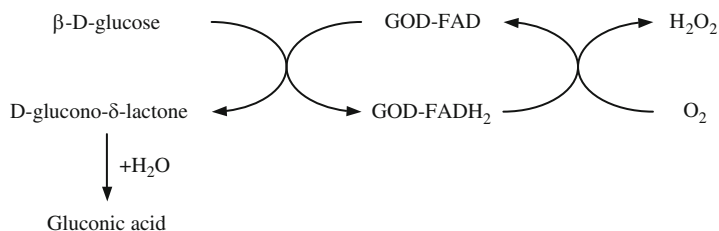


Fig. 16.5. Oxidation of glucose to form gluconolactone, catalysed by glucose oxidase (GOD). This is accompanied by reduction of the prosthetic group FAD to form FADH₂, which is then oxidised by molecular oxygen to give back FAD and hydrogen peroxide

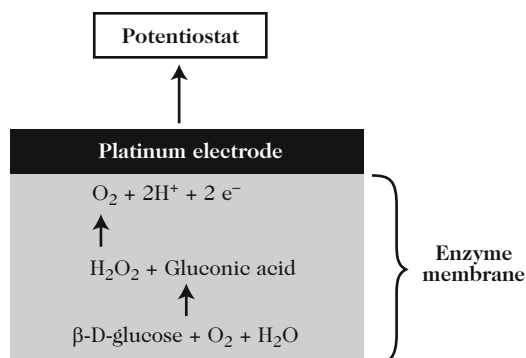


Fig. 16.6. Glucose electrode. The hydrogen peroxide produced during the oxidation of glucose, catalysed by glucose oxidase immobilised on a membrane, is detected by a platinum electrode with a bias of 650 mV relative to Ag/AgCl

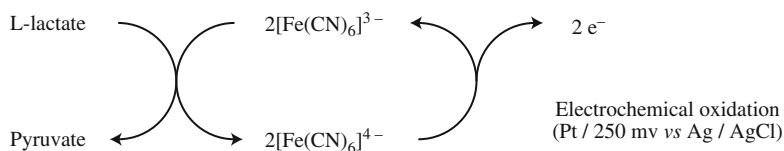


Fig. 16.7. Reactions involved in the enzymatic oxidation of L-lactate, catalysed by lactate dehydrogenase in the presence of ferricyanide

formation of hydrogen peroxide, and can therefore be used to devise different enzyme electrodes.

Biosensors implementing this idea can give a response that depends not only on the concentration of the target substance, but also on the partial oxygen pressure. Moreover, mainly for the detection of H₂O₂, electrochemical interference is not uncommon, insofar as other compounds may be oxidised at this working potential (usually between +600 and +700 mV relative to Ag/AgCl). In order to avoid such interference, three types of solution have been put forward:

- A differential arrangement with two sensors, one being the enzyme electrode proper and the other a similar sensor to the first, but incorporating a membrane without enzyme which will thus detect only the electrochemical interference. The specific response is then obtained by subtracting the response of the non-enzyme sensor from the response of the enzyme sensor.
- Use of a selective membrane that only allows small uncharged molecules to diffuse to the metal electrode, exploiting the fact that interfering species are generally charged. This is the solution proposed by Yellow Spring Instruments (YSI) in the United States, the first company to have commercialised a glucose analyser incorporating an enzyme electrode, in 1974.
- Use of redox mediators, either in solution or bound on the electrodes, that can be oxidised at low potentials. This obviously minimises the risks of electrochemical interference. For the enzymatic oxidation of glucose by glucose oxidase, the mediator in oxidised form (MED_{ox}) substitutes for oxygen:



The reduced mediator (MED_{red}) is electrochemically reoxidised. The strength of the oxidation current is thus correlated with the glucose concentration. Some dyes, the quinones, but also ferrocene and its derivatives, are used as mediators.

Apart from oxidase electrodes, one can also design dehydrogenase electrodes. There are several hundred dehydrogenases which catalyse various redox reactions in the presence of a particular coenzyme, viz., nicotinamide adenine dinucleotide (NAD):



NAD^+ and NADH are respectively the oxidised and reduced forms of the coenzyme. During the enzyme reaction, the reduced substrate is oxidised in the presence of NAD^+ to give the oxidised reaction product and NADH plus a proton. Direct electrochemical reoxidation of NADH is theoretically possible. However, the high detection potential for NADH (≥ 800 mV relative to Ag/AgCl) places a restriction on this use. To get around the difficulty, redox mediators can be used to reoxidise the NADH into NAD^+ , in which case the oxidation current of the reduced mediator is measured.

An interesting case is provided by the oxidation of L-lactate, which can be detected by the action of lactate oxidase using the same idea as for glucose oxidase, but also by the action of lactate dehydrogenase. The mammalian muscle enzyme functions with the pair NAD^+/NADH , while lactate dehydrogenase from microorganisms can function in vitro with the pair ferri-cyanide/ferrocyanide. An L-lactate electrode was thus devised using this idea. The chain of redox reactions is depicted in Fig. 16.7.

Enzyme electrodes using a mediator are called second generation electrodes. With the third generation of enzyme electrodes, which will not be

Table 16.2. Examples of enzyme systems involving oxygen consumption or hydrogen peroxide production

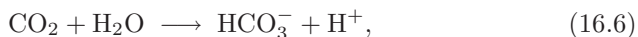
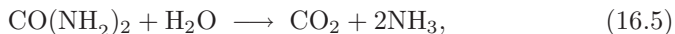
Analyte	Enzyme reactions	Enzymes used
Cholesterol	$\text{Cholesterol} + \text{O}_2 \longrightarrow \text{cholestenone} + \text{H}_2\text{O}_2$	Cholesterol oxidase
D-glutamate	$\text{D-glutamate} + \text{O}_2 \longrightarrow \alpha\text{-ketoglutarate} + \text{NH}_3 + \text{H}_2\text{O}_2$	Glutamate oxidase
L-lactate	$\text{L-lactate} + \text{O}_2 \longrightarrow \text{pyruvate} + \text{H}_2\text{O}_2$	L-lactate oxidase
Pyruvate	$\text{Pyruvate} + \text{HPO}_4^{2-} + \text{O}_2 \longrightarrow \text{acetyl phosphate} + \text{CO}_2 + \text{H}_2\text{O}_2$	Pyruvate oxidase
Creatinine	$\text{Creatinine} + \text{H}_2\text{O} \longrightarrow \text{creatinine}$	Creatininase
	$\text{Creatine} + \text{H}_2\text{O} \longrightarrow \text{urea} + \text{sarcosine}$	Creatinase
	$\text{Sarcosine} + \text{O}_2 \longrightarrow \text{glycine} + \text{formaldehyde} + \text{H}_2\text{O}_2$	Sarcosine oxidase
Lactose	$\text{Lactose} + \text{H}_2\text{O} \longrightarrow \beta\text{-D-galactose} + \beta\text{-D-glucose}$	Galactosidase
	$\beta\text{-D-glucose} + \text{O}_2 \longrightarrow \text{D-glucono-}\delta\text{-lactone} + \text{H}_2\text{O}_2$	Glucose oxidase
Maltose	$\text{Maltose} + \text{H}_2\text{O} \longrightarrow 2\beta\text{-D-glucose}$	Glucosamylase
	$\beta\text{-D-glucose} + \text{O}_2 \longrightarrow \text{D-glucono-}\delta\text{-lactone} + \text{H}_2\text{O}_2$	Glucose oxidase
Saccharose	$\text{Saccharose} + \text{H}_2\text{O} \longrightarrow \beta\text{-D-fructose} + \alpha\text{-D-glucose}$	Invertase
	$\alpha\text{-D-glucose} \longrightarrow \beta\text{-D-glucose}$	Mutarotase
	$\beta\text{-D-glucose} + \text{O}_2 \longrightarrow \text{D-glucono-}\delta\text{-lactone} + \text{H}_2\text{O}_2$	Glucose oxidase

discussed in this chapter, electrons are transferred directly between the electrode and the catalytic site of the enzyme.

The number of possibilities can be extended by exploiting multienzyme systems, which involve series of reactions. Some examples of mono- and multienzyme systems of this kind are listed in Table 16.2.

Urea Electrode

The hydrolysis of urea, catalysed by urease, leads to the production of carbon dioxide and ammonia, which are ionised in an aqueous medium. The percentage of ionised species then depends on the pH of the reaction medium. The various reactions involved here are as follows:



Potentiometric detection can be used for several of the species involved:

- CO_2 by a gas diffusion electrode sensitive to carbon dioxide.
- NH_3 by a gas diffusion electrode sensitive to ammonia.
- NH_4^+ by a specific glass membrane electrode.

The use of a gas diffusion electrode sensitive to CO_2 is complicated by the possible interference of CO_2 in the atmosphere. The NH_4^+ electrode is also liable to interference, notably from Na^+ and K^+ ions that are difficult to eliminate from the sample.

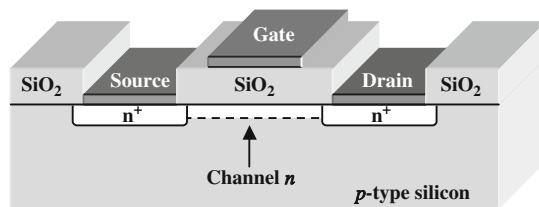


Fig. 16.8. Schematic view of a MOSFET

Although theoretically possible, the use of an NH_3 electrode raises the problem of pH compatibility between the enzyme and the electrochemical sensor. Indeed, at pH 7, the optimal pH for the catalytic action of urease, ammonia is totally ionised in the form NH_4^+ . Now the maximal sensitivity of the NH_3 electrode is obtained at a strongly alkaline pH (≥ 12). This value is totally incompatible with the catalytic activity of urease, which is irreversibly denatured at pH > 8.5 . Possible solutions are to work with a pH of about 8, but a much reduced sensitivity of the electrochemical sensor, or to carry out the enzyme stage and electrochemical stage separately, each at its optimal pH value.

The latter solution has been adopted to design an automated analyser with a urease microreactor in which the enzyme is immobilised on microspheres in a column of length about 1 cm and diameter about 0.5 cm. The sample to be assayed is injected at the head of the column, in which the pH is equal to 7. At the column outlet, the ammonium ions produced by hydrolysis of urea at neutral pH are totally transformed into ammonia by an alkaline solution with pH > 12 . An NH_3 electrode is placed downstream of the region where the alkaline solution mixes with the outflow of the column. The sensor then operates at its optimal pH with maximum sensitivity.

16.3.2 ENFET or Enzyme ISFET

Enzyme biosensors using field effect transistors (enzyme field effect transistor or ENFET) combine an enzyme preparation and an ion sensitive field effect transistor (ISFET). These transistors have the structure of a metal oxide semiconductor field effect transistor (MOSFET) comprising a *p*-type silicon substrate with its surface oxidised to form an insulating layer of SiO_2 . Using photolithography, two regions of the insulating layer are cleared to create two zones doped with *n*-type impurities, i.e., the source and the drain (see Fig. 16.8).

Silicon is the basic material used to make semiconductors. In pure (undoped) silicon, there is little or no electron conduction. To render it conducting, impurities are introduced in atomic form in the crystal. A type *n* semiconductor is made by doping the crystal with atoms of group V in the periodic table of the elements, e.g., P or As. These atoms have one more valence electron than silicon and can therefore supply electrons to the crystal.

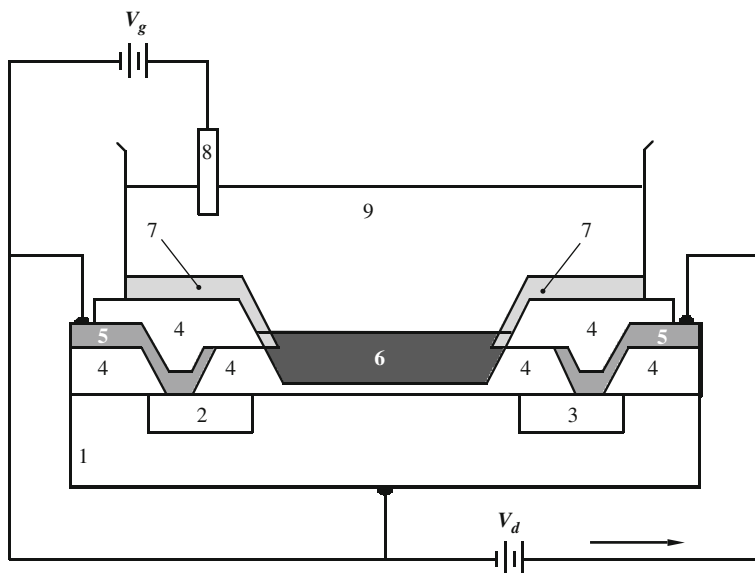


Fig. 16.9. Schematic view of an ISFET: (1) substrate, (2) source, (3) drain, (4) insulating layer of SiO_2 , (5) metal contact, (6) selective membrane, (7) insulating resin, (8) reference electrode, (9) reaction medium

By doping the crystal with group III atoms such as B or Al, a type p semiconductor is obtained. These atoms have one less valence electron than silicon and their presence within the silicon crystal lattice creates immobilised traps which capture free electrons. The source and drain are coated with metal to provide electrical contact. Above the channel between the source and drain, the SiO_2 insulating surface is also coated with metal. This constitutes the gate. By applying a weak positive potential or bias V_d to the drain and a large enough potential V_g to the gate, the electric field created in the substrate will induce a current in the channel between source and drain. With a suitable electrical setup, the current will be directly proportional to V_g .

The idea of an ISFET was first described in 1970 by Bergveld [13]. The gate is replaced by a selective membrane sensitive to ions and, since the device must operate with solutions, the gate potential is imposed using a reference electrode (see Fig. 16.9).

ISFET devices sensitive to pH are mainly used to design ENFET devices. These biosensors make use of enzyme reactions during which H^+ ions are consumed or produced. The enzyme is immobilised on the selective membrane in the form of a thin film of thickness a few μm . Among the enzyme reactions used, one should mention the oxidation of glucose by glucose oxidase (see Fig. 16.5, detection of gluconic acid), the hydrolysis of urea by urease [reactions (16.5)–(16.7)], the hydrolysis of penicillin by penicillinase [reaction

(16.8)], and the hydrolysis of acetylcholine by acetylcholinesterase [reaction (16.9)]:



The possibility of miniaturisation is the main advantage of these biosensors. ISFET devices used to develop biosensors generally have widths in the range of a few tenths of a millimeter and lengths of a few millimeters. Another positive point, a consequence of these small dimensions, is the possibility of making multibiosensors, i.e., combining several biosensors with different specificities into the same device. The concentrations of several different species can then be measured in the same operation. The low cost of FET technology, due to mass production to meet the needs of the microelectronics industry, constitutes another advantage for these transducers. On the other hand, the current weakness of ENFET devices lies in the difficulty in making a thin enough enzyme film and depositing it accurately enough on the selective membrane. Moreover, an ISFET requires a reference electrode in order to operate, and at the present time these cannot be made with the dimensions of a field effect transistor.

16.4 Mass Transducer Biosensors

The physical phenomenon underlying this kind of transducer is the piezoelectric effect observed in some materials, such as quartz. Mass detection by a piezoelectric crystal uses the fact that the natural frequency of vibration of an oscillating crystal depends on the mass of substance adsorbed on its surface (quartz microbalances are discussed in more detail in Chap. 9). Under certain well defined conditions, the change in the resonance frequency of the crystal is directly proportional to the amount of matter present at the surface.

All crystals have a natural vibration frequency called the resonance frequency, which depends on their chemical constitution, their dimensions, their shape, and their mass. With a piezoelectric crystal, vibrations generate an oscillating electric field of the same frequency as those vibrations. Conversely, a piezoelectric crystal placed in an oscillating electric field will vibrate at the same frequency as the electric field. The empirically established Sauerbrey equation expresses the relation between the masses of thin metal films deposited at the surface of a quartz crystal and the corresponding change in the resonance frequency of the crystal:

$$\Delta f = \frac{k f^2 \Delta m}{A}, \quad (16.10)$$

where Δf is the change in frequency (Hz), k is a constant whose value depends on the type of crystal, f is the resonance frequency (Hz) of the crystal before the deposit, Δm is the deposited mass (g), and A is the area coated

Table 16.3. Applications of biosensors using mass transduction

Liquid phase analysis	
Target substance	Bioreceptor
<i>Candida albicans</i>	Ab anti- <i>Candida</i>
<i>Salmonella thyphimurium</i>	Ab anti- <i>Salmonella</i>
Human transferrin	Ab anti-transferrin
Human albumin	Ab anti-albumin
<i>S. thyphimurium</i> DNA	<i>S. thyphimurium</i> DNA
Organophosphorus and carbamate pesticides	Acetylcholinesterase
Gas phase analysis	
Target substance	Bioreceptor
Formaldehyde	Formaldehyde dehydrogenase
Cocaine	Ab anti-cocaine
Parathion	Ab anti-parathion

with the deposit (cm^2). This equation is used to calculate the sensitivity of a given crystal. For example, for quartz crystals with 9 MHz and 15 MHz, the sensitivities are 400 Hz/ μg and 2,600 Hz/ μg , respectively. If the frequency can be measured with an accuracy of 0.1 Hz, the detection limit is in the range 10–100 pg. Substances other than metals can modify the resonance frequency of a piezoelectric crystal, and this has been exploited to design piezoelectric transduction biosensors.

Quartz crystals with resonance frequencies 5, 9, or 15 MHz are the most commonly used to make biosensors. They are usually in the shape of a disk with diameter a few mm and thickness about 0.15 mm. The metal electrodes, usually gold, used to induce an oscillating electric field normal to the disk surface, have thickness in the range 0.3–1 μm and diameter a few mm, less than the thickness of the quartz disk, and are deposited on each face. The disk is thus sandwiched between these two electrodes.

In practice, a piezoelectric detection system comprises two oscillating circuits: an oscillating detection crystal and an oscillating reference crystal. The reference crystal is physically identical to the measurement crystal, but carries no immobilised bioreceptor. A frequency counter is connected to each oscillating circuit and the difference of frequency between the two crystals is then obtained.

To carry out the measurement proper, the frequency difference Δf_1 between the measurement crystal and the reference crystal is first determined. After setting the sample in contact with the measurement crystal, the latter is rinsed and dried and the frequency difference Δf_2 is once again measured. The change $\Delta f = \Delta f_2 - \Delta f_1$ in the frequency difference is then correlated with the amount of substance adsorbed onto the measurement crystal.

The main applications concern immunoassays, in liquid or gas phase, and the detection of certain volatile substances (see Table 16.3). For immunoassays, an antigen or antibody (Ab) is immobilised on the crystal and the measurement is carried out after the reaction between the bioreceptor and the target substance. One then speaks of direct immunoassays. The detection of volatile substances by a biosensor using a piezoelectric crystal was developed for formaldehyde and organophosphorus pesticides. In the case of formaldehyde, the bioreceptor is formaldehyde dehydrogenase. The organophosphorus pesticides can be detected either by measuring the inhibition of acetylcholinesterase, which is then immobilised on a crystal, or by immunoassaying after immobilisation of a specific antibody. Finally, it should be mentioned that one of the first DNA sensors was a mass transduction sensor for detecting salmonella, in which an oligonucleotide complementary to a specific sequence of salmonella was immobilised on a quartz crystal.

16.5 Enzyme Thermistors

Enthalpy changes ΔH accompanying enzyme reactions are in the range -25 to -100 kJ/mol and can be detected by calorimetry. The amount of heat Q released during a reaction is proportional to the change in enthalpy of the reaction and the number of moles n involved in the reaction:

$$Q = n\Delta H. \quad (16.11)$$

The amount of heat can also be expressed as a function of the heat capacity C of the system and the change in temperature ΔT :

$$Q = C\Delta T. \quad (16.12)$$

Hence the temperature change is proportional to the change in enthalpy and the number of moles in the reaction, and inversely proportional to the heat capacity:

$$\Delta T = \frac{n\Delta H}{C}. \quad (16.13)$$

Measurements of the temperature change induced by an enzyme reaction can thus be exploited to design biosensors. However, these temperature changes are very small, as exemplified by the oxidation of glucose catalysed by glucose oxidase. The enthalpy change for this reaction is -80 kJ/mol. In a continuous flow system fed by an aqueous solution with a flow rate of 1 mL/min, introducing a 1 mL sample of glucose at 1 mM would produce a temperature variation of only 0.01°C . This means that very sensitive temperature sensors are required, and the thermistor is at the present time the most suitable temperature transducer.

In most devices designed to use heat transduction, the biological component consists of a column of enzyme immobilised on microbeads (either glass

or Sepharose). The internal diameter of the column is a few millimeters and the height is in the range 15–30 mm. The column is housed in a heat-insulated container and fed by a flow of reagents in which the sample is injected via an external system. Most of the heat produced by the enzyme column is carried out of the column by the flow of reagents. The flow temperature is measured at the column outlet by a thermistor and, in order to improve the stability of the base line, a differential measurement is made by placing a reference thermistor at the column inlet.

Enzyme thermistors can only carry out a global measurement of the enthalpy change in the system, and apart from the enzyme reaction, this includes the interactions between the various species present in the reaction medium. So the measurement is not strictly specific to the relevant enzyme reaction. One solution is to use an inactive reference column, so that measurements can be made in the absence of the enzyme reaction and these subtracted from the values obtained with the column containing the bioreceptor.

Some enzyme reactions produce a very small enthalpy change and the measurements are not sensitive enough. In this case, several solutions have been put forward to amplify the temperature variation, viz., chemical amplification, enzyme amplification, and the use of organic solvents:

- Chemical amplification means using a compound whose interaction with one of the products of the enzyme reaction is accompanied by a large enthalpy change, e.g., tris-hydroxymethylaminomethane (Tris), whose protonation enthalpy is -47.5 kJ/mol and which reacts with an acidic reaction product. For example, the change in enthalpy for the hydrolysis of acetylcholine by acetylcholinesterase [see (16.9)] is close to zero. Despite this, choline can be assayed by calorimetry, because the enzyme hydrolysis of acetylcholine yields a proton which can interact with Tris if the latter is present in the reaction medium. Under these conditions, the overall reaction is highly exothermic.
- Enzyme amplification involves co-immobilising one or more enzymes which work in sequence with the specific enzyme reaction. Each reaction contributes to the enthalpy change in the system, and the measurement corresponds to the sum of the reaction enthalpies. For example, with oxidase enzymes, the catalytic action is generally accompanied by production of H_2O_2 and one can then associate catalase, which catalyses the transformation of hydrogen peroxide into oxygen and water with an enthalpy change of -100 kJ/mol . The amount of heat given off by these multienzyme systems is very large and the measurement becomes all the more sensitive.
- Enzyme reactions can also be carried out in an organic solvent. Since the temperature change is inversely proportional to the heat capacity of the system, one can significantly increase the detection sensitivity by working in the presence of organic solvents, because they have a heat capacity that is generally about three times lower than water. However, not all enzymes can operate in an organic medium. Most of the reactions investigated

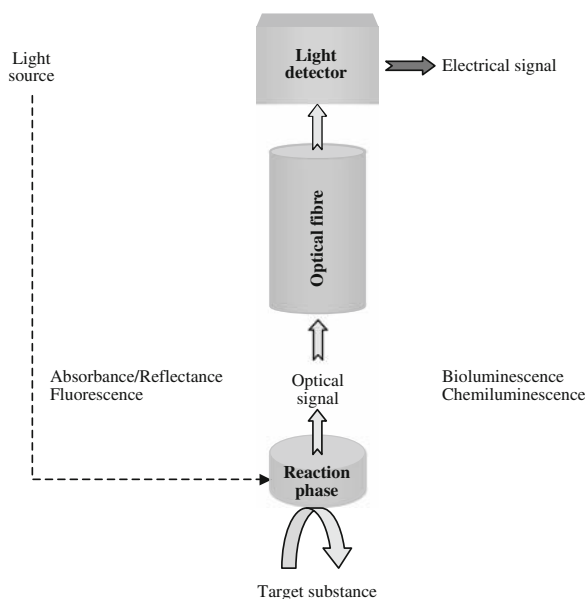


Fig. 16.10. General setup for a fibre optic sensor. The target substance is specifically recognised by a bioreceptor in the reaction phase. This specific recognition leads to emission or modification of a light signal transmitted by an optical fibre or a bundle of such fibres to a light detector

involve lipid substrates and lipases. For example, lipoprotein lipase is used to measure the concentration of triglycerides and, when the reaction is set up in cyclohexane, the resulting signal, i.e., the amount of heat produced, is about 2.5 times greater than in an aqueous phase.

Generally speaking, whatever reaction system is set up to develop enzyme thermistors, the detection limits are usually in the range 1–10 μM .

16.6 Fibre Optic Biosensors

A fibre optic biosensor can be defined as an association of the reaction phase with a transduction system made from an optical fibre or a bundle of optical fibres coupled to a light detector [14]. The reaction phase comprises an immobilised biological component, sometimes associated with one or more coreagents, themselves immobilised. Biospecific recognition then directly or indirectly generates an optical signal which is transmitted to the light detector by the optical fibre. These biosensors mainly exploit fluorescence, absorbance/reflectance, bioluminescence, and chemiluminescence. When fluorescence or absorbance/reflectance measurements are used, a light source must be integrated into the setup (see Fig. 16.10).

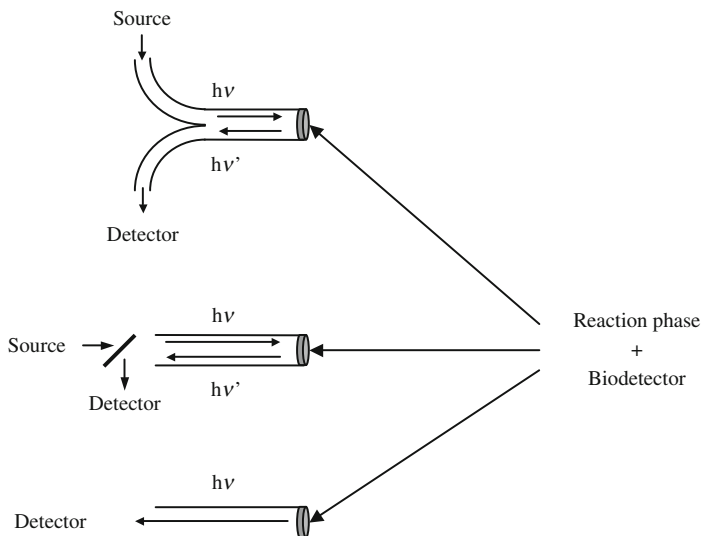


Fig. 16.11. The different configurations of a fibre optic biosensor depending on the type of detection used. *Top:* Bundle of branching fibres. *Middle:* The same fibre or bundle of fibres guides the light from the light source to the reaction phase, then from the reaction phase to the light detector. These two configurations are used for fibre optic biosensors with detection of fluorescence or absorbance. *Bottom:* Light emitted directly by the reaction phase is guided to the detector by the optical fibre or a bundle of optical fibres. This configuration is used for fibre optic biosensors using bioluminescence or chemiluminescence

16.6.1 Fibre Optic Chemical Sensors

Many fibre optic chemical sensors have been devised, mainly based on fluorescence or fluorescence extinction, absorbance and/or reflectance [14]. Among these devices, most fibre optic biosensors use pH sensors, oxygen sensors, or NH_3 sensors.

pH Sensors

The changes in absorbance of phenol red at different values of the pH have been exploited. After immobilisation, the indicator dye, with a pK_a in solution of 7.9, has a pK_a of 7.6. The basic form of the indicator absorbs at 558 nm, while no form absorbs at 600 nm. The ratio of the reflected intensity at 558 nm to the reflected intensity at 600 nm depends on the pH, which can be measured to within 0.01 between 7.0 and 7.4.

Some fluorophores can have different spectral properties depending on their acidic or basic form. The acidic and basic forms of 4-methylumbelliferone emit at different wavelengths for the same excitation wavelength. The ratio of the intensities measured at these two wavelengths is proportional to the

pH. For another fluorophore, 8-hydroxy-1,3,6-pyrenetrisulfonate (HPTS), the basic form is selectively excited at 470 nm, whereas the acidic form is excited at 405 nm. In this case, it is the ratio of the fluorescence intensities resulting from excitations at 405 nm and 470 nm which is correlated with the pH. The response range for these sensors is limited to 1 or 2 pH units around the pK_a of the indicator.

NH₃ Sensors

Optical ammonia sensors are pH sensors modified by setting them in contact with a microreservoir containing an ammonium salt solution. Since the wall of the reservoir is only permeable to gases, the ammonia present in the sample diffuses into the enclosed solution and modifies its pH. At equilibrium, the pH measured in the enclosed solution is proportional to the partial pressure of ammonia in the sample.

Oxygen Sensors

These sensors are generally based on fluorescence extinction measurements. The fluorophores used are pyrene butyric acid, perylene dibutyrate, and 9,10-diphenylanthracene. The fluorescence intensity is inversely proportional to the oxygen partial pressure.

16.6.2 Setups for Fibre Optic Biosensors

The configuration of a fibre optic biosensor depends on whether or not a light source is required. The reaction phase containing the bioreceptor is located at the end of the fibre or bundle of fibres. The reagent can be immobilised by adsorption or covalence on a synthetic membrane, or confined behind a semi-permeable membrane.

When a light source is required, the light is guided to the reaction phase using either a branching fibre device, or a single fibre (or single bundle of fibres), also guiding the optical signal produced in the reaction phase. In the first case, one end of the bundle of fibres splits into two parts. One of these is coupled to the light source and the other to the detector. At the undivided end, the strands in the bundle are usually randomly distributed and held in contact with the reaction phase. When the same fibre transports the light from the light source to the reaction phase, then from the reaction phase to the detector, a separator is used to orient the light emerging onto the detector. Finally, when the light is emitted directly by the reaction phase, e.g., in bioluminescence and chemiluminescence reactions, the fibre or bundle of fibres is directly connected to the light detector (see Fig. 16.11).

Table 16.4. Examples of fibre optic biosensors incorporating chemical sensors

Target substance	Enzyme	Enzyme reaction	Chemical sensor
Cholesterol	Cholesterol oxidase	$\text{Cholesterol} + \text{O}_2 \longrightarrow \text{cholestenone} + \text{H}_2\text{O}_2$	O_2
Ethanol	Alcohol oxidase	$\text{Ethanol} + \text{O}_2 \longrightarrow \text{acetaldehyde} + \text{H}_2\text{O}_2$	O_2
Glucose	Glucose oxidase	$\text{Glucose} + \text{O}_2 + \text{H}_2\text{O} \longrightarrow \text{gluconate} + \text{H}^+ + \text{H}_2\text{O}_2$	O_2 , pH
Lactate	Lactate monooxygenase	$\text{L-lactate} + \text{O}_2 \longrightarrow \text{acetate} + \text{CO}_2 + \text{H}_2\text{O}$	O_2
Penicillin	Penicillinase	$\text{Penicillin} + \text{H}_2\text{O} \longrightarrow \text{penicilloate} + \text{H}^+$	pH
Urea	Urease	$\text{Urea} + \text{H}_2\text{O} \longrightarrow \text{CO}_2 + 2\text{NH}_3$	pH, NH_3

16.6.3 Enzyme Fibre Optic Biosensors

These biosensors involve an enzyme reaction in which one of the substrates or products of the reaction is detected directly by fluorescence or absorbance, or else indirectly by a chemical fibre optic sensor (pH, O_2 , NH_3) if there is no optically measurable property. The enzyme is deposited on a membrane placed at the end of a fibre or bundle of fibres, or confined at the end of the biosensor.

Indirect Detection by Chemical Sensor

When the enzyme reaction involves the consumption or production of protons, oxygen, or ammonia, these species can be detected with a suitable sensor. Some examples of fibre optic biosensors made with chemical sensors are listed in Table 16.4.

Direct Detection by Fluorescence

As mentioned previously, reduced nicotinamide adenine dinucleotide (NADH) is the coenzyme for many dehydrogenases [see reaction (16.4)], and its fluorescence properties have been put to use to develop fibre optic biosensors associated with an immobilised specific dehydrogenase. The rate of production or consumption of the reduced coenzyme can be monitored by fluorimetry and correlated with the concentration of the second substrate in the reaction.

Several setups exploiting this idea have been described for assaying glucose, lactate, pyruvate, and ethanol. In general, the sensitivity of these sensors is poor and the detection limit is in the millimolar range or a few tens of micromoles per litre.

Direct Detection by Absorbance

The appearance of the reaction product from an immobilised enzyme can be monitored by UV or visible spectrophotometry. This type of fibre optic biosensor has been reported for glucose measurements. The idea is to use

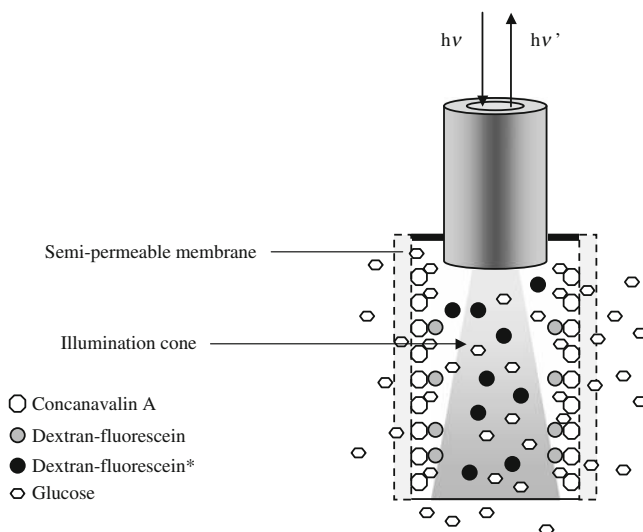


Fig. 16.12. Fibre optic affinity biosensor for glucose assays. The free glucose in the sample competes with dextran labelled by fluorescein to bind onto immobilised concanavalin A. When the dextran–fluorescein conjugate is in the illumination cone, the fluorescein is excited and the fluorescence increases

spectrophotometry to measure the hydrogen peroxide produced during the enzymatic oxidation of the glucose, using an auxiliary reaction catalysed by peroxidase in the presence of a chromogenic substrate. The glucose oxidase is immobilised on a nylon membrane positioned at the end of a branching bundle. The latter is placed in a small measurement cell containing peroxidase and the chromogenic reagent in solution. The appearance of the coloured product is then monitored using a bundle of fibres coupled to a spectrophotometer, and the relation between the absorbance and the glucose concentration is linear from 10 μM to 1 mM.

16.6.4 Affinity Biosensors

Lectin Biosensors

Lectins are proteins that can establish specific reversible bonds with certain mono- or oligosaccharides. An implantable fibre optic biosensor has been developed to measure glucose *in vivo*. The idea is to use the competition between glucose and dextran (a polymer of glucose) labelled by fluorescein isothiocyanate (FITC–dextran) to bind onto a lectin (concanavalin A) immobilised on the inner walls of a semi-permeable membrane [15]. Concanavalin A can bind just as well with either free glucose or polymerised glucose. The semi-permeable membrane is adjusted at the end of an optical fibre of diameter 300 μm and, in the absence of glucose, the dextran, which cannot diffuse

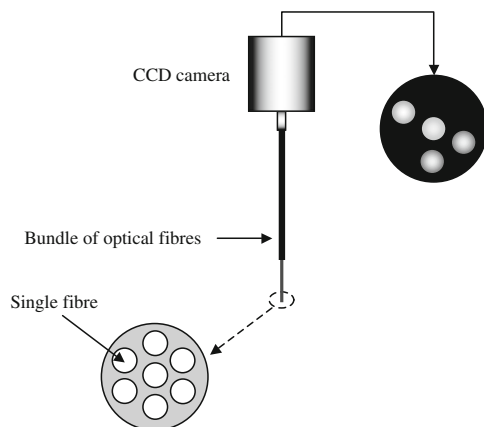


Fig. 16.13. Fibre optic biosensor for detecting DNA sequences. The oligonucleotide probes are immobilised at one end of each fibre. The other ends of the fibres are connected to a CCD camera in order to visualise those fibres in which there has been hybridisation of complementary strands labelled by fluorescein

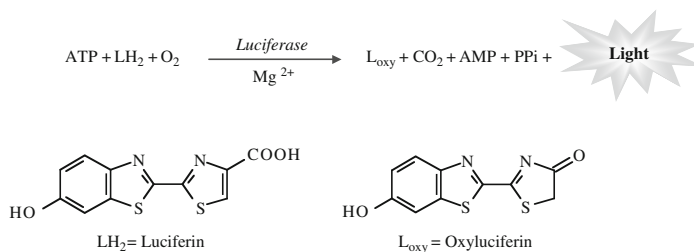


Fig. 16.14. Bioluminescence reaction catalysed by firefly luciferase. Luciferin, a specific substrate produced by the firefly, and which can be synthesised chemically, undergoes oxidative decarboxylation into oxyluciferin. The light emitted in this reaction is yellow–green with a peak at 560 nm

through the walls of the semi-permeable membrane, binds onto the lectin. The immobilised lectin is then outside the illumination cone of the optical fibre and the fluorescence is minimal (see Fig. 16.12). In a medium containing glucose in solution, this sugar diffuses through the semi-permeable membrane and can bind onto the concanavalin A instead of the FITC–dextran. As the latter is no longer bound to the lectin, it lies in the region illuminated by the fibre and the resulting increase in fluorescence can be correlated with the concentration of glucose in the medium, between 0.5 g/L and 2 g/L.

Fibre Optic Biosensors for Detecting DNA

A bundle of optical fibres is assembled, in which each fibre of diameter 200 μm carries a different oligonucleotide probe at the end. The hybridisation of

complementary oligonucleotides labelled by fluorescein is detected by measuring the fluorescence intensity with a CCD (charge coupled device) camera placed at the non-reacting end of this biosensor (see Fig. 16.13). The feasibility of simultaneously analysing several DNA sequences has been reported [16].

16.6.5 Biosensors Based on Chemiluminescent or Bioluminescent Detection

These biosensors exploit the light emission phenomenon occurring during certain reactions [17]. Chemiluminescence is light emission from a chemical reaction. Bioluminescence, the light emission phenomenon observed in certain living organisms, can be considered as a special case of chemiluminescence, in which a protein, usually an enzyme (the enzyme is luciferase here), is involved in the light-producing reaction.

Many bioluminescent organisms have been identified, but only a few have been subjected to detailed biochemical investigation. The most widely used bioluminescent systems in the field of analysis are the firefly system, used for specific and sensitive detection of adenosine triphosphate (ATP) (see Fig. 16.14), the bienzyme system of certain (mainly marine) bacteria used to detect reduced nicotinamide adenine dinucleotide (NADH) (see Fig. 16.15), and the system of the jellyfish *Aequorea* used to detect calcium ions.

The genes for the luciferases which catalyse light emitting reactions in the firefly and in light-emitting bacteria are also used in molecular biology as marker genes. A marker gene is inserted inside another gene and expressed at the same time as the gene in which it has been inserted. In a medium containing the substrates required for the bioluminescence reaction, light is emitted, serving as witness to the expression of the other, less easily detectable gene.

Apart from the direct analysis of ATP and NADH, many metabolites can be assayed by combining these bioluminescence reactions with other specific enzyme reactions.

Chemiluminescence reactions involve synthetic compounds and many light-emitting compounds have been studied. However, only a few have been used in biological and biochemical analysis. The most widely used reaction in biosensors is the chemiluminescence of luminol. This reaction occurs in an alkaline medium and in the presence of hydrogen peroxide, and this source of light emission can thus be coupled with enzyme reactions generating H_2O_2 , such as the oxidation of glucose catalysed by glucose oxidase. Working in suitable conditions, luminescence can thus be used to measure the concentration of glucose in a sample.

Other applications for these chemiluminescence reactions are immunoassay techniques and more generally, techniques for detecting affinity pairs. There are two main ways of going about this: either the light emitter is used to label a compound by chemical combination with the compound (analogous to radioactive labelling), or one of the non-luminescent compounds involved

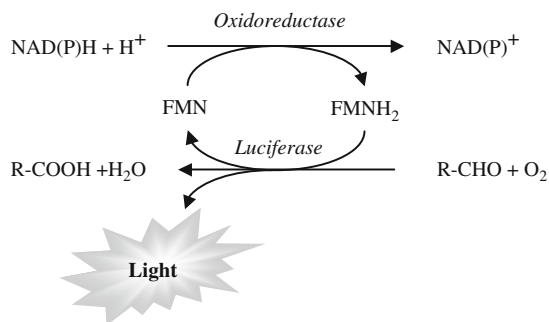


Fig. 16.15. Bienzyme bacterial system used to detect NADH or NADPH by bioluminescence. R-CHO is a long-chain aliphatic aldehyde and R-COOH the corresponding fatty acid. In vitro, the aldehyde generally used is decanal. Blue light is emitted, with a peak at 490 nm

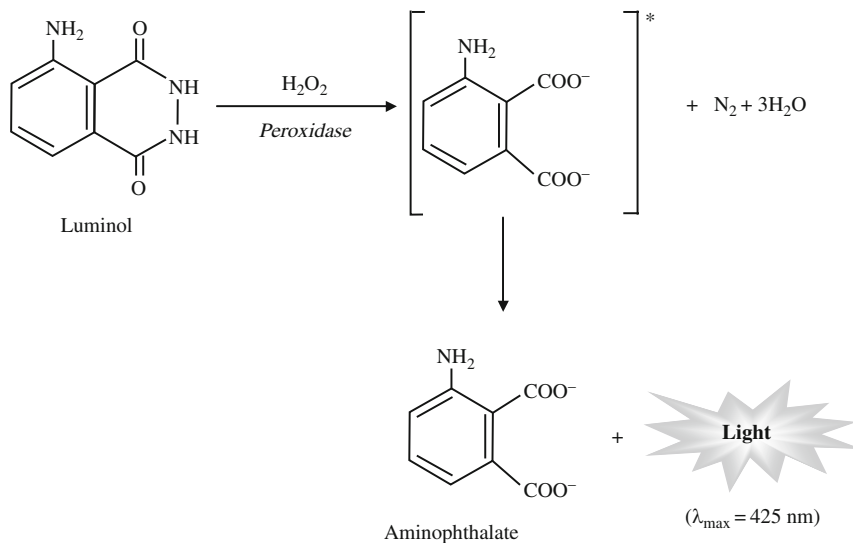


Fig. 16.16. Chemiluminescence reaction of luminol catalysed by peroxidase

in the light emission reaction or able to produce a light emitter is used as a label.

This is the case with horseradish peroxidase, which can catalyse the luminol chemiluminescence reaction in the presence of hydrogen peroxide (see Fig. 16.16). A peroxidase bioconjugate can thus be detected by adding luminol and hydrogen peroxide to the reaction medium. Alkaline phosphatase, an enzyme catalysing the hydrolysis of phosphoric esters, can also be used as a label. In this case, a stable non-luminescent synthetic substrate is used (a derivative of 1,2-dioxetane). This is hydrolysed by the action of alkaline phosphatase into an unstable product that emits light when it decomposes.

Bio- and chemiluminescence reactions are useful because of the great sensitivity of analytical techniques exploiting these reactions. Indeed, light emission can be measured very sensitively today, so compounds involved in these light emitting reactions can be detected at lower levels than by spectrophotometric, fluorimetric, or electrochemical methods. Furthermore, analyses based on light emission can handle cloudy samples, where spectrophotometric assays could not be used.

Biosensors based on bioluminescence or chemiluminescence reactions have usually been adapted to continuous flow systems for measurements by flow injection analysis (FIA). Total or partial automation of the analysis process is then possible (see Fig. 16.17).

Chemiluminescent Biosensors

Biosensors based on chemiluminescent detection have been developed to measure hydrogen peroxide [17, 18]. Chemiluminescence reactions become particularly interesting when coupled with reactions catalysed by oxidases of the kind used with amperometric detection sensors and generating hydrogen peroxide [20–25].

Electrochemiluminescent Biosensors

There is an alternative to peroxidase to trigger the chemiluminescence of luminol in the presence of hydrogen peroxide. This is the electro-oxidation of luminol, applying a voltage of the order of +400 to +500 mV relative to Ag/AgCl. This possibility has been put to use recently to develop electrooptical biosensors for detecting glucose, cholesterol, lactate, and choline [26–29]. These devices have been miniaturised, in particular, using screen printed carbon electrodes [30, 31].

Bioluminescent Biosensors

Two kinds of light-emitting reaction are involved in biosensors based on detection of bioluminescence: the firefly bioluminescence reaction for ATP assays, and the oxidoreductase–luciferase bioluminescent system for detection of NADH or NADPH. With such a system, the detection limits for ATP and NADH are 20 pM and 0.3 nM, respectively [19, 32, 33]. Furthermore, by joint immobilisation of specific dehydrogenases with the bacterial bioluminescent system, concentrations of sorbitol, ethanol, or oxaloacetate can be measured with a detection limit of nanomolar order [34–36]. Likewise, by coupling the firefly bioluminescence reaction with enzyme reactions involving ATP, optical biosensors for ADP and AMP have been devised [37, 38].

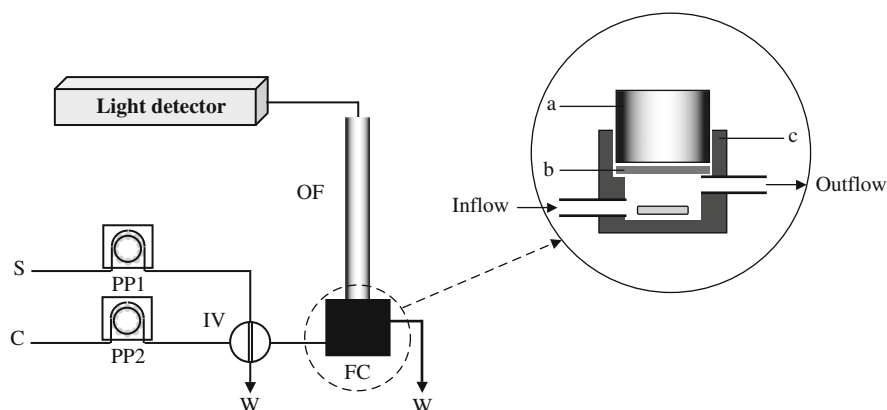


Fig. 16.17. Fibre optic biosensor inserted in a flow injection analysis device for bioluminescence or chemiluminescence measurements. OF optical fibre, FC flow cell, S sample, C flow of reagents, PP1 and PP2 peristaltic pumps, IV injection valve, W waste outlet. Details of the flow cell: (a) optical fibre, (b) sensitive membrane containing the bioreceptor, (c) continuous flow cell cutting out all external light. The bioreceptor can be firefly luciferase for measuring ATP, the oxidoreductase bacterial bienzyme system for measuring NADH or NADPH, or horseradish peroxidase which catalyses the luminol chemiluminescence reaction in the presence of hydrogen peroxide

16.7 Biochips

Biochips or microarrays are miniaturised devices for parallel analysis. The general idea of a biochip is to exploit the specific interaction properties of receptor–ligand type occurring between certain biomolecules. These devices comprise a support measuring a few square centimeters, structured into micro-zones or spots numbering anything from several hundred to several thousand per chip, on which the receptors are immobilised. When brought into contact with the sample, which may or may not contain the corresponding ligands, each receptor will capture its ligand, thereby forming a specific complex on the spot that can subsequently be detected. At the present time, there exist several categories of these biochips: DNA microarrays, protein microarrays, sugar microarrays, and whole-cell microarrays.

The first examples appeared at the beginning of the 1990s. These were DNA chips. It was the enormous project of sequencing the human genome that led to the development of this technology. The genome is the set of all genes carried by the chromosomes of a given organism. Expression of the genome leads to the synthesis of proteins with a primary structure (the amino acid sequence) determined by that of the DNA (deoxyribose nucleic acid). There are two main mechanisms for achieving this expression, namely transcription

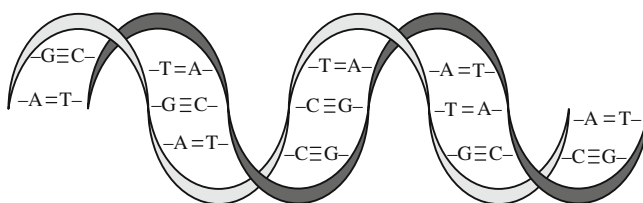


Fig. 16.18. Schematic view of double helix DNA

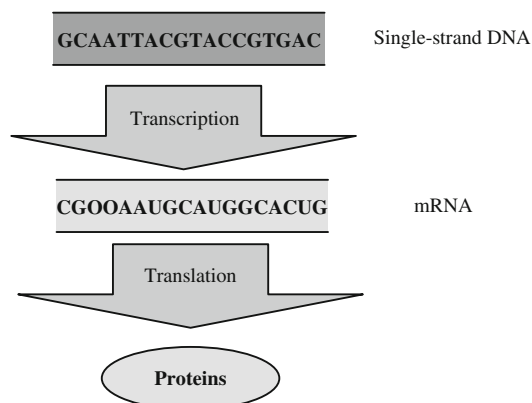


Fig. 16.19. Transcription followed by translation, the process leading to the synthesis of proteins

and translation, which bring in other nucleic acids, the ribonucleic acids or RNA.

The nucleic acids (either ribonucleic or deoxyribonucleic acids) are polymers built up from four different nucleotides (or monophosphate nucleosides) derived from four different bases. Three bases are common to RNA and DNA, viz., adenine (A), cytosine (C), and guanine (G). The fourth base is thymine (T) for DNA and uracil (U) for RNA. The order in which the various bases join up to form a nucleic acid constitutes its primary structure.

DNA is the seat of all genetic information, made up of two complementary strands of nucleic acids, wound into a double helix (double-strand DNA), in which the bases of one strand pair up with those of the other strand by means of hydrogen bonds, always according to the same complementary pairs, C with G and A with T. Three hydrogen bonds form between C and G and two between A and T (see Fig. 16.18).

In transcription, the genetic information carried by DNA is transcribed into RNA, in particular, messenger RNA (mRNA). During this process, the two strands separate locally and temporarily, generating a short single-strand region which serves as a template for the synthesis of mRNA by forming a short DNA–RNA hybrid via the complementary pairs C/G and A/U. The primary structure of the transcribed RNA molecules is thus parallel to that

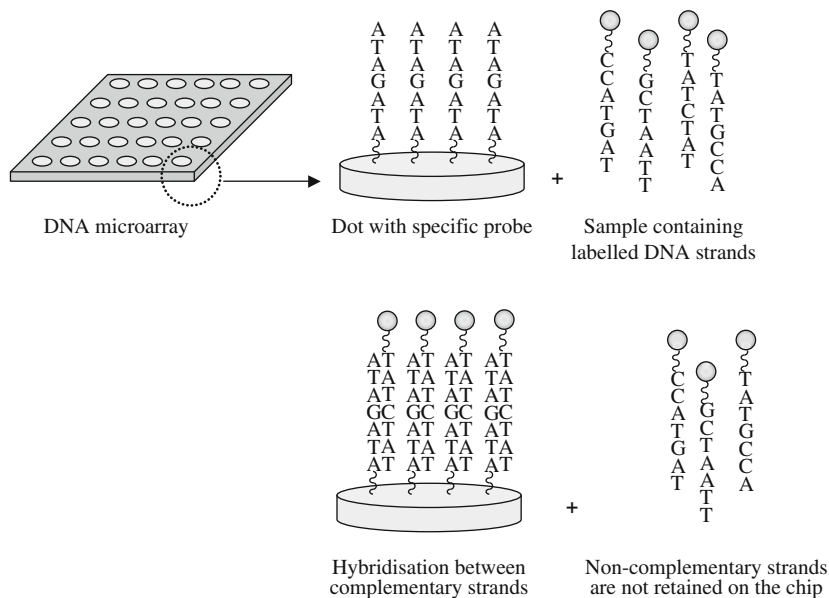


Fig. 16.20. Principle of the DNA microarray

of the original DNA (see Fig. 16.19). The full set of mRNA that can be transcribed from the genome under definite conditions (physiological state, cell type) is called the transcriptome. Technologically, a DNA with complementary sequence to an mRNA template can be synthesised in the laboratory. A synthetic DNA prepared in this way is called complementary DNA (cDNA).

In translation, the structure transcribed on certain RNA molecules is expressed by the synthesis of proteins whose primary structure translates the information carried by the primary structure of the original DNA into amino acids. The set of all proteins expressed in a given cell type under well-defined conditions is called the proteome.

The human genome contains some 30,000 genes, to which there correspond as many different proteins. The genome is the same in each cell, but the genes may have different expression, depending on the stage of development, the cell type, and the normal, pathological, or stressed state of the cell. By analysing the transcriptome, one can thus determine the level of expression of the genes in a cell in a given state. However, information about the concentrations of mRNA is not enough to analyse the regulation of gene expression. The level of activity of the proteins informs as to the variations in cell functions.

It is easy to see that analysis of the genome, transcriptome, and proteome, so important for understanding how a cell works, is extremely complex and requires powerful tools able to carry out thousands of measurements in parallel. This is the task that can be fulfilled by microarray technology.

16.7.1 DNA Microarrays

DNA microarrays (also called DNA chips or gene microarrays) make use of molecular hybridisation by the pairing of complementary sequences of nucleic acid strands (see Chap. 17). The DNA strands immobilised on the microarray are probes that can associate specifically by hybridisation with complementary strands present in a sample. Strictly speaking, there is a difference between a DNA microarray and a DNA chip based on the nature of the immobilised DNA fragments. A DNA microarray is a microsystem integrating cDNA probes of 500–5,000 bases, while a DNA chip comprises oligonucleotide spots of 20–80 bases. Common usage disregards this distinction and the two terms are used interchangeably.

The possibility of miniaturisation provided by microarray technology allows faster and more sensitive testing. Hundreds or thousands of parameters and/or samples can be analysed simultaneously by parallelisation. Existing and potential applications of DNA microarrays concern many different fields, including medicine, the pharmaceutical industry, the food industry, and the environment, as well as fundamental scientific research. Apart from sequencing by hybridisation mentioned above, other applications in which biochips provide a more powerful tool than any conventional technology include research on mutations, the diagnosis of genetic disorders or infectious disease, the identification of targets for therapeutic research, the detection of sequences from genetically modified organisms in crops or foods, and the detection of infectious agents in food, water, or air.

Substrates for DNA Microarrays and Immobilisation of Probes

The basic material used to make DNA microarrays is generally silicon, glass, or a synthetic polymer. Probes can be immobilised directly on these substrates by a suitable chemical treatment setting up a covalent bond between the DNA strands and the chip surface. In some biochips, the surface is structured in such a way that each spot is in fact a gold or platinum microelectrode. The probes can be immobilised either chemically or electrochemically in this case.

In order to fabricate high capacity DNA microarrays, containing several hundred or several thousand spots and hence a corresponding number of different probes, the immobilisation process must be automated. The main challenge here is the problem of addressing, i.e., being able to immobilise a given probe on a spot of given position. Many copies of the same probe are in fact fixed on a given spot. In order to achieve this in a reliable and accurate way, there are basically two different approaches: directly grafting previously synthesised probes, or synthesising the probes *in situ*.

The latter technique, developed by Affymetrix, the world leader in biochip manufacture today, builds up the probes by successively depositing the four constitutive nucleotides of DNA onto a glass substrate. The surface of the spots is modified by binding photolabile chemical groups that can be activated

by exposure to light. Once activated, these groups react with a nucleotide, itself protected by a photolabile group, thereby causing it to bind. Lithographic masks of specific configuration are used to expose and hence activate the spots on the microarray at which a particular nucleotide must be immobilised (A, T, C, or G) in a selective manner. This operation is repeated successively with each nucleotide and with a particular mask until the probes have been fully synthesised. This process can produce microarrays carrying probes of maximum length 30 bases.

Longer probes can be tethered to the spot by directly grafting previously synthesised oligonucleotides. In this case, the probes can be as long as sixty bases. Addressing can then be carried out mechanically using an automated micropipette which delivers to the desired address a tiny volume, of the order of a few nanoliters, of a solution containing the oligonucleotides to be immobilised. A sophisticated variant here is to use inkjet printing technology. Very small droplets of liquid, with volume less than the nanoliter, are projected onto the substrate. Cis-Bio International and the French atomic energy authority (CEA) have developed an electrochemical addressing process for a silicon chip carrying gold spots. A molecule of pyrrole is bound onto each pre-synthesised probe. The probes are then attached by electrochemical copolymerisation with free pyrrole. Addressing is achieved by applying a voltage to a definite electrode, which triggers polymerisation and hence binds the chosen oligonucleotide.

Reading the DNA Microarray

Reading a microarray means identifying addresses at which hybridisation has occurred, and thereby identifying the type, and hence the sequence, of the DNA fragments that have reacted in the given sample. Fluorescence detection is currently the most widely used method and has been almost systematically adopted for the fabrication of DNA chips. This technique requires prior labelling of all potential targets by a fluorophore. Fluorescence signals will only be emitted by spots where labelled targets have hybridised with probes of complementary sequence (see Fig. 16.20).

Other detection methods are currently under development. Rather than using a fluorophore, labelling can be achieved by a molecule that can be detected by chemiluminescence or electrochemiluminescence [39–42]. Some read processes under investigation do not use labelling at all, but they still suffer from a lack of sensitivity and are subject to interference from the sample template. These techniques are based either on the variation of an electrical quantity observed during hybridisation, e.g., a change in impedance or a change in the charge density detected by a field effect transistor, or by surface plasmon resonance.

16.7.2 Protein and Other Microarrays

Protein microarrays are developed both to analyse expression at the protein level, and to analyse specific interactions of certain proteins with other components (see Chap. 18). In other words, some such devices are designed to identify and assay proteins in a sample, while others are developed to study protein–protein, protein–DNA, protein–lipid, enzyme–substrate, or more generally protein–ligand interactions. The basic idea is the same as in DNA microarrays, i.e., the immobilisation of receptors on an array of spots, exploiting similar means of immobilisation but different chemistry.

It is often more difficult to immobilise proteins than to immobilise nucleic acids, owing to the greater fragility of these structures as compared with DNA, and also their greater difference in reactivity. Indeed, a nucleic acid is immobilised via its terminal nucleotide, which has similar chemical reactivity whether it be A, T, C, or G. On the other hand, proteins are immobilised via side chains of the amino acids. Now the involvement of the side chain of an amino acid in a chemical bond can modify the structural and functional properties of a protein. This is why, whenever possible, proteins are immobilised via their *N*-terminal or *C*-terminal amino acid, or via a short peptide sequence artificially adjoined to the protein sequence, or again via their carbohydrate part in the case of glycoproteins.

Different types of protein microarrays can be made depending on the nature of the immobilised detectors. Indeed, depending on the application in mind, the microarray can be designed by immobilisation of different proteins as receptors, or by immobilisation of protein receptors, where these receptors may themselves be proteins, e.g., antibodies, or not as the case may be. In fact, a protein microarray is a ligand–receptor chip, where the proteins can play one or the other role depending on the type of interaction under investigation.

The type of detection used for protein microarrays is the same as for DNA microarrays, with fluorescence being the most common. Other ways of reading the microarray are currently under investigation, including chemiluminescence and electrochemiluminescence methods [43, 44].

For the study of specific sugar–protein interactions, particular sugars in the form of oligomers or polymers are immobilised on a chip in order to detect proteins that interact specifically with these osidic structures. One can then speak of sugar microarrays.

Cells on chips are microsystems designed for individual analysis of living cells with dimensions of the order of a few tens of micrometers (see Chap. 19). Hence the toxic or beneficial effects of such and such a compound can be analysed within a single cell, but in a parallel way. The advantage of miniaturisation is that great savings can be made with regard to what are often very costly samples. The microarray format which achieves this parallelism increases the efficiency of this approach.

16.8 Conclusion

We have described the principal types of biosensor developed to-date. However, this survey does not claim to be exhaustive, not just for didactic reasons, but also because it would be impossible in such a short space to go into all the details of the approaches outlined here.

The dual nature of biosensors, associating a biological element with a physical transducer, undoubtedly confers upon them certain attractive features. The operational character of a biosensor is usually assessed on the basis of four criteria, viz., specificity, stability of the bioreceptor, reusability, and detection threshold:

- Specificity is determined by the bioreceptor, usually an enzyme. It is an intrinsic property of an enzyme and is almost impossible to adjust in any way. Some enzymes are very specific indeed. This is the case for glucose oxidase which recognises only β -D-glucose. Others have a slightly broader specificity, such as the phosphatases, which can catalyse the hydrolysis of a great many natural or synthetic phosphoric monoesters.
- Stability is the ability to maintain specific recognition properties and, in the relevant cases, catalytic properties, over a period of time. Stability is also an intrinsic property of the bioreceptor. However, some experimentally controllable parameters such as the pH, the kind of buffer, or the temperature, can affect the stability. The stability of bioreceptors can be adjusted, preferably positively, by using certain additives. However, these are usually only one-off solutions, valid for a particular bioreceptor or protein.
- Reusability of the sensitive element in the biosensor is made possible by the fact that it is immobilised. However, depending on the type of application in mind (discrete or continuous analysis, warning sensor), the rate of analysis required, and constraints related to the regeneration of the sensitive element after use, this particularity is not necessarily exploited. The idea of a single use, throwaway sensor is becoming more and more common.
- The detection threshold is not always the determining factor. It is often more important to consider the dynamic range to see whether the biosensor is suitable to meet the requirements of the user. For example, an ethanol biosensor for monitoring alcohol fermentation does not require a low threshold, since the samples to be assayed have an alcohol content in the range 0.5–12% (vol/vol), i.e., $0.08\text{--}2\text{ mol L}^{-1}$. On the other hand, other applications will require exceptionally low detection limits, in the picomolar or femtomolar range.

Biosensors have very variable performance. The detection limit and dynamic range depend on intrinsic properties of the biosensor, such as the kinetic characteristics for an enzyme, and the affinity constant for an antibody, but

also on the kind of transducer associated with the molecular recognition. Here are some typical values for the detection limit:

- $0.1 \mu\text{mol L}^{-1}$ for many enzyme electrodes.
- $10\text{--}100 \mu\text{mol L}^{-1}$ for many ENFET devices.
- A few ng for mass transduction biosensors.
- $1 \mu\text{mol L}^{-1}$ for enzyme thermistors.
- $10 \mu\text{mol L}^{-1}$ for optical biosensors with fluorescence detection.
- $0.01 \mu\text{mol L}^{-1}$ for optical biosensors with chemiluminescence detection.
- 10 pmol L^{-1} to 0.1 nmol L^{-1} for optical biosensors with bioluminescence detection.

The main applications of biosensors are in the control of bioprocesses, medical analysis, and the environment. These encompass a wide range of different situations and it is difficult to define a universal biosensor for any given target substance. Depending on the use intended, the architecture of the device can vary significantly. As an example, it is important to monitor blood glucose levels in diabetics, but it is also important to monitor glucose levels in an industrial fermentation process. Starting from the same basic idea, the instrumentation will clearly be designed quite differently, because specifications for optimal use will be very different in the two cases.

Some applications will only require a simple warning biosensor to indicate when some important threshold has been exceeded. It will be important to know whether one is in some specific range of values below the threshold, and in this case the detection limit will be the key feature, rather than the measurement accuracy (environmental control). In other situations, it may be important to know as accurately as possible the concentration of some compound in the medium in order to run a regulation system.

The miniaturisation of biosensors is a permanent concern. In the case of electrochemical biosensors, ISFET devices are already miniaturised transducers for which scale reduction is still possible [45]. Miniaturised amperometric detection sensors are easily designed, either by the technique known as screen printing, or by photolithography. By using masks, screen printing provides a way of depositing films of different composition and geometry on a solid substrate. Various structured molecular assemblies can be constructed by immobilising biomolecules in or on these films. Electrodes with dimensions in the range of a few tens of μm can be obtained by screen printing, while techniques in microelectronics can reach much smaller length scales. In this area, carbon nanotubes could very probably be used to make nanoscale electrodes for the design of nanobiosensors.

Optical sensors also lend themselves well to miniaturisation. Indeed, devices using optical fibres with diameter a few tens of μ have already been reported. In addition, optical sensors without fibres, in biochip format, have been described recently, in which the substrate used for immobilisation is the light sensor itself (CMOS photodetector) [46].

Miniaturising the transducer involves particular methods for immobilising the bioreceptor. Indeed, for areas of a few μm^2 , or even a few nm^2 , it is impossible to immobilise the bioreceptors on synthetic membranes which are then intimately associated with the transducer. Innovation is required here. Various solutions have already been put forward and used for biochips, wherein a very small volume ($<1\text{nL}$) of bioreceptor solution can be deposited in a very precisely defined position. Other ideas currently under investigation use artificial lipid nanostructures in monolayers or bilayers (see Chap. 2).

The development of the lab on a chip (see Chap. 20), a genuine miniaturised laboratory, involves not only the development of biochips, but also miniaturised bioreceptors.

Finally, it must be emphasised that the elaboration of these miniaturised systems can only be achieved by the cross-disciplinary collaboration of biologists, biochemists, physicists working in microelectronics, optics, and fluid mechanics, and chemists.

References

1. Clark, L.C., Jr., Lyons, C.: Ann. NY Acad. Sci. **102**, 29–45 (1962)
2. Updike, S.J., Hicks, G.P.: Nature **214**, 986–988 (1967)
3. Guilbault, G.G., Montalvo, J.G.: J. Am. Chem. Soc. **91**, 2164–2165 (1969)
4. Blum, L.J., Coulet, P.R.: *Biosensors. Principles and Applications*, Marcel Dekker, New York (1991)
5. Yang, V.C., Ngo, T.T.: *Biosensors and their Applications*, Kluwer Academic, New York (2000)
6. Cooper, J., Cass, T.: *Biosensors. A Practical Approach*, 2nd edn., Oxford University Press, Oxford (2004)
7. Stevenson, J.D., Thomas, N.R.: Nat. Prod. Rep. **17**, 535–577 (2000)
8. Haupt, K.: Analyst **126** (6), 747–756 (2001)
9. Bell, T.W., Hext, N.M., Khasanov, A.B.: Pure and Appl. Chem. **70** (12), 2371–2377 (1998)
10. Mosbach, K.: *Methods in Enzymology*, Vol. 44, *Immobilized Enzymes*, Academic Press, New York (1976)
11. Mosbach, K.: *Methods in Enzymology*, Vols. 135–137, *Immobilized Enzymes and Cells*, Academic Press, New York (1987)
12. Taylor, R.F.: *Protein Immobilization*, Marcel Dekker, New York (1991)
13. Bergveld, P.: IEEE Trans. Biomed. Eng. BME **17**, 70–71 (1970)
14. Wolfbeis, O.S.: *Fiber Optic Chemical Sensors and Biosensors*, Vols. 1 and 2, CRC Press, Boca Raton (1991)
15. Schultz, J.S., Mansouri, S., Goldstein, I.J.: Diabetes Care **5**, 245–253 (1982)
16. Ferguson, J.A., Boles, T.C., Adams, C.P., Walts, D.R.: Nature Biotechnology **14** (13), 1681–1684 (1996)
17. Blum, L.J.: *Bio- and Chemiluminescent Sensors*, World Scientific, Singapore (1997)
18. Freeman, T.M., Seitz, W.R.: Anal. Chem. **50**, 1242–1246 (1978)
19. Blum, L.J., Gautier, S.M., Coulet, P.R.: Anal. Lett. **21**, 717–726 (1988)

20. Blum, L.J.: Enzyme Microb. Technol. **15**, 407–411 (1993)
21. Berger, A., Blum, L.J.: Enzyme Microb. Technol. **16**, 979–984 (1994)
22. Hlavay, J., Guilbault, G.G.: Anal. Chim. Acta **299**, 91–96 (1994)
23. Lapp, H., Spohn, U., Janasek, D.: Anal. Lett. **29**, 1–17 (1996)
24. Tsafack, V.C., Marquette, C.A., Pizzolato, F., Blum, L.J.: Biosensors Bioelectronics **15**, 125–133 (2000)
25. Degiuli, A., Blum, L.J.: J. Med. Biochem. **4**, 32–42 (2000)
26. Marquette, C.A., Blum, L.J.: Anal. Chim. Acta **381**, 1–10 (1999)
27. Marquette, C.A., Ravaut, S., Blum, L.J.: Anal. Lett. **33**, 1779–1796 (2000)
28. Marquette, C.A., Leca, B.D., Blum, L.J.: Luminescence **16**, 159–165 (2001)
29. Tsafack, V.C., Marquette, C.A., Leca, B., Blum, L.J.: Analyst **125**, 151–155 (2000)
30. Leca, B., Blum, L.J.: Analyst **125**, 789–791 (2000)
31. Leca, B., Verdier, A.M., Blum, L.J.: Sensors Actuators B **74**, 190–193 (2001)
32. Blum, L.J., Gautier, S.M., Coulet, P.R.: Anal. Chim. Acta **226**, 331–336 (1989)
33. Gautier, S.M., Blum, L.J., Coulet, P.R.: Anal. Chim. Acta **235**, 243–253 (1990)
34. Gautier, S.M., Blum, L.J., Coulet, P.R.: J. Biolumin. Chemilum. **5**, 57–63 (1990)
35. Michel, P.E., Gautier, S.M., Blum, L.J.: Anal. Lett. **29**, 1139–1155 (1996)
36. Michel, P.E., Gautier, S.M., Blum, L.J.: Enzyme Microb. Technol. **21**, 108–116 (1997)
37. Michel, P.E., Gautier, S.M., Blum, L.J.: Anal. Chim. Acta **360**, 89–99 (1998)
38. Michel, P.E., Gautier, S.M., Blum, L.J.: Talanta **47**, 167–181 (1998)
39. Marquette, C.A., Thomas, D., Degiuli, A., Blum, L.J.: Anal. Bioanal. Chem. **377**, 922–928 (2003)
40. Marquette, C.A., Blum, L.J.: Anal. Chim. Acta **506**, 127–132 (2004)
41. Marquette, C.A., Blum, L.J.: Biosensors Bioelectronics **20**, 197–203 (2004)
42. Calvo-Muñoz, M.L., Dupont-Filliard, A., Billon, M., Guillerez, S., Bidan, G., Marquette, C., Blum, L.J.: Bioelectrochemistry **66** (1–2), 139–143 (2005)
43. Marquette, C.A., Degiuli, A., Imbert-Laurenceau, E., Mallet, F., Chaix, C., Mandrand, B., Blum, L.J.: Anal. Bioanal. Chem. **381**, 1019–1024 (2005)
44. Marquette, C.A., Imbert-Laurenceau, E., Mallet, F., Chaix, C., Mandrand, B., Blum, L.J.: Anal. Biochem. **340** (1), 14–23 (2005)
45. Galdin-Retailleau, S., Bournel, A., Dollfus, P.: In: *Nanoscience, Vol. 1, Nanotechnologies and Nanophysics*, Springer, Berlin Heidelberg New York (2007) Chap. 11
46. Mallard, F., Marchand, G., Ginot, F., Campagnolo, R.: Biosensors Bioelectronics **20**, 1813–1820 (2005)

DNA Microarrays

C. Nguyen and X. Gidrol

17.1 Introduction

Genomics has revolutionised biological and biomedical research. This revolution was predictable on the basis of its two driving forces: the ever increasing availability of genome sequences and the development of new technology able to exploit them. Up until now, technical limitations meant that molecular biology could only analyse one or two parameters per experiment, providing relatively little information compared with the great complexity of the systems under investigation. This gene by gene approach is inadequate to understand biological systems containing several thousand genes. It is essential to have an overall view of the DNA, RNA, and relevant proteins. A simple inventory of the genome is not sufficient to understand the functions of the genes, or indeed the way that cells and organisms work. For this purpose, functional studies based on whole genomes are needed. Among these new large-scale methods of molecular analysis, DNA microarrays provide a way of studying the genome and the transcriptome. The idea of integrating a large amount of data derived from a support with very small area has led biologists to call these chips, borrowing the term from the microelectronics industry.

At the beginning of the 1990s, the development of DNA chips on nylon membranes [1, 2], then on glass [3] and silicon [4] supports, made it possible for the first time to carry out simultaneous measurements of the equilibrium concentration of all the messenger RNA (mRNA) or transcribed RNA in a cell. These microarrays offer a wide range of applications, in both fundamental and clinical research, providing a method for genome-wide characterisation of changes occurring within a cell or tissue, as for example in polymorphism studies, detection of mutations, and quantitative assays of gene copies. With regard to the transcriptome, it provides a way of characterising differentially expressed genes, profiling given biological states, and identifying regulatory channels.

With this technology, the characteristic molecular signatures of certain types of human tumour have been established. Examples are leukemia [5]

and breast cancers [6, 7]. Characteristics of new groups of patients that are undetectable to conventional histological methods have also been identified. There are many medical applications of DNA chips. It should be possible to extend the characterisation and use of molecular signatures in expressed genes to many pathologies, helping clinical medicine to meet the challenge of early diagnosis, prognosis, and therapeutic orientations.

DNA Microarray: Definition

DNA microarrays are solid substrates on which thousands of DNA molecules corresponding to the genes under investigation are deposited in an ordered arrangement. Based on the idea of molecular hybridisation, these arrays can carry out studies that were previously done using standard techniques of molecular biology (Northern or Southern blots), but with two major differences:

- The known genomic sequence is immobilised on the solid substrate.
- The system is miniaturised in such a way that many probes can be integrated in parallel.

Several tens of thousands of probes on a support measuring a few square centimeters achieve such a high level of integration that these studies can be carried out on all the genes of a complex organism in a single experiment.

However, in order to carry out genome-wide studies of complex biological systems, new tools, and in particular, new biochips are required. Today it has become possible to use DNA chips for other applications than the study of the transcriptome. Varied combinations of probes, i.e., biological entities deposited or synthesised on the chip, and targets, biological entities to be studied with the chip, open the way to a wide range of applications, some of which are listed in Table 17.1.

17.2 Analysing the Transcriptome

17.2.1 Basic Idea

DNA chip technology is based on the hybridisation of an ordered set of DNA molecules (the probes) bound onto a solid support with some labelled target, prepared from the DNA or RNA of a biological sample. This exploits the remarkable ability of the sequences in two complementary strands of DNA to pair up with perfect fidelity, despite the great complexity of the molecule. Studies carried out in solution in the 1960s showed that the level of reassociation of the duplex depended on the percentage of G + C and the salt concentration [8], and that a single strand fixed on a membrane inhibits reassociation with its complementary strand, while allowing hybridisation with its complementary RNA in solution [9].

During hybridisation, the labelled molecules in solution will bind onto the corresponding DNA molecules on the microarray. After rinsing and exposing,

Table 17.1. Different types of DNA microarray. Probes are biological entities deposited or synthesised on the chip, and targets are biological entities to be assayed using the chip. Abbreviations: CGH = comparative genome hybridisation, ChIP on chip = chromatin immunoprecipitation on chip, siARN = short interfering RNA, BAC = bacterial artificial chromosome, AB = antibody

Type	Application	Probe	Target
DNA microarray	Transcriptome	cDNA, oligonucleotides of the reading frame	Messenger RNA
ChIP on chip	To analyse sites where proteins bind to the genome	cDNA, oligonucleotides of promoter or intergenic regions	Genomic DNA immunoprecipitated by a specific AB
CGH array	To analyse chromosome rearrangements	cDNA clones of BAC	Genomic DNA
siRNA microarray	To study genes with gain of function	Double-stranded DNA, plasmids	Cells
Expression microarray	To study genes with loss of function	Plasmids	Cells

an image of the hybridisation can be obtained with quantifiable signals, whose intensities vary with the spot, and which in fact represent the level of amplification or expression of the corresponding gene [2, 10, 11]. These intensities are then quantified and normalised, and the results analysed and visualised using the ever more sophisticated tools of biocomputing. The whole process is depicted in Fig. 17.1.

17.2.2 Different Types of DNA Microarray for Transcriptome Analysis

There are two types of DNA microarray: those made by depositing complementary DNA (cDNA) and those made using oligonucleotides, either by deposition on a solid substrate as for cDNA, or by in situ synthesis.

Different Types of Support and Fabrication

The first microarrays were made by depositing bacterial clones as-is on solid substrates. In this case, the support was a nylon membrane, often treated to bind the DNA [1]. These arrays, referred to formally as high density filters, in which the clones are deposited robotically in an ordered arrangement, were the precursors of macroarrays fabricated from the amplification products of these same clones by PCR. These systems fall into two classes, depending on the level of integration of the probes: macroarrays integrating a few tens of probes up to a few hundred probes per square centimeter (with spots

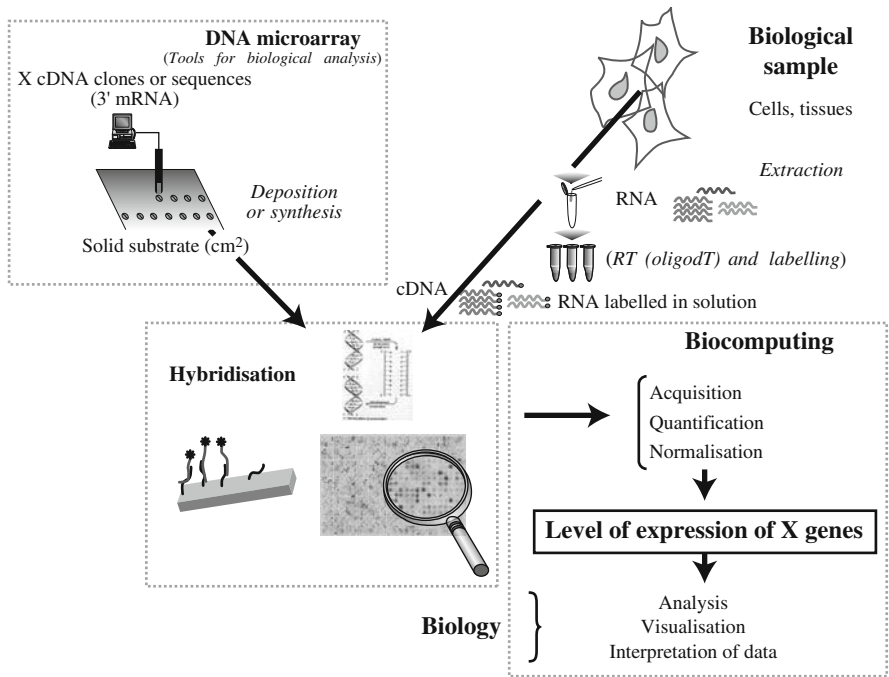


Fig. 17.1. The five components of DNA microarray analysis: (1) The DNA microarray, which reveals the event. (2) The biological sample and its treatment. (3) The hybridisation underlying the mechanism that reveals the event. (4) Computer analysis. (5) Understanding one’s biological model for the purpose of interpretation

one or two millimeters apart), and microarrays integrating several thousand probes per square centimeter [12]. These porous substrates provide a large area for deposition and, thanks to their physicochemical characteristics, they are reusable (dehybridisation/rehybridisation), allowing repeated, independent, and cheaper experiments [13].

At the same time, cDNA microarrays were developed on previously treated glass slides, since this technology, associated with the detection of fluorescent signals, allows a higher probe integration density. These slides, often microscope slides, without asperities, rigid, and transparent, improve image acquisition and processing. They are coated with a poly-lysine film, or a layer of amino-reactive silanes, which increase both the hydrophobicity and the adhesion of the DNA deposited there. Furthermore, they limit the diffusion of the deposited DNA droplet [14]. This support has the advantage of being rigid and being able to tolerate high temperatures without deformation [15, 16]. Owing to its non-porous nature, it permits a very low volume of hybridisation and thereby increases the reaction rate. However, it does have the disadvantage that the available binding area is lower and therefore reduces the number

of probes per μm^2 . It has negligible autofluorescence, making it the choice support for signal detection by fluorescent labelling. The latter method can be used to label two or more different RNA molecules with different fluorochromes, and analyse them simultaneously.

cDNA Microarrays

In this case, one must first establish a collection of probes specific to all the genes one hopes to analyse, e.g., the 25,000 genes currently identified in the human genome. This collection is made from cDNA clone libraries obtained by reverse transcription of the messenger RNA from a given cell or tissue. Clones are selected in terms of their sequence in the form of an expressed sequence tag (EST), which can be used to identify them [17,18]. In this context, the international approach proposed by the IMAGE consortium, which has produced several million partially sequenced clones, has greatly stimulated, if not laid the foundations for the microarray transcriptome approach [19,20]. By sorting these pieces of sequence using the tools of biocomputing, it has been possible to establish minimal sets of clones representative of single genes [21]. The advantage of these arrays is their flexibility, with the possibility of tailoring them to contain the relevant genes, at a rather low cost compared with oligomer microarrays. However, a certain level of care is required at all times when managing the clones and PCR products, to obtain and maintain the collections in a good state.

The difficulties encountered with cDNA methods stem from the fact that there is 10–20% error in the correspondence between the cDNA and the gene it came from, combined with the fact that cDNA does not provide a way of differentiating the responses of homologous genes belonging to multigene families, and, above all, that these methods require constant updating of cDNA libraries. For these reasons, the scientific community has tended to favour oligonucleotide methods, which are easier to control.

Oligomer Microarrays

The second type of microarray, the oligomer microarray, is based entirely on knowledge of the genome sequence under investigation and requires neither the prior establishment of clone collections nor the preparation of PCR products. Generally speaking, the oligomer microarrays were developed on silicon or glass substrates, pretreated so that nucleic acids could be bound covalently onto them. One difficulty with this approach lies in the choice of specific oligonucleotides of each gene. The latter must have similar behaviour during hybridisation (close melting temperature), while being highly specific to the genes under investigation [22,23].

In this category of microarray, one finds once again fabrication by direct deposition of previously synthesised oligonucleotides and fabrication by *in situ* synthesis of the oligomer sequences, based on several stages of combinative

coupling. In the first case, different types of distributor deliver previously synthesised oligomers to the substrate surface, depositing with solid or hollow pins [15, 24–26], or by inkjet [27]. The latter deposition technique is also used to synthesise the oligomers directly on the support [28].

The second type of microarray exploits systems in which oligonucleotides are synthesised *in situ* by a series of reactions in which the reactive groups are alternately protected then deprotected, until all the various oligomers have been obtained to the required length. There are various systems for deprotecting the reactive groups.

The technology developed by Affymetrix uses light emission through a protective mask, only letting the light through to selected sites on the surface [29]. This synthesis, carried out on a silicon substrate by photolithography, derives from methods used in microprocessor fabrication. Extremely localised photodeprotection provides a way of carrying out the chemical reactions of the synthesis in a bath, focussing on selected, deprotected nucleotides that are already covalently bound to the substrate. The system is iterated until the desired oligomer length has been obtained [30]. Oligonucleotides of length 20–25 mers can be obtained, specific to each gene. This choice of length (20–25 mers) is due to fabrication constraints, compelling one to represent each gene by several oligonucleotides (about twenty) to compensate for the relative lack of specificity of these elements for hybridisation.

Another approach to fabrication by photolithography has been developed without using masks. This is the maskless array synthesiser (MAS), which simply replaces the traditional chrome mask by a virtual mask generated by a computer-oriented system of micromirrors (the digital micromirror device or DMD) [32]. The software synchronises the light exposure and the ensuing cycles of chemical synthesis. The steps are repeated with different virtual masks until the desired length of oligonucleotide is obtained, and this according to the required pattern. The advantage here is the considerable flexibility for tailoring the microarray to desired specifications, the fact that it does not require any special environment, and the very precise synthesis, able to produce very high resolution microarrays. An array with 76,000 spots has been synthesised on an area of $16\text{ }\mu\text{m}^2$, with a system which should theoretically allow the synthesis of half a million oligonucleotides with different sequences. This technology has a very high integration capacity and is cheaper than the Affymetrix technology, but requires laser detection systems associated with confocal microscopy.

The synthesis can also be achieved by confining the chemical components physically, using masks or barriers, in the same way as before. This process can synthesise highly complex arrays in several stages using combinatorial methods [16]. One of the advantages with this method is that one can test the quality of the oligomer synthesis and study the hybridisation characteristics of the nucleic acids very precisely.

All these methods aim for a very high probe integration density, up to several hundred thousand different oligonucleotides per square centimeter,

with a high potential for miniaturisation, and they have the advantage of being able to synthesise the chosen sequence anywhere on the substrate surface.

Third Generation Microarrays

The aim in the last section was not to give an exhaustive review of all the developments in the field of DNA microarrays. However, it should be stressed that each method has its strengths and weaknesses, and that today much still remains to be done to meet the ideal fabrication criteria for a DNA microarray. One aim is to be able to specify and fabricate an array very quickly (in a few hours). Another is to be able to program this step in such a way as to specify the sequence and the site for each experiment. One would like to have a spatial resolution that would make it possible to reduce the size and the amount of biological material used in hybridisation reactions. One would like to be able to bind the probes in 3 min or 5 min, with or without intercalators and with the desired dimensions, in order to be able to test the different methods of enzyme extension and/or annealing.

Clearly, no method can meet all these requirements, so the way is open for further development. As an example, let us briefly describe the synthesis of oligonucleotides by direct electrochemical addressing and the principle of arrays on beads.

The direct synthesis of oligomers by electrochemistry is achieved on a solid substrate in contact with a microelectrode array able to induce an extremely localised production of acid by passing a current through each electrode and thereby deprotecting the target molecule. With this system, the reaction can be controlled in a bath, by minimising the formation of secondary products at each step required for the synthesis.

Using this approach, it is easy to access any point on the surface and to vary the sequences as required from one fabrication to the next. Since with this system there is no displacement between each coupling step, there is no loss of information between the printing tool and the surface to be processed, thereby avoiding any degradation of the chemical image of the final compound. This is an application of fast, conventional chemical methods, in which the synthesis does not depend on the orientation of the oligonucleotide or its analogue, and for which standard reagents are available. This is clearly a method that can be applied to the synthesis of other molecules to generate other types of array [33], but which does not yet fulfill all the criteria mentioned above.

The idea of an assembly of beads on a surface, in which each bead carries many copies of the sequence of a gene, may well be the approach that revolutionises future genome analysis, in terms of systematic sequencing, and analysis of polymorphisms and the level of gene expression.

The beads have a diameter of 3 μm . On each bead is fixed: (1) n times the same oligomer corresponding to the gene sequence one seeks to assay, (2) a sequence for identifying the bead (see Fig. 17.2) [36]. A first study was carried out with 1,536 different beads, where each type of bead could bind up to 70,000

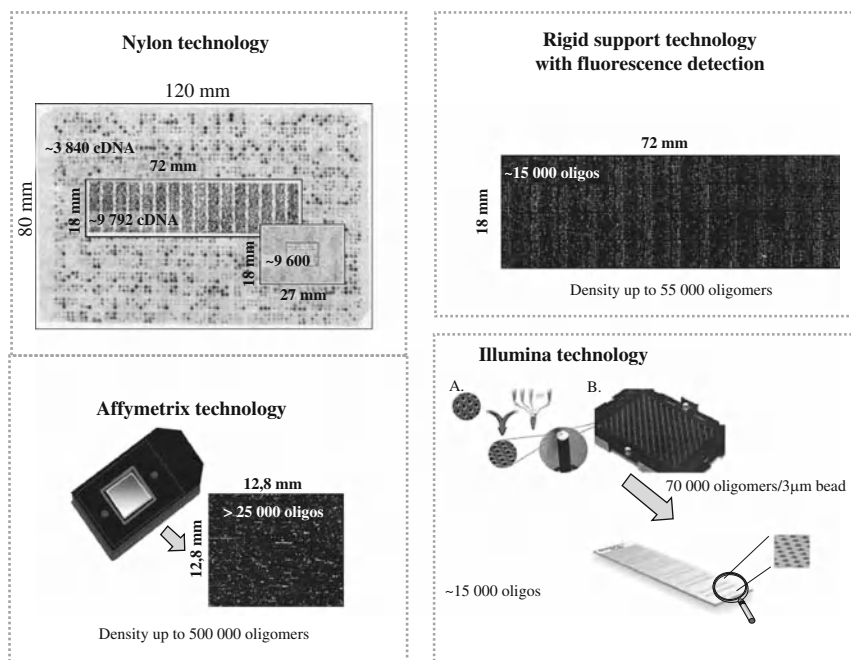


Fig. 17.2. Different types of microarray in use today. Since the nylon technologies appeal to either radioactive or colorimetric detection, their integration capacities depend on the spotting system, the physicochemical reaction of the substrate, and the detection system. Among the technologies using rigid substrates, one finds the counterpart of the previous technologies, but with very high integration densities, developed in both academic and industrial laboratories. The Affymetrix and Illumina technologies represent the world of industry. From [34–36]

copies of the relevant oligomer. These beads were then assembled in boxes of different shapes. Since these boxes were able to contain up to 50,000 different beads in the first experiment, this amounts to having each category of bead 30 times per box on average. While the viability of the concept was first proved using standard systems, i.e., 96-pit plates, the idea developed very quickly on glass supports with the dimensions used by all other technologies, whence it was possible to use read systems already available in research laboratories. This technology has a real potential for miniaturisation. One can work with very small amounts of sample and obtain results based on studies with very robust statistics, generally difficult to achieve with other approaches. Indeed, this kind of analysis depends on the number of experiments that can be carried out and hence on the size and number of biological samples [37].

17.2.3 Some Applications of DNA Microarrays

To conclude, let us mention a certain number of examples that will not be further discussed in this chapter. One is the detection of pathogenic microorganisms within biological samples using an array that integrates specific targets of each of these microorganisms [38, 39]. This approach can be particularly useful for determining the presence of pathogenic organisms in food products or during infections, or simply for detecting certain strains of bacteria, and so on.

Another application based on the idea of sequencing by hybridisation [40], used to check known sequences, can now be applied to *de novo* sequencing thanks to the bead technology [41].

The idea of sequencing by hybridisation has also been applied to the detection of mutations by resequencing [42], in which one tries to characterise, within a DNA sample, the mutations of the relevant genes, e.g., involved in some kind of pathology. This type of microarray has been used to characterise variants of microorganisms with different degrees of pathogenicity [43], the mutations of the human mitochondrial genome, the mutations of genes such as BRCA1 [44], or again the HIV protease gene [45], etc.

A final application is the characterisation, within a genome, of sites of single nucleotide polymorphism (SNP). These SNP microarrays can simultaneously investigate the thousands of polymorphism sites across the whole genome and draw up a map [46]. They constitute one of the tools for characterising multifactorial disease, especially for research into genetic linkage between polymorphism sites and phenotypes [47, 48].

17.2.4 Some Remarks Concerning Transcriptome Data Analysis

Data analysis remains today the most difficult aspect of microarray assays. It can be divided into different steps. The first, related to the choice of technology, are relatively standardised, but this is not the case for methods used to interpret the data.

To identify candidate genes for a biological study or related to some pathology, the first approaches were based on methods for filtering the data, the aim being to reduce the sheer quantity of data generated by the microarray, and also to eliminate as many false positives as possible.

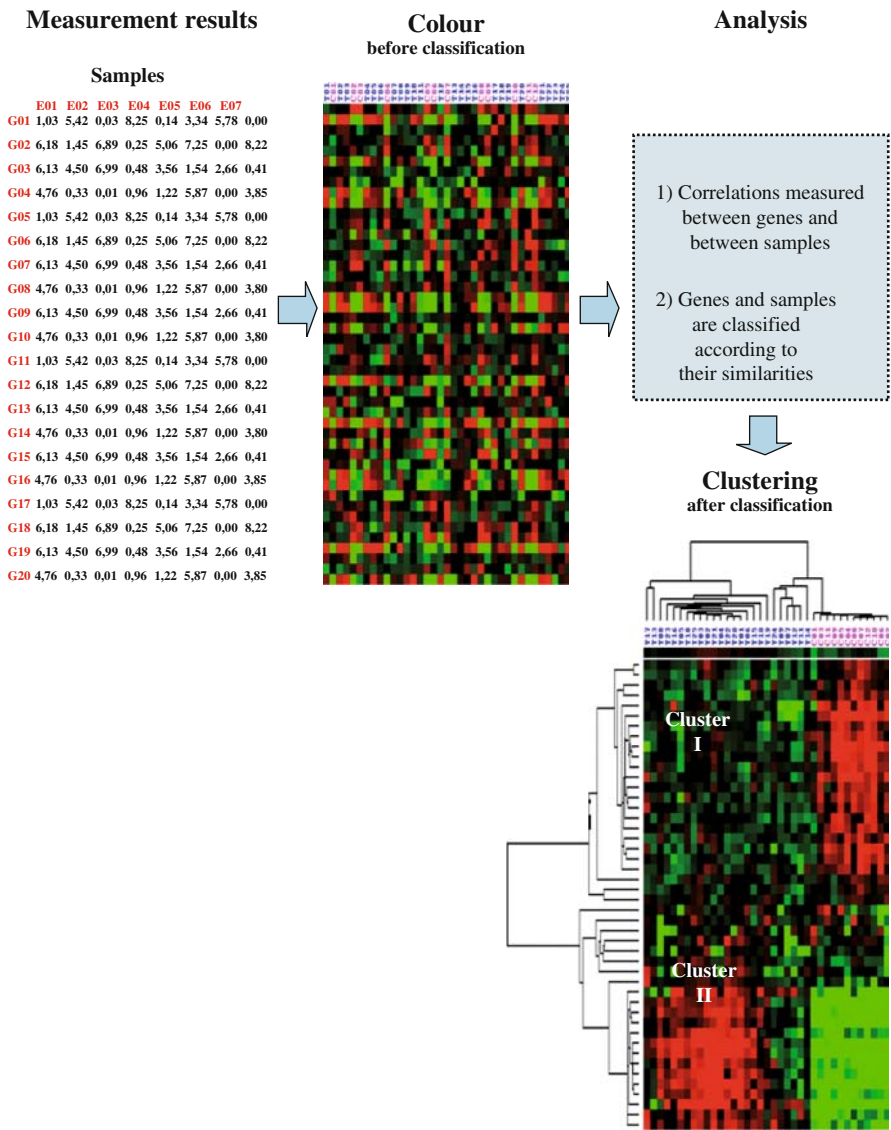
Since then, there have been many developments, with the advent of new methods of statistical analysis, whose relevance to the problems of biology has not always been demonstrated. The methods require cooperation between statisticians and biologists, a difficult problem in itself, due to the lack of mutual understanding between two communities using such different languages. Today, while some of the difficulties encountered in data analysis have been resolved, there is still no ready-made solution. This has done nothing to stop the scientific community from tackling the transcriptome problem, as attested by the large number of publications in PubMed between 1998 and 2004 [49].

17.2.5 Transcriptome Applications

It is the possibility of simultaneously assaying the expression of several thousand genes in a biological sample that has thrust the DNA microarray to the fore in this area of biology. It provides an overview of the workings of cells and tissues in a normal, pathological, or experimental environment, used originally to identify genes of unknown function. The next task was to establish a catalogue of genes expressed in each cell or tissue of interest, so that the researcher could select the best candidate genes for whatever biological question had been posed. This is quantitative differential screening. Whatever technology was chosen, this approach was used with success, as attested by the many publications, some of which will be cited here as examples. Differential screening was used to identify genes with specific expression in muscle tissue [50], genes associated with a pathological particularity such as tumour characters [51], and genes playing a fundamental role in some aspect of development, such as thymic ontogenesis. In the latter case, work is based on transgenic models in mice [52]. The main limitation here is that the method does not inform as to the reasons for the observed modulation of expression. In particular, the expression of some genes varies in ways that are unrelated to the original question. So an assay involving several samples of the same type, subjected to several stimulation and/or inhibition treatments and sampled at different times, supplies more biological data than a simple comparison between two states.

This observation led biologists to the idea of gene expression profiling, where the objective is molecular characterisation of a given state given various samples associated with a phenotype, e.g., the different tumour states [6, 53]. In this approach, comparative analysis of the profiles aims to classify samples on the basis of the transcriptome, in the hope of spotting new subtypes not yet identified by conventional factors and relevant to diagnosis and/or prognosis. Whatever biocomputing tools are used, the idea is to group together the most similar samples on the basis of their expression profile and also to group together the genes with similar expressions in the studied samples. Among the various classification systems that are still the subject of a great deal of research in biocomputing, the most widely used up to present has been the hierarchical classification software developed by Eisen and coworkers at Stanford: the cluster program (see Fig. 17.3) [54]. This method measures the degree of similarity of samples and genes by the Pearson correlation coefficient, calculated from the level of gene expression. These correlation distances provide a way of grouping together the most clearly related profiles, in the form of a similarity/distance graph. A colour scale is used to visualise the classification of the results, which are generally represented in the form of a colour scale matrix.

In the rather complex situations of human multifactorial pathologies, expression data processing provides as efficient a classification as standard histoclinical methods [55, 56]. But more importantly, in some cases it leads to



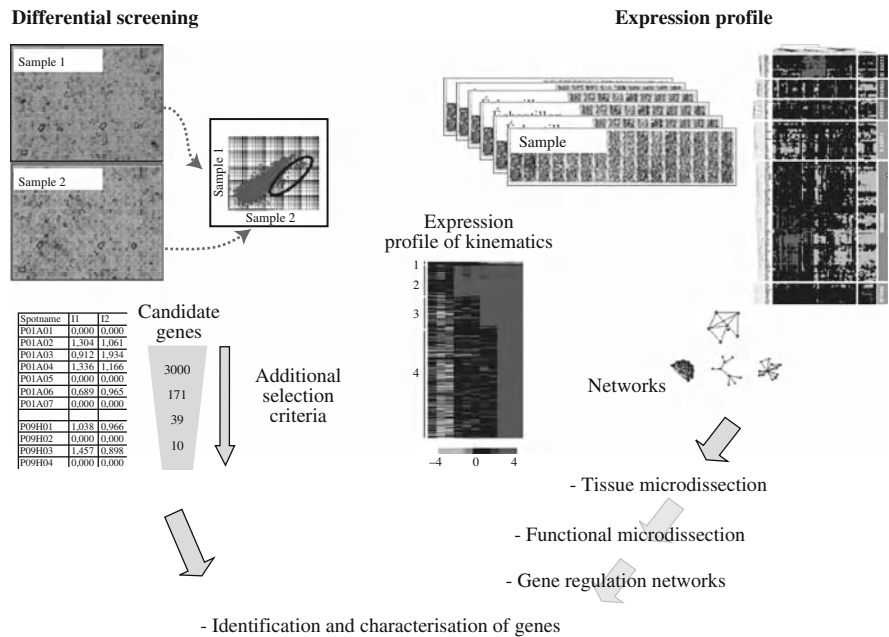


Fig. 17.4. Different applications depending on the type of analysis based on gene expression

differential screening, some groups have physically microdissected tissue samples in order to assay expression profiles of homogeneous regions, the idea being to minimise the number of irrelevant genes resulting from the complexity of these tissues [59,60]. In parallel, the opposite idea has been developed, in which one tries to achieve microdissection of the tissues through the expression profile. This novel idea, using cell lines isolated from different tissues, exploits the fact that there must be a logical connection between the function of a gene and its expression profile in the various tissues. A corollary is that the expression of a gene with known function can reveal a new feature of the cell or tissue phenotype [61].

This brings us to the idea of virtual microdissection, as described in the research by our group [62]. This work uses models of transgenic mice invalidated for the genes involved in thymocyte ontogeny. This study showed that regions of an organ as complex as the thymus could be subjected to virtual microdissection. The expression profiles identified different cell compartments and cell types in the various stages of their development. Moreover, the study showed that the method can reveal some of the regulatory channels which control the functioning of cells and tissues.

In this last application of the transcriptome, which concerns the characterisation of regulatory channels, there is no difficulty in building up a catalogue

of studies in which expression profiles are associated with regulatory channels (see Fig. 17.4). One finds studies of gene inactivation mutants blocking the signalling of a metabolic cascade, and all those approaches concerned with effector response kinetics. The latter reveal successive waves of activated and repressed genes in the metabolic channel of the target molecule, and also on other coregulated channels. This type of analysis not only establishes a clear view of cell activity under varied conditions, but also reveals possible points for therapeutic intervention within the cell.

To conclude, it looks more and more likely that genes with similar expression profile, i.e., coexpressed genes, have a high probability of sharing regulation mechanisms. One of the coregulation mechanisms involves target sequences in the gene promoters, and these sequences may be identical or similar, and recognised by one or more shared transcription factors, as has been shown for yeast [63]. This type of approach, combining DNA microarrays, hierarchical classification algorithms, and the systematic search for DNA consensus sequences in potential promoters can determine transcriptional regulation networks in sequenced organisms [64].

In the first part of this chapter, devoted to DNA microarrays, we have discussed at length the applications to transcriptome analysis. In the following, we shall review other applications of DNA microarrays (see Table 17.1), going beyond the transcriptome, explaining how they work and giving examples to illustrate their use in biomedical research.

17.3 Beyond the Transcriptome

17.3.1 CGH Microarrays

Basic Idea

Comparative genome hybridisation (CGH) is used to study the large chromosome rearrangements, such as gene duplications or deletions, in metaphase cells. This method uses a reference DNA and a test DNA labelled with two different fluorescent labels, which are then hybridised competitively onto metaphase normal chromosomes in cells deposited on a microscope slide [65]. The detected fluorescence ratio indicates the relative number of each copy of DNA in the test DNA as compared with the reference DNA. However, the limiting factor with this approach and its applications is the low resolution, somewhere between 5 and 10 Mbases for deletions and around 2 Mb for amplifications. More recently, this problem has been partly resolved by the CGH microarray, i.e., CGH on a chip, which provides much better resolution. This technology is based on the same idea as conventional CGH, but genomic clones (fragments of DNA cloned in a plasmid) are ordered then deposited on a chip (see Fig. 17.5). The first CGH microarray was reported in 1997 [66]. The first chip covering the whole human genome with 3,360 cDNA clones was used to

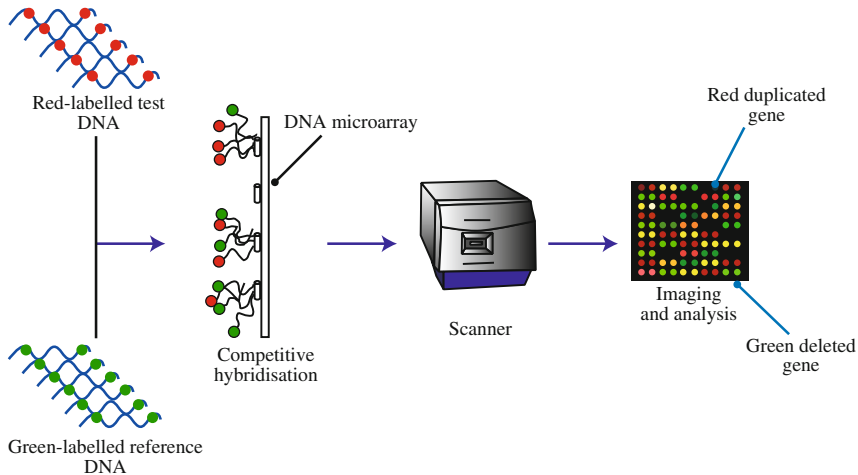


Fig. 17.5. CGH microarray. Deposition is carried out by a robot equipped with an array of 48 hollow pins, each with diameter $80\text{ }\mu\text{m}$. These deposit about 1 nL of probes on each spot of the array. The probes deposited on the array are either bacterial artificial chromosomes (BAC) or oligonucleotides. The target is genomic DNA extracted from cells. The aim is to study chromosome rearrangements (duplication or deletion of genes) in the tested genome

map the amplifications and deletions of genes in human breast cancer [67]. The CGH microarray is currently the most efficient way of simultaneously detecting and localising both gains and losses of genetic material. Many studies use this technique, especially in cancerology.

CGH microarrays have many advantages. For example, one does not have to work on dividing cells as in conventional CGH. The whole genome can be analysed in a single experiment, rather than with a thousand independent fluorescent in situ hybridisation (FISH) experiments. The resolution is much improved, varying between 1 and 5 Mb, but as low as 100 kb if more clones are deposited on the chip. The chip is highly specific and generates very few false positives and practically no false negatives. It also provides better sensitivity than conventional CGH. Finally, this technology is fast because part of the procedure is automatic. The number of clones on the chip, and also the choice of method for amplifying the clones, are crucial for the final quality of the experiments. CGH microarrays also have a number of limitations, e.g., they cannot detect chromosome abnormalities that are not reflected by significant changes in the number of copies of the genes. The background noise and quality of the clones can also limit their potential.

Applications

Today, CGH is the most efficient method for detecting and localising all sorts of chromosome abnormalities resulting from the loss or gain of genetic material. However, we shall only mention here a few specific applications of CGH biochips.

The Prader–Willi syndrome (PWS) results from deletions of 3–5 Mb on chromosome 15. These deletions cannot be detected with conventional CGH. Using a chip that covers the whole human genome with a resolution of 1 Mb, deletions of 1.5 to 2.9 Mb have been detected in three patients, in the suspected region of chromosome 15 [68]. Likewise, in collaboration with the team at the Gustave Roussy Institute (in France), we have developed a chip able to pinpoint abnormalities that cannot be detected by traditional methods in neuroblastoma patients [69].

Congenital aural atresia (CAA) is often reported in patients with chromosome abnormalities, in particular, deletions in the long arm of chromosome 18. However, the region specifically concerned remains unknown. Using an array specific to chromosome 18, it has been possible to specify a candidate region for CAA. Indeed, a common 5-Mb region has systematically been found absent in 20 patients [70].

17.3.2 ChIP on Chip

Basic Idea

The name ChIP on chip gives a relatively explicit rendering of the principles involved here, viz., the immunoprecipitation of chromatin on a chip, as illustrated in Fig. 17.6. Such arrays are used to study the binding sites of proteins on DNA. This is the case for example with transcription factors, histones, DNA repair enzymes, and so on. Today, with these chips, all the binding sites of a given protein at a given instant of time can be characterised in a single experiment, and across the whole genome. The cells are lysed and the chromatin is isolated and fragmented by sonication to generate DNA fragments of around 1 kb. The chromatin fragments are then immunoprecipitated by a specific antibody against the relevant protein. The proteins are eliminated from the precipitate and the DNA is then amplified, incorporating labelled nucleotides in order to monitor the targets. This amplified and labelled DNA is hybridised to the arrays, on which all the relevant regions situated outside the open reading frames (ORF) have been previously deposited using a microdeposition robot. These regions may be intergenic regions, in yeast for example, or promoter regions as in higher eukaryotes. After hybridisation, the chips are read with a scanner traditionally used to read transcriptome microarrays. The problem then is to determine those regions that have been enriched by immunoprecipitation. Indeed, these are regions where the relevant protein binds preferentially.

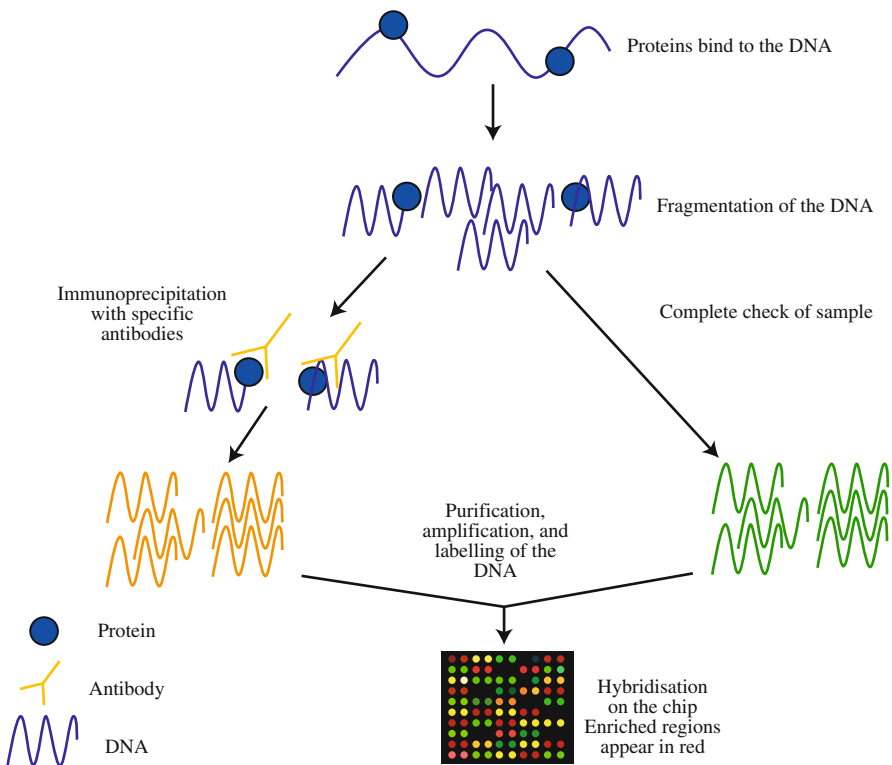


Fig. 17.6. Chromatin immunoprecipitation on a chip (ChIP on chip) technology. Probes are oligonucleotides or PCR products specific to intergenic regions, or specific regions outside genes. These are deposited on the chip by a robot equipped with an array of pins, under similar conditions to those used for the CGH microarray. Targets are PCR products resulting from chromatin immunoprecipitation with a specific antibody of the relevant protein. Using this technology, one can locate and study all binding sites of a given protein across the whole genome

Applications

Gene Regulation Networks

Transcription factors are proteins controlling differentiation, development, the cell cycle, and other important biological functions, by modulating the expression of the genes. To do this, they act on regulatory elements upstream of the genes, either by directly binding onto these elements, or indirectly, by interacting with other factors bound to these regulation elements. There are several ways of identifying the genes targeted by a transcription factor. One of these is to characterise potential regulation elements upstream of certain genes by biocomputer analysis of the sequences. However, owing to the degeneracy of the genetic code, these regulation elements can be difficult to identify. The

Table 17.2. Different ChIP on chip technologies. AB stands for antibody and CpG islands are regions of DNA rich in G and C

Type of array	Advantages	Disadvantages
ChIP on CpG islands	Chips easy to fabricate	Poorly defined targets, poor clone annotation, limited by AB performance, low resolution
ChIP on promoter regions	Chips easy to fabricate, probes perfectly characterised and annotated	Limited by AB performance, presence of a bias
ChIP on oligonucleotide chips	High resolution, less bias	Very expensive, high consumption of biological material

biocomputing approach is also limited in its ability to characterise more complex regulation elements. In addition, if the regulator operates by interacting with other proteins, the regulation elements will remain undetectable by biocomputing. The ChIP on chip method then provides a particularly useful, even indispensable way of identifying the genes targeted by a transcription factor.

The ChIP on chip has been used recently, e.g., with yeast, to study the binding sites of RNA polymerase III on the yeast genome [71], and also the binding sites of transcription factors [72]. In humans, however, the situation is much more delicate, because intergenic regions are very long and poorly defined. Several approaches can get around this problem. On an array, one can use PCR products corresponding to promoter proximal regions. This was the strategy chosen for our own research. Other groups have used chips on which CpG islands are deposited, considering that these are particularly present in gene promoter regions. With these tools, various projects have been able to characterise the regulation network of the cell cycle [73] or those for muscle formation [74].

Chromatin Modelling and Epigenetics

Other proteins can also be studied on the same type of array. This is possible for example with chromatin proteins, especially histones, or DNA-modifying enzymes like the methylases. Hence, the binding sites of proteins interacting specifically with 5mCpG sites (methylation of cytosine at 5') can be mapped to identify new targets for epigenetic inactivation in cancers [75]. Genome maps of the modification of histones in mammal cells are also beginning to appear. An example is the di- and trimethylation of the Lys4 of the protein H3 and the acetylation of H3 in human cells. Finally, the polycomb group of proteins (PcG) form heterochromatin complexes. Several binding sites of the

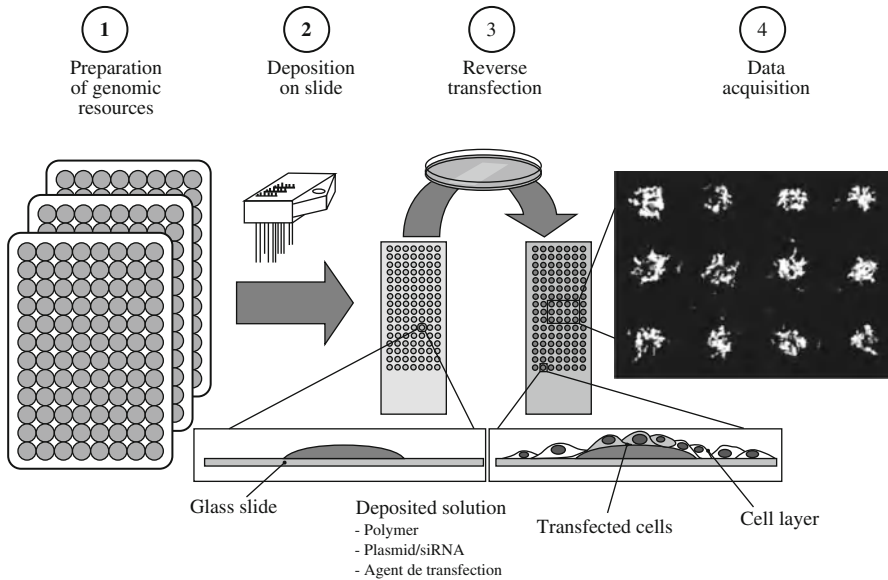


Fig. 17.7. Cell microarray. Probes are nucleic acids (expression vector or interfering RNA) deposited mechanically on the chip, as for the ChIP on chip. Deposits have diameters in the range 200–400 μm . Targets are cells deposited in a monolayer on the chip. The aim is to obtain thousands of cell islands transfected by these different nucleic acids and to study their phenotype

PcG have been identified in humans by combining the ChIP on chip results with transcriptome analysis [76].

17.3.3 When DNA Microarrays Become Cell Microarrays

Highly Parallel Transfection in Cell Microarrays

High-throughput analysis of the gain and loss of function of a gene in cell cultures, e.g., to determine the function of unknown genes, involves a stage of simultaneous reverse transfection of several thousand different nucleic acids on an array [77]. In this very flexible and cheap format, derived from DNA microarrays, plasmids mixed with a polymer and a transfection agent are deposited on a glass slide by a robot traditionally used to fabricate DNA microarrays. The microarray is then placed in a culture dish and coated with a monocellular layer of adhesion cells. The nucleic acids deposited on the slide will then diffuse locally and enter the cells in direct contact with the deposit, whence the name reverse transfection (see Fig. 17.7) [78].

Expression Microarrays

Using highly parallel transfection on a microarray, complete collections of human genes, cloned in an expression vector, can be transfected to study

the gain-of-function effect of each gene, via the synthesis of the corresponding protein in the islands of transfected cells. The relevant phenotypes (apoptosis, proliferation, etc.) resulting from this excess of protein in the cell islands can then be analysed. In this way, our group has obtained a collection of 16,000 complementary DNA (cDNA) (of which 70% are full length), cloned under the control of a strong promoter with enhanced synthesis of the gene products.

siARN Array

The discovery of interfering RNA (iRNA) allows systematic silencing in vitro of the expression of a gene in eukaryote cells, something that was previously rather difficult to do [79,80]. More recently, it has been shown that short interfering RNA (siARN) allows extremely efficient silencing in mammal cells [79]. Even though the use of iRNA differs from the knockout technology, insofar as complete silencing of gene expression is difficult, this feature is not necessarily a problem and may even constitute a major advantage when studying proteins whose total absence would be lethal for the cell. In addition, the gradual induction of silencing using iRNA can be used to monitor the progressive silencing of the relevant gene and in some cases to observe the initial effects. Cell microarrays can thus be used to transfect thousands of siRNA in parallel and analyse the relevant phenotypes resulting from the absence of the protein in the islands of modified cells. Today we have available a collection of 1,800 siRNA and we hope eventually to obtain a collection containing one siARN for each human gene. Alternatively, it is also possible to transfect plasmids allowing transcription in vivo of short hairpin RNA (shRNA). These vectors can silence the expression of a gene in the same way as an siARN.

Applications

Functional Exploration of Genomes

The perturbations generated in cells cultivated on this type of array can be monitored in several ways. Once the cell layer has been bound, the phenotype analysis resulting from the gain or loss of function of the genes can be carried out by immunofluorescence (IF) using specific antibodies for the relevant targets, in situ hybridisation to localise messenger RNA on the subcellular level, or autoradiography to monitor the biomolecules or an active constituent labelled by a radionuclide. But this analysis can also be carried out on living cells and in real time to monitor different cell functions, such as division, synthesis and translocation of proteins, intracellular calcium flows, ATP synthesis, etc.

Several aspects of cell life can also be elucidated on microarrays. Cell proliferation (primary cells or cell lines) is monitored by immunofluorescence directed against the proliferating cell nuclear antigen (PCNA), naturally expressed in proliferating cells, but also against a modified nucleotide added to the medium (5-bromo-2'-deoxyuracil) incorporated in cell DNA during the S

phase. Differentiation is also monitored by IF using other specific labels for the type of cell under investigation, e.g., desmin in myogenic cells. Apoptosis is monitored with a caspase (cysteinyI aspartate cleaving protease) inhibitor labelled by a fluorophore. While it cannot provide precise indications concerning the molecular mechanisms involved in the studied phenotypes, this analysis should nevertheless significantly speed up the functional annotation of the genome.

Other Applications

The applications of cell arrays are not limited to functional exploration. Different types of cell event can be studied using reporter genes. Cell microarrays can be adapted to measure the potential affinity of a ligand for the receptor proteins expressed at the surface of transfected cells. To do this, the idea is to deposit the expression vectors of the relevant protein collection on the array and monitor the binding of the ligand onto the islands of cells expressing them. The possibility of high-throughput screening with very small sample volumes makes this format particularly attractive to the pharmaceutical industry. On the same basis, one can also envisage characterising unknown viral receptors on human cell membranes.

17.3.4 Prospects and Conclusion

A new generation of high-resolution high-throughput DNA microarrays is currently emerging. In the future, it seems likely that arrays of long oligonucleotides (with 50–100 bases) will replace those with bacterial artificial chromosomes (BAC) or PCR products, as has already happened for transcriptome analysis. These developments will considerably increase the resolution and genome coverage, and at reasonable cost. In parallel, more efficient fluorescent probes should appear, together with chemical surfaces able to enhance fluorescence, and the sensitivity and signal-to-noise ratio should also improve. As a consequence, these arrays should be better reproducible. Finally, the number of spots per unit area should increase to reach millions of spots on a standard microarray. It thus seems likely that these future developments will allow a more or less simultaneous analysis of the transcriptome, chromosome rearrangements, and proteins binding to the DNA, on a single array and for an entire genome, with a resolution of 1–3 kb.

References

1. Gress, T.M., Hoheisel, J.D., Lennon, G.G., Zehetner, G., Lehrach, H.: Hybridization fingerprinting of high-density cDNA-library arrays with cDNA pools derived from whole tissues, *Mamm. Genome* **3**, 609–619 (1992)

2. Nguyen, C., Rocha, D., Granjeaud, S., Baldit, M., Bernard, K., Naquet, P., Jordan, B.R.: Differential gene expression in the murine thymus assayed by quantitative hybridization of arrayed cDNA clones, *Genomics* **29**, 207–216 (1995)
3. Schena, M., Shalon, D., Davis, R.W., Brown, P.O.: Quantitative monitoring of gene expression patterns with a complementary DNA microarray, *Science* **270**, 467–470 (1995)
4. Lockhart, D.J., Dong, H., Byrne, M.C., Follettie, M.T., Gallo, M.V., Chee, M.S., Mittmann, M., Wang, C., Kobayashi, M., Horton, H., et al.: Expression monitoring by hybridization to high-density oligonucleotide arrays, *Nat. Biotechnol.* **14**, 1675–1680 (1996)
5. Golub, T.R., Slonim, D.K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J.P., Coller, H., Loh, M.L., Downing, J.R., Caligiuri, M.A., et al.: Molecular classification of cancer: Class discovery and class prediction by gene expression monitoring, *Science* **286**, 531–537 (1999)
6. Bertucci, F., Houlgatte, R., Benziane, A., Granjeaud, S., Adelaide, J., Tagett, R., Lloriod, B., Jacquemier, J., Viens, P., Jordan, B., et al.: Gene expression profiling of primary breast carcinomas using arrays of candidate genes, *Hum. Mol. Genet.* **9**, 2981–2991 (2000)
7. Perou, C.M., Sorlie, T., Eisen, M.B., van de Rijn, M., Jeffrey, S.S., Rees, C.A., Pollack, J.R., Ross, D.T., Johnsen, H., Akslen, L.A. et al.: Molecular portraits of human breast tumours, *Nature* **406**, 747–752 (2000)
8. Bishop, J.O.: Molecular hybridization of ribonucleic acid with a large excess of deoxyribonucleic acid, *Biochem. J.* **126**, 171–185 (1972)
9. Gillespie, D., Spiegelman, S.: A quantitative assay for DNA–RNA hybrids with DNA immobilized on a membrane, *J. Mol. Biol.* **12**, 829–842 (1965)
10. Schena, M., Shalon, D., Heller, R., Chai, A., Brown, P.O., Davis, R.W.: Parallel human genome analysis: Microarray-based expression monitoring of 1000 genes, *Proc. Natl. Acad. Sci. USA* **93**, 10614–10619 (1996)
11. Takahashi, N., Hashida, H., Zhao, N., Misumi, Y., Sakaki, Y.: High-density cDNA filter analysis of the expression profiles of the genes preferentially expressed in human brain, *Gene* **164**, 219–227 (1995)
12. Granjeaud, S., Bertucci, F., Jordan, B.R.: Expression profiling: DNA arrays in many guises, *Bioessays* **21**, 781–790 (1999)
13. Lloriod, B., Victorero, G., Nguyen, C.: *cDNA Micro (and Macro) Arrays with Radioactive Detection*, Springer, Berlin Heidelberg New York (2000)
14. Duggan, D.J., Bittner, M., Chen, Y., Meltzer, P., Trent, J.M.: Expression profiling using cDNA microarrays, *Nat. Genet.* **21**, 10–14 (1999)
15. Cheung, V.G., Morley, M., Aguilar, F., Massimi, A., Kucherlapati, R., Childs, G.: Making and reading microarrays, *Nat. Genet.* **21**, 15–19 (1999)
16. Southern, E., Mir, K., Shchepinov, M.: Molecular interactions on microarrays, *Nat. Genet.* **21**, 5–9 (1999)
17. Adams, M.D., Kelley, J.M., Gocayne, J.D., Dubnick, M., Polymeropoulos, M.H., Xiao, H., Merril, C.R., Wu, A., Olde, B., Moreno, R.F., et al.: Complementary DNA sequencing: Expressed sequence tags and human genome project, *Science* **252**, 1651–1656 (1991)
18. Okubo, K., Hori, N., Matoba, R., Niiyama, T., Matsubara, K.: A novel system for large-scale sequencing of cDNA by PCR amplification, *DNA Seq.* **2**, 137–144 (1991)

19. Auffray, C., Behar, G., Bois, F., Bouchier, C., Da Silva, C., Devignes, M.D., Duprat, S., Houlgatte, R., Jumeau, M.N., Lamy, B., et al.: IMAGE: Molecular integration of the analysis of the human genome and its expression, *C.R. Acad. Sci. III* **318**, 263–272 (1995)
20. Lennon, G., Auffray, C., Polymeropoulos, M., Soares, M.B.: The IMAGE Consortium: An integrated molecular analysis of genomes and their expression, *Genomics* **33**, 151–152 (1996)
21. Schuler, G.D., Boguski, M.S., Stewart, E.A., Stein, L.D., Gyapay, G., Rice, K., White, R.E., Rodriguez-Tome, P., Aggarwal, A., Bajorek, E., et al.: A gene map of the human genome, *Science* **274**, 540–546 (1996)
22. Lockhart, D.J., Dong, H., Byrne, M.C., Follettie, M.T., Gallo, M.V., Chee, M.S., Mittmann, M., Wang, C., Kobayashi, M., Horton, H., et al.: Expression monitoring by hybridization to high-density oligonucleotide arrays [see comments], *Nat. Biotechnol.* **14**, 1675–1680 (1996)
23. Rouillard, J.M., Zuker, M., Gulari, E.: OligoArray 2.0: Design of oligonucleotide probes for DNA microarrays using a thermodynamic approach, *Nucleic Acids Res.* **31**, 3057–3062 (2003)
24. Bertucci, F., Bernard, K., Lioriod, B., Chang, Y.C., Granjeaud, S., Birnbaum, D., Nguyen, C., Peck, K., Jordan, B.R.: Sensitivity issues in DNA array-based expression measurements and performance of nylon microarrays for small samples, *Hum. Mol. Genet.* **8**, 1715–1722 (1999)
25. Schena, M., Shalon, D., Davis, R.W., Brown, P.O.: Quantitative monitoring of gene expression patterns with a complementary DNA microarray [see comments], *Science* **270**, 467–470 (1995)
26. Shalon, D., Smith, S.J., Brown, P.O.: A DNA microarray system for analyzing complex DNA samples using two-color fluorescent probe hybridization, *Genome Res.* **6**, 639–645 (1996)
27. Blanchard, A.P., Kaiser, R.J., Leroy, L.E.: Synthetic DNA arrays, *Biosensors and Bioelectronics* **11**, 687–690 (1996)
28. Hughes, T.R., Mao, M., Jones, A.R., Burchard, J., Marton, M.J., Shannon, K.W., Lefkowitz, S.M., Ziman, M., Schelter, J.M., Meyer, M.R., et al.: Expression profiling using microarrays fabricated by an ink-jet oligonucleotide synthesizer, *Nat. Biotechnol.* **19**, 342–347 (2001)
29. McGall, G., Labadie, J., Brock, P., Wallraff, G., Nguyen, T., Hinsberg, W.: Light-directed synthesis of high-density oligonucleotide arrays using semiconductor photoresists, *Proc. Natl. Acad. Sci. USA* **93**, 13555–13560 (1996)
30. Lipshutz, R.J., Fodor, S.P., Gingeras, T.R., Lockhart, D.J.: High density synthetic oligonucleotide arrays, *Nat. Genet.* **21**, 20–24 (1999)
31. Singh-Gasson, S., Green, R.D., Yue, Y., Nelson, C., Blattner, F., Sussman, M.R., Cerrina, F.: Maskless fabrication of light-directed oligonucleotide microarrays using a digital micromirror array, *Nat. Biotechnol.* **17**, 974–978 (1999)
32. Sampsel, J.: Digital micromirror device and its application to projection displays, *J. Vac. Sci. Technol. B* **12**, 324263246 (1994)
33. Egeland, R.D., Southern, E.M.: Electrochemically directed synthesis of oligonucleotides for DNA microarray fabrication, *Nucleic Acids Res.* **33**, e125 (2005)
34. www.affymetrix.com/index.affx
35. www.illumina.com
36. Gunderson, K.L., Kruglyak, S., Graige, M.S., Garcia, F., Kermani, B.G., Zhao, C., Che, D., Dickinson, T., Wickham, E., Bierle, J., et al.: Decoding randomly ordered DNA arrays, *Genome Res.* **14**, 870–877 (2004)

37. Kuhn, K., Baker, S.C., Chudin, E., Lieu, M.H., Oeser, S., Bennett, H., Rigault, P., Barker, D., McDaniel, T.K., Chee, M.S.: A novel, high-performance random array platform for quantitative gene expression profiling, *Genome Res.* **14**, 2347–2356 (2004)
38. Behr, M.A., Wilson, M.A., Gill, W.P., Salamon, H., Schoolnik, G.K., Rane, S., Small, P.M.: Comparative genomics of BCG vaccines by whole-genome DNA microarray, *Science* **284**, 1520–1523 (1999)
39. Salazar, N.M., Caetano-Anolles, G.: Nucleic acid scanning-by-hybridization of enterohemorrhagic *Escherichia coli* isolates using oligodeoxynucleotide arrays, *Nucleic Acids Res.* **24**, 5056–5057 (1996)
40. Drmanac, S., Kita, D., Labat, I., Hauser, B., Schmidt, C., Burczak, J.D., Drmanac, R.: Accurate sequencing by hybridization for DNA diagnostics and individual genomics, *Nat. Biotechnol.* **16**, 54–58 (1998)
41. Brenner, S., Johnson, M., Bridgham, J., Golda, G., Lloyd, D.H., Johnson, D., Luo, S., McCurdy, S., Foy, M., Ewan, M., et al.: Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays, *Nat. Biotechnol.* **18**, 630–634 (2000)
42. Hacia, J.G.: Resequencing and mutational analysis using oligonucleotide microarrays, *Nat. Genet.* **21**, 42–47 (1999)
43. Troesch, A., Nguyen, H., Miyada, C.G., Desvarenne, S., Gingeras, T.R., Kaplan, P.M., Cros, P., Mabilat, C.: Mycobacterium species identification and rifampin resistance testing with high-density DNA probe arrays, *J. Clin. Microbiol.* **37**, 49–55 (1999)
44. Hacia, J.G., Brody, L.C., Chee, M.S., Fodor, S.P., Collins, F.S.: Detection of heterozygous mutations in BRCA1 using high density oligonucleotide arrays and two-colour fluorescence analysis, *Nat. Genet.* **14**, 441–447 (1996)
45. Kozal, M.J., Shah, N., Shen, N., Yang, R., Fucini, R., Merigan, T.C., Richman, D.D., Morris, D., Hubbell, E., Chee, M., et al.: Extensive polymorphisms observed in HIV-1 clade B protease gene using high-density oligonucleotide arrays, *Nat. Med.* **2**, 753–759 (1996)
46. Wang, D.G., Fan, J.B., Siao, C.J., Berno, A., Young, P., Sapolsky, R., Ghandour, G., Perkins, N., Winchester, E., Spencer, J., et al.: Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome, *Science* **280**, 1077–1082 (1998)
47. Chesler, E.J., Lu, L., Shou, S., Qu, Y., Gu, J., Wang, J., Hsu, H.C., Mountz, J.D., Baldwin, N.E., Langston, M.A., et al.: Complex trait analysis of gene expression uncovers polygenic and pleiotropic networks that modulate nervous system function, *Nat. Genet.* **37**, 233–242 (2005)
48. Hubner, N., Wallace, C.A., Zimdahl, H., Petretto, E., Schulz, H., Maciver, F., Mueller, M., Hummel, O., Monti, J., Zidek, V., et al.: Integrated transcriptional profiling and linkage analysis for identification of genes underlying disease, *Nat. Genet.* **37**, 243–253 (2005)
49. Loring, J.F.: Evolution of microarray analysis, *Neurobiol. Aging* (2005)
50. Pietu, G., Alibert, O., Guichard, V., Lamy, B., Bois, F., Leroy, E., Mariage-Sampson, R., Houlgatte, R., Soularue, P., Auffray, C.: Novel gene transcripts preferentially expressed in human muscles revealed by quantitative hybridization of a high density cDNA array, *Genome Res.* **6**, 492–503 (1996)
51. DeRisi, J., Penland, L., Brown, P.O., Bittner, M.L., Meltzer, P.S., Ray, M., Chen, Y., Su, Y.A., Trent, J.M.: Use of a cDNA microarray to analyse gene

- expression patterns in human cancer [see comments], *Nat. Genet.* **14**, 457–460 (1996)
52. Wurbel, M.A., Philippe, J.M., Nguyen, C., Victorero, G., Freeman, T., Wooding, P., Miazek, A., Mattei, M.G., Malissen, M., Jordan, B.R., et al.: The chemokine TECK is expressed by thymic and intestinal epithelial cells and attracts double- and single-positive thymocytes expressing the TECK receptor CCR9, *Eur. J. Immunol.* **30**, 262–271 (2000)
 53. Alizadeh, A.A., Eisen, M.B., Davis, R.E., Ma, C., Lossos, I.S., Rosenwald, A., Boldrick, J.C., Sabet, H., Tran, T., Yu, X., et al.: Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling, *Nature* **403**, 503–511 (2000)
 54. Eisen, M.B., Spellman, P.T., Brown, P.O., Botstein, D.: Cluster analysis and display of genome-wide expression patterns, *Proc. Natl. Acad. Sci. USA* **95**, 14863–14868 (1998)
 55. Bertucci, F., Salas, S., Eyseries, S., Nasser, V., Finetti, P., Ginestier, C., Charafe-Jauffret, E., Lloriod, B., Bachelart, L., Montfort, J., et al.: Gene expression profiling of colon cancer by DNA microarrays and correlation with histoclinical parameters, *Oncogene* **23**, 1377–1391 (2004)
 56. Steenman, M., Chen, Y.W., Le Cunff, M., Lamirault, G., Varro, A., Hoffman, E., Leger, J.J.: Transcriptomal analysis of failing and nonfailing human hearts, *Physiol. Genomics* **12**, 97–112 (2003)
 57. Bertucci, F., Nasser, V., Granjeaud, S., Eisinger, F., Adelaide, J., Tagett, R., Lloriod, B., Giaconia, A., Benziene, A., Devillard, E., et al.: Gene expression profiles of poor-prognosis primary breast cancer correlate with survival, *Hum. Mol. Genet.* **11**, 863–872 (2002)
 58. Magrangeas, F., Nasser, V., Avet-Loiseau, H., Lloriod, B., Decaux, O., Granjeaud, S., Bertucci, F., Birnbaum, D., Nguyen, C., Harousseau, J.L., et al.: Gene expression profiling of multiple myeloma reveals molecular portraits in relation to the pathogenesis of the disease, *Blood* **101**, 4998–5006 (2003)
 59. Nakamura, T., Furukawa, Y., Nakagawa, H., Tsunoda, T., Ohgashi, H., Murata, K., Ishikawa, O., Ohgaki, K., Kashimura, N., Miyamoto, M., et al.: Genome-wide cDNA microarray analysis of gene expression profiles in pancreatic cancers using populations of tumor cells and normal ductal epithelial cells selected for purity by laser microdissection, *Oncogene* **23**, 2385–2400 (2004)
 60. Zhu, G., Reynolds, L., Crnogorac-Jurcevic, T., Gillett, C.E., Dublin, E.A., Marshall, J.F., Barnes, D., D'Arrigo, C., Van Trappen, P.O., Lemoine, N.R., et al.: Combination of microdissection and microarray analysis to identify gene expression changes between differentially located tumour cells in breast cancer, *Oncogene* **22**, 3742–3748 (2003)
 61. Ross, D.T., Scherf, U., Eisen, M.B., Perou, C.M., Rees, C., Spellman, P., Iyer, V., Jeffrey, S.S., Van de Rijn, M., Waltham, M., et al.: Systematic variation in gene expression patterns in human cancer cell lines, *Nat. Genet.* **24**, 227–235 (2000)
 62. Puthier, D., Joly, F., Irla, M., Saade, M., Victorero, G., Lloriod, B., Nguyen, C.: A general survey of thymocyte differentiation by transcriptional analysis of knockout mouse models, *J. Immunol.* **173**, 6109–6118 (2004)
 63. Tavazoie, S., Hughes, J.D., Campbell, M.J., Cho, R.J., Church, G.M.: Systematic determination of genetic network architecture, *Nat. Genet.* **22**, 281–285 (1999)

64. Brazma, A., Vilo, J.: Gene expression data analysis, *FEBS Lett.* **480**, 17–24 (2000)
65. Kallioniemi, A., et al.: Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors, *Science* **258**, 818–821 (1992)
66. Solinas-Toldo, S., et al.: Matrix-based comparative genomic hybridization: Biochips to screen for genomic imbalances, *Genes Chromosomes Cancer* **20**, 399–407 (1997)
67. Pollack, J.R., et al.: Genome-wide analysis of DNA copy number changes using cDNA microarrays, *Nat. Genet.* **23**, 41–46 (1999)
68. Visser, L.E., et al.: Array-based comparative genomics hybridization for the genome-wide detection of submicroscopic chromosomal abnormalities, *Am. J. Hum. Genet.* **73**, 1261–1270 (2003)
69. Guillaud-Bataille, M., et al.: Detecting single DNA copy number variations in complex genomes using one nanogram of starting DNA and BAC-array CGH, *Nuc. Acid. Res.* **32**, e112 (2004)
70. Veltman, J.A., et al.: Definition of a critical region on chromosome 18 for congenital aural atresia by array-CGH, *Am. J. Hum. Genet.* **72**, 1578–1584 (2003)
71. Harismendy, O., et al.: Genome-wide location of yeast RNA polymerase III transcription machinery, *EMBO J.* **22**, 4738–4747 (2003)
72. Ren, B., et al.: Genome-wide location and function of DNA binding proteins, *Science* **290**, 2306–2309 (2000)
73. Ren, B., et al.: E2F integrates cell cycle progression with DNA repair, replication and G(2)/M check-points, *Genes Dev.* **16**, 245–256 (2002)
74. Blais, A., et al.: An initial blueprint for myogenic differentiation, *Genes Dev.* **19**, 553–569 (2005)
75. Ballestar, E., et al.: Methyl-CpG binding proteins identify novel sites of epigenetic inactivation in human cancer, *EMBO J.* **22**, 6335–6345 (2003)
76. Kirmizis A., et al.: Silencing of human polycomb target genes is associated with methylation of histone H3 Lys 27, *Genes Dev.* **18**, 1592–1605 (2004)
77. Ziauddin J., et al.: Microarrays of cells expressing defined cDNAs, *Nature* **411**, 107–110 (2001)
78. Baghdoyan, S., et al.: Quantitative analysis of highly parallel transfection in cell microarrays, *Nucleic Acids Res.* **32**, e77 (2004)
79. Elbashir, S.M., Harborth, J., Lendeckel, W., Yalcin, A., Weber, K., Tuschl, T.: Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells, *Nature* **411**, 494–498 (2001)
80. Fire, A., Xu, S., Montgomery, M.K., Kostas, S.A., Driver, S.E., Mello, C.C.: Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*, *Nature* **391**, 806–811 (1998)

Protein Microarrays

S. Ricard-Blum

Proteins are key actors in the life of the cell, involved in many physiological and pathological processes. Since variations in the expression of messenger RNA are not systematically correlated with variations in the protein levels, the latter better reflect the way a cell functions. Protein microarrays thus supply complementary information to DNA chips. They are used in particular to analyse protein expression profiles, to detect proteins within complex biological media, and to study protein–protein interactions, which give information about the functions of those proteins [3–9]. They have the same advantages as DNA microarrays for high-throughput analysis, miniaturisation, and the possibility of automation.

Section 18.1 gives a brief overview of proteins. Following this, Sect. 18.2 describes how protein microarrays can be made on flat supports, explaining how proteins can be produced and immobilised on a solid support, and discussing the different kinds of substrate and detection method. Section 18.3 discusses the particular format of protein microarrays in suspension. The diversity of protein microarrays and their applications are then reported in Sect. 18.4, with applications to therapeutics (protein–drug interactions) and diagnostics. The prospects for future developments of protein microarrays are then outlined in the conclusion.

The bibliography provides an extensive list of reviews and detailed references for those readers who wish to go further in this area. Indeed, the aim of the present chapter is not to give an exhaustive or detailed analysis of the state of the art, but rather to provide the reader with the basic elements needed to understand how proteins are designed and used.

18.1 Overview of Proteins

Proteins are macromolecules essential to the functioning of living organisms. They comprise a succession of amino acids bound to one another in a well-defined order. There are about twenty different amino acids. Within a given

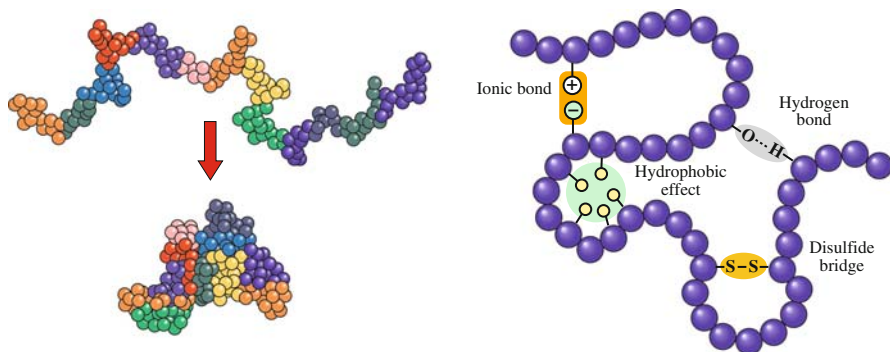


Fig. 18.1. (a) A protein folds into its 3D structure. (b) Bonds involved in the 3D structure [1]

protein, two successive amino acids are bound together covalently by what is called a peptide bond. The molecular masses of proteins vary between a few thousand and a few hundred thousand daltons (Da). The dalton is a unit of molecular mass, named after John Dalton who developed an atomic theory of matter. One dalton is exactly $1/12$ of the mass of a carbon-12 atom.

Proteins fold up to adopt precise 3D conformations (see Fig. 18.1a) and have a variety of shapes. Some proteins like albumin are globular, while others like collagen are rod-shaped and can assemble into fibres. The biological function of a protein is closely related to its shape, which is maintained by low energy bonds (hydrogen, hydrophobic, or ionic bonds) formed between the amino acids of the folded protein (see Fig. 18.1b). The 3D structure of some proteins is also maintained by covalent bonds, the disulfide bridges. When these bonds are broken by chemical reagents, organic solvents, or changes in temperature and pH, the protein can unfold. This change of shape is generally accompanied by a loss of function of the protein, which is said to be denatured.

Proteins have a wide range of functions. They can play a structural role, especially in the assembly of cellular and extracellular structures, and are involved in muscle contraction. They are also involved in the transport of chemical substances in the blood, e.g., oxygen transported by hemoglobin, or through the membranes of cells in which they form pores, e.g., aquaporins. Hormones such as insulin or growth hormone are proteins. Enzymes, which catalyse chemical reactions in living organisms, are also proteins. Other proteins, antibodies, participate in the defence of the organism against pathogens (see Fig. 18.2).

Every protein is synthesised according to a blueprint in the corresponding gene. This information is translated by the cell machinery into a linear chain of amino acids to build up the protein. A given protein can be produced *in vitro* by inserting its gene into bacteria, yeast, insect, or mammal cells. A protein obtained in this way is called a recombinant protein.

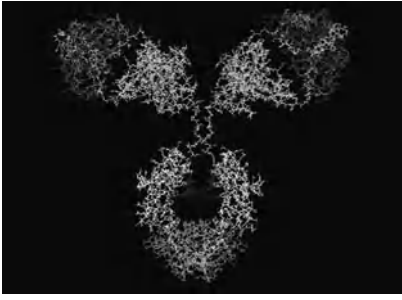


Fig. 18.2. An antibody molecule [2]

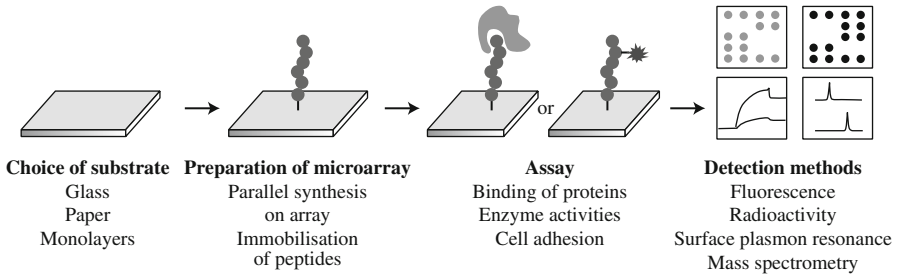


Fig. 18.3. Protein or peptide microarray. The proteins or peptides are spotted on the support. The array is then set in contact with the soluble samples to be assayed. The partners present in the sample bind specifically to the immobilised proteins. Any binding that occurs is then detected by one of the following methods: fluorescence, radioactivity, surface plasmon resonance, and mass spectrometry (shown in the *insert*), or atomic force microscopy [16]

18.2 Fabricating a Protein Array on a Flat Support

A protein microarray is usually produced on a solid flat substrate, on which the proteins are immobilised. The proteins will then bind selectively to molecules present in the sample to be studied. The reagents and samples are delivered to the surface of the array by a microfluidic system. There are several different ways of detecting the binding of molecules to these immobilised proteins. Some involve labelling the target molecules, while others do not. The signal obtained is then analysed using computational tools, but data processing and management will not be discussed here. Figure 18.3 gives a schematic view of an experiment carried out with a protein microarray.

18.2.1 Preparation of Purified Proteins

The proteins used to build a protein microarray must be available in soluble native form, i.e., suitably folded. Given the extreme diversity of protein structures, it is difficult to produce and purify them on a large scale by a single method, and several approaches are possible [9]. One of these is to extract

and purify proteins from tissue, culture medium, or biological liquids, but it is a long and tedious process that requires careful development of appropriate purification conditions for each protein and cannot be used on a large scale. The production of recombinant proteins is currently the most favoured method for building protein microarrays. Another approach is to transfect cells immobilised on the array with the relevant genes and then produce the proteins *in situ*, directly on the microarray.

Recombinant Proteins

The most widely used method is the production of recombinant proteins with a tag located at the N- or C-terminal of the protein. Addition of a tag allows the use of the same purification procedure for all proteins. The protein is purified by affinity chromatography using the tag and captured on the array via the tag. The drawback with this approach is that post-translational modifications of recombinant proteins are not generally the same as those of natural proteins. The tags used are glutathion S-transferase (GST, 26 kDa) and a polyhistidine sequence of 4–6 consecutive residues. Histidine can bind to a nitrilotriacetic acid–nickel (NTA–Ni) complex, in which the nickel is bound by chelation to the nitrilotriacetic acid. The nickel has six free sites in its coordination sphere, and four of these six sites are occupied when it is chelated by an NTA group. The two vacant sites can bind two histidine residues through their nitrogen atom. This property is used to bind the proteins with a polyhistidine tag onto a substrate containing the NTA–Ni complex. It has been used on a large scale to capture 6,566 protein samples representing 5,800 yeast proteins on a glass slide coated with nickel [10].

Some protein microarrays are made, not with full-length proteins, but with protein domains [9, 11, 12]. This allows faster identification of binding site(s) in the protein [12].

In Situ Protein Production

Proteins are produced *in situ* in recombinant form with a tag. The major advantage lies in the fact that the proteins are bound onto the array by the specific capture system, so that they do not need to be purified. This is useful in the case of membrane proteins, which are particularly insoluble and difficult to produce and purify.

There are two ways of producing proteins directly on the array. The first one does not require cells. It is carried out in the presence of DNA coding for the relevant protein and rabbit reticulocyte lysate to achieve the transcription and synthesis of the proteins *in vitro*. The cell-free Discern Array technology amplifies genes by polymerase chain reaction (PCR) or reverse transcriptase PCR (RT-PCR), then synthesises the proteins *in vitro*, in the presence of rabbit reticulocyte lysate, with a polyhistidine tag used to capture them on microplates or magnetic beads coated with the NTA–Ni complex. The proteins

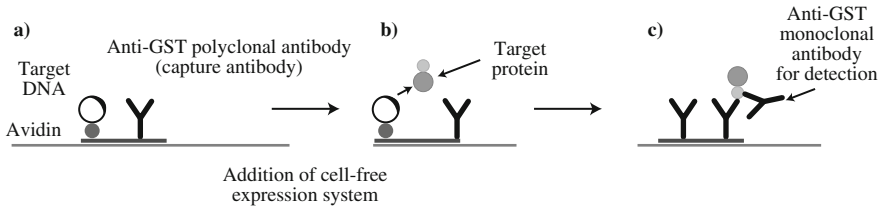


Fig. 18.4. Transfection and production of recombinant proteins in situ (nucleic acid programmable protein array or NAPPA). (a) Biotinylated DNA is immobilised on the surface by avidin, which is bound covalently onto the surface, together with the anti-GST polyclonal antibody. (b) in situ expression of recombinant proteins carrying the GST tag in the presence of rabbit reticulocyte lysate. (c) Expressed proteins are captured by the anti-GST antibody immobilised on the surface. The immobilisation of recombinant proteins can be checked by adding a monoclonal antibody directed against GST [14]

are immobilised in situ after synthesis, without a purification stage [13]. A similar acellular approach has been developed to produce and capture proteins containing a GST tag on a glass slide. These arrays are called nucleic acid programmable protein arrays or NAPPA (see Fig. 18.4) [14].

The second approach requires the presence of cells which are transfected with the relevant genes on the array, where they produce the proteins with an appropriate tag. The proteins are then captured on the substrate of the microarray. Ziauddin and Sabatini [15] deposited plasmids on a glass slide, on which they had incubated cells. One of the advantages of these transfected cell microarrays is the possibility of screening the effects of expression of a given gene on living cells.

Peptide Production

In many cases, the biological activity of a protein is found in a short amino acid sequence. This has led to the development of peptide microarrays, prepared either by in situ synthesis of peptides on a membrane, or by immobilising previously synthesised peptides on the array [16, 17]. The two major techniques used for parallel synthesis of many peptides on a solid substrate are photolithography and the SPOT method. Photolithography uses photolabile protector groups which allow synthesis only on regions of the substrate that are illuminated. An array of 1,024 peptides has been designed in this way [16]. In the SPOT method the peptides are synthesised by sequential deposition of small volume (of microlitre order) of activated amino acids on a porous membrane [18].

Alternatively, the relevant peptides are synthesised by chemical reactions or produced in a recombinant form, depending on their size, before immobilisation on the substrate [16]. It is possible to add a cysteine residue at the peptide end to use the same immobilisation chemistry on the substrate for all

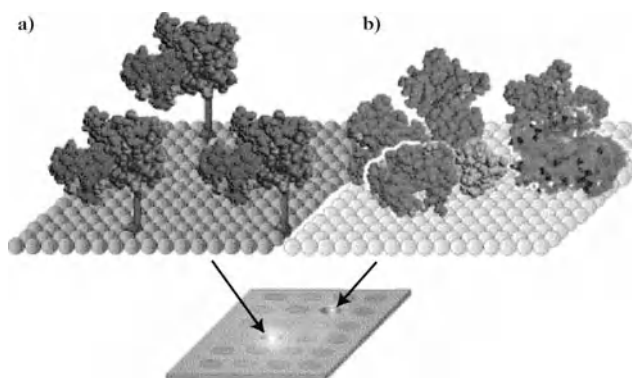


Fig. 18.5. Proteins immobilised on an array. **(a)** Idealised immobilisation of proteins. Proteins have uniform orientation and are properly folded, with optimal spacing to allow protein–protein interactions to occur. **(b)** Non-ideal situation in which the proteins are immobilised in various orientations and with various degrees of denaturation, with proteins adsorbed non-specifically onto the surface of the array [22]

the peptides. Biotinylated peptides can be captured on a surface previously coated with avidin or streptavidin. Peptide microarrays are used in immunology to map epitopes, to analyse kinase substrates [19], and to design protease inhibitors [20].

18.2.2 Substrates for Protein Microarrays

The substrate must give good quality spots, be easy to manipulate, and give a weak non-specific binding with the proteins. Indeed, this is of crucial importance for the specificity of the binding events.

Flat Supports

Nitrocellulose membranes are used as a substrate to immobilise proteins. Membranes called DiscoverLight Protein Array Kits allowing the deposition of 384 spots are available from Pierce. The most commonly used support for protein microarrays are glass slides, while silicon is currently less widely used [21]. These two rigid supports are the best suited to automation. The Ciphergen system uses aluminium functionalised to capture cell, tissue, or biological fluid extracts.

Non-Flat Supports

Non-planar supports such as microspheres and nanoparticles, which form suspended arrays, will be discussed in Sect. 18.3.

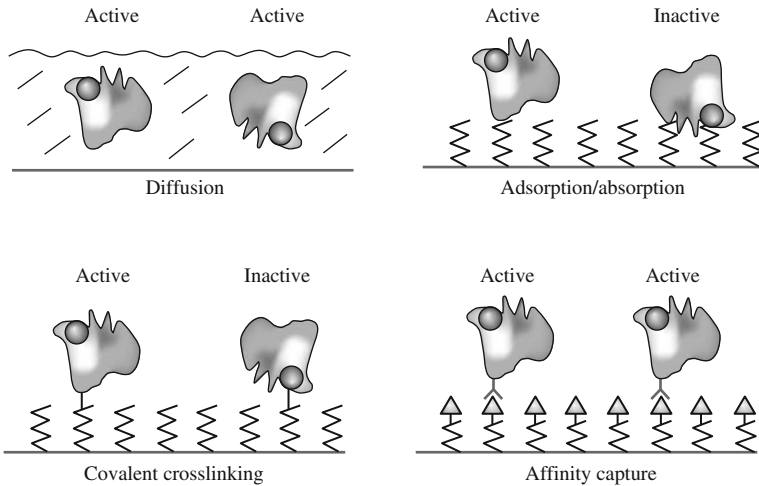


Fig. 18.6. Different methods for immobilising proteins on a solid substrate: diffusion in a gel, adsorption/absorption, covalent crosslinking, and affinity capture (*circles* indicate functional sites of the protein) [5]

18.2.3 Immobilising Proteins on the Array

When a protein binds to the solid substrate, this should alter neither its functionality nor its stability. It is essential to preserve the native 3D structure of the proteins. The ideal situation is to obtain well folded and well spaced proteins with uniform orientation across the surface of the array (see Fig. 18.5a) [22]. In practice, the proteins generally have a range of orientations and degrees of denaturation and some proteins may adsorb non-specifically onto the substrate (see Fig. 18.5b).

Proteins can be bound covalently onto the support using chemical reagents which can modify their structure, or they can be bound non-covalently by methods which respect their integrity, providing a suitable orientation at the surface, as well as a uniform distribution. The various types of immobilisation are shown in Fig. 18.6.

Covalent Immobilisation

Proteins can be immobilised on a solid substrate by covalent bonds which provide a strong binding and high protein density. This can be done on glass slides activated by epoxy, primary amino, or aldehyde groups. More than 10,000 proteins have been deposited with a density of 1,600 spots/cm² on glass slides activated by aldehyde groups [23]. This method requires chemical modification of the proteins, which can alter their native structure and functionality, rendering them partially or totally inaccessible to their partners.

Non-Covalent Immobilisation

Non-covalent immobilisation preserves the integrity, structure, and activity of the proteins. It can be achieved in several different ways. Immobilisation by adsorption/absorption on a membrane (nitrocellulose or polyvinylidene difluoride) does not modify the protein, and these membranes have a good binding capacity. The proteins can also be inserted without chemical modification in a 3D hydrogel which also has a good binding capacity. These two methods have the disadvantage that the proteins are not oriented (see Fig. 18.6), and this reduces their ability to interact with their partners.

Capture by Affinity

There are several capture reagents [24] for immobilising proteins in an oriented manner on the substrate (see Fig. 18.6).

Monoclonal antibodies, directed either against the protein to be captured, or against the tag of recombinant proteins, are used to capture proteins on a solid substrate. This requires immobilisation of the antibody on the support, followed by specific, non-covalent capture of the relevant proteins. In this way, 5,800 yeast proteins expressing the polyhistidine tag have been captured on a glass slide coated with nickel [10].

Proteins can also be captured by aptamers or photoaptamers, which are highly specific capture reagents (see Fig. 18.7) [25]. They are able to recognise a wide variety of targets (proteins, small molecules, nucleic acids) with a high affinity and high specificity. Aptamers are designed like antibodies, with a constant, invariable part which exposes a peptide loop of variable sequence. They are selected for their ability to specifically recognise a target protein.

Streptavidin immobilised on a solid substrate is used to capture biotinylated proteins. This system is based on the strong affinity of streptavidin for biotin ($K_D \sim 10^{-13}$ M). A method of immobilisation on silicon has been developed along these lines [21]. The silicon is coated with an inorganic material, which confers a negative charge on the chip surface, and then with an organic polymer, poly(L-lysine)-grafted-poly(ethylene glycol), in which the polyethylene glycol part has been biotinylated. The presence of biotin allows the capture of proteins conjugated with streptavidin. The spots obtained have diameter 50 μm .

18.2.4 Spotting Proteins

Spotters are used to make protein microarrays. The proteins are deposited on the surface by extremely fine pins which are in contact with the surface and deliver very small and precise volumes of protein solutions. Another approach is to spray the proteins in the form of very fine droplets onto the surface of the array. This technique, which does not involve contact between the protein delivery system and the chip surface, is sometimes referred to as the inkjet

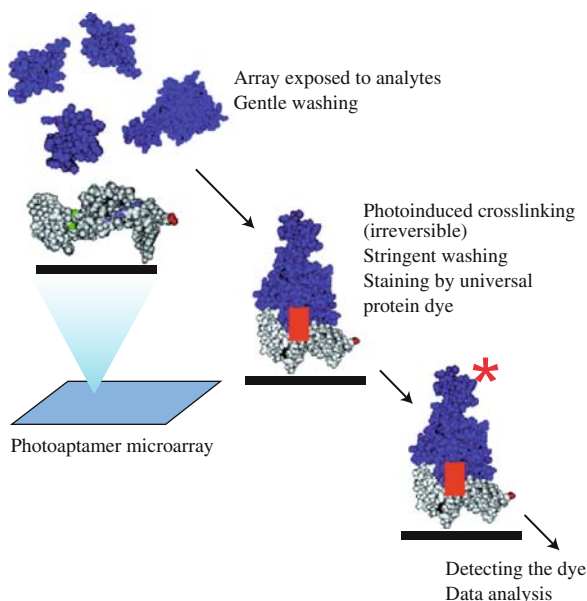


Fig. 18.7. Photoaptamer array assay. The aptamer array is incubated with the biological sample, e.g., a serum, then rinsed to remove non-specifically bound molecules. The array is then subjected to a photoinduced crosslinking process to covalently bind specifically bound analytes to the aptamers [25]

method, by analogy with the well known system used in desktop printers. The size of the spots depends on the spotting method, but also the solubility of the protein. It is generally in the range 150–200 μm , and volumes of proteins delivered are less than the nanolitre.

Packard Bioscience has developed an original non-contact technology called tipnology, exploiting the piezoelectric effect. It is able to deliver picodroplets of sample (about 333 pL) on a substrate. This technology is remarkable for its reproducibility and flexibility, but also for the uniformity of the spots [26].

Proteins can be deposited on a microarray surface by electrospray [27,28]. In this ionisation technique the proteins in solution are subjected to a strong electric field inducing the formation of charged microdroplets. One advantage of this technique is that the protein deposits can be considered as dry, given that the microdroplets dry in one millisecond under optimal conditions [28]. This is important because the activity and solubility of proteins can be decreased if they are stored in a wet environment. Electrospray deposition thus preserves the structure and function of proteins during fabrication of the array. Volumes of one microlitre are ionised by electrospray and spots of diameter 30–40 μm have been obtained with a spacing of 125 μm [28].

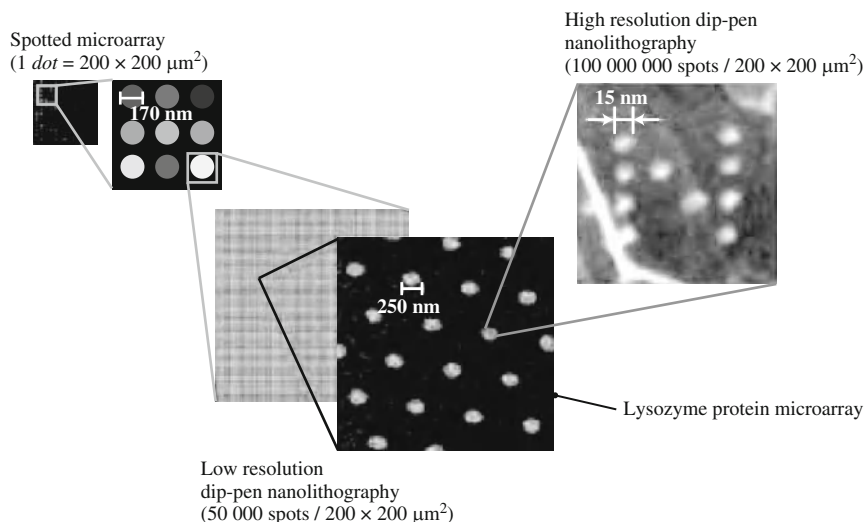


Fig. 18.8. Typical resolution on protein microarrays obtained by dip-pen nanolithography [30]. Copyright Wiley-VCH Verlag. Reproduced with kind permission from the publisher

Lithography is also used to spot proteins on the array surface. Dip-pen nanolithography is a new scanning probe-based direct write tool for generating surface-patterned functionality on the sub-100 nm length scale [29, 30]. These new, high performance etch and deposition techniques are decisive for printing structures of very small dimensions and miniaturising deposits on protein microarrays (see Fig. 18.8).

Dip-pen nanolithography has been used to fabricate protein microarrays [30]. This is a sequential writing technique in which a particle beam is narrowly focused on a point in space, then displaced pixel by pixel to produce the desired pattern. An atomic force microscopy tip is used to deliver chemical reagents and proteins directly to nanoscopic regions of the substrate [30, 31]. Dip-pen nanolithography with a microscope tip has been used to create spots of diameter 200–350 nm, on which G immunoglobulins and lysozyme have been adsorbed [32]. Collagen deposits of width 30–50 nm have been made with this technique on a gold-coated mica substrate (see Fig. 18.9) [33].

The immobilisation of histidine-tagged proteins by dip-pen nanolithography requires some modifications. Indeed, the roughness of glass slides coated with NTA–Ni complex does not lend itself well to accurate nanoscale deposits and the use of mica as a smoother substrate is unsatisfactory because the pretreatments lead to nickel removal from the surface. Proteins are therefore deposited on ionised regions of the nickel coating by applying an electrical potential to the tip of the scanning tunnelling microscope [34].

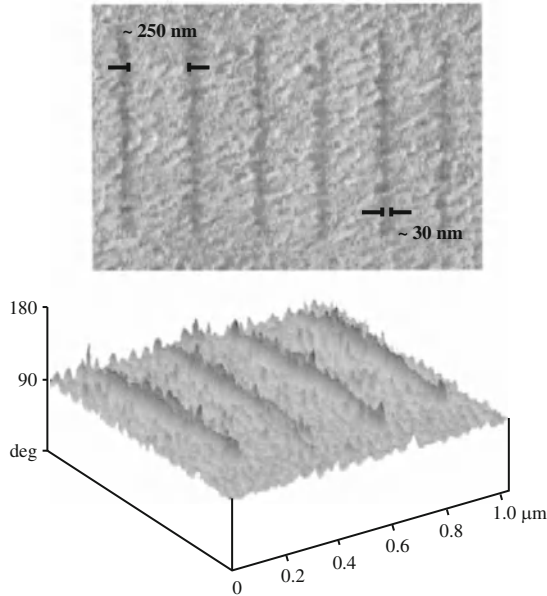


Fig. 18.9. Phase image of patterned collagens at 30 ± 4.6 nm line widths made by dip-pen lithography using a tip [33]. Copyright (2001) National Academy of Sciences, USA

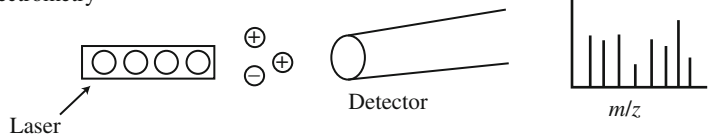
18.2.5 Detection Systems

There are many systems for detecting the binding of partners to proteins spotted on microarrays (Fig. 18.10) [35]. Several methods, viz., colorimetry, radioactivity, and fluorescence, use molecules labelled by a reporter group. This group can be a chromogen, a fluorophore, or a radioactive isotope. Chemiluminescence has also been applied to detect immunoglobulin E in human serum using an allergen microarray [36]. Radioactivity has been used for a ‘universal’ microarray to detect protein–protein, protein–DNA, and protein–RNA interactions [37]. Proteins immobilised on a nitrocellulose membrane react sequentially with a protein probe labelled by $[\gamma\text{-}^{32}\text{P}]\text{ATP}$, a single-strand DNA probe labelled by ^{32}P , and a messenger RNA probe labelled by ^{32}P . Blocking and washing steps were optimised for each type of probe and the signal was visualised by autoradiography [37]. Kinase-type enzyme activities in the yeast proteome have also been analysed with ATP labelled by ^{32}P [38].

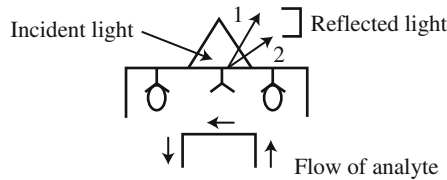
Detection can be direct, indirect, or use a sandwich-like array (Figs. 18.10c, d and e). Direct detection is achieved by tagged antibodies which recognise the proteins immobilised on the chip. The sensitivity of this detection can be enhanced by an amplification stage with the avidin–biotin system (see Fig. 18.10c). In a variant, a mixture of proteins was immobilised on each spot of the microarray, and the mixture was then probed with a single tagged antibody. This is called a reverse phase protein microarray [39]. Indirect detection

Without labelling

a) Mass spectrometry

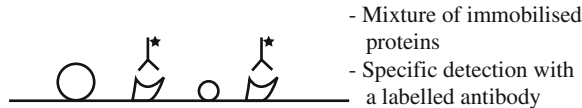


b) Surface plasmon resonance (SPR)



Labelled molecules

c) Direct



d) Indirect



e) Sandwich

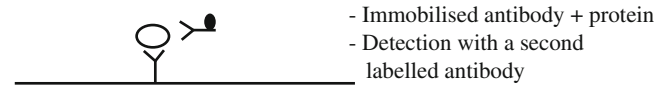


Fig. 18.10. Detection methods used for protein microarrays. (a) and (b) without labelling. (c), (d), and (e) with labelling [35]

uses an antibody that specifically captures tagged molecules in a complex medium (see Fig. 18.10d). The sandwich method uses two antibodies, one of which is bound to the substrate and captures the relevant molecule, while the other detects the resulting antigen–antibody complex (see Fig. 18.10e).

Other methods do not require labelling of the molecules, thereby avoiding the risk of altering their structure or biological activity. These methods are mass spectrometry, surface plasmon resonance (Figs. 18.10a and b), and atomic force microscopy.

Detection by Fluorescence

This is the most widely used method because it is extremely sensitive and provides very high resolution. Fluorescent detection molecules are required. These bind onto the molecules that have reacted with the proteins immobilised on the array surface. The fluorescence is measured by specific equipment which generally uses a laser to excite fluorophores bound to the array. There are two possible detection modes. One can either take a snapshot of the array with a CCD camera (array imager), or read the array using a photomultiplier tube (array scanner). A microarray on which 5,800 yeast proteins were immobilised has been analysed by fluorescence [10].

A technique known as rolling-circle amplification (RCA) can be used to amplify the fluorescence signal [40–42]. An oligonucleotide is bound to an anti-biotin antibody which binds in turn to the biotinylated antibody used for detection (see Fig. 18.11). DNA polymerase extends the oligonucleotide in the presence of nucleotides and circular DNA as template. The oligonucleotide, still attached to the anti-biotin antibody, is then hybridised with short fluorescent oligonucleotides which carry a large number of fluorophores, thereby increasing the detection sensitivity. A detection limit of 0.5 pg/mL is achieved with the best pairs of antibodies in a sandwich format [41].

Another method for increasing the sensitivity of fluorescence detection is based on the planar waveguide technique devised by Zeptosens [43, 44]. A 150-nm thick film of Ta₂O₅, a dielectric material widely used to make optical multilayers, is deposited on a transparent glass substrate [44]. A strong evanescent field, with limited penetration depth (200 nm) in the adjacent medium, is produced at the surface, along the light path in the waveguide (see Fig. 18.12). All the molecules located in the evanescent field are excited simultaneously and only the fluorophores confined to the surface of the array are excited for fluorescence emission. The resulting surface-restricted detection allows in situ measurements of proteins bound to the array surface without washing steps. The detection limit is 2 pM for an antibody tagged by a fluorophore and deposited on the spot. This concentration corresponds to 0.8 zeptomole (0.8×10^{-21} mole) of the antibody [44].

Fluorescent quantum dots conjugated with antibodies can be used as fluorescent reporter systems. These inorganic nanoparticles are much brighter than organic fluorophores. In addition, they have a much longer light emission time, which significantly limits photobleaching effects, allow increased sensitivity, and are well suited to multicolour labelling experiments due to their spectral properties. Streptavidin and biotin, as well as antibodies conjugated to quantum dots, are commercially available. This detection method has already been reported for oligonucleotide microarrays [45].

Detection by Surface Plasmon Resonance

Surface plasmon resonance (SPR) is an optical method that does not require proteins to be labelled. This is an advantage because the attachment of

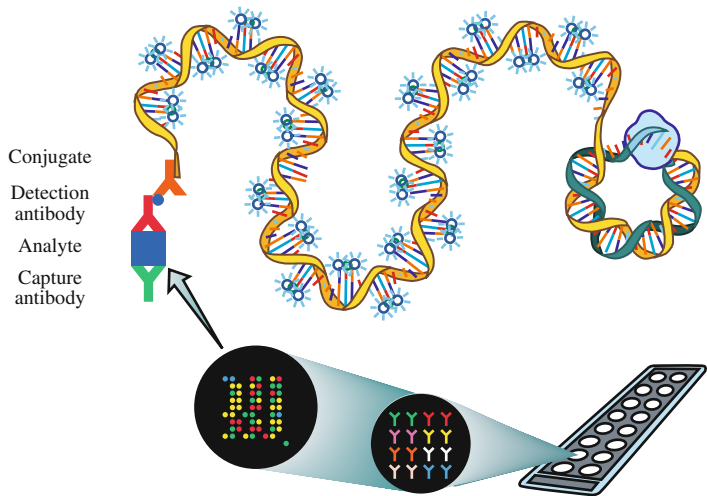


Fig. 18.11. Microarray immunoassay with rolling-circle amplification (RCA). The analyte is captured by an antibody immobilised on the array. It is detected by a secondary biotinylated antibody (detection antibody). The latter recognises an anti-biotin antibody–oligonucleotide conjugate pre-annealed with circular DNA which serves as a template. RCA replication generates a long single-strand DNA which is hybridised with fluorescent oligonucleotide [42]

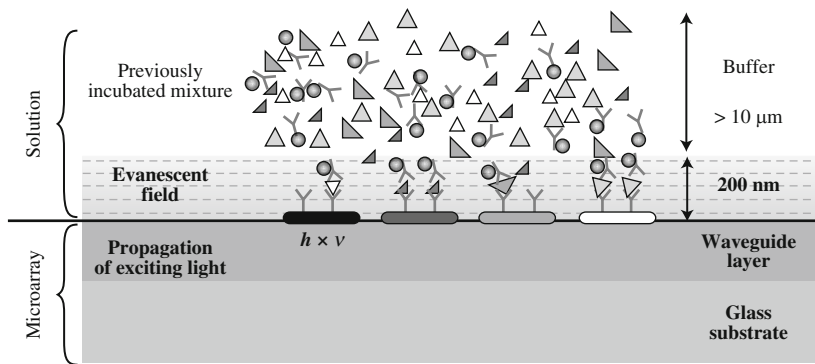


Fig. 18.12. Application of planar wave guide technology for detecting fluorescence confined at the surface of a protein microarray. The antibodies are immobilised on the array, which is incubated with a mixture of sample and detection antibodies coupled to fluorophores. Only the surface-bound fluorophores are selectively excited. Signals, correlated with the amount of analyte captured on each spot, are recorded by a CCD camera [43]

reporter groups or molecules on a protein can modify its ability to interact with other molecules, either by causing a conformational change, or by interfering with the site or sites of interaction with its partners. SPR requires the

microarray surface to be coated with a metal rich in electrons. Gold is often chosen because of its low reactivity with biological molecules.

This detection method has been used to study interactions between peptides or proteins and antibodies [46–49], and between proteins and DNA [47], and to detect the presence of pathogens such as *Escherichia coli* O157:H7, *Salmonella typhimurium*, *Legionella pneumophila*, and *Yersinia enterocolitica* using antibodies [50]. Up to 400 antigen–antibody interactions can be detected simultaneously by surface plasmon resonance using a CCD camera (the Biacore Flexchip system, commercially available from Biacore, GE Healthcare). The 400 spots are distributed over an area of 1 cm^2 and the volume of sample delivered to the array surface is about $50\text{ }\mu\text{L}$ [51]. Surface plasmon resonance imaging lends itself well to automation.

Surface plasmon resonance can monitor interactions in real time, giving access to rate and affinity constants. Qualitative and quantitative information on molecular interactions can be obtained using surface plasmon resonance imaging as detection system for protein microarrays [52]. A detection limit of 0.5 nM has been reported for an antigen–antibody pair, and the adsorption coefficient K_{abs} can be calculated and compared with the equilibrium dissociation constant obtained in solution [46]. Kinetic data (association and dissociation rate constants) have been determined for protein–peptide interactions and for enzyme cleavage by the Xa factor of peptides immobilised on a microarray [48].

Surface plasmon resonance can be combined with other methods such as mass spectrometry to identify proteins that have reacted with the protein microarray, or atomic force microscopy to determine the surface topology of the protein microarray [52].

Detection by Atomic Force Microscope

Atomic force microscopy does not require the proteins to be labelled, and provides information about topological modification of the surface occurring when molecules bind to the immobilised proteins. Data concerning the orientations of the immobilised and captured proteins, and concerning the stoichiometry of the interactions, can be obtained [32]. As an example, atomic force microscopy has been used to measure the change in height occurring during the interaction between an antigen and an antibody [32, 53].

Figure 18.13 shows a height profile obtained when a rabbit anti-immunoglobulin G (anti-IgG) antibody binds onto an immunoglobulin G. The IgG molecule, which is Y-shaped, has height 14.5 nm , width 8.5 nm , and thickness 4.0 nm . The height of immunoglobulin G molecules deposited on the array is measured by atomic force microscopy to be $6.5 \pm 0.9\text{ nm}$ ($n = 10$, Fig. 18.13a), indicating the existence of a single IgG monolayer adsorbed onto the array. When the array is set in contact with rabbit anti-IgG, there is an increase in height of each IgG deposit to $12.1 \pm 1.3\text{ nm}$ ($n = 10$, Fig. 18.13b). This value

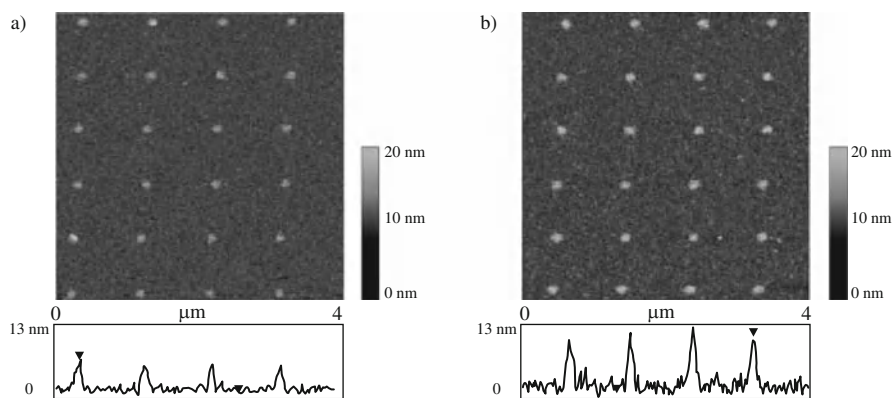


Fig. 18.13. AFM tapping mode image and height profile of rabbit IgG molecules deposited on a film of 16-mercaptohexadecanoic acid, with a pattern generated by dip-pen nanolithography and bound on a glass slide coated with gold: (a) before and (b) after contact with a solution containing lysozyme, goat/sheep anti-IgG, human anti-IgG, and rabbit anti-IgG [32]

suggests that a single anti-IgG molecule binds onto an IgG molecule and that the stoichiometry of the interaction is 1:1 [32].

Detection by Mass Spectrometry

This method does not require labelling of the proteins under investigation. The company Ciphergen has developed a technique combining chromatography chemistry and specialised affinity capture surfaces to immobilise proteins on protein chip arrays with SELDI-TOF (surface-enhanced laser desorption ionisation time-of-flight) mass spectrometry to detect and characterise the bound proteins [54]. The arrays used as support are low density protein microarrays, carrying 8–16 spots of diameter 2 mm. The detection threshold is in the femtomole range.

The main application of this technique is to study protein expression profiles. Comparing the protein profile of two samples expected to exhibit some differences (normal vs. pathological, or treated vs. non-treated samples) provides a way of identifying therapeutic targets, diagnostic markers, or markers for monitoring the effects of a treatment. The raw biological extract, e.g., serum, urine, culture medium, tissue extract, is deposited on the array. After washing to eliminate non-specifically bound proteins, the array is inserted in the SELDI-TOF detector, which calculates the molecular mass of the proteins that have been captured on the array. This approach can also be used to study molecular interactions, to investigate post-translational modifications, and to map epitopes.

18.3 Other Formats for Protein Microarrays

Protein microarrays do not all come in a flat format. Indeed, some have been developed using microbeads [55–57] or nanoparticles in suspension [57, 59]. One advantage is that each element of the array is independent. This flexibility means that a limited number of proteins can be tested without having to fabricate a new array with all the proteins. Beads are identified by a fluorescent code, while nanoparticles (nanotubes) carry a microscale optical code or bar code. An intrinsic optical parameter replaces the physical localisation used for flat arrays (see Fig. 18.14).

Qiagen, now joined up with Luminex, has developed protein microarrays in a liquid phase, a system called the LiquiChip protein suspension array. This approach uses a suspension of microbeads of diameter $5.6\text{ }\mu\text{m}$ on which capture molecules, e.g., antibodies, are immobilised. The latter will interact with their partner (the analyte) in solution. The analyte bound to the bead is detected by a reporter molecule which is specific to it and which is labelled by a fluorophore (see Fig. 18.15a).

Each microbead has a specific spectral signature obtained by varying the ratio of two organic dyes. The captured analyte is detected by a third dye bound to the so-called reporter molecule. The capture analyte is detected using a flow cytometer comprising two lasers. One laser is used to detect the microbead and the other to detect the fluorophore associated with the analyte captured at the surface of the microbead (see Fig. 18.15b). This multiplex system can simultaneously assay 100 parameters in a sample without the need for washing, by associating 100 microbeads carrying different capture molecules. It can detect about one picogram of analyte and is highly sensitive. The LiquiChip array can analyse all interactions involving proteins, in particular protein–protein, receptor–ligand, and protein–nucleic acid interactions, as well as immunoassays and enzyme assays (kinase, phosphatase, and protease).

18.4 Applications of Protein Microarrays

There are a wide range of applications for protein microarrays. Although somewhat arbitrary, one classification makes a distinction between analytical and functional microarrays [5, 52].

18.4.1 Analytical Microarrays

Analytical microarrays, i.e., antigen or antibody arrays (see Fig. 18.16), are used for large scale detection of proteins in complex media and to determine variations in protein expression by incubating the antibody microarray with biological extracts [55, 60]. These are powerful tools for proteomics. Protein expression profiles can be analysed on the scale of a genome, a cell, or a tissue,

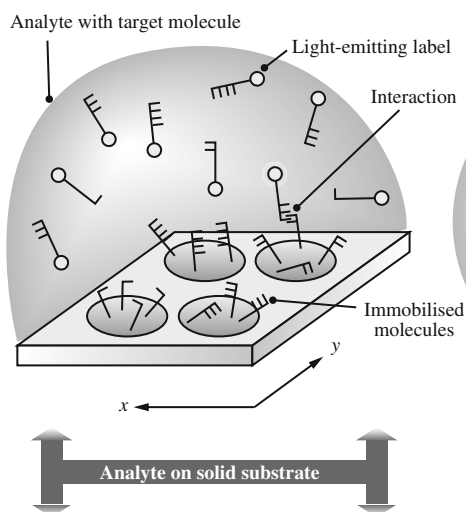
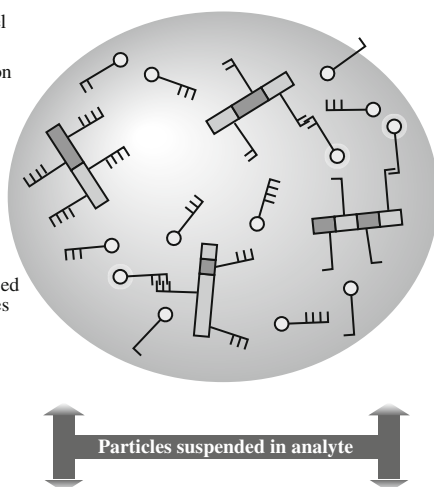
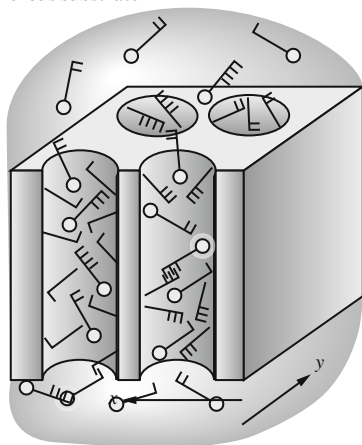
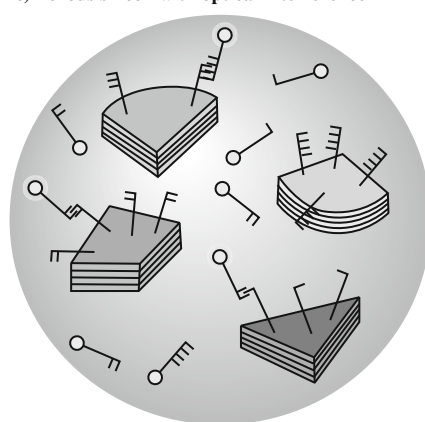
a) Flat substrate**b) Bar-coded metal tubes****d) Porous substrate****c) Porous silicon with optical interference**

Fig. 18.14. Protein microarrays can be made on solid substrates (**a**, **d**) or in suspension (**b**, **c**). The molecules binding the analyte are identified (**a**) by their coordinates x and y on a flat support, (**b**) using barcoded metal rods, (**c**) by interference colours of porous silicon fragments, and (**d**) by their coordinates x and y in a porous substrate [57]

or by comparing two samples, e.g., normal vs. pathological profile. The latter approach is used for diagnosis or prognosis.

Antibody microarrays are used to detect proteins in complex media and can be used either for direct detection, or with a sandwich format which requires selection of a pair of antibodies for each protein [61]. Many

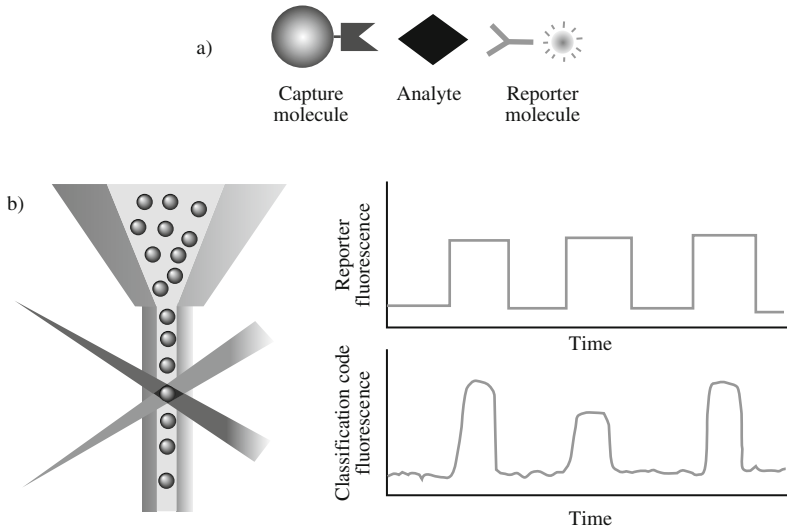


Fig. 18.15. (a) The LiquiChip system. Capture molecules are immobilised on beads by standard techniques, e.g., covalent coupling or affinity using the avidin–biotin system or a polyhistidine tag. (b) Simultaneous detection of fluorescence from the microbeads and fluorescence of the reporter molecule [58]

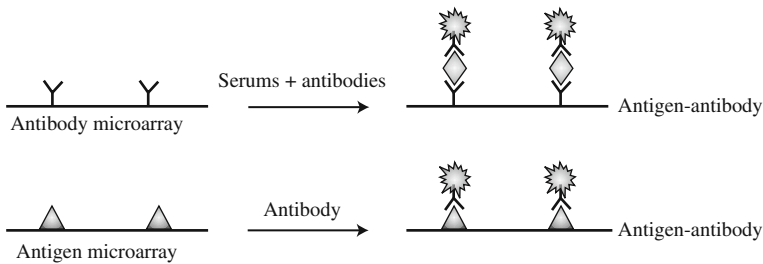


Fig. 18.16. Analytical protein microarrays [5]

manufacturers offer antibody microarrays. An array containing 512 monoclonal antibodies immobilised on a glass slide (Ab Microarray 500) is available from Clontech.

Antigen microarrays detect serum antibodies in infectious and autoimmune diseases [62–64], and allergic reactions. A chip with 196 autoantigens spotted on a glass slide has been used to detect autoantibodies present in the serum of patients suffering from autoimmune diseases such as lupus erythematosus or rheumatoid arthritis. Detection is made by fluorescence [63]. Microarrays have been designed to represent the proteomes of tissues targeted by autoimmune diseases. Examples are the synovial proteome microarray (650

potential autoantigens), a microarray for corrective tissue diseases [62], and a myelin microarray (500 proteins and peptides) used to monitor autoantibodies in autoimmune encephalomyelitis [65]. Microarrays made of 430 proteins and peptides of the proteome of the acquired immunodeficiency viruses of humans and monkeys have been used to monitor the immune response in vaccine tests [66].

Allergen microarrays have been developed to characterise the (IgE) antibody profile of patients with respect to a wide range of allergens in a single assay [67]. Work on tree and grass allergens has been carried out with recombinant allergens immobilised covalently in triplicate on a glass slide. Spot diameters lie in the range 180–250 μm and correspond approximately to 1 ng of protein per spot. The allergen microarray is exposed to a very small volume of serum (20 μL) and the IgE in the serum of the patient binds to the corresponding allergens. They are then detected by the anti-IgE conjugated to a fluorophore, and the fluorescence from the microarray is read by a fluorescence scanner. This technology can simultaneously analyse up to 400 allergens with 20 μL of serum, while conventional tests require 50 μL of serum per allergen [68].

18.4.2 Functional Microarrays

Functional microarrays are used for functional studies of proteins and are also very useful tools for diagnosis and identification of new therapeutic targets (see Fig. 18.17).

Protein microarrays are used for high-throughput analysis of molecular interactions involving proteins, to map interaction sites, and to obtain information about protein functions [4, 23, 69]. Most interactions studied are protein–protein and protein–DNA interactions [9, 70]. A microarray with 5,800 proteins of the yeast *Saccharomyces cerevisiae* has been made to study protein–protein and protein–phospholipid interactions, leading to the identification of 150 proteins binding to phospholipids [10]. This microarray has also been used to screen the specificity and cross-reactions of antibodies [71] and to study protein–DNA interactions [72]. Two hundred proteins have been identified which bind to DNA, half of them previously not known to do this [72]. Following the work by Zhu et al. [10], this microarray (Yeast ProtoArray TM) was commercialised by Protometrix–InVitrogen Corporation.

Enzyme–substrate interactions can also be studied using protein microarrays. The characterisation of 119 kinase proteins of the yeast *S. cerevisiae* with 17 different substrates has been achieved using a protein microarray. Substrates were immobilised on the array, which was then incubated with a kinase and ATP labelled by ^{33}P [38]. Kramer et al. [73] identified 21 potential substrates of the CK2 α kinase among 768 proteins immobilised on a chip.

Protein microarrays are also useful for identifying new therapeutic targets and new drugs [74]. They are used to study the interactions between proteins and small molecules [23]. Fields of application include pharmacogenomics

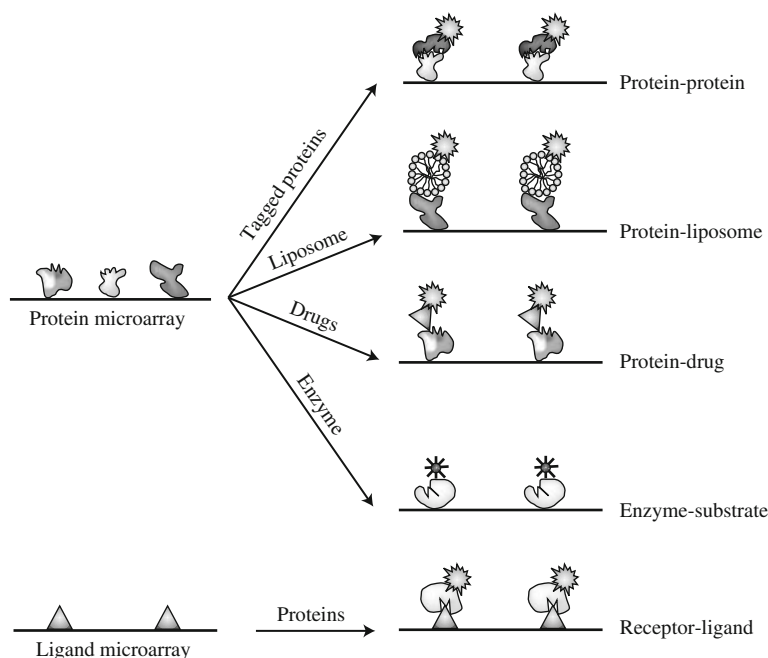


Fig. 18.17. Different functional protein microarrays [5]

and pharmacogenetics, which involve among other things the examination of protein–drug interactions.

Pharmacogenomics is the identification of genes involved in the efficiency of a therapeutic molecule and its side-effects (toxicogenomics). It leads to a better understanding of the mechanisms whereby the drug acts, by studying the interactions between the drug and the products of whatever genes are involved, i.e., proteins, which are the potential targets of such drugs. Protein microarrays, which can be used to monitor protein–drug interactions, are therefore useful tools for pharmacogenomics.

Pharmacogenetics seeks to characterise the influence of genetic variability on the patient’s response to therapeutic treatment. The study of the variation in pharmacological response due to heredity will eventually lead to the development of personalised treatment. This will require the use of protein microarrays to characterise protein–drug interactions and/or to monitor protein expression profiles during treatment. Protein microarrays can therefore be useful in customising and improving treatment, especially cancer treatment, and reverse phase protein microarrays [39] have been developed with this in mind. These arrays are made by immobilising the serum or cell lysate prepared from biopsies on the array [75].

18.5 Conclusion

Protein microarrays have been developed in research laboratories for the high-throughput analysis of protein interactions. Some microarrays are not made using full-length proteins, but individual protein domains, in order to map interaction sites. Peptide microarrays are used to determine the linear sequences involved in interactions or epitopes in the case of antigens.

Protein microarrays are also more and more commonly used in health applications, especially for diagnostics, the identification of new therapeutic targets, and the screening of molecules likely to have pharmacological activity. Automated systems have been commercialised. These make it easier to implement this technology by standardising protocols and improving the reproducibility of results. Reproducibility and the determination of standards are crucial in order to be able to compare results between laboratories, but also for the routine use of protein microarrays as diagnostic tools. The improvement of the stability of the proteins immobilised on the arrays is a prerequisite for array storage, and for their routine use in clinical applications. Microarrays containing all the proteins of a tissue or all the proteins linked to a pathology are under development and one of the main fields of application of protein microarrays in the future may be personalised medicine [76].

Technological evolution of microarrays will involve miniaturisation, an increase in the number of analyses carried out simultaneously, and a reduction in the volume of biological sample required. Eventually, microarrays will be replaced by nanoarrays, which use approximately 1/10,000th of the area occupied by a standard microarray. About 1,500 nanospots could be placed on an area occupied by just one spot of a microarray.

Addressable and barcoded nanospheres provide a promising way of making microarrays in a liquid phase or in suspension, which will make them more flexible than arrays fabricated on rigid, flat supports. Detection sensitivity and specificity will be further increased by using quantum dots and improving chemiluminescence techniques. The need to distribute samples and reagents at the surface of the protein microarray will lead to the inclusion of microfluidic systems, and this area will benefit from the development of the lab-on-a-chip.

Regarding the solid support, the future may lie in the BioCD, a prototype of which has been made by David Nolte at Purdue university (USA). Rather than containing digital data, the surface of a BioCD would contain molecules able to measure the protein concentration in serum [77]. In ten years or so, the BioCD would thus be able to carry out cheap and fast tests which would 'screen' for a large number of pathologies. Eppendorf Array Technologies already offers DNA chips on a compact disk (BioCD). The CD format has also been developed recently to design microlaboratories (Gyrolab Workstation, Gyros AB, Uppsala, Sweden).

References

1. ici.cegep-ste-foy.qc.ca/profs/gbourbonnais/pascal/fya/chimcell/notesmolecules/proteines_2.htm
2. www.physics.purdue.edu/people/faculty/nolte.shtml
3. Templin, M.F., Stoll, D., Schrenk, M., Traub, P.C., Vohringer, C.F., Joos, T.O.: Protein microarray technology, *Drug Discov. Today* **7**, 815–822 (2002)
4. Schweitzer, B., Predki, P., Snyder, M.: Microarrays to characterize protein interactions on a whole-proteome scale, *Proteomics* **3**, 2190–2199 (2003)
5. Zhu, H., Snyder, M.: Protein chip technology, *Curr. Opin. Chem. Biol.* **7**, 55–63 (2003)
6. LaBaer, J., Ramachandran, N.: Protein microarrays as tools for functional proteomics, *Curr. Opin. Chem. Biol.* **9**, 14–19 (2005)
7. Lueking, A., Cahill, D.J., Mullner, S.: Protein biochips: A new and versatile platform technology for molecular medicine, *Drug Discov. Today* **10**, 789–794 (2005)
8. Poetz, O., Schwenk, J.M., Kramer, S., Stoll, D., Templin, M.F., Joos, T.O.: Protein microarrays: Catching the proteome, *Mech. Ageing Dev.* **126**, 161–170 (2005)
9. Sakanyan, V.: High-throughput and multiplexed protein array technology: Protein–DNA and protein–protein interactions, *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* **815**, 77–95 (2005)
10. Zhu, H., Bilgin, M., Bangham, R., Hall, D., Casamayor, A., Bertone, P., Lan, N., Jansen, R., Bidlingmaier, S., Houfek, T., Mitchell, T., Miller, P., Dean, R.A., Gerstein, M., Snyder, M.: Global analysis of protein activities using proteome chips, *Science* **293**, 2101–2105 (2001)
11. Espejo, A., Cote, J., Bednarek, A., Richard, S., Bedford, M.T.: A protein-domain microarray identifies novel protein–protein interactions, *Biochem. J.* **367**, 697–702 (2002)
12. Newman, J.R., Keating, A.E.: Comprehensive identification of human bZIP interactions with coiled-coil arrays, *Science* **300**, 2097–2101 (2003)
13. He, M., Taussig, M.J.: DiscernArray technology: A cell-free method for the generation of protein arrays from PCR DNA, *J. Immunol. Methods* **274**, 265–270 (2003)
14. Ramachandran, N., Hainsworth, E., Bhullar, B., Eisenstein, S., Rosen, B., Lau, A.Y., Walter, J.C., LaBaer, J.: Self-assembling protein microarrays, *Science* **305**, 86–90 (2004)
15. Ziauddin, J., Sabatini, D.M.: Microarrays of cells expressing defined cDNAs, *Nature* **411**, 107–110 (2001)
16. Min, D.H., Mrksich, M.: Peptide arrays: Towards routine implementation, *Curr. Opin. Chem. Biol.* **8**, 554–558 (2004)
17. Panicker, R.C., Huang, X., Yao, S.Q.: Recent advances in peptide-based microarray technologies, *Comb. Chem. High Throughput Screen* **7** (6), 547–556 (2004)
18. Frank, R.: The SPOT-synthesis technique. Synthetic peptide arrays on membrane supports: Principles and applications, *J. Immunol. Methods* **267**, 13–26 (2002)
19. Schutkowski, M., Reineke, U., Reimer, U.: Peptide arrays for kinase profiling, *Chembiochem.* **6**, 513–521 (2005)

20. Eichler, J.: Synthetic peptide arrays and peptide combinatorial libraries for the exploration of protein–ligand interactions and the design of protein inhibitors, *Comb. Chem. High Throughput Screen* **8**, 135–143 (2005)
21. Ruiz-Taylor, L.A., Martin, T.L., Zaugg, F.G., Witte, K., Indermuhle, P., Nock, S., Wagner, P.: Monolayers of derivatized poly(L-lysine)-grafted poly(ethylene glycol) on metal oxides as a class of biomolecular interfaces, *Proc. Natl. Acad. Sci. USA* **98**, 852–857 (2001)
22. Lee, Y.S., Mrksich, M.: Protein chips: From concept to practice, *Trends Biotechnol.* **20** (12 Suppl.), S14–S18 (2002)
23. MacBeath, G., Schreiber, S.L.: Printing proteins as microarrays for high-throughput function determination, *Science* **289**, 1760–1763 (2000)
24. Elia, G., Silacci, M., Scheurer, S., Scheuermann, J., Neri, D.: Affinity-capture reagents for protein arrays, *Trends Biotechnol.* **20** (12 Suppl.), S19–S22 (2002)
25. Petach, H., Gold, L.: Dimensionality is the issue: Use of photoaptamers in protein microarrays, *Curr. Opin. Biotechnol.* **13**, 309–314 (2002)
26. Lounici, S., Tahar, J.: La technologie piézo-électrique et la production de biopuces, *Forum Labo et Forum Biotech* (2002)
27. Morozov, V.N., Morozova, T.Ya.: Electrospray deposition as a method for mass fabrication of mono- and multicomponent microarrays of biological and biologically active substances, *Anal. Chem.* **71**, 3110–3117 (1999)
28. Avseenko, N.V., Morozova, T.Ya., Ataullakhanov, F.I., Morozov, V.N.: Immobilization of proteins in immunochemical microarrays fabricated by electrospray deposition, *Anal. Chem.* **73**, 6047–6052 (2001)
29. Vieu, C.: La course vers l’infiniment petit: Pourquoi et comment?, *Signaux. Journal de l’Association des Anciens Elèves de l’ISEP*, no. 93, December 1998 (volume devoted to nanotechnology)
30. Ginger, D.S., Zhang, H., Mirkin, C.A.: The evolution of dip-pen nanolithography, *Angew. Chem. Int. Ed. Engl.* **43** (1), 30–45 (2004)
31. Wouters, D., Schubert, U.S.: Nanolithography and nanochemistry: Probe-related patterning techniques and chemical modification for nanometer-sized devices, *Angew. Chem. Int. Ed. Engl.* **43**, 2480–2495 (2004)
32. Lee, K.B., Park, S.J., Mirkin, C.A., Smith, J.C., Mrksich, M.: Protein nanoarrays generated by dip-pen nanolithography, *Science* **295**, 1702–1705 (2002)
33. Wilson, D.L., Martin, R., Hong, S., Cronin-Golomb, M., Mirkin, C.A., Kaplan, D.L.: Surface organization and nanopatterning of collagen by dip-pen nanolithography, *Proc. Natl. Acad. Sci. USA* **98**, 13660–13664 (2001)
34. Agarwal, G., Naik, R.R., Stone, M.O.: Immobilization of histidine-tagged proteins on nickel by electrochemical dip pen nanolithography, *J. Am. Chem. Soc.* **125** (24), 7408–7412 (2003)
35. Espina, V., Woodhouse, E.C., Wulfkuhle, J., Asmussen, H.D., Petricoin, E.F. 3rd, Liotta, L.A.: Protein microarray detection strategies: Focus on direct detection technologies, *J. Immunol. Methods* **290**, 121–133 (2004)
36. Fall, B.I., Eberlein-Konig, B., Behrendt, H., Niessner, R., Ring, J., Weller, M.G.: Microarrays for the screening of allergen-specific IgE in human serum, *Anal. Chem.* **75**, 556–562 (2003)
37. Ge, H.: UPA, a universal protein array system for quantitative detection of protein–protein, protein–DNA, protein–RNA and protein–ligand interactions, *Nucleic Acids Res.* **28** (2), e3 (2000)

38. Zhu, H., Klemic, J.F., Chang, S., Bertone, P., Casamayor, A., Klemic, K.G., Smith, D., Gerstein, M., Reed, M.A., Snyder, M.: Analysis of yeast protein kinases using protein chips, *Nat. Genet.* **26**, 283–289 (2000)
39. Liotta, L.A., Espina, V., Mehta, A.I., Calvert, V., Rosenblatt, K., Geho, D., Munson, P.J., Young, L., Wulfkühle, J., Petricoin, E.F. 3rd: Protein microarrays: Meeting analytical challenges for clinical applications, *Cancer Cell* **3** (4), 317–325 (2003)
40. Schweitzer, B., Wiltshire, S., Lambert, J., O'Malley, S., Kukanskis, K., Zhu, Z., Kingsmore, S.F., Lizardi, P.M., Ward, D.C.: Inaugural article: Immunoassays with rolling circle DNA amplification: A versatile platform for ultrasensitive antigen detection, *Proc. Natl. Acad. Sci. USA* **97**, 10113–10119 (2000)
41. Kingsmore, S.F., Patel, D.D.: Multiplexed protein profiling on antibody-based microarrays by rolling circle amplification, *Curr. Opin. Biotechnol.* **14**, 74–81 (2003)
42. Shao, W., Zhou, Z., Laroche, I., Lu, H., Zong, Q., Patel, D.D., Kingsmore, S., Piccoli, S.P.: Optimization of rolling-circle amplified protein microarrays for multiplexed protein profiling, *J. Biomed. Biotechnol.* **5**, 299–307 (2003)
43. Joos, T.O., Stoll, D., Templin, M.F.: Miniaturised multiplexed immunoassays, *Curr. Opin. Chem. Biol.* **6**, 76–80 (2002)
44. Pawlak, M., Schick, E., Bopp, M.A., Schneider, M.J., Oroszlan, P., Ehrat, M.: Zeptosens' protein microarrays: A novel high performance microarray platform for low abundance protein analysis, *Proteomics* **2**, 383–393 (2002)
45. Liang, R.Q., Li, W., Li, Y., Tan, C.Y., Li, J.X., Jin, Y.X., Ruan, K.C.: An oligonucleotide microarray for microRNA expression analysis based on labeling RNA with quantum dot and nanogold probe, *Nucleic Acids Res.* **33**, e17 (2005)
46. Wegner, G.J., Lee, H.J., Corn, R.M.: Characterization and optimization of peptide arrays for the study of epitope–antibody interactions using surface plasmon resonance imaging, *Anal. Chem.* **74**, 5161–5168 (2002)
47. Wegner, G.J., Lee, H.J., Marriott, G., Corn, R.M.: Fabrication of histidine-tagged fusion protein arrays for surface plasmon resonance imaging studies of protein–protein and protein–DNA interactions, *Anal. Chem.* **75**, 4740–4746 (2003)
48. Wegner, G.J., Wark, A.W., Lee, H.J., Codner, E., Saeki, T., Fang, S., Corn, R.M.: Real-time surface plasmon resonance imaging measurements for the multiplexed determination of protein adsorption/desorption kinetics and surface enzymatic reactions on peptide microarrays, *Anal. Chem.* **76**, 5677–5684 (2004)
49. Kanda, V., Kariuki, J.K., Harrison, D.J., McDermott, M.T.: Label-free reading of microarray-based immunoassays with surface plasmon resonance imaging, *Anal. Chem.* **76**, 7257–7262 (2004)
50. Oh, B.K., Lee, W., Chun, B.S., Bae, Y.M., Lee, W.H., Choi, J.W.: The fabrication of protein chip based on surface plasmon resonance for detection of pathogens, *Biosens. Bioelectron.* **20**, 1847–1850 (2005)
51. Usui-Aoki, K., Shimada, K., Nagano, M., Kawai, M., Koga, H.: A novel approach to protein expression profiling using antibody microarrays combined with surface plasmon resonance technology, *Proteomics* **5**, 2396–2401 (2005)
52. Yuk, J.S., Ha, K.S.: Proteomic applications of surface plasmon resonance biosensors: Analysis of protein arrays, *Exp. Mol. Med.* **37**, 1–10 (2005)

53. Jones, V.W., Kenseth, J.R., Porter, M.D., Mosher, C.L., Henderson, E.: Microminiaturized immunoassays using atomic force microscopy and compositionally patterned antigen arrays, *Anal. Chem.* **70**, 1233–1241 (1998)
54. Tang, N., Tornatore, P., Weinberger, S.R.: Current developments in SELDI affinity technology, *Mass Spectrom. Rev.* **23**, 34–44 (2004)
55. Zhou, H., Roy, S., Schulman, H., Natan, M.J.: Solution and chip arrays in protein profiling, *Trends Biotechnol.* **9** (10 Suppl.), S34–S39 (2001)
56. Nolan, J.P., Sklar, L.A.: Suspension array technology: Evolution of the flat array paradigm, *Trends Biotechnol.* **20**, 9–12 (2002)
57. Lehmann, V.: Barcoded molecules, *Nat. Mater.* **1**, 12–13 (2002)
58. www1.qiagen.com/Products/Protein/Assay/LiquiChipSystem/
59. Cunin, F., Schmedake, T.A., Link, J.R., Li, Y.Y., Koh, J., Bhatia, S.N., Sailor, M.J.: Biomolecular screening with encoded porous-silicon photonic crystals, *Nat. Mater.* **1**, 39–41 (2002)
60. Jenkins, R.E., Pennington, S.R.: Arrays for protein expression profiling: Towards a viable alternative to two-dimensional gel electrophoresis?, *Proteomics* **1**, 13–29 (2001)
61. Nielsen, U.B., Geierstanger, B.H.: Multiplexed sandwich assays in microarray format, *J. Immunol. Methods* **290**, 107–120 (2004)
62. Hueber, W., Utz, P.J., Steinman, L., Robinson, W.H.: Autoantibody profiling for the study and treatment of autoimmune disease, *Arthritis Res.* **4**, 290–295 (2002)
63. Robinson, W.H., DiGennaro, C., Hueber, W., Haab, B.B., Kamachi, M., Dean, E.J., Fournel, S., Fong, D., Genovese, M.C., de Vegvar, H.E., Skriner, K., Hirschberg, D.L., Morris, R.I., Muller, S., Puijn, G.J., van Venrooij, W.J., Smolen, J.S., Brown, P.O., Steinman, L., Utz, P.J.: Autoantigen microarrays for multiplex characterization of autoantibody responses, *Nat. Med.* **8**, 295–301 (2002)
64. Robinson, W.H., Steinman, L., Utz, P.J.: Protein arrays for autoantibody profiling and fine-specificity mapping, *Proteomics* **3**, 2077–2084 (2003)
65. Robinson, W.H., Fontoura, P., Lee, B.J., de Vegvar, H.E., Tom, J., Pedotti, R., DiGennaro, C.D., Mitchell, D.J., Fong, D., Ho, P.P., Ruiz, P.J., Maverakis, E., Stevens, D.B., Bernard, C.C., Martin, R., Kuchroo, V.K., van Noort, J.M., Genain, C.P., Amor, S., Olsson, T., Utz, P.J., Garren, H., Steinman, L.: Protein microarrays guide tolerizing DNA vaccine treatment of autoimmune encephalomyelitis, *Nat. Biotechnol.* **21**, 1033–1039 (2003)
66. Neuman de Vegvar, H.E., Amara, R.R., Steinman, L., Utz, P.J., Robinson, H.L., Robinson, W.H.: Microarray profiling of antibody responses against simian–human immunodeficiency virus: Postchallenge convergence of reactivities independent of host histocompatibility type and vaccine regimen, *J. Virol.* **77**, 11125–11138 (2003)
67. Harwanegg, C., Hiller, R.: Protein microarrays in diagnosing IgE-mediated diseases: Spotting allergy at the molecular level, *Expert. Rev. Mol. Diagn.* **4**, 539–548 (2004)
68. Deinhofer, K., Sevcik, H., Balic, N., Harwanegg, C., Hiller, R., Rumpold, H., Mueller, M.W., Spitzauer, S.: Microarrayed allergens for IgE profiling, *Methods* **32**, 249–254 (2004)
69. Smith, M.G., Jona, G., Ptacek, J., Devgan, G., Zhu, H., Zhu, X., Snyder, M.: Global analysis of protein function using protein microarrays, *Mech. Ageing Dev.* **126**, 171–175 (2005)

70. Korf, U., Wiemann, S.: Protein microarrays as a discovery tool for studying protein–protein interactions, *Expert. Rev. Proteomics* **2**, 13–26 (2005)
71. Michaud, G.A., Salcius, M., Zhou, F., Bangham, R., Bonin, J., Guo, H., Snyder, M., Predki, P.F., Schweitzer, B.I.: Analyzing antibody specificity with whole proteome microarrays, *Nat. Biotechnol.* **21**, 1509–1512 (2003)
72. Hall, D.A., Zhu, H., Zhu, X., Royce, T., Gerstein, M., Snyder, M.: Regulation of gene expression by a metabolic enzyme, *Science* **306**, 482–484 (2004)
73. Kramer, A., Feilner, T., Possling, A., Radchuk, V., Weschke, W., Burkle, L., Kersten, B.: Identification of barley CK2alpha targets by using the protein microarray technology, *Phytochemistry* **65**, 1777–1784 (2004)
74. Zhou, F.X., Bonin, J., Predki, P.F.: Development of functional protein microarrays for drug discovery: Progress and challenges, *Comb. Chem. High Throughput Screen.* **7**, 539–546 (2004)
75. Wulfschuhle, J., Espina, V., Liotta, L., Petricoin, E.: Genomic and proteomic technologies for individualisation and improvement of cancer treatment, *Eur. J. Cancer.* **40** (17), 2623–2632 (2004)
76. Jain, K.K.: Applications of biochips: From diagnostics to personalized medicine, *Curr. Opin. Drug Discov. Devel.* **7**, 285–289 (2004)
77. news.uns.purdue.edu/UNS/html4ever/2004/040518.~Nolte.CD.html

Cell Biochips

B. Le Pioufle and N. Picollet-D'Hahan

19.1 Biochips for Analysing and Processing Living Cells

A cell biochip is a microsystem, equipped with electronic and microfluidic functions, designed to manipulate or analyse living cells. The first publications in this emerging area of research appeared toward the end of the 1980s. In 1989 Washizu described a biochip designed to fuse two cells by electroporation of the cytoplasmic membrane [1].

19.1.1 From Single Cells to Reconstituted Tissue

Research centers have devised a whole range of cell chip structures, for simultaneous or sequential analysis of single cells, cell groups, or cell tissues reconstituted on the chip. The cells are arranged in a square array on a parallel cell chip for parallel analysis, while they are examined and processed one by one in a microchannel in the case of a series cell chip. In contrast to these biochips for high-throughput analysis of a large number of cells, single-cell chips focus on the analysis of a single isolated cell.

Parallel Cell Biochips

As in DNA microarrays, where a large number of oligonucleotides are ordered in a matrix array, parallel cell chips order living cells in a similar way. At each point of the array, the cells can be isolated, provided that the cell type allows this, e.g., blood cells, or cultivated in groups (most adhesion cells can only survive in groups). The aim is to allow massively parallel analysis or processing. Le Pioufle et al. describe a microdevice for the culture of single cells or small groups of cells in a micropit array [2]. Each pit is equipped to stimulate the cell or group of cells either electrically or fluidically. Among the applications envisaged are gene transfer, cell sorting, and screening in pharmacology. A complementary approach, combining the DNA microarray and

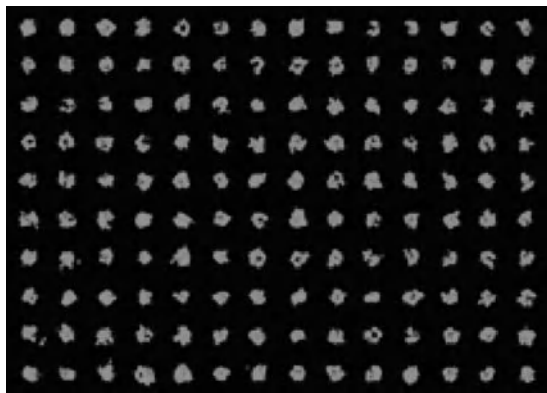


Fig. 19.1. Reverse transcription by cDNA of green fluorescent protein (GFP) arranged in an array on a chip. Cells cultivated on this substrate express the gene in a way that depends on their position in the array

cell biochip ideas, has been put forward by Bailey et al. [3]. Genes previously arrayed on the chip transfect the cultured cells on the substrate depending on their position in the array (see Fig. 19.1). This way of achieving differential lipofection on a chip was then taken up again by Yoshikawa et al. [4] with primary cells, more relevant for the medical applications under investigation. This idea of cell transfection on a DNA microarray is described in more detail in Chap. 17.

Series Biochips

Series biochips handle the analysis, processing, and sorting of cells sequentially. This approach to the cell chip can be likened to miniature flow cytometry. The cells move through a microchannel, separated from one another, and can be shunted into other microchannels, sorted, or temporarily immobilised for future analysis or processing, using the physical principles described below. An example of such microdevices for conveying cells by microfluidics is discussed in [5, 6].

Biochips for Manipulating and Analysing Single Cells

These microdevices can be used to isolate and precisely position a single cell in a region of the chip where it can be analysed and/or processed. In the work by Huang et al. [7], the target cell is carried to and then immobilised in a microchamber, by virtue of a micro-aperture used to generate a suction flow. The chip is designed so that an electric field can be focused on, and hence concentrated on, the immobilised cell permeabilising its cytoplasmic membrane for the subsequent transfer of genes (see Fig. 19.2). In the same university (Berkeley), other researchers have put forward an alternative method for the

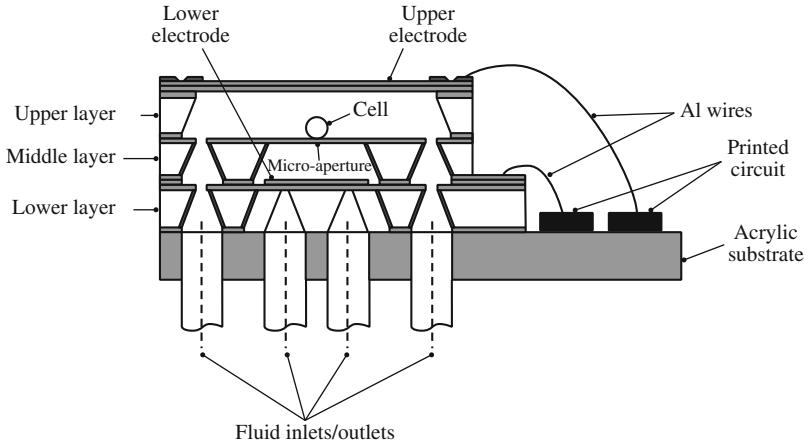


Fig. 19.2. Biochip for electroporation of a single cell. The cell is brought precisely to the fluidic aperture in the upper chamber and subjected to an electroporation field. Observations are carried out through the upper electrode, which is transparent. Copyright Elsevier (2001). With the kind permission of Boris Rubinski

electroporation of single cells or the measurement of transmembrane potentials in an ensemble of single cells (patch clamp), using suction to attract the cell or cells onto the opening of a polymer microfluidic channel [8,9]. An alternative method, combining microfluidics and electric fields, has been proposed by Seger et al. [10]. The single cell is isolated and in equilibrium between two force fields, being pushed by a microfluidic current onto a repulsive force field produced by negative dielectrophoresis. Once again, the application in mind is electroporation of a single cell.

Tissue Models on a Chip

In this case, the aim is to reconstitute relevant physiological models on the chip. In the end one hopes to analyse behavioural physiological models *ex vivo*, or reconstruct an artificial organ on the chip. Tissue or cell architecture cultures are built on a chip equipped to receive the relevant biochemical or electrical physiological signals. Most work in this area concerns neural networks on chip [11], but some research groups are working with cardiac cell tissues [12], liver cell tissues [13], muscle cell tissues [14], and the association and communication of different cell tissues [15].

19.1.2 Cell Micromanipulation Methods

Cell biochips can convey, guide, link, sort, and precisely position cells to be analysed and/or processed on the active part of the chip. There are several methods for micromanipulating cells, in series or in parallel.

Dielectrophoretic Methods

An electric field gradient generates a force on the cell due to its different complex dielectric permittivity from the surrounding medium. In the work reported in [5] and [16], the cells are immobilised sequentially in a dielectrophoretic cage to analyse and hence sort them. In [17], the authors describe a parallel cell chip to simultaneously arrange cells into an array by dielectrophoresis. The idea of dielectrophoresis, described in 1951 by Pohl and used to displace electrically neutral but polarisable particles [18], is indeed applicable to living cells. Microdevices have shown the feasibility of using this idea for cell separation [19], cell fusion [20], and cell sorting [21,22]. The expression for the dielectrophoretic force exerted on a cell (modelled to a first approximation by a sphere of radius r , permittivity ε_c , and conductivity σ_c) is [23]

$$\mathbf{F} = 2\pi r^3 \varepsilon_m \operatorname{Re} [K(\omega)] \nabla (E^2),$$

where

$$\varepsilon_c^* = \varepsilon_c - j \frac{\sigma_c}{\omega}, \quad \varepsilon_m^* = \varepsilon_m - j \frac{\sigma_m}{\omega}, \quad K(\omega) = \frac{\varepsilon_c^* - \varepsilon_m^*}{\varepsilon_c^* + 2\varepsilon_m^*},$$

and ε_c , ε_m , and σ_c represent the respective dielectric constants and conductivities of the cell and the medium in which it is immersed, with ω and $\mathbf{E}$ the angular frequency and effective strength of the electric field.

In the case where $\operatorname{Re}[K(\omega)] < 0$, the cells are brought by the negative dielectrophoretic force toward the minima of the field. In the opposite case of positive dielectrophoresis, the cells are brought toward the maxima of the field.

Mechanical or Fluidic Methods

A microdevice applies a mechanical force to the cell to set it in motion. A ferromagnetic microthruster, driven by an external magnetic force, conveys a single cell (an oocyte in the case studied) [24]. The mechanical force can also be produced by a fluid movement. The cells are then driven toward the desired positions on the chip, using the suction force produced by micro- or nanocapillaries [25,26]. These capillaries go through the quartz substrate [25] or silicon substrate [26] of the chip. A low pressure applied under the substrate creates a flow of culture medium through these openings, carrying the cells along until they are micropositioned in such a way as to block the flow. In [25], the micro-openings are obtained by ion bombardment of the quartz substrate, followed by chemical etching with hydrofluoric acid (HF). In [26], anisotropic KOH and TMAH etching of orthogonal channels on the two faces of a silicon substrate produces these submicron openings at the intersection of the pair of channels.

Biochemical Methods

Micropatterning of the substrate guides cell adhesion on predefined surfaces. In [27], the cells proliferate on surfaces functionalised by affinity proteins, delimited by zones that are repulsive for the cells, e.g., polyethylene glycol (PEG). In [28], the active surfaces of the chip are functionalised by antibodies with specific recognition of different cell types. These methods are used to precisely position cells on the chip by means of self-assembled molecular layers on very precisely delimited regions of the chip. To do this, polymer buffers, on which the molecules making up the precursor layer are adsorbed, are generally used to apply these molecules to the substrate. An alternative method is to graft this first layer of molecules on a gold surface by thiol bonds. In this case, the gold surfaces are marked out on the substrate by photolithography. The ensuing layers are self-assembled, up to the layer of adhesion or recognition molecules of the relevant cells.

Optical Methods

Optical tweezers are used [29] to lift and immobilise a single cell at the focal point of a laser. The various immobilisation methods can be combined, as in [30] where the dielectrophoresis force is combined with optical tweezers.

19.1.3 Methods for Characterising Microcultured Cells on Chip

Cell biochips are equipped to characterise cells or analyse their behaviour under environmental stimuli. The biochip must interpret the relevant physical data best characterising the cell response (protein response, transcriptome, mechanotransduction, electrical response, response of an ion channel, etc.). Microcultured cells on the chip are stimulated optically, mechanically, or electrically. Likewise, the cell response is observed optically, mechanically, or electrically. Here are some examples:

- The protein expression of a cell can be detected and studied dynamically on a chip. To do this, one detects by optical means the expression of a transgene associating the promoter of the relevant gene with a gene coding for a fluorescent protein, viz., green fluorescent protein (GFP) in most cases. The cell chips proposed by Huang et al. [7] use this idea to check the efficiency of gene transfer by electroporation. Likewise, the cell chips proposed by Wu et al. [31] use this idea to assess the efficiency of a pharmacological product on an array of transgenic cells.
- Cell biochips can prove very useful for studying the mechanical interaction between the cell and its culture substrate, i.e., mechanotransduction, and also for studying cell motility. Various mechanical structures have been proposed in the literature to measure the forces exerted by the cell on its support. An array of micro-cantilevers can be used to measure with very great accuracy the forces exerted by fibroblasts migrating on the chip

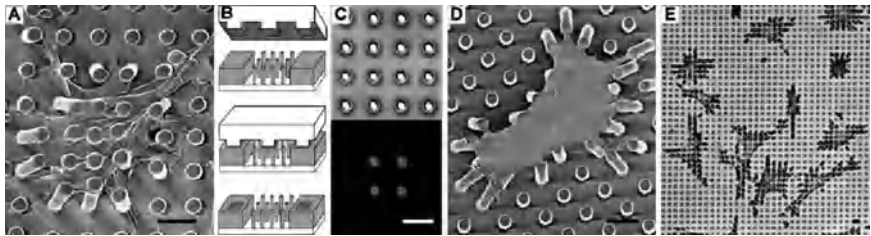


Fig. 19.3. Measuring the forces exerted by a cell on the substrate by deformation of polymer spots. The spots are functionalised by fibronectin for cell adhesion. Copyright 2003 National Academy of Sciences, USA. With the kind permission of Christopher Chen

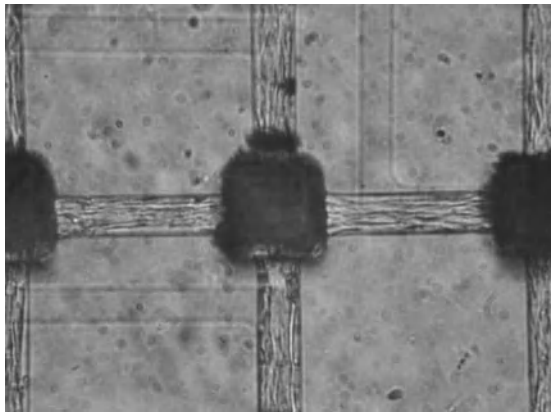


Fig. 19.4. Neuronal network on a MEDPROBE chip (Panasonics). The membrane potentials of the groups of neurons are detected by electrodes. The groups of neurons communicate via bundles of axons guided by polydimethylsiloxane microfluidic channels. From F. Morin et al. [45]

(a measurement accuracy of 0.2 nN has been achieved) [32]. Fluorescent nanobeads are included in the polymer supporting the cell culture [33]. The elastic deformations of the polymer induce measurable displacements of the nanobeads and this can be used to measure the forces exerted by the migrating cell. One can thereby obtain a map of the forces on the whole cell interaction region. A parallel approach has been proposed [34] in which the cells adhere and exert driving forces on an array of polymer spots functionalised by an adhesion protein. The deformation of the spots allows one to deduce the mechanical stresses (see Fig. 19.3). More recently, cell microsystems have found an application in microrobotics. Xi et al. [35] have demonstrated the feasibility of a microrobot with legs made of muscle tissues cultured on chip.

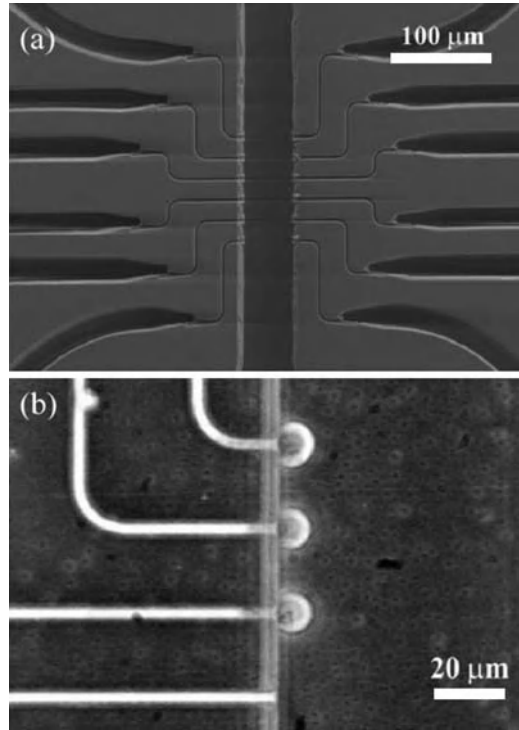


Fig. 19.5. Electrophysiology on a microfluidic chip. Patch-clamp array with medium (a) and high (b) integration density. With the kind permission of Luke P. Lee

- Since cell biochips are the result of technology derived from microelectronics, they naturally integrate functionalities for analysing cells by detecting their electrical response, characterising them by their impedance or membrane potential. Gawad et al. [36] have proposed a microdevice for the flow characterisation of non-adhesive cells by impedance spectroscopy on a series biochip. Before characterisation, cells are first focused between the electrodes by negative dielectrophoresis, which improves the reproducibility of the measurement and makes it possible to sort the cells after analysis. With regard to adhesive cells, a great deal of work is described in the literature, concerning measurement of the membrane potential of on-chip neuron cultures [37–39]. In [40], signals are detected with ex vivo neural architectures defined by microfluidic confinement on the chip (see Fig. 19.4). There are potential applications in pharmacology and toxicology [41]. With this in mind, analysis of the impact of a tested compound is usefully complemented by measuring the activity of the ion channels of these cultured neurons. Patch-clamp techniques (see Sect. 19.2), giving access to this data, are amenable to miniaturisation, as attested in several publications (see Fig. 19.5) [42–44].

19.2 Patch-Clamp Microarrays

19.2.1 Motivations

Limitations of the Patch-Clamp Technique in the Pharmaceutical Industry

The generation and transmission of electrical signals in cells depend on specialised transmembrane proteins, viz., ion channels (see Chap. 11). These channels are involved in both physiological and pathological processes [46]. They handle the emission and propagation of activating electrical inputs, transform physical and chemical stimuli into electrical signals which propagate to the nervous system, and maintain communications between the cell and the surrounding medium. Furthermore, the ion channel is currently identified as one of the molecular elements in which the pathologies known as channelopathies arise. Indeed, some neurological, cardiovascular, renal, endocrine, exocrine, muscular and other anomalies are a consequence of spontaneous mutations affecting sodium, potassium, calcium, and chloride channels.

Functional analysis of these channel proteins has become a genuine bottleneck in the process of discovering new active pharmacological compounds, and constitutes a major challenge for the pharmaceutical industry, which aims to discover and validate innovative molecules specifically targeting these channels and other membrane transport systems [47, 48].

Although the patch-clamp technique remains the reference for studying these channels [49], providing direct access to the currents generated by ion flow, this technique is not well suited to high-throughput screening, since it can be neither miniaturised nor automated [50, 51]. Although it is very effective, it is indeed extremely vulnerable to interference effects, requiring expert handling for experimental implementation and providing a low test efficiency. But faced with the ever-increasing number of candidate molecules and mutant proteins, generated by new techniques of combinatorial chemistry and combinatorial genetics, respectively, there is now an urgent need for reliable and efficient new high-throughput screening (HTS) tools that require very small sample volumes.

Locating Techniques for Studying Ion Channels in the Drug Development Cycle

As discussed in Chap. 11, the patch-clamp technique is the only direct way to study electrical properties of individual ion channels or natural or recombinant protein receptors. No matter how elegant and effective it is, there are some major disadvantages which prevent it from revolutionising the discovery of receptors and drug development. This research and development is carried out according to a value chain in which the techniques used must fulfill a certain number of criteria. The synthesis of therapeutic molecules and testing

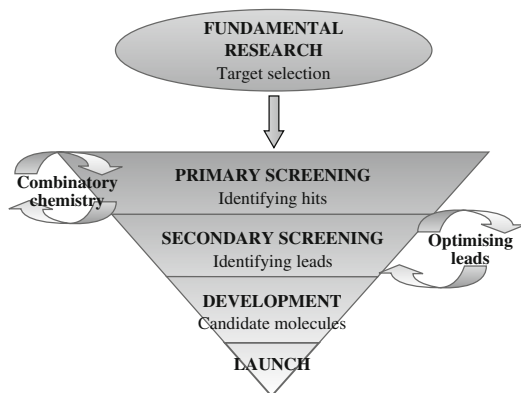


Fig. 19.6. Research and development cycle and screening cycle for the synthesis of candidate drugs

Table 19.1. Techniques for studying ion channels in the value chain of research and development of active molecules. ++ widely used, + sometimes used, – never used

Stage	Patch-clamp	Binding	Flow	Fluorescence
Fundamental research	++	++	++	+
Primary screening	–	++	++	++
Secondary screening	++	–	++	++
Toxicological screening	++	–	+	++

of molecules stored in vast chemistry (or compound) libraries (see below) requires this panel of techniques at different stages. In other words, even the indirect techniques discussed in Chap. 11 are used at certain stages of the research and development process, because they are sufficient to answer the questions relevant at this particular stage. The patch-clamp technique, although more effective and more direct, has too low a throughput to be able to satisfy the medium- and high-throughput needs of the tests in the pharmaceutical industry (see Fig. 19.6).

Throughput and Screening

- *Chemical Library.* This is a storage bank containing hundreds or thousands of chemical compounds, which can be divided up into libraries devoted to different uses. The storage format varies. The main body of the chemical library is often in powder form. These powders are then used to make liquid compound libraries in pitted plates (with 96, 384, etc., pits), which are used in high-throughput screening (HTS).
- *Screening.* This is the selection of molecules with potential therapeutic interest that are likely to become drugs.
- *High-Throughput Screening (HTS).* This is used by the pharmaceutical industry to seek out candidate drugs in chemistry libraries with 96, 384, or 1536

experiments running simultaneously. From 10,000 data per day: 1 piece of data = 1 drug concentration per test.

- *Primary Screening.* 500,000/1 million compounds per run. Selection of hits.
- *Secondary Screening.* Capacity is less important here. The aim is now to validate hits and identify leads, i.e., candidate molecules, by the practice known as cherry-picking, i.e., take the hits, reformat the plates, and reconfirm.
- *Counter-Screening.* Checking on non-transfected cells.
- *Safety Screening Profiling.* Carried out on hERG channels. Test capacities are increased to detect side-effects as soon as possible.
- *Low-Throughput.* In academic research laboratories, with 100 measurements per day.
- *Key Screening Parameters.* Speed, robustness, and cost. At the present time, the maximum acceptable cost is 0.5 dollars per HTS data.

19.2.2 Emergence of New Patch-Clamp Platforms

Introduction

In the area of ion channel research, the first decade of this century has been marked by the arrival of new technologies, in two main directions: automated recording on cells using conventional electrodes, and patch-clamp array systems [50, 52].

The recent history of the automated patch clamp is characterised by very rapid evolution, in terms of both technology and economics, with the appearance of a dozen or so spin-off and start-up companies between 2000 and 2003, in Europe (Germany, Sweden) and the United States. In 2005, only two or three companies lead the market, having strengthened their position by mergers and buyouts. Despite the appearance of the first commercialised systems in 2003, technical progress is expected, as we shall see at the end of this chapter, to answer certain user needs that are not yet satisfied. The present aim of research endeavour in this field, e.g., at the French atomic energy authority in Grenoble, is to fill various technological gaps in such a way as to satisfy the demands of the pharmaceutical and biotechnology industries, but also the needs of academic research centers. We shall see that the different approaches do in fact set out from the same basic idea, but differ with respect to certain building blocks, often related to some specific knowhow or skills, or to some locally well developed infrastructure.

Microelectrodes and Automated Recording of Cell Currents

Let us begin with a brief historical review, mentioning some of the original hints of innovation in the 1990s. One of the first ideas aimed to automate the patch clamp while preserving its current configuration, i.e., the glass micropipette format (see Fig. 19.7a). The idea was to control a complex and repetitive manual process automatically by computer, with a view to improving the screening efficiency and reproducibility. Even in the 1990s, robotic

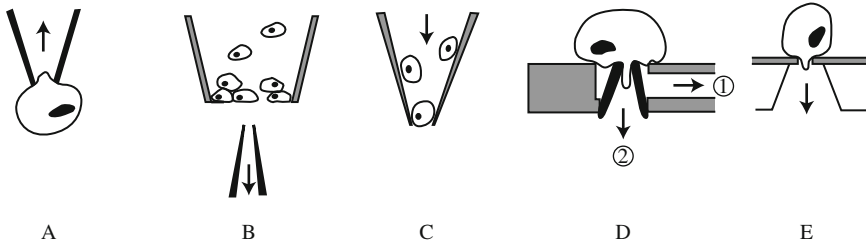


Fig. 19.7. Different patch-clamp configurations. (A) Conventional patch clamp (micropipette). (B) Interface patch clamp (micropipette in contact with a cell suspended by a gradient density). (C) Flip-the-tip arrangement (cells positioned within a pipette). (D) Cells positioned (suction channel 1) and sealed (suction channel 2) by two distinct channels. (E) Plane patch-clamp (cells positioned and sealed on a micropatterned chip). The *arrow* indicates the direction of suction

systems were able to automate the patch-clamp approach on mammal cells. For example, there was an automatic cell recognition and analysis system with 8 individualised cells in separate chambers. The test efficiency was between 30 and 100 results per day. However, since the cells were cultured and patched on solid surfaces, the system required accurate selection of the cell and a pipette positioning system designed to ensure that the glass recording electrodes would not break. Another system was devised in Great Britain for blind patching of mammal cells (see Fig.19.7b). The system takes the cells to be patched to the electrode, working at the air–liquid interface in a glass capillary. The advantage with this interface patch clamp is that it avoids problems due to vibrations and does not require ultra-precise positioning of the electrode. The present device, comprising a single recording chamber, has a similar test efficiency to the manual patch clamp, with more than 50% success in the whole-cell configuration. Generations with 18 chambers required a huge parallelisation effort. It is no easy matter to coordinate 48 pipettes.

Still following the idea of forming an array of pipettes, an automated system of parallel measurements with 8 electrodes and 96 pits was introduced in Germany [53]. The idea was rather novel and unexpected. The seal was made on cells placed inside the micropipettes (see Fig. 19.7c). Here is another example of a field that would like to defy the patch-clamp dogma.

Innovation is also to be found in the environment of the conventional patch clamp itself and in the introduction of accessories which optimise the way it is implemented. For example, a high-speed laminar flow microfluidic system using multichannels has applications involving several compounds on the same patched cell, and is well suited to tests on channels that desensitise quickly [54].

As discussed in Chap. 11, the xenopus oocyte is widely used in the patch-clamp approach, with the advantage of being rather large. In the double

electrode technique, the experimenter impales the oocyte with two glass microcapillaries, in which the electrodes are bathed in physiological liquid. An obvious idea is to replace the two arms of the experimenter by two robot arms, and some laboratories have taken it up with a view to automated recording on oocytes [55,56]. In this way, several automated measurement systems were independently invented in the 1990s in several pharmaceutical laboratories. In 1994, a completely automated prototype was proposed for sequential measurement on oocytes, recording on 96-pit plates. Other devices followed, e.g., a double electrode recording system able to analyse 8 oocytes simultaneously, and carrying out several hundred measurements per day.

But the technological race was only at its beginnings. It turned out that neither the automated patch clamp nor robotic measurement systems on oocytes were able to converge upon a high enough level of parallelism or miniaturisation. Further developments with high-throughput screening systems become more and more difficult due to the complexity involved in controlling and positioning the multiple recording electrodes. In this context, much of the industrial and academic research effort turned to on-chip ion channel measurement systems. In this chapter, we shall be concerned only with the patch-on-a-chip systems, rather than fluorescence measurement systems.

On-Chip Measurement Systems

Over the past few years, microfabrication technologies have been used to develop biochips with a view to significantly improving the efficiency of biological tests (see the review in [57]). Examples are DNA microarrays for sequencing, detection of mutations, and gene expression studies. By immobilising a large number of DNA probes at different anchor sites, typically arranged in a lattice of rows and columns, it has been possible to carry out massively parallel DNA hybridisation experiments. Likewise, biochips have been developed for electrophoretic separation of molecules and dielectrophoresis, and any other cell manipulation procedure based on a physical approach. Finally, there is currently a great interest in the lab on a chip, i.e., laboratories in the form of integrated and functional chips. So it is natural enough to find the extension of biochips into the field of ion channel analysis. As we shall see, the aim is not to reinvent the very principles of electrophysiology, but rather to reorganise them by adapting them to a new format.

A Brief History

The patch clamp on a chip is the result of an international academic research effort. The companies that have recently been involved in developing the skills for and selling this technique are witness to the fact, and the scientific publications appear as a series of landmarks, each one suggesting some possible revolution of the conventional patch-clamp technique.

In Switzerland, research teams have innovated in the techniques for positioning the cell to be analysed (electrophoretic flow) [58]. At Yale in New

Haven (USA), a new technology has been developed for the fabrication of the electrodes [59,60], and this research was supervised by Fred Sigworth, who contributed to the original development of the patch-clamp approach with Bert Sackmann and Erwin Neher. This technology led to the design of an industrial test platform integrating planar chips that could simultaneously carry out multiple electrical measurements on 16 single cells [61]. This technological race, as mentioned above, is currently moving at full tilt. It was in 2004 that the world leader on the planar electrophysiology market established its position [62], closely followed by three European start-up companies [63–67].

Plane Patch-Clamp Array

Construction of a micropipette array soon reaches its limits, faced with the fragility and bulkiness of the system, and the challenge of parallelisation. In this context, a novel idea was born, together with a new concept. The innovation was to replace the traditional recording micropipettes (see Fig. 19.7a) by a planar chip, with an average size of a few square millimeters and thickness less than one millimeter, complete with an array of orifices of diameter 1–2 μm . The idea was to retain only the key part of the micropipette, i.e., its tip of diameter 1–2 μm directly interfacing the cell membrane. The rest of the micropipette was in fact fluidic, playing the roles of reservoir for the saline solution and measurement electrode. The chip format then had a certain number of advantages (see below) and opened the way to new prospects.

Plane Patch Clamp Array: Advantages and Questions

Some advantages of the planar format are:

- Several measurement sites on the same substrate $\implies$ parallelisation.
- Mature fabrication methods $\implies$ accuracy and reproducibility of measurement sites.
- Cells on a planar substrate $\implies$ improved stability and lifetime of the seal.
- Small and compact chip $\implies$ reduction in electron interference effects.
- Small volume of chambers $\implies$ reduced volume of reagents.
- Microfluidic approach $\implies$ easier to dispense drugs.
- Wide range of dielectric materials $\implies$ wide variety of substrates.
- Similarity with slide format $\implies$ combination with microscope techniques.

Some questions raised by the patch-clamp array are:

- Is it possible to form the gigaseal on any other material than glass?
- How do the hole geometry, surface roughness, and other properties affect the gigaseal?
- How can the measurement sites be controlled individually?
- How can recording be automated?
- How can electronic interference between the different sites be minimised?
- How can the vast amount of generated data be processed?
- What is the cost of such an instrument and how much does it cost per test?

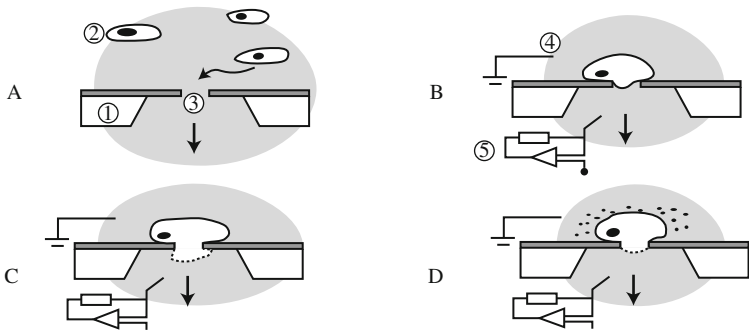


Fig. 19.8. Whole-cell recording (see Chap. 11) on a patch-clamp array. (A) Positioning a cell on the microhole of a planar chip by suction (1 chip, 2 cell, 3 micro-opening). (B) Formation of the gigaseal (4 Ag/AgCl electrode, 5 preamplifier). (C) Transition to whole-cell configuration by increased suction. (D) Whole-cell recording with drug dispensing in the upper chamber

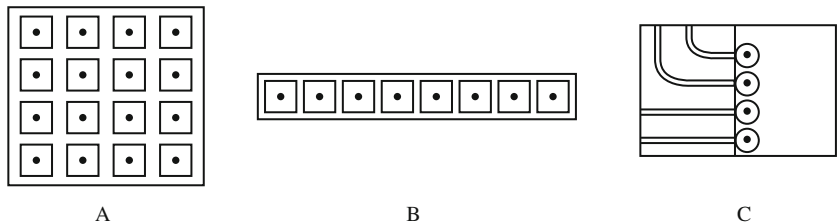


Fig. 19.9. Different chip configurations. (A) Array chips, e.g., 4 × 4 measurement sites. (B) Rod-shaped chips, e.g., 8 measurement sites. (C) Transverse chips in which each measurement site (there are 4 here) is formed at the intersection of the fluidic channel and the glass slide

This innovative idea led to a new concept. Indeed, thanks to the planar configuration, a single cell would be brought to and immobilised on the measurement site (microhole), rather than bringing the end of a micropipette to an immobilised cell at the bottom of a Petri dish. This reverse patch-clamp idea was destined to introduce new technological constraints.

In practice (see Fig. 19.8), the cell suspension is placed on the chip by a microfluidic system and a cell on the microhole addressed by various procedures. Once the cell is immobilised on the microhole, contact between cell and orifice is set up in a few seconds and the high resistance seal or gigaseal is produced, often favoured by a slight pressure reduction via a suction channel. The chip carries several micro-openings for rapid positioning of several cells and parallel measurement of the generated currents. The solutions on either side of the chip can be exchanged quickly by a microfluidic system, whence the measurement conditions can be varied for the same set of cells. This system overcomes the experimental constraints of the patch-clamp approach, viz., anti-vibration table, micromanipulation, drawing of pipettes, manual dexterity of

the experimenter, etc. It aims to optimise test efficiency without reducing the temporal resolution of the measurements.

While this concept has now become the norm, innovations in the technology or procedure are implemented to try to get around technological bottlenecks. One major problem lies in the crucial stage of the patch-clamp method, viz., the formation of the gigaseal. It is easy to see the importance of the material when designing and fabricating these chips. Another source of difficulties comes in designing the electronics of a measurement site (as in standard patch clamp) or n measurement sites. Likewise, microfluidic constraints are introduced by the need to achieve parallelisation and independence of measurements. However, not all solutions put forward require the revocation of proven theories. For example, one does not have to reinvent the electronics of the patch clamp, but rather to adapt it to the new format, miniaturising, integrating, and so on. Let us examine each of these technological bottlenecks more closely, and consider the various solutions that have been proposed.

Different Configurations for Plane Devices

Array Chips. The most common arrangement for measurement sites is an array (see Fig. 19.9a). For example, some chips have an array of 48 micro-openings, i.e., an 8×6 array, with 16 pits analysed in parallel [63]. This format can be adapted to conventional automated pipettes (2 rows of 8 pipettes) which are already used in industry to fill or empty pit plates in which chemical libraries are stored. So using a robotic arm of this kind, cards with 48 micro-openings are filled (drug dispensing) in 3 stages. Likewise, some cards with 8×48 measurement sites are handled by a pipette arm with 12 channels [68].

Rod-Shaped Chips. Some groups have developed rods of 16 pits to be adaptable to the pit plate format for drug dispensing, e.g., $12 \times 8 = 96$ pits (see Fig. 19.9b) [61]. This format allows a certain flexibility in implementing the fluidics, while the array format soon encounters a problem of overcrowding in the architecture of the electrical wiring or fluid channels one seeks to integrate. Other groups have also produced a rod of 4 chambers and another with 16 that is under development [66].

Transverse Chips. A microfluidic network with circular format has been proposed by a research team at Berkeley USA (see Fig. 19.9c) [69, 70]. In one of their systems, the micropore is in fact replaced by the junction between microfluidic channels in a disk format radiating out from a round central chamber and a glass slide that closes the channels. This format reduces the volumes used, and reduces the unwanted capacitive coupling between the channel and the chamber, as well as facilitating exchange of the media during the experiment.

Solutions for Technological Bottlenecks

Choice of Material. As discussed in Chap. 11, the molecular mechanisms involved in the formation of the gigaseal between the micropipette and the

Table 19.2. Performance of the automated patch-clamp pipette technology. The whole-cell configuration provides access to the totality of the ion currents (see Chap. 11)

Innovation	Gigaseal success rate	Transition to whole-cell	Whole-cell efficiency
Patch-clamp interface	Not available	Suction	50–60%
Cell in pipette [53]	75–82%	Suction (losses < 10%)	80%
[63]	Not available	Suction	Not available

Table 19.3. Performance of the patch-clamp chip technology. DEP = dielectrophoresis

Innovation	Gigaseal success rate	Transition to whole-cell	Whole-cell efficiency
Glass chip with surface modification [61]	1 Gohm with DEP	Suction (losses < 1%)	> 75%
PDMS chip [59, 60]	50% under suction, 13% gigaseal on oocytes	Suction	50%
Glass chip [65]	Double suction		
Chip made from Si + Si ₃ N ₄ + charges [58]	Electrophoresis		
Hybrid plastic chip with glass surface. Plate 384 [62, 68]	Suction 20–250 Mohm	Amphotericin	60–80% 95%
Glass chip [67]	Suction 30–50 Gohm	Suction	30–50%
Silicon chip with biocompatible non-conducting material [63]	Planar fluidic channels and electroosmotic pumps		

membrane are still rather poorly understood. For this reason, it is hard to establish precise specifications for the ideal material. Only experience in the field allows one to identify a certain number of parameters, known to affect the quality and behaviour of the seal. For example, the surface state of the material in contact with the cell is crucial. It must be clean, smooth, and hydrophilic, and rendered adhesive by some means. At the present time, three types of material stand out for this application: glass, polydimethylsiloxane (PDMS), and silicon.

- *Glass Chips.* Glass is the standard material since it is a very good dielectric (dielectric constant equal to 4) compared to silicon ($\epsilon_r = 10$), and interacts strongly with lipid membranes. Furthermore, it is relatively cheap and its optical properties make it well suited to coupling with spectroscopy

and imaging techniques. The main constraint with these quartz chips lies in the microfabrication of apertures with very precise geometry. Etching techniques using heavy ion bombardment have been developed [71–73]. However, it is hard to imagine electronic integration with this type of material. Current performance in terms of the seal and success rate is summarised in Tables 19.2 and 19.3

- *Polymer Chips.* Sigworth and coworkers at Yale (USA) were the first to demonstrate the feasibility of PDMS as substrate for patch-clamp chips with xenopus oocytes [59]. The probability of recording ion currents on small mammal cells nevertheless remains low (around 10%) (see Tables 19.2 and 19.3). On the other hand, ease of fabrication makes it well suited to laboratory use. Once again, these authors exploited and transposed the interesting dielectric properties of PDMS, well known to electrophysiologists under the name of Sylgard (see Chap. 11), to the planar format. The PDMS is used here after an oxygen plasma treatment to make the surface hydrophilic. This change of state is temporary, since the polymer molecules restore the hydrophobic nature of PDMS. The reversion time depends on the duration of the surface treatment and the environment in the storage place (for storage in air, the chip should be used within half an hour of treatment, and in water, within 2 h). This surface treatment of the PDMS for electrophysiological measurements rather complicates the experimental procedure, because one must integrate a plasma system into the patch-clamp setup. PDMS is a transparent material. This optical neutrality means that methods of microscopy could be integrated into the system to carry out combined measurements. However, its instability and the difficulty in industrialisation significantly reduce the chances for such an application. Studies with this material are better viewed as fundamental research to validate ideas than as industrialisation. Other polymers are also used, such as polyimide polyethylene terephthalate, in which micro-openings are etched. To increase the seal resistance, the surfaces of these chips receive a silicon oxide deposit [62].
- *Silicon Chips.* The silicon is coated with a thermal oxide layer (SiO_2) or silicon nitride layer (Si_3N_4 , $\epsilon_r = 4.5$) to reduce the capacitance (from a hundred or so picofarads down to about ten, compared with glass which has a capacitance of about 1 picofarad) and make the surface hydrophilic. As the microfabrication procedures are well understood, silicon has all the advantages of an industrialisable material [63, 74, 75]. Silicon has an advantage in terms of cost, due to the large-scale fabrication of chips, and also due to the possibility of integrating active elements, including novel electronic functions, within the chip (e.g., integration of preamplifiers [63]).

One idea is to take the best of each of the materials and put them together in order to best exploit their properties. This has led to hybrid chips, with polymer parts for the packaging aspect of the chip, often non-active parts which may cover large areas, and silicon parts for measurement sites, since these are

either active parts in which electronics may be integrated, or micropatterned parts with precisely specified geometry and surface state. Work by Pantoja [76] provides a good illustration of the combination of silicon with PDMS for the microfluidics. Chips are generally sold encapsulated in a polymer support.

Positioning and Capture of a Single Cell. Within the context of research and development of cell biochips, the problems associated with manipulating a single cell are a common theme (see Chap. 11). The challenge with planar patch-clamp systems is, starting with a cell population, to attract, accurately guide, then immobilise a single cell on a micro-aperture [77, 78].

The most widely used protocol is the one used in conventional patch-clamp applications, namely suction. This time this is done from below the cell rather than from above, and it tends to be automated to overcome the random and awkward features of sucking with the mouth. Note in passing that this randomness may be the solution for obtaining a good seal. Indeed, the exact pressure difference to be applied is left to the appreciation of the experimenter, who proceeds in stages, stopping and starting in response to the behaviour of the cell. As a result, the experimenter effectively establishes a pressure/seal resistance feedback system, since he/she modulates the suction while visualising the value of the resistance, and desists when the required value is achieved. It is thus very difficult to set up any standard pressure cycle, which would in principle be different from one experimenter to another. And yet such a pressure scheme would be very useful for automation of the seal. Suction must be implemented rather quickly to avoid any sedimentation around the microhole. Cells adhering around the edge of the aperture would not be sensitive to the pressure drop and would never reach the orifice. The sphere of influence of the suction around the aperture can in fact be evaluated experimentally.

Other capture procedures have been proposed, often combined with suction for the seal stage. One innovative positioning method by electrophoretic flow was developed by C. Schmidt [58]. A membrane vesicle, viz., a liposome or artificial lipid structure, was successfully brought to a microhole. The idea is as follows. A potential difference of a few hundred millivolts is imposed across a silicon nitride membrane of thickness 100 μm , and this electric field guides small charged particles toward the microhole. A further layer of 20 nm of SiO_2 modified by aminosilanes or by adsorption of polylysines makes it possible to form gigaseals on liposomes, but not on eukaryotic cells. The stronger electric fields required would have damaged the cells.

Another procedure, a kind of dielectrophoretic trap, can immobilise a cell in a potential well produced by means of a quadrupole located above the micro-opening [79].

A concentric suction procedure [65] was shown to dissociate the guiding and sealing of the cell in an elegant way. This is done with a quartz chip containing two concentric openings, each connected to a different microfluidic channel. By sucking fluid through the channel of aperture 10 μm , the cell is

guided toward the inner opening of the channel, with diameter about $1\text{ }\mu\text{m}$. This inner channel functions like the inside of a patch-clamp pipette.

Another chip configuration, developed by a Swiss group [80] and by a group at Berkeley in the USA [81], has a micronozzle at its surface which stands in for the tip of the micropipette and helps the cell to settle on the measurement site. Whatever device is used, a final suction is always applied to invaginate the membrane on the walls of the hole.

Making the Seal: The Key Parameters. Whether a chip is used as a planar patch clamp or to detect the electronic properties of neurons (see below), the aim is the same: to minimise the gap between the chip surface and the cell membrane in order obtain high-resolution recordings. A neuron adheres to an inert surface with a residual gap of 60–100 nm [82]. The thickness of the fluid film between cell and substrate depends on the type of cell, the surface treatment, and so on. The question is whether the suction will suffice to eliminate this gap. In the case of the patch clamp, the answer is that it will not always do so. Experience shows that the creation of the seal will depend on a range of parameters related in particular to the substrate, but also, as we shall see below, to the type and quality of the cell. In the literature, one can find attempts to collect together the optimal surface parameters for an ideal chip. It turns out that favourable parameters are a smooth, hydrophilic, dielectric surface and a micro-opening of round cross-section, of diameter about $2\text{ }\mu\text{m}$, with smooth, non-angled edges and thickness between 2 and $10\text{ }\mu\text{m}$. Experiments have shown that a silicon nitride membrane thickness of 120 nm does not procure a sufficient contact area between the walls of the hole and the membrane to form a gigaseal [71]. In any case, the perfect recipe, judiciously combining all these attributes, has not yet been found. For this reason, the success rate for obtaining the seal remains low (between 40 and 80%) and poorly reproducible. One of the building blocks attesting to the importance of specific knowhow lies in the surface chemistry or processing, often patented and kept a closely guarded secret. Efforts are now converging toward the development of adequate topology and surface coating, formed by very thin films of polymers or bioactive molecules which tend to improve the seal and also to guide cells toward the measurement site.

Note, however, that in devices that are currently under development, the notion of gigaseal is altogether relative! Indeed, some researchers do not adopt this standpoint, and do not seek to achieve the gigaseal at any cost. A seal resistance of the order of 100 Mohm is sometimes taken as adequate, or at least acceptable for the analysis of sufficiently large currents. Moreover, a statistical approach referring to the implementation of n seals in parallel is regarded as an advantage for increasing the probability of success! Fifty percent success rate, i.e., the efficiency of conventional patch-clamp methods, on 96 cells at the same time, instead of just one, would always be considered a good performance. Furthermore, only those cells that succeed with the seal and the transition to whole-cell recording will receive the drug, and this reduces loss of compounds. Another strategy for increasing the chances of the seal being produced is

to carry out redundancy tests with quadruplicates for each compound, even though this procedure is rather long. To get around this problem, the authors have adopted a current recording system averaged over 384 cell populations arranged at the bottom of a pit plate, where each pit is provided with several analysis sites.

In all cases, experience has shown that once the cell (or cell debris) has adhered to the microhole, there remains a residue that is very difficult to remove and prevents any further seal. This is why these chips are viewed as consumables, whence the need to succeed immediately at the first contact between cell and measurement site. There can be no second chance.

Micro- or Macrofluidics? In current devices, the fluidics supplying the chip are usually in capillary format and use robotic pipettes to dispense drugs and/or cells. Some devices include a planar fluidic system with electroosmotic pumps to handle fluids, which removes the need to apply a pressure (see Fig. 19.10a). Drug dispensing constrains the type of material, because drugs are often highly hydrophobic compounds. The material must not adsorb the drug as this would bias dose-response tests. Furthermore, it must be inert, without releasing chemical compounds or molecules adsorbed during previous tests. Appropriate surface treatments are envisaged to get around these problems, e.g., a silica deposit on the polymer card. Planar fluidic channels, a small volume for the analysis chamber, and dispensing in a laminar flow are some of the possibilities for reducing the required volume of precious samples, i.e., cells and drugs, and for optimising the washing and cycle times [83]. While it is useful to miniaturise fluidic channels in order to save precious molecules, a macrofluidic scale can clearly be retained upstream of the chip, for reservoirs holding electrophysiological medium, washing medium, water, alcohol, and so on. However, the microfluidics should be integrated in such a way that solutions can be exchanged very quickly on the two faces of the chip. Moreover, this flow must be able to create a laminar ligand front to target the cell, while avoiding any turbulence and any mechanical destabilisation of the patched cell.

Integration of Electrical Measurements. Two types of measurement are available here, viz., mix and read (see Fig. 19.10b) and continuous flow measurement (see Fig. 19.10c). Mix and read desynchronises the drug dispensing and electrical measurement stages. For this reason, the analysis of certain types of channel such as ligand-gated (chemically sensitive) channels cannot be analysed in real time by fast fluidic exchange (often around twenty milliseconds). On the other hand, the continuous flow system (see Fig. 19.10c) uses a more integrated microfluidic system which handles all stages of the test without discontinuity, from measurement without the drug, to drug dispensing, and complete washing of the cell, and which allows finer analysis of the functional interactions of chemical compounds on the ion channels. In current systems, the electronics remains outside the device. For example, some systems include 8 superposed two-channel amplifiers outside the measurement platform. There

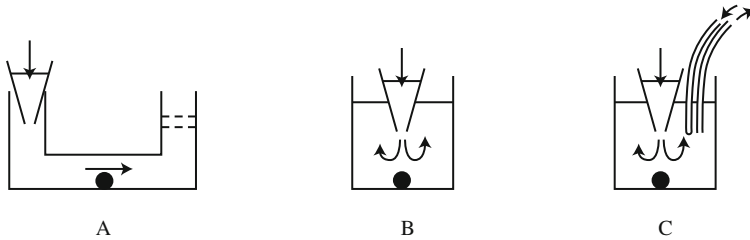


Fig. 19.10. Ligand perfusion systems. Adapted from [90]. **(A)** Electroosmotic pump guiding the flow. **(B)** Mix and read. Electrical measurements are not synchronised with drug delivery. **(C)** Continuous flow system. The solution is modified during the electrical measurement

must be some real advantage in integration. Indeed, it seems sensible to try to integrate the preamplifiers or current–voltage converters so that they can be as close as possible to the measurement sites. Integrated on the chip, they would also be less sensitive to external interference. Moreover, miniaturisation means reducing capacitances and hence increasing the resolution of recorded signals. Sigworth suggests the long-term aim of monolithic amplifier arrays. The first steps toward integrating the electronics were taken with cards integrating 48 preamplifiers or amplifiers containing 48 units but not integrated into the chip. A mobile support moves this amplifier during the measurements.

Compatibility of Cell Preparation and On-Chip Measurement. Reversing the conventional patch-clamp method also introduces constraints on the biological sample to be analysed. The idea now is to move a single cell toward a chip which remains fixed, rather than bringing a pipette to a cell immobilised at the bottom of a Petri dish. This seemingly slight difference means selecting the right cell and manipulating it independently of the others. But what means do we have at our disposal for choosing this cell? None whatever! The systems on offer work blind. With luck, suction will trap a cell! And only electrical control by measuring the seal resistance can tell whether a cell has been positioned, then sealed. The problem then is to optimise the chances of success for this random step. To do this, efforts are made upstream in the cell suspension, to ensure that it is homogeneous and free of aggregates, membrane debris, dust, and dead cells. The cells have to have a round morphology and good membrane integrity, and they must be well separated from one another. Protocols aiming to improve conditions of culture, digestion, resuspension, and centrifugation have been established [84]. Other work has concentrated on the judicious choice of cell medium to facilitate recovery and also conservation of the cells before dispensing. Current systems use cell lines expressing the same type of ion channel in a rather uniform way (see Chap. 11). Hence, if a cell is trapped, one knows a priori the electrical characteristics. What happens if one is concerned with transiently transfected cells which express the relevant channel at the 50% level? Once again, statistically speaking, there is

a certain probability of capturing a transfected cell, but the efficiency is low. This probability is all the higher as the number of cells per analysis chamber is increased. While this need for a large amount of sample remains a lesser constraint for lines with stable expression, which are usually readily available, it constitutes a real limitation for the analysis of primary cultures or transient lines. Current research aims to reduce sample volumes, as we have seen with the integration of microfluidics. Effort is also being brought to bear on cell sorting upstream, or planar patch systems that could combine microscopic observation with electrical measurement. With this in mind, transparent substrates like glass, quartz, silica, or certain polymers will be preferred.

Interfacing Living Beings and Microelectronics

The work of Peter Fromherz at the Max Planck Institute in Germany has demonstrated the feasibility of the first cell–electronic junction [85–87]. Placing snail neurons, which are large and easily isolated, between probimide spots on an electronic circuit, he showed that the signal emitted by the electronic chip transits via the two neurons connected by a synapse (connection established by chemical molecules or neurotransmitters) and is once again transferred to the chip. This is a perfect interface between cell and microelectronics! The chip includes a silicon stimulator and the signal is recovered by a field-effect transistor (FET). The probimide spots serve to constitute a cage around the cell body of the neuron, leaving the dendrites sticking out to form synapses with neighbouring neurons. The cell bodies are thus immobilised on the active parts of the silicon and no longer tend to move around randomly during dendritic extension. The extracellular signals recorded remain somewhat noisy [85]. However, the coupling between living system and semiconductor has been clearly demonstrated. As far as applications are concerned, by positioning these devices near isolated cells, it is possible to measure the cell response to the presence of drugs or hormones. This discovery opens up a whole new world for the development of sensors in medicine. This electronic system should be considered as both the extracellular electrode and the current amplifier. When the cell sends an electrical impulse, ions circulate in the extracellular region and this ion flow induces charges in the conduction channel of the FET by capacitive coupling. These charges then modulate the current between the source and drain. The authors are developing new systems based on the use of semiconducting polymers, which have the advantage of being transparent and thus allowing optical access to the cell–substrate interface. In addition, these polymers are used to send electrical signals to the cells and set up non-toxic communication in both directions. Another advantage lies in the chemical functionalisation of these substrates and the grafting of active groups like peptides or antigens, which can provide better control over and stabilise the adhesion of the cell to the FET.

These extracellular interfaces represent an alternative to the invasive aspect of the patch clamp or intracellular electrodes, and also to the toxic nature

of voltage-sensitive fluorescent probes, for example. For this reason, long term recording is possible, to monitor the evolution of the electrical profile of a cell under the influence of drugs.

Microelectrode Arrays (MEA)

Many devices have been developed in which cells and also tissues are cultured and studied in vitro. This is in fact one of the first applications of integrated systems in cellomics [88]. Cell cultures placed in analysis chambers are used to study the effects of drugs, osmotic response, cytogenetics, immunology, metabolism, and so on. Chips integrated with fluidic channels designed to dispense nutrients have been presented by Heuschkel [89]. Multielectrode arrays (MEA) are interfaced with the cell culture to record and stimulate membrane signals. Three-dimensional arrays of electrodes in the form of microtips have also been proposed in order to record deep cell layers in organotypic slices. These electronic microchambers allow highly localised electrical interaction between neurons or cardiac cells and can be used to study the effects of drugs on ion channels [90,91]. These cell cultures can also be integrated on multi-parametric sensors [92].

19.2.3 A Cultural Revolution? Prospects

‘Patchers Versus Screeners’

This review title by J. Comley [93], which appeared in 2003, stresses the different opinions regarding high-throughput electrophysiology. This change of scale in electrophysiology gave rise to many hopes and expectations within the ion channel community. The author spells out the different points of view which oppose traditional electrophysiologists and screeners, i.e., supporters of high throughput, especially with regard to expected specifications and the operating modes of these new instruments (see Table 19.3). But is there really such a gulf between those developing the technology and the end user [94]?

Patchers give priority to the quality of recordings. They do not therefore adhere to the idea of separate measurements before and after administering the drug. The discontinuity in this sequence leads to loss of information, such as the very rapid response kinetics of some channels. On the other hand, screeners claim to be more pragmatic in seeking a compromise, accepting a seal resistance of 150 Mohm, which can nevertheless provide a basic pharmacological answer identical to the one that would be obtained by conventional patch-clamp techniques. And it is this yes/no answer that the patchers find inadequate. Only the discovery of new targets by this system could convince the patchers of the worth of this technology. Some manufacturers [47] think that these new tools may revolutionise the development process for new drugs, viewed as a kind of funnel with a million molecules to screen at the outset and only one that will eventually find its way onto the market! High-throughput electrophysiology can shorten certain stages of the process, such

Table 19.4. The pros and cons of the conventional patch clamp

Advantage	Disadvantage
High information content: • voltage control • single cell • temporal resolution	Requires long training and skilled handling
High sensitivity: • picoamperes (pA) • single molecule	Low efficiency (<10 cell/day)
Flexibility: • change of solution	Experimental complexity
Applicability: • any channel type	Ill-suited to primary screening
Few false negatives compared with indirect methods	Ill-suited to automation and parallelisation

as primary screening, or at least a first screening of small specific chemistry libraries, by giving not just a simple yes/no answer, but information about kinetic properties, ion specificity, voltage sensitivity, and the active lifetime of drugs. According to some authors [47, 63], its impact would also be manifest in the validation stage, or confirmation of hits, and in the optimisation stage for candidate molecules (see Fig. 19.11). Likewise, by launching counter-screening programmes (e.g., on hERG channels,¹ as discussed in Chap. 11), high-throughput electrophysiology could detect the side-effects of drugs very early on, even before the pharmacovigilance stage.

Although the patchers may agree regarding technological performance, they nevertheless think that these systems will never replace the patch-clamp expert. The latter will keep his/her place in the development of tests and interpretation of results. In this trust developing period, it will be the ability of a system to generate highly resistant and reproducible seals that will be the most important factor in convincing electrophysiologists that they should adopt a given technology. However, according to Netzer [95], one should first assess the projects of the pharmaceutical industry or biotech corporations in order to select a technology that precisely satisfies their needs. Hence, to meet the needs of a screening campaign, the best strategy will probably be to combine several technologies. High-throughput screening of 30,000, 100,000, or 500,000 compounds using binding or fluorescence techniques (see Chap. 11) will lead to the detection of the first hits. The compounds thereby selected

¹ The potassium channel (human ether-a-go-go related gene) responsible for the repolarisation of the ventricular action potential. By targeting these channels, one can eliminate candidate molecules that might have secondary effects on the heart.

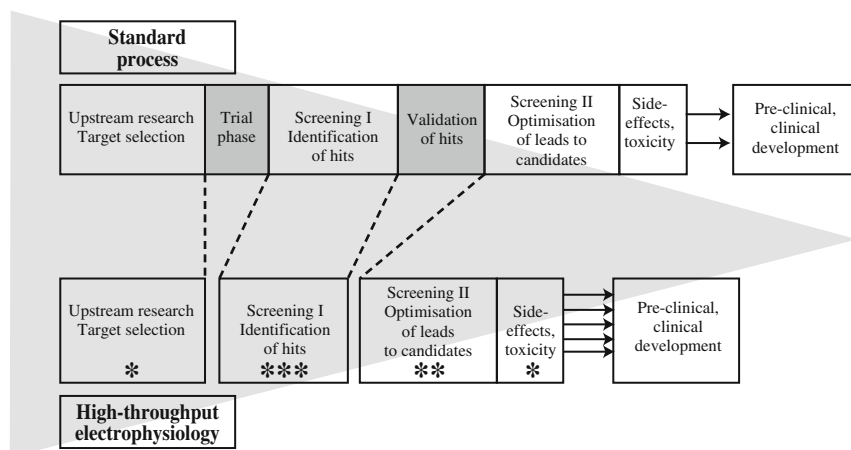


Fig. 19.11. Discovery of new ion channel modulators and expected impact of high-throughput electrophysiology on certain steps in the process. Adapted from [47]

(1–2% of the total) and these same compounds synthesised chemically will then be assessed by medium-throughput automated electrophysiological systems to be optimised by the user.

Through a better understanding of their genetics and molecular biology, the expectation is that more and more pathologies will be diagnosed as being related to an ion channel disorder. High-throughput electrophysiology would become an essential tool in the discovery of new channels among those coded by the 300 genes identified in the human genome program [96–98].

Expected Technological Progress

Combining Electrophysiology and Fluorescence

In the cell-on-chip field, current effort is converging upon a multiparametric analysis of single cells in order to obtain a complete and relevant functional and phenotypic signature in real time. But also in the ion channel field, attempts are being made to associate a phenotypic signature with this electrical signature using optical detection of morphological and membrane characteristics, recording of chemical signals, or ligand analysis. In short, the aim is to obtain a multiple molecular and electronic signature. Fluorescent probes are very useful tools for monitoring molecular motions of proteins. For example, by simultaneously recording the change in fluorescence and ion currents, it has been possible to understand how voltage-gated channels ‘feel’ the voltage. Such studies have been carried out on channel populations, but successfully correlating the spectroscopic changes in a fluorophore with individual ion currents would be a whole lot more informative [99].

Even with the conventional patch clamp, experiments were carried out in 2002 that couple the glass micropipette with a microscope nanopipette imaging the membrane surface [100]. Indeed, this was motivated by the fact that few or no tools provide an approach to ion channels that is both functional and spatial. Now cell functions are often governed by the spatial distribution of these proteins at the membrane surface. In this context the smart patch clamp or patch-clamp scanner is a high resolution (100 nm) spatio-functional technique for studying the ion channels in subcellular compartments and nanostructures. Briefly, a nanopipette mounted on an x, y, z piezo obtains the topography of the cell surface under the control of a scanning ion conductance microscope (SICM). Once the relevant structure has been identified, the nanopipette is used to record the ion currents in the cell-attached or inside-out configuration. This approach allows one to reach intracellular nanostructures that remain invisible to conventional microscopy. It can be used for any functional channel without necessarily knowing the molecular structure and without having to bind the cell. In contrast to biochemical labelling techniques, this approach obtains the functional characteristics of the protein distribution. Finally, it can be used on small cells and submicrostructures that are inaccessible to the conventional patch clamp. Possible applications also include the mapping of mechanically- and ligand-gated channels, insofar as the nanopipette can deliver chemical, electrical, or mechanical stimuli.

A Technological and Biological Race

The achievements of microtechnology, chemistry, and tools for microscopic characterisation, e.g., atomic force microscopy (AFM) [101], may one day make it possible to determine the ideal candidate chip bringing together all the properties required of the gigaseal [102]. Experience also shows that a large part of the variability in the results lies in the biological variability of the sample, and that the cell type, cell size, membrane integrity, and level of channel expression are all factors affecting the level of success of tests. Pushing these observations to the extreme, the answer would be to design, develop, and implement one chip for each given cell model [103]. The idea of a universal chip seems utopian, even though various groups are developing flexible analysis systems which tend to adapt to any chip geometry. It has thus been observed that Chinese hamster ovarian (CHO) cells patch very well on a chip, in contrast to human embryonic kidney (HEK) cells, while the opposite is observed using patch-pipette.

A useful trend in this field would be to focus efforts on optimising cell lines for the expression of target ion channels, and the winner in this technological race will no doubt be the one to propose a quality service in cell biology in association with new tools.

In a review in 2003, Wang and Li see the future in terms of predictions, and their vision would be a sophisticated, multifunctional system capable of

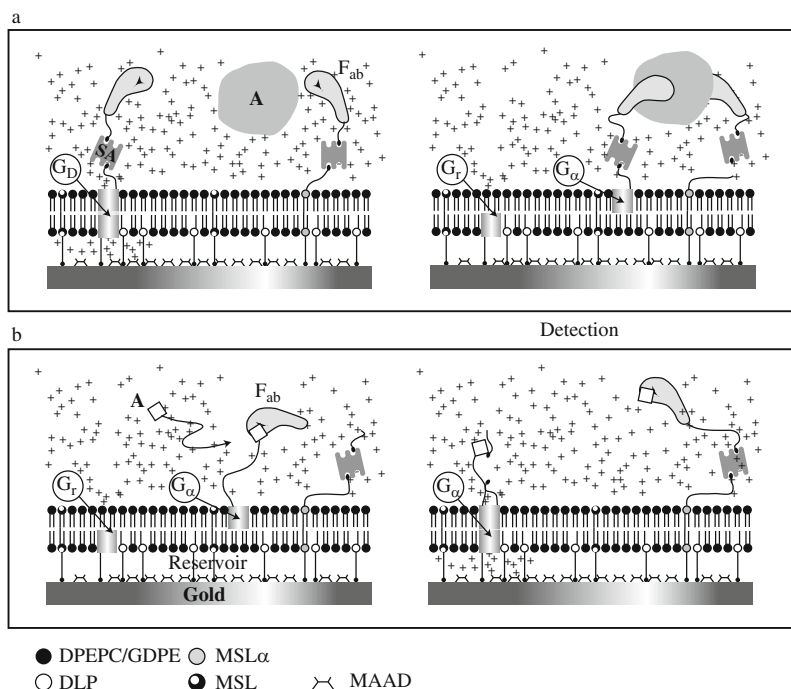


Fig. 19.12. A biosensor using the ion channel as switch. Model of a molecular biosensor combining a biological recognition mechanism with a signal transduction technique. Taken from [104]

producing large amounts of high quality data [51]. This system would also be equipped with better integrated microfluidics, including active functions such as dielectrophoresis, in order to sort the relevant cells from a heterogeneous population and then transport them to recording sites. In short, the electrophysiologist's dream.

Ion Channels or Biosensors?

Imagine a molecular interaction resulting in a change in electrical conductance, and a single molecule activating an electrical switch, sensitive enough to convert this chemical signal into a very weak detectable and amplifiable current. This is precisely what one would call a biosensor, i.e., a sensor integrating a biological sensitive element or a derivative of such. The ion channel seems by its biophysical properties, i.e., specificity, selectivity, stability, and so on, to fulfill the requirements of this definition. A very elegant system was described by Cornell in 1997 (see Fig. 19.12) [104], showing that the conductance of a population of ion channels is modified following a biological recognition event. The size of the biosensor can be reduced to become a component of

a microelectronic circuit. It can be used for cell typing and the detection of large proteins, viruses, antibodies, DNA, electrolytes, drugs, pesticides, or any other component of low molecular weight. Cornell's ion channel biosensor comprises a lipid membrane containing gramicidin (a bacterial membrane polypeptide whose double helix forms a channel through which sodium and potassium ions are exchanged between the bacterium and its environment) and attached to a gold electrode. The gramicidin is itself bound to antibodies specifically recognising the analyte. Gramicidin has two subunits (here, one is anchored to the membrane and the other mobile) which, once assembled, form a conducting channel. On the other hand, when the analyte is present, it binds to the antibodies and forms a complex which holds the mobile channel subunit away from its partner anchored in the membrane. This prevents the conducting dimer from forming and reduces the electrical conductance of the membrane. This goes to make an effective sensor, since a single molecule of gramicidin allows the flow of more than a million ions a second.

Conclusion

The technological achievements and the first available systems clearly show the worth of the patch clamp on chip. However, high throughput has not yet been achieved in pharmacology and certain difficulties remain to be resolved, such as the integration and coupling of microfluidics, electronics, optics, and so on, not to mention the question of data processing. Some dream of a complete system able to procure multifaceted molecular signatures in real time and on the scale of a single cell [76].

The aim of these new test platforms is to accelerate the discovery of new active molecules, specifically targeting ion channels. This is the critical point, because ion channels display a strong molecular similarity in the 'pore' region and for this reason any small hydrophobic molecule will interact 'easily' with the protein. One must therefore work on molecule libraries, specifying a broad enough range of structures to succeed in finding the one that will specifically modulate the activity of the channel. It is these developments in chemistry which, combined with high-throughput screening, will allow us to improve the identification of quality candidate molecules very early on in the drug discovery process [105].

References

Section One. Microarrays for Analysing and Processing Living Cells

1. Washizu, M.: In: *Proc. of Integrated Micro-Motion Systems* (1989) pp. 417–432
2. Le Pioufle, B., Surbled, P., Chun, H.S., Fujita, H., Nagai, H., Murakami, Y., Tamiya, E.: In: *Proc. of Transducers'99*, Sendai, Japan (1999) pp. 768–771
3. Bailey, S.N., Wu, R.Z., Sabatini, D.M.: *Drug Discovery Today* **7**, S113–S118 (2002)

4. Yoshikawa, T., Uchimura, E., Kishi, M., Funeriu, D.P., Miyake, M., Miyake, J.: *Journal of Controlled Release* **96**, 227–232 (2003)
5. Gradl, G., Müller, T., Pfennig, A., Shirley, S., Schnelle, T., Führ, G.: In: *Proc. of Micro Total Analysis Systems* (2000) pp. 443–446
6. Huang, Y., Rubinsky, B.: *Sensors and Actuators A: Physical* **104**, 205–212 (2003)
7. Huang, Y., Rubinsky, B.: *Sensors and Actuators A* **89**, 242–249 (2001)
8. Khine, M., Lau, A., Ionescu-Zanetti, C., Jeonggi, S., Lee, L.P.: *Lab on a Chip* **5**, 38–43 (2005)
9. Seo, J., Ionescu-Zanetti, C., Diamond, J., Lal, R., Lee, L.P.: *Applied Physics Letters* **84**, 1973–1975 (2004)
10. Seger, U., Gawad, S., Scheer, A., Renaud, P.: In: *Proc. of MicroTAS* (Nara) **2**, 796–798 (2002)
11. Fromherz, P.: *Eur. Biophys. J.* **31**, 228–231 (2002)
12. Steinhoff, G., Baur, B., Wrobel, G., Ingebrandt, S., Offenhausser, A., Dadgar, A., Krost, A., Stutzmann, M., Eickhoff, M.: *Applied Physics Letters* **86**, 033901 (2005)
13. Leclerc, E., Sakai, H., Fujii, T.: In: *Proc. of IEEE-EMBS* (2002) pp. 104–105
14. Tourovskaia, A., Figueroa-Masot, X., Folch, A.: *Lab on a Chip* **5**, 14–19 (2005)
15. Griscom, L., Château, Y., Pennec, J.-P., Misery, L., Le Pioufle, B.: In: *Proc. of IEEE-EMBS*, (2005) pp. 184–187
16. Cui, L., Morgan, H.: *J. Micromech. Microeng.* **10**, 72–79 (2000)
17. Frénéa, M., Lhermite, H., Le Pioufle, B., Fujita, H.: In: *Proc. of IEEE-EMBS (USA)* (2002) pp. 40–143
18. Pohl, H.A., Kaler, K.: *Cell Biophysics* **1**, 15–28 (1979)
19. Wang, X.B., Huang, Y., Burt, J.P.H.: *J. Phys. D: Appl. Phys.* **26**, 1278–1285 (1993)
20. Masuda, S., Washizu, M., Nanba, T.: *IEEE Trans. Industry Applicat.* **25**, 732–737 (1989)
21. Washizu, M., Nanba, T., Masuda, S.: *IEEE Trans. Industry Applicat.* **26**, 352–356 (1990)
22. Fiedler, S., Shirley, S.G., Schnelle, T.: *Anal. Chem.* **70** (9), 1909–1915 (1998)
23. Jones, T.B.: *Electromechanics of Particles*, Cambridge University Press (1995)
24. Gauthier, M., Piat, E.: In: *Proc. of IEEE-EMBS* **2**, 3199–3204 (2002)
25. Fertig, N., George, M., Klau, M., Meyer, C., Tilke, A., Sobotta, C., Blick, R.H., Behrends, J.C.: *Receptors Channels* **9**, 29–40 (2003)
26. Tixier-Mita, A., Mita, Y., Cozic, K., Frenea, M., Le Pioufle, B., Fujita, H.: In: *Proc. of MicroTAS* **2**, 888–890 (2002)
27. Ostuni, E., Chen, C.S., Ingberg, D.E., Whitesides, G.M.: *Langmuir* **17**, 2828–2834 (2001)
28. Revzin, A., Sekin, K., Sin, A., Tompkins, R.G., Toner, M.: *Lab on a Chip* **5**, 14–19 (2005)
29. Ashkin, A., Dziedzic, J.M., Yamane, T.: *Nature* **330**, 769–771 (1987)
30. Fuhr, G.R., Reichle, C.: In: *Proc. of MicroTAS* (Enschede), 261–264 (2000)
31. Wu, R.Z., Bailey, S.N., Sabatini, D.M.: *Trends in Cell Biology* **12**, 485–488 (2002)
32. Galbraith, C.G., Sheetz, M.P.: *PNAS* **94**, 9114–9118 (1997)
33. Pelham, R.J. Jr., Wang, Y.-L.: *Mol. Biol. Cell* **10**, 935–945 (1999)
34. Tan, J.L., Tien, J., Pirone, D.M., Gray, D.S., Bhadriraju, K., Chen, C.S.: *Proc. Natl. Acad. Sci. USA*, **100**, 1484–1489 (2003)

35. Xi, J., Schmidt, J.J., Montemagno, C.D.: *Nature Materials* **4**, 180–184 (2005)
36. Gawad, S., Batard, P., Seger, U., Metz, S., Renaud, P.: In: *Proc. of MicroTAS* (Nara) **2**, 649–651 (2002)
37. Claverol-Tinture, E., Pine, J.: *Journal of Neuroscience Methods* **117**, 13–21 (2002)
38. Denyer, M.C.T., Riehle, M., Britland, S.T., Offenhauser, A.: *Medical and Biological Engineering and Computing* **36**, 638–644 (1998)
39. Degenaar, P., Le Pioufle, B., Griscorn, L., Tixier, A., Murakami, Y., Yokoyama, K., Tamiya, E., Fujita, H.: *Journal of Biochemistry* **130**, 367–376 (2001)
40. Morin, F., Denoual, M., Griscorn, L., Le Pioufle, B., Fujita, H., Tamiya, E.: In: *Proc. of MicroTAS* (Nara), **2**, 515–517 (2002)
41. Pancrazio, J.J., Gray, S.A., Shubin, Y.S., Kulagina, N., Cuttino, D.S., Shaffer, K.M., Eismann, K., Curran, A., Zim, B., Gross, G.W., O'Shaughnessy, T.J.: *Biosensors and Bioelectronics* **18**, 1339–1347 (2003)
42. Fertig, N., Blick, R.H., Behrends, J.C.: *Biophys. J.* **82**, 3056–3062 (2002)
43. Klemic, K.G., Klemic, J.F., Reed, M.A., Sigworth, F.J.: *Biosens. Bioelectron.* **17**, 597–604 (2002)
44. Sigworth, F.J., Klemic, J.F.: *IEEE Transactions on Nanobioscience* **4**, 1536–1241 (2005)
45. Morin, F., Denoual, M., Griscorn, L., Le Pioufle, B., Fujita, H., Tamiya, E.: In: *Proc. of MicroTAS* (Japan) **1**, 578–580 (2002)

Section Two. Patch-Clamp Microarrays

46. Niemeyer, B.A., Mery, L., Zawar, C., Suckow, A., Monje, F., Pardo, L.A., Stuhmer, W., Flockerzi, V., Hoth M.: Ion channels in health and disease, 83rd Boehringer Ingelheim Fonds International Titisee Conference, *EMBO Rep.* **2** (7), 568–573 (2001)
47. Willumsen, N.J., Bech, M., Olesen, S.P., Jensen, B.S., Korsgaard, M.P., Christophersen, P.: High throughput electrophysiology: New perspectives for ion channel drug discovery, *Receptors Channels* **9** (1), 3–12 (2003)
48. Wood, C., Williams, C., Waldron, G.J.: Patch clamping by numbers, *Drug Discovery Today* **9** (10), 434–441 (2004)
49. Hamill, O.P., Marty, A., Neher, E., Sakmann, B., Sigworth, F.J.: Improved patch-clamp techniques for high-resolution current recording from cells and cell-free membrane patches, *Pflugers Arch.* **391** (2), 85–100 (1981)
50. Xu, J., Chen, Y., Li, M.: High-throughput technologies for studying potassium channels: Progress and challenges, *Drug Discovery Today* **3** (1), 32–38 (2004)
51. Wang, X., Li, M.: Automated electrophysiology: High throughput of art, *Assay Drug Dev. Technol.* **1** (5), 695–708 (2003)
52. Owen, D., Silverthorne, A.: Channeling drug discovery, current trends in ion channel discovery research, *Drug Discovery World*, 348–361 (2002)
53. Lepple-Wienhues, A., Ferlinz, K.: Flip the Tapp: An automated, high quality, cost-effective patch-clamp screen, *Receptors and Channels* **9**, 13–17 (2003)
54. Sinclair, J., Pihl, J., Olofsson, J., Karisson, M., Jardebrand, K., Chiu, D.T., Orwar, O.: A cell based bar code reader for high-throughput screening of ion channel–ligand interactions, *Anal. Chem.* **74**, 6133–6138 (2002)
55. Shieh, C.C., Trumbull, J.D., Sarthy, J.F., McKenna, D.G., Parihar, A.S., Zhang, X.F., Faltynek, C.R., Gopalakrishnan, M.: Automated parallel oocyte

- electrophysiology test station (POETs): A screening platform for identification of ligand-gated ion channel modulators, *Assay Drug Dev. Technol.* **1** (5), 655–663 (2003)
56. Trumbull, J.D., Maslana, E.S., McKenna, D.G., Nemcek, T.A., Niforatos, W., Pan, J.Y., Parihar, A.S., Shieh, C.C., Wilkins, J.A., Briggs, C.A., et al.: High throughput electrophysiology using a fully automated, multiplexed recording system, *Receptors Channels* **9** (1), 19–28 (2003)
 57. Andersson, H., Van den Berg, A.: Microfluidics devices for cellomics: A review, *Sensors and Actuators B* **92**, 315–325 (2003)
 58. Schmidt, C., Mayer, M., Vogel, H.: A chip-based biosensor for the functional analysis of single ion channels, *Angew. Chem. Int. Ed. Engl.* **39** (17), 3137–3140 (2000)
 59. Klemic, K.G., Klemic, J.F., Reed, M.A., Sigworth, F.J.: Micromolded PDMS planar electrode allows patch clamp electrical recordings from cells, *Biosens. Bioelectron.* **17** (6–7), 597–604 (2002)
 60. Sigworth, F.J., Klemic, K.G.: Patch clamp on a chip, *Biophys. J.* **82** (6), 2831–2832 (2002)
 61. Xu, J., Guia, A., Rothwarf, D., Huang, M., Sithipong, K., Ouang, J., Tao, G., Wang, X., Wu, L.: A benchmark study with sealchip planar patch-clamp technology, *Assay Drug Dev. Technol.* **1** (5), 675–684 (2003)
 62. Schroeder, K., Neagle, B., Trezise, D.J., Worley, J.: Ionworks HT: A new high-throughput electrophysiology measurement platform, *J. Biomol. Screen.* **8** (1), 50–64 (2003)
 63. Asmild, M., Oswald, N., Krzywkowski, K.M., Friis, S., Jacobsen, R.B., Reuter, D., Taboryski, R., Kutchinsky, J., Vestergaard, R.K., Schroder, R.L., et al.: Upscaling and automation of electrophysiology: Toward high throughput screening in ion channel drug discovery, *Receptors Channels* **9** (1), 49–58 (2003)
 64. Stett, A., Bucher, V., Burkhardt, C., Weber, U., Nisch, W.: Patch-clamping of primary cardiac cells with micro-openings in polyimide films, *Med. Biol. Eng. Comput.* **41** (2), 233–240 (2003)
 65. Stett, A., Burkhardt, C., Weber, U., van Stiphout, P., Knott, T.: Cytocentering: A novel technique enabling automated cell-by-cell patch clamping with the Cytopatch chip, *Receptors Channels* **9** (1), 59–66 (2003)
 66. Bruggemann, A., George, M., Klau, M., Beckler, M., Steindl, J., Behrends, J.C., Fertig, N.: High quality ion channel analysis on a chip with the NPC technology, *Assay Drug Dev. Technol.* **1** (5), 665–673 (2003)
 67. Brueggemann, A., George, M., Klau, M., Beckler, M., Steindl, J., Berhends, J.C., Fertig, N.: Ion channel drug discovery and research: The automated nano-patch-clamp technology, *Current Drug Discovery*, 191–196 (2004)
 68. Kiss, L., Bennett, P.B., Uebele, V.N., Koblan, K.S., Kane, S.A., Neagle, B., Schroeder, K.: High throughput ion-channel pharmacology: Planar-array-based voltage clamp, *Assay Drug Dev. Technol.* **1** (2), 127–135 (2003)
 69. Seo, J., Ionescu-Zanetti, C., Diamond, J., Lal, R., Lee, L.P.: Integrated multiple patch-clamp array chip via lateral cell trapping junctions, *Applied Physics Letters* **84** (11), 1973–1975 (2004)
 70. Ionescu-Zanetti, C., Shaw, R.M., Seo, J., Jan, Y.N., Jan, L.Y., Lee, L.P.: Mammalian electrophysiology on a microfluidic platform, *PNAS* **102** (26), 9112–9117 (2005)

71. Fertig, N., Tilke, A., Blick, R.H., Kotthaus, J.P., Behrends, J.C., ten Bruggen-cate, G.: Stable integration of isolated cell membrane patches in a nanomachined aperture, *Applied Phys. Lett.* **77**, 1218–1221 (2000)
72. Fertig, N., Blick, R.H., Behrends, J.C.: Whole cell patch clamp recording performed on a planar glass chip, *Biophys. J.* **82** (6), 3056–3062 (2002)
73. Fertig, N., George, M., Klau, M., Meyer, C., Tilke, A., Sobotta, C., Blick, R.H., Behrends, J.C.: Microstructured apertures in planar glass substrates for ion channel research, *Receptors Channels* **9** (1), 29–40 (2003)
74. Picollet-D'Hahan, N., Sordel, T., Garnier-Raveaud, S., Sauter, F., Ricoul, F., Pudda, C., Marcel, F., Chatelain, F.: A silicon-based 'multi-patch' device for ion channel current sensing, *Sensor Letters* **2** (2), 91–94 (2004)
75. Sordel, T., Garnier-Raveaud, S., Sauter, F., Pudda, C., Marcel, F., De Waard, M., Arnoult, C., Vivaudou, M., Chatelain, F., Picollet-D'Hahan N.: Hour-glass SiO₂ coating increases the performance of planar patch-clamp, *J. of Biotechnology* **125**, 142–154 (2006)
76. Pantoja, R., Nagaraj, J.M., Starace, D.M., Melosh, N.A., Blunck, R., Bezanilla, F., Heath, J.R.: Silicon chip-based patch-clamp electrodes integrated with PDMS microfluidics, *Biosens. Bioelectron.* **20** (3), 509–517 (2004)
77. Thielecke, H., Stieglitz, T., Beutel, H., Matthies, T., Ruf, H.H., Meyer, J.U.: Fast and precise positioning of single cells on planar electrode substrates, *IEEE Eng. Med. Biol. Mag.* **18** (6), 48–52 (1999)
78. Cheung, K.C., Kubow, T., Lee, L.P.: Individually addressable planar patch clamp array, *IEEE Eng. Med. Biol. Mag.* (2002)
79. WO 02/077259 A2
80. Lehnert, T., Gijs, M.A.M., Netzer, R., Bischoff, U.: Realization of hollow SiO₂ micronozzles for electrical measurements on living cells, *Applied Physics Letters* **81** (26), 5063–5065 (2002)
81. Kubow, T.M., Cheung, K.C., Bentley, L.F., Lee L.P.: Microfabricated patch-clamp array for neural MEMS application, *Materials Research Society*, 729197–729202 (2002)
82. Jenkner, M., Muller, B., Fromherz, P.: Interfacing a silicon chip to pairs of snail neurons connected by electrical synapses, *Biol. Cybern.* **84** (4), 239–249 (2001)
83. Matthews, B., Judy, J.W.: Characterization of a micromachined planar patch-clamp for cellular electrophysiology, *IEEE-EMBS International Conference on Neural Engineering*, 648–651 (2003)
84. Tao, H., Guia, A., Sithipong, K., Miu, P., Ligutti, J., Wu, L., Xu, J.: Cell isolation for high quality electrophysiological measurements on planar electrodes (2004)
85. Straub, B., Meyer, E., Fromherz, P.: Recombinant maxi-K channels on transistor, a prototype of iono-electronic interfacing, *Nat. Biotechnol.* **19** (2), 121–124 (2001)
86. Kaul, R.A., Syed, N.I., Fromherz, P.: Neuron–semiconductor chip with chemical synapse between identified neurons, *Phys. Rev. Lett.* **92** (3), 038102 (2004)
87. Braun, D., Fromherz, P.: Imaging neuronal seal resistance on silicon chip using fluorescent voltage-sensitive dye, *Biophys. J.* **87** (2), 1351–1359 (2004)
88. Hutzler, M., Fromherz, P.: Silicon chip with capacitors and transistors for interfacing organotypic brain slice of rat hippocampus, *Eur. J. Neurosci.* **19** (8), 2231–2238 (2004)

89. Heuschkel, M.O., Guerin, L., Buisson, B., Bertrand, D., Renaud, P.: Buried microchannels in photopolymer for delivering of solutions to neurons in a network, *Sensors and Actuators B: Chemical* **48**, 356–361 (1998)
90. Heuschkel, M.O., Fejtl, M., Raggenbass, M., Bertrand, D., Renaud, P.: A three-dimensional multi-electrode array for multi-site stimulation and recording in acyte brain slices, *J. of Neurosciences Meth.* **114**, 135–148 (2002)
91. Stett, A., et al.: Biological applications of microelectrode arrays in drug discovery and basic research, *Anal. and Bioanal. Chem.* **377**, 486–495 (2003)
92. Brischwein, M., et al.: Functional cellular assays with multiparametric silicon chips, *Lab on a Chip* **3**, 2324–2400 (2004)
93. Comley, J.: Patchers V screeners: Divergent opinion on high throughput electrophysiology, *Drug Discovery World* (2003)
94. Chien, K.: Twenty questions: The state of ion channel research in 2004, *Nature Reviews: Drug Discovery* 3239–3278 (2003)
95. Netzer, R.: The new generation of patch-clamp equipment, *DDT* 9, **14**, 593 (2004)
96. www.ornl.gov/sci/techresources/Human_Genome/home.shtml
97. Venter, J.C., et al.: The sequence of the human genome, *Science* **291**, 1304–1351 (2001)
98. Worley, J.F., Main, M.J.: An industrial perspective on utilizing functional ion channel assays for HTS, *Receptors and Channels* **8**, 269–282 (2002)
99. Sigworth, F.J., Klemic, K.G.: Microchip technology in ion-channel research, *IEEE Transactions in Nanobioscience* 4, **1**, 121–127 (2005)
100. Gorelik, J., Gu, Y., Spohr, H.A., Shevchuk, A.I., et al.: Ion channels in small cells and subcellular structures can be studied with a smart patch-clamp system, *Biophys. J.* 83, **6**, 3296–3303 (2002)
101. Gullo, M.R., Akiyama, T., Fredreix, P.L., Tonin, A., Staufer, U., Engel, A., de Rooij, N.F.: Towards a planar sample support for in situ experiments in structural biology, *Microelectronic. Eng.* **78–79**, 571–574 (2005)
102. McDowell, M., Gray, E.: An optimal cell detection technique for automated patch clamping, *NASA/TM*, 1–8 (2004)
103. Pandey, S., Mehrotra, R., Wykosky, S., White, M.H.: Characterization of a MEMS BioChip for planar patch-clamp recording, *Solid-State Electronics* **48** (10–11), 2061–2066 (2004)
104. Cornell, B.A., Braach-Maksvytis, V.L.B., King, L.G., Osman, P.D.J., Raguse, B., Wieczorek, L., Pace, R.J.: A biosensor that uses ion-channel switches, *Nature* **387**, 580–583 (1997)
105. Lachnit, W., Xie, M., Curtis, R., Castle, N.: Drug discovery technology for ion channels, *DDT* **6** (12), S17–S18

Lab on a Chip

P. Puget

20.1 The General Idea

The reliable and fast detection of chemical or biological molecules, or the measurement of their concentrations in a sample, are key problems in many fields such as environmental analysis, medical diagnosis, or the food industry. There are traditionally two approaches to this problem. The first aims to carry out a measurement in situ in the sample using chemical and biological sensors. The constraints imposed by detection limits, specificity, and in some cases stability are entirely imputed to the sensor. The second approach uses so-called total analysis systems to process the sample according to a protocol made up of different steps, such as extractions, purifications, concentrations, and a final detection stage. The latter is made in better conditions than with the first approach, which may justify the greater complexity of the process. It is this approach that is implemented in most methods for identifying pathogens, whether they be in biological samples (especially for in vitro diagnosis) or samples taken from the environment. The instrumentation traditionally used to carry out these protocols comprises a set of bulky benchtop apparatus, which needs to be plugged into the mains in order to function. However, there are many specific applications (to be discussed in this chapter) for which analysis instruments with the following characteristics are needed:

- Possibility of use outside the laboratory, i.e., instruments as small as possible, consuming little energy, and largely insensitive to external conditions of temperature, humidity, vibrations, and so on.
- Possibility of use by non-specialised agents, or even unmanned operation.
- Possibility of handling a large number of samples in a limited time, typically for high-throughput screening applications.
- Possibility of handling small samples.

At the same time, a high level of performance is required, in particular in terms of (1) the detection limit, which must be as low as possible, (2) specificity,

i.e., the ability to detect a particular molecule in a complex mixture, and (3) speed.

Faced with these needs, over the past thirty years, the development of microtechnology has led to extraordinary progress in electronics, with improved performance at lower cost. The characteristic of microtechnology is that objects consisting of motifs with typical dimensions in the micrometer range are fabricated in large numbers on a slice of material called a wafer, thereby cutting the cost of a number of fabrication stages by several orders of magnitude. Apart from the basic microelectronics of integrated circuits, these fabrication processes have been implemented for other applications, e.g., to make miniaturised devices or microsystems, which constitute realistic technological solutions to meet the above requirements for measurement instruments. The contributions of microtechnology to chemical and biological analysis are many and varied. Some examples are as follows, although the list is probably not exhaustive:

- Mass assembly and machining for fabricating objects with microscale motifs or smaller.
- The principles of electromechanical or mechanical actuation, e.g., electrostatic or piezoelectric.
- Optoelectronic functions, either detectors or sources (useful for all detection based on fluorescence or chemiluminescence).
- Other transduction methods, e.g., magnetic (with giant magnetoresistances, able to detect the presence of magnetic micro- or nanoparticles traditionally used in medical and biological analysis) or electrochemical.
- The possibility of making digital or analog electronic functions for signal acquisition or processing.

It was in this context, with the need for new measurement systems and at the same time a flood of innovations from microtechnology, that the idea of a lab on a chip or micro total analysis system (μ TAS) was proposed for the first time by Andreas Manz in 1990 [1]. Since then, the lab on a chip has come to mean a microsystem implanting one or more steps in a chemical or biological measurement protocol. This area developed enormously in the 1990s, in terms of the number of actors, scientific publications, patents, companies, and commercial products. There are several very good review articles in the literature [2–5], and some specialised works [6], where the interested reader will find more detail to supplement the present discussion.

Over the last ten years, the areas that have benefited most from the development of the lab on a chip are mainly in the life sciences. In particular, in the field of *in vitro* diagnosis, analysis is prescribed by a doctor and carried out by a laboratory (especially in Europe). In some cases, the sample is taken in one place and analysed in another. This means that there is a significant time lapse between the prescription of the test and the return of the result. Furthermore, it implies a complex organisation to manage samples and results, and this is not only expensive but can lead to medical error. There

is thus a need for rapid tests allowing the doctor to take almost immediate decisions, whether for a standard consultation, an emergency, or during an operation. Such diagnostic systems are called point-of-care systems [7]. This diagnostic requirement transposes almost immediately to veterinary medicine, where there is a similar need for fast, easy-to-use analysis systems, e.g., in the stable. Concerning health and safety requirements in the food industry, detection systems are needed to look for contaminating pathogens [8], while regulatory restrictions on information regarding product composition require the detection of genetically modified organisms [9].

In environmental control, whether it be analysis of the air [10] or water [11], there is a need for high-performance portable systems to measure the level of industrial pollutants, but also to detect biological and chemical threats in the defence against bioterrorism [12, 13]. Space research, especially in the search for extraterrestrial life forms, has financed the development of automated systems to look for traces of amino acids in dust on the surface of Mars [14].

Finally, the lab on a chip is useful for high-throughput screening, mainly in the pharmaceutical industry, where each chemical or biological test has to be repeated many times while varying the input reagents. In this case, portability becomes less important in comparison with

- the ability to carry out tests in parallel so as to minimise the time required for a test campaign,
- the possibility of reducing the unit cost by decreasing the amount of reagent needed.

In contrast to the applications mentioned previously, these devices are not single use, but can generally be reused for a large number of tests.

20.2 Implanted Functions

As discussed above, a lab on a chip carries out one or more operations of a measurement protocol. In the case of a biological analysis, a complete protocol generally consists of three main operations (see Fig. 20.1):

- The critical stage is molecular recognition, where the analyte in question, or target, usually a protein or fragment of nucleic acid, is reacted with a known molecule present in the system, called the probe, capable of reacting specifically with the target. This probe can be an antibody (in immunoassays), another protein, or a DNA sequence complementary to the target nucleic acid.
- Analyte extraction aims to transfer the target analyte from a matrix or gaseous, liquid, or solid initial state into some suitable buffer solution.
- Sample preparation aims to bring the target analyte into optimal physicochemical conditions for molecular recognition: concentration, interferent elimination, ionic strength, pH, temperature, addition of other chemical

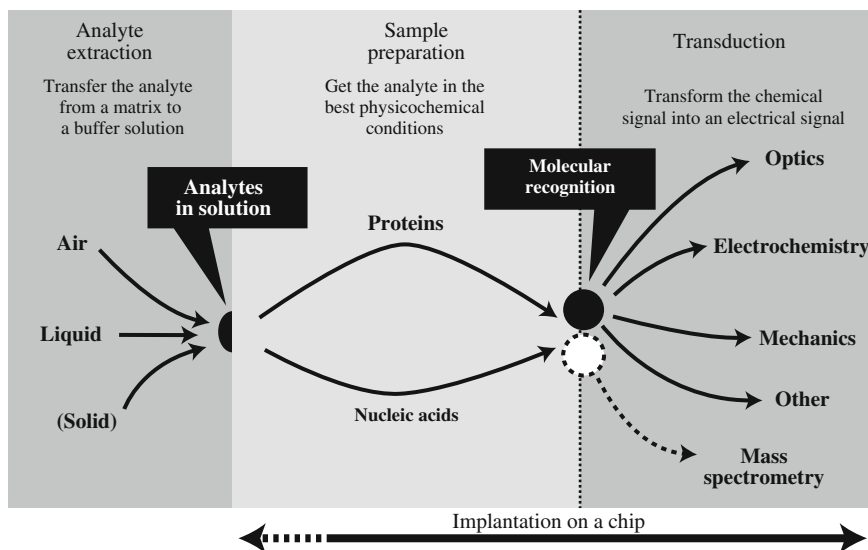


Fig. 20.1. Main steps in a chemical or biological measurement protocol

compounds, etc. In the particular case where the analyte is a nucleic acid, sample preparation may involve an amplification stage, usually by polymerase chain reaction (PCR) [15].

In the general case, going beyond biological analysis, implanted protocols may differ from the above scheme, but the idea remains similar and the basic operations discussed below are relevant.

The diagram in Fig. 20.1 applies to molecular analyses. There are other labs on chips specifically designed to handle cells [16,17]. These were discussed in Chap. 19 and will not be mentioned further here.

20.2.1 Sample Preparation

Analyte extraction and sample preparation are generally achieved by well known methods of analytical chemistry which have simply been miniaturised with the help of microtechnology. Several thorough review articles focus on these operations. In particular, [18] is the most general and systematic, illustrating with many examples, while [19] pays more attention to the underlying physicochemical principles of thermodynamics and kinetics. In this section, we shall see how the main functions relating to sample preparation can be implemented on a chip.

Filtering

Filtering aims to eliminate debris or solid particles present in a solution. Filters can be inserted into a lab on a chip by various means. The simplest in principle

is to etch pillars with a smaller spacing than the diameter of the particles to be retained. An example of this kind of system is described in [20]. Figure 20.2 shows the details of such a filter, designed to retain particles. Using a slightly different approach, but paying the price of a more sophisticated technology, the filter can be a membrane in the plane of the substrate [21].

The most commonly mentioned limitation of these two approaches is that the efficiency of the filter depends closely on the spatial resolution with which the structures are realised. It is the characteristics of the fabrication process, possibly including both lithography and forming (etching, moulding, or some other kind of machining), which determine the spatial resolution finally obtained. Commonly used processes can achieve a spatial resolution in the micrometer range, which fixes the minimal size of the particles that can be filtered out by these filters at something like $1.5\text{ }\mu\text{m}$. Typical fabrication methods can achieve much better controlled spatial resolution in the direction normal to the plane of the substrate, i.e., regarding the thicknesses of deposited layers. By making sacrificial layers, the minimal size of particles that can be removed by filtering can thus be reduced. This is exemplified in [22]. The filter made in this way obtains a very good pore reproducibility, with pore diameters around 25 nm.

Separation by Transport Phenomena

While the last approach is based on a mechanical principle, molecules or nanoparticles can also be separated from larger entities by exploiting other phenomena that do not require the microfluidic system to be set in contact with the solid fractions of the sample, thus avoiding the risk of clogging up the filter, for example.

A first idea uses discrimination by diffusion in a laminar flow [23]. Indeed, any flow in a submillimetric capillary at a speed less than one centimeter per second is such that the flow will be laminar and completely free of turbulence (see Chap. 12). In an H-shaped capillary system like the one in Fig. 20.3, the two flows entering the system only mix by diffusion in the central section. As a consequence, a particle introduced by the upper arm will have a greater chance of coming out via the lower arm if it has a high diffusion constant. More precisely, the average times taken by two objects (molecules or particles) to cover the same length L are inversely proportional to the ratio of their respective diffusion constants. For a protein in pure water at room temperature, the diffusion constant is of the order of $5 \times 10^{-11}\text{ m}^2\text{s}^{-1}$, for a single-strand DNA comprising 20 bases, it is $15 \times 10^{-11}\text{ m}^2\text{s}^{-1}$, and for a sphere of diameter $1\text{ }\mu\text{m}$, it is $10^{-13}\text{ m}^2\text{s}^{-1}$, i.e., factors of 500 and 150 times smaller than for the previous molecules. Flow conditions can thus be created such that, at the outlet of the device in Fig. 20.3, the concentration of light fractions is almost uniform throughout the whole capillary, while the bigger particles have hardly begun to diffuse into the lower part of the capillary system. It is then possible to collect part of the sample freed from these bigger particles.

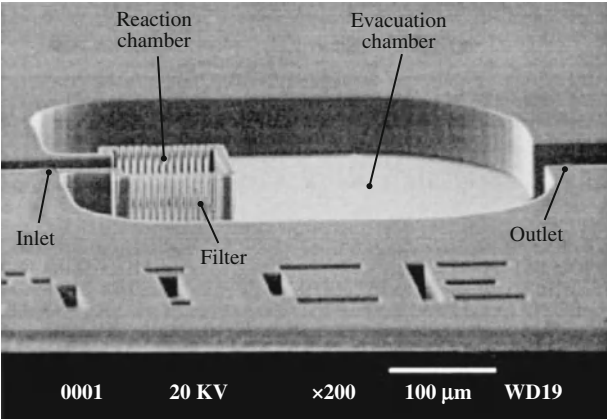


Fig. 20.2. Pillar structures for filtering beads in a sample [20]

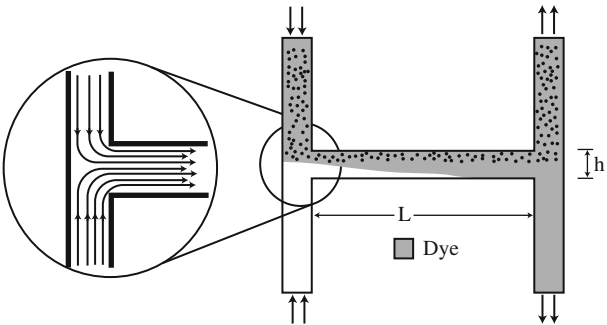


Fig. 20.3. Separation by diffusion in an H-shaped capillary system [23]

This idea has also been exploited to carry out an immunological test [24]. In this work, the authors use the different diffusion constants of an antigen (light) and an antibody, free or bound to a particle. They show that the profile of the concentration gradient between the veins of antibody and antigen solutions is not only detectable, but also depends on the concentration of antigens, whence it is possible to carry out an immunological test.

An active variant of this principle consists in combining the flow with an electrostatic attraction on the entities to be separated [25]. The electric field is perpendicular to the flow axis. The particles are thus deflected from their straight trajectory through an angle that depends on the electrophoretic mobility of the molecules. This separation method, invented originally in macroscopic configurations, has been implanted in microfluidic devices.

Solid Phase Extraction

The idea of affinity chromatography is to retain the relevant analytes present in a solution on the surface of a solid phase, while allowing unwanted

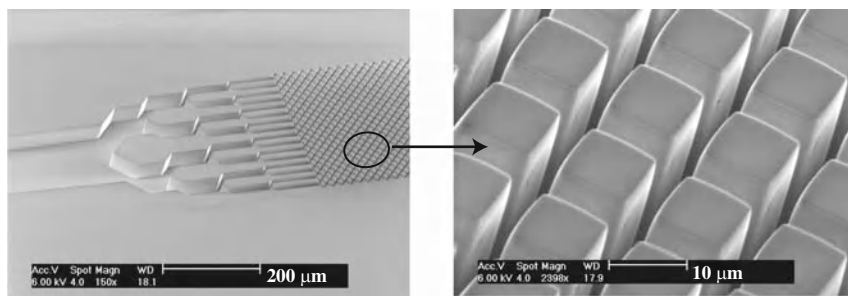


Fig. 20.4. Chromatography microcolumn with a pillar structure functionalised with C18 [27]

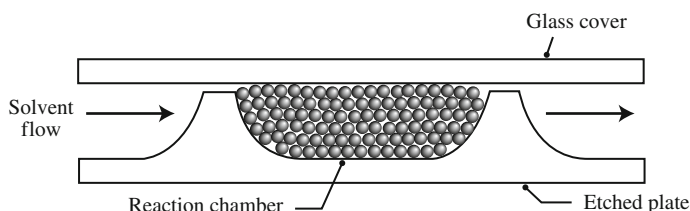


Fig. 20.5. Functionalised beads immobilised in a reaction chamber [28]

compounds or solvents to pass by unaffected. In a second stage, the device is rinsed or eluted with a small amount of buffer within which the analytes exist in a purified and generally more concentrated state. In the case of liquid chromatography, the solid phase interacts with the analytes present in the sample in such a way that their transit times in the column differ from one another, whence they may be separated. These two ideas have been successfully implanted in labs on chips. Several different approaches have been put forward. They have the common goal of maximising the interactions between the molecules present in the solution and the solid phase constituting the device. The different steady state phases and materials are compared in [26].

In the first approach, the solid phase is shaped by etching the material. The most typical structure is a forest of pillars through which the sample is made to flow. The component shown in Fig. 20.4 has this structure, functionalised with C18 to make a liquid chromatography column [27].

In the second approach, previously functionalised beads are immobilised in a reaction chamber. For the case described in [28], the authors etched a reaction chamber of volume 330 picolitres, i.e., an area of side about 200 μm for a depth of 10 μm , bounded by 'walls' leaving an interstice of only 1 μm between the bottom of the component and its lid, thereby blocking the beads within (see Fig. 20.5). The beads are introduced into the device via a side channel.

In the third approach, proposed around 2000 by Jean Fréchet and coworkers at Berkeley in the USA, a porous polymer plays the role of the solid phase in the component. Once the capillaries have been made, they are filled with a solution of monomers which are then polymerised by UV exposure. The choice of monomer allows one to design a porous polymer with well defined features, i.e., with regard to hydrophobicity, surface charge, various functionalisations, etc. The first studies concerned silica or glass labs on chips [29], but were then extended to polymers [30].

Comparing the three approaches, the surface-to-volume ratios that can be obtained are, according to [31], less than 10^6 m^{-1} for functionalised structures, of the order of 10^7 m^{-1} for immobilised functionalised beads, and between 10^6 and 10^7 m^{-1} for porous polymers.

Magnetic Particles

The above filtering and chromatographic approaches are based on the use of an immobilised solid phase. It is also possible to use a fractionated solid phase in the form of particles with sizes in the micrometer range or smaller. These particles are usually functionalised. Compared with an immobilised solid phase, they have the advantage of remaining in suspension, provided that they are small enough (micrometer or less). Chemical reactions between the analytes in the buffer and the surface of the particles are therefore faster, especially if the liquid is stirred. However, if one wishes to rinse these particles and elute the analytes as in the case of affinity chromatography, it is also necessary to bind the particles from time to time. This is why magnetic particles are used. In the relevant size range, they are superparamagnetic, which means that they can sediment out under the effect of a magnetic field and lose all magnetisation, whence they will redisperse once again when the field is removed. The action on these beads can be achieved by external permanent magnets [32] or integrated microcoils [33].

Liquid–Liquid Extraction

Liquid–liquid extraction is a commonly used process in chemical engineering. Two immiscible liquid phases, e.g., an aqueous solution and an organic solvent, are set in contact in an H structure of the kind shown in Fig. 20.3. Molecules present in the first phase and having greater affinity for the second will thus move from the first to the second and are thereby extracted. Once again, the extraction yield will depend on geometric features of the contact zone and the diffusion constants of the relevant molecules in the liquids. In particular, size reduction decreases the diffusion time and increases the surface-to-volume ratio, resulting in a very large increase in the extraction efficiency [34].

Polymerase Chain Reaction (PCR)

For nucleic acids, amplification, especially by PCR, is one of the key steps in the analysis. Indeed, once the sample has been amplified, detection is simplified by the fact that it no longer needs to be so sensitive. The particularity of PCR from the point of view of its implementation is the need to subject the sample to a series of heating cycles through 2 or 3 temperature levels, and this around thirty times. So the implantation of PCR within a lab on a chip, apart from the usual issues of storing the reagents, mixing them together, and ensuring their biocompatibility, represents a tough problem of thermal engineering. There are several approaches for subjecting the sample to heat cycles. The first is naturally to heat and cool either the whole device, or just the part where the sample is located. Another solution is to create several regions of different temperatures within the lab on a chip, these being the temperature levels required for the PCR process. The sample is then subjected to the appropriate heat cycles by circulating it in the correct order through the different regions [35].

There are also several ways of heating and cooling a device. Contact with a thermostatically controlled body (which can use the Peltier effect) is the most common solution, for both heating and cooling. Heating by the Joule effect is clearly feasible and usually easy to integrate. There are also some more novel approaches. One example is illumination by infrared radiation [36], and another the use of endo- or exothermic chemical reactions [37]. For further details, the reader is referred to [38].

20.2.2 Transduction

The problem of transduction in a lab on a chip lies in the generation of a measurable signal (usually optical or electric) which reflects whether a specific chemical reaction has taken place or not. The different modes of transduction available are as follows [39]:

- *Optical.* This concerns cases where the reaction generates photons (chemiluminescence), the colour of the reaction medium changes, fluorescent labels are activated or accumulate, or the optical response of an optical interface is modified.
- *Mechanical.* Here a hybridisation reaction at the surface of a material perturbs characteristics such as the damping factor or resonance frequency of a quartz oscillator (quartz microbalance or QMB), the frequency of a surface acoustic wave (SAW), or the deflection of a microcantilever [40].
- *Electrical.* This is used when the chemical reaction can modify an electrochemical potential at an electrode or an impedance between two electrodes, or create a field effect at the gate of a transistor [41, 42].
- *Magnetic.* Here the target molecule is labelled by a submicrometer magnetic particle that can be detected by a magnetic sensor integrated into the microsystem [43].

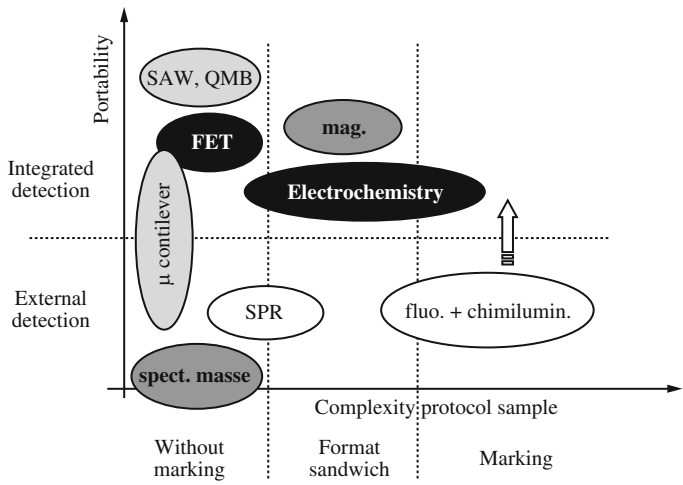


Fig. 20.6. Comparison between different detection methods. Optical, electrical, mechanical, and magnetic methods are shown in white, black, light grey, and dark grey, respectively

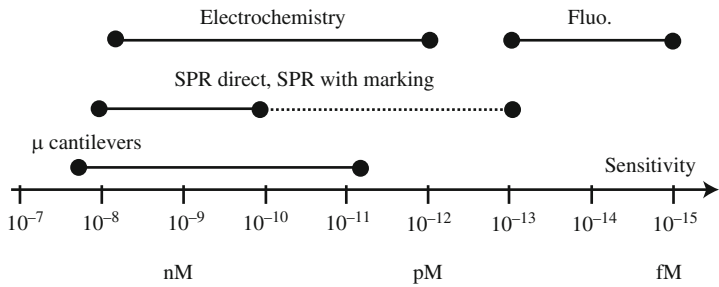


Fig. 20.7. Comparing the sensitivity of different detection methods. Values are taken from the literature and are not the result of a strict comparison on a single biological model

The detection system is chosen with respect to several criteria regarding performance, cost, and portability, whose relative importance will depend on the application. To weigh up the pros and cons of the various methods, they can be plotted on a 2D graph on which one dimension represents the complexity imposed by detection on the sample preparation protocol. The least favourable case is the one where the detection principle requires chemical tagging of the molecule to be detected. This extra step would lead to increased complexity of the relevant protocol and a consequent increase in complexity of the resulting device. The other axis in the diagram must reflect the potential portability of the read system. The two extremes here are a read system that can be completely integrated into the microsystem on the one hand, and a full-scale read system, necessarily in the form of a benchtop instrument, on the other. In the

graph of Fig. 20.6, it is clear that fluorescence methods, currently the most widespread, are also the most handicapping for this kind of system, in terms of both sample preparation and portability. Conversely, mechanical methods would appear to be the most interesting with regard to these two points.

However, when one considers the performance of these methods, especially in terms of sensitivity (see Fig. 20.7), it is clear that fluorescence is considerably more efficient than the competing methods cited above. This is why today the competition between methods remains very open. Other methods using functionalised nanostructures such as nanowires [44, 45] or biomimetic membranes functionalised by membrane proteins [46, 47], allowing single-molecule detection without labelling, seem to be the way of the future.

A particular detection method is mass spectrometry, which is unequalled in terms of sensitivity and the ability to discriminate. This is used to detect both nucleic acids [48] and proteins [49]. We consider this an exceptional case, because it looks likely that it will only ever be used as a full scale instrument, impossible to integrate either totally or in part within any microsystem.

20.3 Technological Aspects

Having considered the lab on a chip from a functional point of view, it is also useful to view it from a technological angle. In this section, we discuss the materials and the various technologies used to form, assemble, and functionalise the surfaces of these chips. Once again, the reader is referred to the literature for further details [50, 51].

The choice of material and fabrication process result from optimisation with respect to a set of criteria, taking into account the functions and expected performance of the device, its cost, and the volume of production. The following must be considered:

- Chemical properties (for biocompatibility and chemical functionalisation of surfaces), thermal properties, and in some cases optical and electrical properties (mainly for detection purposes) of the materials.
- Processes used for moulding/machining, assembly, chemical surface functionalisation, and packaging, which must be mutually compatible and compatible with the relevant biological material.
- The total production cost of the device, taking into account the cost of raw materials and the cost of the processes. The unit production cost may of course depend heavily on the volume of production.

The materials most commonly used are polymers in various forms [52], glass, quartz, and silicon [53]. The advantage with polymers is their cost and ease of implementation, especially with regard to shaping. Silicon has the advantage that electrical or electronic functions are easy to implant, e.g., for detection. Furthermore, it is a good heat conductor and micrometric structures can be made with well understood processes [54, 55]. However, its main disadvantage

Table 20.1. Technology and main operations required to fabricate a lab on a chip for different commonly used materials. DRIE stands for deep reactive ion etching

Step	Glass	Si	SU8	PDMS	
				elastomer	PMMA, PC
Shaping	Wet etch, DRIE, moulding	Photolithography + etch		Casting	Die stamping, moulding, machining
Surface treatment, functionalisation	Silanisation	SiO ₂ + id, glass		Plasma activation	UV activation
Assembly and packaging	High temperature processes			Low temperature bonding processes	

is the relatively high cost of fabricating objects even of modest area (anything beyond a few mm²), if the objects made are not mass produced. The main properties of the various materials are summarised in Table 20.2 along with their strengths and weaknesses.

With regard to the processes involved, a complete production unit for fabricating labs on chips must necessarily include the following steps:

- *Shaping the Material.* The material has to be transformed from the raw format to an object with the required shape and dimensions. The starting point may be a wafer or a slab in the case of inorganic materials, or a volume, a film, a slab, particles, or a liquid in the case of polymers.
- *Surface Treatment and Functionalisation.* Here molecules or biological probes are grafted onto the surface to carry out molecular recognition, e.g., as in the case of DNA chips (see Chap. 17).
- *Assembly and Packaging.* This concerns the assembly of the various parts or subsystems making up the full lab on a chip. A standard operation is the assembly of a plane cover on some lower part containing open reaction chambers and channels, etched in some way depending on the material.

Table 20.1 lists the technology associated with each step for commonly used materials. Information concerning surface treatment and functionalisation already shown in Table 20.2 is not repeated here.

Inorganic Materials

For shaping and assembly, glass and silicon processes are usually adapted directly from those used to fabricate micro electromechanical systems (MEMS) [53–55]. Structures are produced by etching patterns defined by a mask of photoresist. The latter is deposited on the surface, exposed, and then developed. This yields excellent spatial resolution (much better than micrometric) and reproducibility. With regard to assembly and packaging, processes traditionally used for MEMS are high-temperature processes, which raises a problem

of compatibility for biological applications. Biological material, e.g., molecular probes, is destroyed at these temperatures. One thus has the choice of functionalising the device after assembly or using low-temperature bonding methods borrowed from polymer processing technology.

Polymer Materials

There are many polymer technologies, especially for shaping and structuring. The epoxy resin SU8 is photosensitive and can be worked by common white room techniques, as seen earlier. It is more and more often used as a structural material today.

However, most polymers are shaped by replication techniques using a metal or silicon mould. Elastomers and in particular PDMS are shaped by polymerisation of the liquid form previously introduced into a mould (casting). This technique, invented by Whitesides and coworkers at Harvard (USA), can be implemented without sophisticated equipment. It is particularly well suited to making prototypes [56]. Embossing techniques, and in particular hot embossing techniques, involve impressing a mould in a thermoplastic polymer softened by raising its temperature [57]. Injection moulding techniques are also used for certain materials such as polycarbonate [58].

Finally, polymers can also be worked by ablation techniques. One option is photoablation, using a UV excimer laser to illuminate the material through a mask. Another, more interesting for mass production, is plasma etching of a polyimide sheet coated with copper [59].

Microfluidics

Microfluidics, the art of moving, storing, and distributing liquid samples and reagents in a miniaturised device, typically a lab on a chip, is a key technology for designing and making such a system, in particular when the problem at hand involves running together several steps in some experimental protocol. A whole chapter of this book is devoted to this topic, and the reader is thus referred to Chap. 12.

20.4 Conclusion

The research and development of labs on chips is an extremely active field. In fifteen years or so, many elementary functions have been studied and implanted in miniaturised devices. These functions concern in particular molecular recognition and transduction, as well as certain processes relating to sample preparation. For each elementary function, several different approaches have been proposed, and it is not always easy, as we have seen for example in the particular case of detection methods, to decide which one is better than the other.

Table 20.2. Comparing the properties of materials used to make labs on chips

Function	Glass	Silicon	Polymer	
			SU8 photoresist	PDMS elastomer
				PMMA, polycarbonate
Structure/shaping	Good, cheap structural material, but more difficult to shape than silicon	Expensive material, but structuring processes are well understood, reliable, and mature as far as industry is concerned. In the case of mass production, costs are relatively low	The raw material is relatively cheap, making polymers the choice material for large devices (in the centimeter range). Shaping processes for structures above a few tens of micrometers are well understood and technologically mature. However, post-processing procedures, e.g., implantation of electrical functions, packaging, are relatively expensive	
Optical	Excellent	Unsuitable		Very good
Electrical		Electronic functions relating to transduction and signal processing are easily implanted on the chip	Difficult today, although research is underway to achieve electronic functions compatible with polymer substrates	
Thermal		Very good heat conductor. Interesting for thermal cycling as in PCR (see Chap. 15)	Poor heat conduction	
Functionalisation	Glass surface chemistry is well understood. Biological molecules are easily grafted onto glass via a self-ordered intermediate layer of silane	Silicon is naturally coated with amorphous oxide which can be functionalised like glass. Recently, ways have been devised to graft molecules covalently directly onto silicon	Surface properties are not always easy to control. Until recently, functionalisation procedures were less well understood than for inorganic materials, but methods here should eventually catch up	

The integration of a complete analysis protocol remains a major challenge, and few examples of such have yet seen the light of day, even at the proof-of-concept level. Among the more remarkable, one could cite [60–62], as well as the various achievements of Steve Quake and coworkers, e.g., [63, 64]. In order to run together the various steps, a complete mastery of microfluidics is essential, and this remains a difficult point.

Finally, from the point of view of industrialisation, a certain number of products have reached this stage over the past few years. The most popular is the Bioanalyser made by Agilent. More recently, this company has also commercialised a liquid chromatography microcolumn. These products only carry out one specific function, and cannot yet handle a complex protocol. Other companies, usually start-up companies, such as Fluidigm resulting from the work of Steve Quake and his team, offer more complex products designed for some specific application. The particular technology used would not allow them, at least in the short term, to be used as tools in the field. The mass production of labs on chips devoted to a particular application is not due for a few years yet and will not happen until there is a mass market for these products, which may be for in vitro diagnosis, health and safety in the food industry, or environmental control.

References

1. Manz, A., Graber, N., Widmer, H.M.: Miniaturized total chemical analysis systems: A novel concept for chemical sensing, *Sensors and Actuators B: Chemical* **1** (1–6), 244–248 (1990)
2. Bange, A., Halsall, H.B., Heineman, W.R.: Microfluidic immunosensor systems, *Biosensors and Bioelectronics* **20**, 2488–2503 (2005)
3. Vilkner, T., Janasek, D., Manz, A.: Micro total analysis systems. Recent developments, *Analytical Chemistry* **76** (12), 3373 (2004)
4. Reyes, D.R., et al.: Micro total analysis systems. 1. Introduction, Theory, and Technology, *Analytical Chemistry* **74** (12), 2623–2636 (2002)
5. Auroux, P.-A., et al.: Micro total analysis systems. 2. Analytical standard operations and applications, *Analytical Chemistry* **74** (12), 2637–2652 (2002)
6. Oosterbroek, R.E., van den Berg, A.: *Lab-on-a-Chip, Miniaturized Systems for (Bio)Chemical Analysis and Synthesis*, ed. by R.E. Oosterbroek and A. van den Berg, Elsevier (2003) p. 394
7. Tüdös, A.J., Besselink, G.A.J., Schasfoort, R.B.M.: Trends in miniaturized total analysis systems for point-of-care testing in clinical chemistry, *Lab on a Chip* **1** (2), 83–95 (2001)
8. Hobson, N.S., Tothill, I., Turner, A.P.F.: Microbial detection, *Biosensors and Bioelectronics* **11** (5), 455–477 (1996)
9. Ahmed, F.E.: Detection of genetically modified organisms in foods, *Trends in Biotechnology* **20** (5), 215–223 (2002)
10. Koester, C.J., Simonich, S.L., Esser, B.K.: Environmental Analysis, *Analytical Chemistry* **75**, 2813–2829 (2003)

11. Richardson, S.D.: Water analysis: Emerging contaminants and current issues, *Analytical Chemistry* **75** (12), 2831–2857 (2003)
12. Iqbal, S.S., et al.: A review of molecular recognition technologies for detection of biological threat agents, *Biosensors and Bioelectronics* **15** (11–12), 549–578 (2000)
13. Wang, J.: Microchip devices for detecting terrorist weapons, *Analytica Chimica Acta* **507** (1) 3–10 (2004)
14. Skelley, A.M., et al. Development and evaluation of a microdevice for amino acid biomarker detection and analysis on Mars, *P.N.A.S. USA* **102** (4), 1041–1046 (2005)
15. White, T.J.: The future of PRC technology: Diversification of technologies and applications, *Trends in Biotechnology* **14** (12), 478–483 (1996)
16. Le Pioufle, B., Frenea, M., Tixier, A.: Biopuces pour le traitement de cellules vivantes: Micromanipulation des cellules par voie électrique ou microfluidique, *Comptes Rendus Physique* **5** (5), 589–596 (2004)
17. El-Ali, J., Sorger, P.K., Jensen, K.F.: Cells on chips, *Nature* **442** (7101), 403 (2006)
18. Lichtenberg, J., de Rooij, N.F., Verpoorte, E.: Sample pretreatment on micro-fabricated devices, *Talanta* **56** (2), 233–266 (2002)
19. Pawliszyn, J.: Sample preparation: Quo Vadis?, *Analytical Chemistry* **75** (11), 2543–2558 (2003)
20. Andersson, H., et al.: Micromachined flow-through filter-chamber for chemical reactions on beads, *Sensors and Actuators B: Chemical* **67** (1–2), 203–208 (2000)
21. Xing, X., et al.: Micromachined membrane particle filters, *Sensors and Actuators A: Physical* **73** (1–2), 184–191 (1999)
22. Desai, T.A., et al.: Nanoporous anti-fouling silicon membranes for biosensor applications, *Biosensors and Bioelectronics* **15** (9–10), 453–462 (2000)
23. Brody, J.P., Yager, P.: Diffusion-based extraction in a microfabricated device, *Sensors and Actuators A: Physical* **58** (1), 13–18 (1997)
24. Hatch, A., et al.: A rapid diffusion immunoassay in a T-sensor, *Nature Biotechnology* **19** (5), 461–465 (2001)
25. Raymond, D.E., Manz, A., Widmer, H.M.: Continuous sample pretreatment using a free-flow electrophoresis device integrated onto a silicon chip, *Anal. Chem.* **66**, 2858–2865 (1994)
26. Stachowiak, T.B., Svec, F., Frechet, J.M.J.: Chip electrochromatography, *Journal of Chromatography A* **1044** (1–2), 97–111 (2004)
27. Sarrut, N., et al.: Enzymatic digestion and liquid chromatography in micro-pillar reactors; hydrodynamic versus electroosmotic driven flow. In: *Photonics West, Microfluidics, BioMEMS and Medical Microsystems III*, San Jose, California: SPIE-Int. Soc. Opt. Eng. (2005)
28. Oleschuk, R.D., et al.: Trapping of bead-based reagents within microfluidic systems: On-chip solid-phase extraction and electrochromatography, *Analytical Chemistry* **72** (3), 585–590 (2000)
29. Yu, C., et al.: Monolithic porous polymer for on-chip solid-phase extraction and preconcentration prepared by photoinitiated in situ polymerization within a microfluidic device, *Analytical Chemistry* **73** (21), 5088–5096 (2001)
30. Stachowiak, T.B., et al.: Fabrication of porous polymer monoliths covalently attached to the walls of channels in plastic microdevices, *Electrophoresis* **24** (21), 3689–3693 (2003)

31. Lion, N., et al.: Microfluidic systems in proteomics, *Electrophoresis* **24** (21), 3533–3562 (2003)
32. Fan, Z.H., et al.: Dynamic DNA hybridization on a chip using paramagnetic beads, *Analytical Chemistry* **71** (21), 4851–4859 (1999)
33. Choi, J.-W., et al.: Development and characterization of microfluidic devices and systems for magnetic bead-based biochemical detection, *Biomedical Microdevices* **3** (3), 191–200 (2001)
34. Tokeshi, M., et al.: Continuous-flow chemical processing on a microchip by combining microunit operations and a multiphase flow network, *Analytical Chemistry* **74** (7), 1565–1571 (2002)
35. Hashimoto, M., et al.: Rapid PCR in a continuous flow device, *Lab on a Chip* **4** (6), 638–645 (2004)
36. Giordano, B.C., et al.: Polymerase chain reaction in polymeric microchips: DNA amplification in less than 240 seconds, *Analytical Biochemistry* **291** (1), 124–132 (2001)
37. Gijjt, R.M., et al.: Chemical and physical processes for integrated temperature control in microfluidic devices, *Lab on a Chip* **3** (1), 1–4 (2003)
38. Roper, M.G., Easley, C.J., Landers, J.P.: Advances in polymerase chain reaction on microfluidic chips, *Analytical Chemistry* **77** (12), 3887–3893 (2005)
39. Schwarz, M.A., Hauser, P.C.: Recent developments in detection methods for microfabricated analytical devices, *Lab on a Chip* **1** (1), 1–6 (2001)
40. Fritz, J., et al.: Translating biomolecular recognition into nanomechanics, *Science* **288** (5464), 316–318 (2000)
41. Drummond, G., Hill, M.G., Barton, J.K.: Electrochemical DNA sensors, *Nature Biotechnology* **21** (10), 1192–1199 (2003)
42. Bakker, E., Qin, Y.: Electrochemical Sensors, *Anal. Chem.* **78** (12), 3965–3984 (2006)
43. Miller, M.M., et al.: Detection of a micron-sized magnetic sphere using a ring-shaped anisotropic magnetoresistance-based sensor: A model for a magnetoresistance-based biosensor, *Applied Physics Letters* **81** (12), 2211–2213 (2002)
44. Cui, Y., al.: Nanowire nanosensors for highly sensitive and selective detection of biological and chemical species, *Science* **293**, 1289–1292 (2001)
45. Patolsky, F., Zheng, G., Lieber, C.M.: Nanowire-based biosensors, *Analytical Chemistry* **78** (13), 4261 (2006)
46. Gu, L.-Q., Cheley, S., Bayley, H.: Capture of a single molecule in a nanocavity, *Science* **291**, 636–640 (2001)
47. Saleh, O.A., Sohn, L.L.: An artificial nanopore for molecular sensing, *Nano Letters* **3** (1), 37–38 (2003)
48. Gut, I.G.: Automation in genotyping of single nucleotide polymorphisms, *Human Mutation* **17**, 475–492 (2001)
49. Aebersold, R., Mann M.: Mass spectrometry-based proteomics, *Nature* **422**, 198–207 (2003)
50. Hierlemann, A., et al.: Microfabrication techniques for chemical/biosensors, *Proceedings of the IEEE* **91** (6), 839–863 (2003)
51. Verpoorte, E.M.J., de Rooij, N.F.: Microfluidics meets MEMS, *Proceedings of the IEEE* **91** (6), 930–953 (2003)
52. Becker, H., Locascio, L.E.: Polymer microfluidic devices, *Talanta* **56** (2), 221–378 (2002)

53. Spearing, S.M.: Materials issues in microelectromechanical systems (MEMS), *Acta Materialia* **48** (1), 179–196 (2000)
54. Kovacs, G.T.A., Maluf, N.I., Petersen, K.E.: Bulk micromachining of silicon, *Proceedings of the IEEE* **86** (8), 1536–1551 (1998)
55. Bustillo, J.M., Howe, R.T., Muller, R.S.: Surface micromachining for microelectromechanical systems, *Proceedings of the IEEE* **86** (8), 1552–1574 (1998)
56. Duffy, D.C., et al.: Rapid prototyping of microfluidic systems in poly (dimethylsiloxane), *Analytical Chemistry* **70** (23), 4974–4984 (1998)
57. Hecke, M., Schomburg, W.K.: Review on micro molding of thermoplastic polymers, *Journal of Micromechanics and Microengineering* **14** (3), R1–R14 (2004)
58. Noerholm, M., et al.: Polymer microfluidic chip for online monitoring of microarray hybridizations, *Lab on a Chip* **4** (1), 28–37 (2004)
59. Rossier, J.S., et al.: Plasma etched polymer microelectrochemical systems, *Lab on a Chip* **2** (3), 145–150 (2002)
60. Anderson, R.C., et al.: A miniature integrated device for automated multistep genetic assays, *Nucleic Acids Research* **28** (12), E60 (2000)
61. Liu, R.H., et al.: Self-contained, fully integrated biochip for sample preparation, polymerase chain reaction amplification, and DNA microarray detection, *Analytical Chemistry* **76** (7), 1824–1831 (2004)
62. Easley, C.J. et al.: A fully integrated microfluidic genetic analysis system with sample-in-answer-out capability, *Proceedings of the National Academy of Sciences* **103** (51), 19272–19277 (2006)
63. Hong, J.W., et al.: A nanoliter-scale nucleic acid processor with parallel architecture, *Nature Biotechnology* **22** (4), 435–439 (2004)
64. Hong, J.W., et al.: Molecular biology on a microfluidic chip, *Journal of Physics: Condensed Matter* **18** (18) (2006)

Polyelectrolyte Multilayers

P. Schaaf and J.-C. Voegel

21.1 The Idea

21.1.1 Construction and Properties

The films known as polyelectrolyte multilayers are made by alternating deposition of polyanions (negatively charged polymers) and polycations (positively charged polymers). The development of these films, invented in the 1990s [1,3], has seen a considerable burst of interest, in particular due to their many applications. Indeed, these films are used to make electroluminescent diodes [4], anti-reflecting surfaces [5], water filtering substrates [6], and substrates for the separation of chiral molecules [7]. The alternating deposition of positive and negative species can also be used to make films with a mechanical strength close to that of steel [8]. Applications to biosensors and especially biomaterials are currently under investigation [9]. This is the last example discussed in the present chapter.

Polyelectrolytes are charged polymers, usually soluble in an aqueous solution. When a surface, supposed negatively charged, is set in contact with a solution of polycations (positively charged polyelectrolytes), the chains will immediately interact with the surface via electrostatic interaction and adsorb onto it. Like any other polymer, polyelectrolytes do not adsorb lengthwise against the surface, but form loops and tails. This adsorption is generally irreversible, and replacing the polycation solution by the solvent (water) alone will only lead to very slight desorption. This irreversibility of adsorption results from the formation of many anchoring points with the surfaces along the long polymer chains. Even if the interaction energy between a monomer, the basic building block of the polymer, and a surface is small, the fact that a number of contact points are set up makes the overall interaction between a polymer and a surface rather strong. Furthermore, in order for a chain to desorb, all the anchor points on the surface must be broken simultaneously, and such an event is highly improbable.

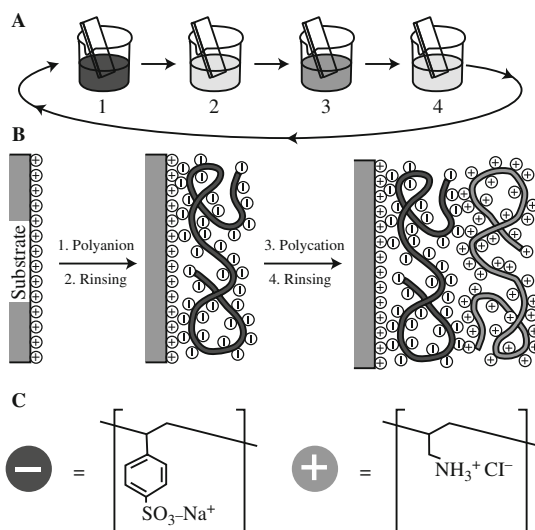


Fig. 21.1. Construction of polyelectrolyte multilayers. **(A)** A charged surface is immersed in a solution of polyanions, shown in *dark grey* (1). After adsorption, the surface is rinsed with an aqueous solution (2), then immersed in a solution of polycations, shown in *light grey* (3). This is then followed by another rinsing stage (4), and the cycle continued. **(B)** Each deposition stage leads to the formation of a polyelectrolyte layer which adds to the film, thereby producing a polyelectrolyte multilayer. **(C)** Example of a much studied polyanion/polycation combination. The polyanion is polystyrene sulfonate (PSS) and the polycation is poly(allylamine) hydrochloride (PAH). Taken from [3]

Up until the 1980s, it was thought that the adsorption of a polyelectrolyte on a surface of opposite charge would stop as soon as the surface was globally neutral, i.e., when the charges brought by the polyelectrolyte exactly compensated those initially present on the surface. However, this is not the case. When the adsorption of a polyelectrolyte on a surface stops, the surface is once again charged, but with the opposite sign. It then has the same sign as the polyelectrolyte adsorbed during the last deposition. During adsorption, the charge is therefore overcompensated, due to the formation of loops and tails created by the polymer chains. This charge surplus, positive in this case, can thus serve as an anchoring point for the deposition of other polyelectrolytes, this time polyanions. And once again, when the deposition process comes to an end, the surface now has a negative charge surplus. Hence, with each polyelectrolyte adsorption, there is an overcompensation of charge, and this allows the construction process to be continued at will. The result is a film called a polyelectrolyte multilayer [3]. The process underlying the construction of these layers is shown schematically in Fig. 21.1.

It is clear that what drives the construction is the overcompensation of charges in each deposition cycle, as revealed by the evolution of the

zeta potential during the construction [10]. Figure 21.2 shows an example of this evolution during each adsorption cycle for a poly(styrene sulfonate)/poly(allylamine) multilayer.

21.1.2 Physical Origin of Interactions Between Polyanions and Polycations

The formation polyanion/polycation complexes or polyelectrolyte multilayers arises mainly due to electrostatic interactions between oppositely charged polymer chains. These interactions do not usually result directly from attraction between the negative charges of the polyanions and the positive charges of the polycations. If this were the case, the complexation between a polyanion and a polycation would be exothermic. However, this reaction can be endothermic. In fact, the origin of this interaction is rather entropic. Indeed, since the polyelectrolyte chain is charged, it is surrounded by counterions which distribute themselves around it over a distance of the order of the Debye length. This is the characteristic length beyond which the charges are screened in the solution. Hence, if a polyanion chain is immersed in an aqueous solution containing NaCl, there will be an excess of Na^+ ions in the neighbourhood of the chain, thereby reducing the mobility of these ions. Conversely, in the neighbourhood of a polycation chain, there will be an excess of Cl^- ions. These small ions can come near to the charges carried by the chains and hence significantly reduce the electrostatic interaction energy. When the polyanion chain comes near the polycation chain, the neutralisation of some charges is not achieved exclusively by the counterions, but can also occur through the groups of opposite charges carried by the other chain. This mechanism leads to a release of counterions located in the neighbourhood of the two chains (see Fig. 21.3).

Through this release, the counterions become more mobile, and this increases the entropy S of the solution, i.e., $\Delta S > 0$. Very often, the charges carried by the chains are also bulkier than the small ions that play the role of counterions. The positive charges carried by the polycations cannot therefore come as close to the negative charges carried by the polyanions as the Cl^- ions. The polyanion–polycation interaction is then energetically unfavourable (endothermic reaction), i.e., the change in enthalpy ΔH (essentially the change in internal energy) is then positive. However, if the reaction is to occur, it is the change in free energy ΔG that must be negative:

$$\Delta G = \Delta H - T\Delta S, \quad (21.1)$$

where T is the temperature of the system. In general, the term $T\Delta S$ dominates over the term ΔH and the change in free energy is negative. The complexation reaction between a polyanion and a polycation, and hence also the formation of polyelectrolyte multilayers, is thus mainly of entropic origin.

The counterions of the polyelectrolyte chains thus play a crucial role in the electrostatic interactions. Hence an increase in the salt concentration in

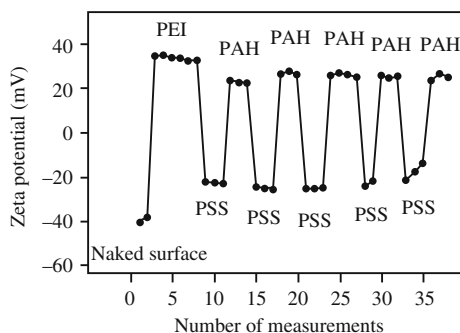


Fig. 21.2. Evolution of the zeta potential during the construction of a poly(styrene sulfonate)/poly(allylamine) multilayer. Taken from [10]

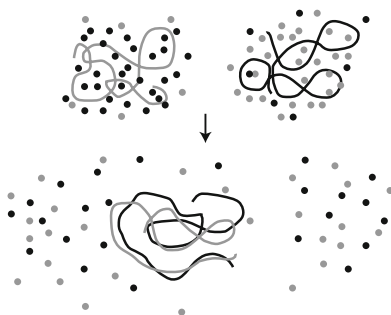


Fig. 21.3. Interaction of a polyanion and a polycation leads to the release of counterions in the solution. Before interaction, the polyanions in *light grey* (respectively, the polycations in *dark grey*) are surrounded by cations shown as *dark grey dots* (respectively, anions shown as *light grey dots*), which mainly constitute the counterion cloud. During the interaction between polyanions and polycations, some of these small ions are released. This release is accompanied by an increase in the entropy of the system and constitutes one of the mechanisms driving the attractive interaction between polyanions and polycations

the polyelectrolyte solutions during film construction will reduce the Debye length and the extent of the counterion clouds. A direct consequence is a reduction in the characteristic interaction length. Furthermore, the change in entropy due to release of counterions will also be reduced, thereby implying a reduction in the interaction between the polyelectrolyte chains. An increase in the salt concentration will also reduce the interactions between charges of the same sign in the same chain. The latter will then lose some of its rigidity, and its persistence length will be decreased. A consequence of this effect is the formation of thicker and less dense polyelectrolyte multilayers. For a poly(styrene sulfonate)/poly(allylamine) (PSS/PAH) multilayer, Ladam et al. [10] found that the increase in thickness per bilayer was 2.7 nm for an NaCl concentration of 0.01 M and 6 nm for an NaCl concentration of 0.9 M.

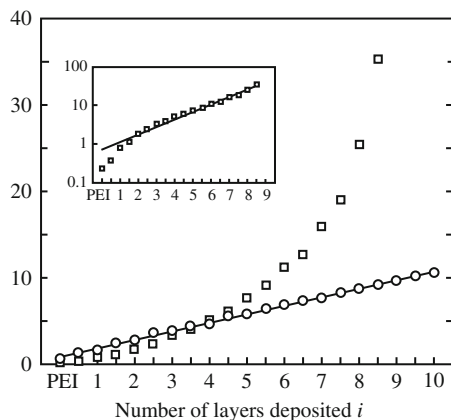


Fig. 21.4. Optical thickness of polyelectrolyte multilayers as a function of the number of layers deposited. *Circles* represent a poly(styrene sulfonate)/poly(allylamine) multilayer, which has linear growth, and *squares* a poly(L-lysine)/poly(L-glutamic acid) multilayer, which has exponential growth. The optical thickness is defined as the product of the thickness d_A and the difference between the refractive index n_A of the film and the refractive index n_c of the solution. This quantity is directly proportional to the mass of the film and is the quantity that can be most accurately measured using optical techniques. Graph taken from [2]

21.2 Linear Growth and Exponential Growth of Polyelectrolyte Films

21.2.1 Linear Growth

The possibility of constructing films by alternate adsorption of polyanions and polycations was discovered in 1992 by G. Decher and coworkers [1]. The first studies showed that the thickness of these films increased linearly with the number of deposition cycles, the thickness increment for each pair of layers being typically in the nanometer range (see Fig. 21.4). The structure of such films was studied by neutron reflectivity.

In that case, deuterated polyelectrolytes are used instead of their hydrogenated counterparts in some stages of the construction (see Fig. 21.5a). These polyelectrolytes have the same chemical and physicochemical properties, but interact differently with neutrons. Now a beam of neutrons of the same momentum behaves like a monochromatic light beam. The wavelength associated with such a beam is typically in the range 10–50 nm, i.e., the right length range for investigating multilayers. When such a beam is directed on a multilayer containing deuterated layers that are distributed periodically through the film, constructive interference is observed for some angles of incidence, giving rise to Bragg peaks (see Fig. 21.5b). The phenomenon is analogous to the diffraction patterns observed when X-rays are reflected by crystals.

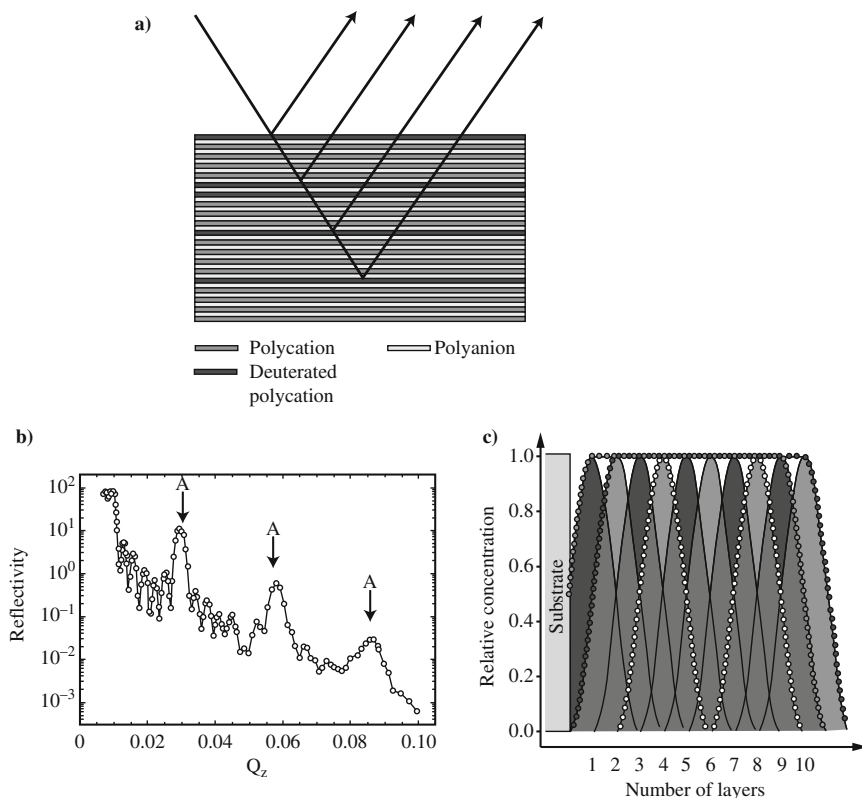


Fig. 21.5. (a) Schematic view of a multilayer of polyanions (*light grey*) and polycations (*dark grey*), for which some polycation layers have been replaced by deuterated polycation layers (*black*). (b) When the neutron beam reaches such a film, the deuterated layers reflect more neutrons. When the deuterated layers are distributed periodically through the film, constructive interference occurs for certain reflected directions, and this gives rise to Bragg peaks. (c) Examination of these curves suggests a periodic model for the distribution of polyanions and polycations in the film. Each layer interpenetrates only its closest neighbours. The quantity Q_z in Fig. 21.5b is the diffusion vector given by $Q_z = (2\pi/\lambda) \sin(\theta/2)$, where λ is the wavelength of the neutron beam and θ the angle of diffraction. A indicates the Bragg peaks. Curve (c) is taken from [3]

Accurate analysis of the data shows that each layer interpenetrates only its closest neighbours, leading to a kind of layered molecular structure (see Fig. 21.5c) [3]. This suggests that, in each adsorption stage, the polyelectrolytes in the solution only interact with those of opposite charge forming the outermost layer of the film. In particular, no polyelectrolyte diffuses normally to the film. However, lateral diffusion of the polyelectrolytes has been observed

within their adsorption layer. The formation of these films corresponds in fact to the formation of polyanion/polycation complexes at the surface.

One of the main questions raised here concerns the way electrical neutrality is achieved within these films. There appear to be two non-exclusive possibilities:

1. During the construction, the masses of the polyelectrolytes that adsorb are such that the number of positive charges brought by the polycations is equal to the number of negative charges brought by the polyanions.
2. There may be a surplus of one of the two types of charge, which is then compensated by counterions of opposite charge within the film [11].

Experiments carried out using radioactive ions tend to validate the first hypothesis, but the question has not yet been decisively answered.

Films with linear growth can be constructed with two types of polyelectrolyte: strong polyelectrolytes, for which the degree of ionisation is barely sensitive to pH variations, and weak polyelectrolytes, for which the charge is sensitive to the pH. In both cases, the ionic strength of the polyelectrolyte solutions used to construct the films is an important parameter, strongly influencing the thickness of a deposited bilayer. As we have already seen, the greater the ionic strength, the greater the increase in thickness of the bilayer. This arises due to the screening of charges when the ionic strength is great, one consequence of this being the more globular shape of the polyelectrolytes dissolved in the solution. For multilayers constructed using weak polyelectrolytes, the pH of the solution during construction is also an important factor affecting the film thickness [12], with the general rule that, the more highly charged the two polyelectrolytes, the more likely they are to produce thin layers, while a low charge carried by one of the two polyelectrolytes during film construction leads to a greater thickness. There is, however, a critical value for the level of ionisation, below which the film will not be constructed at all.

21.2.2 Exponential Growth

Alternate deposition of polyanions and polycations does not only lead to films whose thickness increases linearly with the number n of layers. In fact, the thickness of some such films grows exponentially with n . These films were discovered at the end of 1990, when multilayers of polypeptides and/or polysaccharides were considered for the first time [13].

In a limited number of deposition cycles with this growth mode, thicknesses anywhere between a few hundred nanometers and a few tens of micrometers can be reached. The mechanism leading to this exponential growth is totally different to the one for films with linear growth. Indeed, it has been shown that, for exponential growth, at least one of the two polyelectrolytes used in the construction of the film must diffuse in and out of the architecture during each stage of the construction [14]. The growth mechanism is shown schematically in Fig. 21.6.

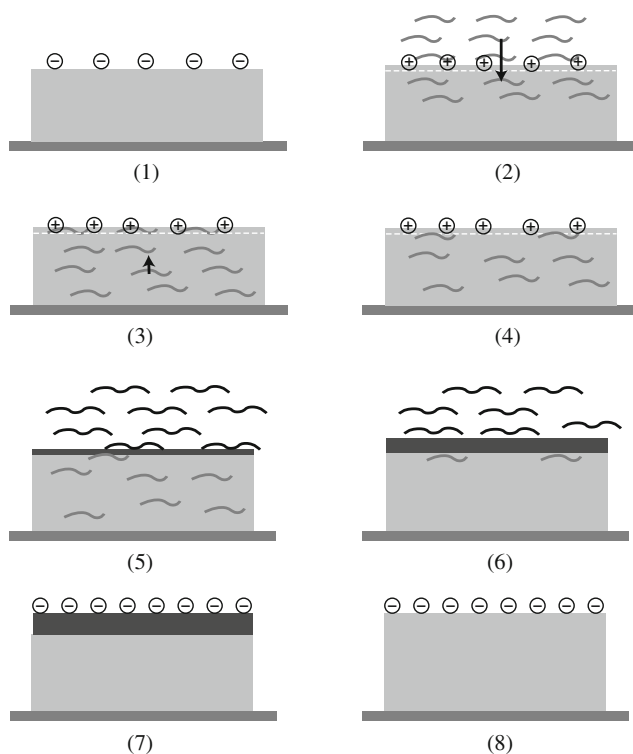


Fig. 21.6. Schematic view of the physical process leading to exponential growth. The case shown here is a film in which the polycation (*curve with one wave*) can diffuse in and out of the film, while the polyanion (*curve with two waves*) does not diffuse. (1) Film under construction after adsorption of the negatively charged polyanion. (2) Addition of polycations which diffuse into the construction. (3) Having removed the supernatant, the free polycations begin to diffuse out of the film. (4) Not all the polycations leave the film, since a certain proportion remain in free form inside the film. (5) Addition of polyanions accompanied by diffusion of free polycations outside the construction. (6) All free polycations leave the film and form complexes with the polyanions at the interface between the film and the solution, contributing to the growth of the film. (7) A new layer forms on the construction. (8) It is negatively charged. The process can now continue by addition of polycations

Suppose the polycation is the diffusing species and consider the film after the rinsing stage that follows deposition of the polyanion (see Fig. 21.6 step 1). There is then an excess of negative charges at the top of the film. When it is set in contact with a polycation solution, the chains first arriving in contact with the film interact with the polycations at the interface between the film and the solution by a similar mechanism to the one observed with linear growth films. However, in contrast to the latter, the polycations also diffuse inside the film to constitute what we shall call free chains (see Fig. 21.6 step 2). The

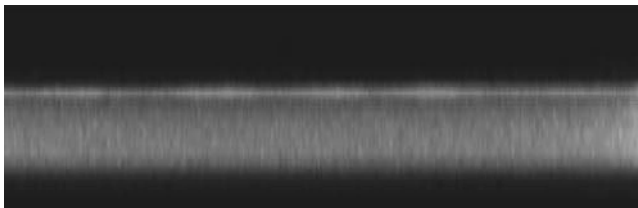


Fig. 21.7. Cross-section of a (poly(L-lysine)/hyaluronic acid)₂₀ multilayer imaged by confocal microscopy. The last step in the construction was carried out with hyaluronic acid labelled with Texas Red and poly(L-lysine) labelled with (green) fluorescein. This image proves that the poly(L-lysine) diffuses throughout the film during this step (see text for details). The film has thickness about 4 μm . Taken from [14]

adsorption process stops when the chemical potential of the free polycations in the film equals the chemical potential of the polycations in the solution. Since the film is in contact with the solution of polycations, an excess of positive charges will appear at the film/solution interface (see Fig. 21.6 step 3). When rinsing, some of the free polycations in the film can diffuse out of the film and into the solution (see Fig. 21.6 step 4). The excess of positive charges in the film also creates an electrostatic potential barrier that the polycation chains must cross. Now the diffusion of free polycations out of the film reduces their concentration within the film, as well as their electrochemical potential, whereas the energy barrier to be crossed in order to diffuse out of the film is on the increase. When the height of this barrier goes over kT , i.e., the thermal energy, diffusion out of the film slows down very rapidly, and the process comes to a total halt when the height of this barrier becomes very much greater than kT . There are still free polycations in the film at the end of the rinsing step. When this film is set in contact with a solution of polyanions, the negative chains will first interact with the surplus positive charges at the film/solution interface, causing the electrostatic energy barrier to disappear, whence all the free polycations in the film will diffuse out of it (see Fig. 21.6 step 5). When they reach the film/solution interface, they form complexes with the polyanions in the solution and close to the interface. These complexes remain bound to the multilayer and constitute the next layer. The process of polyanion deposition stops when all the free polycations in the film have formed complexes (see Fig. 21.6 step 6). It is therefore observed that the mass of complexes adding to the film is proportional to the amount of free polyanions in the film before it is set in contact with the polycation solution (see Fig. 21.6 step 7). This amount is itself proportional to the thickness of the film. This is the classic situation for exponential growth, where the mass increment at each polycation addition stage is proportional to the mass, or the thickness, of film already formed. Such a film will therefore grow exponentially with the number of deposits.

This growth mechanism has been confirmed by experiments using confocal optical microscopy. Figure 21.7 shows a cross-section of a hyaluronic acid/poly-L-lysine (HA/PLL) multilayer. To construct this film, 19 bilayers of poly(L-lysine/hyaluronic acid) were first deposited using unlabelled polyelectrolytes. The twentieth deposit was made using hyaluronic acid labelled with Texas Red, a fluorescent molecule emitting in the red, and poly-L-lysine labelled with fluorescein, a fluorescent molecule emitting in the green. It was observed that the whole film appeared green, while a red strip appeared at the top (a thin lighter strip in Fig. 21.7). It can be deduced that the PLL diffuses into the film, whereas the hyaluronic acid does not. Similar experiments have shown that PLL also diffuses out of the film after setting it in contact with a solution of hyaluronic acid.

Films with exponential growth seem more similar to gels formed from complexes than well organised periodic structures like the linear growth multilayers. By abuse of language, the term ‘multilayer’ is nevertheless used to refer to these films, by reference to the deposition method.

21.2.3 Fabrication of Polyelectrolyte Multilayers

Most films described in the literature are obtained by immersing the substrates alternately in polyanion and polycation solutions, with the two adsorption stages being separated by a rinsing stage (see Fig. 21.1). The latter merely eliminates the excess of polyelectrolytes in contact with the surface but only weakly bound to the film. During each deposition stage, the substrate is usually in contact with the polyelectrolyte solutions for 5–20 min, while the rinsing stage lasts between 5 and 10 min. Another method is to bring the polyelectrolyte solutions into contact with the substrate by pouring them over the surface. This approach is widely used for studies in which film construction is monitored in situ. The characteristic times of the different stages are the same as in the immersion method. Polyelectrolyte adsorption kinetics are limited by their diffusion from the solution into the film. It may thus take more than 10 h to construct a film with twenty bilayers.

To speed up multilayer formation, one can also use the method known as spin coating [15]. In this method, the polyelectrolyte solutions are alternately deposited on a substrate rotating at high speed (several hundred revolutions per minute). In each deposition, the droplet spreads over the surface by the effect of the centrifugal force to form a film with thickness in the micrometer range. As they spread, the polyelectrolyte chains anchor themselves onto the layer of opposite charge. At the same time as the spreading, evaporation is observed, leading to an increased concentration of the polyelectrolyte solution above the film. This effect seems to lead to densification of the multilayers. The chains in the excess solution are carried along by the liquid. After each deposition stage, there is a rinsing stage in which the polyelectrolyte solution is replaced by a pure buffer. Rinsing eliminates surplus polyelectrolytes and prevents the formation of polyanion/polycation complexes in solution above

the film. The method is much faster than the standard immersion process, but deposition has to be carried out on a plane substrate.

A third way of making multilayers is to deposit the polyelectrolytes by alternately spraying solutions of polyanions and polycations onto a vertically oriented substrate [16]. This orientation is necessary in order to allow the natural removal of excess polyelectrolytes near the film by drainage under gravity. The result is that only a small fraction of the sprayed polyelectrolytes remains at the surface. This fraction typically dries with a few minutes. The very fast spreading and drainage of the fine droplets arriving on the surface considerably accelerates the diffusion–convection phenomenon of the polyelectrolytes toward the surface and hence increases their adsorption kinetics. The different spray stages can also be very short (typically between 1 and 20 seconds). This is a key difference with the immersion method of film construction and considerably reduces the contact time required to deposit a layer. Each deposition stage is followed by rinsing, during which the buffer is sprayed onto the substrate. Despite the differences between the immersion and spray methods, the multilayers obtained by the two approaches have very similar growth and structure. It has even been shown that the rinsing stages can be dispensed with, producing films that grow continuously but turn out to be rougher. Finally, one can not only dispense with rinsing, but simultaneously spray the polyanions and polycations [17]. One then obtains films that can no longer be called polyelectrolyte multilayers, but whose thickness increases with the spray time. Note, however, that simultaneous spraying is not possible with all polyanion/polycation combinations and we cannot yet predict *a priori* which pairs will be amenable to this method of deposition. The spray methods for fabricating polyelectrolyte films are interesting, because they are easily transposed from the laboratory to large scale production. They are therefore likely to see significant development in the future.

21.3 Biological Functionalisation

There are different ways of functionalising polyelectrolyte multilayers. Some films can be rendered totally bio-inert by choosing suitable polyelectrolytes and deposition conditions [18]. It is also possible to make them biologically active by inserting proteins [19], active peptides [20,21], and even drugs [22]. We shall now discuss the various ways of functionalising these films.

21.3.1 Biologically Inert Films

Bio-inert films are films that resist protein adsorption and/or cell adhesion. This is one of the main objectives when making materials to be set in contact with blood, such as implants or biosensors, for example. One of the most widely used methods for achieving this aim is to coat materials with poly(ethylene glycol) (PEG) chains. However, methods for preparing these

surfaces are complex and cannot be applied to all substrates and all geometries. In 1999, Elbert et al. [23] deposited alginate/poly-(L-lysine) multilayers on various solid substrates, viz., gelatin, extracellular matrix produced by fibroblasts, and type I collagen. These films were the first example of exponential growth multilayers that were bio-inert with regard to human fibroblasts. Other multilayers using polysaccharides, like hyaluronic acid/poly-(L-lysine) and hyaluronic acid/chitosan films, also proved to be non-adhesive with respect to other types of cells, such as chondrosarcomas [24] or osteoblasts. The particularity of these films is that they are highly hydrated. Non-adhesive films have also been obtained with polyacrylic acid and poly(allylamine).

Films for which the polyanion and polycation were both deposited at pH 2.0 proved to be highly resistant with regard to cell adhesion, whether the film terminated with the polycation or the polyanion layer. Furthermore, no correlation between the adhesion of cells and the adsorption of proteins, such as fibrinogen or lysozyme, was observed for these films. Similar behaviour had been noted with regard to these same cells when polyacrylic acid was replaced by polymethacrylic acid, poly(styrene sulfonate), or poly(diallyldimethylammonium) chloride. A direct correlation was demonstrated between the non-adhesive character of the films and their ability, once dried, to swell when set in contact with an aqueous solution. The greater the ability of a multilayer to swell, the more the film appears to be non-adhesive for cells [18]. This result also suggests that the non-adhesive character of a multilayer polyelectrolyte film will become more significant as its level of hydration increases.

Micromanipulation studies have been carried out to determine the adhesive forces between chondrosarcomas and (poly-(L-lysine)/poly-(L-glutamic acid))_n multilayers [24]. Whereas for films terminating with the polyglutamic acid layer the adhesive forces are very weak, even zero, they are much stronger in films terminating by the polylysine layer, and decrease with the number *n* of bilayers. It has also been shown that such films can prevent the adhesion and proliferation of bacteria, in particular (poly-(L-glutamic acid)/poly-(L-lysine))_n films terminated by three (polylysine/PGA-PEG) bilayers, where PGA-PEG are poly(L-glutamic) acid chains on which poly(ethylene glycol) chains of mass 2,000 Da have been covalently coupled [25]. These architectures reduce the adhesion of bacteria like *Escherichia coli* by more than 92%. Other multilayers resisting protein adsorption have also been described. Hence, at pH 7.4, proteins like lysozyme, immunoglobulins, human serum albumin (HSA), and fibronectin are practically unable to adsorb onto the surface of films carrying opposite charge to those of the protein for architectures based on the alternating deposition of poly(ethyleneimine) and polyacrylic acid or poly(ethyleneimine) and polymaleic acid, while they adsorb significantly onto films including strong bases, or poly(ethyleneimine)/poly(vinylsulfate) films [18, 26].

The fabrication of non-adhesive multilayers is not just an aim in itself, but fits into a broader strategy which seeks to subsequently functionalise

these surfaces in a selective way, to induce the adhesion of one specific cell type. To achieve such an aim, non-adhesive films will be functionalised by specific adhesion peptides for a given cell type. Studies are currently underway in several research centers, with a view to biomedical applications for such polyelectrolyte films, in particular for coating biomaterials.

21.3.2 Functionalisation by Protein Insertion

Multilayers can be functionalised by adsorbing proteins at the top of an architecture or by inserting one or more proteins, the same or different, at different levels in the construction while it is being built up. However, these proteins must be able to conserve a structure close to their native structure and, even when buried within the film, still be able to interact with cells deposited at the top of the film.

In the last section we described multilayers with non-adsorbent behaviour with regard to proteins. These are nevertheless exceptions, and proteins are generally observed to adsorb at the surface of multilayer films. Studies of the adsorption of different proteins at the surface of poly(styrene)/poly(allylamine) multilayers has shown that the amounts adsorbed are greater when the overall charge of the proteins is opposite to that of the multilayer, although the amounts adsorbed are not zero, and far from being negligible, when the proteins are of the same sign as the film (see Fig. 21.8) [27]. This is likely to be because any protein, although it has some net charge at any given pH, still carries positive and negative groups.

The construction of the film can be pursued on the layer of adsorbed proteins, whereupon they will be buried within the final structure. By using fluorescent proteins, it has been shown that proteins like protein A, buried within a linear growth film like poly-(L-lysine)/poly(L-glutamic) acid were unable to diffuse in the direction normal to the film and therefore remained confined within their deposition layer [19]. However, lateral diffusion has been observed for a large fraction of human serum albumin molecules adsorbed or inserted within linear growth films like poly(styrene sulfonate)/poly(allylamine) [28, 29].

Clearly, in order for proteins inserted within these films to conserve their biological activity, they must not be denatured. Polyelectrolytes in general and polyelectrolyte multilayers in particular seem to successfully fill this role. For example, it has been shown that glucose isomerase, glucosamylase, glucose oxidase, and peroxidase included in polyelectrolyte multilayers fully conserve their enzyme activity. Likewise, immunoglobulin G buried under a small number of polyelectrolyte layers conserves its activity and interacts with its antigens [30]. Fourier transform infrared spectroscopy has been used to show that proteins inserted in multilayer films conserve the secondary structure of their native form [31]. In addition, the presence of the polyelectrolytes prevents the formation of intermolecular β sheets and stabilises proteins like fibrinogen from thermal denaturation.

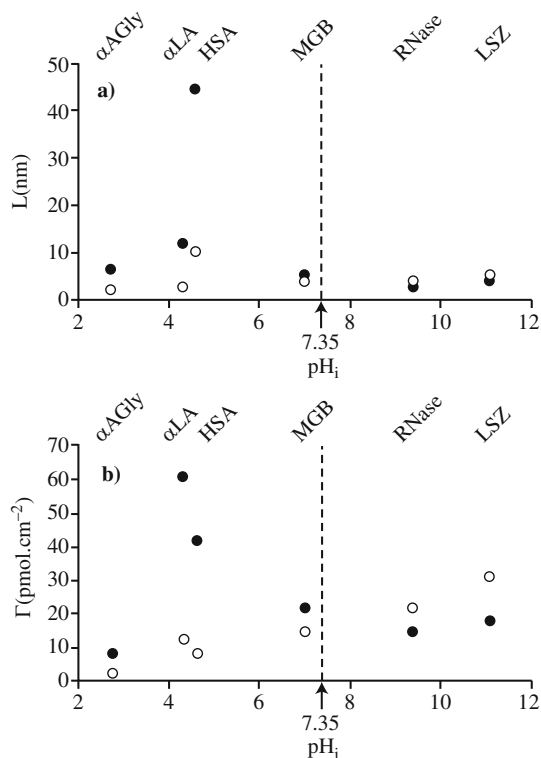


Fig. 21.8. Thickness L and mass Γ of layers of different proteins adsorbed onto PEI-(PSS-PAH)₃ (●) and PEI-(PSS-PAH)₃-PSS (○) multilayers as a function of the isoelectric point pH_i of the proteins. Adsorption occurs at pH 7.35, 0.15 M NaCl. Abbreviations are PEI for poly(ethyleneimine), PSS for poly(styrene sulfonate), PAH for poly(allylamine), αAGly for α₁-glycoprotein acid, αLA for α-lactalbumin, HSA for human serum albumin, MGB for myoglobin, RNase for ribonuclease A, and LSZ for lysozyme. Taken from [27]

Communication between cells deposited at the top of the structure and proteins buried in the polyelectrolyte assembly is a key factor for this type of functionalisation. This question has been studied for a model system [19]. Protein A is a molecule with pro-inflammatory properties. So the presence of protein A in contact with monocytes causes them to produce TNF- α . The latter can therefore serve to indicate that contact has occurred between monocytes and protein A. Experiments have been carried out in which protein A was inserted at different levels in poly(glutamic) acid/poly-(L-lysine) (PLL/PGA) multilayers and monocytes were then deposited at the top. the production of TNF- α was then monitored. It was observed that production of this substance increases for the first 2 h of contact between the cells and the film when the protein A is buried under 10 pairs of PGA/PLL layers. The production of

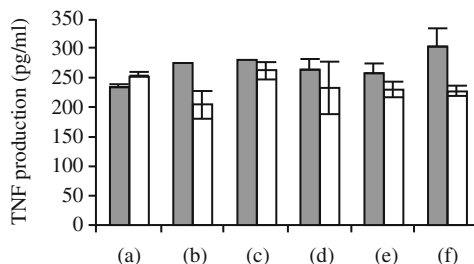


Fig. 21.9. Production of TNF- α by monocytes deposited on different multilayers containing protein A. The latter is deposited on a PEI-PGA-PLL precursor film and inserted under n PGA-PLL bilayers. (a) $n = 0$. (b) $n = 10$. (c) $n = 15$. (d) $n = 20$. (e) $n = 25$. (f) $n = 30$. *Dark columns* show measurements after 4 h of contact between cells and film, while *light columns* show measurements after 12 h of contact. PEI poly(ethyleneimine), PGA poly-(L-glutamic) acid, PLL poly-(L-lysine). Data extracted from [19]

TNF- α then saturates over a period of 12 h. Experiments have also been carried out in which the production of TNF- α was measured after 4 h and 12 h of contact between the cells and the films, for different buried depths between 10 and 30 PGA/PLL bilayers. The production of TNF- α is roughly the same whatever the depth of the protein A in the multilayer (see Fig. 21.9).

Finally, it has been shown that TNF- α production ceases when PLL is replaced by its D homologue (PDL). Furthermore, confocal optical microscopy has been used to show that the protein A does not diffuse vertically through the film. All these results suggest that the monocytes come into contact with the protein A buried in the multilayer by locally degrading the film. This hypothesis was confirmed by the observation of pseudopods formed by cells in contact with the protein A layer. This degradation is probably due to enzymes, which explains why there is no TNF- α production when PLL is replaced by PDL.

The fact that there is no degradation of the film deposited on the protein A when D form polypeptides are used instead of L form polypeptides suggests a way of modulating the destruction of the film by means of the D and L forms for polylysine and polyglutamic acid. It turns out that these mixtures, associated with variations in the number of bilayers covering the protein A, provide a good way of modulating the response in a time lapse of 0–6 h (see Fig. 21.10). Note in particular that this approach can trigger the full response in a time lapse of 1 h [32].

Constructions have also been made by alternating proteins and polyelectrolytes [33, 34]. The film can be stabilised by using glutaraldehyde to form covalent bonds between proteins, and uncoupled polyelectrolytes can then be removed by rinsing. For this purpose, constructions have been made containing only albumin, or albumin and heparin [35]. These films are designed to coat medical equipment used in contact with blood plasma. Surface passivation

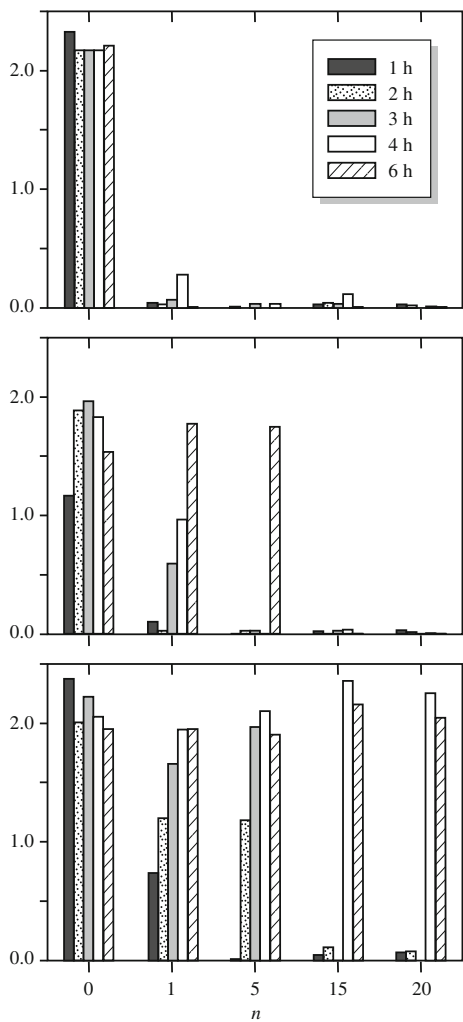


Fig. 21.10. Production of TNF- α for monocytes deposited on architectures made up of a mixture of L and D forms of polylysine and polyglutamic acid (respectively, PLL and PDL or PLGA or PDGA). Mixtures are expressed with x % of PDL or PDGA and $(1-x)$ % of PLL or PLGA in the solutions used to prepare the films. The architectures used are $(\text{PLL}/\text{PLGA})_5\text{-PLL-PA-}(\text{PDL}_x\text{PLL}_{1-x}/\text{PDGA}_x\text{PLGA}_{1-x})_n\text{-DL}_x\text{PLL}_{1-x}$, with $n = 0, 1, 5, 15, 20$, and x varying between 0 and 1. When $n = 0$, there is no bilayer at the top of the $\text{PDL}_x\text{PLL}_{1-x}$ film. *Top:* $x > 0.5$. *Center:* x close to 0.3 and 0.4. *Bottom:* x close to 0.1 and 0.2. Results extracted from [32]

by this type of film, made up exclusively of albumin, reveals itself through the almost total lack of adsorption of IgG from blood plasma on these substrates. This kind of treatment would thus appear to prevent direct contact between the surface and the cells or proteins of the blood plasma. For their part, constructions terminating by a heparin layer seem to have anticlotting properties revealed by their reaction with antithrombin. By using natural polyelectrolytes such as chitosan and dextran sulfate to build up films, it is possible to modulate pro- or anticlotting properties with respect to the blood medium [36, 37]. The ionic strength of the polyelectrolyte solutions used to construct films plays an essential role in their resulting properties. The film structure varies in particular when salt concentrations lie in the

range 0.2–0.5 M NaCl. Increasing the salt concentration leads to an increase in the thickness of each deposited layer, and also an increase in the (negative) charge excess observed after each dextran deposit. So after deposition of the third pair of layers and for a salt concentration greater than 0.5 M, the film is proclotting when terminated by chitosan and anticlotting when terminated by dextran. The anticlotting properties are explained by the charge excess introduced by the outer dextran layer. For the other conditions (lower salt concentration and hence thinner films), proclotting properties are still observed. Similar constructions obtained by successive deposition of chitosan and heparin dissolved in solutions of NaCl (1 M) exhibit significant anticlotting properties, independently of the nature of the last deposited layer.

Films with exponential growth, in particular HA/PLL or HA/chitosan multilayers, often reach a thickness of several micrometers after deposition of a limited number of pairs of layers (typically 15–20). These films, built from what are in principle highly biocompatible biopolymers, nevertheless display very low potential in terms of cell adhesion and proliferation (in particular for different types of primary cells, such as chondrocytes, fibroblasts, epithelial cells, osteoblasts, and so on). The main reason for this property which limits their use is without doubt their very high level of hydration, preventing cells from anchoring strongly onto their surface. One way of solving the problem is to crosslink the films chemically [38]. For films containing polyanions with carboxylic groups and polycations with amine groups, this can be done for example using carbodiimide combined with N-hydroxysulfo-succinimide, which leads to the formation of amide bonds. In this way, it has been shown that crosslinking an (HA/PLL) multilayer leads to good adhesion and proliferation of chondrosarcomas, whereas when the same architecture is not crosslinked, it strongly resists adhesion of these cells.

21.3.3 Functionalisation by Peptides

Another approach to functionalisation is to include polyelectrolytes modified by covalent coupling with a peptide in the architecture of the film (see Fig. 21.11). This idea has been demonstrated by measuring the response of melanocytes (B16-F1) deposited on (PGA/PLL)_n architectures in which the poly-(L-lysine) (PLL) molecules have been replaced in some layers by other PLL chains on which α -melanocortin (α -MSH), a biologically active peptide, has been coupled [20]. This peptide hormone is a potential stimulator of melanogenesis and, when it comes into contact with melanocytes, it can cause a cascade reaction within the cells. After a few hours, cyclic AMP is produced, and after a few days melanin. These two molecules have been used to reveal the biological activity of films including chains that contain PLL- α -MSH. It has been observed that, on short time scales, the activity of such films, assessed by cyclic AMP production, in which PLL- α -MSH molecules have been buried, depends on the depth of the modified polyelectrolyte, while on long time scales (melanin production), the response does not change until



Fig. 21.11. Example of a peptide (α -MSH) that could be coupled with a polypeptide chain in order to functionalise a multilayer film by insertion

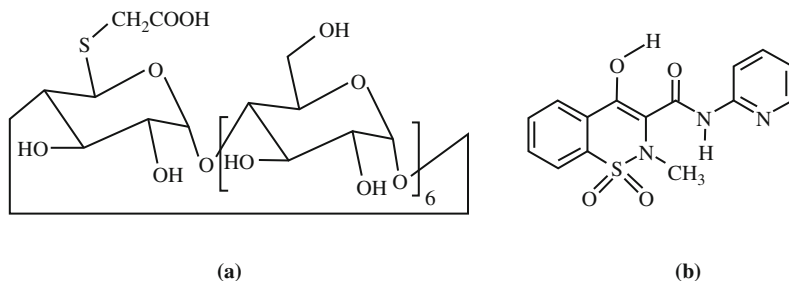


Fig. 21.12. (a) Chemical structure of 6A-carboxymethylthio- β -cyclodextrin. (b) Chemical structure of pyroxicam

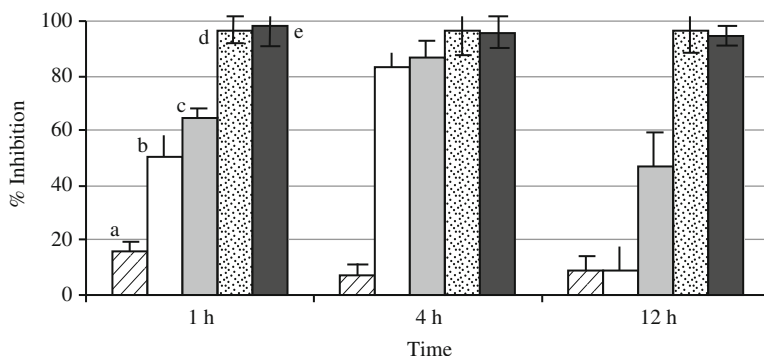


Fig. 21.13. Inhibition of the production of TNF- α secreted by monocytes incubated with 10 ng mL^{-1} of lipopolysaccharide and deposited on multilayers containing cyclodextrin-pyroxicam (cCDPx) complexes. Cells were deposited on different films: (a) PEI-(PGA-PLL)₆-Px. (b) PEI-(PGA-PLL)₆-cCDPx. (c) PEI-(PGA-PLL)₆-cCDPx-PLL-(PGA-PLL)₃. (d) PEI-(PGA-PLL)₃-cCDPx-PLL-(PGA-PLL)₃-cCDPx-PLL-(PGA-PLL)₃. (e) PEI-(PGA-(PLL-cCDPx)₃-PLL). Incubation times in the presence of lipopolysaccharide were 1 h, 4 h, and 12 h before measuring the TNF- α production. Px pyroxicam, cCDPx cyclodextrin-pyroxicam complex. Taken from [22]

the PLL- α -MSH has been covered by at least 25 bilayers of (PGA/PLL). In this study, the cell communication mechanism was not rigorously identified and may have several origins. (PGA/PLL) films have exponential growth and it is known that the two polyelectrolytes are able to diffuse both in and out of the film during each construction step.

Cell communication may also result from diffusion of PLL- α -MSH toward other cells in contact with the film. Moreover, cell communication may also be due to degradation of the multilayers by the cells, similar to what happens for the interaction of monocytes with protein A inserted in films made with the same polyelectrolyte system. However, this second possibility cannot be the only signalling channel, because insertion in the (PGA/PLL) $_n$ architecture of several layers of (PSS/PAH) non-degradable by enzymes does not fully prevent cell communication. It has also been shown that PLL/PGA multilayers including PLL- α -MSH chains also have anti-inflammatory properties [21].

21.3.4 Functionalisation by Drugs

Polyelectrolyte multilayers can also be functionalised by inserting drugs. For example, pyroxicam, a molecule known for its anti-inflammatory properties, has recently been inserted in (PGA/PLL) $_n$ films [22]. This molecule is shown in Fig. 21.12b. Pyroxicam is highly soluble in water. It is thus generally administered in the form of complexes with cyclodextrin molecules. These occur in the form of cages (see Fig. 21.12a), in which the pyroxicam molecules can be inserted.

To be able to insert these complexes in polyelectrolyte multilayers, a charged cyclodextrin, viz., 6A-carboxymethylthio- β -cyclodextrin, was used. Anti-inflammatory properties were demonstrated by coating films functionalised by pyroxicam with monocytes. The cells were then stimulated by a lipopolysaccharide. If there were no anti-inflammatory substance, this would naturally lead to the production of TNF- α , an inflammation marker. Now, where there is pyroxicam in the multilayers, one observes a significant decrease in the production of TNF- α by the monocytes, and this is an indication of the anti-inflammatory activity of these films (see Fig. 21.13).

21.3.5 Development of Nanoreactors

We have seen how proteins can be included within multilayer films while conserving their biological activity. The same goes for enzymes. The preservation of the activity of enzymes included in polyelectrolyte multilayer films was first investigated for enzyme/polyelectrolyte architectures, e.g., PEI/GOD or POD/PSS, where PEI is poly(ethyleneimine), PSS is poly(styrene sulfonate), GOD is glucose oxidase, and POD is peroxidase [39–41]. The first enzyme (GOD) catalyses the reaction between β -glucose and O $_2$, with production of H $_2$ O $_2$ and D-glucono- δ lactone, which oxidises the dye DA67. POD catalyses this last reaction. Constructions involving both the enzymes GOD and POD were then made. These consisted of 4 PEI-PSS (precursor film) bilayers, on which two POD-PSS bilayers and two PEI-GOD bilayers were deposited one after the other. The reaction in two successive steps involving the two enzymes is initiated by adding glucose and the dye DA67 to the above solution for the construction. Multilayers made with insertion of glucamylase (GA),

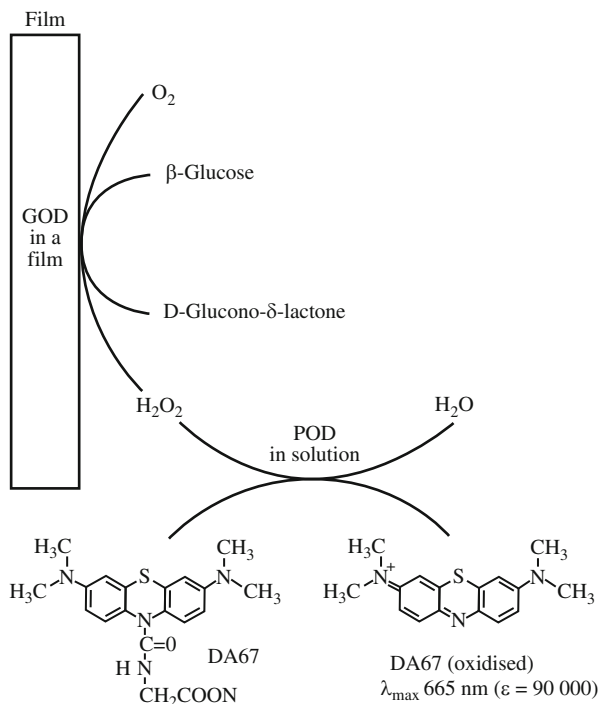


Fig. 21.14. Sequence of enzyme reactions using glucose oxidase (GOD) immobilised in the film, and peroxidase (POD) in solution (3 mL, pH 7, and 25°C): glucose 56 mM (0.01 mg/mL), POD 0.004 mg/mL, DA67 10 mM (0.004 mg/mL). Taken from [42]

which catalyses the conversion of starch into β -glucose, were also built on substrates made from ultrafiltration membranes. So an assembly comprising, from top to bottom, two bilayers of PEI/GA and two bilayers of PEI/GOD, is once again able to produce H_2O_2 by conversion of starch to β -glucose, this reaction being catalysed by GA. The β -glucose is in turn transformed into D-glucono- δ galactone (when it comes into contact with GOD) with production of H_2O_2 which, in solution in the presence of POD, leads to oxidation of the dye DA67, as can be seen from Fig. 21.14.

Constructions producing a cascade of two or three enzyme reactions can thus be obtained by a judicious assembly of polyelectrolytes and enzymes, to make a structure in which the product of intermediate reactions subsequently reacts with molecules located in the same neighbourhood. The yield of the multistep reaction can be increased by optimising the separation between the enzyme layers. These multienzyme constructions imitate natural organelles and could perhaps be used to carry out reactions in several steps with unstable intermediate reaction products. Such reactions are difficult to carry out in solution when the enzymes are dissolved in solution. In addition, this kind of assembly requires only a small amount of enzymes.

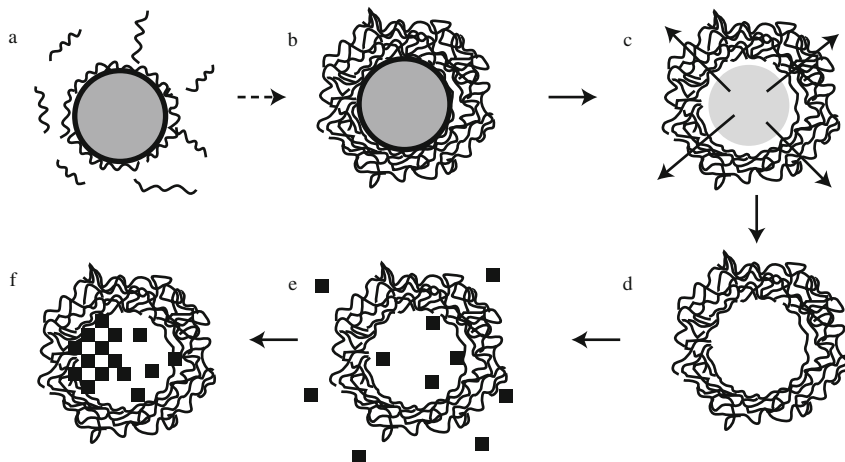


Fig. 21.15. Sequential adsorption of (a) positively charged and (b) negatively charged polyelectrolytes on negatively charged particles. After dissolving the core of the particle (c), hollow particles are obtained (d). By adding a dye to the suspension (e), the interior of the particles can be loaded by varying the ionic strength or the solvent composition. Taken from [43]

21.4 Making Hollow Particles from Multilayers

One can also cover colloidal particles with polyelectrolyte multilayers, dissolving the particle core after depositing the film. The fabrication process is shown schematically in Fig. 21.15.

The substrate particles can be synthetic (inorganic or organic) or biological. One of the first substrates used for this application was melamine formaldehyde [44], which is easy to dissolve in HCl. Another solution was to use protein precipitates, destroying the central core by means of sodium hypochlorite. One can also adjust the solubility or add a complexing agent like EDTA in the case of CaCO_3 or CdCO_3 [45]. The calcination of SiO_2 particles coated with a multilayer film produces hollow and porous particles, with a residual film of SiO_2 under the polyelectrolytes.

The wide range of materials that can be used to fabricate hollow particle membranes means that their permeability can be controlled. For applications, one often requires the controlled release over a long period, or fast and sporadic release, of some active ingredient. This release can be modulated over an order of magnitude in time by varying the film thickness, but a rapid release is easier to implement. Rapid release of the protein contents can be achieved by including proteins (enzymes) in crystalline form after gradual dissolution. Release can also be induced by rupturing the particle wall, e.g., by osmotic shock, while the architecture can be destroyed by simple variation of the ionic strength if low mass polyelectrolytes have been inserted. An

alternative solution includes heat-, light- or pressure-sensitive derivatives in these constructions.

Reversible and repeatable release can be obtained by modulating the film permeability. Another novel approach may be to include ion channels similar to those in cell membranes. The diffusion of bulkier species can also be induced by pH variations [46]. For example, fluorescein-tagged dextran molecules adsorb onto the outside of PSS/PAH capsules prepared at pH 8 and, by reducing the pH to 6, these molecules diffuse through the core of the particles. When the pH reaches 8, the tagged molecules are observed solely within the particles. Alternating the pH causes reversible reactions and the change in permeability results from a partial ionisation of the polyelectrolyte at pH 8, which becomes total at pH 6. Repulsion between the polyelectrolytes produces a certain level of porosity in the film constructed at pH 8, whence the dextran molecules are able to diffuse through the membrane.

Hollow capsules could have many applications provided that they could be easily loaded with active molecules. Chemical and physical methods have been imagined to this end. For example, a Donnan equilibrium can be set up using capsule walls that are impermeable to the polyelectrolytes. For such constructions, addition of a polyacid outside allows the diffusion of protons into the particles, inducing pH variations by 1 to 4 units. The pH difference causes porosity variations which could then be exploited to load hollow particles. A different way to imprison an enzyme is to adsorb it onto the substrate on which the film is built, but before depositing the first layers, or again to include it in the first layers of the architecture (toward the inside of the hollow capsule). If these first layers are made from degradable polyelectrolytes [47], enzymes can be release inside the capsules after dissolving the inner layers. A key feature of porous particles follows from the property of small molecules to diffuse through films, while this is not possible for higher molecular masses. For example, encapsulating enzymes can protect them against the reactions of high molecular weight proteases. This property has been demonstrated for chymotrypsin and horseradish peroxidase [48]. This form of protection would appear to be effective, but it is nevertheless accompanied by a reduction of biological activity. More recently, the idea has also been implemented by coating colloidal particles of nanometric dimensions with gold [49]. The core of these particles can then be dissolved using HF.

21.5 The Route to More Complex Architectures

As discussed earlier, films can be constructed with linear growth, e.g., PSS/PAH, or with exponential growth, e.g., HA/PLL. Apart from the significant difference in thickness after depositing a given number of bilayers, linear growth films are harder and do not allow the diffusion of proteins, polypeptides, or molecular complexes perpendicularly to the plane of the film. Conversely, exponential growth films allow significant diffusion of these same species, in

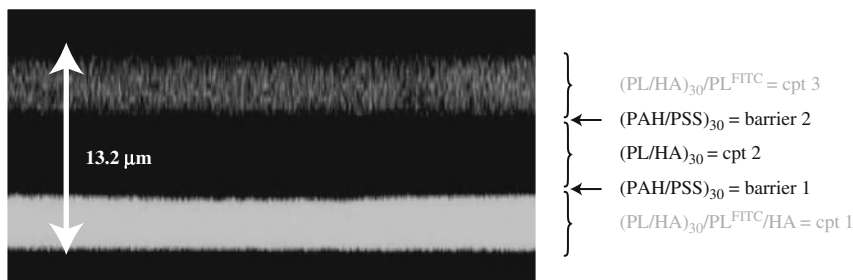


Fig. 21.16. Polyelectrolyte multilayer film formed by domains constituted by reservoirs obtained by alternating hyaluronic acid and polylysine (exponential growth film) and other domains (barriers) formed by alternating polystyrene sulfonate and polyallylamine (linear growth film). Adding PL^{FITC} to the last bilayer of the reservoir clearly indicates that the reservoir, but only that reservoir, completely fills with fluorescent polylysine. Note also how effective the PSS/PAH barriers are, preventing the diffusion of PLL from one reservoir to the next. Reproduced from [50] with the kind permission of the publisher. Copyright (2004) American Chemical Society

particular polypeptides and molecules of low molecular weight. These properties can be exploited to make compartmented films, by alternating domains of linear growth with others of exponential growth. This kind of architecture has been made by depositing $(\text{PSS}/\text{PAH})_m$ films on $(\text{HA}/\text{PLL})_n$ films [50]. Structures comprising three (HA/PLL) reservoirs separated by two (PSS/PAH) barriers have also been made (see Fig. 21.16).

By using fluorescein-tagged PLL to make the last PLL layer in the HA/PLL compartments 1 and 3, it has been shown that fluorescence is localised inside each compartment. These experiments show that each linear growth PSS/PAH film does indeed play the required role of barrier. The idea now is to load each compartment with different drugs which could then be released gradually. In addition, each reservoir could be loaded with just the right amount, depending on the width of the compartment and the concentration used in solution, to fill the corresponding compartment. One major problem that remains to be solved concerns degradation of the barrier films.

Another important way of loading polyelectrolyte films or multifunctionalising constructions is to include phospholipid vesicles in the films. For example, vesicles formed mainly from 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphatidylcholine (POPC) have been described in the literature [51, 52]. They were stabilised by coating them with a layer of poly-(D-Lysine) before inserting them in exponential growth polyelectrolyte multilayers. It has been shown that these vesicles can be integrated without rupture or coalescence (see Fig. 21.17).

These vesicles are therefore interesting tools for functionalising films. Multifunctionalisation can be achieved by mixing liposomes enclosing different reagents and including them within the architecture at some precise level in

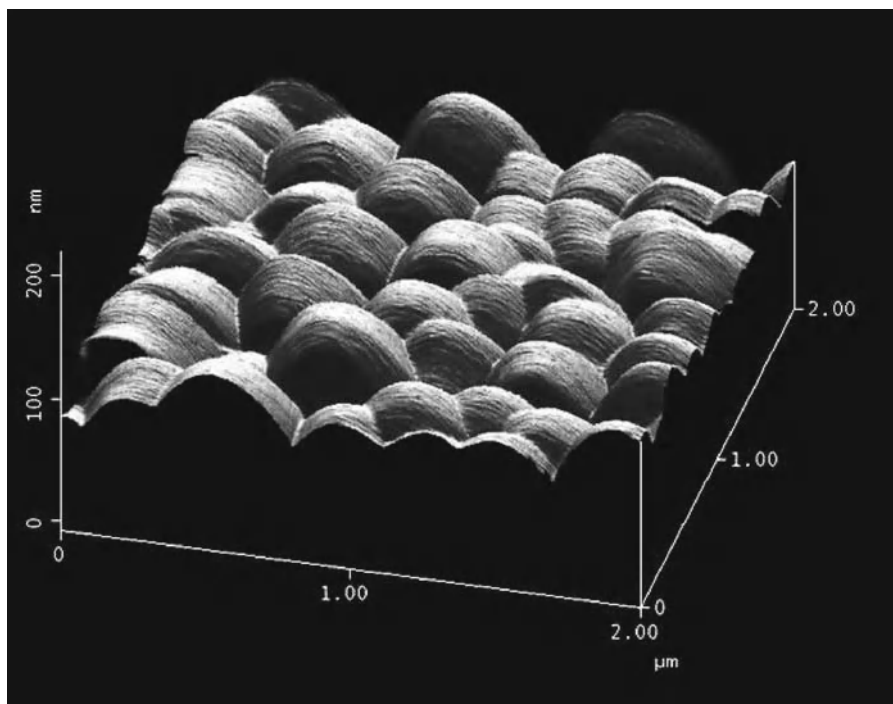


Fig. 21.17. AFM image of vesicles obtained using 1-palmitoyl-oleoyl-sn-glycero-3-phosphatidylcholine (POPC), 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphatidyl-DL-glycerol (POPG), and glycerol. The vesicles are stabilised by coating with PDL. The vesicles are coated with two PGA-PAH bilayers. Provided by M. Michel, Institut Charles Sadron (CNRS), Strasbourg, France

the construction. One can also envisage including liposomes containing different drugs or reagents in strata, at different levels throughout the structure, in the case of biomedical applications. There is then a problem in triggering the reaction. For biomedical applications, cells must be able to interact with the drugs, which requires an opening in, or gradual dissolution of, the vesicle envelope. Liposomes in films can also contain reagents producing unstable derivatives or derivatives with limited stability, but used by the medium as soon as it forms. In this case, the vesicles can be triggered to open by applying a force at the top of the structure, for example.

21.6 Prospects

Polyelectrolyte multilayers provide a very flexible tool for functionalising or multifunctionalising biomaterials. Studies carried out so far have mainly been concerned with proof of concept, but work is now aiming toward certain

specific applications, e.g., surfaces with antibacterial or anti-inflammatory properties. Another type of application concerns films in which growth factors or adhesion peptides are incorporated. Applications of polyelectrolyte multilayers to biomaterials should therefore see further growth over the next few years.

References

1. Decher, G., Hong, J.D., Schmitt, J.: *Thin Solid Films* **210**, 831–835 (1992)
2. Lavalle, P., Gergely, C., Cuisinier, F.J.G., Decher, G., Schaaf, P., Voegel, J.C., Picart, C.: *Macromolecules* **35**, 4458–4465 (2002)
3. Decher, G.: *Science* **277**, 1232–1237 (1997)
4. Fou, A.C., Onitsuka, O., Ferreira, M., Rubner, M.F., Hsieh, B.R.: *J. Appl. Phys.* **79**, 7501–7509 (1996)
5. Hiller, J., Mendelsohn, J.D., Rubner, M.F.: *Nature Materials* **1**, 59–63 (2002)
6. Stanton, B.W., Harris, J.J., Miller, M.D., Bruening, M.L.: *Langmuir* **19**, 7038–7042 (2003)
7. Rmaile, H.H., Schlenoff, J.B.: *J. Am. Chem. Soc.* **125**, 6602–6603 (2003)
8. Mamedov, A., Kotov, N.A., Prato, M., Guldi, D.M., Wicksted, J.P., Hirsch, A.: *Nature Materials* **1**, 190–194 (2002)
9. Schultz, P., Vautier, D., Richert, L., Jessel, N., Haikel, Y., Schaaf, P., Voegel, J.-C., Ogier, J., Debry, C.: *Biomaterials* **26**, 2621–2630 (2005)
10. Ladam, G., Schaaf, P., Voegel, J.C., Schaaf, P., Decher, G., Cuisinier, F.J.G.: *Langmuir* **16**, 1249–1255 (2000)
11. Schlenoff, J.B., Ly, H., Li, M.: *J. Am. Chem. Soc.* **120**, 7626–7634 (1998)
12. Shiratori, S.S., Rubner, M.F.: *Macromolecules* **33**, 4213–4219 (2000)
13. Picart, C., Lavalle, P., Hubert, P., Cuisinier, F.J.G., Decher, G., Schaaf, P., Voegel, J.C.: *Langmuir* **17**, 7414–7424 (2001)
14. Picart, C., Mutterer, J., Richert, L., Luo, Y., Prestwich, G.D., Schaaf, P., Voegel, J.C., Lavalle, P.: *Proc. Natl. Acad. Sci. USA* **99**, 12531–12535 (2002)
15. Lee, S.S., Hong, J.D., Kim, C.H., Kim, K., Koo, J.P., Lee, K.B.: *Macromolecules* **34**, 5358–5360 (2001)
16. Schlenoff, J.B., Dubas, S.T., Farhat, T.: *Langmuir* **16**, 9968–9969 (2000)
17. Porcel, C.H., Izquierdo, A., Ball, V., Decher, G., Voegel, J.-C., Schaaf, P.: *Langmuir* **21**, 800–802 (2005)
18. Mendelsohn, J.D., Yang, S.Y., Hiller, J., Hochbaum, A.I., Rubner, M.F.: *Biomacromolecules* **4**, 96–106 (2003)
19. Jessel, N., Atalar, F., Lavalle, P., Mutterer, J., Decher, G., Schaaf, P., Voegel, J.C., Ogier, J.: *Adv. Mater.* **15**, 692–695 (2003)
20. Chluba, J., Voegel, J.C., Decher, G., Erbacher, P., Schaaf, P., Ogier, J.: *Biomacromolecules* **2**, 800–805 (2001)
21. Benkirane-Jessel, N., Lavalle, P., Meyer, F., Audouin, F., Frisch, B., Schaaf, P., Ogier, J., Decher, G., Voegel, J.-C.: *Adv. Mater.* **16**, 1507–1511 (2004)
22. Benkirane-Jessel, N., Schwinté, P., Falvey, P., Darcy, R., Haïkel, Y., Schaaf, P., Voegel, J.-C., Ogier, J.: *Adv. Funct. Mater.* **14**, 174–182 (2004)
23. Elbert, D.L., Herbert, C.B., Hubbell, J.A.: *Langmuir* **15**, 5355–5362 (1999)
24. Richert, L., Lavalle, P., Vautier, D., Senger, B., Stoltz, J.-F., Schaaf, P., Voegel, J.-C., Picart, C.: *Biomacromolecules* **3**, 1170–1178 (2002)

25. Boulmedais, F., Frisch, B., Etienne, O., Lavalle, P., Picart, C., Ogier, G., Voegel, J.-C., Schaaf, P., Egles, C.: *Biomaterials* **25**, 2003–2011 (2004)
26. Muller, M., Rieser, T., Dubin, P.L., Lunkwitz, K.: *Macromol. Rapid Comm.* **22**, 390–395 (2001)
27. Ladam, G., Schaaf, P., Cuisinier, F.G.J., Decher, G., Voegel, J.-C.: *Langmuir* **17**, 878–882 (2001)
28. Szyk, L., Schaaf, P., Gergely, C., Voegel, J.-C., Tinland, B.: *Langmuir* **17**, 6248–6253 (2001)
29. Szyk, L., Schwinté, P., Voegel, J.-C., Schaaf, P., Tinland, B.: *J. Phys. Chem. B* **106**, 6049–6055 (2002)
30. Caruso, F., Furlong, D.N., Ariga, K., Ichinose, I., Kunitake, T.: *Langmuir* **14**, 4559–4565 (1998)
31. Schwinté, P., Voegel, J.-C., Picart, C., Haikel, Y., Schaaf, P., Szalontai, B.: *J. Phys. Chem. B* **105**, 11906–11916 (2001)
32. Benkirane-Jessel, N., Lavalle, P., Hübsch, E., Holl, V., Senger, B., Haikel, Y., Voegel, J.-C., Ogier, G., Schaaf, P.: *Adv. Funct. Mater.* **15**, 648–654 (2005)
33. Houska, M., Brynda, E.: *J. Coll. Int. Sci.* **188**, 243–250 (1997)
34. Brynda, E., Houska, M.: *J. Coll. Int. Sci.* **183**, 18–25 (1996)
35. Brynda, E., Houska, M.: In: *Protein Architecture: Interfacing Molecular Assemblies and Immobilization Biotechnology*, ed. by H. Moehwald, Marcel Dekker, New York (2000) pp. 251–285
36. Serizawa, T., Yamaguchi, M., Matsuyama, T., Akashi, M.: *Biomacromolecules* **1**, 306–309 (2000)
37. Serizawa, T., Yamaguchi, M., Akashi, M.: *Biomacromolecules* **3**, 724–731 (2002)
38. Richert, L., Boulmedais, F., Lavalle, P., Mutterer, J., Ferreux, E., Decher, G., Schaaf, P., Voegel, J.-C., Picart, C.: *Biomacromolecules* **5**, 284–294 (2004)
39. Schuler, C., Caruso, F.: *Macromol. Rapid Comm.* **21**, 750–753 (2000)
40. Caruso, F., Schuler, C.: *Langmuir* **16**, 9595–9603 (2000)
41. Ariga, K., Kunitake, T.: In: *Protein Architecture: Interfacing Molecular Assemblies and Immobilization Biotechnology*, ed. by H. Moehwald, Marcel Dekker, New York (2000) pp. 169–191
42. Onda, M., Ariga, K., Kunitake, T.: *J. Biosci. Bioeng.* **87**, 69–75 (1999)
43. Sukhorukov, G., Dähne, L., Hartmann, J., Donath, E., Möhwald, H.: *Adv. Mater.* **12**, 112–115 (2000)
44. Gao, C.Y., Moya, S., Lichtenfeld, H., Casoli, A., Fiedler, H., Donath, E., Moehwald, H.: *Macromolecular Materials and Engineering* **286**, 355–361 (2001)
45. Antipov, A.A., Sukhorukov, G.B., Donath, E., Mohwald, H.: *J. Phys. Chem. B* **105**, 2281–2284 (2001)
46. Sukhorukov, G.B., Antipov, A.A., Voigt, A., Donath, E., Mohwald, H.: *Macromol. Rapid Comm.* **22**, 44–46 (2001)
47. Radtchenko, I.L., Sukhorukov, G.B., Mohwald, H.: *Colloids and Surfaces A: Physicochemical and Engineering Aspects* **202**, 127–133 (2002)
48. Tiourina, O.P., Antipov, A.A., Sukhorukov, G.B., Larionova, N.L., Lvov, Y., Mohwald, H.: *Macromolecular Bioscience* **1**, 209–214 (2001)
49. Schneider, G., Decher, G.: *Nano Letters* **4**, 1833–1839 (2004)
50. Garza, J.M., Schaaf, P., Muller, S., Ball, V., Stoltz, J.F., Voegel, J.-C., Lavalle, P.: *Langmuir* **20**, 7298–7302 (2004)
51. Michel, M., Izquierdo, A., Decher, G., Voegel, J.-C., Schaaf, P., Ball, V.: *Langmuir* **21**, 7854–7859 (2005)
52. Michel, M., Vautier, D., Voegel, J.-C., Schaaf, P., Ball, V.: *Langmuir* **20**, 4835–4839 (2004)

Biointegrating Materials

J. Amédée, L. Bordenave, M.-C. Durrieu, J.-C. Fricain, and L. Pothuaud

22.1 Cell and Tissue Engineering

The extraordinary increase in human longevity explains the growing need for replacement organs. The remarkable successes of conventional transplants (associated with the development of effective antirejection drugs and improved control of their administration) are also accompanied by certain drawbacks. First on the list is an inadequate supply of replacement organs: the number of candidates for transplants grows larger, opposition to the removal of organs increases, and the number of transplants has reached a ceiling. Furthermore, it has come to light over the past few years that organ transplants carry a significant risk of transmitting pathogens. Finally, the main drawback lies in the need to pursue an immunosuppression treatment. Scientists and doctors have long been in search of alternatives to human organ transplants.

According to the definition drawn up in Chester in 1986 at the Consensus Conference organised under the aegis of the European Society for Biomaterials, biomaterials are non-viable materials used in a medical device and destined to interact with biological systems, whether they contribute to the constitution of a diagnostic device, a tissue or organ substitute, or a device designed to provide functional assistance or replacement.

From single-use medical equipment to permanently implanted prostheses, a wide range of products involve biomaterials. The biomaterials sector includes 4,000 different products, resulting from technology and materials generally developed for other purposes, and it is interesting to note that few if any biomaterials have arisen from technology specific to biomedical applications [1]. The common feature of biomaterials is biocompatibility, and the multiparametric character of this property makes it difficult to measure or assess objectively.

When biomaterials and implantable materials are used, the level of biocompatibility required varies depending on the device. The relevant parameters are the total time span over which the material is confronted with the patient organism, the anatomical site, and the area of the surface in contact with

biological tissue. The biocompatibility of an implant is the extent to which it is accepted by the host tissue, or indeed integrated within the latter, whence the term 'biointegration' commonly used to refer to the ideal result that few (if any) synthetic biomaterials are able to achieve.

Clearly, the observable result is conditioned by the evolution of the acute inflammatory reaction resulting from implantation and by the behaviour of cells in tissues surrounding the implant. The materials must have structural properties suited to the intended function and surface properties guaranteed to ensure positive interactions at the material–tissue interface. Depending on the kind of material, the structure and physicochemical properties of its surface, and its micromorphology, cells will be able to adhere to the material, spread out, migrate, proliferate, synthesise and secrete the components of a new extracellular matrix, and contribute to the construction of a new tissue within which the implant is perfectly integrated. This integration may or may not be accompanied by resorption of the implant.

To endow materials with suitable means for integrating within host tissues, the idea of tissue engineering is to associate within the same system an artificial component of synthetic or natural origin (construed from one or more rationally functionalised biomaterials) and a cell or tissue component of autologous origin (one or more cell types), suitable for taking part in the formation of replacement tissue, or repairing or ensuring the function of an organ that needs to be replaced. This is in fact a rather recent field of research and development in biotechnology, combining various features of medicine and biology, in particular, cell biology and molecular biology, materials science, and engineering, with the goal of regenerating or replacing diseased or damaged tissue. Tissue engineering is a rapidly growing interdisciplinary field, at the crossroads between the life sciences and engineering, which uses biomimetism. Among other things it draws upon the progress achieved by specialists in developmental and cell biology (the possibility of differentiated cell cultures) and progress made by specialists in the field of biomaterials and process engineering. Once the choice of support and the choice and implementation of the cell material have been established, the development of such strategies should lead to new families of products, alongside drugs, medical devices, and organ or tissue transplants (whose limitations and risks are well understood), which doctors and surgeons will be able to use to the benefit of their patients to improve their quality of life.

Tissue engineering undoubtedly represents the next generation of biomaterials and implantable materials. Treatments for disease or damage to skin, cartilage, bone, cardiac valves, and blood vessels feature among current applications in this area of regenerative medicine. With this in mind, tissue engineering aims to fulfill three goals:

- To develop in situ tissue regeneration or repair processes, stimulated or induced by supplying proficient cells, or suitable biologically active factors (the approach known as cell therapy). Note that these factors could be

supplied in a self-controllable way by cells that are not necessarily the ones taking part in the tissue reconstruction, or by intelligent substrates.

- To create in vitro draft versions of organ or tissue substitutes (for implantation), combining a matrix support and cells (usually of autologous origin).
- To create in vitro bioartificial equivalents of organs or tissues designed to serve as study models in pharmacology or cosmetology.

There are two main types of design for getting cells to recognise biomaterials on the molecular scale. The first endows biomaterials with the property of bioactivity, by incorporating soluble biologically active molecules like growth factors, antibodies, and so on, into the substrates, in order to subsequently release these molecules and thereby trigger or modulate new tissue formation. The other approach is to incorporate proadhesive peptides in the biomaterials by chemical or physical modifications, during surface or core modifications to the biomaterials [2].

22.2 Modifying Material Surfaces

22.2.1 Using Nanoparticles to Deliver Active Ingredients

The general aim is to endow implantable devices with the ability to oppose the development of infections and/or inflammatory processes following their implantation. Today, to alleviate these effects, a drug is administered generally, while antibiotics are administered locally in bone surgery:

- When a drug is administered generally, it is distributed throughout the organism and its concentration at the target site, i.e., at the location of an infectious, inflammatory, or neoplastic process, can only exceed the efficiency threshold if the dose administered is high enough, at risk of exposing the patient to toxic side-effects. Substances administered orally or parenterally often have too short a half-life to achieve the desired local effect and represent a significant risk of toxicity.
- Since 1970, antibiotic cements have been used in articular prosthetic surgery. Ten years after their introduction [3], a Swedish study showed that in 1688 patients with a full hip replacement, only the gentamicin-containing cements were as effective as systemic antibioprophyllaxis. The cement impregnated with antibiotics is widely used in Europe [4]. In France, two preparations are commercialised, using either gentamicin or an association of erythromycin and colimycin. An antibiotic cement can also be made with vancomycin, in the operating theatre under non-standardised conditions. The limiting factor with this method is the uncontrolled release (in terms of concentration and duration) of the active ingredient. To allay these difficulties, drug delivery systems (DDS) have been developed. The idea is to deliver pharmacologically active substances

in situ, at a steady rate over prolonged periods, and in sufficient amounts without being toxic.

What are the limitations of existing methods or products in satisfying these needs?

- There is no control over the kinetics underlying the release of the active ingredient since there is no device that could be used to adjust the rate of release and hence perpetuate its action over some predetermined period.
- Part of the active ingredient is not released since it is trapped too deeply within the cement.

Stimuli-responsive polymers with reactive functions obtained by encapsulation or adsorption of active ingredients directly within the material or in beads, themselves adsorbed or grafted onto the material, have already been described [5, 6]. However, adsorption does not allow controlled release of the active ingredient. With regard to encapsulation, although it does allow controlled release of the active ingredient, it is nevertheless incompatible with prolonged use and/or situations in which the material is subjected to large stresses, e.g., flow, friction, etc.

For effective local delivery of drugs in humans, the main challenges are: (1) the choice of active ingredient, (2) assessment in the animal of the systemic dose previously established as being locally effective, and (3) identification of the biocompatible vector capable of delivering the active ingredient.

A Detailed Example: Design and Preparation of Surfaces for Controlled Delivery of an Active Ingredient

Ideally, an implant should be able to release a selected active ingredient, covalently bound to its surface, at the site where the biomaterial is implanted, and for an adjustable length of time (see Fig. 22.1) [7, 8].

Nanoparticles (see Fig. 22.2) undoubtedly provide the perfect solution because they increase the specific surface area of the material, which guarantees a high enough concentration of bioactive molecules, while at the same time providing a simple way of introducing several chemical functions or active ingredients at the surface of the biomaterial (see Fig. 22.1).

Nanoparticles are synthesised in a single step [9–11]. These polymer nanoparticles are responsive and have acid, amine, alcohol, acid chloride, etc. reactive functions (F) on the outside which allow them to bind onto the surface of the prefunctionalised biomaterial (see Fig. 22.3). These particles are obtained by ring-opening olefin metathesis polymerisation (ROMP) in a dispersed medium (emulsion, mini- and microemulsion, dispersion, suspension) of cyclic olefins with macromonomers α - ω -functionalised by a polymerisable entity and a reactive function (A) or an active ingredient (B) (see Fig. 22.3).

The basic monomer is a cyclic olefin called norbornene. Highly reactive in metathesis polymerisation, it leads to high polymer yields. Furthermore,

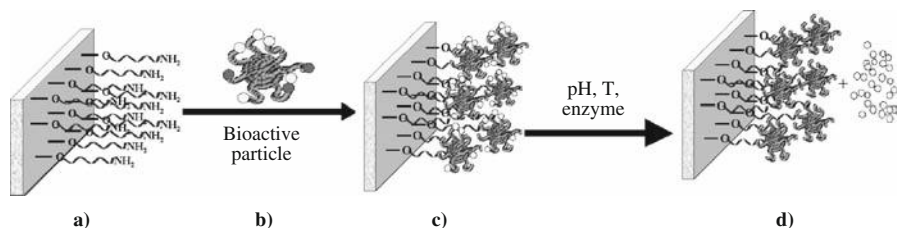


Fig. 22.1. Schematic view of the surface of a biomaterial designed to deliver active ingredients. (a) Functionalisation or grafting of the biomaterial. (b) Synthesis of nanoparticles functionalised by the active ingredient. (c) Binding the nanoparticles onto the biomaterial. (d) Release of the active ingredient

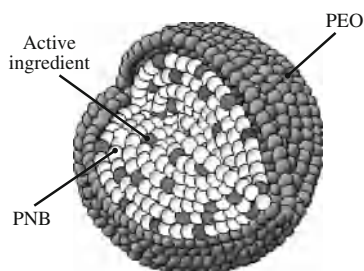


Fig. 22.2. Schematic view of the structure of nanoparticles functionalised by an active ingredient

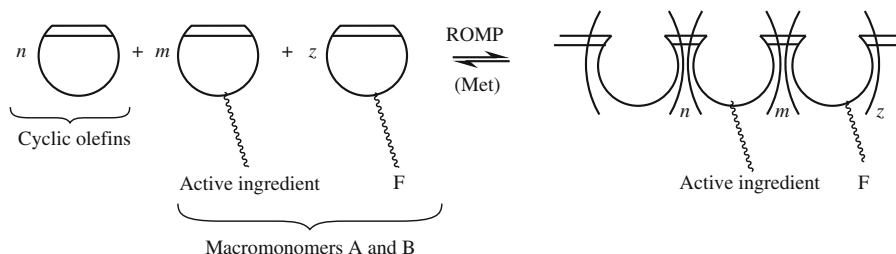


Fig. 22.3. Synthesis of bifunctionalised nanoparticles

polynorbornene (PNB) is a non-cytotoxic polymer and its use has already been envisaged to make tracheotomy probes. The macromonomers A and B are oligomers of polyethylene oxide (PEO). They are functionalised at one end by a norbornenyl unit, an entity chosen for its high reactivity in metathesis polymerisation, and at the other by an acid reactive function (A) or the active ingredient (B) via a cleavable amide bond.

Particles are covalently bound onto the material by condensation of two antagonistic reactive functions, one on the material and the other on the particles. Examples are the pairs acid/amine, acid/alcohol, acid/acid chloride, etc. Prior to this binding, it is clearly necessary to modify the surface of the material in order to bind the desired functions onto the surface. Functionalisation creates reactive chemical functions at the surface of the substrate,

i.e., functions that can be used to covalently bind another molecule (see Sect. 22.2.2) [12–21].

A reaction cleaving the bond between particle and active ingredient, brought about by contact between the material and the physiological medium, a change in pH, or a local change in temperature, would release the native form of the bioactive molecule in a controlled way (see Fig. 22.1).

22.2.2 Macroscale Functionalisation of Biomaterial Surfaces

Over the last few decades, scientists in the areas of materials science, surface engineering, chemistry, physics, biology, biochemistry, and medicine have been working on the problem of functionalising materials with a view to obtaining specific cell–material interactions. Among the methods used, one category concerns functionalisation based on non-specific chemical modifications. This is the case when a plasma is used in the presence of various atmospheres, the nature of which conditions the resulting modification. It is also the case when a surface is activated by ionising electromagnetic radiation (β or γ rays), or ionising particle bombardment (by electrons or ions). The result is often difficult to control and multifunctional. The introduction of carboxylic acid or amino acid functions, fluorine-bearing groups, etc., is standard and used in industry [22, 23]. The resulting modifications can improve cell adhesion and also repel proteins or cells depending on the functional group or groups that have been introduced. For example, an Arg-Gly-Asp (RGD) sequence favours the biospecific binding of cells [24], while PEG grafts reduce the interaction with proteins and platelets [25].

More specific activations can be achieved by exploiting the coupling chemistry of functional groups initially present or introduced as intermediates for synthesis. For example, the presence of carboxylic acid or primary amine groups leads to coupling with primary or secondary amine compounds or carboxylic acid, respectively, with the help of coupling agents such as dicyclo-carbodiimide or DCC. In the literature, there are many examples of coupling involving peptides, especially peptides carrying the RGD sequence, enzyme proteins, peptide hormones, antibodies, and so on.

A Detailed Example: Functionalising Materials with RGD Peptides

On the fundamental and experimental levels, the design of hybrid artificial organs associating a cell or tissue component and an artificial matrix component of natural or synthetic origin, provides a way of improving the biointegration properties of these substitutes.

The sequence Arg-Gly-Asp (RGD) is by far the most effective and most widely used peptide sequence for stimulating cell adhesion at the surface of materials. Since the discovery that this peptide sequence had a mediating role in cell attachment [26], many scientists have deposited peptides containing

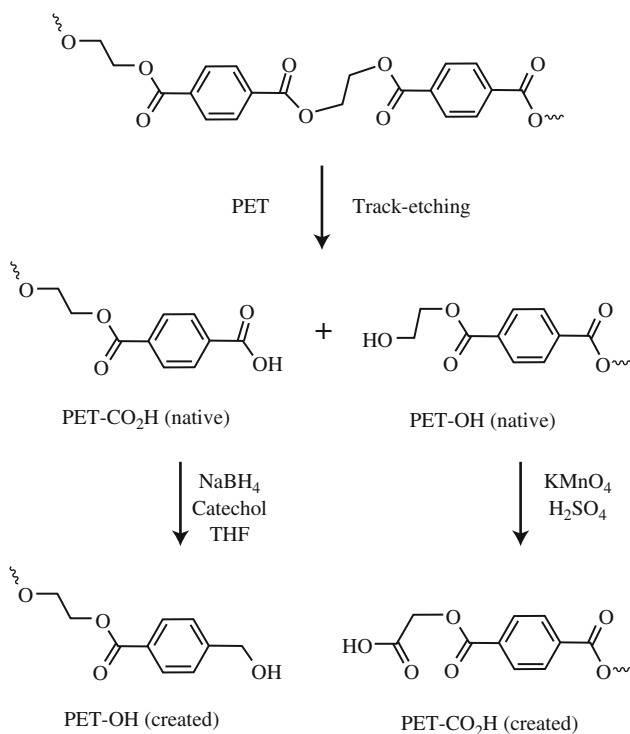


Fig. 22.4. Functionalising PET by track etching and chemical treatments [68]

this particular sequence on the surface of biomaterials to promote cell attachment [18–20, 27–31]. Membrane receptors belonging to the integrin superfamily recognise this RGD sequence and favour attachment. The integrins [32] are heterodimers comprising two subunits, α and β , which contain specific binding sites in their extracellular portion, whose activity requires the presence of calcium, and which recognise the tripeptide consensus sequence Arg-Gly-Asp (RGD). This motif is the central structure of the ligands of all members of the integrin family, the amino acid sequence surrounding the RGD motif determining the specificity of each protein in the matrix for one or more integrins. Cell adhesion achieved via the integrins involves a cascade of four events [33]: cell attachment, cell spreading, organisation of the actin cytoskeleton, and the formation of focal contacts.

Strong cell adhesion is only possible if the RGD-containing peptides have a stable bond with the surface. Indeed, focal adhesions can only form if the ligands can withstand the contractile forces of the cells [34–36]. Mere adsorption of RGD-bearing peptides on the surface only provides a very weak cell attachment [37–40]. Moreover, the drawbacks with adsorption techniques are the relatively large amount of proteins/peptides, their possible unwanted release and/or the diffusion of bioactive molecules far from the tissue–implant

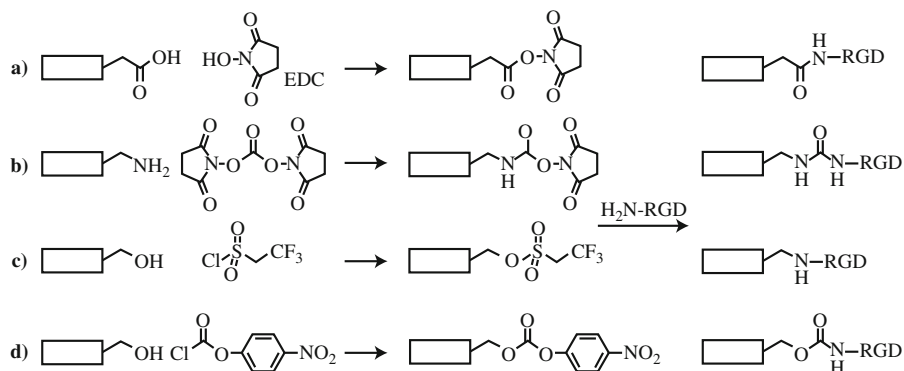


Fig. 22.5. RGD-containing peptides react with the various groups present at the material surface by virtue of their N-terminal function. (a) Carboxylic functions preactivated with a carbodiimide and NHS to produce an active ester. (b) Amine groups preactivated with DSC. (c) Hydroxyl groups preactivated to form tresylates. (d) Hydroxyl groups preactivated to form p-nitrophenyl carbonate [68]

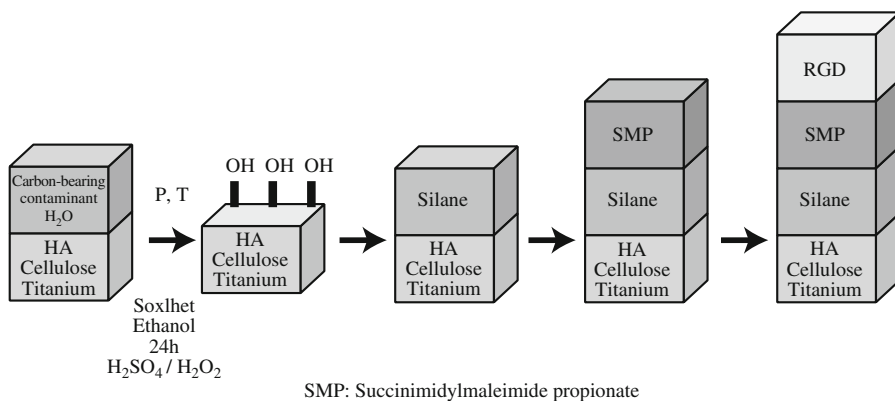


Fig. 22.6. Strategy for functionalising materials by RGD-containing peptides via two intermediate molecules [19]

interface where these biomolecules are required. In contrast, coupling techniques can induce immobilisation and orientation of biomolecules, whence it is possible to obtain a faster and more specific physiological response by exposing the appropriate active sites at the interface.

To obtain stable peptide/surface bonding, the RGD-bearing peptides must be grafted covalently onto the material, i.e., via functional groups such as carboxylic, hydroxyl, or amine functions. Given that not all materials have such functional groups on their surface, these must be introduced by physical or chemical treatments, hydrolysis in an alkaline medium, reduction, or oxidation.

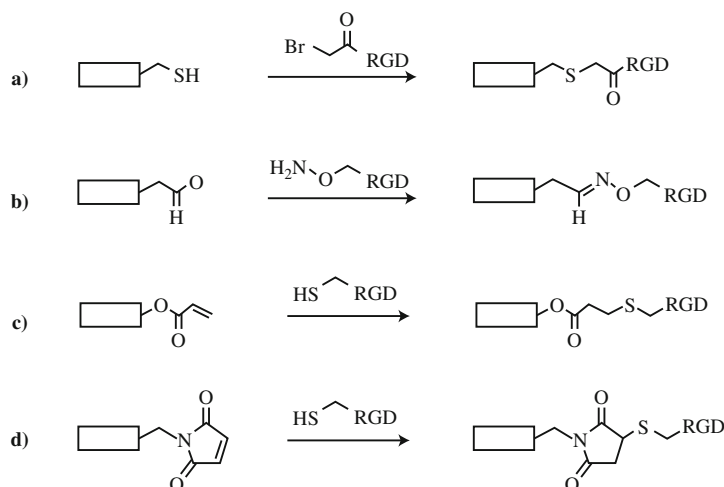


Fig. 22.7. Chemical coupling between two antagonistic functions. (a) Reaction between a thiol function and a bromoacetyl function of the peptide. (b) Reaction between an aldehyde function and an aminooxy function of the peptide. (c) Reaction between an acrylate function and a thiol function of the peptide. (d) Reaction between a maleimide function and a thiol function of the peptide [68]

In the case of polymers, the functional groups needed to immobilise peptides can be generated by blending with other polymers which do contain the functional groups. For example, poly(tetrafluoroethylene) (PTFE) [41], polystyrene (PS) [42], and poly(lactic acid) [43] have been coated with poly(L-lysine) (PLL) by adsorption. The free amine functions of PLL can be used for other modifications.

For example, hydroxyl functions are created at the surface of poly(vinyl acetate) by hydrolysis [44] in an alkaline medium. Aminolysis of poly(γ -methyl L-glutamate) by diamines such as hydrated hydrazine, ethylenediamine, or hexamethylenediamine produces surfaces functionalised by amine functions, which can be used to immobilise peptides [45]. The carbonyl groups present at the surface of poly(arylether ether ketones) (PEEK) can be reduced by means of sodium borohydride to generate hydroxyl functions with a view to further functionalisation [46].

Poly(ethylene terephthalate) (PET) can be treated with sodium borohydride to obtain a uniform functionalisation by hydroxyl functions, or oxidised to obtain carboxylic terminal functions (see Fig. 22.4) [47]. Subsequent modifications can be used to obtain materials with different chemical functions, e.g., NH_2 , NCS, NCO, OSO_3H .

In most cases, RGD peptides are bound to materials via a stable amide bond. This is usually created by means of a reaction between the active carboxylic functions on the surface of the material and the terminal nucleophilic nitrogen of the peptide. The carboxylic functions can be activated

by a coupling agent such as 1-ethyl-3-(3-dimethylaminopropyl)-carbodiimide (EDC, also referenced as a water-soluble carbodiimide WSC), dicyclohexylcarbodiimide (DCC), or carbonyl diimidazole (CDI) (see Fig. 22.5).

In order to functionalise titanium surfaces with bioactive organic molecules, titanium materials can be treated chemically so as to clean them in a reproducible way and reoxidise them [19, 48]. Controlled oxidation is beneficial, since the oxide film and the hydroxyl functions are required for subsequent coupling of molecules and biomolecules. Silane and thiol molecules are probably the most widely used for functionalising metal surfaces. A biomimetic modification of a titanium alloy (TA6V) has been carried out under controlled atmosphere, temperature, and pressure, in order to avoid the presence of physisorbed water or contamination by carbon-bearing compounds [19]. One approach for grafting peptides is to modify the hydroxyl sites at the surface and involves grafting an aminosilane on the titanium surface, binding a bi-functional intermediate agent, viz., succinimidyl maleimidopropionate (SMP), onto the amine functions, and binding an RGD peptide (C is cysteine) onto the maleimide function (see Fig. 22.6) [19].

Figures 22.5 and 22.7 show different possible ways of binding RGD peptides onto the surface of materials.

22.2.3 The Relevance of Controlled Nanotopochemistry and Nanodomains

When designing implantable medical devices able to ensure long-lasting functional service, one is always on the look-out for biomaterials that will facilitate the biointegration of such devices. The word ‘biointegration’ is taken to mean *the absence of chronic inflammatory reaction* such as might follow the acute inflammatory reaction resulting from the implantation of the device, and *the reconstruction of host tissues* around constitutive biomaterials (or within them if they have adequate porosity), according to a scheme in conformity with the original model of those tissues. Such a result is obviously conditioned by the behaviour of peri-implant cells, for which any intrusive biomaterial should immediately constitute a base for construction of the desired tissues.

This tissue construction involves several interdependent processes:

- Cleansing of the site by polynuclear neutrophils.
- Adhesion, spreading, migration, proliferation, matrix synthesis by cells specific to the relevant tissue.
- Recruitment of endothelial cells or their precursors for reconstruction of neovascularisation.

All these processes can only get started, run, and achieve the expected result if the relevant cells have spatial and molecular relations with the biomaterial similar to those they have with their natural matrix. Apart from its role in regulating the availability of various cytokines and growth factors, this matrix

determines the behaviour of cells and their phenotypic expression by exposing proadhesive ligands that are specifically recognisable by the cell integrins. The adhesion cells then find themselves in a strain field, whose spatial distribution they optimise by an organised concentration of membrane receptors (a phenomenon known as clustering), while at the same time exploiting the plasticity of the extracellular matrix. This optimised strain field in its turn induces mechanotransduction effects which stimulate the phenotypic expression of the cells.

There are different strategies for obtaining the adhesion of differentiated cells on a biomaterial. For example, a matrix protein known for its proadhesive properties can be readsorbed onto the biomaterial. However, the lack of plasticity of the substrate limits the possibilities for optimising the strain field produced by adhesion bonds established by the cells, and from this point of view it is easy to appreciate the advantage of a substrate with some level of relief, i.e., able to provide a 3D environment for the cells. Another strategy is to biofunctionalise the substrate by associating a proadhesive peptide. The latter can theoretically be distributed uniformly with optimal density, or better distributed in a heterogeneous way in packets according to some topographically determined distribution.

These considerations show why it is useful to design materials with a surface whose topography and topochemistry are defined on the same scale as the cells and molecules involved in the processes determining tissue reconstruction around these materials. A great deal of work [49] on materials endowed with surface relief in the form of grooves, spots, or pits of micrometric or sub-micrometric dimensions, distributed according to some predefined geometric layout, has demonstrated the sensitivity of cells scattered on the surface of these materials to such topographic details. Note, however, that apart from their influence on the spatial environment of the cells, these details can also be distinguished by local physicochemical features that differ from those of the rest of the material surface. Indeed, other studies [49], although less common because harder to actually carry out, have demonstrated the influence of the local physicochemical characteristics of the surface on cell behaviour. An example is provided by the segregation effects characterising the microstructure of poly(ether urethane urea) (PEUU). These underlie the formation of microdomains revealing themselves at the surface of these polymers and distinguished by their corresponding surface free energy. By varying the molecular mass, i.e., the chain length, of the polyethylene glycol contributing to the structure of these PEUU, Lyman showed [49] that the surface free energy of the latter barely varies. On the other hand, the size of the microdomains resulting from the segregation of the PEG segments (soft segments) does vary, and this could explain the quantitative and qualitative differences between the adsorption of proteins from serum observed with these materials, as well as the different behaviours of cells cultivated on these materials, since it is known that this behaviour is sensitive to the nature and conformation of the adsorbed proteins. However, in this particular example, the nanoscale

surface topochemistry, which seems indirectly to influence cell behaviour, is characterised by a statistical distribution of two types of nano/microdomains, whose sizes are determined by the primary structure of the original polymer. It was immediately clear that it would be interesting to have materials in which the surface nano/microtopochemistry was characterised by a distribution, conforming to some predetermined but variable geometry, of two types of nano/microdomains, differing from one another by their surface free energy, or better, by their exposure or otherwise to specific ligands. These new study models would provide a way of defining the distribution and functionality of the nano/microdomains most favourable to the adhesion, proliferation, or differentiation of any given type of cell. By judicious choice of the type of functionalisation for each type of microdomain, one can reasonably hope to obtain the selective adhesion of one cell type on one of them and another cell type on the other. This would then provide coplanar coculture models.

Considering the plasticity of the extracellular matrix, the product and natural support of the cells, whose positive role has already been described, one can envisage artificial substrates providing cells with a support endowed with similar advantages. Indeed, it is possible to graft a polymer characterised by a glass transition temperature T_g close to 37°C onto the surface of a polymer material, the most favourable situation, but also onto the surface of other materials [49]. Topographically determined functionalisation of the grafted layer at a temperature below this value of T_g , by ligands that are proadhesive with respect to the cells, should lead to a support that they could adapt to more easily from a mechanical standpoint, to the benefit of their phenotypic expression in standard culture conditions at 37°C . Note also that this surface plasticity on the molecular scale could be usefully combined with the favourable consequences of a controlled microtopography.

22.3 Applications of Biointegrated Biomaterials

22.3.1 Applications to Bone Tissue

The materials used for bone reconstruction must conform to the expectations of clinical medicine, which requires an easily available material. It must be free of all risk to the host and allow tissue development and/or local and deferred release of pharmacologically active substances, designed to modulate the inflammation response, reduce the risk of infection, or stimulate the formation of new bone. The design and fabrication of such materials for bone reconstruction must also take into account a whole set of requirements regarding the surface state of the material, relatively high porosity to allow fast vascularisation of the implant, and good mechanical properties. Indeed, for the implantation of bone substitutes, and more particularly, in the case of macroporous materials, their successful integration is closely linked to the neovascularisation of these implants. In this context, one of the main difficulties of bone tissue engineering is to recreate large volumes of bone tissue,

given that the implants or transplants must be rapidly vascularised to ensure, in particular, the survival of cells able to reconstruct the tissue (supply of nutrients, oxygen, etc.). However, it is important to mention that, outside the fields of diagnosis and imaging, applications of nanoparticles to bone reconstruction remain at the exploratory stage today. On the other hand, it is clear that these are important issues for developing innovative materials.

One strategy that has been explored experimentally is the functionalisation of materials by nanoparticles loaded with active ingredients such as growth factors that can activate the vascularisation of implants. Among these factors, it seems relevant today to use VEGF angiogenic factors or oligopeptides corresponding to their sequence [50] to functionalise a biomaterial, in order to favour bone formation and vascularisation of the implant. Although the feasibility of this idea has been demonstrated on a fundamental level [50], the way these growth factors can be conveyed and delivered remains problematic. The use of hydroxyapatite particles as vectors is beginning to receive description in the literature [51].

A coprecipitation method has been used to prepare nanospheres loaded with osteoinductive proteins, viz., bone morphogenetic proteins (BMP). Three polymers with different charges have been used as packaging material for making the BMP delivery device, in the presence of phosphatidylcholine as stabiliser [52].

A more conventional research program today seeks to use nanoparticles to deliver antibiotics [53]. Recently, calcium sulfate nanoparticles have been used as delivery systems for vancomycin and gentamicin to alleviate the risks of infection due to implantation of materials [54]. In the same way, antibiotics have been associated with calcium carbonates (aragonite) with different degrees of porosity, allowing the gradual and local diffusion of active ingredients: metronidazole used in parodontology or gentamicin sulfate, used in orthopedics to treat centers of infection.

A second approach for improving the biointegration of bone substitutes involves associating osteogenic cells with these materials, capable of constructing mineralised tissue within the matrix. The surface state of the material in part conditions the response of cells attempting to colonise it, and the development of a bioactive surface will be crucial for the osteoconductive properties of the material. Indeed, any surface modification must respect a great many biological, mechanical, physical, and chemical demands. It is important to remember that the first biological events associated with the implantation of a biomaterial involve the adsorption of passing proteins. Indeed, at the first contact with the surrounding medium, water molecules, ions, and solutes with low molecular weights quickly adsorb onto the surface. The adsorption of proteins is the result of all these solid-liquid interactions. These phenomena partly condition cell adhesion and spreading. It is thus important to control the surface state of the material used in implantology and to provide a bioactive surface for the host tissue and cells. Surface treatments for modifying the topography, roughness, and chemistry of the surface, or for immobilising

ligands on the surface, are strategies used in materials science to favour the interaction of the material with the living medium.

The functionalisation of materials by adhesive ligands such as proteins in the extracellular matrix [55] is a line of research that has been widely explored as a way of stimulating cell adhesion, spreading, and colonisation with regard to these materials [56]. This strategy may eventually be improved with the help of nanotechnologies, or by developing macrodomains able to recruit different cell populations involved in the reconstruction and remodelling of bone tissue. The creation of macrodomains of different diameters, e.g., 10, 50, 100, and 200 μm , within which different peptide sequences, possibly containing the arginine-glycine-aspartic acid (RGD) sequence, could be grafted, should provide a way of selecting a specific sequence of a given cell type and the optimal domain size for promoting cell adhesion [57].

Finally, surfaces with rough microtopographies and peptide sequences seem to have a strong effect on osteoblast activity [58]. Indeed, Tosatti et al. [59] have studied differentiation in an osteoblast line (MG63) cultivated on smooth and rough surfaces modified by adsorbed monomolecular layers of poly(L-lysine)-graft-polyethylene glycol (PLL-g-PEG) functionalised by the RGD-containing peptide GCRGYGRGDSPG. Their results show increased differentiation on surfaces associating rough microtopography with these peptide sequences.

Catledge et al. [60] have demonstrated the interest in using nanostructured cobalt–chromium or titanium alloys to promote the adhesion of osteogenic mesenchymal cells and improve the biointegration of orthopedic and dental implants. In the same way, Webster et al. [61] have demonstrated significant adhesion of osteoblast cells on nanophase CoCrMo, $\text{Ti}_6\text{Al}_4\text{V}$, and Ti alloys as compared with unstructured alloys.

22.3.2 Applications to the Vascular System

One of the first causes of mortality in western countries is the deterioration of the vascular system, of multifactorial origins. This slow and irreversible disorder leads to either the obstruction or the dilation of vessels, putting the functions and even the life of the patient at risk. Current solutions treat only symptoms, restoring the blood circulation by replacing the defective length of artery by a synthetic or biological vascular substitute.

For several decades now, vascular surgeons have been using polyester knitted or woven tubular ducts for arterial revascularisation, and more recently expanded polytetrafluoroethylene, despite the fact that these materials do not fully comply with the specifications for the ideal vascular graft. For clinical use, all available materials are potentially thrombogenic and any improvement observed in their functional longevity, still largely inadequate, is due to progress made by the antithrombosis therapies to which patients are submitted.

The prostheses currently used have proven their adequacy in the aortic and aorto-iliac positions, but with uncertain durability for small calibre revascularisations (less than 6 mm in diameter): coronary and peripheral arteries under the crural arch, which represents some 35,000–40,000 operations a year in France. For femoro-popliteal or leg bypasses, the autologous internal saphenous vein is the most effective substitution material, provided that it is still available and of good enough quality. Apart from these cases, there is no choice but to use prosthetic revascularisation, even though it is well known that permeability is significantly better with the saphenous vein.

The development of vascular substitutes able to provide permanent functional service without the help of antithrombosis therapies is still a major challenge in the twenty-first century [62,63]. Any solution must provide a way of controlling the relation between the blood and the wall of the prosthetic duct, and this control will be achieved by developing materials that combine surface properties ensuring their hemocompatibility and structural properties suited to the fabrication of tubular ducts satisfying the criteria of mechanical behaviour required of a vascular substitute.

Several strategies are conceivable [64] for improving the long-term permeability of vascular grafts, and research in this area appeals to chemistry, physics, and biology, one of the common aims being to reduce the thrombogenic character of the surface of the prosthetic duct, and another to reduce the intimal hyperplastic reaction that tends to occur in the distal perianastomotic region.

Among the many different lines of research currently underway, one that predominates seeks to imitate the natural vascular wall in its ability to control thrombogenesis. An attractive solution would appear to be to line the endoluminal surface of arterial prostheses, prior to implantation, by a new endothelium created from endothelial cells from the host. This is one of the strategies of tissue engineering.

Lining the inner wall of the prosthesis with a continuous endothelial monolayer is a key condition for obtaining revascularisation with long-lasting permeability. Indeed, the vascular endothelium is a particularly active organ metabolically speaking and contributes significantly to maintaining the balance between thrombosis and hemostasis, as well as maintaining the homeostasis of smooth muscle cells and other cell types. The physiological endothelial luminal surface is perfectly hemocompatible and thromboresistant, and no biomaterial used for vascular purposes can equal the thromboresistance properties of the normal endothelium. In addition, while spontaneous endothelialisation of a vascular graft is usual for many animal models, this phenomenon is not observed in humans, although a few rare cases have been reported.

There have been two procedures for achieving endothelialisation, depending on the *modus operandi*: the single-step or preoperation protocol, now abandoned, and the two-step or deferred protocol [65]. The most recent results of clinical trials reported in the literature encourage optimism with regard to

the use of such cellularised structures, provided that the endothelial lining of the substitute is complete at the time of implantation [66].

To favour *in vitro* adhesion of endothelial cells to the particularly hydrophobic expanded polytetrafluoroethylene, to incite their proliferation, and to improve their resistance to being removed under the effects of shear strain, it is essential to treat the inner face of the graft [2, 67, 68].

22.4 In Vivo Assessment of Tissue Engineering Products

Tissue engineering products (TEP), complex by their very nature, require a great many preclinical tests before therapeutic trials can be carried out on humans. While a considerable number of tests can be carried out *in vitro*, e.g., physicochemical characterisation, cytotoxicity, *in vitro* hemocompatibility, animal models must be used to prove the biocompatibility of the product, whether it be a simple medical device or a tissue engineering product.

22.4.1 Animal Models

Why are animal models required to evaluate tissue engineering products? Quite apart from the demands of government regulations (FDA, EC standards) which are in a permanent state of evolution, animal models are the only ones able to mimic the clinical situation for which the TEP is intended. They thus allow a preclinical validation of the design, whatever the field of application: skin, cartilage, bone, vessels, pancreas, etc. This validation must provide the means for selecting the most effective device, demonstrating the clinical benefits of the product, and assessing the risks related to its use. Moreover, animal models are the only ones providing a way of testing and improving the surgical procedure for implanting bioactive substitutes. Finally, animal models are needed to understand and demonstrate certain properties of a TEP.

As an example, products resulting from bone tissue engineering are characterised by three main properties: resorption, osteoconductivity (the ability of the substitute to guide bone formation upon contact), and osteoinductivity (ability of the substitute to induce bone formation at ectopic sites). Resorption can be assessed *in vitro* in model solutions imitating interstitial fluids or the osteoclast resorption chamber. However, these media cannot provide a good prediction for the resorption of the material: taking the example of cellulose regenerated by oxidation of viscose-type cellulose by periodic acid, this is completely degraded in 48 h *in vitro*, whereas it is still present after one month of implantation in the rabbit femoral condyle [69]. If the material is colonised by cells, these resorption processes cannot be studied *in vitro*, insofar as the conditions of the culture medium cannot be correlated with the degradation media.

Concerning osteoconductivity, specific media able to induce the formation of calcium phosphates have been developed [70]. However, while these media are useful for calcium phosphate coating, they cannot be used to select materials in terms of their osteoconductive properties. Macroporous cellulose, which is soon coated with a layer of calcium phosphate in these solutions, does not lead to new bone formation within the macropores after one month of implantation in the rabbit femoral condyle [69].

Finally, by definition, osteoinductivity can only be explored using in vivo study models, usually in subcutaneous or intramuscular sites [71].

22.4.2 Which Animal Model for Which Application?

Various parameters determine the choice of animal model:

- The issue in question, e.g., biocompatibility, toxicological and pharmacological aspects, function and clinical performance of the device.
- The nature of the tissue engineering device.
- The implantation technique.
- Purchase, maintenance, and intervention costs.
- Lifespan of the animals.
- Ethical considerations.

The in vivo study of TEP occurs in several steps. In the first stages, small animals are used to discriminate the potential components of the device (matrix, growth factors, and cells) and to test the feasibility of the device. In the second stage, TEP are assessed on clinically relevant animal models [72].

Whatever the field of application, the animals most commonly used are mice, rats, rabbits, dogs, sheep, goats, pigs, and primates. Each has its own characteristics, determining its relevance or otherwise for testing TEP [73–75].

- *Mice*. The small size of these animals means that they do not provide a relevant clinical model. However, for ethical and financial reasons, mice are often used as a first step to test the feasibility of TEP and examine certain properties such as osteoinductivity for cellularised bone substitutes [71]. Moreover, the use of novel mutant mouse models represents a promising line of investigation in the future for understanding the pathology and developing relevant TEP.
- *Rats*. These have the same advantages as mice, while their slightly larger size facilitates manipulation. In the field of vascular research, they are used to study myocardial tissue engineering, and to a lesser extent, owing to their small size, vessel engineering. In the field of bone research, they are used to test the effect of the TEP on the cicatrization of newly formed defects, and to evaluate the concept of local genetic therapy, or the topical use of growth factors.
- *Rabbits*. For anatomical and financial reasons, the rabbit is a useful intermediate animal between small and large animals. It is used for early studies

of arterial [73] and osseous [76] TEP. However, the size of the vessels and the absence of osseous mechanical solicitation limit the clinical relevance of this model, justifying the recourse to larger animals in a second phase.

- *Dogs.* Dogs have been widely used for vascular and bone studies because they live long enough to carry out long-term studies, and they are large enough to facilitate surgical operations. However, they now tend to be used less often, for reasons of noise, the pressure of public opinion, and cost.
- *Sheep and Goats.* Their lifespan and weight in adulthood are compatible with long-term studies. Adult sheep weigh about 70 kg, implying mechanical solicitation on a level comparable to humans, making it a useful orthopedic model. The sheep is also an animal well suited for testing cardiac valves, insofar as this animal is similar to humans in terms of both the mechanical properties of the valves and hemodynamic flow parameters. The goat has a long neck, allowing easy access to the carotid artery, where vascular TEP are often implanted.
- *Pigs.* This animal is commonly used in animal experimentation, because it has a similar anatomy and physiology to humans. However, its rapid growth and its weight are limitations for long-term studies. Special breeds of miniature pigs can be used for experimental purposes, but their cost and rarity limit their use.
- *Primates.* These are the animals which most closely resemble humans from the anatomical and physiological point of view. Their long lifetime and appropriate weight allow long-term studies. However, for ethical and financial reasons, their use is prohibited when an animal lower in the phylogenetic order can be used instead. For these reasons, primates play only a marginal role in the field of tissue engineering.

22.4.3 Standard Methods for in Vivo Evaluation of Tissue Engineering Products

The procedures used for in vivo study of TEP can be divided into in vivo and ex vivo methods. The method used depends on the type of TEP.

In vivo methods usually stem from the field of medical research. Although they are undeniably useful, these approaches are rarely used, undoubtedly due to the dichotomous separation between the worlds of science and medicine, and also due to the difficulty in gaining access to the necessary equipment. However, over the past few years, imaging platforms dedicated to research have appeared on the scene in France, suggesting that these promising techniques may become more common in the future. This would allow a substantial saving in terms of animals, and the development of specific techniques for monitoring TEP, even in humans. In the field of vascular TEP, the techniques used are angiography and Doppler. For bone TEP, the techniques used are radiography, tomodensitometry, scintigraphy, and more recently MRI, positron emission tomography, and microscanning. One of the present limits of these investigative

methods is their resolution, which is not always sufficient compared with the dimensions of macropores in the matrices used in tissue engineering.

Ex vivo methods are the most widely used in vascular and osseous tissue engineering. The most common analytical method involves qualitative and quantitative histological examination combined with image analysis systems [76]. Observations can be made on sections that have simply been coloured to distinguish cell and matrix elements, or on sections treated with antibodies to reveal more specific structures. Bone tissue is rather particular in that it can be studied by calcified or uncalcified histology. In the absence of decalcification, a resin inclusion must be introduced and sections cut with a suitable microtome. Non-decalcified sections can be used for qualitative and quantitative study of bone frameworks, but do not allow a good analysis of cell structures owing to the fact that the sections are generally rather thick. Fluorochromes with bone tropism can be used in a complementary way to monitor new bone formation [77]. Likewise, radioactive tracers that can be incorporated in bone, e.g., calcium 45, can serve as functional labels and be measured quantitatively on autoradiographs or directly on sections using a beta imager [77]. Finally proton-induced X-ray emission (PIXE) analysis has been carried out on non-decalcified sections to reveal modifications on the atomic level. More classically, decalcified histology is used to characterise the cell response to the TEP (inflammation, degeneracy, necrosis, *ad integrum* cicatrisation, etc.), sometimes associated with histochemistry or immunocytochemistry techniques. Today, techniques of molecular biology such as Western blot, in situ hybridisation, quantitative PCR, etc., are also commonly used to analyse tissue repair mechanisms in the presence of the TEP.

Despite the development of ever more effective in vitro or simulation techniques, the recourse to animal models to analyse TEP is unfortunately still commonplace. These models must be used with due consideration for the ethics of animal experimentation, which concerns the kind and number of animals as well as the experimental conditions. In the future, it would be useful to define animal study models and standard methods in such a way as to be able to make comparisons between the different studies.

22.5 Investigative Methods Associated with Tissue Engineering

Tissue engineering provides a wide range of possibilities for combining cells, proteins, and materials within the same biomaterial dedicated to tissue repair, such as bone repair or vascular repair. Imaging techniques play a major role in the development of new tissue engineering technologies – by supplying tools for elaborating and optimising biomaterials – and also for their in situ validation in preclinical animal models and human clinical trials.

Different microscopic techniques (optical, acoustic, scanning electron, atomic force, etc.) can be used to evaluate biological tissues on the cellular and/or

microstructural level, the surface of materials and adsorbed molecules, and the tissue/material interface.

With regard to *in situ* experimentation on animal models, one is mainly concerned with the integration and efficiency of implanted biomaterials. High-resolution 3D imaging techniques such as synchrotron radiation microtomography [78], X-ray microtomography [79,80], or magnetic resonance microscopy [81] are used to reconstitute volumes of newly formed tissue and extract information regarding mineralisation, microarchitecture, and microvascularisation of these newly formed tissues.

One of the challenges in the years to come lies in the development and use of non-invasive imaging techniques for morphological and/or functional assessment during a longitudinal study of the same animal, leading to a significant reduction in intragroup biological variations, a reduction in the number of sacrifices, and a reduction in the cost of experimentation. High-resolution imaging techniques dedicated to experimental animal models, such as X-ray microtomography [79] or magnetic resonance microscopy [82,83], have made considerable progress both in terms of techniques and associated materials and in terms of exploitation and data analysis software. High-resolution nuclear imaging techniques, such as positron emission tomography (PET) or single photon emission computed tomography (SPECT), have also seen significant development, especially with the development of new tracers which allow many more applications to experimental animal models [84].

Imaging techniques used clinically on humans are still rather limited at the present time. The most widely used examinations for monitoring the integration of biomaterials are X-ray scanning and radiography [85]. These techniques have limited diagnostic and prognostic value. Indeed, their resolution and image quality do not allow an accurate assessment. However, it seems highly likely that current developments, oriented toward non-invasive high-resolution imaging techniques dedicated to experimental animal models, such as X-ray microtomography or magnetic resonance microscopy [86], will lead in the future to a considerable improvement in the performance of these techniques for applications to humans [87].

The optimal use of digital images resulting from the various imaging techniques lies mainly in the development of image analysis tools allowing quantitative evaluation [76]. Hence, any quantitative imaging technique integrates not only the imaging technique, but also specific techniques for quantification by image analysis, as well as specific techniques for interpreting with a view to diagnosis and/or monitoring. Digital modelling and/or simulation techniques are also exploited in association with imaging techniques. This association is particularly beneficial for optimising material design [88,89].

References

1. Baquey, C., Marmey, P., Guillemot, F., Porté-Durrieu, M.C.: Des matériaux aux biomatériaux: Une conversion qui passe par des modes d'élaboration et de traitement de surface appropriés, *Ann. Chim. Sci. Mat.* **28**, 109–121 (2003)
2. Shin, H., Jo, S., Mikos, A.G.: Biomimetic materials for tissue engineering, *Biomaterials* **24** (24), 4353–4364 (2003)
3. Josefsson, G., Lindberg, L., Wiklander, B.: Systemic antibiotics and gentamicin containing bone cement in the prophylaxis of postoperative infections in total hip arthroplasty, *Clin. Orthop.* **159**, 194–200 (1981)
4. Hanssen, A.D., Osmon, D.R., Nelson, C.L.: Prevention of deep periprosthetic joint infection, *J. Bone Joint Surg.* **78 A**, 458–471 (1996)
5. Voegel, J.C., Decher, G., Schaaf, P.: Multicouches de polyélectrolytes dans le domaine des biotechnologies, *Actualité chimique* November 2003, 30–34 (2003)
6. Kavanagh, C.A., Rochev, Y.A., Gallagher, W.M., Dawson, K.A., Keenan, A.K.: Local drug delivery in restenosis injury: Thermoresponsive co-polymers as potential drug delivery systems, *Pharmacology & Therapeutics* **102**, 1–15 (2004)
7. Durrieu, M.C., Héroguez, V., Quemener, D.: Synthèse de nanoparticules polymères rendues stimulables et réactives grâce à l'ancrage covalent à leur périphérie d'entités réactives et de principes actifs à libération contrôlée, France patent 04/06707, 2004 (21 June 2004)
8. Durrieu, M.C., Héroguez, V., Quemener, D., Baquey, C.: Biomatériaux bioactifs pour le relargage contrôlé de principes actifs, France patent 4/06708, 2004 (21 June 2004)
9. Chemtob, A., Heroguez, V., Gnanou, Y.: Dispersion ring-opening metathesis polymerization of latex particles using PEO based stabilizers, *Macromolecules* **35**, 9262 (2002)
10. Quemener, D., Chemtob, A., Heroguez, V., Gnanou, Y.: Synthesis of latex by ring-opening metathesis polymerization, *Polymer* **46**, 1067–1075 (2005)
11. Quemener, D., Heroguez, V., Gnanou, Y.: Synthesis of acid-sensitive lattices by ring-opening metathesis polymerization, *J. Polym. Sci. Part A: Polym. Chem.* **43**, 217–229 (2004)
12. Bartouilh de Taillac, L., Porté-Durrieu, M.C., Labrugère, C., Bareille, R., Amédée, J., Baquey, C.: Grafting of RGD peptides to cellulose to enhance human osteoprogenitor cell adhesion and proliferation, *Composites Science and Technology* **64**, 827–837 (2004)
13. Baquey, C., Palumbo, F., Porté-Durrieu, M.C., Legeay, G., Treassaud, A., d'Agostino, R.: Plasma treatment of expanded polyterafluoroethylene offers a way to biofunctionalization of its surface, *Nuclear Instruments and Methods in Physics Research B* **151**, 255–262 (1999)
14. Durrieu, M.C.: Conception, élaboration et caractérisation de matériaux bioactifs, *ITBM-RBM* **26**, 229–237 (2005)
15. Labrugère, C., Porté-Durrieu, M.C., Baquey, C.: Suivi du greffage peptidique de biomatériaux par spectrométrie XPS, *Le Vide: Science, Technique et Applications* **308**, 324–333 (2003)
16. Porté-Durrieu, M.C., Aymes-Chodur, C., Vergne, C., Betz, N., Baquey, C.: Surface treatment of biomaterials by gamma and swift heavy ion grafting, *Nuclear Instruments and Methods in Physics Research B* **151**, 404–415 (1999)

17. Porté-Durrieu, M.C., Aymes-Chodur, C., Betz, N., Brouillaud, B., Rouais, F., Le Moël, A., et al.: Synthesis of biomaterials by swift heavy ion grafting: Preliminary results of haemocompatibility, *Nuclear Instruments and Methods in Physics Research B* **131**, 364–375 (1997)
18. Porté-Durrieu, M.C., Pallu, S., Labrugère, C., Bareille, R., Amédée, J., Dard, M., et al.: Grafting RGD containing peptides onto hydroxyapatite to promote osteoblastic cell adhesion, *J. Mater. Sci.: Mater. in Med.* **15**, 779–786 (2003)
19. Porté-Durrieu, M.C., Guillemot, F., Pallu, S., Labrugère, C., Brouillaud, B., Amédée, J., et al.: Cyclo-DfKRG RGD peptides grafting onto Ti-6Al-4V: Physical characterization and interest towards human osteoprogenitor cells, *Biomaterials* **25**, 4837–4846 (2003)
20. Baquey, C., Palumbo, F., Porté-Durrieu, M.C., Legeay, G., Tressaud, A., D'Agostino, R.: Plasma treatment of expanded PTFE offers a way to a bio-functionalization of its surface, *Nuclear Instruments and Methods in Physics Research B* **151**, 255–262 (1999)
21. Porté-Durrieu, M.C., Aymes-Chodur, C., Bareille, R., Brouillaud, B., Betz, N., Baquey, Ch.: Development of heparin-like polymers using swift heavy ion and gamma radiation. Part I: Preparation and characterization of the material, *J. Biomedical Material Research* **52** (1), 119–227 (2000)
22. Hulak, I., Legeay, G.: An industrial three-dimensional surface treatment system and its applications, *Le Vide* **54** (284), 211–214 (1995)
23. Favia, P., D'Agostino, R., Palumbo, F.: Grafting of chemical groups onto polymers by means of RF plasma treatments: A technology for biomedical applications. In: Bordage, M.C., Gleizes, A. (Eds.), *J. Physique IV*, C4.199–C4.208 (1997)
24. Liu, L., Li, S., Garreau, H., Vert, M.: Selective enzymatic degradations of poly(L-lactide) and poly(ϵ -caprolactone) blend films, *Biomacromolecules* **1** (2), 350–359 (2000)
25. Cheo, S.H.Y., Wang, P., Tan, K.L., Ho, C.C., Kang, E.T.: Surface modification of natural rubber latex films via grafting of poly(ethylene glycol) for reduction in protein adsorption and platelet adhesion, *J. Mater. Sci.: Mater. Med.* **12** (5), 377–384 (2001)
26. Pierschbacher, M.D., Ruoslahti, E.: Cell attachment activity of fibronectin can be duplicated by small synthetic fragments of the molecule, *Nature* **309**, 30–33 (1984)
27. Ruoslahti, E., Pierschbacher, M.D.: New perspectives in cell adhesion: RGD and integrins, *Science* **238**, 491–497 (1987)
28. Xiao, S.J., Textor, M., Spencer, N.D., Sigrist, H.: Covalent attachment of cell-adhesive, (Arg-Gly-Asp)-containing peptides to titanium surfaces, *Langmuir* **14**, 5507–5516 (1998)
29. Rezania, A., Thomas, C.H., Branger, A.B., Waters, C.M., Healy, K.E.: The detachment strength and morphology of bone cells contacting materials modified with a peptide sequence found within bone sialoprotein, *J. Biomed. Mater. Res.* **37**, 9–19 (1997)
30. Rezania, A., Johnson, R., Lefkow, A.R., Healy, K.E.: Bioactivation of metal oxide surfaces. 1. Surface characterization and cell response, *Langmuir* **15**, 6931–6939 (1999)
31. Bartouilh de Taillac, L., Porté-Durrieu, M.C., Labrugère, C., Bareille, R., Amédée, J., Baquey, C.: Grafting of RGD peptides to cellulose to enhance

- human osteoprogenitor cell adhesion and proliferation, *Composites Science and Technology* **64**, 827–837 (2004)
32. Van Der Flier, A., Sonnenberg, A.: Function and interactions of integrins, *Cell Tissue Res.* **305**, 285–298 (2001)
 33. Lebaron, R.G., Athanasiou, K.A.: Extracellular matrix cell adhesion peptides: Functional applications in orthopedic materials, *Tissue Eng.* **6**, 85–103 (2000)
 34. Katz, B.Z., Zamir, E., Bershadsky, A., Kam, Z., Yamada, K.M., Geiger, B.: Physical state of the extracellular matrix regulates the structure and molecular composition of cell-matrix adhesions, *Mol. Biol. Cell* **11**, 1047–1060 (2000)
 35. Pelham, R.J. Jr., Wang, Y.L.: Cell locomotion and focal adhesions are regulated by the mechanical properties of the substrate, *Biol. Bull.* **194**, 348–350 (1998)
 36. Choquet, D., Felsenfeld, D.P., Sheetz, M.P.: Extracellular matrix rigidity causes strengthening of integrin–cytoskeleton linkages, *Cell* **88**, 39–48 (1997)
 37. Harrison, D., Johnson, R., Tucci, M., Puckett, A., Tsao, A., Hughes, J., et al.: Interaction of cells with UHMWPE impregnated with the bioactive peptides RGD, RGE or poly-L-lysine, *Biomed. Sci. Instrum.* **34**, 41–46 (1997)
 38. McConachie, A., Newman, D., Tucci, M., Puckett, A., Tsao, A., Hughes, J., et al.: The effect on bioadhesive polymers either freely in solution or covalently attached to a support on human macrophages, *Biomed. Sci. Instrum.* **35**, 45–50 (1999)
 39. Healy, K.E., Tsai, D., Kim, J.E.: Osteogenic cell attachment to degradable polymers, *Mater. Res. Soc. Symp. Proc.* **252**, 109–114 (1992)
 40. Dettin, M., Conconi, M.T., Gambaretto, R., Pasquato, A., Folini, M., Di Bello, C., et al.: Novel osteoblast-adhesive peptides for dental/orthopedic biomaterials, *J. Biomed. Mater. Res.* **60** (3), 466–471 (2002)
 41. Walluscheck, K.P., Steinhoff, G., Kelm, S., Haverich, A.: Improved endothelial cell attachment on ePTFE vascular grafts pretreated with synthetic RGD-containing peptides, *Eur. J. Vasc. Endovasc.* **12**, 321–330 (1996)
 42. Varani, J., Inman, D.R., Fligiel, S.E., Hillegas, W.J.: Use of recombinant and synthetic peptides as attachment factors for cells on microcarriers, *Cytotechnology* **13**, 89–98 (1993)
 43. Yang, X.B., Roach, H.I., Clarke, N.M., Howdle, S.M., Quirk, R., Shakesheff, K.M., et al.: Human osteoprogenitor growth and differentiation on synthetic biodegradable structures after surface modification, *Bone* **29**, 523–531 (2001)
 44. Breuers, W., Klee, D., Hocker, H., Mittermayer, C.: Immobilization of a fibronectin fragment at the surface of a polyetherurethane film, *J. Mater. Sci.: Mater. Med.* **2**, 106–109 (1991)
 45. Kugo, K., Okuno, M., Masuda, K., Nishino, J., Masuda, H., Iwatsuki, M.: Fibroblast attachment to Arg-Gly-Asp peptide-immobilized poly(g-methyl L-glutamate), *J. Biomater. Sci. Polym. Ed.* **5**, 325–337 (1994)
 46. Porté-Durrieu, M.C., Labrugère, C., Villars, F., Lefebvre, F., Dutoya, S., Guette, A., et al.: Development of RGD peptides grafted onto silica surfaces: XPS characterization and human endothelial cell interactions, *J. Biomed. Mater. Res.* **46**, 368–375 (1999)
 47. Marchand-Brynaert, J.: Surface modifications and reactivity assays of polymer films and membranes by selective wet chemistry, *Recent Res. Polym. Sci.* **2**, 335–361 (1998)
 48. Guillemot, F., Porté-Durrieu, M.C., Labrugère, C., Baquey, C.: Ti^{4+} to Ti^{3+} conversion of TiO_2 uppermost layer by low-temperature vacuum annealing:

- Interest for titanium biomedical applications, *J. Colloid and Interface Science*, **255**, 75–78 (2002)
49. Adamo, C.T., Mailhot, J.M., Smith, A.K., Borke, J.L.: Connexin-43 expression in oral-derived human osteoblasts after transforming growth factor-beta and prostaglandin E2 exposure, *J. Oral Implantol.* **27** (1), 25–31 (2001)
 50. Narmoneva, D.A., Oni, O., Sieminski, A.L., Zhang, S., Gertler, J.P., Kamm, R.D., et al.: Self-assembling short oligopeptides and the promotion of angiogenesis, *Biomaterials* **26** (23), 4837–4846 (2005)
 51. Matsumoto, T., Okazaki, M., Inoue, M., Yamaguchi, S., Kusunose, T., Toyonaga, T., et al.: Hydroxyapatite particles as a controlled release carrier of protein, *Biomaterials* **25** (17), 3807–3812 (2004)
 52. Jiang, B.B., Gao, C.Y., Hu, L., Shen, J.C.: Water-dispersed bone morphogenetic protein nanospheres prepared by co-precipitation method, *J. Zhejiang Univ. Sci.* **5** (8), 936–940 (2004)
 53. Patka, P., Haarman, H.J., Bakker, F.C.: Bone transplantation and bone replacement materials, *Ned. Tijdschr. Geneesk.* **142** (16), 893–896 (1998)
 54. Rauschmann, M.A., Wichelhaus, T.A., Stirnal, V., Dingeldein, E., Zichner, L., Schnetzler, R., et al.: Nanocrystalline hydroxyapatite and calcium sulphate as biodegradable composite carrier material for local delivery of antibiotics in bone infections, *Biomaterials* **26** (15), 2677–2684 (2005)
 55. Geissler, U., Hempel, U., Wolf, C., Scharnweber, D., Worch, H., Wenzel, K.: Collagen type I-coating of Ti6Al4V promotes adhesion of osteoblasts, *J. Biomed. Mater. Res.* **51** (4), 752–760 (2000)
 56. Gronowicz, G., McCarthy, M.B.: Response of human osteoblasts to implant materials: Integrin-mediated adhesion, *J. Orthop. Res.* **14** (6), 878–887 (1996)
 57. Hasenbein, M.E., Andersen, T.T., Bizios, R.: Micropatterned surfaces modified with select peptides promote exclusive interactions with osteoblasts, *Biomaterials* **23** (19), 3937–3942 (2002)
 58. El-Ghannam, A.R., Ducheyne, P., Risbud, M., Adams, C.S., Shapiro, I.M., Castner, D., et al.: Model surfaces engineered with nanoscale roughness and RGD tripeptides promote osteoblast activity, *J. Biomed. Mater. Res. A* **68** (4), 615–627 (2004)
 59. Tosatti, S., Schwartz, Z., Campbell, C., Cochran, D.L., Van de Vondede, S., Hubbell, J.A. et al.: RGD-containing peptide GCRGYGRGDSPG reduces enhancement of osteoblast differentiation by poly(L-lysine)-graft-poly(ethylene glycol)-coated titanium surfaces, *J. Biomed. Mater. Res. A* **68** (3), 458–472 (2004)
 60. Catledge, S.A., Vohra, Y.K., Bellis, S.L., Sawyer, A.A.: Mesenchymal stem cell adhesion and spreading on nanostructured biomaterials, *J. Nanosci. Nanotechnol.* **4** (8), 986–989 (2004)
 61. Webster, T.J., Ejiofor, J.U.: Increased osteoblast adhesion on nanophase metals: Ti, Ti6Al4V, and CoCrMo, *Biomaterials* **25** (19), 4731–4739 (2004)
 62. Vara, D.E.S., Salacinski, H.J., Kannan, R.Y., Bordenave, L., Hamilton, G., Seifalian, A.M.: Cardiovascular tissue engineering: State of the art, *Pathologie Biologie* **53** (10), 599–612 (2005)
 63. Kannan, R.Y., Salacinski, H.J., Butler, P.E., Hamilton, G., Seifalian, A.M.: Current status of prosthetic bypass grafts: A review, *J. Biomed. Mater. Res. B: Appl. Biomater.* **74** (1), 570–581 (2005)
 64. Thomas, A.C., Campbell, G.R., Campbell, J.H.: Advances in vascular tissue engineering, *Cardiovasc. Pathol.* **12** (5), 271–276 (2003)

65. Bordenave, L., Remy-Zolghadri, M., Fernandez, P., Bareille, R., Midy, D.: Clinical performance of vascular grafts lined with endothelial cells, *Endothelium* **6** (4), 267–275 (1999)
66. Meinhart, J.G., Deutsch, M., Fischlein, T., Howanietz, N., Froschl, A., Zilla, P.: Clinical autologous in vitro endothelialization of 153 infrainguinal ePTFE grafts, *Ann. Thorac. Surg.* **71** (5 Suppl.), S327–S3231 (2001)
67. Li, C., Zheng, Y., Imran, M.: In vitro study of cell-promoting multiple-armed peptides, *J. Biomed. Mater. Res. A* **71** (1), 134–142 (2004)
68. Hersel, U., Dahmen, C., Kessler, H.: RGD modified polymers: Biomaterials for stimulated cell adhesion and beyond, *Biomaterials* **24** (24), 4385–4415 (2003)
69. Laurence, S., Bareille, R., Baquey, C., Fricain, J.C.: Development of a resorbable macroporous cellulosic material used as hemostatic in an osseous environment, *J. Biomed. Mater. Res. A* **73** (4), 422–429 (2005)
70. Kokubo, T., Kushitani, H., Sakka, S., Kitsugi, T., Yamamuro, T.: Solutions able to reproduce in vivo surface-structure changes in bioactive glass-ceramic A-W, *J. Biomed. Mater. Res.* **24** (6), 721–734 (1990)
71. Bareille, R., Lafage-Proust, M.H., Faucheux, C., Laroche, N., Wenz, R., Dard, M., et al.: Various evaluation techniques of newly formed bone in porous hydroxyapatite loaded with human bone marrow cells implanted in an extra-osseous site, *Biomaterials* **21** (13), 1345–1352 (2000)
72. Hunziker, E.B.: Tissue engineering of bone and cartilage. From the preclinical model to the patient, *Novartis Found Symp.* 2003 **249**, 70–78; discussion 78–85, 170–174, 239–241 (2003)
73. Rashid, S.T., Salacinski, H.J., Hamilton, G., Seifalian, A.M.: The use of animal models in developing the discipline of cardiovascular tissue engineering: A review, *Biomaterials* **25** (9), 1627–1637 (2004)
74. Buma, P., Schreurs, W., Verdonchot, N.: Skeletal tissue engineering: From in vitro studies to large animal models, *Biomaterials* **25** (9), 1487–1495 (2004)
75. Khan, S.N., Lane, J.M.: Spinal fusion surgery: Animal models for tissue-engineered bone constructs, *Biomaterials* **25** (9), 1475–1485 (2004)
76. Pothuau, L., Fricain, J.C., Pallu, S., Bareille, R., Renard, M., Durrieu, M.C., et al.: Mathematical modelling of the distribution of newly formed bone in bone tissue engineering, *Biomaterials* **26** (33), 6788–6797 (2005)
77. Fricain, J.C., Granja, P.L., Barbosa, M.A., de Jeso, B., Barthe, N., Baquey, C.: Cellulose phosphates as biomaterials. In vivo biocompatibility studies, *Biomaterials* **23** (4), 971–980 (2002)
78. Weiss, P., Obadia, L., Magne, D., Bourges, X., Rau, C., Weitkamp, T. et al.: Synchrotron X-ray microtomography (on a micron scale) provides three-dimensional imaging representation of bone ingrowth in calcium phosphate biomaterials, *Biomaterials* **24** (25), 4591–4601 (2003)
79. Gasser, J.A., Ingold, P., Grosios, K., Laib, A., Hammerle, S., Koller, B.: Non-invasive monitoring of changes in structural cancellous bone parameters with a novel prototype micro-CT, *J. Bone Miner. Metab.* **23** Suppl., 90–96 (2005)
80. Jiang, Y., Zhao, J., Liao, E.Y., Dai, R.C., Wu, X.P., Genant, H.K.: Application of micro-CT assessment of 3-D bone microstructure in preclinical and clinical studies, *J. Bone Miner. Metab.* **23** Suppl., 122–131 (2005)
81. Traore, A.S., Woerly, S., Doan, V.D., Marois, Y., Guidoin, R.: In vivo magnetic resonance imaging and relaxometry study of a porous hydrogel implanted in the trapezius muscle of rabbits, *Tissue Eng.* **6** (3), 265–278 (2000)

82. Pothuaud, L., Majumdar, S.: Magnetic resonance imaging. In: *The Physical Measurement of Bone*, ed. by C.M. Langton and C.F. Njeh, Institute of Physics Publishing (IoP) (2004) pp.475–510
83. Miraux, S., Serres, S., Thiaudiere, E., Canioni, P., Merle, M., Franconi, J.M.: Gadolinium-enhanced small-animal TOF magnetic resonance angiography, *Magma* **17** (3–6), 348–352 (2004)
84. Green, M.V., Seidel, J., Vaquero, J.J., Jagoda, E., Lee, I., Eckelman, W.C.: High resolution PET, SPECT and projection imaging in small animals, *Comput. Med. Imaging Graph.* **25** (2), 79–86 (2001)
85. Beltrame, F., Cancedda, R., Canesi, B., Crovace, A., Mastrogiacomo, M., Quarto, R., et al.: A simple non invasive computerized method for the assessment of bone repair within osteoconductive porous bioceramic grafts, *Biotechnol. Bioeng.* **92** (2), 189–198 (2005)
86. Jiang, Y., Zhao, J., White, D.L., Genant, H.K.: Micro CT and micro MR imaging of 3D architecture of animal skeleton, *J. Musculoskelet. Neuronal Interact.* **1** (1), 45–51 (2000)
87. Wang, G., Zhao, S., Yu, H., Miller, C.A., Abbas, P.J., Gantz, B.J., et al.: Design, analysis and simulation for development of the first clinical micro-CT scanner, *Acad. Radiol.* **12** (4), 511–525 (2005)
88. Sun, W., Darling, A., Starly, B., Nam, J.: Computer-aided tissue engineering: Overview, scope and challenges, *Biotechnol. Appl. Biochem.* **39** (Pt. 1), 29–47 (2004)
89. Lin, C.Y., Kikuchi, N., Hollister, S.J.: A novel method for biomaterial scaffold internal architecture design to match bone elastic properties with desired porosity, *J. Biomech.* **37** (5), 623–636 (2004)

Viral Vectors for in Vivo Gene Transfer

E. Thévenot, N. Dufour, and N. Déglon

23.1 Introduction

23.1.1 In Vivo Gene Transfer

The transfer of DNA into the nucleus of a eukaryotic cell (gene transfer) is a central theme of modern biology. The transfer is said to be somatic when it refers to non-germline organs of a developed individual, and germline when it concerns gametes or the fertilised egg of an animal, with the aim of transmitting the relevant genetic modification to its descendants [1]. The efficient introduction of genetic material into a somatic or germline cell and the control of its expression over time have led to major advances in understanding how genes work in vivo, i.e., in living organisms (functional genomics), but also to the development of innovative therapeutic methods (gene therapy). The efficiency of gene transfer is conditioned by the vehicle used, called the vector. Desirable features for a vector are as follows:

- Easy to produce high titer stocks of the vector in a reproducible way.
- Absence of toxicity related to transduction (transfer of genetic material into the target cell, and its expression there) and no immune reaction of the organism against the vector and/or therapeutic protein.
- Stability in the expression of the relevant gene over time, and the possibility of regulation, e.g., to control expression of the therapeutic protein on the physiological level, or to end expression at the end of treatment.
- Transduction of quiescent cells should be as efficient as transduction of dividing cells.

Vectors currently used fall into two categories: non-viral and viral vectors. In non-viral vectors, the DNA is complexed with polymers, lipids, or cationic detergents (described in Chap. 3). These vectors have a low risk of toxicity and immune reaction. However, they are less efficient in vivo than viral vectors when it comes to the number of cells transduced and long-term transgene expression. (Naked DNA transfer or electroporation is rather inefficient in

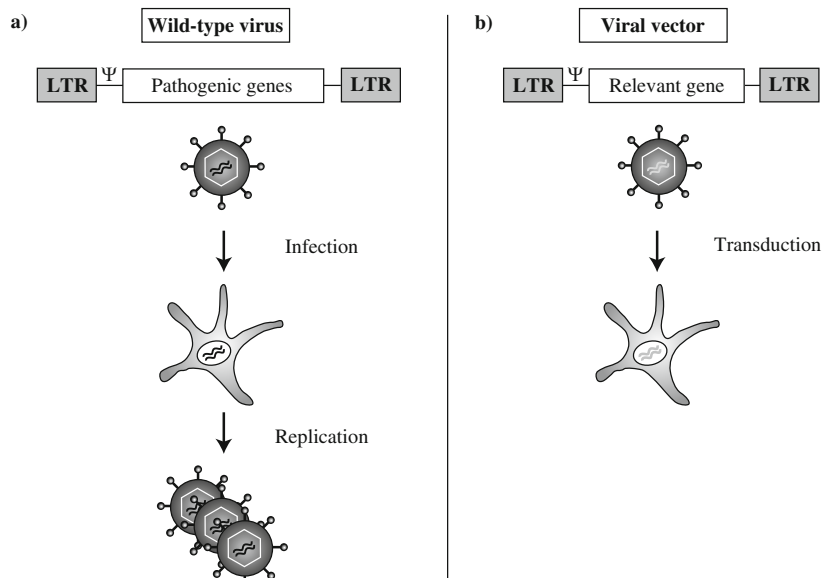


Fig. 23.1. Constructing a viral vector from a wild-type virus by modifying its genome. A viral vector (**b**) is derived from a wild-type virus (**a**) by modifying its genome. The genes required for replication and those responsible for its pathogenicity are excised and replaced by the gene of interest (transgene). However, some non-coding sequences are kept, e.g., in the case of retroviruses, the Ψ sequence which is essential during production to ensure that the genome is encapsidated within a virus particle, and the LTR sequences which, during transduction, are necessary for reverse transcription of the genome, and hence the expression of the relevant gene

Table 23.1. Four main types of viral vector used for in vivo gene transfer. Abbreviations: ss single strand, ds double strand, TU transducing unit (number of cells transduced, defined on p.1076). Among the reviews discussing these four viral vectors, a good selection would be [6] for retroviral vectors, [7] for lentiviral vectors, [8] for AAV, and [17] for adenoviral vectors

Vector	Virus family	Size [nm]	Titer [TU/ml]	Maximal capacity [kb]
MLV	Retrovirus (RNA)	100	10^6 – 10^9	7–8
Lentiviral	Retrovirus (RNA)	100	10^6 – 10^9	7–8
Adeno-associated	Parvovirus (ss DNA)	20–30	10^9 – 10^{13}	4.5
Adenoviral	Adenovirus (ds DNA)	80–120	10^{10} – 10^{12}	20–35

the organism. This type of gene transfer will not be discussed here, and the interested reader is referred to the review [2].) For this reason, it is mainly viral vectors that are used for gene transfer in animals and humans.

23.1.2 Viral Vectors

Recombinant viruses, which have dimensions in the range 20–100 nm, play an important role in health technologies, both in understanding what genes do in situ in the organism and in developing novel therapeutic strategies.

In this chapter, we discuss the main types of viral vectors currently used for gene transfer and the common principles ensuring the production of efficient and non-pathogenic recombinant viruses. We shall also describe the specific features of these different vectors, which help the experimenter to make a choice suited to the targeted cell type. We shall then illustrate the two main applications of viral vectors, viz., therapeutic gene transfer and the generation of animal models. Finally, we shall consider recent improvements regarding these vectors, whereby it is now possible to monitor by imaging and regulate transgene expression once it has been transferred within the organism.

23.2 Main Types of Viral Vector

Viruses possess genetic material that is essential for their replication. Viral vectors are derived from viruses by genetic engineering, replacing this genetic material by the relevant cassette, which typically contains a promoter and the relevant transgene (see Fig. 23.1). As a consequence, the viral vector will still be able to enter the target cell and get the transgene expressed, but it will not be able to replicate. The common principle for building a viral vector from a wild-type virus thus consists in excising from its genome the sequences responsible for its pathogenicity, as well as those involved in its replication. Viral genes will be expressed solely during the transient transfection production phase to generate the viral particles by complementation.

All virus families have been used to generate viral vectors. Indeed, each one has its own characteristic features, and the wealth of vectors available is a key factor for responding to the specificities of each pathology (size of the transgene to be transferred, target cell type, stability of expression over time, chromosomal integration of the transgene). To begin with, we shall discuss the most commonly used viral vectors, which derive from retroviruses and lentiviruses, adenoviruses and adeno-associated viruses¹ (see Table 23.1). We shall then be able to detail their biomedical applications to humans and animals for treating and modelling various pathologies.

¹ A fifth type of viral vector, derived from the herpes simplex virus (HSV), has the advantage of a high theoretical capacity for cloning (several tens of kb). However, the use of these vectors is limited by their toxicity and immunogenicity, and by the drop in transgene expression observed in certain tissues, such as the brain. These vectors will not be discussed in this chapter, in which we have chosen to describe only those vectors that are currently the most effective for gene transfer. However, the interested reader may refer to the review [3].

23.2.1 Retroviral and Lentiviral Vectors

Retroviral Vectors

Retroviruses differ from other viruses by the reverse transcription and chromosomal integration mechanisms of their genome in the target cell. During the infection cycle of a wild-type retrovirus, its RNA viral genome is reverse transcribed into DNA, thereby forming the proviral genome. The latter then integrates into the genome of the target cell and uses the transcription and translation machinery of the host. Morphologically, virions are enveloped particles of diameter about 100 nm, with an inner capsid protecting a genome comprising two identical RNA molecules of around 10 kb. Retroviruses have three essential gene classes (see Fig. 23.2):

1. *gag* codes structure proteins, especially the capsid protein,
2. *pol* codes the reverse transcriptase enzyme, used to convert the RNA genome into DNA in the reverse transcription cycle, and integrase enzyme, required to integrate the proviral genome into the host cell,
3. *env* codes the glycoprotein envelope.

Their genome also contains *cis*-regulatory sequences² (see Fig. 23.2) required for its encapsidation within the virus particles (Ψ sequence) and, after infection of the target cell, its reverse transcription, its chromosomal integration, and the expression of the three viral gene classes (sequences located in the long terminal repeat or LTR).

The tropism of a retrovirus, i.e., its target cell type, is determined by the envelope glycoprotein, which binds to the receptor of the target cell. For example, the Moloney virus of murine leukemia (MLV) only infects mice cells. The target cell population of the viral vector can be diversified by replacing *env* by the envelope gene of another virus, i.e., by pseudotyping the vector. For example, by using the G protein of the vesicular stomatitis virus (VSV-G), one can generate a viral vector capable of transducing a wide variety of mammalian cell types (amphotropic). A second advantage of VSV pseudotyped viral particles is their improved mechanical strength, which means that they can be concentrated by more than a factor of 1,000 by ultracentrifuging.

Vectors derived from the MLV have been obtained by excising the *gag*, *pol*, and *env* genes from the viral genome. Only those viral sequences required for encapsidation (Ψ) of the viral genome during production, and then for reverse transcription and integration (LTR) during transduction, have been conserved. The relevant cassette is inserted in the place of the excised viral sequences (see Figs. 23.1 and 23.2). Consequently, the retroviral vectors cannot

² In molecular biology, *cis*-regulatory sequences refer to non-coding short genome sequences involved, for instance, in the transcription initiation of a nearby gene. The regulatory effect (either gene induction or repression) is triggered by the binding of a specific *trans*-regulatory factor (protein) onto the *cis*-regulatory sequence.

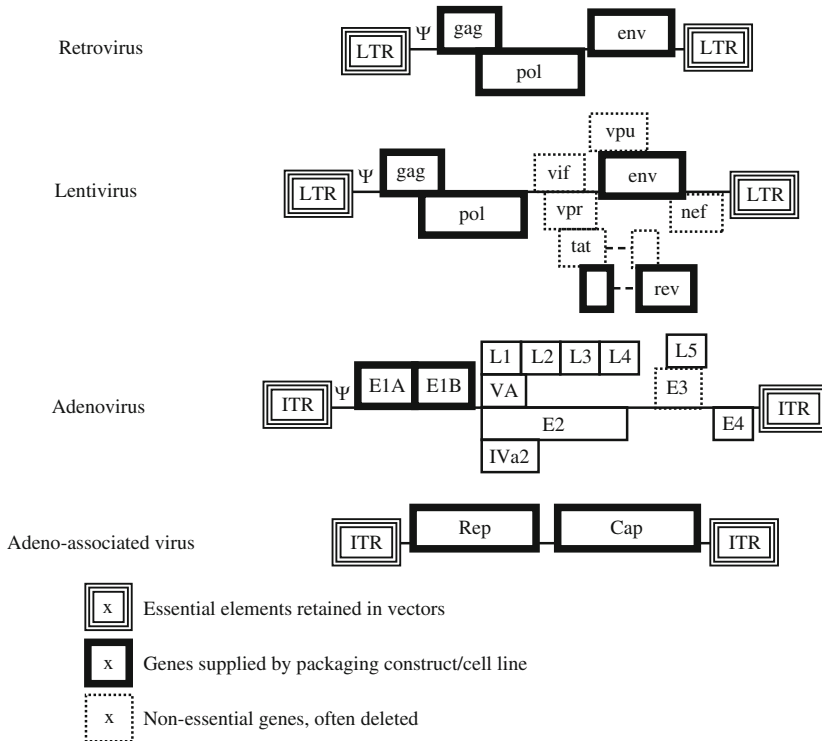


Fig. 23.2. Viruses commonly used to derive recombinant viral vectors. Essential *cis*-regulatory sequences and genes are conserved in the vectors (represented by a *triple boundary* in the diagram) or supplied transiently by complementation during production (*black boundaries*). Non-essential sequences are often excised (*dotted boundaries*). Only the main genetic elements are represented, and genomes are not drawn to scale. In second and third generation adenoviral vectors, the number of deleted viral genes is increased. Ψ encapsidation sequence, LTR (respectively, ITR) long (respectively, inverted) terminal repeat [9]

replicate. They only use the early stages of the viral cycle needed to get the transgene integrated into the cell genome.

The retroviral vector derived from the MLV is currently the most widely used in gene therapy protocols. However, one limitation of simple retroviruses is their inability to infect post-mitotic cells, i.e., cells that do not divide, as is the case for most neurons. Moreover, attempts to generate transgenic animals using these vectors³ have proved fruitless (repression of transgene expression

³ The idea of viral transgenesis is to transduce a fertilised oocyte (zygote, the first cell of the animal) by a retroviral vector, in order to get the transgene integrated in one of the chromosomes of the zygote. From this moment, the transgene is an integral part of the genome of the animal, and as such it is replicated at each cell division and ends up in all the cells of the animal.

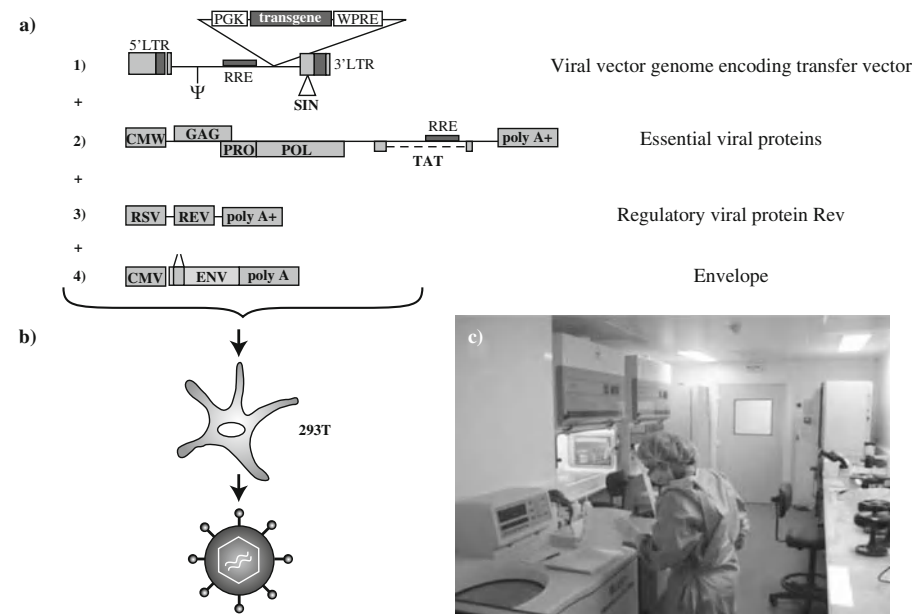


Fig. 23.3. Producing lentiviral stocks by transient transfection. **(a)** The four transfected plasmids contain on the one hand the modified genome of the virus including the relevant transgene (transfer vector), and on the other hand code for the viral proteins required for *trans*-complementation during production. **(b)** After cotransfection in the cell line 293T, the cells produce viral particles in which the genome comprises only sequences coded by the transfer vector. These particles thus maintain their ability to penetrate a host cell and express the relevant transgene within it, but they are unable to replicate. **(c)** The lentiviral particles are gathered from the culture supernatant by ultracentrifuging in an L3 confined environment

has been observed, possibly due to methylation of promoter sequences during development of the transgenic animal). These limitations led to the development since 1996 of vectors from a specific subfamily of retroviruses, namely the lentiviruses.

Lentiviral Vectors

Introduction

Lentiviruses are retroviruses with a more complex genome than the MLV. Their name comes from the long incubation period of the illnesses they cause. The best known and best studied of these is the human immunodeficiency virus (HIV). As well as *gag*, *pol*, and *env*, their genome includes two genes which directly regulate expression of the viral genes (*tat* and *rev*) and four accessory genes (*vpr*, *vif*, *vpu*, and *nef*), which have important functions

during the viral propagation cycle and in the pathogenicity of the virus (see Fig. 23.2). Lentiviruses have the particular feature that they can infect both mitotic and post-mitotic cells, thanks to an active mechanism for nuclear import of the proviral genome.

Recombinant Vector

The development of non-replicating vectors derived from the HIV goes back to the mid-1990s. Several improvements have been made with later generations of lentiviral vectors in the areas of viral titer, biosafety, and transgene expression level [4, 5]. One point is that the most recent vectors, called self-inactivating vectors (SIN), include a deletion of the viral promoter located in the U3 region of the LTR 3' (see Fig. 23.3). This excision has two advantages: first, it eliminates viral transcription elements that could interfere either with the transgene promoter or with the the cell genes in the vicinity of the integration site; second, it limits the risks of recombination with a wild-type virus. Another point is that, adding the woodchuck hepatitis virus posttranscriptional regulatory element (WPRE), transgene expression can be increased by a factor of 2 to 5 in different cell types [5].

Lentiviral vectors transduce both undifferentiated cells (pluripotent stem cells) and differentiated cells, and transgene expression is stable for months and even years. Their efficiency has already been proven in the nervous system, the hematopoietic system, and to a lesser extent in the liver. Vectors have also been derived from lentiviruses of other species (feline, equine, simian, and bovine lentiviruses).

Producing Lentiviral Vectors

Stocks of lentiviral particles are routinely produced within a week. The first step is to transfect a cell line with the genetic material needed to produce viral particles. The supernatant is then collected and the particles concentrated by ultracentrifugation [4, 7]. The transferred genetic material is in most cases distributed over four plasmids (for biosafety reasons, the genome must be physically split between different plasmids to prevent recombination and creation of wild-type viruses):⁴

1. A plasmid (called the transfer vector) coding the viral RNA to be encapsidated.
2. Two *trans*-complementation plasmids coding the viral proteins needed for the production of the viral particles (especially the structure proteins *gag* and the enzymes *pol*).

⁴ The term 'plasmid' refers to circular DNA molecules able to replicate autonomously once transferred inside a bacterium. In molecular biology, plasmids are routinely used, especially for stocking some particular gene. The integrity of the sequence of this gene is then preserved and it is easily amplified to produce a very large number of identical copies, by transfecting bacteria with the plasmid and growing the bacterial culture.

3. A plasmid coding the envelope glycoprotein (see Fig. 23.3).

Only the transfer vector codes the viral RNA genome to be encapsidated in the lentiviral particles. The three other complementation plasmids serve merely to fabricate viral particles. The absence of the encapsidation sequence Ψ in these plasmids ensures that the viral sequences they contain, in particular the genes *gag*, *pol*, and *env*, are not encapsidated in the particles. In addition, the physical separation of these sequences in four different plasmids reduces the risk of reconstituting the full-length sequence coding the wild-type viral genome by recombination. Indeed, such an event would require three successive recombinations between the different complementation plasmids and the transfer vector. The calcium phosphate coprecipitation protocol in the cell line 293T provides a simple and highly efficient transfection of the four plasmids.⁵

Viral production is carried out in level L3 confinement conditions according to French legislation.⁶ Supernatants are collected 48 h after transfection and viral particles are sedimented out by ultracentrifuge. The viral stock, aliquoted and frozen at -80°C , can be kept for several years. At this stage, the viral vector can be downgraded to level L2 for in vitro experiments, i.e., with culture cells, and in vivo experiments, provided that the transgene is not itself classified as L3 and that the total amount of vector manipulated corresponds to less than $2\text{ }\mu\text{g}$ of p24 (see below for the definition of this unit).

Titer

The most commonly used titration method for lentiviral stock is to quantify the p24 capsid protein using an ELISA assay. After ultracentrifuging, which can concentrate by a factor of 1,000, the resulting titer is generally of the order of 10^5 ng of p24 per milliliter. It is also possible to assay the viral genome RNA by quantitative RT-PCR [10] (see Chap. 15).

The functionality of the vector can be determined in vitro, provided that one has a method for visualising the transgene expression (antibody or enzyme activity). The idea is to infect a reference cell line (e.g., 293T cells) with a serialised dilution of the viral stock, then count the number of cells expressing the transgene. The titer obtained (the number of transduced cells, called the transducing unit or TU), which can vary depending on the type and size of the transgene, is generally of the order of 10^8 TU/mL.

Checking for Replication-Competent Viruses

To ensure the biosafety of the batch, the absence of contamination by replication-competent lentiviruses (RCL) is systematically checked after production [11]. One approach is to infect the cell line with the viral vector and

⁵ Attempts to avoid the transfection step by generating stable cell lines expressing the complementation proteins have come up against problems of toxicity with VSV-G and certain lentiviral proteins (integrase).

⁶ Principles of the *Commission de Génie Génétique* for classifying operations involving vectors derived from lentiviruses.

then analyse, after a change of medium, the absence of the p24 protein (or the viral RNA genome) in the supernatant [12].

23.2.2 Adenoviral Vectors

Wild-Type Virus

Adenoviruses are medium-sized viruses isolated in various species of birds and mammals. In humans, they are responsible for benign respiratory infections. Adenoviral particles have diameters in the range 80–120 nm without envelope. The adenovirus can infect both mitotic and post-mitotic cells. They remain in an episomal form, i.e., not integrated into a chromosome, in the nucleus. The adenoviral genome comprises a 36-kb double-stranded linear DNA molecule coding for more than a dozen genes.

The adenovirus enters into the host cell by binding to the cell surface via its interaction with the coxsackievirus and adenovirus receptor (CAR). Infection is a very fast and efficient process, beginning when the adenovirus enters the cell by endocytosis: about 40% of internalised particles form complexes at the nuclear pores and release their DNA into the nucleus just 2 h after infection. The first phase of protein expression, called the early phase, then begins, followed by a second phase, the late gene expression phase, beginning some 10 h later. Replication of the viral genome occurs between the two phases. The ends of the genome contain inverse terminal repeats (ITR) of about 100 bp, serving as the basis for replication (see Fig. 23.2). At the end of the late phase, 10^3 – 10^4 virions are produced within the infected cell nucleus. Finally, 30–40 h after the beginning of the infection, the cell dies and the virions are released, marking the end of the reproductive cycle.

Among the genes with early expression (indicated by the letter E for ‘early’), E1A (a gene that is said to be immediate early) controls the expression of all other early genes. E1B, in association with E1A, serves to transform the host cell.⁷ E3 codes for proteins that block the inflammatory response and which allow infected cells to escape the immune system.

Recombinant Vector

At the beginning of the 1970s, the adenovirus was first used as an orally administered vaccine. The absence of major side effects and the innocuousness of adenoviral infections in immunocompetent individuals were key arguments to justify the use of this virus as a vector for gene therapy in humans. The first in vivo gene transfer using an adenoviral vector was carried out in 1990 [13]. Up to the present time, adenoviral vectors have been used in 25% of the 1,347

⁷ Cell transformation is the acquisition by a cell of one or more characters proper to the malignant cell, e.g., loss of contact inhibition and/or unlimited growth (immortality).

clinical gene therapy trials recorded in the *Journal of Gene Medicine* [28]. They are deleted from the E1 (E1A and B) region of the viral genome, responsible for activating viral transcription and replication (see Fig. 23.2). The E3 region can also be deleted, as its presence is not essential for viral production nor for cell transduction. The so-called first generation adenoviral vectors Ad Δ E1 Δ E3 have a cloning capacity of 7.5 kb. One of the main advantages of these vectors is the possibility of producing high titer stocks (typically 10^{11} TU/mL) relatively easily.

The principles of construction and production are essentially the same as for other viral vectors. Recombinant adenoviruses are produced in cell line 293, which expresses proteins E1A and E1B constitutively and hence allows *trans*-complementation of the activity of the E1 region. Classically, adenoviral vectors are generated by in situ homologous recombination in this cell line. The vector is then purified by phage lysis, in order to select a cloned population, then amplified to produce a viral stock. However, production of DNA viruses (adenoviruses and adeno-associated viruses) differs from production of retroviruses by the fact that the viral particles accumulate, not in the culture medium, but in the cells, which must then be lysed to recover the vector.

Adenoviral vectors have broad tropism, including both proliferating and quiescent cells. In 1993, the remarkable efficiency of these vectors for gene transfer in cells of mammalian central nervous systems (CNS) was demonstrated [14]. Since then, many studies have confirmed that the recombinant adenovirus is a choice vector for gene transfer in the CNS [15,16].

Despite the deletion of the E1 region, the viral genes are expressed at a low level in cells transduced by first generation vectors. This results in a certain direct toxicity of the vector and immunogenicity of the products of the viral genes. The immune response of cytotoxic T lymphocytes can thus lead to elimination of the transduced cells. This is why new high-capacity (gutless) adenoviral vectors have been developed. All viral sequences are eliminated from these vectors, apart from the ITR and the encapsidation signal Ψ [17].

23.2.3 Adeno-Associated Vectors

The adeno-associated virus (AAV) is among the smallest (20 nm) and simplest of the eukaryotic viruses.⁸ It is not pathogenic for humans, and the vectors derived from it turn out to be extremely efficient for gene transfer. It belongs to the family of dependoviruses and, a unique feature in the world of animal viruses, requires further genes from auxiliary viruses in order to replicate. These genes can be supplied by the adenovirus (whence the name, since the AAV was discovered in association with this virus) or by the herpes virus. The viral particles are made up mainly of proteins and DNA, in equal proportions.

⁸ Eukaryotes refer to the group of organisms (including animals and plants) whose cells have their genetic material enclosed within a structure called the nucleus, as opposed to prokaryotes (e.g., bacteria).

The capsid provides the virions with a considerable stability, making them extremely resistant to high temperatures (1 h at 50°C). This feature means that AAV vectors can be lyophilised after purification, which simplifies their conservation and transport.

The AAV genome is made up of linear single-stranded DNA about 5 kb long and carrying two genes: *rep* codes for the viral replication and integration functions, while *cap* codes for the structural proteins of the capsid (see Fig. 23.2). *Rep* and *cap* are framed by inverse terminal repeats (ITR). The ITR are required for encapsidating the genome in viral capsids and for integrating it within the cell chromosome. The viral vector is constructed by replacing the *rep* and *cap* genes by the expression cassette of the relevant transgene. The viral particles are produced by plasmid cotransfection in cell line 293, in order to produce in *trans* the *rep* and *cap* gene products, as well as the helper adenoviral genes required for the replication of the AAV.

Wild-type AAV can integrate into a well-defined site of human chromosome 19 under the action of the Rep protein. However, the vectors, in which the *rep* gene has been deleted, integrate randomly. AAV-derived vectors can infect both dividing cells and quiescent cells. In mice, dogs, and monkeys, the expression of transferred genes by AAV vectors is stable in cells with long lifetimes, such as cells in muscle, the liver, and the brain. Owing to the efficiency and stability of the transduction of muscle cells, an AAV2 vector, for example, has been successfully used to correct a mutation in dystrophin in a murine model of myopathy [18]. The long-term expression of AAV transgenes results both from the integrated viral genomes and also from the fraction of genomes remaining in extrachromosomal form.

The main drawback with AAV-derived vectors is their low cloning capacity. Indeed, only transgenes with sizes less than 4.7 kb can be included within these vectors, which restricts their use to transferring small genes under the control of equally small promoter sequences. However, this limitation can be overcome in some cases by using ‘shortened’ genes [19], such as the cystic fibrosis transmembrane conductance regulator (CFTR) minigene, silenced in patients suffering from cystic fibrosis. Recently, an alternative has been proposed, based on the concatemerisation of AAV genomes before integration. Cotransduction by two AAV vectors induces concatemers which, if they have been designed with suitably placed splice sites, allow the reformation of a whole molecule with length greater than 4.7 kb by a trans-splicing process after transcription [20].

23.3 Biomedical Applications of the Viral Platform

Gene transfer by viral vectors is used on humans and animals to treat and to model various pathologies [21]. Gene therapy refers to any application of gene transfer for therapeutic purposes. Conversely, the transfer of disease genes is used in animals to generate models of genetic disorders. These models help

to understand pathological mechanisms and to validate new therapeutic approaches.

23.3.1 Gene Therapy

Gene therapy can be defined as the deliberate transfer of genetic material in the patient organism to correct some specific defect at the origin of a pathology, as either a curative or a preventive measure [1]. In curative strategies, the gene transferred is usually a functional copy of the defective gene. Some anticancer approaches are based on the use of a suicide transgene that is toxic for tumour cells. Preventive strategies concern for example the transfer of genes coding for growth factors, e.g., the glial cell line-derived neurotrophic factor (GDNF), to limit the neuronal death that occurs in certain neurodegenerative processes. For ethical reasons, gene transfer in humans is limited to somatic cells, and excludes all modifications of germline cells that would then be transmitted to the patient's descendants.

The mutation at the origin of the pathology may correspond to the loss of function of a gene, or quite the opposite, to a gain of function. In the first case, the therapeutic strategy will be to overexpress the deficient gene by introducing a functional copy. In the second, one seeks to inhibit the expression of the mutated protein. This second approach is developing quickly thanks to the recent discovery of a general mechanism for inhibiting the expression of a particular gene, viz., RNA interference.

Overexpression

The most natural application of gene transfer is the expression of some chosen gene in a target cell (or the overexpression if the endogenous gene is already present). The proof of principle of the complementation of a genetic defect and the correction of an abnormal phenotype was obtained in 1983: a retroviral vector coding hypoxanthine phosphoribosyltransferase (HPRT) was able to restore the activity of the enzyme and normalise the metabolism of purines in the cells of patients suffering from the Lesh–Nyhan syndrome [22].

MLV-derived retroviral vectors were the first to be used in a gene therapy trial in 1990 and remain today the most frequently used. The patients taking part in this first trial were two children suffering from a disease affecting the immune system, viz., adenosine deaminase (ADA) deficiency. In this disorder, the lack of the ADA enzyme due to mutations in the corresponding gene causes an accumulation of adenosine or deoxyadenosine which is toxic for lymphocytes. Although the success of the therapy was not decisively demonstrated because the patients were receiving a conventional treatment at the same time, this protocol supplies, with more than 10 years' median duration, valuable information concerning the safety and efficiency of retroviral gene transfer [23].

The year 2000 was crucial for clinical gene therapy, with the announcement that two children suffering from X-chromosome-related human severe combined immunodeficiency (SCID-X1), commonly known as bubble babies, treated at the Necker Hospital in Paris, had been cured by grafting genetically modified hematopoietic stem cells [24]. The correction of the severe combined immune deficiency remains to this day the greatest success of gene therapy and illustrates the power, as well as the current limits, of this approach. Ten patients have been treated by transfer of the gene γ_c using an MLV-derived vector in hematopoietic stem cells ex vivo, i.e., cells were taken from the patient and then treated in the laboratory before being reimplanted. A stable correction of the immune deficiency was obtained in seven of them, with a median duration exceeding 4.5 years for the oldest trials [25], and this has been confirmed with other children by British and Italian teams. A similar strategy has been successfully used to treat two children suffering from ADA deficiency [26]. These examples illustrate the benefits of gene therapy where there is a selective advantage for the proliferation of corrected cells (here the lymphocytes originating from the transduced stem cells) over non-corrected cells (the patient's own non-functional lymphocytes). However, two of the children treated for SCID-X1 developed leukemia three years after the treatment, which then necessitated chemotherapy for both of them [25]. In both cases, the mechanism behind this was insertional mutagenesis within the locus of a proto-oncogene LMO-2: insertion of the provirus upstream of the gene causes an enhancer effect of the viral LTR on transcription and aberrant expression of LMO-2, probably responsible for the cell transformation.

These results demonstrate the importance of accurately weighing up the benefits against the risks when using gene therapy in each pathological context. For SCID-X1, it seems that the efficiency of gene therapy wins out over the traditional bone marrow transplant, in the absence of an identical HLA donor, where the toxicity, in particular, the reaction of the transplant against the host, is far from being negligible. To improve therapeutic safety, the use of lentiviral vectors, in which the LTR is inactivated, seems promising. In the longer term, another attractive approach will be targeted integration of the transgene [27], in particular by using an integrase with few targets in the human genome. This line of research is all the more attractive in that one may hope one day to correct genetic mutations directly in situ, by replacing the mutated sequence by the functional gene in the chromosomal locus.

The most comprehensive list of worldwide gene therapy clinical trials is available at the website of the *Journal of Gene Medicine* [28]. For the United States, an exhaustive list can be consulted at the site of the Office of Biotechnology Activities, NIH [29]. Over the period 1989–2007, 1,347 clinical gene therapy trials have been recorded around the world (total number of all approved, current, or completed protocols). More than 66% of these protocols concern cancer and are proposed to patients in the terminal phase. Most are

phase I or II trials, with around 3% in phase III.⁹ Viral vectors represent 70% of the different types of vectors used. Among these, retroviral and adenoviral vectors are the most common (22.3% and 24.8% respectively). There are now five international scientific reviews dedicated solely to gene therapy, which attests to the importance of this field in current biomedical research.

Concerning gene therapy for diseases of the nervous system, a great deal of effort has been devoted to experimental and preclinical research, although clinical protocols remain few and far between. Recently, a promising phase I trial has begun for Alzheimer's disease. The idea is to implant in the brain fibroblasts taken from the patient's skin and transduced *ex vivo* by an MLV vector coding the nerve growth factor (NGF) [31]. Generally speaking, the types of neurological disease that could benefit from such treatments are cancers, e.g., glioblastomas, monogenic diseases, e.g., lysosomal diseases, Huntington's disease, but also etiologically complex disorders such as Parkinson's and Alzheimer's diseases. In the case of Parkinson's disease, for example, our group has shown as part of an international collaboration [32] that expression of the glial cell line-derived neurotrophic factor (GDNF) by a lentiviral vector in a primate model can stop neurodegeneration and favour functional recovery (see Fig. 23.4). When preclinical studies on primates show that the benefit expected for a human patient is greater than with conventional treatment by pharmacology or neurosurgery, the gene therapy protocol is proposed for a human clinical trial.

Inhibition of Expression by RNA Interference

While gene transfer lends itself naturally to gene overexpression, it is much more difficult to inhibit the expression of some given gene. Now some genetic disorders are caused by the expression of a mutant protein, e.g., expression of the mutated huntingtin in Huntington's disease, which one would like to inhibit. Likewise in functional genomics, inhibiting the expression of a given gene in the animal proves to be a very effective approach for identifying the function of the corresponding protein, or for generating a model of the pathology. Consequently, the possibility of devising a general technique for inactivating given genes is a major undertaking. For a few years now, RNA interference (RNAi) has gradually been developing into a choice method in this area (in 2006, A. Fire and C. Mello were awarded the Nobel Prize for the discovery of RNA interference). RNAi is a physiological mechanism that inhibits gene expression at the translational level: production of small interfering RNA strands (siRNA) complementary to nucleotide sequences of the targeted mRNA and subsequent base-pairing between the two RNAs lead to their degradation by

⁹ In 2004, the first authorised commercial distribution of a gene therapy treatment appeared on the market in China, namely, an adenoviral vector coding for the protein p53, used to treat patients suffering from carcinomas on the head and neck [30].

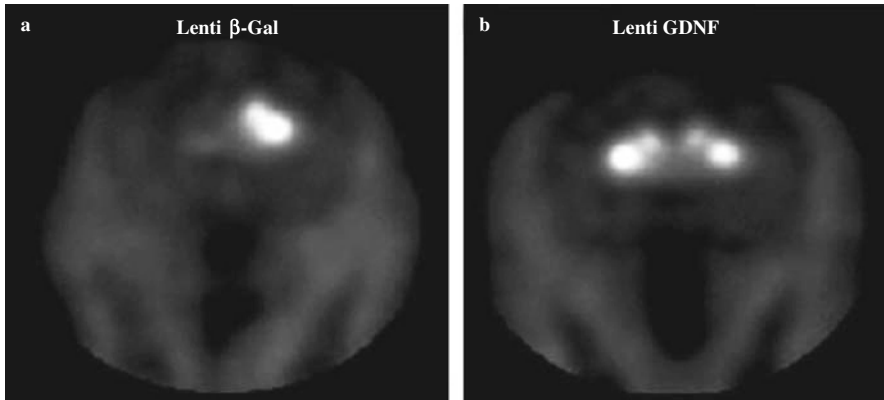


Fig. 23.4. Functional restoration in a primate by lentiviral transfer of the gene coding for GDNF. Brain activity is measured by positron emission tomography (PET) of fluorodopa (precursor of dopamine radio-tagged by the positron emitter ^{18}F). In the primate model of Parkinson's disease used here, the neuron activity of the left striatum is considerably diminished. Injection of the lentiviral vector coding for the glial cell line-derived neurotrophic factor (GDNF) restores neuron activity (b), in contrast to the control vector (a) containing the β -galactosidase reporter gene [32]

a specific enzymatic complex (see below; other less studied silencing outcomes of the RNA binding event are translation blocking and epigenetic changes to the gene itself). RNAi is a conserved process that is present in most eukaryotes in which its original function was a defense mechanism against viruses and foreign nucleic acids. Recent genetic studies suggest that it could also participate in the regulation of expression of endogenous genes.

The ability of double-stranded RNA to induce degradation of single-stranded RNA containing the same sequences was initially demonstrated in the nematode *Caenorhabditis elegans* [33]. Using data obtained from plants, fungi, the worm *C. elegans*, drosophila, and more recently, mammals, the main lines of the underlying mechanism have been brought to light (see Fig. 23.5). The double-stranded RNA molecules are cleaved by a double-stranded ribonuclease called Dicer, producing fragments containing about 21 nucleotides, the short interfering RNA (siRNA). These short RNA are then incorporated in the form of a single strand in a ribonucleoprotein complex, viz., the RNA-induced silencing complex (RISC), where they serve as a guide for recognition of the target mRNA. Perfect pairing with the target mRNA determines a sequence-specific nuclease activity of the complex with endonucleolytic cleavage and degradation of the fragments.

A further step was taken when it was shown that a specific gene inhibition could be obtained in mammalian cells by introducing synthetic siRNA containing 21 nucleotides [34]. Data since obtained by various groups has shown

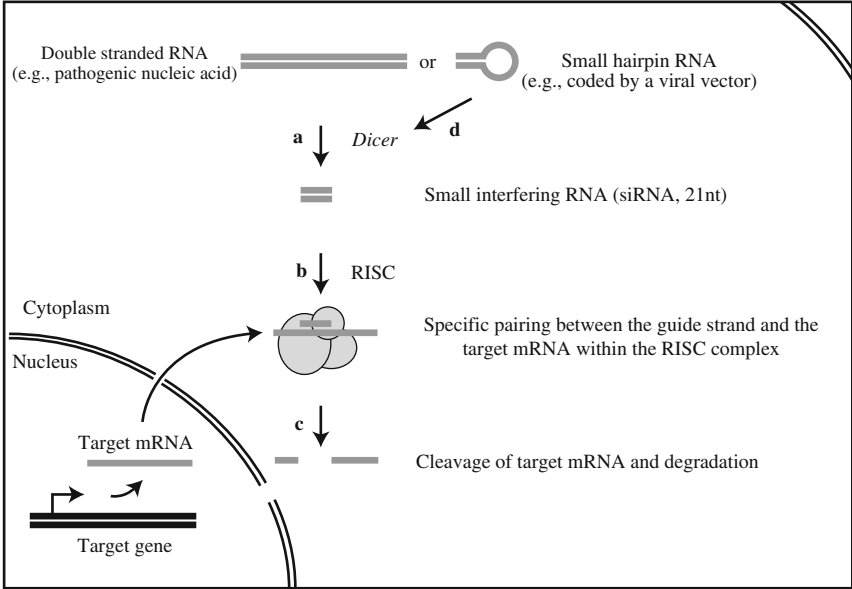


Fig. 23.5. The RNA interference mechanism and its use for inhibiting expression of a target gene. (a) Dicer cleaves the long double-stranded RNA into short interfering RNA (siRNA). (b) After associating with the RISC complex, the two strands dissociate and the remaining RNA strand guides the complex to localise the target messenger RNA (mARN). (c) After pairing with the siRNA, the mRNA is cleaved. (d) It has been shown that the RNA interference mechanism can be generalised to inhibit any chosen gene in vivo. To do this, a short RNA complementary to the relevant mRNA and with a hairpin structure is expressed in the cell via the viral vector. This RNA is then recognised by Dicer and cleaved into functional siRNA

statistically that the arbitrary choice of an siRNA sequence within a given gene produces a biological effect in almost half of the cases, demonstrating the efficiency of this approach.

In mammals, in contrast to plants, fungi, and nematodes, there does not seem to be an amplification or regeneration mechanism for double-stranded RNA. As a consequence, introducing short interfering RNA only temporarily silences expression, for anything between a few days and two weeks, depending on the amount of transfected RNA and the proliferation of the cells. To silence expression over longer periods, the simplest strategy is to use an expression vector which produces a short hairpin-shaped oligonucleotide, or short hairpin RNA (shRNA), in the cell. This is then cleaved by Dicer to form functional siRNA duplexes [35]. In a quite natural way, lentiviral vectors have been combined with the efficiency of RNA interference and a series of vectors coding for shRNA under the control of polymerase III (which transcribes most short cellular RNA) has been generated [36]. These vectors efficiently inhibit the

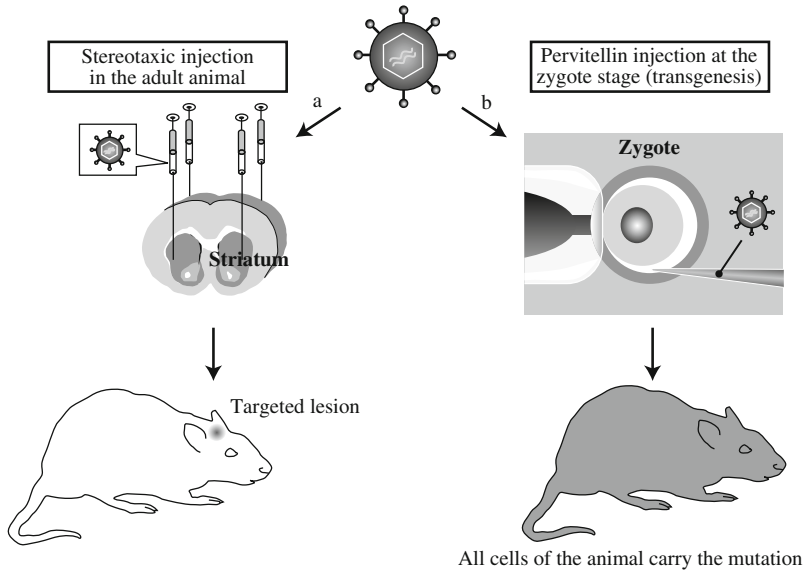


Fig. 23.6. Generation of animal models by somatic or germline lentiviral transfer. Two types of animal model can be generated using lentiviral vectors. (a) Stereotaxic injection is used to carry out rapid studies in the adult animal and in a targeted structure of the overexpression or inhibition of a given gene. (b) Infection of the zygote serves to transfer the chosen gene or interfering sequence to all cells in the animal. Lentiviral transgenesis has recently been validated in mice and rats. Stereotaxic injection is also used in primates

expression of cell genes in cell lines, hematopoietic stem cells, and also in transgenic mice [37].

One feature of RNA interference is the partial nature of the inhibition of gene expression. This partial inhibition may be considered as one of the strong points of the approach, because it provides a way of analysing the function of essential genes, and more generally, through hypomorphic alleles, studying the relation between the level of expression and cell physiology. Thanks to its ease of implementation, RNA interference has in just a few years become the best technique for inhibiting the expression of a gene in cultured mammalian cells, either temporarily or in the long term, and perhaps even in vivo. The first clinical trials for siRNA were presented during the 2005 conference of the American Society for Gene Therapy, for patients suffering from macular degeneration.

23.3.2 Animal Models of Human Pathologies

In order to understand the mechanisms underlying neurodegeneration and thereby develop new therapeutic methods, it is essential to have relevant animal models in rodents and, for preclinical studies, in primates. Since the 1980s,

the advent of transgenic mice has provided a way of studying the pathological consequences of the mutation of a protein on the organism. However, the classical transgenesis technique is extremely inefficient in mammals other than mice. Therefore, new approaches needed to be developed for rats and primates, which are better suited for the investigation of complex behaviour, and whose larger body size facilitates imaging (spatial resolution of positron emission tomography is in the millimeter range).

The development of lentiviral vectors provides new techniques to generate rat and primate models by virtue of the direct transfer of the gene into the brain of the adult animal [38], and more recently [39], thanks to a new technique of transgenesis by infection of the zygote (see Fig. 23.6).

Intracerebral Injection

In the central nervous system, stereotaxic techniques are used to inject the viral vector in a precise way into the relevant cerebral structure by using spatial coordinates listed in an atlas (see Fig. 23.7). This technique has been used to study the consequences of the expression of a mutated gene on the functioning of the brain. Compared with conventional transgenesis in mice, it allows one to work with rats and primates. Moreover, it is very fast. Cloning of the transgene, production of the vector, and injection of the vector in the animal can all be achieved within a few weeks, compared with several months for conventional transgenesis. Finally, the resulting degeneration is severe, probably due to strong expression of the mutated transgene, and the later phases of the disease can be observed in a much shorter lapse of time [40].

Lentiviral and adeno-associated vectors have been used to develop models of disease in the human nervous system [41]. For example, transfer of mutated α -synuclein in the substantia nigra brain region of healthy rats or primates leads to specific dopaminergic neuron loss in the striatum, accompanied by the appearance of cell inclusions characteristic of Parkinson's disease [42, 43].

With the same approach, our group also developed a model of Huntington's disease in rats.¹⁰ We demonstrate that overexpression of the mutant form of huntingtin in the striatum triggers the formation of nuclear aggregates and striatal degeneration characteristic of Huntington's disease [38]. Quite

¹⁰ Huntington's disease is a hereditary disorder characterised by specific degeneration of GABAergic neurons in the striatum, a particular cerebral structure. This degeneration leads to motor and cognitive problems. Huntington's disease often begins at an age between 30 and 50 years. It results in death after evolving for some fifteen years. In western countries, about one person in 10,000 is affected. At the present time, there is no commercially available treatment for stopping or even slowing down the neurodegenerative process. In 1993, the gene responsible for Huntington's disease was identified on chromosome 4 by an international consortium. This gene codes for huntingtin (Htt) and carries a CAG triplet coding for glutamine in its N-terminal region. Alleles containing more than 36 repetitions of the triplet are pathogenic.

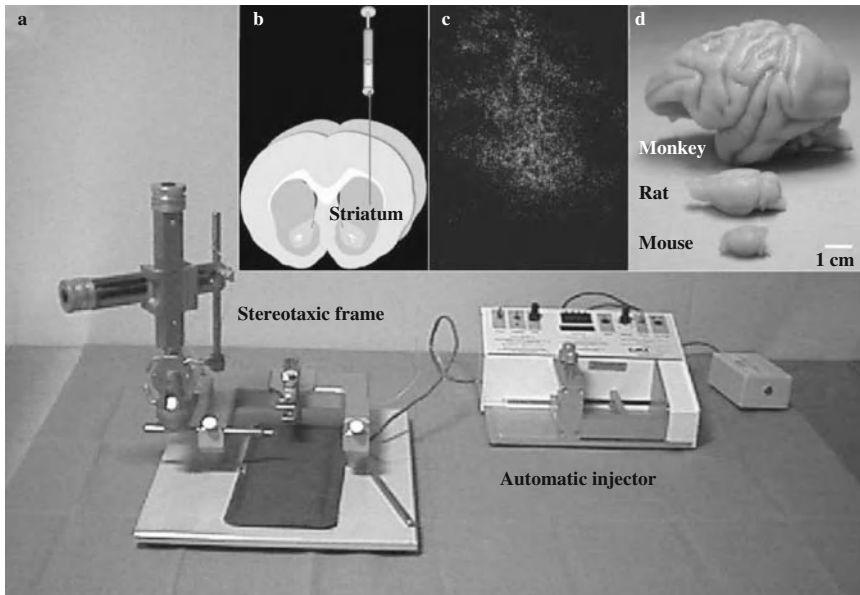


Fig. 23.7. Intracerebral injection by stereotaxy. (a) Stereotaxy is a general neuro-surgical technique for humans and animals. (b) The idea is to introduce a nozzle into the desired structure of the brain, e.g., striatum, using spatial coordinates listed in an atlas. (c) By injecting a few microlitres of concentrated lentiviral vector, several hundred thousand cells can be transduced. (d) Comparing the size and morphology of the brain in monkeys and rodents

remarkably, we observed that the pathology induced is more precocious and more serious as the number of repetitions of the CAG triplet increases (in agreement with clinical observations) or when the level of expression of the mutated protein increases.

Transgenesis

The idea of transgenesis is to have the gene of interest inserted into the genome at the zygote stage in order to subsequently generate an animal whose cells all contain the transgene. This is a complementary approach to somatic gene transfer by direct injection of viral vectors. While the injection approach is best suited for restricted transgene expression in one specific structure, transgenesis is helpful when broad expression in all the tissues of the animal is required. So for example, the transgenic mice generated as a model for Huntington's disease express the mutant huntingtin not only in the striatum, but also in the cortex, another region playing an important role in the development of the disease in humans. Another advantage of transgenesis is that one can generate animal lines expressing the transgene in a way that is strictly

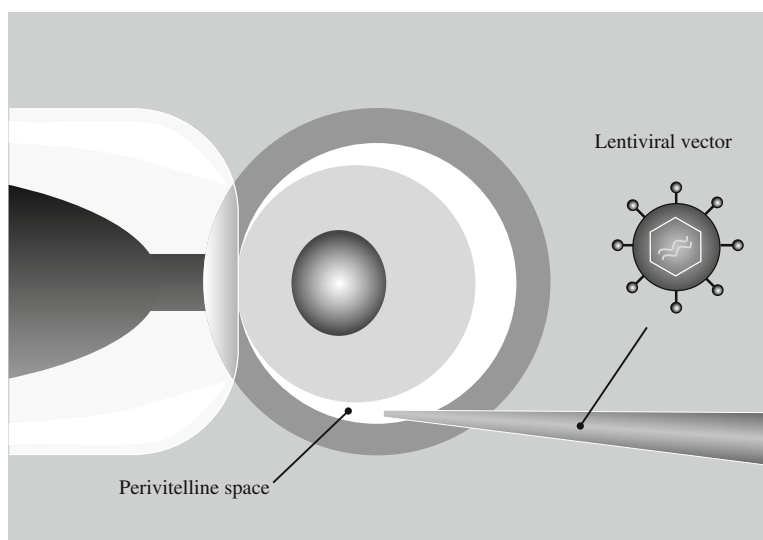


Fig. 23.8. Transgenesis by lentiviral infection of the zygote. A few picolitres of concentrated lentiviral stock are injected into the perivitelline space of the zygote between the cytoplasm and the zona pellucida

identical from one animal to another, thereby avoiding any risk of variability that may occur during the intracerebral injection procedure. Furthermore, transgenic lines are maintained by simple cross-breeding, without the need to operate on each animal. Finally, applications of transgenesis in cattle open up prospects for generating animal lines that are more resistant to certain disorders [44, 45].

In 1980 the technique whereby linear DNA is micro-injected into the pronucleus of the fertilised oocyte produced the first transgenic mice. Since then, the standard technology for studying the expression of a given gene in the organism is the generation of a transgenic mouse lineage. However, one of the limitations of pronuclear DNA injection is its low efficiency: the transgene is integrated and expressed in less than 2% of injected and transferred embryos [46]. In farm animals like pigs and bullocks, the efficiency is even lower for technical reasons, in particular, the opacity of the zygote cytoplasm makes it more difficult to inject into the pronucleus. Cattle thus represent a small proportion of publications on transgenesis (less than 5% in 1996) and production costs are high (30,000 US dollars for a pig and 300,000 US dollars for a bullock).

In 2002, transgenesis became a new field of application for lentiviral vectors. Indeed, two groups showed in mice and rats that infection of a zygote by a lentiviral vector can generate a transgenic animal [39, 47]. The technique involves injecting the viral solution in the perivitelline space between the cytoplasmic membrane and the zona pellucida (outer layer protecting the embryo)

(see Fig. 23.8). This new technique has the advantage of being more efficient than conventional pronuclear injection [46]. For example, in 2004, our own research center, as part of a consortium for lentiviral transgenesis, obtained an average of 16% transgenic animals among all reimplanted embryos, representing a yield several times higher than with pronuclear injection, and leading to the successful generation of several rat transgenic lines.

Apart from the results obtained in laboratory animals (rat and mouse), several recent publications describe applications of the perivitelline lentiviral injection technique to cattle and chickens. Pfeifer et al. generated transgenic pigs with 27 times greater efficiency than micro-injection [48], and obtained preliminary results with cattle [49]. With regard to birds, transgenic chickens have been generated by injecting a horse lentiviral vector into the egg [50]. Using transgenesis and the complete sequencing of the whole chicken genome in the near future, these birds will become a choice model for functional studies of genes in vertebrates.

One restriction on lentiviral transgenesis as compared with pronuclear DNA injections is the size of the transgene, which must be less than 8.5 kb. However, this is large enough for most regions coding genes and also for interfering RNAs. Another question concerning perivitelline lentiviral injection is whether this technique will finally provide a way of generating transgenic monkeys. The first results by blastocyst infection have been able to produce integration and expression of GFP in the placenta, but not in the newly born animal [51].¹¹ Alternative approaches are also being developed, such as in vitro transduction of male germline stem cells (spermatogonia) before reimplanting in the animal to generate mature spermatozooids [53]. Lentiviral vectors have become a key player in the field of transgenesis over the past few years.

23.4 Controlling and Visualising Transgene Expression

In order to control in vivo gene transfer even more precisely, two types of improvement can be made to the cassette containing the transgene. On the one hand, regulation systems have been developed over the past ten years in order to induce or repress the expression of the transgene. On the other hand, imaging techniques have recently been adapted to visualise and monitor the expression of the transgene in vivo after transfer into the organism.

23.4.1 Controlling Transgene Expression

The possibility of regulating the expression of a transgene or interfering sequence is essential not only for fundamental research when studying the reversible nature of a phenotype, but also for clinical research, e.g., to set up a

¹¹ Preliminary results on the successful generation of a transgenic primate model of Huntington's disease by lentiviral transgenesis have been reported recently [52].

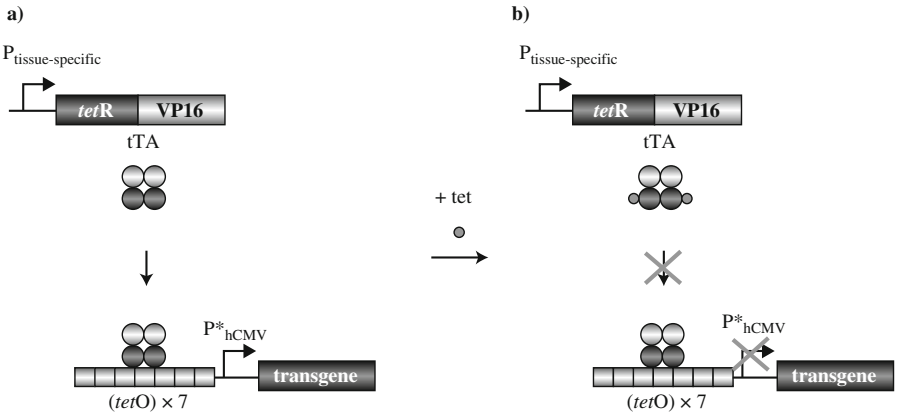


Fig. 23.9. Tet-OFF tetracycline regulation system. **(a)** This is a two-component system. On the one hand it includes the chimeric gene coding the transcriptional activation factor tTA. This protein is a fusion between (1) the DNA binding domain of the bacterial repressor TetR, and (2) the transactivator domain of the protein VP16 of the herpes simplex virus (HSV). By using a specific promoter $P_{\text{tissue specific}}$, the restricted cell expression of the tTA gene can be controlled. On the other hand, the relevant gene is placed under the control of a minimal promoter P^*_{CMV} combined with a concatamer of tetO sequences, binding sites of TetR. In the absence of tetracycline (or its analogue doxycycline), tTA binds to the tetO sequences and activates expression of the given gene. **(b)** In the presence of the antibiotic, tTA cannot bind to the operator sequences and transcription of the transgene is repressed

feedback control of the expression of a therapeutic protein via some physiological regulation (as for insulin-dependent diabetes) or to adjust expression of a drug gene to the desired dosage.

The tetracycline regulation system is the most widely used and hence the best understood, thanks to its relative simplicity and the good tissue penetration and tolerance of this antibiotic (and its analogue doxycycline) in vivo. This system can trigger repression of a given transgene by systemic administration of the antibiotic in the organism (the tet-OFF system). It was developed by Bujard and coworkers using regulation elements of the tetracycline-resistance operon encoded in Tn10 of *E. coli* [54]. It comprises a transactivator tTA able to bind onto the tetO sequences and activate expression of the transgene located downstream (see Fig. 23.9). In the presence of tetracycline, tTA cannot bind to the tetO sequences, whence transcription of the transgene is stopped. An analogous system, this time for inducing expression of the transgene (the tet-ON system), has been obtained by isolating a mutant of the tTA gene called reverse tTA (rtTA) which, in contrast to tTA, requires the presence of the antibiotic to bind onto the tetO sequences and activate expression of the transgene [55].

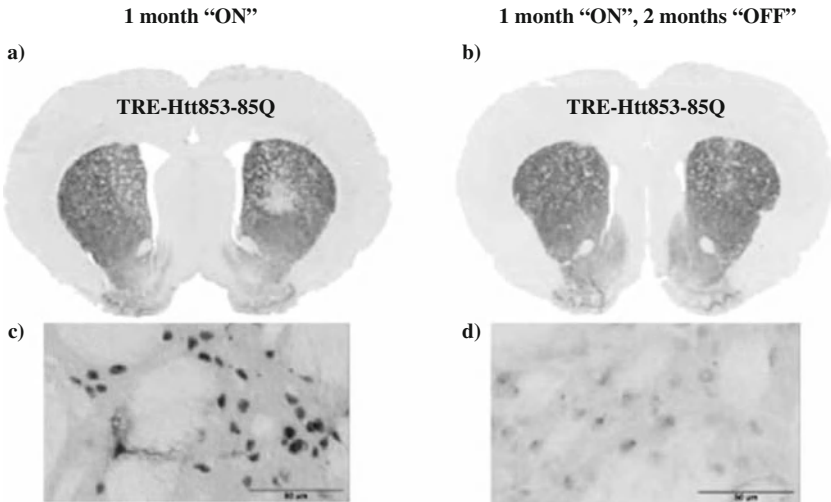


Fig. 23.10. Reversibility of the Huntington neuropathology induced by a regulated lentiviral vector. The transgene TRE-Htt853-82Q codes a mutant form of huntingtin (Htt) containing an expansion of 82 glutamines. It is placed downstream of the tetO sequences of the Tet-OFF system. This vector is coinjected into the rat striatum with a second vector coding the transactivator tTA. (a) When the animals are not treated with doxycycline, mutant Htt is expressed (ON) and a lesion corresponding to loss of the neuronal marker DARPP-32 appears one month after injection. (c) Accumulation of aggregates containing Htt observed in these animals confirms the advance of the pathological process. (b) and (d) In contrast, in rats which have in addition received doxycycline in their drinking water during the two following months to repress expression of the mutated transgene (OFF), the recovery of 61% of the DARPP-32 marker, accompanied by the disappearance of the Htt aggregates, is observed [40]. With the kind permission of Oxford University Press

In viral vectors, the two cassettes of the tetracycline system (the *trans*-regulator and the relevant transgene) can be incorporated separately in two vectors, which are then co-injected [40], or encoded together by a single vector [56].

Regulation systems are widely used in therapeutic strategies to control the level of transgene expression, but also in animal models to study the physiological effects in ON and OFF conditions. For models of Huntington's disease in rodents, for example, our group has shown that silencing the expression of mutant huntingtin after one month's expression leads to the disappearance of aggregates and the reappearance of neuron survival markers (see Fig. 23.10). These results, and those obtained in mice, also using the tetracycline system, show that continuous expression of the mutant protein is needed to maintain inclusions and symptoms and suggests that the degenerative process of Huntington's disease is in fact reversible.

23.4.2 Imaging Transgene Expression

A burgeoning field of research today concerns the *in vivo* imaging of transgene expression once the transgene has been transferred into the organism. Indeed it is essential to be able to monitor expression of the transgene after injecting the viral vector, in particular to control its biodistribution, expression level, and stability over time. Up until the last few years, transgene expression could only be studied post mortem, owing to the lack of imaging equipment with adequate resolution. Today, this aim can be achieved thanks to the development of more powerful equipment designed for use on small animals in the various imaging modalities now available.¹² Transgene expression imaging falls within the more general context of molecular imaging. This discipline, at the intersection between molecular biology and *in vivo* imaging, aims to detect expression of a gene, either endogenous or transferred into the organism, by means of a non-invasive procedure in animals and humans [57].

One way of visualising the transgene is to insert a second so-called reporter gene into the vector. The product of this second gene causes accumulation in the cell of a probe that can be detected by imaging. In PET imaging, the gene for viral thymidine kinase HSV1-*tk* has thus been used as a reporter gene with radio-tagged probes [¹²⁴I]-FIAU (derivative of uracil) or [¹⁸F]-FHBG (derivative of pencyclovir). In the transduced cell, the viral thymidine kinase phosphorylates the probe which can then no longer efficiently pass through the membrane and is thus trapped within the cell [58]. In MRI, the magnetic properties of ferritin, the main protein for iron storage, have recently been used to monitor expression of this protein over time and in the same animal after transferring the corresponding gene in the brain by means of an adenoviral vector [59]. The combined development of multimodality fusion reporter genes and new ‘intelligent’ contrast agents, able to target certain cell types, is currently the basis for longitudinal monitoring of gene transfer in humans and animals.

23.5 Prospects

In little over ten years, viral vectors have become an indispensable platform for *in vivo* gene transfer. Current research in vectorology focuses in particular on vector tropism and control of the chromosome integration site. With the

¹² There are three main types of imaging modality (apart from ultrasound, whose applications to molecular imaging are nevertheless more restrictive). Nuclear imaging, and especially positron emission tomography (PET), is based on the idea of injecting a radioactive tracer, and has the highest sensitivity. Magnetic resonance imaging (MRI) uses the magnetic properties of hydrogen atoms in water and organic compounds. It is non-invasive and has high spatial resolution. Finally, optical imaging (bioluminescence) stands out by its relative ease of use and low cost.

development of molecular imaging, biology must combine with other fields, such as physics, e.g., for the development of new MRI contrast agents, but also mathematics and computing for data processing and image reconstruction. Progress in controlling target cell transduction and regulating and visualising transgene expression in situ are major advances toward successful clinical treatment through gene transfer.

Acknowledgements

The authors would like to thank Pascal Le Masson and Chamsy Sarkis, together with Aurélie Delzor, Mario Lepore, Jean-Charles Robillard, Anne-Sophie Chaplault, Sandro Alvès, Raymonde Hassig, and Carole Escartin for their invaluable contributions to this chapter.

References

1. Gros, F.: La thérapie génique, Rapport Académie des Sciences **36** (1995)
2. Andre, F., Mir, L.M.: DNA electrotransfer: Its principles and an updated review of its therapeutic applications, *Gene Ther.* **11** (Suppl. 1), S33–S42 (2004)
3. Goins, F.W., et al.: Delivery using herpes simplex virus: An overview, *Methods Mol. Biol.* **246**, 257–299 (2004)
4. Naldini, L., et al.: In vivo gene delivery and stable transduction of nondividing cells by a lentiviral vector, *Science* **272** (5259), 263–267 (1996)
5. Zufferey, R., et al.: Woodchuck hepatitis virus posttranscriptional regulatory element enhances expression of transgenes delivered by retroviral vectors, *J. Virol.* **73** (4), 2886–2892 (1999)
6. Barquinerio, J., Perez-Melgosa, M.: Retroviral vectors: New applications for an old tool, *Gene Ther.* **11** (Suppl. 1), S3–S9 (2004)
7. Delenda, C.: Lentiviral vectors: Optimization of packaging, transduction and gene expression, *J. Gene Med.* **6** Suppl. 1, S125–S138 (2004)
8. Tenenbaum, L., et al.: Recombinant AAV-mediated gene delivery to the central nervous system, *J. Gene Med.* **6** (Suppl. 1), S212–S222 (2004)
9. Verma, I.M., Weitzman, M.D.: *Annu. Rev. Biochem.* **74**, 711–738 (2005)
10. Sastry, L., et al.: Titering lentiviral vectors: Comparison of DNA, RNA and marker expression methods, *Gene Ther.* **9** (17), 1155–1162 (2002)
11. Escarpe, P., et al.: Development of a sensitive assay for detection of replication-competent recombinant lentivirus in large-scale HIV-based vector preparations, *Mol. Ther.* **8** (2), 332–341 (2003)
12. Sastry, L., et al.: Certification assays for HIV-1-based vectors: Frequent passage of gag sequences without evidence of replication-competent viruses, *Mol. Ther.* **8** (5), 830–839 (2003)
13. Stratford-Perricaudet, L.D., et al.: Evaluation of the transfer and expression in mice of an enzyme-encoding gene using a human adenovirus vector, *Hum. Gene Ther.* **1** (3), 241–256 (1990)
14. Le Gal La Salle, G., et al.: An adenovirus vector for gene transfer into neurons and glia in the brain, *Science* **259** (5097), 988–990 (1993)

15. Corti, O., et al.: Long-term doxycycline-controlled expression of human tyrosine hydroxylase after direct adenovirus-mediated gene transfer to a rat model of Parkinson's disease, *Proc. Natl. Acad. Sci. USA* **96** (21), 12120–12125 (1999)
16. Thévenot, E., et al.: Targeting conditional gene modification into the serotonin neurons of the dorsal raphe nucleus by viral delivery of the Cre recombinase, *Mol. Cell Neurosci.* **24** (1), 139–147 (2003)
17. Volpers, C., Kochanek, S.: Adenoviral vectors for gene transfer and therapy, *J. Gene Med.* **6** Suppl. 1, S164–S171 (2004)
18. Goyenvalle, A., et al.: Rescue of dystrophic muscle through U7 snRNA-mediated exon skipping, *Science* **306** (5702), 1796–1799 (2004)
19. Sirninger, J., et al.: Functional characterization of a recombinant adeno-associated virus 5-pseudotyped cystic fibrosis transmembrane conductance regulator vector, *Hum. Gene Ther.* **15** (9), 832–841 (2004)
20. Nakai, H., Storm, T.A., Kay, M.A.: Increasing the size of rAAV-mediated expression cassettes in vivo by intermolecular joining of two complementary vectors, *Nat. Biotechnol.* **18** (5), 527–532 (2000)
21. Deglon, N., Hantraye, P.: Viral vectors as tools to model and treat neurodegenerative disorders, *J. Gene Med.* **7** (5), 530–539 (2005)
22. Miller, A.D., et al.: A transmissible retrovirus expressing human hypoxanthine phosphoribosyltransferase (HPRT): Gene transfer into cells obtained from humans deficient in HPRT, *Proc. Natl. Acad. Sci. USA* **80** (15), 4709–4713 (1983)
23. Muul, L.M., et al.: Persistence and expression of the adenosine deaminase gene for 12 years and immune reaction to gene transfer components: Long-term results of the first clinical gene therapy trial, *Blood* **101** (7), 2563–2569 (2003)
24. Cavazzana-Calvo, M., et al.: Gene therapy of human severe combined immunodeficiency (SCID)-X1 disease, *Science* **288** (5466), 669–672 (2000)
25. Hacein-Bey-Abina, S., et al.: LMO2-associated clonal T cell proliferation in two patients after gene therapy for SCID-X1, *Science* **302** (5644), 415–419 (2003)
26. Aiuti, A., et al.: Correction of ADA-SCID by stem cell gene therapy combined with nonmyeloablative conditioning, *Science* **296** (5577), 2410–2413 (2002)
27. Bushman, F.D.: Targeting survival: Integration site selection by retroviruses and LTR-retrotransposons, *Cell* **115** (2), 135–138 (2003)
28. www.wiley.co.uk/genmed/clinical
29. www4.od.nih.gov/oba/rdna.htm
30. Pearson et al.: *Nat. Biotechnol.* **2**, 3–4 (2004)
31. Tuszynski, M.H., et al.: A phase 1 clinical trial of nerve growth factor gene therapy for Alzheimer disease, *Nat. Med.* **11** (6) 551–555 (2005)
32. Kordower, J.H., et al.: Neurodegeneration prevented by lentiviral vector delivery of GDNF in primate models of Parkinson's disease, *Science* **290** (5492), 767–773 (2000)
33. Fire, A., et al.: Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*, *Nature* **391** (6669), 806–811 (1998)
34. Elbashir, S.M., et al.: Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells, *Nature* **411** (6836), 494–498 (2001)
35. Brummelkamp, T.R., Bernards, R., Agami, R.: A system for stable expression of short interfering RNAs in mammalian cells, *Science* **296** (5567), 550–553 (2002)
36. Abbas-Terki, T., et al.: Lentiviral-mediated RNA interference, *Hum. Gene Ther.* **13** (18), 2197–2201 (2002)

37. Robinson, D.A., et al.: A lentivirus-based system to functionally silence genes in primary mammalian cells, stem cells and transgenic mice by RNA interference, *Nat. Genet.* **33** (3), 401–406 (2003)
38. de Almeida, L.P., et al.: Lentiviral-mediated delivery of mutant huntingtin in the striatum of rats induces a selective neuropathology modulated by polyglutamine repeat size, huntingtin expression levels, and protein length, *J. Neurosci.* **22** (9), 3473–3483 (2002)
39. Lois, C., et al.: Germline transmission and tissue-specific expression of transgenes delivered by lentiviral vectors, *Science* **295** (5556), 868–872 (2002)
40. Regulier, E., et al.: Early and reversible neuropathology induced by tetracycline regulated lentiviral overexpression of mutant huntingtin in rat striatum, *Hum. Mol. Genet.* **12** (21), 2827–2836 (2003)
41. Kirik, D., Bjorklund, A.: Modeling CNS neurodegeneration by overexpression of disease-causing proteins using viral vectors, *Trends Neurosci.* **26** (7), 386–392 (2003)
42. Lo Bianco, C., et al.: Alpha-synucleinopathy and selective dopaminergic neuron loss in a rat lentiviral-based model of Parkinson's disease, *Proc. Natl. Acad. Sci. USA* **99** (16), 10813–10818 (2002)
43. Kirik, D., et al.: Nigrostriatal alpha-synucleinopathy induced by viral vector-mediated overexpression of human alpha-synuclein: A new primate model of Parkinson's disease, *Proc. Natl. Acad. Sci. USA* **100** (5), 2884–2889 (2003)
44. Wall, R.J., et al.: Genetically enhanced cows resist intramammary *Staphylococcus aureus* infection, *Nat. Biotechnol.* **23** (4), 445–451 (2005)
45. Denning, C., et al.: Deletion of the alpha(1,3)galactosyl transferase (GGTA1) gene and the prion protein (PrP) gene in sheep, *Nat. Biotechnol.* **19** (6), 559–562 (2001)
46. Fassler, R.: Lentiviral transgene vectors, *EMBO Rep.* **5** (1), 28–29 (2004)
47. Pfeifer, A., et al.: Transgenesis by lentiviral vectors: Lack of gene silencing in mammalian embryonic stem cells and preimplantation embryos, *Proc. Natl. Acad. Sci. USA* **99** (4), 2140–2145 (2002)
48. Hofmann, A., et al.: Efficient transgenesis in farm animals by lentiviral vectors, *EMBO Rep.* **4** (11), 1054–1060 (2003)
49. Hofmann, A., et al.: Generation of transgenic cattle by lentiviral gene transfer into oocytes, *Biol. Reprod.* **71** (2), 405–409 (2004)
50. McGrew, M.J., et al.: Efficient production of germline transgenic chickens using lentiviral vectors, *EMBO Rep.* **5** (7), 728–733 (2004)
51. Wolfgang, M.J., et al.: Rhesus monkey placental transgene expression after lentiviral gene transfer into preimplantation embryos, *Proc. Natl. Acad. Sci. USA* **98** (19), 10728–10732 (2001)
52. Yang, S.H., et al.: Towards a transgenic model of Huntington's disease in a non-human primate, *Nature* **453**, 921–924 (2008)
53. Hamra, F.K., et al.: Production of transgenic rats by lentiviral transduction of male germ-line stem cells, *Proc. Natl. Acad. Sci. USA* **99** (23), 14931–14936 (2002)
54. Gossen, M., Bujard, H.: Tight control of gene expression in mammalian cells by tetracycline-responsive promoters, *Proc. Natl. Acad. Sci. USA* **89** (12), 5547–5551 (1992)
55. Urlinger, S., et al.: Exploring the sequence space for tetracycline-dependent transcriptional activators: Novel mutations yield expanded range and sensitivity, *Proc. Natl. Acad. Sci. USA* **97** (14), 7963–7968 (2000)

56. Vogel, R., et al.: A single lentivirus vector mediates doxycycline-regulated expression of transgenes in the brain, *Hum. Gene Ther.* **15** (2), 157–165 (2004)
57. Tavitian, B.: In vivo imaging with oligonucleotides for diagnosis and drug development, *Gut* **52** Suppl. 4, 40–47 (2003)
58. Shah, K., et al.: Molecular imaging of gene therapy for cancer, *Gene Ther.* **11** (15), 1175–1187 (2004)
59. Genove, G., et al.: A new transgene reporter for in vivo magnetic resonance imaging, *Nat. Med.* **11** (4), 450–454 (2005)

Pharmaceutical Applications of Nanoparticle Carriers

B. Heurtault, F. Schuber, and B. Frisch

24.1 Introduction to Drug Delivery in Pharmaceuticals

Once it has been administered, an active principle still has to face many physiological barriers on the way to its target, and this may significantly affect its efficiency. These different barriers depend to a great extent on the active ingredient itself and on the way it is administered. They may be constituted by enzymes, an acidic or basic pH, or cell membranes that must be crossed. As a consequence, the active principle may be degraded or distributed to organs other than the therapeutic target. This can reduce the efficiency of the administered dose, or even lead to toxicity with regard to organs other than the target. For example, this situation is observed in trials for the oral administration of insulin (for treating type I diabetes). One point is that this molecule is weakly absorbed by the digestive epithelium (first barrier). Secondly, it undergoes enzymatic degradation by gastric proteases (second barrier). As a consequence, the free form of the molecule cannot be administered orally. This is why insulin is mainly administered subcutaneously, so that it attains the blood circulation directly. However, such a means of administration requires specific training of the patient. This example shows that lack of efficiency and/or difficulties in using certain molecules are not necessarily due to their pharmacology, but rather in some cases to their physicochemical properties.

Drug delivery involves the use of a vehicle whose role is to transport the molecule to the target, while protecting it and masking its physicochemical properties so that it may pass through physiological barriers. The distribution of the carrier is governed entirely by its own physicochemical properties. This is why active principles that are only weakly soluble in water are particularly relevant to drug delivery. Returning to the insulin example, protecting it by carriers like polymer nanoparticles has led to encouraging results. This may mean that it could be administered by much simpler means, e.g., orally or intranasally [1–4].

The aim of drug delivery is thus to improve the efficiency of a molecule, while limiting its toxicity. Having encapsulated or bound the active molecule, the main characteristics of the ideal carrier are:

- to protect the active molecule, e.g., from enzyme attack or acid pH,
- to transport the active molecule through cell membranes in cases where its physicochemical properties prevent it from doing so alone,
- to target the region to be treated and concentrate the active molecule there,
- to release the active principle in a controlled way.

To favour the diffusion of carriers through the organism and in the relevant cases to favour their passage through cell membranes, small dimensions are a major asset. This is why we shall be concerned in this chapter with nanoparticle carriers, bearing in mind however that microparticle carriers are also used, in the form of implants, for example. In addition, given the desired pharmaceutical application, the carrier must not exhibit any intrinsic toxicity, whatever means of administration is chosen. And in the field of drug delivery, all administrative routes remain suitable, i.e., cutaneous, oral, ocular, nasal, injection, etc.

Nanoparticle carriers in pharmaceuticals belong to the fast-developing field of nanotechnology. For this reason, the delivery of molecules for therapeutic, diagnostic, or control purposes is subsumed under what has recently been called nanomedicine by the National Institutes of Health [5].

In the following, we begin by presenting the various types of nanoparticle carriers and their classification, characteristics, and ability to encapsulate active molecules. Then, in Sect. 24.3, we discuss surface modifications involved in the preparation of the so-called stealth particles and carriers able to deliver an active principle in a specific manner. Finally, in Sect. 24.4, we describe the applications of these carriers, classifying them in terms of the relevant medical speciality.

24.2 Nanoparticle Carriers

24.2.1 The Main Nanoparticle Carriers

Particle carriers are organised assemblies of molecules. These carriers can be classified in terms of their size, from nanoparticles (sizes less than $1\text{ }\mu\text{m}$) to microparticles (sizes greater than $1\text{ }\mu\text{m}$). Microparticles with their large dimensions are generally used as implants, or administered via channels other than the veins to avoid any risk of embolisation. These ‘large’ carriers will not be discussed in this chapter. For their part, nanoparticles have the advantage that they can be administered by any desired route.

But size is not the only criterion used to distinguish these particles. They can also be classified in terms of their structure: matrix structure for spheres

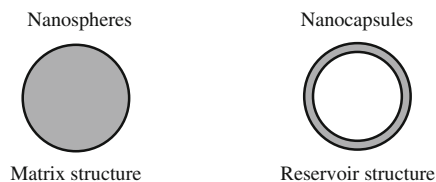


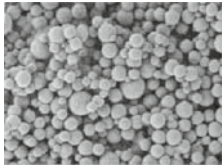
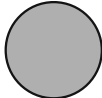
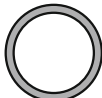
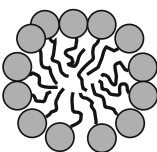
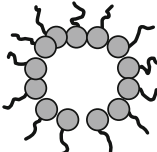
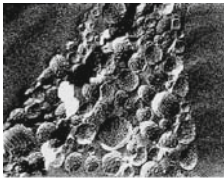
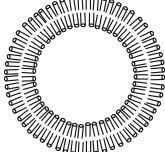
Fig. 24.1. Nanoparticle structures

or reservoir structure for capsules (see Fig. 24.1). The matrix structure is homogeneous, in contrast to the reservoir structure, which consists of a polymer or lipid wall surrounding a hydrophilic or lipophilic core. The type and amount of active principle encapsulated will depend on this structure. In the case of particles with nanometric sizes, one speaks of nanoparticles, which includes nanospheres with matrix structure and nanocapsules with reservoir structure.

Another important criterion is the composition of the nanoparticles. Some examples are:

- Polymer nanoparticles made from natural or synthetic polymers. In pharmaceuticals, the polymers used are mainly biodegradable polymers such as poly(lactic acid) (PLA), poly(glycolic acid) (PGA), or their copolymers (PLA/PGA), or again bioresorbable (bioerodable) polymers such as the polyalkylcyanoacrylates [6].
- Solid lipid nanoparticles (SLN), resulting from technological developments in the 1990s [7]. SLN are made from solid lipids at the physiological temperature. They are usually triglycerides or waxes [8]. Their stability and high tolerance make them excellent carriers, especially for parenteral administration [8]. The advent of SLN led to the development of lipid drug conjugates (LDC) and nanostructured lipid carriers (NLC), which will be described on p. 1104 in the discussion of the carrier–active principle association.
- Micelles made from surfactant molecules arranged in spherical monolayers in an aqueous medium, which lead to the formation of a hydrophobic reservoir. In contrast, reverse micelles have structures which, in an organic medium, contain a hydrophilic reservoir. Di- or triblock copolymers such as poly(ethylene oxide)–poly(propylene oxide)–poly(ethylene oxide) (PEO–PPO–PEO) can self-assemble in an aqueous solution to form polymer micelles [9]. Micelle solutions are used in particular to solubilise lipophilic or amphiphilic molecules. They form totally transparent (isotropic) solutions.
- Liposomes are the carriers with the longest history. Indeed, these were first studied in the 1960s by Bangham [10]. They are made from phospholipids which arrange themselves into spherical unilamellar bilayers (small or large unilamellar vesicles, SUV or LUV) or into multilamellar vesicles (MLV) in an aqueous medium. These liposomes can measure several micrometers across. Many reviews are devoted exclusively to liposomes and one of the most recent describes the evolution of these structures over the

Table 24.1. Characteristics of the main nanoparticle carriers. **(A1)** PLGA/PVA nanoparticles observed by scanning electron microscope. Average diameter 300 nm, $\times 25,000$ [13]. **(A2)** Schematic view of a nanosphere. **(A3)** Schematic view of a nanocapsule. **(B1)** Schematic view of a micelle, comprising a lipophilic core (apolar lipid chains of the surfactant), surrounded by the polar heads of the surfactant forming an external ‘membrane’. **(B2)** Schematic view of a reverse micelle, comprising a hydrophilic core surrounded by the apolar chains of the surfactant. **(C1)** Liposomes of DMPC/DSPE-PEG2000 at 5% (LUV) observed by electron microscope after cryofracture [14]. **(C2)** Schematic view of a unilamellar liposome

	Nanoparticle	Micelle	Liposome
Composition	Polymers Lipids (SLN)	Surfactants Polymers	Phospholipid bilayer(s)
Structure	Reservoir/matrix	Hydrophobic or hydrophilic reservoir (reverse micelle)	Aqueous reservoir
Microscope image and/or diagram	 A1  A2  A3	 B1  B2	 C1  C2

past 40 years [11]. Owing to the fact that they have similar structure to cell membranes, liposomes, very widely used as carriers, are also useful tools as membrane models [12].

There are many other classes of carriers, in particular, non-self-assembled carriers like cyclodextrins and dendrimers. Cyclodextrins comprise a cyclic polysaccharide which allows the formation of an inner reservoir able to incorporate active molecules. The reader is referred to the literature for various reviews of carriers of this type [15–18]. Dendrimers are polymer carriers,

corresponding to a macromolecule with a tree-like structure formed by regular repetitions of an elementary motif. As a result of this structure, dendrimers have inner cavities, close to the core, that can be used to trap small molecules [19–21]. These systems are also nanometric carriers, but they are not nanoparticles and will not therefore be discussed further in this chapter.

24.2.2 Carrier Characteristics

Importance of Composition

Many techniques have been devised for formulating carriers. They are based on the exploitation of physical, chemical, and/or mechanical phenomena. The choice of preparation method will depend on the molecule to be encapsulated, the end application, and hence the characteristics of the resulting carriers. For each type of carrier, the literature abounds with publications detailing these formulation techniques. In the present discussion, we shall focus on those criteria that need to be taken into account when developing the carrier, criteria determined by the desired pharmaceutical applications.

As indicated above, the components of the carrier must not be toxic for the organism. This concerns the elements making up the particles, in particular the polymers, but also the association of these elements in the form of particles. This is why the main polymers used in this field are biodegradable. These include homopolymers of lactic acid (PLA), copolymers of lactic and glycolic acids (PLA/PGA), poly(butylcyanoacrylates), and poly(epsilon-caprolactone). However, non-biodegradable polymers such as poly(methacrylates) (Eudragit) are still used, mainly for research. Liposomes, with similar structure and composition to cell membranes, and solid lipid nanoparticles have the advantage of being composed of elements that are already present in the organism. In this respect, they are considered to be weakly toxic and have thus been extensively developed in the field of drug carriers.

Not only must the constitutive elements of nanoparticle carriers be non-toxic, but the intermediate reagents needed for their formulation must also meet this requirement. Now formulation techniques and the lipid or polymer composition of nanoparticles often involve the use of organic solvents that will solubilise them. Unfortunately, such solvents, e.g., dichloromethane, chloroform, etc., are toxic after administration, and it is difficult to eliminate them completely from the formulation. Techniques limiting the amounts of solvents, or avoiding them completely, are thus preferred. In order to solubilise the polymer and formulate nanoparticles, some work has been carried out with CO₂ in its supercritical state [22]. For the same reasons, the use of surfactants in the formulation (to stabilise the shape and obtain small particle sizes) must be limited as far as possible [23].

Many other molecules may enter into the composition of the carriers. In the case of liposomes, phosphatidylcholine or phosphatidylglycerol are the main

components. However, it has long been known that cholesterol can stabilise the liposome phospholipid bilayer [24]. Since that discovery, most liposomes have been formulated with cholesterol. Stabilisation of the membrane will depend on the proportion of cholesterol in the bilayer. As a consequence, structural modifications of liposome membranes [25] can lead to a modification in the rate of release of active principles by those liposomes [12].

Importance of Size

By definition, nanoparticles are ordered structures of nanometric dimensions. However, this size range remains rather broad on the scale of interactions between the particles themselves and between the particles and cells.

When nanoparticle carriers are injected intravenously, their distribution will depend greatly on their size. Two key mechanisms allow the carriers to leave the blood flow. The first is an extravasation phenomenon due to the presence of fenestrated epithelia within the organism. For example, liver and spleen epithelia contain pores with diameters of the order of 200 nm, allowing particles of smaller dimensions to pass through them. This is how small liposomes (< 90 nm) can pass through fenestrated epithelia, in particular those of the liver, where they will be widely distributed [26]. Pore sizes are much smaller in other organs and it is the epithelium which separates the central nervous system from blood circulation (blood–brain barrier) which is the most impermeable. It is said to be continuous, meaning that it will not allow such carriers to pass through. The other mechanism whereby nanoparticle carriers may be eliminated from the blood is capture by cells. Some cells such as macrophages even specialise in this task. The effect of the size will depend on the cell type. Prabha and coworkers [27] have shown that particles of size 100 nm can transfer genes *in vitro*, which is not the case for particles measuring 500 nm, nor for a mixture of these two populations. This demonstrates the importance, not only of the average size, but also of the size distribution.

These results are not only relevant to *in vitro* or intravenous studies. Recently, Vila et al. have shown that the size of polymer particles for nasal administration of proteins influences transport through the nasal mucus. Most of the transport is achieved by the smallest nanoparticles (200 nm) in this study [28].

Size thus has consequences on the particle distribution, but also on their interactions with different (cell) surfaces they encounter in the organism. These results show the need to control the sizes of particles obtained in the various formulation techniques. Small nanoparticles are usually obtained by supplying a large amount of energy by vigorous mechanical stirring, application of ultrasound, or increasing the pressure when the particles pass through porous membranes. It is also possible to obtain small unilamellar vesicles by applying ultrasound to multilamellar vesicles.

There are many techniques with differing degrees of sophistication for determining nanoparticle sizes and, in the field of drug delivery, the most widely

used techniques are quasi-elastic light scattering, gel exclusion chromatography, and electron microscopy [29].

Importance of Charge. Zeta Potential

The surface charge of a particle depends on its own nature but also on the surrounding medium. This surface charge affects the ion distribution in the interface region between the particle and the medium. In fact, different electrical layers form around each particle, these involving counterions present in the solution. The first two layers, including the Stern layer, are the most stable. It is only at the outer surface of the second layer that these particles actually interact with one another. This is called the slip plane. One is therefore interested mainly in the electrical potential measured at the slip plane, called the zeta potential, rather than the actual surface charge of the particles. It has the feature of varying with the surroundings and affects the physical, chemical, and biological properties of the particles [30].

For example, the zeta potential affects the stability of carriers suspended in an aqueous medium, e.g., in storage. Indeed, depending on this potential, the particles may interact together to form clusters which then precipitate out. The stability of the particles after intravenous administration also depends on their zeta potential, due to the interactions they may have with blood proteins, which are also electrically charged [31]. This natural phenomenon, known as opsonisation, quickly eliminates particles considered as foreign by the organism. This theme will be further discussed when we consider stealth particles (see p. 1107).

Furthermore, the interactions of these particles with cells, and hence with negatively charged membranes, also depend critically on their charge. This is why many formulations involve cationic components, such as cationic lipids favouring interactions with these membranes [32]. This gave rise to the idea of cationic nanoparticles or liposomes. The negative charge present at the cell surface is thus exploited to increase their contact time with these surfaces and hence favour intracellular penetration (or penetration of the transported active molecule).

Finally, adsorption of the active principle at the particle surface can also be achieved through electrostatic interactions. It thus depends on the surface charge of the particle. The charged molecule the most studied in this context is DNA, which is able to adsorb onto the surface of cationic nanoparticles or liposomes. This theme is discussed in detail in Chap. 3.

It can thus be assumed that the biodistribution of the particles after administration depends to a large extent on their charge, and this whatever form of administration is adopted [33].

The zeta potential is usually measured using electrophoresis. When an electric field is applied to charged particles in suspension in a liquid, they begin to move. The measured speed will depend on the strength of the electric field,

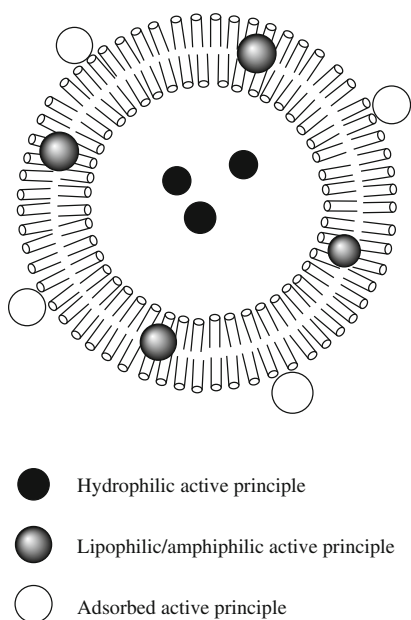


Fig. 24.2. Possible locations of an active principle in a liposome

which is fixed, the known dielectric constant, and the zeta potential, which can thus be determined via Henry's law.

The Carrier–Active Principle Association

Quite generally, and depending on the structure of the carrier, the active principle is either encapsulated in the particle reservoir (liposomes, polymer nanocapsules, micelles), or incorporated in the particle membrane (phospholipid bilayer of liposomes, surfactant monolayer of micelles, etc.), or present at the surface of the structure. Several of these different locations may be used at the same time. The association depends on the formulation technique and the structure of the carrier, and above all on the physicochemical properties of the transported molecule. The aqueous cavity of liposomes can encapsulate a high proportion of hydrophilic molecules, in contrast to the lipophilic core of micelles, for example. Amphiphilic molecules concentrate preferentially in the phospholipid bilayer of liposomes (see Fig. 24.2). This is why in some cases the main purpose of a carrier is to facilitate the solubilisation of a weakly hydrophilic active principle, which cannot therefore be administered by injection [34].

Apart from the spatial localisation of the active principle, the type of interaction/association with the carrier is another important consideration. Indeed, an active principle can be simply adsorbed onto the surface or even within the carrier. A negatively charged active principle, such as DNA, could for example bind by electrostatic interactions onto positively charged carriers.

However, adsorption at the surface can sometimes be neither desired nor controlled, and may concern varying proportions of the active principle that should have been totally encapsulated. This situation explains what is known as the burst effect, where there is a massive release of the active principle, simply adsorbed at the surface, over a very short lapse of time. This may lead to reduced protection of the active ingredient with respect to the metabolism. In parallel, in terms of the release kinetics, this situation may lead to toxicity or simply to the loss of a non-negligible proportion of the active principle when the preparation is stored. In fact, liposomes have themselves recently been microencapsulated within an alginate-based structure in order to limit the burst effect [35]. Another encapsulation method involves covalently linking the active ingredient to the carrier.

It is interesting to note that this nanoparticle–active principle relation underlies the development of solid lipid nanoparticle (SLN). Indeed, SLN preferentially encapsulate lipophilic molecules owing to their initial composition using triglycerides or waxes. However, one major drawback with these structures stems from lipid recrystallisation phenomena. This recrystallisation depends on the time and/or the temperature during formulation or storage. It may lead to the premature expulsion of the active ingredient [36]. For this reason, intermediate systems comprising a mixture of different solid lipids at room temperature, or a mixture of solid and liquid lipids at room temperature, have also been developed with a view to increasing encapsulation levels and avoiding the early release of active principles. These are known as nanostructured lipid carriers (NLC). Unfortunately, this first development does not solve the problem of the low encapsulation levels observed for hydrophilic molecules in SLN. Hydrophilic molecules which could nevertheless be captured within the lipid matrix will only be found in small quantities and must therefore be able to act in low concentrations within the organism [8, 37]. In order to encapsulate high concentrations of hydrophilic molecules, lipid drug conjugates (LDC) have been formulated by forming a salt between the active principle and a fatty acid, or by forming a covalent bond between the active principle and a fatty alcohol forming an ether or an ester. The resulting structure is then transformed into nanoparticles, using high pressure homogenisation and cooling, which leads to recrystallisation of the conjugate [38].

24.3 Development of Carriers for Pharmaceutical Applications

Nanoparticles used as carriers have been adapted to the specific needs of the pharmaceutical industry. Here we shall outline three main trends in this context:

- temperature- or pH-sensitive liposomes,
- stealth particles,
- targeting.

24.3.1 Thermosensitive and pH-Sensitive (Fusogenic) Liposomes

The molecular arrangement of phospholipids forming bilayers can be modified either naturally or artificially in vitro or in vivo by various factors such as the pH or the temperature [39]. It is on the basis of this observation that molecules able to produce pH- or thermosensitive liposomes were incorporated into liposomes. For example, these liposomes can release their contents at a pH below 7 or at a given temperature. The content of these liposomes can be released in a controlled manner into a well-defined environment.

Bearing in mind that acidic pH values are often encountered in tumour tissue or in subcellular compartments in which liposomes find themselves after their capture by cells (endosomes), pH-sensitive liposomes have been used in order to control the release of encapsulated molecules [40–42]. In most cases, dioleoylphosphatidyl ethanolamine (DOPE) was associated with cholesterol-hemisuccinate (CHEMS) in the liposome membrane to make it pH sensitive. Indeed, CHEMS stabilises the bilayers in the lamellar form, due to its anionic polar head group at neutral pH. An acid pH leads to its protonation and a reduction in its hydration. The lamellar structure of the bilayer then evolves toward a hexagonal structure (reverse micelles), totally destabilising the liposome and thereby destroying it. This phenomenon would favour fusion between the membranes of pH-sensitive liposomes and endosomes, and would thus lead to release of the active principle [29].

These pH-sensitive liposomes are said to be fusogenic, since they fuse with endosomes. There is another type of fusogenic liposome able to interact easily with cell membranes, employing a similar mechanism to that used by some viruses, in particular the Sendai virus, to internalise within cells. The Sendai virus invades the host cell at neutral pH by a fusion process involving two glycoproteins of the virus envelope, including the fusion protein F. This type of fusogenic liposome thus incorporates the protein F within its membrane [39, 43].

Heat-sensitive liposomes are made of lipids whose gel/fluid phase transition occurs between 41°C and 43°C in general, such as 1,2-dipalmitoyl-*sn*-glycero-3-phosphocholine. It is during this structural change that the molecules encapsulated within the liposomes are released [44]. The necessary hyperthermia may be caused by a pathological situation, e.g., fever, or hyperthermia in a solid tumour, or it can be triggered in a localised way in order to introduce the active ingredient.

24.3.2 Modifying the Carrier Surface

The administration of carriers loaded with some active principle only rarely leads to a significant improvement in the efficiency of the molecule. Indeed, for intravenous administration for example, the mononuclear phagocyte system, and in particular the macrophages located mainly within the liver and spleen, will capture the particles and eliminate them from the blood flow. This

capture phenomenon is related to opsonisation, i.e., to the rapid coating of particles in the blood by complement proteins. This protein coating favours capture of the particle by macrophages. It is a physiological phenomenon protecting the organism from foreign particles such as bacteria, leading to their ultimate elimination by the macrophages. If the aim is to deliver a molecule for intramacrophagic action, this mechanism allows passive targeting. This is why studies of anti-parasitic drug transport, such as primaquine in polymer nanoparticles, developed for the treatment of leishmaniasis (an intramacrophagic parasite), have given such promising results. Indeed, within a few minutes, more than 90% of the administered particles are eliminated from the blood flow by these macrophages. However, although these results are interesting in the case of intramacrophagic diseases, they are much less so if the molecule must act on other organs than the liver or spleen, or if this molecule is toxic in these organs.

The idea of blocking the macrophage cells either by molecules inhibiting phagocytosis (gadolinium chloride, dextran sulfate), or by an initial overdose of blank particles (not loaded with the active principle), has been envisaged. However, the effect is short-lived since natural compensation mechanisms increasing the number of cells are very quickly implemented by the organism.

Various groups have thus attempted to limit the targeting of the liver and spleen by modifying the carriers themselves. This has led to the development of so-called stealth particles, described below. In parallel, carriers specifically targeting the site to be treated have also been developed.

Stealth Particles

The main mechanism whereby macrophages capture particles is to a large extent understood and shows that this capture is correlated with the prior binding of complement proteins, such as fibronectin, immunoglobulin, etc., called opsonins, at the surface of the particles, in particular via hydrophobic interactions. Opsonins are the ligands of receptors located on the plasma membrane of the macrophages, facilitating the interaction and phagocytosis of the particles by these same macrophages.

The idea of stealth particles comes from observations of natural mechanisms, which show for example that the red blood cells themselves are never captured by the macrophages. In addition, some bacteria manage to deceive the macrophages, avoiding capture and rapid elimination by mononuclear phagocytes. These observations led researchers to examine the composition of the membranes of red blood cells and bacteria. They found that red blood cells have hydrophilic molecules (monosialogangliosides) in their membranes. Furthermore, *Pseudomonas aeruginosa* is able to manufacture a highly viscous, uncharged, and hydrophilic polyuronic polysaccharide molecule. It has been shown that these two molecules are involved in the mechanism used to escape from the macrophages, and thus underlie the stealth of red blood cells and some bacteria.

By the same principle, gangliosides have been incorporated in liposome membranes [45], and the surfaces of polymer nanoparticles have been coated with hydrophilic molecules, such as dextrans, or heparin [46]. For example, poly(methyl methacrylate) particles administered intravenously to mice have a plasma lifetime as short as 3 min. When coated with heparin or dextran, these same particles have a half-life as long as 5 h [46]. These molecules have high steric hindrance, which also restricts interactions with the opsonins. But the main development of stealth carriers has come from the use of poly(ethylene glycol) (PEG), which has high steric hindrance. Moreover, PEG chains are flexible and uncharged. These features lead to a significant reduction in the binding of opsonins to the particle surface, and a consequent extension of the circulation time of the particles. These polymers can be incorporated at the surface of polymer nanoparticles, by using a PLA-PEG copolymer, for example, or incorporated in liposomes (see Fig. 24.3) with the help of PEGylated phospholipids such as phosphatidylethanolamine-PEG or PEGylated cholesterol. They can also be simply adsorbed at the particle surface, using surfactant polymers such as Pluronic or Synperonic. The length of the PEG chains, their density at the particle surface, and the stability of the PEG-particle binding lead to variations in the efficiency of these chains in making the particles stealthy, whether they be polymer particles [48] or liposomes [48].

The use of stealth liposomes has shown that solid tumour tissues have a greater capacity to accumulate and retain nanoparticle structures than normal tissues. After administration, these stealth liposomes do not distribute themselves throughout the whole organism, but tend to concentrate in tumours. This is called enhanced permeability and retention, or the EPR effect [49]. It only concerns macromolecules and lipid particles [50]. Lipid carriers like liposomes or SLN can thus accumulate more easily within tumours. This phenomenon is related to a defective vascular architecture, combined with poor lymphatic drainage due to the need for rapid vascularisation to allow fast growth of the tumour.

Despite the clear increase in the circulation time of stealth carriers, and hence their broader distribution throughout the organism, and apart from tumour targeting due to the EPR effect, simple stealth carriers have no specific action. Surface modifications are thus required to target an organ or increase tumour targeting.

Targeting

Targeting provides a way of avoiding possible toxicity of a molecule for an organ in which it is distributed to no useful purpose, and it also means that a minimum of active principle need be administered for the task at hand, since it will be distributed, in the ideal situation, only within the desired treatment zone.

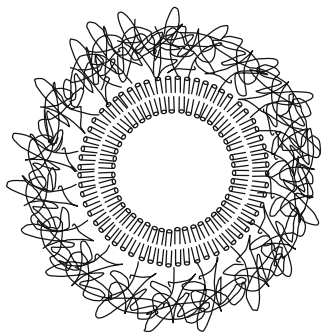


Fig. 24.3. Schematic view of a stealth liposome (obtained by post-insertion of PEGylated lipids in the phospholipid bilayer). The hydrophilic chains orient themselves toward the outer aqueous phase

There are many ways of achieving active targeting. Targeting systems by external intervention, such as encapsulation of magnetic particles (Fe_3O_4) concentrating in a region where a magnetic field is applied, have been studied [51]. But the most encouraging results at the present time concern the use of molecules bound at the surface of the carrier, which allow it to interact with a given receptor on the target cells (see Fig. 24.4).

This kind of carrier is formulated using many techniques for conjugating ligands with the synthesised carriers [52]. These techniques involve bioconjugation reactions between ligands and the previously formed carriers, or anchor the ligand by hydrophobic interaction in the liposome bilayer, for example.

One molecule used in this context is transferrin, a protein whose receptors are overexpressed on the cells of the blood–brain barrier or on the surface of a large number of tumour cells. Transferrin is easily coupled at the surface of the carriers and increases the fraction of them that reach the target. However, monoclonal antibodies or antibody fragments are also used as targeting agents. Indeed, these structures have a natural capacity to interact with specific targets at the surface of cells. This is why they were first exploited by binding the active principle directly on the antibody used as a carrier. Unfortunately, only small amounts of the active principle can be bound under these conditions, and it is difficult to achieve any therapeutic effect [52]. The idea is nevertheless interesting and binding these antibodies at the surface of carriers that can themselves encapsulate more active principle has given promising results. For example, the transferrin antireceptor antibody (OX26) has been successfully used on *in vitro* models [53].

It should be noted that molecules contributing to the stealth of the carrier are often associated with targeting molecules. If this is not the case, targeting is rarely effective since opsonisation phenomena with subsequent capture by macrophages in the liver and spleen then tend to dominate. This is why, by increasing the plasma half-life of liposomes, stealth has provided a way of promoting active targeting using carriers. In the literature, one thus finds stealth immunoliposomes and immunonanoparticles (see Fig. 24.5) [54–56]. Stealth (PEG) immunoliposomes directed against gliofibrillar acid protein (GFAP) have been prepared by coupling with anti-GFAP monoclonal antibodies. In

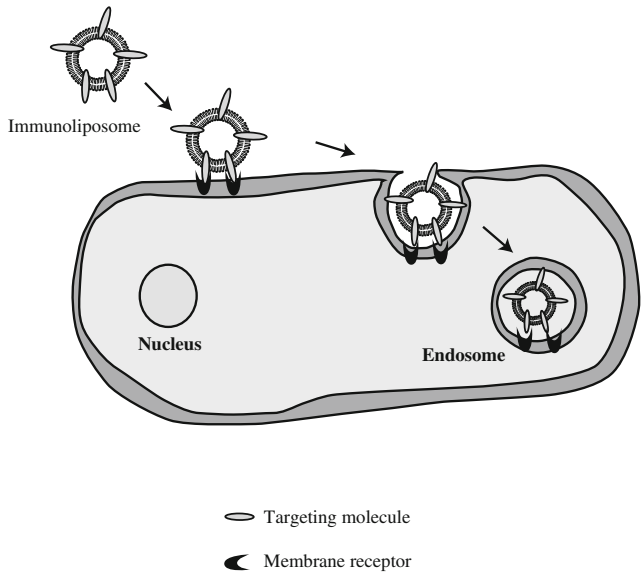


Fig. 24.4. Specific targeting of a cell by a carrier (liposome). The targeting molecules favour contact between the liposome and the cell membrane presenting specific receptors. This triggers endocytosis, internalising the carrier within the endocytotic vesicles (endosomes)

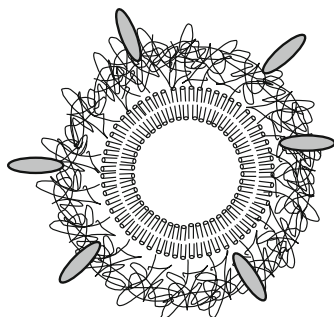


Fig. 24.5. Schematic view of a stealth immunoliposome

vitro studies have revealed a specific and competitive binding of these liposomes with rat brain astrocytes [56].

The EPR mechanism described above is another natural way of targeting carriers to tumours. This situation, associated with a targeting molecule on the carrier, greatly favours tumour targeting.

Current work is also trying to exploit the angiogenesis phenomenon observed in tumours. Indeed, angiogenesis, which allows tumours to be well irrigated by the blood, is a vital element for the development of tumour mass. Now it has been shown that this phenomenon involves regulatory molecules (interleukin 10, interferon γ , etc.) and stimulating molecules (vascular

endothelial growth factor or VEGF). As a consequence, carriers could deliver regulatory molecules or agents inhibiting the stimulating molecules in order to limit angiogenesis [49]. Moreover, nanoparticles specifically targeting endothelial cells involved in tumour angiogenesis have been envisaged. It is known that the integrin $\alpha_V\beta_3$ is overexpressed on these cells. It thus constitutes a potential target. To exploit this, RGD peptides with a particular affinity for this integrin have been coupled at the distal end of the poly(ethylene glycol) chains of stealth liposomes. In vitro results do indeed confirm a targeting of endothelial cells involved in angiogenesis, along with a reduction in the growth of the tumour cells, when these same stealth liposomes encapsulate and deliver doxorubicin [49].

However, no specific targeting carrier is yet available on the market, and this for two main reasons. The first relates to the fact that there are actually only a very few known cell receptors totally specific to a target. The field of drug delivery thus depends on research in biology. Furthermore, this type of technology is still very complex, difficult to implement, and costly. The applications described in Sect. 24.4 thus concern only ‘conventional’ or stealth carriers [8].

24.4 Applications of Carriers

All the ideas discussed in the first part of this chapter have been exploited for pharmaceutical applications. The most interesting of these will be outlined in the following sections. They have been classified in terms of the relevant medical field. This highlights the fact that many different areas of medicine are concerned here, although it should be borne in mind that the most common and most highly developed applications concern cancerology. Note that applications of these carriers to medical diagnosis or treatment belong to what is known as nanomedicine.

24.4.1 Medical Mycology and Parasitology

Amphotericin B, and in particular the speciality Fungizone, is a standard treatment for invasive fungal infections. However, owing to its limited efficiency and poor tolerance (especially in the kidneys), its liposome-associated form has been developed [57–59]. Ambisome (Fujisawa USA Inc. and NeXstar), launched on the US market in 1997, is a speciality in which amphotericin B is encapsulated in small unilamellar liposomes. These essentially limit the problems of intolerance, but may in some cases increase efficiency of the drug.

These liposomes are also used to treat leishmaniasis. Indeed, *Leishmania* is an intramacrophagic parasite which concentrates in the liver and spleen in the visceral form of the disease. As a consequence, nanoparticle carriers encapsulating an anti-parasitic drug seem particularly well suited when administered

intravenously. For this reason, liposomes loaded with amphotericin B (Ambisome) are useful for treating infant visceral leishmaniasis. The conclusions of a study carried out in the south of France on 32 children less than 15 years old show that liposomal amphotericin B was 97% effective and that this form benefited from a particularly good tolerance and ease of administration [60]. These same liposomes had already proved their worth in the treatment of kidney transplant patients suffering from visceral leishmaniasis [61]. Primaquine and atovaquone are other molecules with known anti-parasitic activity *in vitro*, but whose poor efficiency *in vivo* had led to their abandon. By encapsulating them in PLA nanoparticles, groups led by Gaspar, Rodrigues, and Cauchetier were able to increase their efficiency against *Leishmania*, thereby renewing interest in them [62–64]. At the present time, new leishmanicidal molecules such as harmine (an alkaloid isolated from a plant *Peganum harmala*) are being developed. Unfortunately, this molecule is toxic and cannot be used in free form. The use of carriers such as liposomes or nanoparticles encapsulating harmine has once again enhanced its efficiency (by a factor of one and a half to two, depending on the carrier) and reduced its toxicity [65].

In a complementary way, over the past few years several other groups have been working on vaccination against parasites. Protection was observed in BALB/c mice after administering liposomes encapsulating soluble *Leishmania* antigens and coated with oligomannose, for targeting dendritic cells and enhancing the immune response [66].

24.4.2 Ophthalmology

In ophthalmology, liposomes have an enormous therapeutic potential, as reflected in many experimental studies, but only one speciality, used in the case of macular degeneration, is already available commercially (Verteporfin, Visudyne, Novartis Ophthalmis). The relevance of liposomes in ophthalmology lies in their ability to increase the penetration of the active principle within the cornea (anterior segment). By using liposomes in the posterior segment, one can also obtain much longer clearance times, a reduction in the toxicity of the molecules, and a specific release of the molecules in the pigmented retinal epithelium [67]. Other investigations attest to their relevance in the context of gene transfer as applied to ocular tissue [68]. Sterically stabilised liposomes have also proven their usefulness for delivering non-degraded oligonucleotides in the case of retinal diseases [69]. This study led to the incorporation of the same liposomes in a thermosensitive gel in order to achieve a controlled release of oligonucleotides [70]. Therapeutic applications include the treatment of dry eyes, keratitis, uveitis, endophthalmitis, retinopathies, and so on.

In diagnostics, liposomes encapsulating fluorescein, which is only released under the action of a laser, are used to reveal the retinal microvasculature without fluorescence coming from the distribution of fluorescein in the choroid. This is not the only use of these liposomes, since under these conditions they allow one to repeat the study after a single injection. With this in mind,

the liposomes were rendered stealthy so that they could circulate for long enough [71].

24.4.3 Infectious Diseases

Some bacteria are able to get themselves internalised by cells and concentrate in the endosomes. A great many antibiotics administered in free form are unable to act on these intracellular bacteria located in the phagocytotic vesicles, e.g., in macrophages. Indeed, acidic antibiotics such as the β -lactams (penicillins) are negatively charged at physiological pH. As a consequence, they are unable to get through the membranes to reach the bacterial target. Basic antibiotics for their part, uncharged at physiological pH, can get through the membranes, but after penetrating the endosome, where the pH becomes very acidic, they become protonated and are no longer active with regard to the bacteria. For an amphiphilic molecule such as the quinolones, which can diffuse in either direction, i.e., from outside to inside the membranes or from inside to outside, the concentrations reached at equilibrium are not sufficiently high for them to act [72].

By encapsulating these same antibiotics inside carriers, such as liposomes or nanoparticles, they can pass through the membranes (bearing in mind that this passage can be optimised with the help of targeting molecules), while the antibiotic is protected from pH variations that could lead to its being inactivated. It is then delivered in an unmodified form in the vicinity of the bacteria where it will be able to act. The non-negligible toxicity of a certain number of antibiotics could also be avoided or at least limited by virtue of their encapsulation in nanoparticle carriers. A full review of this area of research was given by Couvreur and coworkers in 2000 [72].

At the present time, no formulation of antibiotic nanoparticle carriers has been commercialised, but various preclinical and clinical trials are underway, especially with liposome-encapsulated aminoglycosides used to treat serious infections caused by *Staphylococcus aureus* (MiKasome, NeXstar) [73, 74].

24.4.4 Cancerology

The success of liposome technology in cancer therapy can be explained by certain features discussed above:

- They can encapsulate hydrophilic or lipophilic molecules.
- They can limit the intrinsic toxicity of anti-cancer molecules.
- With the development of stealth liposomes, they can escape the effects of the mononuclear phagocyte system.
- Due to the increased permeability and retention effect, the carriers tend to be retained by tumours.
- Regions of angiogenesis can be targeted.

This is the area in which most commercial applications of liposomes are to be found [11]. Indeed, in 2005 there were 14 liposomal specialities on the market, and 10 of them concerned cancerology. They are used to treat Kaposi sarcomas, lung cancers, ovarian cancers, non-Hodgkin lymphomas, and melanomas. They can encapsulate molecules such as daunorubicin, doxorubicin, vincristine, and annamycin [11].

Doxorubicin, a cytotoxic anthracycline, was first encapsulated in conventional liposomes and commercialised under the name of Myocet for intravenous administration. These same liposomes, made stealthy by coating with poly(ethylene glycol) (DSPE-PEG) molecules, were subsequently developed and commercialised under the names of Doxil and Caelyx. They then had a plasma half-life of the order of 55 h in humans (for Doxil). Owing to their small size, around 100 nm, and extended plasma circulation time, these stealth liposomes can penetrate tumours. After this distribution, encapsulated doxorubicin becomes available in the tumour region by a release mechanism that has not yet been described.

Furthermore, many preclinical and clinical studies with liposomes are currently underway, and it can be observed that cancerology remains the principle field of application [74].

From a fundamental standpoint, the identification of an increasing number of antigens associated with tumours (tumour-associated antigens TAA) makes it possible to envisage new anti-tumour vaccine structures [75]. Some of these exploit peptide epitopes that are representative of these antigens, which could be carried by liposomes. These liposomes will be captured by the dendritic cells specialised in presentation of the epitopes to the cells of the immune system. The advantage of such vaccine formulations lies in their specificity and ease of fabrication, and the possibility of making entirely synthetic and multi-purpose constructs. In addition, many adjuvants devoid of any toxicity and able to enhance the immune response have recently been developed. Among these, it is worth mentioning the toll-like receptor (TLR) ligands which, when incorporated in the liposomes with the peptides, enhance their efficiency. For example, studies carried out *in vivo* on BALB/c mice have shown that the use of such constructs (liposomes containing the peptide epitope ErbB2 and incorporating a lipopeptide anchor adjuvant) provide full and specific protection against tumour development (overexpressing the ErbB2 receptor). Therapeutic vaccination studies are underway [76].

24.5 Conclusion

In order to use nanoparticle carriers in pharmaceuticals, it has been necessary to understand their properties and hence control their formulation. These carriers have evolved to meet the needs of the medical field. Applications are varied, but mainly concern cancerology. Moreover, among the different carriers, liposomes with their longer history of development are currently the

most commonly used. At the present time, many clinical trials are underway, in the hope of developing new drugs that should come on the market in the near future.

References

1. Dange, C., Vranckx, H., Balschmidt, P., Couvreur, P.: Poly(alkyl cyanoacrylate) nanospheres for oral administration of insulin, *J. Pharm. Sci.* **86**, 1403–1409 (1997)
2. Meshi, M.S., Sidhom, M.B., Fasipe, B.: Oral and subcutaneous absorption of insulin poly(isobutylcyanoacrylate) nanoparticles, *Int. J. Pharm.* **288**, 289–293 (2005)
3. Cournarie, F., Auchere, D., Chevenne, D., Lacour, B., Seiller, M., Vauthier, C.: Absorption and efficiency of insulin after oral administration of insulin-loaded nanocapsules in diabetic rats, *Int. J. Pharm.* **242**, 325–328 (2002)
4. Prego, C., Garcia, M., Torres, D., Alonso, M.J.: Transmucosal macromolecular drug delivery, *J. Control Release* **101**, 151–162 (2005)
5. Moghimi, S.M., Hunter, A.C., Murray, J.C.: Nanomedicine: Current status and future prospects, *Faseb. J.* **19**, 311–330 (2005)
6. Bala, I., Hariharan, S., Kumar, M.N.: PLGA nanoparticles in drug delivery: The state of the art, *Crit. Rev. Ther. Drug Carrier Syst.* **21**, 387–422 (2004)
7. Muller, R.H., Mader, K., Gohla, S.: Solid lipid nanoparticles (SLN) for controlled drug delivery: A review of the state of the art, *Eur. J. Pharm. Biopharm.* **50**, 161–177 (2000)
8. Wissing, S.A., Kayser, O., Muller, R.H.: Solid lipid nanoparticles for parenteral drug delivery, *Adv. Drug Deliv. Rev.* **56**, 1257–1272 (2004)
9. Jones, M., Leroux, J.: Polymeric micelles: A new generation of colloidal drug carriers, *Eur. J. Pharm. Biopharm.* **48**, 101–111 (1999)
10. Bangham, A.D., Standish, M.M., Miller, N.: Cation permeability of phospholipid model membranes: Effect of narcotics, *Nature* **208** (17), 1295–1297 (1965)
11. Torchilin, V.P.: Recent advances with liposomes as pharmaceutical carriers, *Nat. Rev. Drug. Discov.* **4**, 145–160 (2005)
12. Tashima, Y., Oe, R., Lee, S., Sugihara, G., Chambers, E.J., Takahashi, M., Yamada, T.: The effect of cholesterol and monosialoganglioside (GM1) on the release and aggregation of amyloid beta-peptide from liposomes prepared from brain membrane-like lipids, *J. Biol. Chem.* **279**, 17587–17595 (2004)
13. Shakweh, M., Besnard, M., Nicolas, V., Fattal, E.: Poly (lactide-co-glycolide) particles of different physicochemical properties and their uptake by Peyer's patches in mice, *Eur. J. Pharm. Biopharm.* **61**, 1–13 (2005)
14. Sriwongsitanont, S., Ueno, M.: Physicochemical properties of PEG-grafted liposomes, *Chem. Pharm. Bull. (Tokyo)* **50**, 1238–1244 (2002)
15. Singh, M., Sharma, R., Banerjee, U.C.: Biotechnological applications of cyclodextrins, *Biotechnol. Adv.* **20**, 341–359 (2002)
16. Eastburn, S.D., Tao, B.Y.: Applications of modified cyclodextrins, *Biotechnol. Adv.* **12**, 325–339 (1994)
17. Buschmann, H.J., Schollmeyer, E.: Applications of cyclodextrins in cosmetic products: A review, *J. Cosmet. Sci.* **53**, 185–191 (2002)

18. Dass, C.R.: Cyclodextrins and oligonucleotide delivery to solid tumours, *J. Drug. Target* **12**, 1–9 (2004)
19. Gillies, E.R., Frechet, J.M.: Dendrimers and dendritic polymers in drug delivery, *Drug Discov. Today* **10**, 35–43 (2005)
20. Boas, U., Heegaard, P.M.: Dendrimers in drug research, *Chem. Soc. Rev.* **33**, 43–63 (2004)
21. Cloninger, M.J.: Biological applications of dendrimers, *Curr. Opin. Chem. Biol.* **6**, 742–748 (2002)
22. Liu, H., Finn, N., Yates, M.Z.: Encapsulation and sustained release of a model drug, indomethacin, using CO(2)-based microencapsulation, *Langmuir* **21**, 379–385 (2005)
23. Xiong, J.Y., Liu, X.Y., Sawant, P.D., Chen, S.B., Chung, T.S., Pramoda, K.P.: Surfactant free fabrication of polymeric nanoparticles by combined liquid–liquid phase separation and solvent/nonsolvent mixing technology, *J. Chem. Phys.* **121**, 12626–12631 (2004)
24. Gregoriadis, G., Davis, C.: Stability of liposomes in vivo and in vitro is promoted by their cholesterol content and the presence of blood cells, *Biochem. Biophys. Res. Commun.* **89**, 1287–1293 (1979)
25. Koklic, T., Sentjurc, M., Zeisig, R.: The influence of cholesterol and charge on the membrane domains of alkylphospholipid liposomes as studied by EPR, *J. Liposome. Res.* **12**, 335–352 (2002)
26. Litzinger, D.C., Buiting, A.M., van Rooijen, N., Huang, L.: Effect of liposome size on the circulation time and intraorgan distribution of amphipathic poly(ethylene glycol)-containing liposomes, *Biochim. Biophys. Acta* **1190**, 99–107 (1994)
27. Prabha, S., Zhou, W.Z., Panyam, J., Labhasetwar, V.: Size-dependency of nanoparticle-mediated gene transfection: Studies with fractionated nanoparticles, *Int. J. Pharm.* **244**, 105–115 (2002)
28. Vila, A., Sanchez, A., Evora, C., Soriano, I., McCallion, O., Alonso, M.J.: PLA–PEG particles as nasal protein carriers: The influence of the particle size, *Int. J. Pharm.* **292**, 43–52 (2005)
29. Torchilin, V.P., Weissig, V.: *Liposomes. A Practical Approach*, Oxford University Press, Boston (2002)
30. Hunter, R.J., Midmore, B.R., Zhang, H.: Zeta potential of highly charged thin double-layer systems, *J. Colloid. Interface Sci.* **237**, 147–149 (2001)
31. Gessner, A., Lieske, A., Paulke, B., Muller, R.: Influence of surface charge density on protein adsorption on polymeric nanoparticles: Analysis by two-dimensional electrophoresis, *Eur. J. Pharm. Biopharm.* **54**, 165–170 (2002)
32. Fenart, L., Casanova, A., Dehouck, B., Duhem, C., Slupek, S., Cecchelli, R., Betbeder, D.: Evaluation of effect of charge and lipid coating on ability of 60-nm nanoparticles to cross an in vitro model of the blood–brain barrier, *J. Pharmacol. Exp. Ther.* **291**, 1017–1022 (1999)
33. Kohli, A.K., Alpar, H.O.: Potential use of nanoparticles for transcutaneous vaccine delivery: Effect of particle size and charge, *Int. J. Pharm.* **275**, 13–17 (2004)
34. Hu, J., Johnston, K.P., Williams, R.O.: Nanoparticle engineering processes for enhancing the dissolution rates of poorly water soluble drugs, *Drug Dev. Ind. Pharm.* **30**, 233–245 (2004)

35. Dhoot, N.O., Wheatley, M.A.: Microencapsulated liposomes in controlled drug delivery: Strategies to modulate drug release and eliminate the burst effect, *J. Pharm. Sci.* **92**, 679–689 (2003)
36. Mehnert, W., Mader, K.: Solid lipid nanoparticles: Production, characterization and applications, *Adv. Drug Deliv. Rev.* **47**, 165–196 (2001)
37. Almeida, A.J., Runge, S., Müller, R.H.: Peptide-loaded solid lipid nanoparticles (SLN): Influence of production parameters, *Int. J. Pharm.* **149**, 255–265 (1997)
38. Muller, R.H., Keck, C.M.: Challenges and solutions for the delivery of biotech drugs: A review of drug nanocrystal technology and lipid nanoparticles, *J. Biotechnol.* **113**, 151–170 (2004)
39. Delattre, J., Couvreur, P., Puisieux, F., Philippot, J.R., Schuber, F.: *Les liposomes. Aspects technologiques, biologiques et pharmacologiques*, Paris (1993)
40. Drummond, D.C., Zignani, M., Leroux, J.: Current status of pH-sensitive liposomes in drug delivery, *Prog. Lipid Res.* **39**, 409–460 (2000)
41. Yatvin, M.B., I., Tegmo-Larsson, I.M., Dennis, W.H.: Temperature- and pH-sensitive liposomes for drug targeting, *Methods Enzymol.* **149**, 77–87 (1987)
42. Bergstrand, N., Arfvidsson, M.C., Kim, J.M., Thompson, D.H., Edwards, K.: Interactions between pH-sensitive liposomes and model membranes, *Biophys. Chem.* **104**, 361–379 (2003)
43. Kunisawa, J., Nakagawa, S., Mayumi, T.: Pharmacotherapy by intracellular delivery of drugs using fusogenic liposomes: Application to vaccine development, *Adv. Drug Deliv. Rev.* **52**, 177–186 (2001)
44. Needham, D., Dewhirst, M.W.: The development and testing of a new temperature sensitive drug delivery system for the treatment of solid tumors, *Adv. Drug Deliv. Rev.* **53**, 285–305 (2001)
45. Allen, T.M., Chonn, A.: Large unilamellar liposomes with low uptake into the reticuloendothelial system, *FEBS Lett.* **223**, 42–46 (1987)
46. Passirani, C., Barratt, G., Devissaguet, J.P., Labarre, D.: Long-circulating nanoparticles bearing heparin or dextran covalently bound to poly(methyl methacrylate), *Pharm. Res.* **15**, 1046–1050 (1998)
47. Mosqueira, V.C., Legrand, P., Gref, R., Heurtault, B., Appel, M., Barratt, G.: Interactions between a macrophage cell line (J774A1) and surface-modified poly (D,L-lactide) nanocapsules bearing poly(ethylene glycol), *J. Drug Target* **7**, 65–78 (1999)
48. Vertut-Doi, A., Ishiwata, H., Miyajima, K.: Binding and uptake of liposomes containing a poly(ethylene glycol) derivative of cholesterol (stealth liposomes) by the macrophage cell line J774: Influence of PEG content and its molecular weight, *Biochim. Biophys. Acta* **1278**, 19–28 (1996)
49. Brannon-Peppas, L., Blanchette, J.O.: Nanoparticle and targeted systems for cancer therapy, *Adv. Drug Deliv. Rev.* **56**, 1649–1659 (2004)
50. Maeda, H., Sawa, T., Konno, T.: Mechanism of tumor-targeted delivery of macromolecular drugs, including the EPR effect in solid tumor and clinical overview of the prototype polymeric drug SMANCS, *J. Control Release* **74**, 47–61 (2001)
51. Zhang, J.Q., Zhang, Z.R., Yang, H., Tan, Q.Y., Qin, S.R., Qiu, X.L.: Lyophilized paclitaxel magnetoliposomes as a potential drug delivery system for breast carcinoma via parenteral administration: In vitro and in vivo studies, *Pharm. Res.* **22**, 573–583 (2005)

52. Nobs, L., Buchegger, F., Gurny, R., Allemann, E.: Current methods for attaching targeting ligands to liposomes and nanoparticles, *J. Pharm. Sci.* **93**, 1980–1992 (2004)
53. Huwyler, J., Cerletti, A., Fricker, G., Eberle, A.N., Drewe, J.: By-passing of P-glycoprotein using immunoliposomes, *J. Drug Target* **10**, 73–79 (2002)
54. Olivier, J.C., Huertas, R., Lee, H.J., Calon, F., Pardridge, W.M.: Synthesis of pegylated immunonanoparticles, *Pharm. Res.* **19**, 1137–1143 (2002)
55. Volkel, T., Holig, P., Merdan, T., Muller, R., Kontermann, R.E.: Targeting of immunoliposomes to endothelial cells using a single-chain Fv fragment directed against human endoglin (CD105), *Biochim. Biophys. Acta* **1663**, 158–166 (2004)
56. Chekhonin, V.P., Zhirkov, Y.A., Gurina, O.I., Ryabukhin, I.A., Lebedev, S.V., Kashparov, I.A., Dmitriyeva, T.B.: PEGylated immunoliposomes directed against brain astrocytes, *Drug Deliv.* **12**, 1–6 (2005)
57. Andres, E., Tiphine, M., Letscher-Bru, V., Herbrecht, R.: New lipid formulations of amphotericin B. Review of the literature, *Rev. Med. Intern.* **22**, 141–150 (2001)
58. Anaissie, E., Paetznick, V., Proffitt, R., Adler-Moore, J., Bodey, G.P.: Comparison of the in vitro antifungal activity of free and liposome-encapsulated amphotericin B, *Eur. J. Clin. Microbiol. Infect. Dis.* **10**, 665–668 (1991)
59. Wiebe, V.J., DeGregorio, M.W.: Liposome-encapsulated amphotericin B: A promising new treatment for disseminated fungal infections, *Rev. Infect. Dis.* **10**, 1097–1101 (1988)
60. Minodier, P., Robert, S., Noël, G., Blanc, P., Petornaz, K., Garnier, J.M.: Amphotéricine B liposomale en première intention dans la leishmaniose viscérale infantile en région Provence-Alpes-Côte-d’Azur-Corse, *Archives de pédiatrie* **12**, 1102–1108 (2005)
61. Boletis, J.N., Pefanis, A., Stathakis, C., Helioti, H., Kostakis, A., Giamarellou, H.: Visceral leishmaniasis in renal transplant recipients: Successful treatment with liposomal amphotericin B (AmBisome), *Clin. Infect. Dis.* **28**, 1308–1309 (1999)
62. Cauchetier, E., Paul, M., Rivollet, D., Fessi, H., Astier, A., Deniau, M.: Therapeutic evaluation of free and nanocapsule-encapsulated atovaquone in the treatment of murine visceral leishmaniasis, *Ann. Trop. Med. Parasitol.* **97**, 259–268 (2003)
63. Gaspar, R., Opperdoes, F.R., Preat, V., Roland, M.: Drug targeting with polyalkylcyanoacrylate nanoparticles: In vitro activity of primaquine-loaded nanoparticles against intracellular *Leishmania donovani*, *Ann. Trop. Med. Parasitol.* **86**, 41–49 (1992)
64. Rodrigues, J.M. Jr., Croft, S.L., Fessi, H., Bories, C., Devissaguet, J.P.: The activity and ultrastructural localization of primaquine-loaded poly (d,l-lactide) nanoparticles in *Leishmania donovani* infected mice, *Trop. Med. Parasitol.* **45**, 223–228 (1994)
65. Lala, S., Pramanick, S., Mukhopadhyay, S., Bandyopadhyay, S., Basu, M.K.: Harmine: Evaluation of its antileishmanial properties in various vesicular delivery systems, *J. Drug Target* **12**, 165–175 (2004)
66. Shimizu, Y., Yamakami, K., Gomi, T., Nakata, M., Asanuma, H., Tadakuma, T., Kojima, N.: Protection against *Leishmania major* infection by oligomannose-coated liposomes, *Bioorg. Med. Chem.* **11**, 1191–1195 (2003)

67. Ebrahim, S., Peyman, G.A., Lee, P.J.: Applications of liposomes in ophthalmology, *Surv. Ophthalmol.* **50**, 167–182 (2005)
68. Dannowski, H., Bednarz, J., Reszka, R., Engelmann, K., Pleyer, U.: Lipid-mediated gene transfer of acidic fibroblast growth factor into human corneal endothelial cells, *Exp. Eye Res.* **80**, 93–101 (2005)
69. Bochot, A., Fattal, E., Boutet, V., Deverre, J.R., Jeanny, J.C., Chacun, H., Couvreur, P.: Intravitreal delivery of oligonucleotides by sterically stabilized liposomes, *Invest. Ophthalmol. Vis. Sci.* **43**, 253–259 (2002)
70. Bochot, A., Fattal, E., Gulik, A., Couarraze, G., Couvreur, P.: Liposomes dispersed within a thermosensitive gel: A new dosage form for ocular delivery of oligonucleotides, *Pharm. Res.* **15**, 1364–1369 (1998)
71. Khoobehi, B., Peyman, G.A., Niesman, M.R., Oncel, M.: Measurement of retinal blood velocity and flow rate in primates using a liposome–dye system, *Ophthalmology* **96**, 905–912 (1989)
72. Pinto-Alphandary, H., Andreumont, A., Couvreur, P.: Targeted delivery of antibiotics using liposomes and nanoparticles: Research and applications, *Int. J. Antimicrob. Agents* **13**, 155–168 (2000)
73. Schiffelers, R., Storm, G., Bakker-Woudenberg, I.: Liposome encapsulated aminoglycosides in pre-clinical and clinical studies, *J. Antimicrob. Chemother.* **48**, 333–344 (2001)
74. Felnerova, D., Viret, J.F., Gluck, R., Moser, C.: Liposomes and virosomes as delivery systems for antigens, nucleic acids and drugs, *Curr. Opin. Biotechnol.* **15**, 518–529 (2004)
75. Rosenberg, S.A.: Cancer vaccines based on the identification of genes encoding cancer regression antigens, *Immunol. Today* **18**, 175–182 (1997)
76. Roth, A., Rohrbach, F., Weth, R., Frisch, B., Schuber, F., Wels, W.S.: Induction of effective and antigen-specific antitumour immunity by a liposomal ErbB2/HER2 peptide-based vaccination construct, *Br. J. Cancer.* **92**, 1421–1429 (2005)

Activatable Nanoparticles for Cancer Treatment. Nanobiotix

V. Simon, A. Ceccaldi, and L. Lévy

25.1 Introduction

With almost 150,000 deaths every year in France (26 times more than on the roads), cancers represent the second cause of mortality after cardiovascular disease. In 2002, 10 million new cases of cancer were registered in the world and there were 6 million deaths (of which 40% in developed countries). Moreover, the World Health Organisation (WHO) predicts a significant increase in the number of new cases between now and 2020 (+40% in developed countries and +100% in developing countries), due mainly to longer life expectations, change in behaviour, and degradation of the environment.

The word ‘cancer’ is a generic term covering a group of more than a hundred diseases, all characterised by the organism losing control over the proliferation of certain cells. These cells then develop in an anarchic way, eventually constituting a tumour which invades surrounding tissue and in many cases ends up disseminating to distant tissues (metastases).

The genetic predisposition of different populations, the natural sensitivity of certain organs, exposure to carcinogenic substances (chemical products, pollution), and certain types of behaviour (smoking, alcohol consumption, a diet rich in saturated animal fats and poor in fruit and vegetables) are the main factors explaining the large variations in the frequency of cancers depending on the affected tissue or organ. Every year, the cancers with the greatest number of new cases are (in decreasing order) lung cancer, stomach cancer, breast cancer, bowel cancer, mouth cancer, and liver cancer. The efficiency of therapy is extremely variable depending on the affected organ: about 95% for testicle cancers, but a five-year survival rate close to zero for cancer

Table 25.1. Cancer statistics. Five-year mortality rates (Infocancer, 2006)

Ovary	Pancreas	Bowel	Bladder	Kidney	Breast	Stomach	Liver	Prostate	Neck of uterus	Oeso-phagus	Thyroid
68%	96%	54%	36%	52%	27%	79%	95%	44%	38%	92%	25%

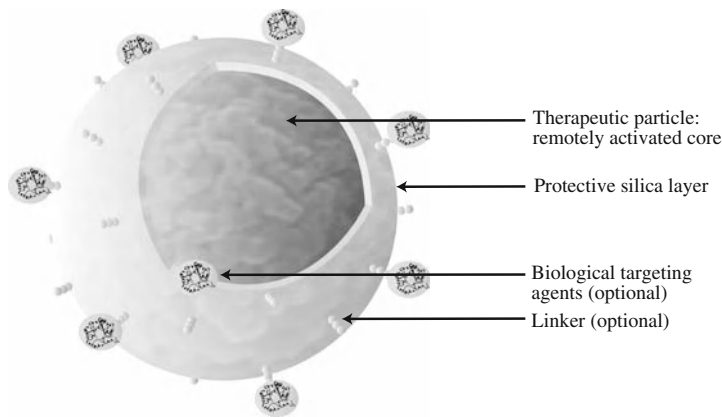


Fig. 25.1. A nanoTherapeutic: size less than 70 nm

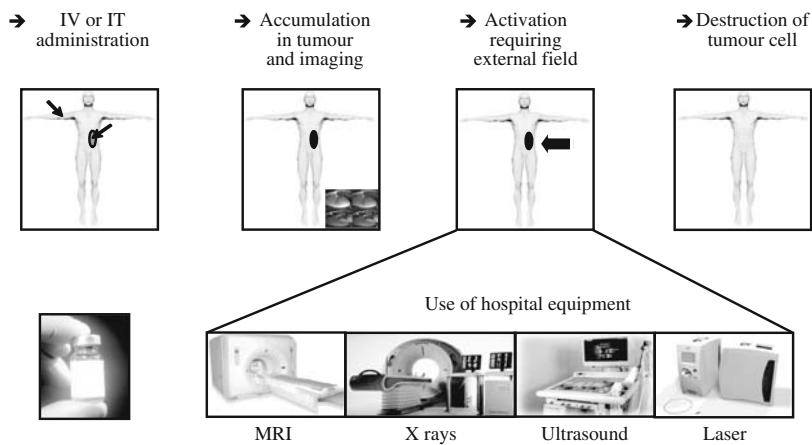


Fig. 25.2. Therapeutic mechanisms of nanoTherapeutics

of the pancreas or the glioblastoma. Table 25.1 shows the 5-year mortality rate for the most common forms of cancer.

In France, advances in therapeutic research over the past twenty years mean that globally 50% of cancers are now cured. However, the average rate for complete remission from a cancer is still much too low.

Many research groups are engaged in the fight against cancer. Some are investigating technologies that have been neglected up to now, in the hope of obtaining radically innovative treatments. For example, the company Nanobiotix is developing new nanotechnological applications for cancer treatment, exploiting related developments in biology, physics, and the chemistry of nano-materials. Nanobiotix technology is built up on two themes: understanding the biological mechanisms and developing the technical ability to make complex structures on the nanoscale.

25.2 NanoTherapeutics

Nanobiotix technology is based on the novel idea of nanoTherapeutics (see Fig. 25.1). These are nanoparticles (controlled diameter less than 70 nm) comprising a therapeutic core that can be remotely activated by an external energy supply. The particles are coated by a layer of silica guaranteeing biocompatibility, in particular, avoiding any immunogenicity. Finally, biological agents can be grafted onto the surface of these particles to obtain highly specific targeting of tumour cells. These targeting agents are not always essential for obtaining a suitable biodistribution of the nanoparticles in the case of passive targeting.

Nanobiotix is developing these nanoTherapeutics with the aim of specifically destroying tumour cells, while preserving healthy tissue, using the general mechanism illustrated schematically in Fig. 25.2. The nanoparticles are injected into the patient intravenously or intratumorally. They are designed to target tumour tissues and take 20–48 h to accumulate selectively in them. Once the particles have been internalised by the cancer cells, an external energy field is applied to activate the nanoTherapeutics: a local physical or chemical effect then destroys the pathological cell. Four types of energy field can be used, each corresponding to one of the technologies developed in parallel by Nanobiotix: a magnetic field (MRI) for the nanoMag technology, a laser for the nanoPDT technology, X-rays for nanoXRay, and ultrasound for nano(U)Sonic. After 72 h, the particles are eliminated via the kidneys.

There are many potential advantages with the nanoTherapeutic system:

- On–off treatment, since the therapy lasts only while the external field is being applied.
- Specific targeting of the tissue to be treated by the nanoparticles or by focusing of the field, thereby limiting side effects.
- Innocuousness of the nanoparticles and non-chronic mode of administration.
- The nanometric dimensions of the particles guarantees efficient diffusion through the tissues, and good elimination.
- Compatibility of the technology, which can be used alone or in synergy with existing treatments such as chemotherapy, surgery, immunotherapy, and radiotherapy.
- Efficiency can be adjusted by modifying the composition of the nanoparticles and their activation parameters.
- Decoupling of the therapeutic effect and biological targeting, which means that the same activatable cores can be used by adapting the targeting to the relevant cancer.
- Particles can be activated by already existing hospital equipment, e.g., MRI, laser, radiotherapy equipment, etc.
- Therapeutic action based on physical rather than biological effects.

Nanotherapeutics	Source	Mechanism	Target
 NanoMag	 MRI	Physical rotation	Tumour Metastases Diagnostic
 NanoPDT	 Laser	Free radicals	Cancer Diagnostic
 NanoXRay	 X rays	Free radicals Heat	Tumour Diagnostic
 Nano(U)Sonic	 Ultrasound	Controlled delivery of active compounds	Tumour Diagnostic

Fig. 25.3. Four families of nanoTherapeutics

The last point is crucial: it means that this highly innovative technology represents a genuine break with therapies based on conventional chemistry or pharmacology and first generation biotechnologies. The common aim of these technologies is to develop an active principle (a small chemical molecule, nucleic acid, enzyme, hormone, monoclonal antibody, etc.) to interact with a targeted biochemical mechanism at the cell level, which the tumour cell is very often able to compensate for or counter. For its part, Nanobiotix technology is based on physical effects which thus drastically reduce the risk of resistance to the treatment and therapeutic escape. It is not just a new mode of action, but a truly new paradigm.

The development of genuinely new solutions for fighting against cancer require a complete rethinking of the very idea of a drug. One of the greatest advances of industrial design came with the understanding that a better final product could be achieved by separating the functions and the attributes of these entities. However, the pharmaceutical industry has not yet taken this road. Today, the approach known as rational drug design has succeeded in making very small therapeutic molecules based on the study of the 3D structure of the human enzyme protein (and in particular its active site) with a view to inhibiting or activating it to treat a given disease. The chemical structure of these drugs is designed to ensure that they will bind onto the human protein (targeting) and also that they will have the desired therapeutic effect. The latter is usually achieved by blocking the enzyme cycle. Therapeutic

monoclonal antibodies produced by first generation biotechnology companies provide the most striking example of this coupling between the 3D structure of the drug, interaction with the ligand (ensuring specific action), and therapeutic effect. These approaches have had many successes, but nevertheless display some weak points. For example, rational drug design requires a precise knowledge of the 3D structure of the targeted human protein, and this is not always possible. Furthermore, it is usually difficult to adapt existing successful molecules to new applications.

Nanobiotix technology has been conceived in a very different way. Indeed, nanoTherapeutics display a high level of decoupling between the therapeutic core, whose task it is to fight the cancer, and the targeting elements which confer specific action on the structure as a whole. This strategy means that nanoTherapeutics will be much more flexible with regard to the various anti-cancer applications. In addition, this approach is much more in tune with the future of human medicine. The use of mass-produced drug treatments will soon make way for more and more solutions tailored to the needs of the individual patient (tailored therapeutics), exploiting the detection and understanding of the molecular process underlying the specific illness of the given individual.

At the present time, many cancers are still difficult to treat, because the potential effective dose in chemotherapy would induce too many side effects. This is particularly true for breast cancers. There are two lines of research that may lead to a solution here: reduce the toxicity of active principles or increase their efficiency. Nanobiotix intends to pursue both at the same time, thereby obtaining benefit-to-risk ratios never before reached. Indeed, the aim here is to break the traditional correlation between therapeutic efficiency and associated toxicity for the organism.

The combined use of nanoTherapeutics and external energy fields also provides *in situ* information about the cells. The development of diagnostic imaging applications can thus proceed in parallel with therapeutic applications, opening the way to a new era of theranostics, i.e., diagnostic therapy for individual patients.

25.3 Different Families of Nanoparticles

In order to obtain a permit to market a drug resulting from the application of a new technology, one must not only demonstrate that it does specifically target cancer tissues, and in a non-toxic way, but also that it is at least as effective as existing drugs. In the case of a nanotechnological product, it is in addition essential to prove the innocuousness of the non-activated particles and demonstrate an intracellular activity.

Nanobiotix is currently developing four potential therapies against cancer (see Fig. 25.3):

- *NanoMag*. The particles are activated by a magnetic field like the one used for magnetic resonance imaging (MRI). Activation is manifested by physical rotation of the particle. These nanoparticles can have applications for deep tumours and metastases. They can also be used for imaging and diagnosis.
- *NanoPDT*. The particles are activated by laser or visible light. The principle is the same as for photodynamic therapy (PDT). Tumour cells are destroyed by the production of free radicals inside the cells. This family of nanoparticles can also be used in diagnostics.
- *NanoXRay*. The particles are activated by X-rays. They also produce free radicals inside the tumour. There are possible applications in imaging and therapy.
- *Nano(U)Sonic*. The particles are activated by ultrasound. They can be used for therapy and diagnosis.

25.4 NanoTherapeutic Action Mechanisms

25.4.1 NanoMag

The treatment involves getting the nanoparticles inside the target cell, then applying a strong enough magnetic field to orient them within the cell. Indeed, nanoMag nanoparticles have an iron oxide core carrying a magnetic moment. During activation, the magnetic moments, which were initial randomly oriented within the cell, line up with the external magnetic field (see Fig. 25.4), transforming the magnetic energy into rotational kinetic energy. The forced orientation of these particles throughout the period of exposure induces directional forces which strain the cell. When the nanoparticle concentration is high enough within the cell, the tumour cell is destroyed. Depending on the level of stress in the cell and/or the resulting damage, the tumour cells enter into apoptosis or necrosis. When the field is switched off, the nanoparticles adopt once again a random orientation and their anti-tumour activity ceases instantaneously.

25.4.2 NanoPDT

Photodynamic therapy (PDT) exploits the photoinduced oxidation of matter. This type of treatment requires a photosensitive molecule which accumulates preferentially within tumours, a light source able to excite this molecule, and an oxygen-rich environment. Singlet oxygen, the main cytotoxic agent produced during PDT, is a highly reactive form of oxygen produced from O_2 molecules in the cell. PDT methods already exist, but only nanoPDT allows specific delivery of the photosensitive molecule to the cancer cells, increasing its therapeutic relevance.

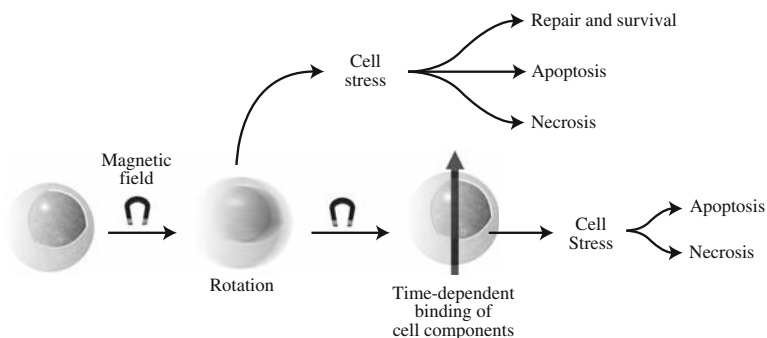


Fig. 25.4. Action of nanoMag on tumour cells

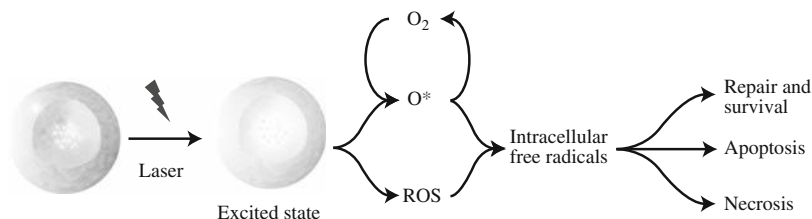


Fig. 25.5. Action of nanoPDT on tumour cells

When exposed to strong specific laser or UV radiation, the photosensitive molecules induce the formation of free radicals. Indeed, the activated chromophores have a photocatalytic effect. This transforms neighbouring molecular oxygen and reactive oxygen species (ROS) by a chain reaction into free radicals, which are highly reactive species causing irreversible damage to the tumour cell (see Fig. 25.5). The first cell organs to be affected are the mitochondria, cell and nuclear membranes (by oxidation of phospholipids), and lysosomes. Singlet oxygen reacts quickly and locally. The oxidation effects induced by PDT are thus confined to a very small region. In addition, the short wavelengths of this low-energy radiation cannot go through any great thickness of tissue, and this means that the treatment is perfectly harmless for the surrounding tissues. For this reason, NanoPDT is designed mainly to treat surface cancers such as cancers of the skin, the oesophagus, the bladder, or the stomach.

25.4.3 NanoXRay

This new approach uses an inorganic matrix sensitive to X-rays. Once excited, it can create free radicals in the cell where it is located. The nanoXRay particles are internalised by the tumour cells, then activated by X-rays. The creation of free radicals during excitation of the particles is accompanied by the local release of heat. This double effect pushes the tumour cells to apoptosis

or necrosis (see Fig. 25.6). The cells are thereby specifically destroyed by heat and by the localised and controlled production of free radicals coming from the interaction of the matrix with water available in the organism, transformed into $\text{HO}^\bullet$.

25.4.4 Nano(U)Sonic

Unlike the other technologies, nano(U)Sonic particles carry cytotoxic agents at the surface. They are activated by focused ultrasound (see Fig. 25.7). The contrast between sound wave reflection by soft tissue (weak reflection) and hard tissue (strong reflection) is exploited by ultrasound imaging techniques (echography). These diagnostic methods often use contrast enhancement agents. These are very diffuse structures such as gas nanosomes, or very dense structures like the nano(U)Sonic particles, which are excellent contrast agents. Apart from the imaging applications of nano(U)Sonic, higher energy sound waves can be used to break the chemical bonds of the particle, so that it releases the cytotoxic agents into the tumour cell. This release can be controlled on the nanoscale in time and space.

Destruction occurs via this release of cytotoxic agents. Their intracellular action condemns the tumour cell to a programmed cell death (apoptosis) or non-programmed destruction (necrosis).

25.5 Synthesising NanoMag Particles

Multifunctional nanoparticles ($< 50 \text{ nm}$) have been fabricated by combining recent progress in the field of ferrofluids, nanotechnology, and targeting. The synthesis of nanoMag particles takes place in four successive stages (see Fig. 25.8):

1. Synthesis of the crystalline magnetic core of Fe_2O_3 , with diameter 20–30 nm, for targeted magnetocytolysis. This is the part responsible for the therapeutic effect, which is in fact physicochemical.
2. Coating with a layer of amorphous silica SiO_2 , ensuring biocompatibility and stability of the nanoparticle.
3. Possible addition of a polycarbonate chain at the particle surface in order to be able to attach a biological targeting agent.
4. Addition of a specific ligand of certain tumour cells, e.g., luteinizing hormone releasing hormone (LHRH), used to treat breast cancer.

Iron dodecyl sulfate $\text{Fe}(\text{DS})_2$ is synthesised by adding $\text{FeCl}_2 \cdot 4\text{H}_2\text{O}$ to water. Sodium dodecyl sulfate (2.3 g) is added to the solution, which is then stirred for several hours. The precipitation of $\text{Fe}(\text{DS})_2$ is obtained by lowering the temperature. The solution is then stored for 24 h. The precipitate is recovered by filtering and washing with cold water. The supernatant is dried in vacuum for 2–3 days. The Fe_2O_3 magnetic core of diameter 20 nm is synthesised using

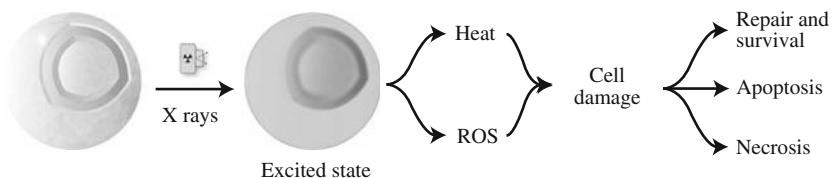


Fig. 25.6. Action of nanoXRay on tumour cells

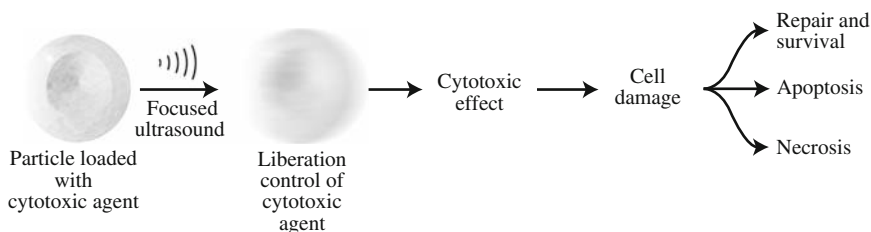


Fig. 25.7. Action of nano(U)Sonic on tumour cells

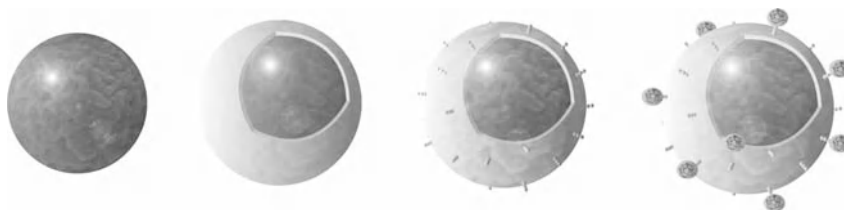


Fig. 25.8. Synthesis of Fe_2O_3

the colloid reaction described by Feltn et al. [1]. $\text{Fe}(\text{DS})$ is then added to 10 ml of water and stirred for several hours, before cooling the solution. The two-photon chromophore ASPI-SH is added to the ferrofluid solution and the final solution incubated at room temperature with stirring for 4 h.

25.5.1 Coating the Fe_2O_3 Particles with SiO_2

The Fe_2O_3 nucleus is coated with a layer of silica. To consolidate the silica and prevent aggregation due to magnetic interaction, sodium silicate is added to the suspension. The latter is stirred for 24 h at room temperature, while sonicating with ultrasound for 5 min every 4 h. The particles are then washed with ethanol/water (1/2) and recovered by centrifuging (10,000g). The washed particles are redispersed in an ethanol/water solution (4/1), and 200 μL of 0.1 M NH_4OH is added. Tetraethyl orthosilicate (TEOS) is added to the particle suspension with stirring. After 14 h, the particles are recovered and washed [2].

25.5.2 Adding the Spacer

The external modification of the silica layer is achieved by adding a binding agent or linker by preparing 3-(triethoxysilanylpropyl-carbamoyl) butyric acid. The latter is synthesised by mixing 3-amine (triethoxysilyl) propyl with glutaric anhydride with 1 mL of *N,N*-dimethylformamide (DMF). Diisopropylamine is added to the solution and the result stirred for 24 h in an argon atmosphere. The resulting product is dispersed in DMF containing 40 mg of particles. After stirring for another 24 h at room temperature, the particles are washed with DMF and dried.

25.5.3 Adding the Ligand

When a ligand is added, e.g., the hormone LHRH (D-Lys), one follows the procedure describe by Wang et al. [3]. This involves adding 2 mg of LHRH and 24 mg of benzotriazol-1-ylol-tris(dimethylamino) phosphonium hexafluorophosphate (BOP) to the nanoparticle suspension. DIPEA is then added dropwise to the solution. The reaction mixture is sonicated at room temperature for 5 h. The product is recovered by centrifuging (10,000*g* for 30 min) and washed with water and DMF. For improved stability, the particles bearing LHRH are dispersed in a large volume of water containing 10% DMSO and concentrated by centrifuging before use [4].

25.6 Advantages of the NanoTherapeutic Families

25.6.1 NanoMag

Using drug delivery with biological ligands, once the nanoparticles have been internalised, tumour cells undergo magnetocytolysis when subjected to a magnetic field. Since this field is not limited with regard to penetration, the nanoMag particles (see Fig. 25.9) are particularly well suited to treat deep tumours such as cancers of the liver, stomach, prostate, and so on. Particles are administered either intratumorally or intravenously in the vicinity of the tumour.

There are two options for targeting tumour cells:

- Biological ligands can be grafted onto the particles in order to target tumour cells through overexpressed receptors.
- Bare particles, without targeting agent on the surface, also accumulate passively in tumour tissues by virtue of the enhanced permeability and retention (EPR) effect. The latter refers to the abnormally high permeability of vessels newly formed by the tumour, combined with ineffective lymphatic drainage.

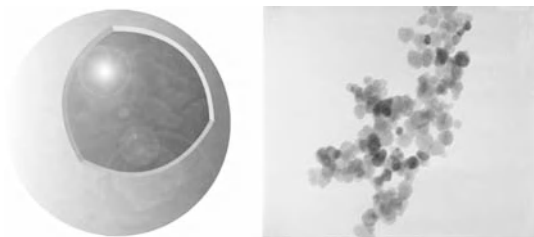


Fig. 25.9. Schematic view and transmission electron microscope (TEM) image of a nanoMag particle

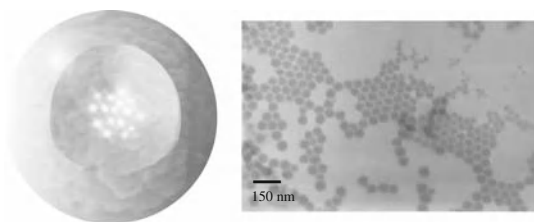


Fig. 25.10. Schematic view of the structure of a nanoPDT particle together with a TEM image

The advantage in using a magnetic field generated by an MRI device is that one can simultaneously treat and visualise the results, e.g., accumulation in the tumour and necrosis. Furthermore, the required strength of the magnetic field (about 1.5 tesla) and the composition of the particles themselves make the treatment perfectly harmless. Nanobiotix developed the nanoMag system in such a way that its use does not require investment in the purchase of costly equipment, designed exclusively for use with nanoMag. The activation source is perfectly well provided for by a standard MRI device. In short, it is a non-chronic treatment requiring few trips to the hospital, which should reduce the overall cost of therapy as compared with traditional chemotherapy.

25.6.2 NanoPDT

The diameter of nanoPDT particles is around 30 nm. They comprise four components (see Fig. 25.10):

1. an organic chromophore,
2. an inert silica layer ensuring biocompatibility, stability, and coating of the nanoparticles,
3. a polycarbonate chain that can be added to the surface in order to graft on biological ligands,
4. a biological agent specific to the targeted organ where necessary.

Photodynamic therapy is a fast-developing field, despite significant side effects reported for the few molecules currently commercialised (Photofrin). Indeed,

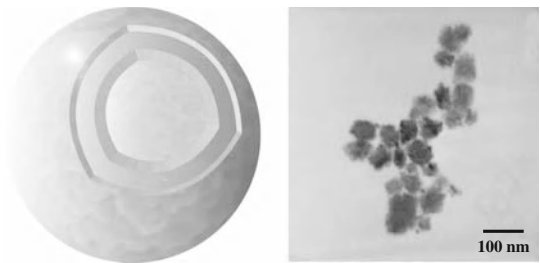


Fig. 25.11. Schematic view and TEM image of a nanoXRay particle

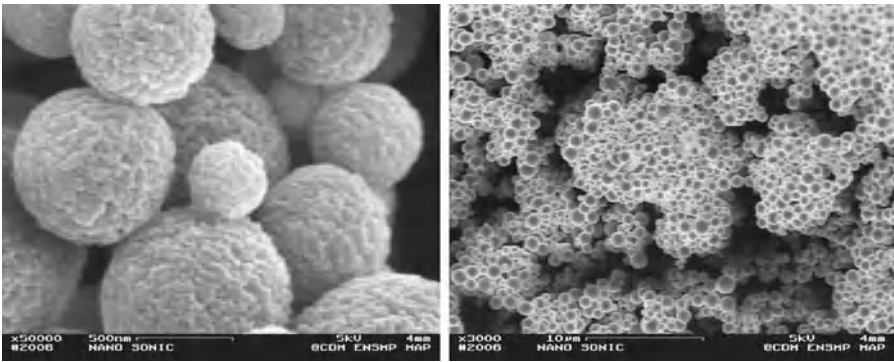


Fig. 25.12. Scanning electron microscope (SEM) image of Trisacryl spheres



Fig. 25.13. Schematic view of a nano(U)Sonic particle

problems have been identified with the stability of the active principle, the potential development of cutaneous photosensitivity in the patient, and toxicity of the molecules. The new method for encapsulating the photosensitising molecules developed by Nanobiotix with silica brings effective solutions to these problems.

NanoPDT particles significantly improve on first generation photodynamic therapy, because they are highly stable and less toxic. Side effects are limited by the fact that the photosensitive molecule is protected by the nanoparticle matrix, preventing this molecule from diffusing through the whole body. The

nanoPDT particles are activated by the standard laser equipment of photodynamic therapy.

25.6.3 NanoXRay

In contrast to lasers, X-rays have wavelengths that allow them to pass through a certain tissue thickness (hard X-rays) and hence reach deeply embedded nanoparticles. Particles can thus be activated in any tissue, even one that is deep or inaccessible to conventional surgery. The targeting mode is the same as for the other two types of nanoparticle.

The technology developed here uses the water present in the human body at physiological concentrations to generate free radicals, in contrast to conventional photodynamic therapy, which may require oxygenation of the treated tissues. The X-rays required for activation can be generated by standard X-ray equipment in any hospital. The nanoXRay particle is represented in Fig. 25.11.

25.6.4 Nano(U)Sonic

The nano(U)Sonic particles are bigger than the other three types of nanoparticle. These are Trisacryl nanospheres, with diameters in the range 1–2 μm . They are very precisely calibrated, spherical, hydrophilic, porous, and made from an acrylic copolymer (Trisacryl), crosslinked with acrylamide. The photographs in Fig. 25.12 were taken by scanning electron microscope (SEM) and show the Trisacryl spheres.

Nano(U)Sonic particles (see Fig. 25.13) can be used as ultrasound contrast enhancers. Targeting agents can be added for therapeutic applications. By activating the particles with ultrasound, cytotoxic agents can be released specifically in tumour cells (after breaking the particle).

25.7 Results (NanoMag)

25.7.1 In Vitro Experiments

Experiments have shown that nanoTherapeutics, when suitably targeted, accumulate selectively in several cell types such as HeLa (carcinoma of the neck of the womb), UCI-107 (carcinoma of the ovaries), KB (carcinoma of the buccal epidermis), and MCF-7 (mammary tumoral epithelial cell line).

NanoMag particles with LHRH have been incubated with breast carcinoma cells MCF-7. The particles contain a fluorophore called ASPISH which can be used to monitor the particles by fluorescence microscopy. With this method, it is possible to visualise the nanoTherapeutics entering tumour cells. Figure 25.14 shows that the accumulation of nanoTherapeutics is time dependent. After 30 min or so, the MCF-7 cells are saturated with particles.

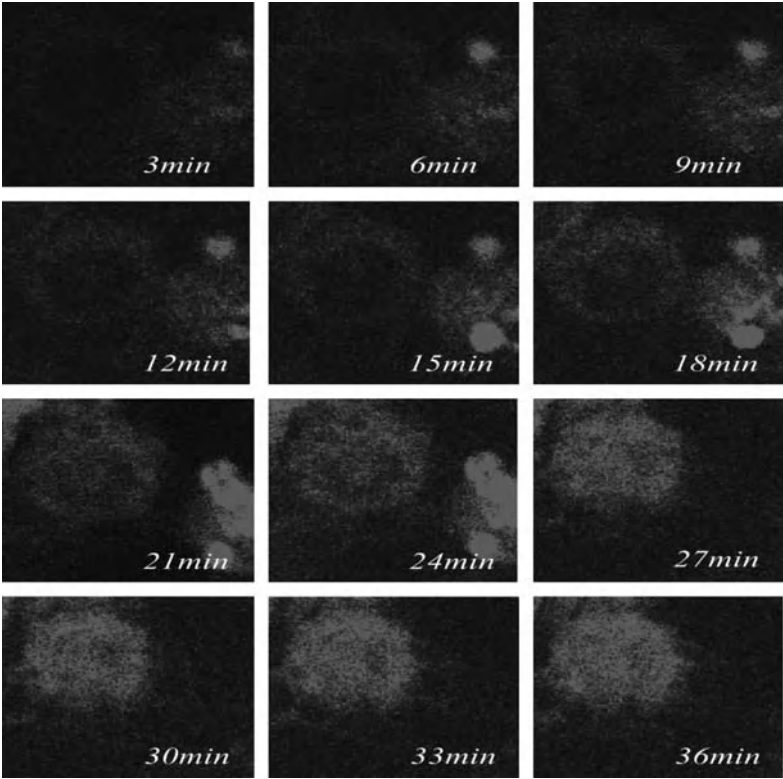


Fig. 25.14. Time-dependent accumulation of nanoTherapeutics in breast carcinoma cells (MCF-7)

In a second experiment, the distribution of the nanoparticles in the cancer cells was investigated. Figure 25.15 shows the distribution 15 min after incubation at 37°C.

Figures 25.16 and 25.17 are scanning electron microscope images of KB cells incubated with nanoMag grafted with LHRH ligands, while Fig. 25.17 shows the surface of a KB cell on which a nanoparticle is visible.

The aim of the following experiment was to determine the relation between the time of exposure to the magnetic field and the activity of the nanoparticles. It was carried out with nanoTherapeutics grafted with the hormone LHRH. The nanoparticles were incubated with MCF-7 cells, then subjected to magnetic fields for different lengths of time. Figure 25.18 reveals the linear dependence of the number of lysed cells on the time of exposure to the magnetic field (up to 30 min). The longer the exposure to the magnetic field, the more efficient the magnetocytolysis.

The effect of the particle concentration was studied in the following experiment. Breast carcinoma cells (MCF-7) were incubated with nanoMag-LHRH

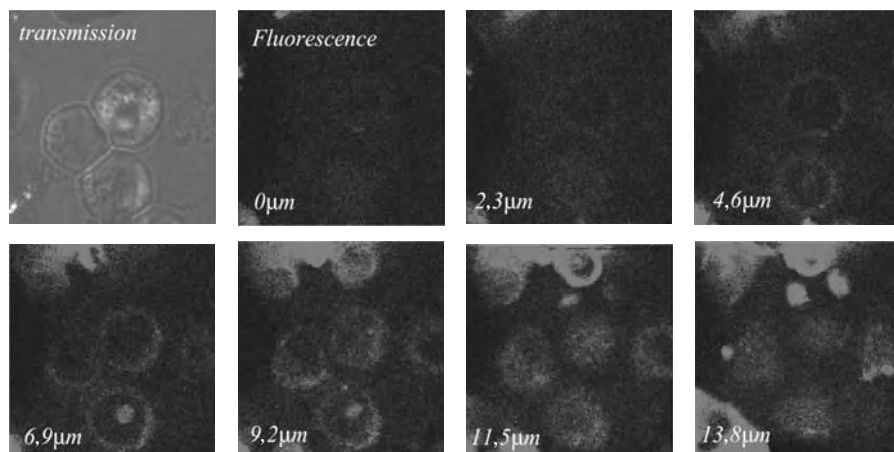


Fig. 25.15. Accumulation of nanoMag particles in cancer cells

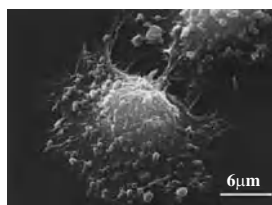


Fig. 25.16. Scanning electron microscope (SEM) image of KB cells

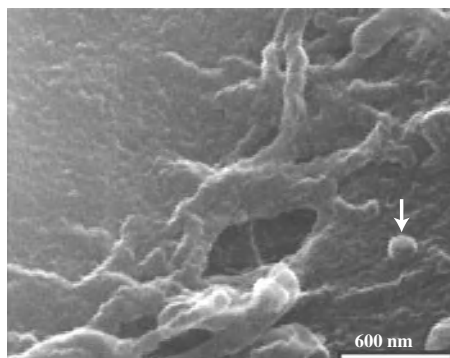


Fig. 25.17. Scanning electron microscope (SEM) zoom of the surface, revealing the presence of nanoparticles. The *arrow* indicates a nanoparticle at the surface of a cell

particles for 24h, then exposed to a magnetic field for 20 min. The control comprised cells incubated with non-activated nanoparticles. Between 0 and 25 particles per cell, it can be considered that there is no difference between the control and the exposed cells (values close to zero). Between 40 and 50 particles per cell are required to observe an effect. The graph in Fig. 25.19 shows that the magnetocytolysis phenomenon depends on the particle concentration.

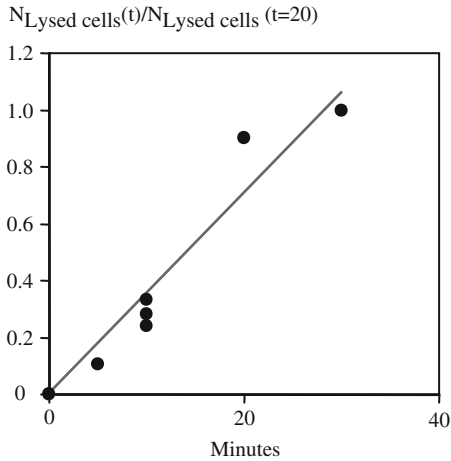


Fig. 25.18. Linear relation between magnetocytolytic activity of nanoMag and time of exposure of MCF-7 cells to the magnetic field

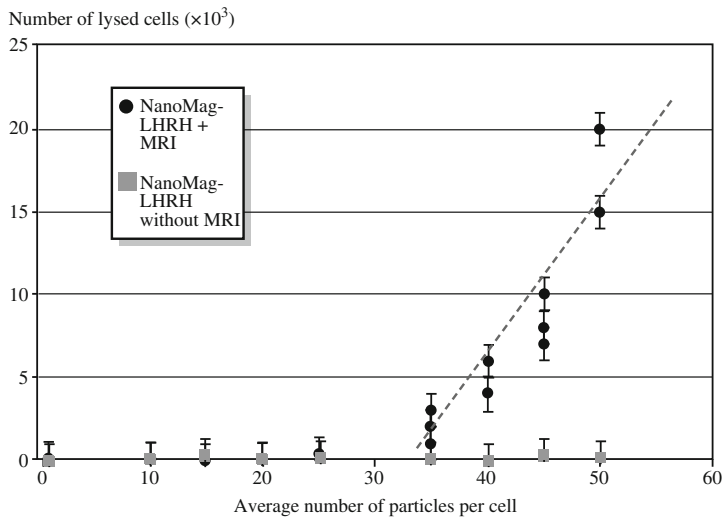


Fig. 25.19. Cell destruction vs. particles/cell

Figure 25.20 compares the magnetisation curves of particles coated with hormones, in the presence or otherwise of cells from the buccal epidermis (KB). The magnetisation M of a particle population can be described by a Langevin function $L(x)$ given by

$$L(x) = M_S \left(\coth x - \frac{1}{x} \right), \quad \text{with } x = \frac{\mu(D_D)H}{kT}, \quad \mu(D_D) = \frac{m_S D_D^3 \pi}{6}, \tag{25.1}$$

where M_S is the magnetisation at saturation, m_S is the magnetisation per unit volume, and D_D is the magnetisation diameter.

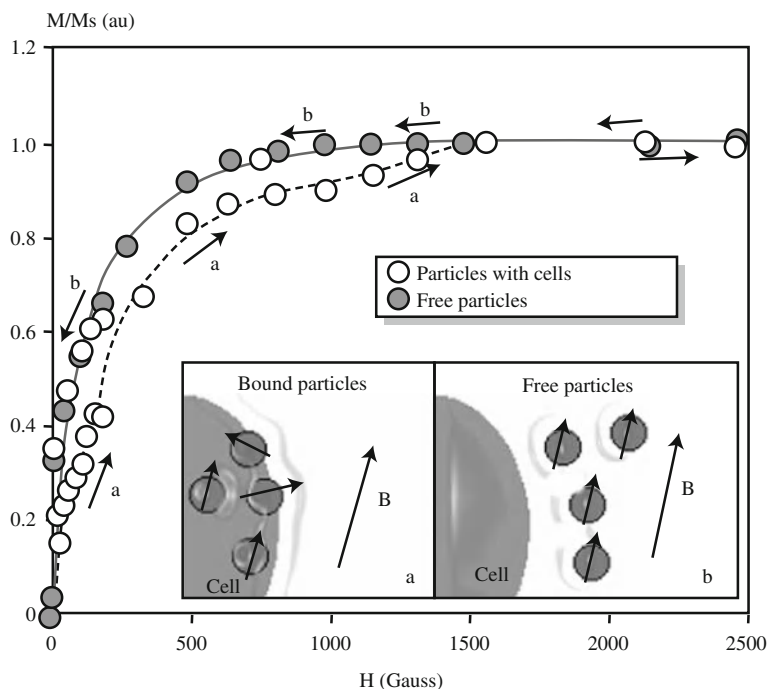


Fig. 25.20. Magnetisation curve of nanoparticles with and without cells. *Continuous curve*: fit with the Langevin equation

The size dispersion is determined using a log-normal distribution $p(D_D)$. The magnetisation curve is then fitted using

$$M = L(D_D)p(D_D)d(D)_D. \quad (25.2)$$

As the field strength increases, the magnetic moments gradually orient themselves along the field lines, until the magnetisation levels out at the saturation value M_S , when all the particles are aligned (continuous curve in Fig. 25.20).

Particles incubated for 2 h with the cells have a different magnetisation curve (the dashed curve in Fig. 25.20) from that due to free particles. The Langevin function can no longer be used to describe this curve. The steps observed can be explained by a resistance to magnetisation, induced by strong interactions between the membrane and the nanoparticles. Physical orientation in a magnetic field then requires more work to rotate bound nanoparticles than to rotate free particles.

These results suggest a probable destruction mechanism by physical orientation of particles subjected to a magnetic field. They also show that the activation of a single particle cannot cause enough damage to destroy a cell, but that many particles are in fact necessary.

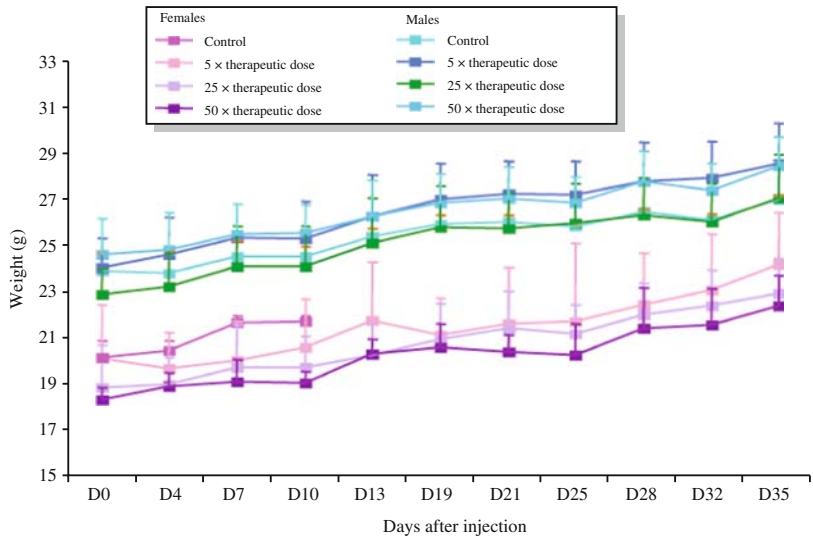


Fig. 25.21. Time evolution in the weight of C57Bl6 mice after injecting increasing doses of nanoMag particles

25.7.2 In Vivo Experiments

Figure 25.21 shows the dependence of male and female mouse weights on the time after increasing intravenous injections of nanoMag particles. For injected doses between 5 and 50 times the therapeutic dose, there is no difference in the weight development of the animals between the control group and the group treated with nanoMag, and this whatever the dose. Moreover, sacrifice and macroscopic and histological observations of the organs during autopsy revealed no sign of toxicity. The nanoMag particles thus seem well tolerated for injected doses between 5 and 50 times the normal dose in mice.

Similar experiments were carried out with repeated injections of nanoMag particles. The graph in Fig. 25.22 follows the weights of mice before and after 5 intravenous injections of nanoTherapeutics. In order to check the tolerance of the particles after activation, some mice were subjected to a magnetic field (MRI). It was observed that there is no difference between the weights of mice in the control group and those of mice subjected to MRI. In addition, the sacrifice of the mice and subsequent autopsy revealed no sign of toxicity after histological examination.

Since the probability of finding nanoparticles in healthy tissues despite having targeted tumour cells is nevertheless nonzero, a check was made to ensure that the concentration of nanoTherapeutics in healthy cells did not induce their destruction when the treatment was applied. Like any therapeutic agent, there is indeed a threshold value of the concentration in the cell, below which there is not enough to induce any affect. It was thus shown that

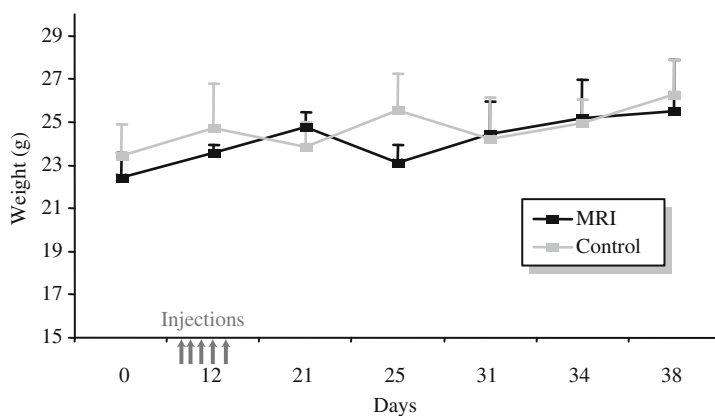


Fig. 25.22. Weights of C57Bl6 mice before and after 5 injections of nanoMag

the magnetocytolysis process is very accurately targeted and controlled both spatially and temporally.

The pharmacodynamics of nanoTherapeutics was assessed *in vivo* on human colon tumours C38 implanted in mice, after intratumoral injection of nanoMag particles or NaCl. The MRI images of the colon tumours in mice shown in Fig. 25.23 were taken after injecting nanoTherapeutics and compared with the control group of mice treated with NaCl. The nanoparticles are easy to see specifically *in vivo* at times 5, 20, and 48 h after injection, due to the positive contrast induced when they accumulate.

This accumulation in the tumour rather than in healthy organs demonstrates the specificity of nanoTherapeutics. Between the maximal accumulation in the tumour tissues and renal clearance, there is a time lapse of a few hours compatible with the implementation of a therapeutic protocol.

The efficiency of nanoTherapeutics has been evaluated on the growth of colon tumours grafted in mice, after MRI activation. The tumour volumes were measured, then compared with the tumour volume on the day of the injection in order to define the relative tumour volume. Figure 25.24 shows the relative tumour volume of mice treated intratumorally with NaCl (control), or with nanoTherapeutics activated by MRI 48 h after injection. The tumour growth decreased significantly after treatment with the nanoparticles. Histological examination (hematoxylin-eosine staining) of the tumours was then carried out, and necrosis was discovered in tumours treated with nanoTherapeutics.

The therapy using nanoMag particles has significant advantages over current forms of therapy, and represents a major break with traditional approaches. Indeed, this type of treatment could be accompanied by diagnosis in order to monitor the patient's prospects in real time. These nanoMag particles would be particularly relevant to the treatment of deep tumours, such as cancers of the liver, pancreas, stomach, and prostate, as well as their metastases. Systemic use of nanoMag with targeting agents will allow the treatment

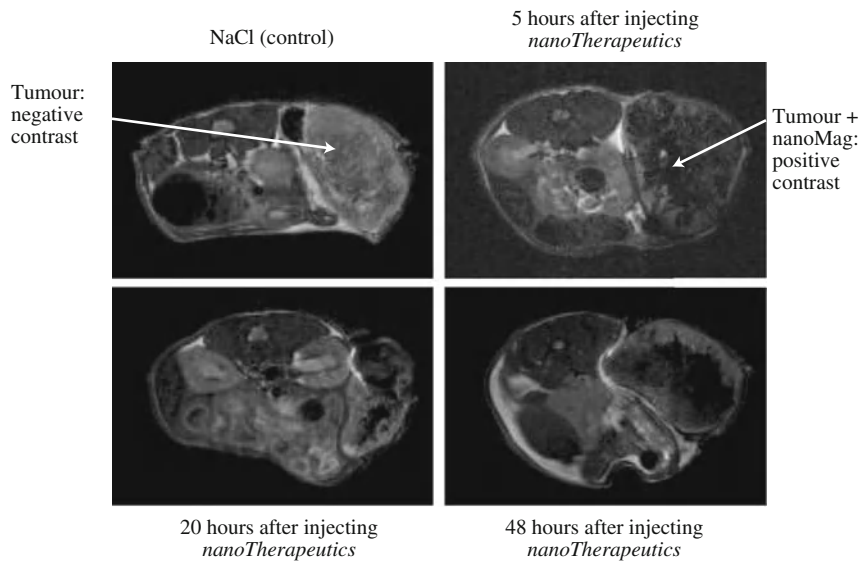


Fig. 25.23. Specific accumulation of nanoMag particles in tumour tissues in vivo 5, 20, and 48 h after injection

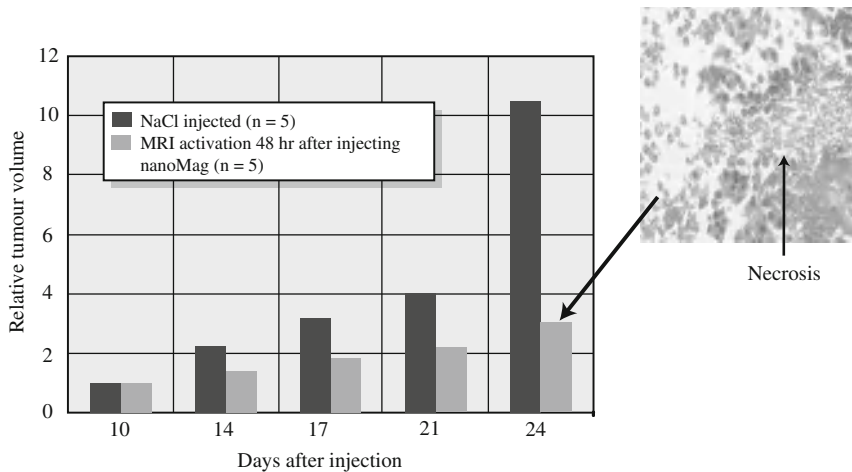


Fig. 25.24. Tumour growth after injecting NaCl or nanoTherapeutics activated by MRI, and histological examination of the tumour

to attain both the tumour and its metastases. Indeed, wherever they may be located, the metastases generally have the same receptors as the cancer that produced them.

25.8 Conclusion

The possibility today of controlling the properties of materials on the nanometric scale has opened up truly novel ways of fighting cancer. These breakthrough innovations give the hope of significantly improving the lives of hundreds of thousands of patients around the world, by fighting tumours against which traditional therapies were ineffective. The very concept of a drug is revolutionised by the chemical nature, structure, and use of nanoTherapeutics.

The different families of nanoTherapeutics developed by Nanobiotix meet needs that remain today largely unsatisfied, by targeting in a specific way both surface cancers and deep tumours. In the mid-term, these nanoparticles should provide a way of limiting the side effects caused by traditional treatments such as chemotherapy and radiotherapy, while improving their efficiency. Potential applications of nanoTherapeutics are many and varied, and could be extended to all the different types of cancer, as well as other pathologies.

References

1. Feltin, N., et al.: *Langmuir* **13**, 3927–3933 (1997)
2. Lévy, L., et al.: *American Chemical Society* **14**, 3715–3721 (2000)
3. Wang, X., et al.: *Proc. Natl. Acad. Sci. USA* **96** (20), 11081–11084 (1999)
4. Krebs, L.J., et al.: *Cancer Research* **60** (15), 4194–4199 (2000)

The Medical, Social, and Economic Stakes of Nanobiotechnology

J. Hache and F. Berger

The preceding chapters have reviewed some theoretical and technological aspects of research in nanobiotechnology. Recall that the latter results from the combination of biology and nanoscale engineering, to manipulate living organisms or to construct materials inspired by biological molecular systems. The aim of this final chapter is to describe a range of examples to illustrate their fields of application and the potential they represent for diagnosis and therapy in the health sector, but also for cosmetics, the food industry, and the environment, bearing in mind the need to foresee the risks, still poorly understood, associated with the use of nanoparticles and nanosystems.

26.1 From Current to Future Applications

It should be pointed out that nanotechnology is already having an impact on products as wide-ranging as new foodstuffs, medical instruments, chemical coatings, personal medical testing kits, sensors for safety systems, and water recycling units, not forgetting their impact on new types of therapy.

26.1.1 Diagnosis and Therapy

In the health sector, there are three main lines of applications for the development of nanobiotechnology:

- In vitro and in vivo diagnosis and imaging, together with high-speed screening of gene transcription and translation, in particular for differential analysis between healthy and pathological cases.
- Therapy, and in particular the development of new drugs based on nanostructured entities, the ‘intelligent’ delivery of drugs targeting well-defined locations in the body, and therapeutic follow-up.
- Compensating for bodily deficiencies by longer-lasting and better suited prostheses, tissue engineering, and the development of new biocompatible materials for regenerating bone, tissues, and nerves.

In Vivo Diagnosis

To improve diagnosis, it is better to check *in vivo*, and this is already a reality. A tiny capsule called M2A [1], the size of a large pill, contains a minute colour video camera able to carry out an endoscopy as it makes its way through the digestive system (see Fig. 26.1). It moves through the digestive tube for 8 h, taking two colour images per second and transmitting them by radio in real time, and thereby producing a genuine film of its journey.

From another point of view, the cells making up our bodies are microscopic. On their length scale, a scalpel is a disproportionately large tool. It is thus useful to imagine a device measuring only a few nanometers, which carries no risk for the organism. Equipped with ultrasensitive sensors, it could send accurate local images back to the doctor, thus facilitating diagnosis.

As an example, some materials, such as cadmium selenide, are fluorescent. Prepared in the form of nanocrystals, called quantum dots, they are subject to quantum effects in such a way that they emit light with a colour that changes with their size when they are excited by ultraviolet radiation. They can thus be used for molecular tagging, so that chemical reactions and biological processes can be tracked in detail in living cells.

These quantum dots have been used to show that the synapse, the junction ensuring communication between two neurons, is in dynamical equilibrium. Receptors on the membrane of the target neuron, which ensure this communication, were tagged with quantum dots, so that they could be monitored by detecting the emitted light. It was shown that, far from being stable in number and position, these receptors continually change their location, their distribution evolving in a specific way depending on whether they are near, outside, or inside the synapse [2].

In Vitro Diagnosis

For a few years now, DNA, protein, and cell microarrays have been used to carry out analyses in parallel, whilst requiring ever small amounts of sample. At the crossroads between microelectronics, nucleic acid or peptide chemistry, image analysis, and data processing, they are used to analyse thousands of samples simultaneously by virtue of their miniaturisation and automation.

They will be even more widely used when detection methods become more sensitive. One proposal for improving detection is to use a stack of dielectric layers, with the help of which fluorescence biochips will be able to achieve unequalled levels of performance and sensitivity (see Fig. 26.2) [3].

Another approach for improving detection abandons the use of fluorescent molecules as tags. The idea is to use inorganic nanowires functionalised with antibodies or nucleic acids. When the target binds onto its receptor, the electrical resistance is modified, allowing a highly sensitive detection [4].

Miniaturisation and automation have also led to tiny laboratories, called labs-on-chips, measuring no more than a few micrometers across (see Fig. 26.3).



Fig. 26.1. M2A minicapsule. Length 30 mm, diameter 11 mm

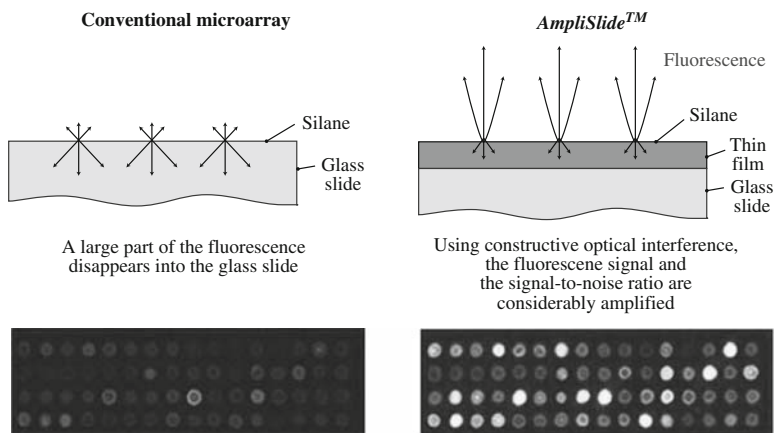


Fig. 26.2. Improving the detection sensitivity of fluorescence biochips by means of an amplifying substrate (AmpliSlide) [3]. (Source Genewave SAS, France)

These use the motion of microvolumes of liquid in capillaries etched into substrates like glass, silicon, or plastic. Liquid/solid interactions between the fluids circulating in the channels and the channel walls are then dominant. These labs-on-chips integrate all the operations required for analysis, from extraction of the analyte to preparation of the sample, molecular recognition, and detection of this recognition. The aim is to reduce costs, particularly with regard to labour, amounts of reagents, and time of analysis.

Therapy

Along with progress in diagnostics, there is now some hope of developing nanomedicine, i.e., machines made up of nanometric components moving around in the human body to take samples or deliver drug molecules to specific targets. This is what is illustrated in Fig. 26.4, an artist's view of a future microrobot introducing its microsyringe into a red blood cell.

While this image only attempts to show an application that may one day be possible, nanobiotechnology is already having a significant impact on the discovery of new molecules, by identifying targets on which to act, then



Fig. 26.3. Lab-on-a-chip. PCR amplification and detection of DNA [5]

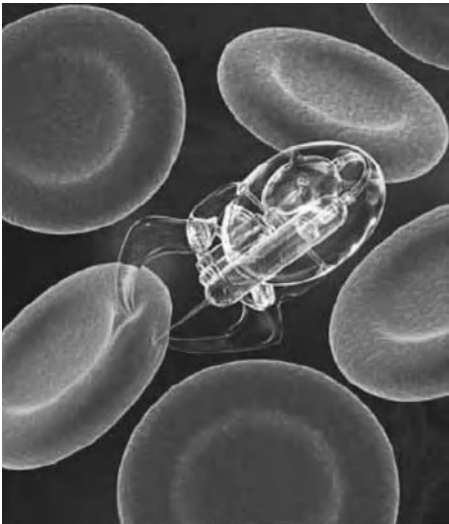
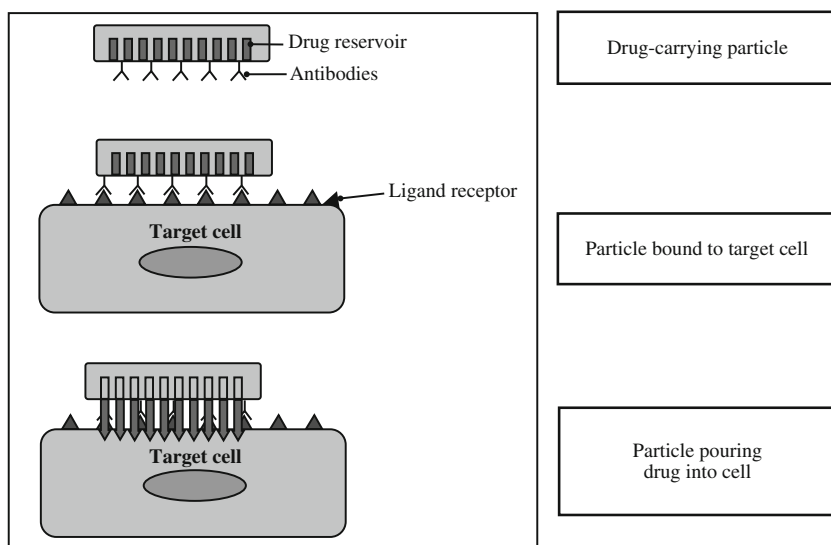
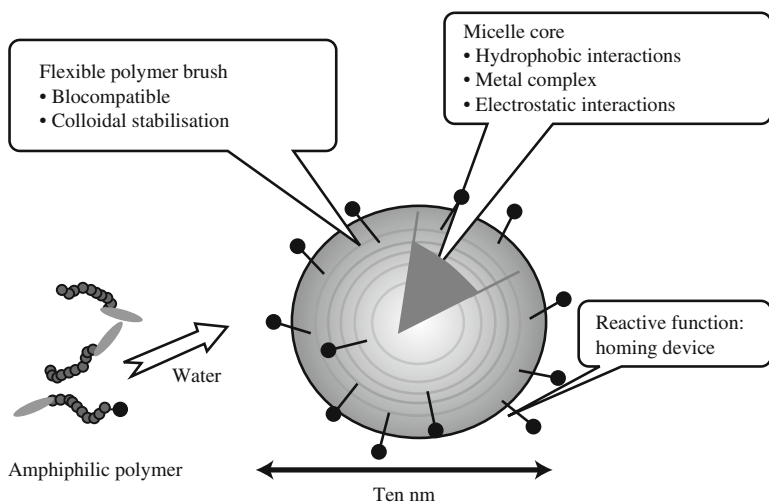


Fig. 26.4. A microrobot equipped with a microsyringe [6]

designing active molecules that can interact with them. The approach is based on differential analysis of gene transcription and translation in pathological and healthy cases.

Nanobiotechnology is especially promising when it comes to the intelligent delivery of new drugs. A conventional drug, distributed throughout the human body by the blood, cannot always easily reach its target tissue or organ. But a drug is only useful if it is present in the right place at the right concentration. This is why drug delivery is such an important concept. By encapsulating

Drug nanocarriers**Fig. 26.5.** Nanocarriers for drugs. Structure and mode of action

an active principle in a nanoporous membrane, it is prevented from being diluted in the blood during transport, and its release is only activated when the nanoporous membrane recognises the desired site for action. The active molecule is thus carried right up to the target. Here are some examples:

- Liposomes and nanoparticles were the first drug carriers to be used, in particular for taking an anti-cancer agent into the liver with the help of macrophage recognition processes.
- Peptide hormones, such as growth hormones, are currently administered using microcapsules. The active peptides are enclosed within a shell made from lactic acids or from polymers composed of lactic and glycolic acids. Hormone release then depends on the rate of degradation of the polymers.
- Surface modifications of the nanocarriers by means of hydrophilic polymers allow drugs to be directed toward biological regions other than the hepatic system.
- Current carriers can now take their contents selectively to cells, using specific receptors or markers at the surface of target cells (see Fig. 26.5). This is exemplified by the use of nanobeads for novel therapeutic strategies. By associating these nanobeads with specific antibodies of the tumour cells, antibodies playing the role of homing device through their affinity for certain targets, they are carried toward the selected tumour cells, which can then be destroyed by some external action, e.g., application of a magnetic field causing mechanical destruction [7] or infrared heating releasing a chemically acting active principle. These strategies provide a better way of destroying targets and thus having specific action without risking toxicity in neighbouring healthy tissues.
- By controlling the surface of an implant to within the nanometer, it should be possible to extend its lifetime and improve its biocompatibility. For example, an artificial hip coated with nanoparticles could bind with increased solidity to bone tissue.
- A connection between an electronic device and cells like neurons will be able to restore a nerve connection, for example, helping a paralysed person to recover some autonomy.
- Finally, nanobiotechnology may one day be able to reconstruct tissues by means of biocompatible materials on the scale of the cell. The idea is to associate biocompatible nanostructured materials, such as organic polymers, with living cells to replace defective tissues.

In the longer term, the implementation of microscopic sensors built up atom by atom would provide an extremely accurate view of tissues and physiological functions as they travel within groups of cells. With such tools, doctors would be able to monitor cell activity with detailed snapshots and an accurate map of the concentrations of various compounds and many physiological parameters.

To end this section, it can be concluded that the combination of memory capacity, mobility due to extreme miniaturisation, and network reconfiguration should open the way to powerful neurotechnical uses, such as repair by remote stimulation using radiofrequency identification (RFID). The full history of a patient can be transmitted to any computerised hospital system using the RFID technology [8].

26.1.2 Cosmetics

Nanotechnology is already used in the cosmetics industry, in particular to produce transparent creams since, by Mie's theory, nanoparticles barely scatter visible light:

- Zinc oxide nanoparticles improve lipstick hold.
- Zirconium oxide nanoparticles are used in nail varnishes.
- Titanium oxide nanoparticles filter ultraviolet light.

Although the visual aspect is generally the most important in cosmetics, other properties such as non-toxicity, hold, how it feels to the touch, how well it retains perspiration, how easy it is to apply, and so on, are also desirable.

- The need for sun creams to hold in contact with water has led to the development of creams containing zinc oxide nanoparticles with hydrophobic properties.
- Silicon and zinc oxides capture enzymes drying out the skin, while antioxidants associated with fullerenes eliminate free radicals, thus resisting aging of the skin.
- Titanium or zinc oxides are non-toxic, chemically inert, and remain at the surface of the skin, unlike some organic sun filters.

The effectiveness of protective creams is an important theme today, now that the general public have become aware of the dangers of exposure to the sun and the need to protect themselves. Compared with organic molecules, hydrophobic nanoparticles containing titanium or zinc oxides block UV radiation at low concentrations, while remaining photostable and penetrating the skin as little as possible, thereby limiting allergic or other cutaneous reactions.

The cosmetics industry thus uses thousands of tonnes of nanoparticles: titanium oxide nanoparticles have replaced former organic molecules in the production of sun creams, silicate nanospheres coated with titanium and iron oxide are used to reduce wrinkles, bismuth oxychloride nanoparticles improve the tactile properties of certain products, and so on.

Furthermore, the cosmetics industry is working closely with pharmaceutical laboratories to find carriers for the products it develops, whereby the action of their bioactive compounds may be better targeted.

The well-known product *Profil inceur* [9] contains an active component associating two molecules extracted from plants: NDGA, an acid extracted from *Larrea divaricata*, and baicalin, an active molecule from the perennial herbaceous plant scutellaria. This combination blocks several key steps in the formation of new fat cells. The action on the adipocytes is optimised by the continuous and targeted delivery deep in the skin: the active material is encapsulated in nanocarriers of diameter 300 nm which guide it by molecular affinity to the place where it must act.

A similar collaboration has been established with the food industry to develop new products.

Vitamin E is known for its ability to trap free radicals and fight aging of the skin. It is thus desirable in sun creams. Unfortunately, by its natural characteristics, it barely penetrates the outer layers of the skin and diffuses poorly through it. To overcome this problem, nanosomes [10] are used to transport pure vitamin E. These tiny intracellular vehicles modify the affinity of the skin, enter it, and release their active ingredients. The vitamin is densely distributed throughout the outer layers of the skin in the form of a gradient. The result is a genuine protective network against the effects of free radicals, with the sun cream playing its role of protecting the skin against ultraviolet radiation. The nanosome is expert in transporting lipophilic, i.e., oil-soluble, active agents. Given that the interstices in the horny layer of the skin measure around 100 nm across, nanocarriers are the best solution to the problem of transporting and concentrating active principles in the skin.

26.1.3 Product Quality and Traceability

Nanotechnology is also widely involved in food safety. There are biochips using DNA to detect microbial contamination in little over an hour, compared with the methods of immunology or bacterial cultures in Petri dishes which require at least 24 h to give their first results.

Example. To detect bacterial contamination very quickly, integrated systems associating the extraction of DNA, PCR amplification, hybridisation, and fluorescence detection [11] are able to simplify and standardise otherwise delicate operations that previously suffered from poor reproducibility.

Systems integrating nanosensors to detect harmful substances in food products could act on the colour in the tab carrying the nanosensors or the packaging, thus warning the consumer that the product is contaminated or out of date.

Example. A light-emitting protein can be modified to bind to the surface of microorganisms like *Salmonella* or *E. coli* [12]. When this protein binds, it emits visible light indicating that the food or drink is contaminated, the emission being all the brighter as the contamination proceeds.

The radiofrequency identification (RFID) technology already mentioned, is very useful for monitoring products. It is a system integrated within objects and allowing them to be identified. It acts like an electromagnetic barcode since, when it receives an electromagnetic pulse, it extracts energy and emits in its turn a response. But unlike the barcode, which must be scanned manually and read individually, an RFID label does not need to be passed in front of

a read system, and hundreds of labels can thus be read automatically every second.

Apart from its impact on packaging and control, nanotechnology is also used to offer tailored functional foodstuffs, incorporating nanocapsules guiding the delivery of certain nutrients.

Example. An Australian industrial baker incorporates nanocapsules containing fish oil, a source of the essential fatty acid omega 3, in its bread. These nanocapsules are designed to survive intact until they reach the stomach, so that the consumer does not suffer the unpleasant taste of fish oil.

With regard to agricultural production, nanostructured catalysts are being developed to increase the efficiency of pesticides and herbicides, so that doses can be reduced.

26.1.4 Environment and Risk Prevention

Nanotechnology opens extraordinary prospects for environmental protection through the design of non-polluting and economical manufacturing processes. It may thus be possible to clean up some of these processes, to limit the use and waste of toxic substances, to forestall or treat pollution, and so on. For example, there is a great deal of work in the following areas:

- Porous nanostructures and catalysts could trap and process pollutants emitted by cars, while coatings made from nanoscale multilayers will be able to reduce friction and reduce the need for lubricants like oil and grease.
- Selective nanomembranes will be able to filter water contaminants and desalinate sea water.
- New materials will open up new opportunities for recycling.

More generally, biochips will be able to identify the genes of microorganisms producing the most active enzymes for natural digestion of pollution and detoxification of different environments. These genes could then be transferred to common bacteria, which would play a considerable role in the preservation of those environments.

Even now, DNA chips, originally developed for genome and post-genome studies, have a natural application in the identification of pollutants. For example, BioMérieux [13] is collaborating with Suez in the development of a chip to seek out the DNA of several hundred bacteria in drinking water. Compared with conventional methods, such a biochip should be able to obtain a response that is:

- faster (a few hours instead of 1 to 2 days),
- more accurate (the sensitivity will be increased by a factor of 1,000),
- cheaper (about 10 times less).

26.2 From Individual Players to Clusters

The academic and industrial worlds are organised as best they can to develop the potential applications offered by nanobiotechnology, with the prospect of seeing, measuring, and manipulating entities on the atomic scale.

26.2.1 Different Players Around the World and the Position of France

All major countries are seeking to develop their own nanobiotechnological program. Government involvement is essential, but in tandem with the creation of new businesses. On the worldwide level, 1,200 new companies have already been set up in the field of nanotechnology, with half of these in the United States. France has many assets in this respect, in particular, a high level of scientific research.

26.2.2 Clusters and Other Poles of Competitiveness

Nanobiotechnology is based on the interaction of many different kinds of knowhow, and development in this area will require a suitable preparation for the next generation of students and research scientists to ensure that they acquire the right cross-disciplinary skills.

Technological clusters are collaboration networks, e.g., biotech/nanotech networks, between organisations of very different sizes but with common aims, complementarity, or interdependence. The point is to carry out under the best possible conditions concrete R & D projects, jointly designed and piloted by a critical mass of industrial manufacturers structured within clearly defined technological fields, with research and training centers (universities and public and private research establishments).

Minatec in Grenoble is the leading center for nanotechnology in France [14]. It is one of the three biggest nanotechnology research centers in the world, along with Albany in the USA and Selete near Tokyo. With 10,000 m² of clean rooms in 45,000 m² of laboratories, it has the advantage of concentrating in the same place a wide range of tools for fabricating nanostructures, together with the facilities for studying and testing them. Downstream of fundamental research activities, but collaborating closely with research institutes, this center is designed primarily to bring nanotechnological innovation to its industrial fruition.

26.3 From Funding to Industrialisation

26.3.1 Patents

The question of industrial ownership is crucial because, while nanobiotechnology raises great prospects, it should not be forgotten that the acquisition of

knowledge and its transformation into technological innovations in the form of new products, processes, or services, is extremely costly in terms of both time and money. The necessary capital investment cannot be obtained without industrial ownership, often the only asset of a new company when it is set up.

At the end of the year 2005, the breakdown of the total number of patents lodged shows that just five countries have filed 81% of these: 40% for the USA, 15% for Germany, 12% for Japan, 9% for France, and 5% for the United Kingdom.

The problem here is that a single innovation may have applications in a wide range of sectors, from aeronautics to pharmacology, whence a company that manages to file a patent early on may put itself in an unassailable position, with an incomparable advantage over its competitors, even to the point of blocking future developments. In this context, one could cite the monopoly on DNA chips established by Affymetrix [15] or the problems raised by patents filed by Myriad Therapeutics for the use of the genes BRCA1 and BRCA2 to detect predisposition to breast cancer [16].

26.3.2 Funding Nanobiotechnological Activity

The prospects for revolutionary applications of nanobiotechnology are such that funding and investment has been growing exponentially, to such an extent that the world market may reach 1,000 billion dollars by 2015. For example, in 2005, the US administration devoted more than a billion dollars to nanotechnology, while the European Union and Japan attribute similar amounts.

In 2004, 8.6 billion dollars were devoted to nanotechnology, including 40% by the United States, 35% from Asia, 22% from Europe, and just 3% from the rest of the world. This reveals US leadership, a clear determination in Asia, and an effort in Europe, with Germany leading the field in Europe. Nanotechnology lies in fourth position on Europe's list of research priorities, which stand as follows:

- life sciences, genomics, and biotechnology for health purposes,
- information technology,
- aeronautics and space,
- nanoscience and nanotechnology,
- food safety and health risks,
- sustainable development, environmental change, and ecosystems,
- citizens and government in the European knowledge-based society.

But the key question remains the same: how can we create wealth from knowledge and understanding produced by research, and convince investors to fund activities, probably profitable in the long term but not without risk? These are heavy investments, with a distant, often unreliable return, of the kind that investors generally shun.

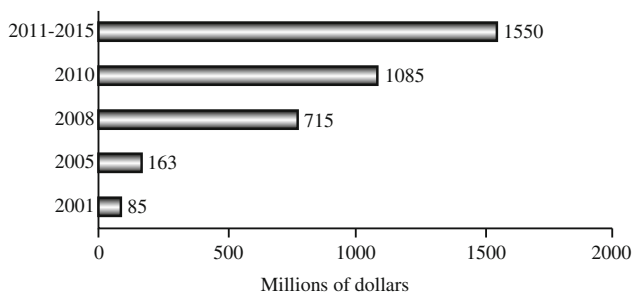


Fig. 26.6. Estimated world nanobiotechnology market

There are many different channels for funding, although the various stages in the lifetime of a product or activity are not covered in the same way:

- Initial funding, when there is still no product and no customers, is very often the weak link. It can be obtained from those directly or closely involved in the project – love money – and private investors – business angels – topped up by public funding, in particular by the ministries of research and industry, OSEO Innovation, and regional authorities.
- Risk capital, arriving in the early stages of development, has stricter selection criteria. It will finance on the basis of the quality of the team and the project (together with industrial ownership), and then usually only if there is a real prospect of a return after 3 or 5 years' funding.
- The stock market presents a contrasting picture, although it looks as though high-tech companies will make a come-back.

There is clearly a need for alliances, in particular between the multinationals and newly created companies. Such an alliance could be defined as a cooperative agreement between two or more competing or potentially competing companies, which commit themselves by contract to a common project, while retaining their legal and strategic autonomy.

26.3.3 The Markets: Between Fantasy and Reality

While the world market for products generated by nanotechnology was 120 billion dollars in 2004, it should reach 1,000 billion dollars, including 170 billion for the life sciences, by 2012, according to the National Science Foundation (see Fig. 26.6) [17].

The chip market, in particular the market for DNA chips, has matured. After a period of sustained growth of the order of 30–50%, the growth rate has stabilised around 10–20% annually, for a turnover of the order of 1,250 million dollars in 2005. The US company Affymetrix is the undisputed leader, with a close to 60% market share.

Other markets are still potential markets, even though promising, since some industrial sources would have it that, taken overall, the nanotechnology

market will be equivalent to the computer and telecommunications markets, and ten times greater than the biotechnology market.

26.4 From Risks to Precautions

Nanobiotechnology has been presented as a scientific, technological, productional, and industrial revolution, and the radical change in the conditions and way of life it promises has a certain fascination, whilst at the same time appearing to many as a threat. As in any field opening up new possibilities for human activity, it is worth being vigilant with regard to unexpected or harmful consequences.

26.4.1 New Risks and Ethical Considerations

The development and use of nanoparticles or nanostructures may bring poorly understood, or even completely unknown health risks. If they are to be used on a large scale, then these risks must be properly assessed [18]:

- What will happen to nanoparticles released into the environment? Might they not create a level of toxicity comparable with asbestos?
- Might bacteria become reservoirs for nanoparticles, thereby facilitating their entry into the food chain?
- How can we prevent nanoparticles, which are able to enter into cells, from accumulating in the organism?

Today, uncertainties over the long-term behaviour of nanoparticles in the environment, their ecotoxicity, and their toxicity for human beings raise major questions [20], although it is clear that nanoparticles have a particular biological reactivity, probably related to the high chemical reactivity of their surface compared with the very small dimensions of these particles. The difficulty lies, however, in the fact that current toxicological data have been gathered from studies of atmospheric pollution.

It would seem that the considerable increase in surface-to-volume ratio in nanosystems raises quite novel toxicological issues. Owing to their small dimensions, nanoparticles can not only reach the deepest ramifications of the respiratory system, but also cross the epithelial barriers with the greatest of ease and enter into the general blood flow. This feature, which is a strong point when such particles are used to release drugs into the brain via the blood–brain barrier, becomes a risk if, by passing into the blood flow, they are at least in part responsible for an increase in cardiovascular disease, as observed in populations exposed to atmospheric pollution.

The toxicity of nanoparticles, if any, is far from being understood, especially as it is not obvious how to extrapolate the results of tests carried out on mice or rats to human beings. However, it should be remembered that toxicity is only defined with respect to a dose, beyond which nanoparticles might become toxic.

On top of these health considerations, there are also ethical concerns:

- The miniaturisation of tests and the development of genetic profiles, even if they are fully justified in the context of personalised medical treatment, may open the door to abuse of personal privacy, with repercussions at work or in the field of insurance.
- The implantation of nanosensors in the human body, used to locate or control an individual (RFID), may also lead to abuse.
- The implantation of chips in the brain may even lead to the idea of remotely controlled people.

These are the preoccupations of developed countries. But what about the developing world?

Although developing countries may benefit from applications in the fields of medical diagnosis, drug distribution, food storage, and water treatment, it would be a mistake to ignore the problem of raw materials. For these countries, raw materials are still the basis of their economy. By developing new materials, nanotechnology will offer a broader choice of raw materials, possibly to the detriment of those proposed by developing countries. This has already happened for the rubber used in car tyres, to such an extent that we may expect the demand for natural rubber to drop sharply in the coming years.

26.4.2 Science Fiction or Future Reality?

As often happens in times of great technological progress, nanobiotechnology has stimulated fear and controversy. The general public has an ambivalent perception of these innovations. On the one hand, they have a certain fascination because a deep understanding of matter and the manipulation of the very atoms constituting its basic building blocks means, for example, that we may envisage the fabrication of machines the size of a virus, able to travel down our capillaries to repair individual cells. On the other hand, this potential is frightening, because one can conceive of molecular assemblies forming systems that may be able to outdo the human brain, or nanorobots with the ability to self-reproduce, which may exceed all the expectations of their designer, getting out of control, and destroying the whole of humanity.

It is no easy matter to predict just where science and technology will take us. It is often science fiction writers who come up with the most realistic attempts. Indeed, science fiction has already turned its attention to nanosystems. For example, in his novel *Prey* which appeared in 2002 [19], Michael Crichton describes nanorobots able to self-assemble and self-replicate using carbon available in their environment. As soon as they escape from human control, they devour everything they come across, including human beings themselves. It is symptomatic of our fear of losing control, or the fear of an irreversible change due to one of humanity's great mistakes.

The prospects opened up by nanotechnology has thus earned it a starring role in works of science fiction. It turns up in futuristic films like Luc Besson's *The Fifth Element*, or literary works like *The Diamond Age* by Neal Stephenson [20], which describes the city of Shanghai in the middle of the twenty-first century, in a future featuring the extensive use of nanotechnology.

However, one should not confuse science fiction with reality or the potential development of nanotechnology. Although some applications have already materialised, e.g., self-cleaning glasses, inkjet printing heads, and beauty nanoproducts, it has already been stressed that the potential applications of nanobiotechnology will cover a broad spectrum of industrial activities, especially those related to health and the environment.

Although some do not hesitate to use living systems, with their self-organising, self-replicating, and self-complexifying properties, to bring them into the service of human ambitions (the aim here is not to make up for some deficiency, but to try to improve living systems, and even construct some kind of superhuman being [21]), it should be remembered that we are still a very long way from such a goal, that we cannot be sure that such a thing could ever be achieved, and that it is better therefore to remain within the realms of reason.

26.4.3 Image and Communication

Nanobiotechnology is still an emerging technology, in the sense that there is still no genuine application in the eyes of the general public. The fantasies that will inevitably accompany its development – to speak of the infinitely small is to speak of the invisible, triggering the imagination in a whole range of different ways – must be met by a broad campaign of information to forestall any negative reaction or the kind of behaviour inspired by the development of genetically modified organisms.. The slightest difference between the reality of progress in nanobiotechnology and the way it is actually perceived by public opinion could indeed be disastrous.

The aim is not of course to conceal potential risks, such as the toxicity of nanoparticles or the problems due to their dissemination in the environment, but rather to demonstrate clearly that research in these areas is being carried out in parallel with development of the applications. It is important to avoid a situation where nanobiotechnology is perceived as a source of both progress and disquiet, and the research scientist as a kind of sorcerer's apprentice. Scientists must publicise the results and the implications of their work, otherwise there is a great risk that the precautionary principle will dominate public debate, applied indiscriminately to become a principle of inertia and inaction. Communication is every bit as important as innovation.

26.4.4 Convergence of Nanoscience and the Life Sciences

Many applications stem from the possibility of studying the way a living cell works with nanoscale resolution, both in the health sector and in everyday

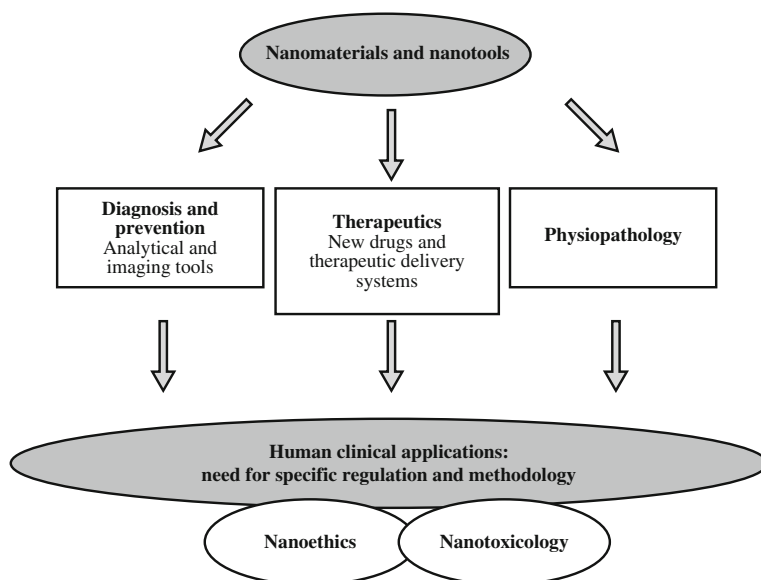


Fig. 26.7. Nanotechnology and nanomedicine

life, opening up the prospect of reconstructing biological entities from their cellular components. But this is only possible when physics, chemistry, and biology are brought together. This meeting is not a new thing, since physical methods already underpinned molecular biology, but it confirms the idea that cross-disciplinary interaction offers much greater prospects for development than the scientific disciplines taken individually.

Nanobiotechnology has the highly promising feature of being one of the areas of cross-disciplinary R & D in which there are many potential biological applications. But this is only the dawn of such foretold revolutions, and it is important not to confuse reality and science fiction. Despite the pressure from manufacturers and investors, knowledge and understanding must continue to move ahead through fundamental research in order to foresee and avoid any excesses.

26.5 The Advent of Nanomedicine

The above discussion serves to illustrate some applications and some of the issues at stake when nanotechnology is applied to life and the quality of life, but it is essential to gauge the importance of nanotechnology in the health sector, and hence the future of nanomedicine. Figure 26.7 gives an overview.

Applications of nanomedicine in diagnostics and imaging are already part of medical practice:

- The discovery of DNA and the development of PCR have made molecular examination possible in pathology, and a change of scale from the microscopic to the nanoscopic.
- Genome-wide analysis allows us to go beyond a monogenic concept, where a single gene is responsible for a given pathology, to the idea of a complex dysfunction of biological systems to explain some pathologies.
- DNA and protein chips make it possible to define a molecular profile constituting a genuine identity card for a pathological process, which can be used for diagnosis, but also prognosis and therapeutic response.
- Imaging methods, already widely used, have become much more accurate with the use of nanoparticles as contrast agents, for example, to the point where we can form images of what is going on in a single cell.

In therapeutics, nanoscience has opened the way to new galenical forms for the intelligent delivery of drugs and to new therapeutic forms which only act at the precise location of the problem.

For drug delivery, three approaches have been envisaged:

- The first is encapsulation of drugs in materials like liposomes or polymers. These materials give better solubility, better diffusion, and control over degradation. They thus improve the bioavailability of the drug.
- The second is drug carrying, which aims to improve the ability of drugs to reach their target. The drug acts specifically and solely at some predetermined site, the carrier of the active principle being equipped with a recognition molecule for the intended site. The great clinical challenge is to target cancer cells.
- The third associates carrying with controlled release of the drug. It provides a way of controlling all the pharmacological parameters. The challenge for the pharmacologist is to connect together all the relevant molecules, viz., active principle, polymer, and carrier, to make multipurpose constructs able to act directly on the biochemistry and physiology of diseased cells. This approach is particularly sought after when repeated administration must be avoided, either because access is difficult, or because it may be traumatising. The brain and the eye are a priority here for this reason.

New forms of therapy are appearing on the scene:

- Nanotechnology can be used to devise drugs that can be activated at will, e.g., by applying a magnetic field.
- Silica nanospheres coated with a thin film of gold can enter inside tumours. By shining an infrared laser on the tumour, the laser beam passes through healthy tissue without affecting it and heats the nanospheres in the tumour tissue, thus destroying only the malignant cells around it.
- Gene therapy offers extraordinary therapeutic prospects, endowing a cell with a missing gene or inhibiting the expression of a pathological protein. It should benefit from nanotechnology for the design of non-viral carriers

able to condense DNA into small enough particles to be internalised within a cell.

- Nanofabrication and molecular self-assembly may be able to provide scaffolding for the fabrication of tissues and cells. We will then be able to make skin for patients with serious burns, and liver cells for those with liver complaints.

In the longer term, medicine will be revolutionised by nanomachines that can be injected into an organism and survive within it, in the same way as a viral or bacterial pathogen, to carry out surveillance and therapy. These entirely synthetic machines, produced using a bottom-up approach, would then play the role of therapeutic viruses like retroviruses, adenoviruses, and more recently lentiviruses.

Finally, nanoscience will allow significant progress in regenerative medicine, which aims to help the body to recover by its own means. After the development of implantable materials, then biodegradable materials, we should see materials combining both bioactivity and biodegradability in a single structure. Added to this will be progress in the control of cell differentiation on nanoscale solid surfaces, which will facilitate the reconstitution of organs from stem cells.

Whether one speaks of diagnosis or therapy, the potential for transfer of nanotechnology to the medical field is fully recognised today. However, two consequences are often invoked: the risks associated with the use of such techniques and a change in the very notion of a disease.

One issue then is the toxicity of nanoparticles. By studying the effect of carbon fibres on rat lungs, it was shown that, three months after the injection, the rats exhibited granuloma inflammations. It is not easy to extrapolate these results to humans, partly due to the doses administered during the tests and partly due to the direct injection into the lungs, both far removed from realistic conditions. But these tests illustrate the need to develop a genuine field of nanotoxicology, applied to suitable preclinical and cell models. This is a classic approach for the introduction of any new therapy or drug since, at the end of the preclinical stage involving animals, clinical stages are set up on humans according to a protocol laid down and overseen by a committee for the protection of the individual, which validates the method for the therapeutic trial both ethically and with regard to the rights of the patient, consistently with known medical data.

The transition from anatomo-clinical medicine, based on the analysis of symptoms and histological studies, to nanomedicine does not for its part merely reflect a change of scale. It induces a change in the very notion of disease, in the sense that the pathology becomes detectable on the level of molecular determinism. In a nutshell, the disease is potentially detected before it can begin to develop.

This nanomedicine, which involves a cross-disciplinary effort between chemists, physicists, biologists, engineers, and doctors, of a kind never before

witnessed, still requires validation of the underlying determinism for anomalies detected on the scale of biological nanosystems, but it marks the way to predictive and preventive medicine, as a real alternative to curative medicine.

References

1. www.givenimaging.com
2. Triller, A., et al.: Biologie cellulaire de la synapse normale et pathologique, *Science* **302**, no. 5644, 17 October (2003)
3. www.genewave.com
4. Lieber, C., et al.: Logic gates and computation from assembled nanowire building blocks, *Science* **294**, no. 5545, 1313–1317, 9 November (2001)
5. www-leti.cea.fr
6. www.conevliav.com
7. Levy, L., et al.: Nanotechnology biomedical applications, *Mol. Cryst. Liq. Cryst.* **34**, 589–598 (2002)
8. Lewis, P.: Chip implants present some intriguing possibilities and raise a host of concerns about privacy and ethics, *Applied Digital*, 27 July (2004)
9. www.ysl.com
10. Le Marois, G.: Les nanomatériaux, *DIGITIP*, 12 July (2004)
11. www.genesystems.fr
12. www.agromicron.com
13. *Gazette Labo*, no. 36, March (1999)
14. www.minatec.com
15. www.affymetrix.com
16. Hache, J.: *Les enjeux des biotechnologies*, EMS editions, Collection Pratiques d'Entreprises (2005)
17. Roco, M.C., Verdana, et al.: Converging technologies for improving human performance: Nanotechnology, biotechnology, information technology and cognitive science, NSF/DOC-sponsored Report, Arlington, VA, National Science Foundation, June (2002)
18. Centre de la prévention et de la précaution: Nanotechnologies, nanoparticules – quels dangers, quels risques?, Note du Ministère de l'écologie et du développement durable, Paris, May (2006)
19. Crichton, M.: *Prey*, HarperCollins (2002)
20. Stephenson, N.: *The Diamond Age*, Bantam Books, London (1995)
21. Benoît Browaeys, D.: Les transhumains s'emparent des nanotechs, *Vivant* no. 3, Paris (2004). www.vivantinfo.com

Index

- absorbance, 894
- abzyme, 874
- acetylation, 20, 927
- acetylcholinesterase, 888, 889
- acidosis, 645
- acridine, 13
- acrylamide, 1133
- actin, 442, 444, 1049
 - assembly dynamics, 177–181
 - ATPase cycle, 179
 - comet, 176, 183, 185
 - depolymerising factor, 177, 179, 182
 - directional polymerisation, 183
 - factory, 176
 - filament, 104, 108, 172, 173, 194–196, 446
 - autocatalytic branching, 190, 191
 - branching, 182
 - dendritic nucleation, 190, 191
 - directed assembly, 174–193
 - flexibility, 200
 - observation, 176
 - unbranched, 183
 - filamentous, 177
 - force produced by, 187–191
 - globular, 177
 - mesoscopic stress model, 192
 - polymerisation, 175–177
 - structure, 174
- activatable probe, 301, 340
- active principle, 1045–1048, 1097–1115, 1150
 - burst effect, 1105
 - encapsulation, 1098, 1104
 - incorporation in carrier, 1104
 - toxicity, 1125
- active principle–carrier association, 1104–1105
- active targeting, 158
- ADA deficiency, 1080, 1081
- adeno-associated vector, 1070, 1078–1079, 1086
- adeno-associated virus, 1078
- adenocarcinoma, 648
- adenosine, 101, 173, 396
 - deaminase, 1080
- adenoviral vector, 1070, 1077–1078, 1082
- adenovirus, 499, 1077
- adherens junction, 86
- adipocyte, 1149
- ADP, 207
- adrenaline, 684
- Aequorea victoria, 290
- aequorin, 290
- aequorin–GFP system, 290
- affinity biosensor, 895–897
- affinity capture, 944
- affinity chromatography, 1004
- Affymetrix, 310, 313, 903, 916, 918, 1154
 - chip, 779
- AFM, *see* atomic force microscopy
- African green monkey, 261
- agarose, 878
- aging, 25

- albumin, 938, 1031
 - bovine serum, 211, 521, 629, 876
 - human serum, 535, 888, 1029, 1030
- alga, 34, 642
- alginate, 1028, 1105
- aliasing, 785
- alkaline phosphatase, 898
- alkoxysilane, 136, 141, 145, 277
- alkylamine, 142
- all-optical histology, 404
- allergen, 956
 - microarray, 956
- alumina surface chemistry, 135
- Alzheimer's disease, 379, 496, 689, 1082
- amide, 873
 - bond, 1047
- amine, 431, 432
- amino acid, 496, 873, 937, 938
 - nitration, 689
 - on Mars, 1001
 - side chain, 879
- aminoglycoside, 1113
- aminolysis, 1051
- aminopeptidase, 595
- ammonia, 884
 - sensor, 893
- amoeba, 175
- AMP, 231, 641
 - cyclic, 1033
- amperometric detection, 675–678, 872, 880
 - double potential step, 677
 - single potential step, 675
- amperometry, 667–695
 - concentration profile, 678–682
 - observation of exocytosis, 686, 688
 - observation of oxidative stress, 690
- amphiphilic molecule, 32, 36, 39, 50, 63, 557, 713, 1099, 1104, 1113
- amphotericin, 662, 665, 1111
- amplicon, 842
- amylase, 388
- amyloid fibril, 379, 496
- analog-to-digital converter, 784, 785
- anesthetic, 645
- angiogenesis, 175, 244
 - in tumours, 1110
- angiography, 163
- animal model, 1058, 1059, 1061, 1071, 1085–1089
- annexin, 87, 88, 91
- anoxia, 641
- antiarrhythmic drug, 869
- antibiotic, 654, 662, 665, 1090, 1113
 - administration, 1045
 - cement, 1045
 - delivery, 1055
 - detection, 500
 - encapsulated, 1113
- antibody, 78, 121, 244, 332, 334, 335, 351, 447, 913, 927, 938, 939, 950, 956, 969, 1001, 1004, 1148
 - A-HAS, 535
 - anti-biotin, 949
 - anti-GST, 490, 491, 941
 - anti-mouse, 333
 - anti-parathion, 888
 - array, 953
 - catalytic, 873, 874
 - detection, 500
 - fluorescent, 610
 - for biosensor, 872
 - immobilised, 889
 - labelling, 257
 - microarray, 311
 - monoclonal, 16, 83, 84, 94, 494, 497, 955, 1109
 - polyclonal, 535, 941
- antibody–antigen recognition, 389, 534, 561, 874
- antidepressant drug, 869
- antigen, 389, 874, 1004
 - array, 953
 - detection, 534
 - immobilised, 889
 - mapping, 388
- antioxidant, 1149
- antitumour molecule, 13
- apoptosis, 645, 648, 929, 930
 - of tumour cell, 1126, 1128
- APS, 137, 138, 164
- aptamer, 232, 944
 - antithrombin, 243
 - array, 945
 - as detection tool, 238–240
 - as purification tool, 237–238
 - as regulatory tool, 240–242

- as research tool, 236–237
- as therapeutic tool, 242–244
- binding, 230
- chip, 239
- clinical trial, 244
- definition, 224
- sequence, 235
- APTES, 332
- aquaporin, 938
- archaeobacterium, 13, 413
- arginine, 110
- argon laser, 317, 318
- argon/krypton laser, 317
- arrhythmia, 647
- artery, 1056
 - revascularisation, 1056
- artificial synapse, 694
- assay, 841, 913
 - enzyme, 953
 - glucose, 871, 881
 - urea, 872
- astrocytoma, 648
- atomic force microscopy, 58, 60, 70, 73, 89, 187, 375–389, 402–434, 786, 951–952
 - AM, 413, 417–422
 - amplitude image, 378
 - basic principle, 376
 - cantilever, 376, 387, 402–403
 - contact mode, 376, 405–412
 - deflection image, 377, 378
 - DNA observation, 427–434
 - FM, 413, 414
 - friction image, 377
 - height image, 377, 378, 427–434
 - imaging, 375
 - in liquid medium, 422–427
 - invention, 402
 - ligand–receptor force measurement, 407
 - membrane imaging, 380–384
 - of cells, 384–387
 - phase image, 378
 - polymer force–extension curve, 407–412
 - quality factor, 417–422
 - recognition imaging, 389
 - tapping mode, 376, 377, 413–434
 - tip, 403
- ATP, 173, 207, 287, 425, 446, 640, 641
 - detection, 897
 - hydrolysis, 173, 196–199, 207, 211, 215
 - synthase, 206–220
 - mechanical model, 218
 - synthesis, 207, 929
- autocorrelation function, 343, 344, 349, 350
- autofluorescence, 264, 276, 307, 915
- autoimmune disease, 955
- autoluminescence, 287
- automated measurement, 783
- automation, 937, 974–976, 980, 999, 1001, 1144
- autoradiography, 947
- avidin, 333, 406
- avidin–biotin system, 73, 74, 520, 955
 - separation force, 405–407, 410
- axon, 175, 646, 970
 - giant, 649, 665
- axoneme, 197
- bacillus, 328
- bacteriology, 862
- bacterium, 12, 103, 188, 206, 284, 379, 384, 499, 1075, 1078, 1151, 1155
 - adhesion, 562, 1028
 - ATP synthase, 209
 - DNA, 13
 - for biosensor, 872
 - halophilic, 85
 - in optical trap, 446
 - in water, 1151
 - intracellular, 1113
 - light-emitting, 897
 - membrane, 694
 - motion, 183, 184
 - pathogenic, 153
 - proliferation, 1028
 - proteins, 873
 - S layers, 412, 413
 - stealth, 1107
- bar code, 1150
 - on metal rod, 954
 - optical, 278, 953
- bay scallop, 203
- Bayesian methods, 794
- bead microarray, 917

- Becker's disease, 646
- benzene, 828
 - molecular dynamics simulation, 828
- benzophenone, 313
- Biacore, 320, 323, 477–479, 483, 490, 495, 496, 500
 - data, 482, 483
 - flow cell, 481
 - RaPID Plot, 499
- bilayer lipid membrane, 46, 75–77
- Binnig, G., 402
- Bio-Beads, 75
- bio-optode, 330
- bio-optoelectronic sensor, 92, 93
- bioactivity, 1045
- bioavailability, 157
- BioCD, 958
- biochip, 78, 92, 307, 872, 900–905, 965–992
 - fluorescence, 1144
 - parallel, 965
 - series, 966
- biocompatibility, 758, 782, 1007, 1009, 1033, 1043
 - of biomaterials, 1148
 - of implant, 1148
 - of inorganic nanoparticles, 95
 - of nanoMag particle, 1128
 - of nanoTherapeutics, 1123
 - of quantum dots, 273
- biocomputing, 920
 - cluster program, 920
- bioconjugation, 1109
- biodistribution, 156, 1103, 1123
 - of drugs, 292, 293, 297
 - of oligonucleotides, 242
- bioelectronic interfacing, 29, 666, 1148
- biofilm formation, 562
- bioinformatics, 776
- biointegration, 1043–1062
 - applications, 1054–1058
 - definition, 1044
 - surface modification, 1045–1054
- bioluminescence, 263, 285–288, 296–297, 891, 892, 897–899, 1092
 - imaging, 298–303, 404
 - commercial systems, 298
- biomacromolecule, 599, 607
 - MS analysis, 621
- biomaterial, 1040, 1043–1045
 - definition, 1043
 - functionalised, 1047–1052
- biomembrane force probe, 393, 397–400
- biomimetics, 192, 719, 771, 872, 1044, 1052
- biopolymer, 1033
- biopsy, 775
- bioreceptor, 872–877
 - artificial, 874–877
 - immobilisation, 877–880
 - by adsorption, 877
 - by confinement, 878
 - by covalent bonding, 879–880
 - by crosslinking, 878
 - by inclusion, 878
 - natural, 873–874
- biosafety, 1075
- biosensor, 77, 86, 89, 92, 477, 667, 697, 871–908, 1027
 - affinity, 895–897
 - antibody, 873
 - array, 92
 - bioluminescent, 899
 - biomimetic, 872
 - chemiluminescent, 899
 - electrochemical detection, 880–887
 - electrochemiluminescent, 899
 - enzyme, 873
 - fibre optic, 891–899
 - enzyme, 894–895
 - experimental setup, 893
 - for DNA detection, 896
 - general arrangement, 871
 - interferometric, 525
 - lectin, 895
 - mass transducer, 887–889
 - mechanical, 525
 - microbial, 874
 - optical, 525–537
 - piezoelectric crystal, 889
 - SPR, 320–323
 - vs ion channel, 991
 - whole-cell, 874
 - with electrical detection, 525
 - with heat detection, 525
- bioterrorism, 312, 783, 1001
- biotin, 63, 74, 332, 333, 399, 406, 494, 520, 944, 955

- birefringence, 513
- bismuth oxychloride, 1149
- black lipid membrane, 75, 76
- blastocyst, 1089
- blastocyte, 285
- BLM, 46, 75–77
- blood, 782, 938, 965, 1056, 1057, 1107, 1146
 - clearance, 156
 - compartment, 154, 158, 163
 - complement, 156, 496
 - plasma, 1031
- blood–brain barrier, 155, 158, 1102, 1109, 1155
- bodipy, 265, 266
- Boltzmann distribution, 750
- bone
 - morphogenetic protein, 1055
 - regeneration, 1143
 - surgery, 1045
 - tissue, 1054–1056
 - engineering, 1054
- BOP, 1130
- Born–Oppenheimer approximation, 805
- bovine serum albumin, 211, 521, 629, 876
- Bragg peak, 1021, 1022
- brain, 605, 641, 666, 1071, 1079, 1082, 1083, 1086, 1087
 - chip implantation, 1156
 - disorder, 776
 - mouse, 611
 - rat, 665
 - substantia nigra, 1086
 - topography, 307
- BRET, 290
- Brewster angle, 557, 559
 - dependence on film thickness, 557
 - index dependence, 556, 557
- Brewster angle microscopy, 556–561
- Brewster image, 558
 - of cholesterol layer, 560
- Brownian motion, 339, 342, 349–351, 436, 448, 450, 454–456, 699, 707, 709
- Brownian ratchet model, 188
- Bruggeman model, 508
- bubble baby, 1081
- buccal cavity, 298
- bullock, 1075
 - transgenic, 1088
- burst effect, 1105
- cadherin, 85, 394, 395, 583, 584
- cadmium arachidate, 544, 549
- cadmium selenide, 1144
- caffeine, 231
- calcitonin, 380
- calcium bridge, 661
- calcium phosphate, 561
- calcium probe, 356
- calibration, 787
 - curve, 787
 - regression band, 787
- calmodulin, 202, 204, 644
- cancer, 25, 103, 120, 129, 153, 164, 244, 303, 500, 645, 647–649, 689, 770, 1081, 1082, 1121
- bladder, 120, 1121, 1127
- bowel, 1121
- breast, 647, 648, 769, 912, 1121, 1125, 1128, 1134
- buccal epidermis, 1133
- colon, 648
- definition, 1121
- five-year mortality rate, 1121
- kidney, 1121
- lesions, 160
- liver, 862, 1121, 1130
- lung, 300, 1121
- mouth, 1121
- neck of womb, 1121, 1133
- oesophagus, 1121, 1127
- ovary, 1121, 1133
- pancreas, 1122
- prostate, 163, 647, 1130
- research, 283, 306
- skin, 1127
- stomach, 1121, 1127, 1130
- testicle, 1121
- therapy, 1121–1141
- thyroid, 1121
- treatment, 167
- cancerogenesis, 19
- cancerology, 862, 924, 1113
- capacitive coupling, 986
- capacitive current, 669

- capillary electrochromatography, 604, 624
- capillary electrophoresis, 604, 624, 721, 744, 764, 770, 847
- capillary isoelectric focusing, 625, 628
- capping protein, 182, 190
- capsaicin, 645
- capsid, 1072
 - protein, 1076
- capture probe, 308, 313, 332
- carbon dioxide, 884
- carbon nanotube, 388, 522, 555, 697, 715
- carbonic anhydrase, 498, 605
- carboxylic acid, 138, 139, 142
- carboxypeptidase, 595
- carcinogenic substance, 1121
- carcinoma, 1082
- cardiac tissue, 666
- cardiology, 292
- cardiopathy, 645, 646
- cardiovascular disease, 689, 1121, 1155
- cardiovascular system, 154
- caspase, 930
- cat, 1075
- catalase, 689
- catecholamine, 684–688
- CCD camera, 322
- CDI, 1052
- cDNA, *see* DNA, complementary
- cDNA microarray, 915
- cell, 101, 104, 640
 - activity, 356
 - adhesion, 31, 107, 394, 400, 442, 443, 561, 583, 645, 970, 1033, 1048, 1049, 1056
 - adhesion force, 386, 400–401
 - AFM imaging, 384–387
 - architecture, 173
 - biochip, 965–992
 - chip, 62
 - chromaffin, 684, 686–688
 - compartmentalisation, 29
 - current, 974–976
 - cycle, 647, 927
 - division, 18, 110, 114, 433, 929
 - endothelial, 155, 689, 1057, 1111
 - engineering, 1043–1045
 - environmental stimuli, 969
 - epithelial, 646, 1033
 - eukaryotic, 499, 640, 1069
 - functions, 644
 - fusion, 968
 - growth, 275
 - growth control, 31
 - HeLa, 443–445, 1133
 - host, 283
 - immobilised, 966
 - KB, 1134–1136
 - labelling, 275
 - lysis, 770
 - manipulation by optical tweezers, 445
 - mast, 688
 - mechanotransduction, 442
 - membrane, 651, 713
 - microarray, 928–930, 965
 - microculture on chip, 969–972
 - micromanipulation, 967–972
 - migration, 175, 387, 442, 443, 645
 - motility, 172, 179–183, 352, 969
 - nucleus, 640
 - oxidative stress, 688–694
 - polarity, 174, 197
 - proliferation, 20, 647, 929, 1033, 1121
 - selection, 654
 - separation, 968
 - sorting, 768–770, 965, 966, 968
 - fluorine-activated, 769
 - spreading, 561, 1049
 - stress, 284
 - targeting, 119, 121
 - therapy, 1044
 - transgenic, 969
- cell on chip, 905
- cellulose, 878
- central nervous system, 1078, 1086, 1102
- ceramide, 34
- CFTR, 646, 654, 1079
- CGH microarray, 923–925
- channelopathy, 639, 645, 666, 972
- channelopsin, 642
- channelrhodopsin, 642
- chaperonin, 379
- Charcot–Marie–Tooth disorder, 198
- charybdotoxin, 647
- chemical library, 973, 979
- chemical sensor, fibre optic, 892

- chemical vapour deposition, 715
- chemiluminescence, 93, 263, 891, 892, 897–899, 904, 1007
- chemiosmotic theory, 208
- chemisorption, 132
 - of water, 134
- chemotherapy, 865, 1123, 1125
- CHEMS, 1106
- cherry-picking, 974
- chicken, 1089
 - genome, 1089
- chili pepper, 645
- chimeric animal, 285
- Chinese hamster ovary, 499, 654, 662
- ChIP on chip, 913, 925–928
- chiral molecule, 1017
- chirality, 236
- chitin, 878
- chitosan, 1028, 1033
- chloroplast, 206, 217
 - membrane, 73
- cholera toxin, 87, 334, 381
- cholesterol, 35, 37, 45, 116, 243, 381, 382, 662, 884
 - as stabiliser, 1102
 - detection, 899
 - layer, 560
 - oxidase, 884
 - PEGylated, 1108
- choline, 93
 - detection, 899
- chondrocyte, 1033
- chondrosarcoma, 1028, 1033
- chromatin, 17, 18, 927
 - compaction, 23, 25
 - immunoprecipitation, 926
- chromatography, 622–624, 703
 - affinity, 1004
 - gel exclusion, 1103
 - HPLC, 623, 625–627
 - ion exchange, 622
 - liquid-phase, 604, 625, 1005
 - online integration, 629
 - peptide, 780
 - reverse-phase liquid, 623
 - steric exclusion, 695
- chromosomal integration, 1071, 1072, 1092
- chromosome, 17, 18, 101, 103, 109, 110, 114, 173, 197, 433, 463, 719, 923, 1073, 1079
 - bacterial artificial, 913, 924, 930
 - X, 18
- chronoamperometry, 675
 - double potential step, 676
 - single potential step, 676
- CIEF, 625, 628
- cilium, 197
- clean room, 1152
- clinical trial, 1081, 1085
- CLIO, 164
- cloning, 841, 1071, 1078, 1086
- CMOS, 322, 328
 - image sensor, 329
 - multispot chip, 330
- coelenteramide, 290
- coelenterazine, 290
- cofilin, 182
- colimycin, 1045
- collagen, 379, 499, 878, 938, 946, 1028
- collodion, 878
- colloid, 390
- colloidal stability, 130, 137, 142
- combinatorial chemistry, 763
- comparative genome hybridisation, 913, 923
- compound library, 973
- compression moulding, 757
- concanavalin, 497, 895
- concentration measurement, 871
 - amperometric, 678–682
 - FCS, 349
 - SPR, 495, 500
- confocal microscopy, 259–261, 316–318, 551, 610, 780
 - fluorescence, 317
 - image restoration, 791
- conformer, 4, 9
- congenital aural atresia, 925
- contamination, 782, 1150
 - bacterial, 1150
- contrast agent, 160, 161, 1093
 - intelligent, 1092
 - magnetic susceptibility, 161–164
 - paramagnetic, 160
- core–corona morphology, 132, 136, 139, 161, 164

- core-shell morphology, 132, 136, 140, 142, 143, 161, 167
- cosmetics, 47, 1045, 1143, 1149–1150
- Cottrell relation, 677
- coumarin, 265
- counterelectrode, 753
- counterion, 125, 130, 699, 720, 750, 1019, 1020, 1103
 - in pore, 700
- CpG island, 927
- CpG sequence, 20–22
 - flexibility, 22
 - methylation of, 19, 21
 - stability, 22
- CRE sequence, 23
- creatininase, 884
- creatinine, 884
- cryo-AFM, 379
- cryo-TEM, 47, 48, 94, 95
- CTAB, 122
- cyanine, 261, 266, 848
 - quantum yield, 267
- cyclic olefin copolymer, 757
- cyclodextrin, 1034, 1035, 1100
- cyclotron resonance, 621
- cysteine, 124, 202, 210, 493, 1052
- cystic fibrosis, 645, 646, 654, 1079
- cytochrome c, 75, 77, 261, 605
- cytochrome P450 2D6, 869
- cytokinesis, 174, 175
- cytomegalovirus, 282
- cytoplasm, 110, 114
 - calcium concentration, 684
- cytosine, 9, 18–21, 23, 101
 - methylation, 927
- cytoskeleton, 173, 174, 258, 397, 442, 444, 446, 642, 1049
 - AFM observation, 379, 382
- cytosol, 110, 114, 335, 349, 644, 662, 664, 665
- dalton, 596, 938
- Dalton, J., 938
- dark-field microscopy, 199
- Darwin, C., 224
- Darwinian theory, 224, 225
- data analysis, 776, 795–798
 - measurement selection, 795
 - transcriptome, 919
- data bank, 785
- data extraction, 790–794
 - differential methods, 792
 - integral methods, 792
 - inverse problem, 793–794
 - regularisation, 794
 - systems approach, 791–792
- data processing, 775–798
 - classification methods, 779, 780, 796–798
 - connectionist approach, 798
 - detection, 781
 - false negative, 781
 - false positive, 781
 - geometrical analysis, 796
 - image restoration, 793
 - imaging, 780–781, 785
 - imaging system, 788, 790
 - learning process, 797
 - observation, 781
 - on-board, 783
 - preprocessing, 782, 786–787
 - quantification, 781
 - sample, 784
 - statistical analysis, 794, 796
 - structural approach, 798
 - thresholding, 779, 781
 - tomography, 780, 786, 791, 793
- data structure, 784–785
- data system, 776–777
 - architecture, 777
- database, 629, 785
- DCC, 1052
- de Broglie wavelength, 805
- de Gennes, P.-G., 700, 704, 709
- Debye length, 671, 699, 712, 720, 751, 1019
- Debye–Waller factor, 579
- Debye–Hückel approximation, 750, 823
- Debye–Hückel electrolyte theory, 699
- dehydrogenase electrode, 883
- deletion, 101, 226, 923
- dendrimer, 110, 111, 1100
- dendrite, 986
- dental implant, 1056
- Derjaguin approximation, 391, 393
- dermatology, 298, 307
- dermis, 689
- desmin, 930

- desmosome, 86
- deuterium, 584
- dextran, 139, 157, 161, 162, 164, 277, 323, 410, 479, 895, 1033
 - coating, 1108
 - conformational change, 411
 - force–extension curve, 411
 - matrix, 482
 - sulfate, 1107
- dextrose, 322, 878
- DGDG, 35
- diabetes, 379, 628, 1090, 1097
- diagnosis, 154, 159–164, 312, 323, 719, 781, 864, 912, 937, 956, 999, 1000, 1111, 1143
 - in vitro, 1144
 - in vivo, 1144
 - point-of-care system, 1001
 - veterinary, 276
- diagnostic marker, 952
- diagnostics, 1013, 1158
- diamine, 1051
- dielectric, 542, 549
 - constant, 538
 - substrate, 546–547, 557, 559
 - transmission spectrum, 551
- dielectric constant effect, 700–701
- dielectrophoresis, 767, 968–969, 980, 982
 - negative, 767, 967, 968, 971
 - on-chip, 976
 - positive, 767, 968
- dielectrophoretic
 - effect, 767, 770
 - force, 968
- diffraction, 529
 - order of, 529
- diffuse layer, 720
- diffusion
 - effects, 782
 - laws of, 675
 - layer, 678
- digital processing, 776
- dioxin, 621
- dip-pen nanolithography, 946, 952
- dislocation, 143
- disulfide bridge, 123, 595, 873, 879, 938
- diversity, 226
- DLVO theory, 390
- DMPC, 74
- DMSO, 492
- DNA, 3–25, 101, 226
 - aptamer, 231
 - B conformation, 4, 9
 - B–Z junction, 10, 11
 - B-to-Z transition, 10, 12
 - base pair, 3, 10, 13, 21, 22, 102, 224, 227
 - biochip, 1150
 - biodegradability, 104
 - biotinylated, 941
 - circular, 12, 15, 1075
 - circulation, 106
 - compaction, 17, 18, 110, 118, 433
 - complementary, 902, 913, 929, 966
 - condensation, 105–106, 122
 - confined in a pore, 705
 - conformation, 4
 - roll, 5
 - tilt, 5
 - twist, 5
 - contour length, 433
 - cruciform, 4, 15, 16
 - detection, 841, 896, 1146
 - double helix, 4–7, 102, 227–228, 458, 901
 - helical parameters, 6, 9
 - elasticity, 406, 408
 - entry into nucleus, 109, 114
 - eukaryotic, 13
 - exogenous, 102
 - extraction, 721, 743, 770
 - hairpin structure, 230
 - Hoogsteen pairing, 228
 - hybridisation, 280, 561, 919, 976, 1150
 - imaging, 429, 430
 - intercalation agent, 13
 - kissing complex, 230
 - loop formed by GalR, 461–462
 - major groove, 22, 23, 227
 - manipulation by AFM, 427
 - methylation of, 17–25
 - microarray, 316, 903–904, 911–930, 966, 976
 - applications, 919
 - definition, 912
 - migration, 723

- minor groove, 13, 227, 848
- monomolecular condensation, 122
- mutation, 846
- nick, 458
- nick detection, 460
- nucleosomal, 13, 14
- observation by AFM, 427–434
- observation by ellipsometry, 521
- observation by magnetic tweezers, 448, 461–465
- palindromic sequence, 15, 16
- persistence length, 409
- polymerase, 841, 847, 849, 850, 855
- prokaryotic, 21
- repair complex, 688
- repair enzyme, 925
- sensor, 889
- separation, 720, 721, 726, 764–768
 - diffusion, 729
 - entropic, 730
- sequence recognition, 278
- sequencing, 312, 703, 718, 719, 731, 919, 976
- stability of helix, 3
- structure, 102
- structure by AFM, 378
- sugar–phosphate backbone, 5, 8, 105, 227
- supercoiled, 12–16, 229–230
 - linking number, 13
 - twist, 13
 - writhe, 13
- testing, 153
- triple helix, 4, 228, 497
- twisting with magnetic tweezers, 458–460
- ultrafast sequencing, 697
- unknotting, 463
- wobble pairing, 228, 230
- wormlike chain model, 455
- Z conformation, 7–12
- Z-to-B transition, 10
- DNA chip, 62, 310, 520, 775, 865, 912
 - multispot, 330
- DNA/PEI complex, 113
- dog, 1079
- DOGS, 117, 118
- dopamine, 684
- DOPC, 60, 63, 381
- DOPE, 109, 118, 120, 1106
- DOPS, 60, 62, 63
- DOTAP, 116, 117
- double-layer model, 130, 669–672, 680, 699, 720, 781
- doxorubicin, 1111
 - encapsulated, 1114
- doxycycline, 1090, 1091
- DPD, 866–868
- DPPC, 381
- DRIE, 629, 1010
- Drosophila, 4, 242, 1083
- drug
 - administration, 1045
 - biotransformation, 866, 869
 - carrying, 1159
 - delivery, 95, 277, 1045, 1046, 1097–1098, 1130, 1145, 1146
 - design, 495, 497, 498, 775, 972
 - discovery, 972, 973
 - side-effects, 988, 1045
 - targeting, 30, 47, 91, 1108–1111
 - active, 158
 - passive, 156, 167, 1107, 1123
- DSC, 1050
- DSPC, 585
- dye
 - indicator, 892
 - organic, 149
- dynein, 173, 197
- dystrophin, 1079
- EDC, 1052
- effective medium theory, 506–508
- effector, 7, 22, 25
- Eigler, R., 402
- elastomer, 417
- electrical detection, 328
 - using nanoparticles, 328
- electrochemical
 - detection, 328, 871, 880–887
 - gradient, 640, 650
 - sensor, 880
- electrochemiluminescence, 93, 904
- electrochemistry, 668
- electrode–electrolyte interface, 669–671
 - capacitance, 671–672
 - double layer, 669, 670

- Gouy–Chapman layer, 671
- inner Helmholtz plane, 670
- outer Helmholtz plane, 670
- redox reaction, 672
- electrolyte, 720
- electromagnetic wave, 501–503, 529, 531, 537–540, 557
 - polarisation, 541
 - reflection, 504–508
- electron beam lithography, 697, 715, 717, 755, 757
- electron bombardment, 598
- electron diffraction, 70, 761
- electron–hole pair, 151, 270
- electroosmosis, 699, 723, 750–752, 760, 769
- electroosmotic flow, 624, 699, 750
- electroosmotic pump, 984, 985
- electrophoresis, 112, 624–625, 627–628, 682, 719–731, 764, 1103
 - capillary, 604, 624, 721, 744, 764, 770, 847
 - different types, 721
 - gel, 721, 731, 845
 - on-chip, 976
 - pulsed-field, 765, 768
- electrophoretic
 - current, 669
 - mobility, 720
- electrophysiology, 639, 643, 652, 682, 971, 976
 - and fluorescence, 989
 - high-throughput, 666, 987, 989
- electroporation, 284, 967, 969, 1069
- electrospray ionisation, 599–606, 620, 625, 627, 945
 - basic principle, 601–602
 - limitations, 604
 - multiply charged species, 603
 - sample preparation, 603
 - sensitivity, 604–606
 - Taylor cone, 602, 606
- electrostatic repulsion, 130, 131, 140
- electrowetting, 752, 753
- ellipsometer
 - imaging, 561
 - laser, 510
 - nulling, 511
 - optical setup, 510
- phase modulation, 512–515
 - analyser, 513
 - detector, 514
 - monochromator, 513
 - optomechanical setup, 513
 - photoelastic modulator, 512
 - polariser, 512
 - rotating element, 511
 - spectroscopic, 511
- ellipsometric angles, 509, 512, 515, 517
- ellipsometry, 58, 70, 90, 320, 389, 500–524, 535
 - applications, 520–523
 - basic principle, 508–509
 - data processing, 515–519
 - instrumentation, 510–515
 - of buried interface, 561
 - of thin films, 509
 - strong points, 510
- ellipticine, 13
- embolisation, 1098
- enantiomer, 236
- encapsulation, 1070
 - sequence, 1073, 1076
- encapsulation, 1098, 1101, 1104, 1149, 1159
 - of fluorophores, 277
 - of magnetic particles, 1109
 - of nanoparticle, 139–142
 - of quantum dots, 272, 274
 - with silica, 1132
- endocytosis, 68, 108, 120, 121, 162, 1110
 - of nanoplex, 127
- endogenous probe, 281–291
- endometrium, 648
- endoplasmic reticulum, 29, 110, 114, 197, 610, 640, 646, 684
- endoscopy, 298, 307, 1144
- endosome, 107, 108, 111, 1106, 1110
- endothelialisation, 1057
- endothelium, 155, 1057
- ENFET, 885–887
- enhanced permeability and retention, 1108, 1130
- entropic recoil, 730
- entropy, 1019
- environment, 263, 276, 312, 783, 907, 999, 1013, 1143, 1151, 1155, 1157
- MS analysis, 621

- enzyme, 14, 78, 85, 103, 104, 180, 206, 223, 282, 283, 302, 328, 444, 459, 461, 496, 644, 688, 690, 865, 866, 1035
 active site, 873
 ADA, 1080
 alkaline phosphatase, 898
 as catalyst, 873
 assay, 953
 conformational change, 497
 digestion, 595, 654
 DNA repair, 925
 DPD, 866–868
 electrode, 871, 880–885
 fibre optic biosensor, 894–895
 for biosensor, 872
 GalR, 461–462
 glucose oxidase, 871
 hepatic, 869
 hydrogen peroxide production, 884
 immobilised, 94, 498, 871
 in polyelectrolyte multilayer, 1035
 metabolic, 654
 NADPH oxidase, 693
 NO synthase, 693
 oxygen production, 884
 QCM observation, 561
 restriction, 720
 substrates, 873
 thermistor, 889–891
 enzyme–substrate complex, 874
 enzymology, 85, 208, 216
 epifluorescence, 336
 microscopy, 201, 259, 314–315, 554
 epigenetic regulation, 18, 25
 epigenetics, 927
 epitaxial growth, 143
 epithelium, 1102
 digestive, 1097
 epitope, 952
 EPR effect, 1108, 1110, 1130
 ergodic system, 805
 ergodicity, 806
 ergosterol, 662
 erythromycin, 1045
 ESCA, 70
 ESI, *see* electrospray ionisation
 ethics, 1155
 ethidium bromide, 13, 118, 845, 847
 euchromatin, 18, 20
 eukaryote, 4, 17, 21, 42, 110, 173, 380, 382–385, 499, 640, 929, 982, 1069, 1078, 1083
 DNA, 13
 European Synchrotron Radiation Facility, 583
 evanescent wave, 414, 442, 525, 536
 evanescent wave fluorescence spectroscopy, 685
 evolution, 224
 molecular, 225–226
 Ewald–Kornfeld summation, 804
 exciton, 142
 exeresis, 303
 exocytosis, 384, 665
 of neurotransmitters, 684–688
 exogenous probe, 263–281
 exon, 866
 expression microarray, 928
 FAB, *see* fast atom bombardment
 facilitated diffusion, 640
 FACS, 769
 FAD, *see* flavine adenine dinucleotide
 faradaic electrochemical detection, 682
 faradaic electrochemistry, 668–678
 faradaic process, 672–675
 electron transfer, 673–674
 equivalent circuit, 672
 impedance, 672
 mass transport, 674–675
 Faraday constant, 672
 Faraday's law, 693
 Faraday, M., 669
 fast atom bombardment, 598–600, 609
 FCCS, 351
 FCS, 58, 334, 341–351, 388
 accuracy, 348
 concentration measurement, 349
 correlator, 347
 cross-correlation, 350
 experimental setup, 345
 spatiotemporal resolution, 346
 time resolution, 347
 femtosecond laser, 304
 Fenton reaction, 689
 ferrocene, 328, 883
 ferrofluid, 134, 135, 1128

- fibrinogen, 535
 - in polyelectrolyte multilayer, 1029
- fibroblast, 258, 442, 662, 689, 1028, 1033, 1082
 - manipulation, 446
 - migration, 969
 - monkey, 382
 - oxidative stress, 690, 691
- fibroin, 878
- fibronectin, 156, 443–446, 970, 1107
- Fick's second law, 676, 677
- field-effect transistor, 872, 885, 904, 986, 1007
- field-flow fractionation, 749
- filopodium, 174
- FISH, 257, 924, 929
- FITC, *see* fluorescein isothiocyanate
- flagellum, 197
- FLAP, 339
- flavine adenine dinucleotide, 881
- flavivirus, 862
- FLIP, 339
- flocculation, 95, 135, 141
- flow cytometry, 654, 847
- flow injection analysis, 899
- fluctuation–dissipation theorem, 820, 822
- fluid mechanics, 745, 746
 - Newtonian fluid, 746
- fluid particle, 745
 - equation of motion, 745
- fluorescein, 265–267, 278, 851, 1025, 1026, 1039
 - isothiocyanate, 895
- fluorescence, 148, 149, 253, 262, 292, 296–297, 894
 - confocal microscopy, 317
 - cube, 256
 - detection, 949
 - enhancement, 324–326
 - imaging, 150, 275, 298–307, 314, 441, 779
 - commercial systems, 298
 - time-resolved, 303–307
 - intensity, 893
 - intensity profile, 443
 - label, 526
 - labelling, 257, 334
 - lifetime, 267, 304, 305, 358
 - microarray, 312
 - microscopy, 176, 199, 253–262, 388, 404, 610, 747, 1133
 - multiple labelling, 257
 - quantum yield, 257, 266
 - quenching, 849, 850, 853
 - reflectance imaging, 298, 299, 302
 - spectroscopy, 320
 - videomicroscopy, 766
- fluorescence resonance energy transfer, 867
- fluorescent protein, 147
- fluorophore, 147, 148, 253, 301, 305, 314, 354, 442, 444, 779, 849, 850, 853, 857, 930, 949, 950, 953
 - acceptor, 849, 851
 - ASPISH, 1133
 - commercially available, 266
 - donor, 849, 851
 - encapsulated, 277
 - excitation, 261
 - HPTS, 893
 - in PCR, 847
 - organic, 263, 265–267, 334
 - perylene dibutyrate, 893
 - photostability, 347
 - pyrene butyric acid, 893
- folic acid, 119, 121, 159
- food industry, 263, 495, 500, 999, 1013, 1143, 1150
- food safety, 276, 1150
- formaldehyde, 888, 889
 - dehydrogenase, 888
- formin, 180, 181
- Fourier transform, 824
- Fourier transform principle, 621
- fractal dimension, 703
- FRAP, 58, 176, 257, 320, 334–341
 - data, 337
 - optical setup, 336–337
- free radical, 1126–1128, 1133, 1149
- Fresnel
 - diagram, 556
 - interface, 537
 - reflection coefficient, 504, 579, 582, 583
 - relations, 546, 559, 579
- FRET, 210, 258, 289, 334, 355–360, 444, 851

- acceptor, 356
- donor, 356–358
- efficiency, 356, 357, 360
- fluorescence intensity, 356
- fluorescence lifetime method, 358
- intermolecular, 357
- measurement methods, 356
- theory, 355
- fringes of equal chromatic order, 392
- FRM1 gene, 21
- fullerene, 1149
- functionalisation, 321
 - by biotin, 74
 - by drugs, 1035
 - by nanoparticles, 1055
 - by peptides, 1033
 - of biomaterial, 1047–1052
 - of labs on chips, 1010
 - of LB films, 78–85
 - of lipid, 72
 - of nanoparticles, 95, 129–168, 269, 276, 277, 279, 1047
 - of polyelectrolyte multilayers, 1027–1036
 - of quantum dots, 272
- funding of nanoscience, 1153
- gadolinium, 160, 161
 - chloride, 1107
- galactose, 461
- galactosidase, 461, 884
- galactoside, 119
- gamete, 1069
- ganglioside, 1108
- GDNF, 1080, 1082, 1083
- GDP, 444
- gel, 527
 - artificial, 765, 770
 - polyacrylamide, 871
- gel electrophoresis, 721, 731, 845
- gelatin, 1028
- gene, 101, 104, 282
 - as blueprint, 938
 - deletion, 923
 - duplication, 923
 - env, 1072
 - exogenous, 110
 - exon, 866
 - expression, 3, 7, 11, 16–18, 20, 283, 292, 497, 923, 966, 976
 - function, 240, 312, 920, 922
 - gag, 1072
 - gain of function, 928
 - heterologous expression, 654
 - in Darwinian theory, 225
 - inhibition, 1082, 1085
 - loss of function, 928
 - marker, 897
 - mutation, 647, 781, 866, 1080
 - nef, 1074
 - overexpression, 1080, 1082, 1085
 - pharmacology, 101
 - pol, 1072
 - promoter, 11, 16, 20, 283, 461, 497, 929, 969, 1071, 1074
 - promoter region, 927
 - regulation, 20, 282, 284
 - regulation network, 926
 - rep, 1079
 - reporter, 113, 263, 281–291, 1083, 1092
 - silencing, 929
 - suicide, 120
 - tat, 1074
 - therapy, 30, 92, 102, 103, 499, 646, 1069, 1077, 1079–1082, 1159
 - transfer, 105–110, 118, 120–121, 965, 969
 - germline, 1069
 - in vivo, 1069–1070
 - somatic, 1069
 - vif, 1074
 - vpr, 1074
 - vpu, 1074
- genetic
 - code, 3
 - correction, 102
 - disorder, 719
 - engineering, 1071
 - information, 3, 101, 103, 104
 - linkage, 919
 - predisposition to cancer, 1121
- genetically modified organism, 903, 1001, 1157
- genome, 4, 312, 785, 900, 911, 953, 1076
 - AAV, 1079
 - adenoviral, 1077

- chicken, 1089
- fully sequenced, 196
- functional exploration, 929
- human, 3, 17, 103, 719, 775, 902
- human mitochondrial, 919
- virus, 862–864, 1070
- genomics, 312, 667, 775, 911
 - functional, 1069
- genotyping, 846, 848, 852, 853, 867
- gentamicin, 1045, 1055
 - sulfate, 1055
- giant magnetoresistance, 1000
- Gibbs monolayer, 557
- Gibbs surface excess, 527
- glass, 320, 321, 980
 - as AFM substrate, 378
 - capillary, 605
 - chip, 980
 - for lab-on-a-chip, 1010, 1012
 - for microfluidics, 756
 - porous, 878
 - substrate, 313, 911
- glioblastoma, 1082, 1122
- gliofibrillar acid protein, 1109
- gluconolactone, 881
- glucosamylase, 884, 1029
- glucose, 895
 - assay, 871, 881
 - detection, 899
 - electrode, 881–884
 - isomerase, 1029
 - oxidase, 871, 881, 883, 886, 897, 1029, 1035, 1036
- glutamate oxidase, 884
- glutaraldehyde, 384, 878, 1031
- glyceroglycolipid, 34
- glycerol, 600, 607
- glycerophospholipid, 31
- glycine, 873, 884
- glycobiology, 498
- glycolipid, 498
- glycoprotein, 328, 879
 - envelope, 1072, 1076
- glycosidic bond, 34
- glycosphingolipid, 34
- goat, 1060
- GOD, *see* glucose oxidase
- gold, 313, 573, 878
 - bead, 211, 388
 - film, 320–323, 479, 521, 555
 - nanoparticle, 145–146, 151, 153, 154, 167, 279, 328, 332
 - particle, 555
 - substrate, 73, 527, 544, 548, 969
 - wire, 753
- Golgi apparatus, 197
- green fluorescent protein, 63, 176, 261, 263, 282, 283, 288–290, 348, 352, 356, 443, 654, 966, 969
 - molecular structure, 289
- GroEL, 425–426
- GroES–GroEL complex, 425, 426
- GST, 490, 941
- GTP, 173, 444, 642
- guanidinium, 231
 - molecular dynamics simulation, 825
- guanine, 5, 11
- guanine–thymine mismatch, 21
- guanosine, 101, 173
- gyrase, 14
- Haber–Weiss reaction, 689
- Hamaker constant, 130, 415
- harmine, 1112
- harmonic oscillator, 563
- health care, 1143, 1155
 - personalisation, 776
- heart disease, 153, 776
- helicase, 14, 462
- Helmholtz free energy, 830
- Helmholtz layer, 670
- hemocompatibility, 1057
- hemoglobin, 155, 535, 938
 - absorption spectrum, 295
 - MS analysis, 621
- hemostasis, 1057
- HeNe laser, 318, 531
- heparin, 1031
 - coating, 1108
- hepatic system, 1148
- hepatitis, 862
- hepatocyte, 119
- herbicide, 1151
- hERG, 645, 648
- heritability, 226
- herpes simplex virus, 1071, 1090
- heterochromatin, 18, 20
 - complex, 927

- hexagonally packed intermediate layer, 412, 413
- high throughput analysis, 782
- histidine, 940, 946
- histidine tag, 63, 72, 73, 75, 87, 88, 210, 394
- histogram equalisation, 788
- histone, 14, 17, 20, 110, 433, 925, 927
 - acetylase, 20
 - acetylation of, 20
 - deacetylase, 20
- HIV, 229, 231, 241, 289, 919, 956, 1074
- Hodgkin–Huxley model, 649–652
- holography, 780
- HOPG, 378
- hormone, 684, 938, 1033, 1148
 - detection, 312, 327, 500
 - growth, 938
- horse, 1075
 - lentiviral injection, 1089
- horse radish peroxidase, 328, 900
- HPLC, 623, 625–627
- HPRT, 1080
- HPTS, 893
- HRTEM, 271
- HTRF, 268, 269
- human immunodeficiency virus, *see* HIV
- human serum albumin, 535, 888, 1028–1030
- huntingtin, 1082, 1091
- Huntington's disease, 1082, 1086, 1089, 1091
- hyaluronic acid, 536, 575, 1025, 1026, 1028
- hybridisation, 779
 - array, 778
 - non-specific, 782
- hydration mantle, 643
- hydrogel, 944
- hydrogen bond, 21, 22, 102, 134, 227, 231, 661, 873, 876, 877, 938
 - in water, 50, 230
- hydrogen peroxide, 689, 881, 882, 884, 897
- hydrophilic/hydrophobic balance, 38
- hydrophobic effect, 38, 873, 877, 938, 1107
- hyperkalemic paralysis, 646
- hypermethylation, 19–21
- hyperosmotic shock, 114
- hyperthermia, 164, 165, 1106
- hypokalemic paralysis, 646
- hypomethylation, 21
- iberiotoxin, 647
- iFRAP, 339
- imaging, 1061
 - medical, 1158
- immune response, 956
 - enhancing, 1112
- immune system, 153, 156, 175, 499, 688, 689, 874, 1077
- immunoassay, 257, 310, 311, 889, 897, 950, 953, 1001
- immunocytochemistry, 151, 610
- immunofluorescence, 929
- immunogenicity, 495, 1071, 1078, 1123
- immunoglobulin, 84, 156, 408, 409, 499, 946, 1107
 - in polyelectrolyte multilayer, 1029
- immunoliposome
 - stealth, 1109, 1110
- immunology, 1004
- immunonanoparticle, 1109
- immunotherapy, 1123
- impedance spectroscopy, 781, 785, 971
- implant, 1044, 1148
- implantable
 - device, 1045
 - material, 1043, 1044
- implanted protocol, 1002
- indium tin oxide, 522, 542
- industrialisation, 1152
- infectious disease, 1113
- inflammation, 645, 646
- infrared laser, 317
- infrared spectroscopy, 70, 541–551
- injection moulding, 754, 757
- insulin, 938, 1097
 - LSIMS analysis, 600
 - PDMS analysis, 600
- integrase, 1072, 1076
- integration density, 971
- integrin, 442, 444, 499, 1049, 1111
- intercalation agent, 10, 118, 847
- interferon, 864, 1110
- interleukin, 1110

- interstitium, 163
- intracerebral injection, 1086
- intramacrophagic disease, 1107
- intramer, 241
- invertase, 884
- ion beam lithography, 697, 717
- ion bombardment, 968
- ion channel, 77, 639–667, 684, 699, 713, 969, 971–973, 984
 - acid-sensing, 645
 - calcium, 649
 - chloride, 641
 - eag, 648
 - endocellular, 642
 - endogenous, 655
 - endomembrane, 665
 - gating, 641, 643, 663
 - hERG, 647, 648, 974, 988
 - indirect methods, 666
 - inward-rectifying, 650
 - ligand-gated, 641
 - light-gated, 642
 - mechanically-gated, 641
 - observation, 649–666
 - outward-rectifying, 650
 - physiological role, 644–645
 - potassium, 641, 644, 647, 648, 652
 - outward-rectifying, 642
 - properties, 641–643
 - research, 974
 - run-down, 665
 - selectivity, 643
 - sodium, 644, 647, 652
 - voltage-gated, 642, 644, 652, 989
 - vs biosensor, 991
- ion current, 669
- ion detection, 263
- ion exchange chromatography, 622
- ion exchange resin, 878
- ion microscopy, 609–611
 - instrumentation, 610
 - sample preparation, 611
- ionic bond, 938
- ionic interaction, 873
- ionisation of biomolecules, 598–611
- IRRAS, 70, 541–547
 - spectrum, 544, 546
- ischemia, 103, 645
- ISFET, 885–887
- ITR sequence, 1073, 1077
- Jablonski diagram, 261, 262
- jellyfish, 288, 897
- JKR theory, 395
- Joule effect, 1007
- Kartaneger's syndrome, 198
- keratinocyte, 689
- kidney, 120, 628, 1123
 - cancer, 1121
- kinase, 498, 646, 947, 953
- kinesin, 173, 198, 446, 447
 - step size, 199, 200
 - structure, 204–206
- Kupffer cell, 156, 163
- L-lactate, 882–884
- lab-on-a-chip, 310, 731, 743, 775, 783, 999–1013
 - fabrication, 1009–1011
 - filtering, 1002
 - functionalisation, 1010
 - glass, 1006
 - liquid–liquid extraction, 1006
 - materials, 1012
 - measurement protocol, 1001
 - polymer technology, 1011
 - polymerase chain reaction, 1007
 - sample preparation, 1002–1007
 - separation by transport, 1003
 - silica, 1006
 - solid phase extraction, 1004–1006
 - surface treatment, 1010
 - transduction, 1007–1009
- lactate
 - dehydrogenase, 882, 883
 - detection, 899
 - oxidase, 883
- lactose, 884
- Ladd method, 804
- lambda phage, 696
- lamellipodium, 174–176, 179, 180, 182, 191, 352
 - artificial, 192
- Langevin equation, 436, 1136, 1137
- Langmuir isotherm, 52–55
- Langmuir model, 482

- Langmuir monolayer, 46, 50–57, 557, 560
 applications, 55–57
 insoluble, 50
 Langmuir trough, 51, 380
 Langmuir–Blodgett film, 46, 62–72, 522–523, 525, 544
 alternating layers, 68
 characterisation, 69–72
 functionalised, 78–85
 phospholipid, 65–68
 protein insertion, 82
 X type, 65
 Y type, 64–66
 Z type, 65
 Langmuir–Blodgett transfer, 57, 63–65, 93
 vertical, 63, 68, 84
 Langmuir–Schaefer transfer, 66, 68, 82
 lanthanide chelate, 264
 luminescence lifetime, 268
 luminescent, 267–269
 structure, 267
 Laplace's law, 397, 398, 748
 laser ablation, 754, 757, 1011
 laser desorption, 606, 607
 laser diode, 322
 latex bead, 768
 lattice mismatch, 143
 LCGreen, 848
 LCST, 167
 lecithin, 713
 lectin, 498
 biosensor, 895
 LEED, 70
 leishmaniasis, 1107, 1112
 lentiviral vector, 1070, 1074–1077, 1083, 1085, 1086, 1089
 production, 1075–1076
 regulated, 1091
 titer, 1076
 lentivirus, 1074
 Lesh–Nyhan syndrome, 1080
 leukemia, 911
 lymphoid, 648
 murine, 1072
 myelogenous, 20
 myeloid, 648
 LHRH, 1128, 1130, 1133, 1134
 ligand perfusion system, 985
 ligand–receptor system, 876
 light emission, 253
 light–matter interaction, 503–504
 light-emitting diode, 331, 1017
 Liouville operator, 808
 lipid, 31, 395, 498, 581, 584, 661
 amphipathic, 36
 bilayer, 30, 31, 41, 104, 117, 380, 381, 394, 401, 498, 528, 561, 640, 651, 696, 699, 713
 patterned, 72
 planar, 75–77
 supported, 46, 57, 72, 381
 biotinylated, 88
 cationic, 116
 drug conjugate, 1099
 for gene transfer, 116
 functionalised, 72
 labelling, 334
 layer, 558
 membrane, 31–45, 114, 713–714
 molecular dynamics simulation, 814, 832
 monolayer, 582–584
 PEGylated, 1109
 peroxidation, 689
 polymorphism, 39–43
 hexagonal phase, 40, 42–44
 lamellar phase, 40, 42, 44
 raft, 68, 90, 381
 recrystallisation, 1105
 shape, 44
 shape theory, 45
 solid, 1099
 vesicle, 397, 401, 561
 lipofection, 966
 lipophilic molecule, 1099, 1105, 1150
 lipoplex, 110, 115–120
 additives, 120
 characterisation, 117
 for gene transfer, 118, 120–121
 stability, 118
 structure, 115
 synthesis, 116
 lipopolyamine, 116
 liposome, 41, 42, 46–49, 129, 161, 498, 1099, 1101, 1148
 cationic, 1103

- deposition, 57
- elasticity, 184
- fusion, 284
- fusogenic, 1106
- gigaseal on, 982
- immobilised, 498
- pH-sensitive, 1106
- plasma half-life, 1109, 1114
- rupture mechanisms, 58
- stealth, 1108
- synthesis, 47
- thermosensitive, 1106
- liquid-phase chromatography, 604, 625
- Lissajous curve, 619
- listeria, 500
- lithography, 1003
- liver, 119, 120, 156, 162, 163, 866, 1079, 1106, 1107, 1109
 - cancer, 862, 1121, 1130
 - cirrhosis, 862
- lobster secretory gland, 596
- look-up table, 788
- LSIMS, 599, 600
- LTR sequence, 1070, 1072, 1073, 1081
- luciferase, 113, 263, 282, 283, 286, 301, 497, 897
 - firefly, 286, 287, 896, 900
 - Renilla, 287
- luciferase-coelenterazine system, 287
- luciferase-luciferin system, 286–287
- luciferin, 301, 896
- luminescence, 262
 - lifetime, 264
- luminol, 93, 897–899
- lung, 120
 - cancer, 1121
 - tumour, 300
- lymph system, 154, 163
- lymphocyte, 107, 155, 384, 387, 648, 689, 866, 1080, 1081
- lymphography, 163
- lysine, 110
- lysosomal disease, 1082
- lysozyme, 946, 952, 1030
- Mach-Zender interferometer, 525
- macroarray, 913
- macrofluidic system, 984
- macromolecule, 57, 937
 - biological, 223
 - motion of, 259
- macrophage, 689, 1102, 1106, 1107, 1109, 1113, 1148
- macular degeneration, 244
- maghemite, 134, 135, 141, 161, 164
- magnetic bead, 447, 448, 940, 1006
- magnetic labelling, 329
- magnetic resonance imaging, 159–164, 292, 404, 1092, 1123, 1126
 - contrast agent, 1093
- magnetic tweezers, 447–467
 - advantages, 458–461
 - basic principle, 447–449
 - bead tracking, 451–453
 - for manipulation, 466–467
 - magnet, 456–458
 - mechanical model, 449–451
 - twisting molecules, 458
- magnetite, 139, 161
- magnetocytolysis, 1130, 1134, 1136, 1139, 1148
- magnetoencephalography, 404
- MALDI, 599, 600, 606–609, 621, 625, 627, 628
 - basic principle, 607–609
 - sample plate, 626
- malto porin, 696
- maltose, 884
- mammography, 298
 - optical, 306, 307
- Marangoni effect, thermal, 752
- mass range, 612
- mass spectrometer, 596
- mass spectrometry, 495, 595–630, 780, 785, 952, 1009
 - analyser, 611–622
 - accessible mass range, 612
 - resolution, 612
 - sensitivity, 612
 - applications, 598–601
 - average mass, 597
 - basic principle, 596
 - error, 597
 - FT-ICR analyser, 599, 605, 606, 609, 611, 612, 620–622
 - FWHM, 598
 - ion trap analyser, 603, 609, 611, 619–620

- 2D, 612
- 3D, 612, 619
- linear, 620
- magnetic sector, 603, 609
- mass accuracy, 597
- mass range, 596
- mass resolution, 598
- mass resolving power, 598
- molecular ion, 597
- monoisotopic mass, 597
- orbitrap analyser, 599, 612
- quadrupole analyser, 599, 601, 603, 611, 612, 616–619
 - practice, 617–619
 - theory, 616–617
- sensitivity, 598
- terminology, 596
- time-of-flight analyser, 599, 601, 603, 609, 611–616
 - with chromatography, 622–624
 - with liquid phase separation, 622–625
- Massart process, 135
- maximum likelihood criterion, 794, 797
- Maxwell's equations, 537
- Maxwell, J.C., 701
- Maxwell–Garnett model, 507
- Mayr, E., 224
- measurement
 - automated, 783
 - complexity, 782
 - continuous flow, 984
 - direct, 786
 - dynamic range, 781, 788
 - filtering, 789
 - geometric localisation, 787
 - ideal, 787
 - independence, 787, 788
 - indirect, 786
 - linearity, 787
 - MIMO, 782
 - mix and read, 984
 - noise, 785, 789, 794
 - normalisation, 788
 - on-chip, 976–987
 - outlier, 787, 789, 797
 - protocol for lab-on-a-chip, 1001
 - reliability, 782
 - reproducibility, 787
 - robustness, 787
 - signal-to-noise ratio, 781, 789
 - SISO, 782
- mechanotransduction, 442, 969
- medical imaging, 95
- medical mycology, 1111
- MEDPROBE chip, 970
- melanin, 1033
- melanocortin, 1033
- melanocyte, 1033
- melanogenesis, 1033
- membrane
 - artificial, 714–718
 - asymmetric, 68
 - bacterial, 694
 - biological, 29–31, 55, 59, 68, 78, 85, 639
 - biomimetic, 30, 65, 82, 83, 92–95, 1009
 - brush border, 499
 - capacitance, 652, 685
 - ceramic, 716, 717
 - curvature, 397
 - cytoplasmic, 206, 383, 965, 966, 1088
 - depolarisation, 684, 690, 693
 - detector, 872
 - dynamics, 388
 - endocellular, 644
 - endosomal, 105
 - functionalised, 184
 - functions, 77
 - imaging, 380–384
 - intracellular, 42
 - leaflet, 46, 50, 68
 - lipid, 31–45, 114, 713–714
 - mimicking, 498
 - models, 46–77, 380–382, 1100
 - motor, 214
 - nanoporous, 1147
 - nitrocellulose, 942, 944, 947
 - Nuclepore, 697, 703, 716
 - nylon, 911
 - organic, 716
 - partitioning, 662
 - perforation, 662, 665, 718
 - permeability, 77
 - photosynthetic, 34, 206
 - plasma, 34, 42, 103, 107, 175, 177, 380, 385, 389, 498, 640, 642
 - polyvinylidene difluoride, 944

- potential, 644, 651, 971
 - depolarisation, 650
 - hyperpolarisation, 650
 - repolarisation, 650
- protein, 89, 382
 - reconstitution of, 91
- protein insertion, 79, 381
- resistance, 659, 660
- selective, 886, 887
- semi-permeable, 878, 895
- silicon carbide, 717
- supported, 57
- suspended, 72–74
- vitelline, 654
- MEMS, *see* microelectromechanical system
- meningioma, 648
- metabolite, 665, 871, 897
- metal alkoxide, 133
- metal colloid, 151
- metastasis, 120, 162, 163, 283, 303, 648, 1121, 1126
- metathesis polymerisation, 1046
- methane, 831
- methyl transferase, 19
- methylase, 927
- methylation, 1074
 - of CpG sequences, 19, 21
 - of cytosine, 9, 18–21, 23, 927
 - of DNA, 17–24
 - symmetric, 23
- metronidazole, 1055
- MGDG, 35
- mica, 62, 91, 391–394, 396
 - as AFM substrate, 378, 381, 383
 - substrate, 946
- micelle, 40, 73, 117, 118, 713, 1099
 - critical concentration, 122
 - cylindrical, 40, 44
 - normal, 40, 44
 - reverse, 42, 44, 1099
 - spherical, 40
- micro total analysis system, 743, 761, 1000
- microarray, 307–320, 775, 778, 900–905, 914
 - allergen, 956
 - analytical, 953
 - antibody, 954
 - cDNA, 915
 - cell, 928–930, 965
 - CGH, 923–925
 - data processing, 787
 - DNA, 316, 903–904, 911–930, 966, 976
 - electrical detection, 328
 - expression, 928
 - fabrication, 312–313
 - for PCR, 847
 - functional, 956
 - high-throughput, 930, 937, 956
 - history, 309
 - immunoassay, 950
 - laser scanning, 316–320
 - myelin, 956
 - oligonucleotide, 915–917, 949
 - patch-clamp, 661, 971–992
 - protein, 495, 905, 937–958
 - read system, 313–320
 - siRNA, 929
 - SNP, 919
 - synovial proteome, 955
 - third generation, 917–918
- microbead, 953
- microcantilever, 1007
- microcapsule, 1148
- microchannel, 748, 966
 - laminar flow, 747
- microdielectrophoresis, 767
- microdissection, 922
- microelectrode, 761, 974–976
 - array, 987
 - functionalised, 781
- microelectromechanical system, 628, 776, 1010
- microelectronics, 744, 1000, 1144
 - interface with living beings, 986
- microESI, 603, 625
- microfluidic chip, 629, 770
- microfluidic flow, 745–753, 781
- microfluidic system, 320, 322, 628, 697, 715, 718, 721, 726–731, 743–771, 939, 966, 967, 970, 971, 979, 982, 984, 1004
 - applications, 761–770
 - data processing, 791
 - detection, 761
 - fabrication, 753–761

- lithography, 754–755
- for lab-on-a-chip, 1011
- for patch-clamp array, 977
- for protein crystallisation, 761–764
- glass technology, 756
- laminar flow, 747, 760
- mixing, 760–761
- plastic technology, 757–758
- silicon technology, 755–756
- silicone elastomer, 758–759, 764
- temperature control, 761
- turbulence, 747
- micromanipulation, 183, 185, 397–402, 448, 656, 719, 967–972, 1028
- micropatterning, 969, 975
- micropipette, 185, 397, 398, 448, 655, 743, 975
- microplate, 310, 313, 940
- micropost array, 727, 765, 768
- micropump
 - membrane, 756, 760
 - peristaltic, 759, 760
- microrobot, 970, 1145
- microsphere, 942
- microspray, 605
- microsyringe, 1145
- microsystem, 1000
- microthruster, 968
- microtubule, 104, 108, 110, 114, 173, 194–196, 199, 446, 447
 - AFM observation, 379
 - disassembly, 198
- microvalve, 758, 760, 763
- Mie's theory, 1149
- MIMD, 833
- miniaturisation, 937, 1000, 1144
 - of biosensors, 907
- minicapsule, 1145
- Minsky, M., 317
- MION, 163
- mitochondrion, 12, 29, 77, 206, 207, 217, 258, 261, 640
 - ATP synthase, 207
 - genome, 919
 - membrane, 73
- mitosis, 110, 114, 173, 197
- mitotic spindle, 173, 197
- Molday process, 139
- molecular dynamics simulation, 7, 23, 803–835
 - accessible length scale, 807
 - accessible thermodynamic ensembles, 818
 - accessible time scale, 807
 - adaptive bias force method, 830, 831
 - all-atom, 813
 - AMBER, 809
 - Andersen equations, 821
 - autocorrelation function, 828, 829
 - basic principle, 806–808
 - bond stretching, 812
 - chemical bond, 812
 - coarse-grained, 813, 832
 - computation time, 824
 - cutoff sphere, 804, 823
 - Debye–Hückel theory, 823
 - diffusion coefficient, 829
 - dihedral angles, 810
 - dynamical properties, 827–829
 - electrostatic potential, 811, 813, 823–826
 - ergodicity, 819
 - Ewald lattice summation, 824–825
 - Ewald summation, 823–825
 - extended Lagrangian method, 819, 821
 - fast Fourier transform, 825, 826
 - free energy, 829–831
 - free energy perturbation, 829
 - Gauss–Seidel scheme, 817
 - Gay–Berne potential, 814
 - generalised reaction field method, 823, 825
 - Hamiltonian, 808, 821, 830
 - holonomic constraints, 817
 - integration step, 806, 813, 832
 - integration with constraints, 817
 - integrators, 814–816
 - isobaric, 821–822
 - isobaric–isenthalpic ensemble, 822
 - isobaric–isothermal ensemble, 821, 822
 - isothermal, 818–821
 - Kirkwood factor, 827
 - Langevin dynamics, 820
 - Langevin equation, 822
 - lattice summation, 804

- leapfrog algorithm, 816
- Lennard-Jones potential, 810, 811
 - modified, 814
- Lorentz–Berthelot mixing rules, 810
- minimal image approximation, 804
- Morse dissociation potential, 812
- multistep integration, 807–808
- Newton’s equations, 806, 814
- Nosé–Hoover algorithm, 819
- of benzene, 828
- of lipid, 814, 832
- of methane, 831
- of phospholipid bilayer, 810, 823
- oxygen–oxygen radial distribution, 826
- parallelisation, 831–834
- parametrisation, 810, 812
- particle mesh algorithm, 826
- particle mesh method, 826
- physical justification, 805–806
- Poisson–Boltzmann equation, 823
- potential energy, 808–814
 - coupling terms, 812
- quantum chemistry, 811
- radial distribution function, 826–827
- relaxation time, 827
- Ryckaert–Bellemans potential, 810
- simulation cell, 804, 821, 833, 834
- spherical cutoff, 825, 827
- structural properties, 826–827
- switching function, 823
- symplectic propagator, 806
- temporal correlation function, 827
- thermodynamic integration method, 830
- torsion potential, 809, 810
- umbrella sampling method, 830
- valence angle bending, 812
- validity, 806
- van der Waals interaction, 810
- velocity–Verlet algorithm, 815, 816
- Verlet algorithm, 815
- virial expansion, 821
 - with ionic species, 823
- molecular evolution, 224–226
 - directed, 232
- molecular hybridisation, 912
- molecular imaging, 160, 240, 1092
- molecular interaction dynamics, 333–360
- molecular modelling, 835
- molecular motor, 193–220, 447
- molecular recognition, 7, 30, 72, 158, 308–309, 1001
- molecularly imprinted polymer, 873, 875
- monkey, 956, 1075, 1079, 1087
 - transgenic, 1089
- monocyte, 689, 1030–1032, 1034, 1035
- mononuclear phagocyte system, 155
- mononuclear phagocyte system, 156, 1106
- morphine, 869
- MOSFET, 885
- motility, 172, 179–183
- motor protein, 196–198
- mouse, 265, 285, 297, 300, 303, 920, 1059, 1072, 1079, 1085, 1108, 1138, 1139
 - brain, 611
 - lentiviral injection, 1089
 - mutant, 1059
 - nude, 300
 - transgenesis, 284
 - transgenic, 922, 1086, 1088
- μ-contact printing, 62
- μFACS, 769
- multibeam interferometry, 392
- multiphoton microscopy, 261, 272, 610
- multiple input/multiple output
 - measurement, 782
- multiple sclerosis, 645
- multiplex analysis, 278, 953
- multiplexing, 782, 791
- murine model, 126
- muscle, 196, 200, 408, 409, 642, 665, 689, 1079
 - contraction, 173, 644, 938
 - cultured tissue, 970
 - enzyme, 883
 - formation, 927
 - ventricular, 647
- muscular dystrophy, 611
- mutagenesis, 387, 1081
- mutarotase, 884
- μTAS, *see* micro total analysis system
- mutation, 101, 198

- myelin
 - microarray, 956
 - sheath, 646
- myoglobin, 82, 1030
- myopathy, 1079
- myosin, 196, 200, 202, 379, 446
 - conformation, 203–204
 - step size, 201
- myotonia, 645, 646

- N-WASP, 180, 182, 187
- NAD, *see* nicotinamide adenine dinucleotide
- nano(U)Sonic, 1128
- nanosarray, 958
- nanobead, 970, 1148
- nanobiosensor, 92
- Nanobiotix
 - nano(U)Sonic technology, 1123, 1126, 1128, 1133
 - nanoMag technology, 1123, 1126, 1130
 - results, 1133–1140
 - nanoPDT technology, 1123, 1126, 1131
 - nanoXRay technology, 1123, 1126, 1127, 1133
- Nanobiotix technology, 1123–1125
- nanocapillary, 968
- nanocapsule, 1099, 1151
- nanocarrier, 91, 1098–1115, 1148, 1149
- nanochannel, 697, 705
 - conductance, 700
 - fabrication, 714
 - hemolysin, 714
 - slipping, 748
- nanodissection, 383, 389
- nanodomain, 1052–1054
- nanodroplet, 62
- nanoelectrode, 683
- nanoelectromechanical system, 776
- nanoESI, 603, 606, 625, 628
- nanoinprinting, 715, 755
 - cold, 757
 - thermal, 757
- nanolithography, 946, 952
- nanomachine, 171, 1160
 - chemical, 172
 - self-replicating, 237
- nanomechanics, 208, 215
- nanomedicine, 172, 277, 279, 776, 1111, 1145, 1158
- nanomembrane, 1151
- nanoneedle, 389
- nanoparticle, 104, 123, 263, 1123
 - activatable, 1121–1141
 - bifunctionalised, 1047
 - bismuth oxychloride, 1149
 - calcium sulfate, 1055
 - carrier, 1045–1048, 1098–1115, 1148
 - pharmaceutical applications, 1105–1111
 - stealth, 1107–1108
 - surface modification, 1106–1111
 - cationic, 1103
 - CdSe/ZnS, 151
 - charge, 1103–1104
 - chemical modification, 133
 - composition, 1099, 1101–1102, 1123
 - copolymer, 1101
 - core–corona, 132, 139, 161, 164
 - core–shell, 132, 140, 142, 143, 161, 167
 - doped rare earth, 275
 - encapsulation of, 139–141
 - ferrimagnetic, 167
 - fluorescent, 332
 - for antibiotic delivery, 1055
 - for biological tagging, 147
 - for cosmetics, 1149–1150
 - for electrical detection, 328
 - for functionalisation, 1055
 - for protein microarray, 942, 953
 - functionalised, 95, 129–168, 269, 276, 277, 279, 1047
 - gold, 145–146, 151, 153, 154, 167, 279, 328, 332
 - hydroxyapatite, 1055
 - in aqueous medium, 143
 - inorganic, 129
 - iron oxide, 133–141, 161
 - lipid, 30
 - maghemite, 141, 161, 164
 - magnetic, 95, 165, 1000, 1006, 1007, 1126
 - magnetite, 139, 161
 - metal, 151, 153, 270, 279–280, 355
 - multifunctional, 1128

- nanoMag synthesis, 1128–1130
- noble metal, 145–146
- optical response, 275, 277, 279
- oxide, 276
- plasma half-life, 157, 158, 1108
- polymer, 270, 277–279, 1099
- probe, 269–280
- semiconductor, 141
- silica, 94, 270, 277–279
- size, 1102–1103
- solid lipid, 1099, 1105, 1108
- structure, 1099
- surface modification, 130–146
- titanium oxide, 1149
- toxicity, 1155, 1157
- zinc oxide, 1149
- zirconium oxide, 1149
- nanopatterning, 62
- nanophase alloy, 1056
- nanophysics, 172
- nanopipette, 990
- nanoplex, 110, 123–126
 - endocytosis, 127
 - for gene transfer, 126–127
- nanopore, 695–719
 - artificial, 714–718
 - biological, 713–714
 - double layer, 699
 - fabrication, 716
 - Gouy–Chapman layer, 699
 - molecular transport via, 694–697
- nanoreactor, 1035
- nanorobot, 1156
- nanosensor, 1150
- nanoSIMS, 610
- nanosome, 1150
- nanosphere, 958, 1055, 1099
 - silicate, 1149
 - Trisacryl, 1133
- nanospray, 605
- nanospray tip, 629
- nanospray tip array chip, 629
- nanostucture, 277
 - fabrication, 1152
 - for detection, 329
 - for DNA separation, 765
 - functionalised, 1009
 - intelligent, 29
 - porous, 1151
- nanostuctured lipid carrier, 1099, 1105
- nanosurgery, 389
- nanosystem, 389
 - observation by vibrational spectroscopy, 555–556
- nanoTherapeutics, 1123–1125
 - advantages, 1123
 - biocompatibility, 1123
 - four families, 1125
- nanotip, 413, 427
- nanotopochemistry, 1052–1054
- nanotoxicology, 1160
- nanotube, 953
- nanotweezers, 697
- nanowire, 1009, 1144
- nanoworld, 4
- NAPPA, 941
- Navier–Stokes equation, 565, 699, 746, 751
- NBIC, 776
- Nd/YAG laser, 607
- NDGA, 1149
- near-field microscopy, 375, 541, 780
- necrosis of tumour cell, 1126, 1128
- Neher, E., 652
- nematode, 4, 1083
- Nernst model, 678, 679
- Nernst potential, 650
- Nernst–Planck equation, 699
- nerve growth factor, 1082
- nervous system, 175
- netropsin, 13
- neural network, 798
- neuroblastoma, 925
- neurodegeneration, 1080, 1082, 1086
- neuron, 34, 151, 352, 354, 641, 642, 645, 665, 666, 684, 688, 983, 1086, 1144
 - connection to electronics, 1148
 - cultured on-chip, 970, 971
 - GABAergic, 1086
 - giant, 596
 - imaging, 290
 - snail, 986
- neuronal network, 970
- neurotransmission, 642, 645, 665, 684–688
- neurotransmitter, 986
- neutron diffraction, 70
- neutron reflectivity, 1021, 1022

- neutron reflectometry, 584–586
- neutron scattering, 705
- neutrophil, 155, 175, 499
- Ni-NTA complex, 72, 73, 75, 87, 940, 946
- nicotinamide adenine dinucleotide, 883, 897, 900
- nitrilotriacetic acid, *see* NTA
- nociception, 645
- noise, 782, 785, 789, 794
 - autocorrelation function, 785
 - power spectral density, 786
- non-faradaic process, 669–672
 - equivalent circuit, 673
- noradrenaline, 684
- norbornene, 1046
- NTA, 72, 73, 75, 581, 940
- nuclear localisation signal, 109
- nuclear magnetic resonance, 159, 202
- nuclear pore, 110
- nucleic acid, 102, 103, 223, 225–226, 1001
 - AFM imaging, 378
 - detection, 847
 - fluorescent probe, 849
 - for biosensor, 872
 - interaction with proteins, 497
 - probe, 928
 - programmable protein array, 941
 - quantification, 852
 - structure, 227–231
- nucleosome, 17, 379
- nucleotide, 102, 227, 232, 431
 - binding site, 203, 208
 - chemical modification, 235–236
 - labelling, 257
- nylon, 911
- nystatin, 662, 665
- oesophagus, 298
- Ogston regime, 723
- OLED, 331, 332
- olefin, cyclic, 1046
- oligomer, 917
- oligonucleotide, 913, 915–917, 924, 926, 927, 930, 965
 - fluorescent, 950
 - microarray, 915–917, 949
- oligopeptide, 1055
- OMCTS, 427
- oncology, 292
- oncoprotein, 490, 491
- ophthalmology, 1112
- opsonin, 1107, 1108
- opsonisation, 156, 157, 161, 1103, 1107, 1109
- optical coherence tomography, 404
- optical fibre, 872, 891
- optical imaging
 - in vivo*, 292–298
- optical spectroscopy, 525–537
 - of biomolecular film, 527–528
- optical stretcher, 446
- optical trap, 434, 768
 - multiple beam, 437
 - quality of, 435
 - stiffness constant, 436–437
- optical tweezers, 183, 199–201, 434–448, 760, 770, 969
 - basic principle, 434–436
 - biological applications, 442–447
 - experimental setup, 441–442
 - holographic, 438–440, 445
 - manipulation of whole cell, 445
 - multiforce, 437, 444
 - multiple trap, 437
 - trapping force, 435
 - with FRET, 444
- optical waveguide lightmode spectroscopy, *see* waveguide spectroscopy
- optoelectronic system, 872
- organ transplant, 1043
- organelle, 29, 35, 42, 177, 197, 642, 644, 667, 872
- organosilane, 136, 137
- organotypic slice, 653, 665
- orthopedic implants, 1056
- orthopedics, 1055
- osmosis, 640
- osmotic forcing, 706–707
- osteoblast, 1028, 1033, 1056
- outlier, 787, 789, 797
- OWLS, *see* waveguide spectroscopy
- oxidase electrode, 883
- oxidative stress, 688–694
- oxygen sensor, 893

- PAF, 340
- PAH, *see* poly(allylamine) hydrochloride
- pain, 645
- palindromic sequence, 15
- pancreatic acine, 384
- parallel processing, 782, 831–834, 965
- parasitology, 1111
- parathion, 888
- Parkinson's disease, 1082, 1083, 1086
- parodontology, 1055
- parvovirus, 1070
- passive targeting, 156, 167, 1107, 1123
- patch-clamp technique, 639, 652–667, 697, 714, 971, 976
- Ag/AgCl electrode, 658
- amplifier head, 658, 659
- automation, 974
- biological preparation, 653–655
- cell-attached configuration, 662
- current recording, 660–664
- experimental implementation, 653–660
- gigaseal, 653, 977, 978, 980, 982, 983
- history, 652
- inside-out configuration, 662
- limitations, 972
- loose, 665
- manual, 975
- microarray, 661, 971–992
- micromanipulation, 656, 662
- micropipette, 655
- outside-out configuration, 663
- planar array, 977–979
- seal, 653, 657, 982–984
- single-channel recording, 653, 661, 663, 664
- variants, 665
- whole-cell recording, 653, 654, 660, 662–665, 978, 980
- patchers, 987, 988
- patents, 1152
- pathogen, 938
- detection, 323, 667
- identification, 999
- pathogenicity, 1075
- pathology, 101
- multifactorial, 920
- paxilin, 444
- PCNA, 929
- PCR, *see* polymerase chain reaction
- PDMS, 322, 332, 599, 754, 980–982
- channel, 715
- for lab-on-a-chip, 1010–1012
- for microfluidics, 718, 758–759, 970
- PEEK, 1051
- PEG, 73, 74, 121, 127, 131, 139, 157, 969
- coating, 1027
- for stealth carriers, 1108, 1111, 1114
- for stealth immunoliposome, 1109
- graft, 1048
- in a pore, 706, 707
- PEGylation, 157, 1108
- PEI, 110–114, 120, 121, 534, 575, 1028, 1030, 1031, 1035
- Peltier effect, 843, 846, 1007
- penicillin, 1113
- PEO, 1047
- peptide, 379, 380, 407, 490, 1148
- adhesion, 1029
- amyloid, 496
- bond, 322, 873, 938
- chromatography, 780
- immobilised, 493, 1051
- MALDI analysis, 606
- mass, 595
- MS analysis, 598, 599, 615, 628
- proadhesive, 1045
- production, 941
- RGD-containing, 1048–1052, 1056, 1111
- separation by chromatography, 622–625
- ultrasensitive MS analysis, 605
- peptisation, 135
- peripheral protein, 85
- perivitelline space, 1088
- peroxidase, 689, 898–900, 1035, 1036
- peroxynitrite, 689
- pesticide, 1151
- carbamate, 888
- detection, 312, 327
- MS analysis, 621
- organophosphorus, 888, 889
- PEUU, 1053
- PGA, *see* poly(L-glutamic) acid
- pH sensor, 892

- phagocytosis, 1107
- pharmaceutical industry, 972, 973, 1105
- Pharmacia, 320
- pharmacodynamics, 1139
- pharmacogenetics, 865–869, 957
- pharmacogenomics, 957
- pharmacology, 312, 971, 1045
- phenol red, 892
- phenylalanine, 497, 646
- phonon, 152
- phosphatase, 953
- phosphate, 111, 118
- phosphatidic acid, 31
- phosphatidylcholine, 713, 1055, 1101
- phosphatidylglycerol, 1101
- phosphoglyceride, 31
- phospholipase, 644
- phospholipid, 956, 1099
 - bilayer, 585, 586, 1102, 1104, 1106
 - PEGylated, 1109
 - membrane, 684
 - molecular dynamics simulation, 810, 823
 - monolayer, 582
 - PEGylated, 1108
 - vesicle, 1039
- phosphonic acid, 138, 142
- phosphorescence, 148, 263
- phosphoric ester, 898
- phosphorylation, 645, 851, 1092
- phosphosphingolipid, 34
- photoactivation, 313, 340
- photoaptamer, 944
 - array, 945
- photobiotin, 313
- photobleaching, 147, 151, 257, 258, 266, 335–337
 - of FRET acceptor, 358
 - of quantum dots, 270, 271
- photodiode array, 322
- photodynamic therapy, 1126, 1131
- photolithography, 62, 754, 755, 759, 885, 916, 941, 1010
- photon counting histogram, 345
- photonic force microscopy, 387
- photoreceptor, 642
- photoresist, 754, 1011
- photosynthesis, 77, 383
- phototaxis, 642
- photothermal effect, 152, 355
- photothermal treatment, 167
- physisorption, 132, 157, 575, 1052
 - of water, 134
- picoforce, 446, 447
- PicoGreen, 847
- piconewton force, 398
- picospray, 605
- pig, 273, 296, 303, 1060
 - transgenic, 1088
- pixel, 788
- plasma activation, 1010
- plasma half-life, 157, 158, 160, 163, 1108, 1109, 1114
- plasmalogen, 33
- plasmid, 103, 105, 109–113, 123, 283, 300, 654, 923, 928, 941, 1075
 - transfection, 929
- plastic for microfluidics, 757–758
- platelet, 155
 - activation, 31
- platinum, 313, 683, 690
 - electrode, 881, 882
- plectoneme, 459
- PLED, 331, 332
- PMIRRAS, 547–551
 - experimental setup, 548
 - on dielectric substrate, 549
 - photoelastic modulator, 547
 - spectrum, 548–550
- PMMA, 277, 321, 417, 755, 757
 - film, 546
 - for lab-on-a-chip, 1010, 1012
 - nanoparticle, 1108
 - resist, 715
- PNB, *see* polynorbornene
- POE, 157
- Poiseuille flow, 748, 749, 762
- Poisson's equation, 701, 750, 824
- Poisson–Nernst–Planck equation, 699
- polarisation of light, 501–508, 541
 - circular, 503
 - elliptical, 503
 - linear, 503
- pollution, 263, 312, 1001, 1151
 - enzymatic digestion, 1151
- poloxamer, 157
- poloxamine, 157

- poly(allylamine) hydrochloride, 535, 536, 1018, 1030
- poly(ethylene terephthalate), 1051
- poly(glycolic acid), 1099, 1101, 1148
- poly(L-glutamic) acid, 497, 535, 536, 1021, 1031
- poly(L-lysine), 110, 497, 536, 575, 1021, 1025, 1026, 1028, 1031, 1033, 1051
 - fluorescein-tagged, 1039
- poly(lactic acid), 1051, 1099, 1101, 1148
- poly(vinyl acetate), 1051
- poly-uracil, 696
- polyacrylamide, 878
- polyacrylic acid, 144, 161, 1028
- polyalkylcyanoacrylate, 1099
- polyamide, 878
- polyamine, 110
- polyanion, 575, 1017, 1018
- polybenzylglutamate, 549
 - film, 548
- polycarbonate, 321, 628, 716, 755, 757, 1128
 - injection moulding, 1011
- polycation, 110, 536, 575, 1017, 1018
 - deuterated, 1022
- polyelectrolyte, 703, 711
 - adsorption kinetics, 1026
 - deuterated, 1021
 - electrophoresis, 764
 - strong, 1023
 - weak, 1023
- polyelectrolyte multilayer, 528, 539, 1017–1041
 - alginate/PLL, 1028
 - bio-inert, 1027–1029
 - charge overcompensation, 1018
 - complex architectures, 1038
 - construction, 1018
 - containing enzymes, 1035
 - exponential growth, 1023–1026, 1033, 1034, 1038, 1039
 - fabrication, 1026–1027
 - functionalised, 1027–1036
 - by drugs, 1035
 - by peptides, 1033
 - HA/chitosan, 1028, 1033
 - HA/PLL, 1033
 - hollow particles, 1037–1038
 - insertion of vesicles, 1039
 - linear growth, 1021–1023, 1038
 - neutron reflectivity observation, 1021, 1022
 - optical thickness, 536, 1021
 - OWLS observation, 534–537
 - PEI/PVS, 1028
 - PLL/HA, 1025, 1026, 1028, 1033
 - PLL/PGA, 1021, 1028–1030, 1033
 - protein adsorption, 1029
 - protein denaturation, 1029
 - protein insertion, 1029–1033
 - PS/PAH, 1029
 - PSS/PAH, 1018, 1020, 1021, 1029
 - QCM observation, 575–577
 - rinsing stage, 1026
 - role of entropy, 1019
 - thickness, 1020, 1027
 - zeta potential, 1019
- polyethylene oxide, *see* PEO
- polyethyleneimine, *see* PEI
- polygalacturonic acid, 878
- polyimide, 716, 1011
- polymaleic acid, 1028
- polymer
 - backbone, 408
 - biodegradable, 1101
 - bioerodable, 1099
 - bioresorbable, 1099
 - carrier, 1099, 1100
 - chain, 1017
 - chain dynamics, 707, 709
 - charged, 1017
 - confined in a pore, 711–713
 - chip, 981
 - confined chain dynamics, 709
 - confined in pore, 703–713
 - dynamic surface-coating, 725
 - entangled, 721
 - entropic elasticity, 408
 - excluded volume, 703
 - film, 755
 - for lab-on-a-chip, 1011, 1012
 - force–extension curve, 407–412
 - gel, 721
 - glass transition temperature, 755, 757
 - granule, 757
 - heat sensitive, 725
 - hot embossing, 1011
 - matrix, 721–725

- examples, 725
 - melt, 757
 - molecularly imprinted, 873, 875
 - nanoparticle, 1099
 - neutral, 704
 - confined in pore, 704–711
 - non-cytotoxic, 1047
 - persistence length, 410, 1020
 - porous, 1006
 - semi-rigid chain model, 408, 409
 - semiconducting, 986
 - stimuli-responsive, 1046
 - substrate, 629
 - synthetic, 878
 - theoretical description, 703
 - thermoplastic, 755, 757
 - wormlike chain model, 409, 455
- polymerase chain reaction, 223, 720, 770, 841, 913, 1002, 1146, 1150
 - amplicon, 842
 - amplification curve, 854
 - annealing, 842
 - temperature, 855, 857
 - applications, 862–869
 - basic principle, 841
 - crossing point, 845, 864
 - cycle threshold, 845
 - denaturation, 842, 856
 - curve, 854, 857
 - temperature, 855
 - effect of pH on fluorescence, 857
 - efficiency, 857–861
 - end-point method, 844, 845
 - extension, 842
 - fluorescence format, 847–853
 - fluorescent marker, 847
 - high-throughput, 852
 - hot start, 854
 - hybridisation probe, 849, 851
 - hydrolysis probe, 849, 850
 - in pharmacogenetics, 865–869
 - kinetics, 843–845
 - molecular beacon, 853
 - multiplex, 847, 850, 861–862
 - on-chip, 1007
 - post-PCR melting curve, 852
 - primer, 841, 842, 845, 847–849, 852, 857
 - quantification of viral genomes, 862–864
 - quantitative real-time, 845–847
 - equipment, 846
 - real-time, 841–869
 - relative quantification, 860–861
 - reverse transcriptase, 940
 - scorpion probe, 849, 853
 - SybrGreen, 847–849, 853, 857
 - TaqMan probe, 850, 852
 - thermocycler, 842, 845, 846
- polymethacrylate, 878, 1101
- polymethacrylic acid, 1028
- polynorbornene, 1047
- polypeptide, 1031
 - chain, 873, 874
 - MS analysis, 599
 - multilayer, 528, 1023
 - ring, 412
 - separation by chromatography, 622
 - synthetic, 549
- polyplex, 110–115
 - for gene transfer, 120–121
 - stability, 112
 - synthesis, 111
- polysaccharide, 408, 411, 479, 499, 878, 1028
 - multilayer, 1023
 - natural, 575
 - sulfated, 498
- polystyrene, 277, 878, 1051
 - sulfonate, 1018, 1028, 1030, 1035
- polyurethane, 878
- polyvinyl alcohol, 878
- POP, 157
- POPC, 1039
- pore, 640, 643, 686, 938
 - conductance, 699
 - conducting cylindrical, 701–703
 - confining a polymer, 703–713
 - conical, 697, 715
 - containing a bead, 701
 - electrical detection of particle transport, 698–703
 - electrical resistance, 698, 701–703
 - hemolysin, 696, 706, 707, 712
 - nanolithographic, 699
 - protein, 699, 706, 765
 - surface charge, 699

- porin, 77
- porphyrin, 716
- positron emission tomography, 160, 240, 293, 404, 1083, 1092
- potentiometric detection, 872, 880, 884
- Prader–Willi syndrome, 925
- pregnancy test, 327
- preprocessing, 782, 786–787
- primate, 1060, 1082, 1083, 1085, 1086
- prion, 379, 496
- probability distribution, 344
- probimide, 986
- processive motor, 180, 181, 183, 198, 199, 201
- profilin, 177, 179, 181–183
- prognosis, 912
- prokaryote, 14, 173, 1078
 - DNA, 21
- prolonged QT syndrome, 647
- promoter sequence, 11, 16, 20, 283, 461, 497, 927, 929, 969, 1071, 1074
- pronuclear injection, 284, 285, 1089
- prostate, 500
 - cancer, 647, 1130
 - hyperplasia, 628
- prosthesis, 1043, 1143
 - articular, 1045
 - vascular, 1057
- protamine, 110
- protease, 953
 - gastric, 1097
- protective cream, 1149
- protein, 101, 103, 104, 937
 - adsorption, 520–521, 561, 573
 - adsorption dynamics, 533, 539
 - affinity capture, 944
 - biotinylated, 944
 - capsid, 1076
 - cargo, 109
 - channel, 639, 640
 - chaperone, 425
 - chimeric, 206
 - covalent immobilisation, 943
 - crystallisation, 761–764
 - expression profile, 952, 953
 - folding, 938
 - hydrodynamic radius, 527
 - immobilised, 581, 582
 - in situ production, 940
 - MALDI analysis, 606
 - microarray, 495, 905, 937–958
 - analytical, 953
 - applications, 953–957
 - detection, 947–952
 - functional, 956
 - immobilisation, 943–944
 - on flat support, 939–952
 - protein spotting, 944–946
 - substrate, 942
 - motor, 108
 - MS analysis, 599, 600, 605
 - mutant, 496
 - non-covalent immobilisation, 944
 - osteoinductive, 1055
 - production, 101
 - purified, 939
 - recombinant, 87, 583, 938, 940
 - Rep, 1079
 - secretion, 694
 - sentinel, 644
 - spotting, 944–946
 - structural, 938
 - structure, 654
 - synthesis, 901
 - translocation, 929
 - transmembrane, 85, 640, 642, 762
 - unfolding, 407, 408, 410, 412, 497, 698
 - two-state model, 410, 412
- protein chip, 62, 238
- Protein Data Bank, 202, 209, 213
- protein organisation, two-dimensional, 85, 89
- protein–lipid assembly, 77–92
- proteome, 775, 785, 902, 955
 - yeast, 947
- proteomics, 495, 775, 953
- proton pump, 85, 114, 383
- proton-motive force, 210–212, 216, 218, 220
- PSS, *see* polystyrene sulfonate
- psychiatry, 865
- PTFE, 1051, 1056
- pulsed laser diode, 304
- pulsed-field electrophoresis, 765, 768
- purine, 101, 103, 1080
- pyridine, 555

- pyrimidine, 101, 103, 866
- pyroxicam, 1034, 1035
 - chemical structure, 1034
- pyrrole, 313, 323
- pyruvate, 884
 - oxidase, 884
- QCM, *see* quartz crystal microbalance
- QT interval, 646
- quality control, 1150–1151
- quantum dot, 95, 141, 142, 148, 149, 151, 269–275, 278, 303, 352, 355, 1144
 - biocompatibility, 273
 - blinking, 272, 277, 355
 - CdSe, 141, 149
 - CdSe/ZnS, 144, 271, 272
 - emission lifetime, 271
 - encapsulation, 274
 - fluorescent, 949
 - functionalised, 150, 271, 272
 - optical response, 147–149, 270, 272
 - PbS, 271
 - photobleaching, 271
 - photostability, 149, 150
 - quantum yield, 149, 270
 - structure, 270
- quantum yield, 142, 143, 149, 270
- quartz, 563, 878
- quartz crystal microbalance, 58, 60, 90, 389, 525, 534, 535, 561–577, 1007
 - basic principle, 561–562
 - crystal in vacuum, 563–564
 - crystal in viscous medium, 564–572
 - experimental setup, 562
 - numerical simulation, 572–574
 - Sauerbrey constant, 564
 - with dissipation monitoring, 576
- quenching, 257, 258, 301, 849, 850, 853
- quinolone, 1113
- quinone, 883
- rabbit, 940, 941, 951, 952, 1059
- radioactive label, 526
- radiofrequency identification, 1148, 1150, 1156
- radiotherapy, 1123
- Radon transform, 793
- Raman microscopy, 553
- Raman scattering, 280
- Raman spectroscopy, 551–555
 - basic principle, 551–553
 - resonance, 554–555
 - signal enhancement, 553–555
 - surface enhanced, *see* SERS
 - weak points, 553
- Randles' model, 781
- random walk, 703
 - self-avoiding, 703
- Ranvier node, 646
- rat, 273, 297, 1059, 1085, 1086
 - brain, 605, 665
 - astrocyte, 1110
 - lentiviral injection, 1089
 - striatum, 1091
 - transgenic, 1088, 1089
- Rayleigh scattering, 152
- reactive ion etching, 755, 757
- receptor, 31, 78, 147, 156, 302, 351, 397–399, 442, 481, 498, 527
 - artificial, 873, 876
 - channel, 641, 644
 - endogenous, 335, 655
 - extracellular, 121
 - GABAergic, 641
 - integrin, 499
 - vanilloid, 645
 - viral, 930
- recombinant
 - adenovirus, 1077
 - virus, 1071
- recycling, 1151
- red blood cell, 48, 68, 155, 384, 397, 605, 1107
 - elasticity, 398
 - MS analysis, 621
 - viscoelasticity, 446
- reflection of light, 504–508
 - by thin film, 505–506
 - Fresnel coefficients, 504
 - plane of incidence, 504
- regenerative medicine, 1160
- replication, 12, 14
- reporter gene, 113, 263, 281–291
- resonance unit, 478
- reticulocyte lysate, 940
- retinopathy, 648

- retroviral vector, 1070, 1072–1074, 1080, 1082
- retrovirus, 1070, 1072, 1074
- Rett's syndrome, 24
- Reynolds number, 746–748
- RFID, 1148, 1150, 1156
- RGD sequence, 1048–1052, 1056, 1111
- RHEED, 70
- rheology, 416
- rhodamine, 265, 266
- rhodopsin, 77, 85, 383, 412
- ribavirin, 864
- ribonuclease, 1030
- ribosome, 807
- riboswitch, 241
- ricin, 328
- RIE, *see* reactive ion etching
- ripening, 137
- RISC, 1083
- risk prevention, 1151
- RNA, 101, 224, 226, 227, 863
 - antisense, 244
 - aptamer, 231
 - interference, 1080, 1082–1085
 - interfering, 928, 929
 - manipulation, 698
 - messenger, 104, 640, 901, 911
 - molecular dynamics simulation, 807
 - polymerase, 87, 241, 378, 461, 927
 - short hairpin, 929
 - short interfering, 913, 929, 1082
 - structure by AFM, 378
 - synthesis, 497
 - transfer, 230
 - ultrafast sequencing, 697
- RNA-induced silencing complex, 1083
- robot, 313
- Rohrer, H., 402
- rolling-circle amplification, 949, 950
- ROMP, 1046
- RPLC, 623
- Ruska, E., 402
- ryanodine, 642
- saccharose, 884
- safety screening profiling, 974
- salmonella, 499, 888, 889, 1150
- sampling, 785
- sarcomere, 408
- sarcosine, 884
 - oxidase, 884
- Sauerbrey film, 572, 575
- Sauerbrey relation, 564
- SAW, *see* surface acoustic wave
- SAXLC, 622
- scale effect, 781
- scanning angle reflectometry, 526
- scanning electron microscopy, 70, 402
- scanning tunneling microscopy, 375, 402, 414
- Schrödinger equation, 805
- SCID-X1, 1081
- science fiction, 771, 1156
- scintigraphy, 160
- screeners, 987
- screening, 312, 495, 865, 965, 973
 - counter, 974, 988
 - high-throughput, 667, 972, 973, 976, 988, 999, 1001, 1143
 - low-throughput, 974
 - parameters, 974
 - primary, 974
 - quantitative differential, 920
 - secondary, 974
- SCXLC, 622
- sea pansy, 287
- selection, 232
 - pressure, 234
- SELEX, 224, 231–236
- self-assembled monolayer, 322
- self-assembly, 29, 31, 36–39, 83, 115, 117, 173, 322, 969, 1099
 - helical, 88
 - of actin, 174
 - of proteins, 495
 - of tubulin, 174
- self-inactivating vector, 1075
- semiconductor colloid, 147
- Sendai virus, 1106
- sensor chip, 496, 498
- SensorChip, 320
- sentinel lymph node, 303
- serotonin, 684
- SERS, 70, 280, 555
- Shannon's theorem, 785
- shear, 746
- sheep, 1060
- SHREC, 202

- shRNA, *see* short hairpin RNA
- signal processing, 1000
- signal transduction, 30, 68, 77
 - on-chip, 1007–1009
- silane, 313, 323, 432, 1052
 - photoactivatable, 313
- silanisation, 136, 332, 1010
- silanol, 624
- silica, 272, 717, 1123
 - amine-grafted, 431
 - as dielectric, 542
 - bead, 199, 200
 - capillary, 605, 624
 - coating, 1129
 - encapsulation, 1132
 - film, 323, 521, 554
 - matrix, 277
 - nanochannel, 705
 - nanoparticle, 94, 270, 277–279
 - nanotube, 715
 - particle, 623
 - photoelastic modulator, 512
 - shell, 140
 - substrate, 62, 63, 91, 528, 629
 - surface chemistry, 135
 - tube, 700
- silica gel, 878
- silicon, 551, 980
 - carbide, 717, 718
 - chip, 981
 - for lab-on-a-chip, 1012
 - for microfluidics, 755–756
 - immobilising proteins on, 944
 - LED, 331
 - micropatterned, 311, 312
 - nitride, 717, 981
 - oxide, 536, 981
 - reconstructed surface, 414
 - stimulator, 986
 - substrate, 67, 313, 554, 560, 561, 586, 911, 968
 - wafer, 757
- siloxane bond, 137
- silver, 151, 279, 313, 320, 333, 391, 555, 880
- SIMS, 600, 606, 609–611
 - biological imaging, 610, 611
- single nucleotide polymorphism, 848, 865, 919
- single particle tracking, 334
- single-input/single-output measurement, 782
- siRNA, *see* RNA, short interfering
- siRNA array, 929
- smart patch clamp, 990
- Smoluchowski equation, 699
- SMP, 1052
- snail, 986
- snake photon, 304
- Snell–Descartes laws, 503, 578
- SNOM, 375, 388, 555
- SNP microarray, 919
- SOD, 689
- soft chemistry, 133
- soft lithography, 754, 758, 763, 764
- solid–liquid interface, 525–528, 537
 - protein adsorption, 561
- space research, 1001
- speckle fluorescence microscopy, 176
- SPECT, 160
- spectrin, 397
- spectroscopic imaging, 780
- spermatogonia, 1089
- spermine, 116, 122
- sphingolipid, 34, 381, 382
- spike effect, 753
- SPIO, 161–163
 - versatile, 164
- spleen, 1106, 1107, 1109
- splicing, 866, 1079
- SPT, 351–353
- starch, 878
- statistical test, 781
- stealth particle, 157, 159, 1098, 1107–1108
- stereolithography, 754
- stereotaxic injection, 1085
- stereotaxy, 1086, 1087
- steric hindrance, 257
- steric repulsion, 131, 139, 142
- Stern layer, 130, 135, 720, 1103
- sterol, 35, 37
- Stöber synthesis, 277
- stochastic process, 794
- Stokes shift, 264, 268
- Stokes' law, 253

- Stokes, G., 253
 streak camera, 305
 streptavidin, 63, 74, 87, 88, 210, 211, 333, 399, 407, 494, 944
 stress tensor, 746
 striatum, 1086, 1087
 rat, 1091
 strip test, 276
 structure–function relation, 380, 383, 386
 sugar, 498
 sulfonic acid, 138
 superhelicity, 10, 12, 15
 superparamagnetism, 161, 1006
 supported phospholipid bilayer, 57
 surface acoustic wave, 1007
 surface force apparatus, 390–402
 surface plasmon resonance, 58, 152, 320–323, 477–500, 780, 904, 949–951
 applications, 495–500
 concentration measurement, 495, 500
 data, 479–484, 489–495
 DMSO calibration, 492
 double referencing, 492
 immobilising the ligand, 493–495
 kinetic conditions, 484–486
 mass transport conditions, 484, 486–487
 non-specific signals, 489
 nucleic acid–protein interactions, 497
 protein structure, 496
 protein–protein interactions, 496
 protein–receptor binding, 498
 protein–sugar interactions, 498
 sensorgram, 478–480, 486–488, 490, 491, 494
 spectroscopy, 526, 527
 surface regeneration, 493
 surface tension, 50, 52, 397, 398
 surface-enhanced Raman spectroscopy, *see* SERS
 surfactant, 1099, 1108
 toxicity, 1101
 SybrGreen, 847–849, 851, 853, 857
 symplecticity, 806, 807
 synapse, 175, 986, 1144
 artificial, 684–695
 chemical, 684
 electrical, 684
 regeneration, 446
 transmission, 644
 synaptic bud, 684
 synaptic gap, 151, 353
 synovial proteome microarray, 955
 synthetic vector, 104, 105, 110

 tachycardia, 647
 tagging, biological, 146–153
 Taylor dispersion, 749, 750, 752, 762
 TCSPC, 305
 technological bottleneck, 781–783, 972, 979
 solutions, 979–986
 TED, 70
 teflon, 881
 temporal point spread function, 304–306
 tet-OFF system, 1090, 1091
 tet-ON system, 1090
 tetracycline, 1090, 1091
 Texas Red, 1025, 1026
 theophylline, 231
 theranostics, 1125
 therapy, 154, 164–167, 292, 497, 639, 645, 864, 912, 937, 952, 956, 1143, 1145–1148
 antithrombosis, 1056, 1057
 cancer, 1121–1141
 gene, 1069, 1079–1082
 new strategies, 1071
 photodynamic, 1126, 1131
 tailored, 1125
 using aptamers, 242–244
 thermistor, 872, 889–891
 thermocouple, 525
 thermoforming, 754
 thiol, 139, 145, 274, 279, 313, 322, 323, 493, 879, 1051, 1052
 Thomsen's disease, 646
 thrombin, 243, 498
 thrombogenesis, 1056, 1057
 thromboresistance, 1057
 thrombosis, 1057
 thymic ontogenesis, 920
 thymidine, 101, 396
 thymocyte ontogeny, 922
 time-of-flight analyser, 603, 609, 611–616

- field-free region, 613
- linear mode, 613
- orthogonal acceleration, 601, 615
- reflectron mode, 613
- resolution, 615
- resolving power, 613
- tipnology, 945
- TIRF microscopy, 199, 200, 404, 442
- TIRF spectroscopy, 526
- tissue
 - bone, 1054–1056
 - engineering, 1043–1045, 1143
 - investigative methods, 1061–1062
 - product, 1058–1061
 - model, 967
 - optical response, 293–296
 - reconstituted, 965
 - substitute, 1045
- titanium
 - alloy, 1052, 1056
 - functionalised, 1052
- titanium dioxide, 528
- titanium oxide, 62, 63, 536
 - nanoparticle, 1149
- titin, 407–410
- tomography, 298, 1092
 - data processing, 780, 786, 791, 793
 - optical, 299
- TOP, 142, 273
- TOPO, 142, 144, 145, 273, 274
- topoisomer, 12
- topoisomerase, 14, 463
- topology, 25
- torsades de pointes, 647
- toxicity, 151, 160, 273, 275, 297, 1045,
 - 1069, 1071, 1076, 1078, 1105, 1108, 1132, 1138, 1155, 1160
 - of active principles, 1125
- toxicology, 971, 1155
- toxin, detection, 667
- TRACE, 268
- traceability, 1150–1151
- tracheotomy probe, 1047
- track etching, 716, 1049
- tracking, 351–355
 - single-particle, 351–353
- transcriptase, reverse, 864, 1072
- transcription, 10, 12, 14, 15, 25, 104, 263, 901, 923, 929
 - decoy molecule, 497
 - factor, 23, 228, 231, 232, 497, 925
 - reverse, 864, 966, 1070, 1072
- transcriptome, 902, 911–923, 969
 - applications, 920–923
 - data analysis, 919
 - microarray, 925
- transducer, 872, 877
 - electrochemical, 872
 - mass, 872
 - optical, 872
 - thermal, 872
- transducing unit, 1070
- transduction, 1069, 1070
- transfection, 102, 115, 283, 335, 443, 654, 928, 966
 - reverse, 928
- transferrin, 121, 888, 1109
- transgene, 969, 1069–1071, 1073
 - controlling expression, 1089–1091
 - imaging expression, 1092
- transgenesis, 1087–1089
- transgenic animal, 283, 285, 920, 922, 1073, 1087
- transgenic cell, 969
- translation, 25, 104, 901, 902, 1072
 - blocking, 1083
- transmembrane
 - electrochemical potential, 650
 - ion flow, 77
 - protein, 85, 640, 642, 762
- transmission electron microscopy, 70, 86
- treadmilling, 176, 179
- triblock copolymer, 416, 1099
- triglyceride, 1099, 1105
- Trisacryl, 1133
- Trotter formula, 808
- tryptophan, 873
- tubulin, 173, 194, 195, 199
 - structure, 174
- tumour, 119, 127, 163, 648, 1106, 1109, 1148
 - angiogenesis, 1110
 - colon, 1139
 - deep, 1126, 1130, 1139
 - detection, 298
 - fast growth, 1108
 - growth, 283, 648
 - imaging, 286

- lung, 300
- magnetocytolysis, 1130
- malignant, 163
- penetration, 1114
- subcutaneous, 301
- targeting, 159, 1110, 1123, 1130, 1138
- tissue, 1106, 1108
- tumour-suppressing gene, 20
- tunnel current, 414, 415
- tunnel effect, 151
- two-photon excitation, 346, 351
- two-photon microscopy, 261, 319

- ubiquinone, 73
- ultramicroelectrode, 679–682, 686–688, 691
 - carbon, 690
 - platinum-coated, 690
- ultrasound, 404, 1102, 1123, 1126, 1128, 1133
- universal chip, 311
- up-converting phosphor, 276
- uracil, 866
- urea, 872, 876
 - electrode, 884–885
 - hydrolysis of, 884
- urease, 871, 884, 885
- uterus, 298
- UV laser, 317
- UVISEL, 513

- vaccine, 956, 1077
- vaccinia virus, 10, 176
- van der Waals interaction, 39, 130, 322, 376, 406, 415, 661, 873, 877
 - between nanoparticles, 131
 - in molecular dynamics simulation, 810
 - measurement, 393
- vancomycin, 1055
- vascular system, 1056–1058
- vascularisation, 1055, 1056, 1108
- vector
 - non-viral, 1069
 - self-inactivating, 1075
 - viral, 1069, 1071–1089
- vectorology, 1092
- VEGF, 1055, 1111
- ventricular fibrillation, 647
- vesicle, 41, 42, 94, 108, 184, 185, 192, 193, 684, 713
 - adhesion energy, 401
 - endocytotic, 1110
 - extruded unilamellar, 48
 - functionalised, 184, 397
 - giant unilamellar, 46, 49
 - large unilamellar, 46, 47, 117, 1099
 - lipid, 397, 401, 561
 - multilamellar, 46, 47, 117, 1099, 1102
 - phagocytotic, 1113
 - phospholipid, 1039
 - small unilamellar, 46, 49, 117, 381, 1099, 1102
- vibrational spectroscopy, 540–556
 - basic principle, 540–541
 - of nanosystems, 555–556
- vinculin, 444
- viral replicon, 104
- viral vector, 1069, 1071–1089
 - biomedical applications, 1079–1089
- virion, 1072, 1077
- virology, 862, 864
- virus, 10, 12, 104, 153, 176, 207, 284, 384, 499, 689, 1071, 1075
 - CMV, 282
 - genome, 862–864, 1070
 - hepatitis C, 862
 - herpes simplex, 1071, 1090
 - HIV, 229, 231, 241
 - Moloney, 1072
 - murine leukemia, 1072
 - particle, 1070
 - promoter, 284
 - recombinant, 1071
 - replication-competent, 1076
 - Sendai, 1106
 - vesicular stomatitis, 1072
 - wild-type, 1070, 1075, 1077
- viscosity, 746
- viscous diffusion coefficient, 747
- vitamin, 500
 - E, 1150
- voltage-clamp technique, 649, 652
- voltammogram, 675, 691

- wafer, 1000
- Warburg diffusion impedance, 781

- water
 - desalination, 1151
 - filtering, 1017, 1151
 - layered structure, 394
 - quality, 323, 327, 621, 1151
 - TiP4P, 826
- wave equation, 538
- waveguide, 330–332, 526, 949, 950
- waveguide spectroscopy, 527–537
 - applications, 534–537
 - basic principle, 528–532
 - resolution, 533
 - signal processing, 532–533
- wavelet, 789
- wax, 1099, 1105
- wetting, 753
- white blood cell, 155
- World Health Organisation, 1121
- WPRE, 1075
- WSC, 1052
- X-ray crystallography, 86, 202, 579
- X-ray diffraction, 10, 70, 117, 761, 763
 - grazing incidence, 70
- X-ray reflection, 578–580
 - critical angle, 578
- X-ray reflectometry, 578–583
- X-ray scattering, 714
- X-ray synchrotron source, 578, 583
- X-ray transform, 793
- xenobiotic, 869
- xenograft, 300
- xenon lamp, 314, 512
- xenopus, 383, 583, 654, 655, 657, 662, 665, 981
- XPS, 70
- YAG laser, 317
- yeast, 3
 - for biosensor, 872
 - genome, 927
 - protein, 949, 956
 - proteome, 947
- zeta potential, 751, 1019, 1020, 1103–1104
- zinc finger protein, 228
- zipper motif, 229
- zona pellucida, 1088
- zwitterion, 625
- zygote, 1073, 1085, 1087
 - transfection, 1086, 1088