

Biophysics

Roland Glaser

Biophysics

An Introduction

Second Edition

 Springer

Roland Glaser
Humboldt-Universität, Berlin
Germany
Roland.Glaser@hu-berlin.de

ISBN 978-3-642-25211-2 e-ISBN 978-3-642-25212-9
DOI 10.1007/978-3-642-25212-9
Springer Heidelberg Dordrecht London New York

Library of Congress Control Number: 2012936486

© Springer-Verlag Berlin Heidelberg 2012

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.

Printed on acid-free paper

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

“Was war also das Leben? Es war Wärme, das Wärmeprodukt formerhaltender Bestandlosigkeit, ein Fieber der Materie, von welchem der Prozeß unaufhörlicher Zersetzung und Wiederherstellung unhaltbar verwickelt, unhaltbar kunstreich aufgebauter Eiweißmolekel begleitet war. Es war das Sein des eigentlich Nicht-sein-Könnenden, des nur in diesem verschränkten und fiebrigen Prozeß von Zerfall und Erneuerung mit süß-schmerzlich-genauer Not auf dem Punkte des Seins Balancierenden. Es war nicht materiell und es war nicht Geist. Es war etwas zwischen beidem, ein Phänomen, getragen von Materie, gleich dem Regenbogen auf dem Wasserfall und gleich der Flamme.”

Thomas Mann, Der Zauberberg

“What then was life? It was warmth, the warmth generated by a form-preserving instability, a fever of matter, which accompanied the process of ceaseless decay and repair of albumen molecules that were too impossibly complicated, too impossibly ingenious in structure. It was the existence of the actually impossible-to-exist, of a half-sweet, half-painful balancing, or scarcely balancing, in this restricted and feverish process of decay and renewal, upon the point of existence. It was not matter and it was not spirit, but something between the two, a phenomenon conveyed by matter, like the rainbow on the waterfall, and like the flame.” (*Translated by H.T. Lowe-Porter, Penguin Books, 1985, pp. 275–276*)

Thomas Mann, The Magic Mountain

Preface to the Second English Edition

More than a decennium has passed since I finished the first English edition of this textbook – a long time for a rapidly developing science! A mass of original publications as well as reviews and books for each of the addressed topics can always be found on the website of the “Web of Knowledge.” Consequently, a full revision of all chapters was necessary, and a number of new results and topics had to be included.

The long time it took in writing this textbook, starting from the first German edition in 1971, reflects an important period of development of biophysics as a modern scientific discipline. It was extremely fascinating to not only observe the progress in biophysics but also to follow the ups and downs of the crucial aspects of its development. At first radiation biophysics dominates. Later one biocybernetics, discussion about entropy and negentropy, extended Onsager matrices of coupled flux equations, dissipative structures, types of kinetic catastrophes, the paradox of spontaneous protein folding, etc. were discussed in extension. All of these approaches and ideas have eventually been well fitted into the complex system of biophysics more or less according to their real importance.

It was not easy to decide on what really should be included in such an introductory textbook, what should be the length of the corresponding sections, how should the plethora of new facts and relations be arranged, and what a student should essentially know to be able to understand the complex framework of biophysics. The author is aware of the subjective nature of these decisions.

At present, biophysical research is mainly focused on molecular structures and processes. This has indeed proved important and helpful in the preparation of this textbook, as new results of molecular biophysics could be included in various sections of this edition. It should be noted, however, that the molecules are embedded in a definite medium. Osmotic and Donnan-osmotic effects control the biological reactions and functions of these molecules. Therefore, the basic figures of “classical” thermodynamics should not be forgotten. The Nernst equation as well as all other basic equations derived by famous physicists – Planck, Fick, Donnan, etc. – still valid and indispensable even in modern research. Therefore,

these are maintained in some sections and, in some parts, are compared with and are strongly oriented to actual applications.

The increasing environmental consciousness worldwide during the last decades also enforces an extensive research on possible influences of physical parameters, such as electromagnetic fields and radiation, on biological systems and human health. This has been discussed in the chapter on environment biophysics, leading to its extension and new structuring.

The enormous upturn of biological systems theory, caused by the abundance of analytical data as well as new methods of data storage and management, thanks to new computer techniques, required a full revision of the last part of this textbook. Earlier graphical methods to analyze data of compartment analyses are outdated and, hence, replaced by corresponding computer softwares. At the same time, some other approaches, for example, the “classical” graph theory of Leonard Euler, have become the new actuality in the age of fast computer techniques. This strongly promotes biological network analysis.

This new edition could not have been possible without the help of friends and colleagues. I would especially like to thank Peter Hegemann and Andreas Herrmann for their help with the chapters on molecular biophysics; Andreas Möglich for revising the sections on light; and Edda Klipp, Hanspeter Herzel, and Werner Ebeling for their help with the theoretical chapters. I also had many discussions and e-mail exchanges on problems of electrical properties of tissues and cells with my former coworker Jan Gimsa. The section of radiobiology has been written in collaboration with Jürgen Kiefer. I wish to express my deepest gratitude to all of them.

Last, but not least, I would like to thank Springer, especially Jutta Lindenborn for her support and, particularly, for her tireless efforts in the preparation of this new edition. The proposal to use color prints, I think, makes the figures more suggestive and comprehensible.

Berlin
March 2012

Roland Glaser

Preface to the First English Edition

When I started teaching biophysics to biology students at the Friedrich Schiller University of Jena in 1965, the questions arose: What actually is biophysics? What should I teach? Only one thing seemed to be clear to me: biophysics is neither “physics for biologists” nor “physical methods applied to biology,” but a modern field of science leading to new approaches of our understanding of biological functions.

Rashevsky’s book *Mathematical Biophysics* (1960), the classical approaches of Ludwig von Bertalanffy (1968), as well as the excellent book by Katchalsky and Curran *Nonequilibrium Thermodynamics in Biophysics* (1965) showed me new ways of looking at biological processes. Thus, I came to the conclusion that it would be worthwhile trying to integrate all these various physical and physicochemical approaches to biological problems into a new discipline called “biophysics.” The first German edition of this textbook, published in 1971, was developed from these considerations.

Meanwhile, I had moved from Jena to the Humboldt University in Berlin, where I organized courses for biologists specializing in biophysics. The idea was: Why should only physicists find their way to biophysics? Why not help biologists overcome the “activation energy” barrier of mathematics and physics and discover this fascinating discipline?

In Berlin, a special group was established (1970) in the Department of Biology with the aim of teaching biophysics. This led to a full university degree course of biophysics, which has developed successfully and attracts an increasing number of students today.

Consequently, my coworkers and I had the responsibility of organizing not only introductory courses to biophysics for biology students but also advanced courses in molecular biophysics, biomechanics, membrane biophysics, bioelectrochemistry, environmental biophysics, and various aspects of theoretical biophysics.

The evolution of this textbook in the following years was the result of these courses. Innumerable discussions with students, colleagues, and friends led to continuous refinement and modification of the contents of this book, resulting in

a second, third, and, in 1996, a fourth German edition. New topics were added and others updated or even deleted. The only sentences that remained unchanged were those of Thomas Mann at the beginning of the Preface.

The philosophy of this book is that biophysics is not simply a collection of physical approaches to biology but a defined discipline with its own network of ideas and approaches, spanning all hierarchical levels of biological organization. The paradigm of a holistic view of biological functions, where the biological system is not simply the sum of its molecular components but is rather their functional integration, seems to be the main concept of biophysics.

While it is easier to realize such an integrated view in a “one-man book,” this has, of course, the disadvantage that the knowledge and experience of many specialists cannot be incorporated. However, to a certain degree, this problem was compensated by discussions with colleagues and friends and by their continuous support over a period of more than three decades. Further problems are the selection of the topics to be included in the book and the emphasis placed on the different aspects, avoiding the underestimation of others. Although I have tried to balance the selection and emphasis of topics by looking at the development of biophysics over the last three decades, I am not sure that I have succeeded. Even if this is the case, this book will at least help to answer the question: What is biophysics? It provides a solid introduction to biophysics. For further reading, books and reviews are recommended at the end of each chapter. The extensive index at the end of the book ensures an easy orientation and will enable this book to be used as a reference work.

As mentioned above, this book is written primarily for biologists and biophysicists with a background in biology. Therefore, some basic knowledge of biology is required, but less knowledge of physics and mathematics is needed. It should encourage biologists to enter the field of biophysics and stimulate further research. The German editions have shown that physicists also can profit from reading this book.

This first English edition is not just a translation of the fourth German edition but is rather a fully revised fifth edition. For an author, it is impossible to translate his book without substantial rewriting and refining. All chapters have been more or less revised, and results which have been published since the last edition have been integrated. Many figures have been redrawn, while some are new; some altogether new chapters have also been included.

Last but not least, I wish to express again my sincere gratitude to all of my colleagues and friends, throughout the world, who helped me with all previous editions and especially with this English edition. Thanks are also extended to the staff of Springer-Verlag for encouraging me to write this English version and for correcting my imperfect English.

Berlin
July 2000

Roland Glaser

Contents

1	Nature and Subject of Biophysics	1
2	Molecular Structure of Biological Systems	5
2.1	Thermal Molecular Movement, Order and Probability	6
2.1.1	Thermodynamic Probability and Entropy	7
2.1.2	Information and Entropy	9
2.1.3	Biological Structures: General Aspects	14
2.1.4	Distribution of Molecular Energy and Velocity at Equilibrium	16
2.1.5	Energy of Activation, Theory of Absolute Reaction Rate	19
2.1.6	Kinds of Thermal Molecular Movement	26
2.2	Molecular and Ionic Interactions as the Basis for the Formation of Biological Structures	33
2.2.1	Some Foundations of Electrostatics	33
2.2.2	The Structure of Water and Ionic Hydration	39
2.2.3	Interaction of Water with Macromolecules	43
2.2.4	Ions in Aqueous Solutions, the Debye–Hückel Radius	47
2.2.5	Intermolecular Interactions	50
2.2.6	Structure of Proteins	56
2.2.7	Protein Folding and Protein Dynamics	59
2.2.8	Ampholytes in Solution, the Acid–Base Equilibrium	63
2.3	Interfacial Phenomena and Membranes	67
2.3.1	Surface and Interfacial Tensions	67
2.3.2	Orientation of Molecules at Phase Boundaries; Self-Assembly of Membranes	70
2.3.3	The Molecular Structure of Biological Membranes	74
2.3.4	Mechanical Properties of Biological Membranes	76
2.3.5	Electrical Double Layers and Electrokinetic Phenomena	80
2.3.6	The Electrostatic Structure of the Membrane	88

3	Energetics and Dynamics of Biological Systems	95
3.1	Some Fundamental Concepts of Thermodynamics	95
3.1.1	Systems, Parameters, and State Functions	96
3.1.2	Gibbs Fundamental Equation	99
3.1.3	Force and Motion	106
3.1.4	Entropy, Stability, and Stationary States	112
3.1.5	Stochastic Resonance and Noise-Enhanced Processes	120
3.1.6	Thermodynamic Basis of Biochemical Reactions	124
3.2	The Aqueous and Ionic Equilibrium of the Living Cell	127
3.2.1	The Van't Hoff Equation for Osmotic Pressure	127
3.2.2	Osmotic Pressure in Cells and Biologically Relevant Fluids	133
3.2.3	Electrochemical Equilibrium: The Nernst Equation	138
3.2.4	The Donnan Equilibrium: Basic Properties	143
3.2.5	The Donnan Equilibrium in Biological Systems	146
3.3	Phenomenological Analysis of Fluxes	150
3.3.1	The Flux of Uncharged Substances	150
3.3.2	Fluxes of Electrolytes	157
3.3.3	The Diffusion Potential	162
3.4	Membrane Transport and Membrane Potential	165
3.4.1	Channels and Pumps: The Variety of Cellular Transport Mechanisms	165
3.4.2	The Network of Cellular Transporters	168
3.4.3	The Membrane Potential	171
3.4.4	The Action Potential	176
3.4.5	Molecular Aspects of Membrane Transport	181
3.5	Electric Fields in Cells and Organisms	185
3.5.1	The Electric Structure of the Living Organism	185
3.5.2	Extracellular Electric Fields and Currents	187
3.5.3	Passive Electrical Properties of Tissue and Cell Suspensions	192
3.5.4	Single Cells in External Electric Fields	197
3.5.5	Manipulation of Cells by Electric Fields	202
3.6	Mechanical Properties of Biological Materials	207
3.6.1	Some Basic Properties of Fluids	208
3.6.2	The Viscosity of Biological Fluids	212
3.6.3	Viscoelastic Properties of Biomaterials	215
3.6.4	The Biomechanics of the Human Body	220
3.7	Biomechanics of Fluid Behavior	225
3.7.1	Laminar and Turbulent Flows	225
3.7.2	Biomechanics of Blood Circulation	228
3.7.3	Swimming and Flying	234
3.8	Allometric Considerations of Structure and Function	239

4 Physical Factors of the Environment	245
4.1 Temperature	246
4.2 Pressure	252
4.3 Mechanical Oscillations	254
4.3.1 Vibration	254
4.3.2 Sound	258
4.3.3 The Biophysics of Hearing	261
4.3.4 Infrasound	268
4.3.5 Ultrasound	269
4.3.6 Biophysics of Sonar Systems	274
4.4 The Static Magnetic Field	278
4.5 The Electrostatic Field	286
4.6 Low-Frequency Electromagnetic Fields	291
4.6.1 Physical Background and Dosimetry	291
4.6.2 Biological Effects and Biophysical Background	294
4.7 Radio- and Microwave Electromagnetic Fields	298
4.7.1 Physical Background and Dosimetry	298
4.7.2 Biophysical Aspects of High-Frequency Field Interaction	300
4.8 Visible and Nonvisible Optical Radiation	303
4.8.1 THz and Infrared: The Vibration-Inducing Frequencies	305
4.8.2 Visible Light: Processes of Excitation and Energy Transfer	307
4.8.3 Visible Light: Photobiological Processes	310
4.8.4 Ultraviolet: The Transition to Ionizing Radiation	314
4.9 Ionizing Radiation	317
4.9.1 Nature, Properties, and Dosimetry of Radiation	317
4.9.2 Primary Processes of Radiation Chemistry	320
4.9.3 Radiobiological Reactions	324
4.9.4 Some Aspects of Radiation Protection	327
4.9.5 Mathematical Models of Primary Radiobiological Effects	329
5 The Kinetics of Biological Systems	333
5.1 Some General Aspects of Systems Theory	334
5.1.1 Basic Equations of Kinetic Processes	334
5.1.2 General Features of System Behavior	338
5.1.3 The Graph-Theory as an Topological Approach to Describe Complex Network Systems	343
5.1.4 Systems with Feedback Regulation	346

5.2 Model Approaches to Some Complex Biological Processes	350
5.2.1 Models of Biochemical Reactions	350
5.2.2 Pharmacokinetic Models	355
5.2.3 Models of Propagation and Ecological Interactions	357
5.2.4 Models of Growth and Differentiation	362
5.2.5 Models of Origin and Evolution of Life	366
5.2.6 Models of Neuronal Processes	370
References	377
Subject Index	391

About the Author



Roland Glaser Born in Jena, Germany in 1935. Studied biology at the Friedrich Schiller University, Jena, Ph.D. in 1961, qualified as a lecturer in 1965. From 1958 to 1961, scientist at the Atomic Energy Agency in Berlin (research: aquatic radioecology), and from 1962 to 1965 scientist at the Institute of Cardiology, Academy of Science in Berlin (with a focus on ion transport). From 1965 to 1970 assistant professor in Jena, from 1970 to 2000 full professor of biophysics at the Humboldt University, Berlin (research: biophysics of cell surfaces, shape and ionic states of erythrocytes, electro-manipulation of cells). In 1977 vice-president and from 1981 to 1985 president of the Society of Physical and Mathematical Biology of the GDR, from 1990 to 1992 dean of the Faculty of Science at the Humboldt University.

Roland Glaser was a member of the UNESCO European expert committee on biophysics (1976–1988), since 1978 a member of various commissions of the council of IUPAB, council member of EBEA (1993–1996) and of the national commission on radiation protection at the German Federal Ministry of Environmental Protection (1992–1998 and 2008–2010). He has also worked as an emeritus adviser for various organizations in the field of electromagnetic safety.

This book is the result of over 35 years of teaching. Following the first German edition in 1971, the textbook was continuously revised and updated in line with his experience from teaching as well as innumerable discussions with students and

colleagues. The fourth German edition was published in 1996, and a fifth revised version was then published as the first English edition in 2001.

With this revised second English edition, the author reflects the enormous strides made in biophysics, especially in the fields of molecular and theoretical biophysics. While the book remains a basic textbook and provides the introductory background needed for fundamental training in biophysics, it also offers new insights into recent advances in biophysical knowledge.

List of Fundamental Constants and Symbols

The numbers in parentheses indicate equations in the text, where the symbols are explained or defined.

\equiv	identical
$=$	equal
$\stackrel{!}{=}$	equal by definition
\approx	approximately equal
\sim	proportional
A	Arrhenius-coefficient (Eq. 2.17)
A	affinity (Eq. 3.75)
A	area
a	chemical activity (Eq. 3.34)
\mathbf{B}	magnetic flux density (Eq. 4.17)
b	electrophoretic mobility (Eq. 2.84)
C	electric capacity (Eq. 3.204)
C^*	complex electric capacitance (Eq. 3.211)
C_l	clearance-constant (Eq. 5.29)
C	heat capacity (Eq. 2.65)
c	molar concentration
c_0	speed of light in vacuum = $2.998 \cdot 10^8 \text{ m s}^{-1}$
D	diffusion coefficient (Eq. 3.131)
E	energy (general expression)
\mathbf{E}	electric field strength (Eq. 2.44)
e	basis of natural logarithm = 2.71828
e	absolute amount of charge on electron = $1.60218 \cdot 10^{-19} \text{ C}$
\mathbf{F}	mechanical force
F	Faraday = $9.6485 \cdot 10^4 \text{ C val}^{-1}$
F	Helmholtz free energy (Eq. 3.22)
f	symbol for an arbitrary function
f	generalized coefficient of friction (Eq. 3.52)

<i>f</i>	activity coefficient (Eq. 3.34)
<i>G</i>	Gibbs free energy (Eq. 3.23)
<i>G</i>	electrical conductivity (Eq. 3.189)
<i>g</i>	specific conductivity (Eq. 3.209)
<i>g</i>	osmotic coefficient (Eq. 3.102)
<i>H</i>	enthalpy (Eq. 3.21)
H	magnetic field strength (Eq. 4.17)
<i>h</i>	Planck's constant = $6.626 \cdot 10^{-34}$ J s = $4.136 \cdot 10^{-15}$ eV s
<i>I</i>	sound intensity (Eq. 4.12)
<i>I</i>	information (Eq. 2.6)
<i>I</i>	ionic strength (Eq. 2.56)
<i>I_A</i>	second moment of area (Eq. 3.232)
<i>I_P</i>	polar second moment of area (Eq. 3.233)
i	unit vector in <i>x</i> -direction
<i>j</i>	imaginary unit = $\sqrt{-1}$
j	unit vector in <i>y</i> -direction
<i>j</i>	electric current density (Eq. 4.23)
J	flux (Eq. 3.125)
<i>J</i>	unidirectional flux in kinetic equations
<i>K_p</i>	equilibrium constant of isobaric chemical reactions (Eq. 3.71)
<i>K</i>	bending ($K = 1/R$), (Sect. 3.6.4)
<i>k</i>	Boltzmann's constant = $1.380658 \cdot 10^{-23}$ J K ⁻¹ = $8.6174 \cdot 10^{-5}$ eV K ⁻¹
<i>k</i>	rate constant
<i>L_i</i>	phenomenological coefficient relating a flow to a force (Eq. 3.49)
<i>L</i>	decibel intensity of sound (Eq. 4.13)
<i>l</i>	distance, length
<i>M</i>	moment of force (Eq. 3.231)
<i>m</i>	mass
<i>N</i>	Avogadro's number = $6.0221367 \cdot 10^{23}$ mol ⁻¹
<i>n</i>	number of particles, individuals etc.
<i>P</i>	mathematical probability (Sect. 2.1.2)
<i>P</i>	permeability coefficient (Eq. 3.133)
<i>P</i>	electrical power density (Eq. 4.23)
<i>p</i>	pressure
<i>Q</i>	heat
<i>q</i>	electric charge
<i>R</i>	molar gas constant = 8.314472 JK ⁻¹ mol ⁻¹
<i>R</i>	radius of curvature ($R = 1/K$), (Eq. 3.227)
<i>R</i>	resistance coefficient relating a flow to a force (Eq. 3.50)
<i>R</i>	Ohm's resistance (reactance), (Sect. 3.5.13)
Re	Reynold's number (Eq. 3.234)
<i>r</i>	radius, radial distance
<i>r</i>	Donnan ratio (Eq. 3.123)
<i>S</i>	entropy (Eqs. 2.4 and 3.10)

T	temperature
t	time
U	internal energy (Eq. 3.9)
V	volume
\bar{V}	partial molar volume (Eq. 3.8)
\mathbf{v}	velocity
W	thermodynamic probability (Eq. 2.5)
W	work (Eq. 3.9)
\mathbf{X}	generalized force (Eq. 3.42)
x	coordinate in an orthogonal system
x	mole fraction (Eq. 3.35)
Y	Young's modulus (Eq. 3.226)
Y^*	electric admittance (Eq. 3.201)
y	coordinate in an orthogonal system
z	coordinate in an orthogonal system
z_i	number of charges
α	electrical polarizability (Eq. 2.48)
β_T	isothermic compressibility (Eq. 2.66)
γ	velocity gradient or shear rate (Eq. 3.220)
γ	surface tension (Sect. 2.3.1)
δ	difference of length
Δ	sign, indicating a difference between two values
ε	mechanical strain (Eq. 3.225)
ε	dielectric constant or permeability number (Eq. 2.41)
ε_0	dielectric permittivity of vacuum = $8.854187817 \cdot 10^{-12} \text{ C V}^{-1} \text{ m}^{-1}$
ζ	electrokinetic potential (Eq. 2.85)
η	viscosity (Eq. 3.221)
κ	Debye–Hückel constant (Eq. 2.55)
λ	thermal conductivity (Eq. 4.1)
λ	wavelength
μ	magnetic permeability (Eq. 4.17)
μ_0	magnetic permeability of vacuum = $1.2566370 \cdot 10^{-6} \text{ V s A}^{-1} \text{ m}^{-1}$ (Eq. 4.17)
μ	electric dipole moment (Eq. 2.47)
μ_i	chemical potential of the component i (Eq. 3.33)
$\tilde{\mu}_i$	electrochemical potential of the salt i (Eq. 3.41)
ν	stoichiometric number (Eq. 3.1.65)
ν	kinematic viscosity ($\nu = \eta/\rho$)
ν	frequency in Hz ($\nu = \omega/2\pi$)
ξ	degree of advancement of a chemical reaction (Eq. 3.73)
π	osmotic pressure (Eq. 3.30)
ρ	density
ρ	charge density in space (Eq. 2.52)
σ	Stefan–Boltzmann constant (Eq. 4.3)
σ	mechanical stress (Eq. 3.224)

σ	entropy production (Eq. 3.63)
σ_0	surface charge density (Eq. 2.86)
σ	Staverman's reflection coefficient (Eq. 3.108)
τ	time constant
τ	sheer stress (Eq. 3.222)
Φ	Rayleigh's dissipation function (Eq. 3.64)
φ	fluidity ($\varphi = 1/\eta$) (Sect. 3.6.1)
χ	magnetic susceptibility (Eq. 4.18)
ψ	electrical potential (Eq. 2.43)
ω	angular frequency ($\omega = 2 \pi \nu$)
ω	mobility factor (Eq. 3.52)

Chapter 1

Nature and Subject of Biophysics

The subjects of Biophysics are the physical principles underlying all processes of living systems. This also includes environmental biophysics, which represents physical influences on physiological functions.

Biophysics is an interdisciplinary science somewhere between biology and physics – as may be concluded from its name – and it is furthermore connected to other disciplines, such as mathematics, physical chemistry, and biochemistry. The term “biophysics” was first used in 1892 by Karl Pearson in his book “The Grammar of Science.”

Does biophysics belong to biology, or is it a part of physics? Biology, by definition, claims to be a comprehensive science relating to all functions of living systems. Hence, biophysics, like genetics, biochemistry, physiology etc., should be considered as a special part of biology. This view has not remained undisputed by physicists, since physics is not confined to subjects of inanimate matter. Biophysics can be considered, with equal justification, as a part of physics. Especially today, when the boundaries between classical disciplines are no longer fixed, it would be futile to try to balance those aspects against each other. Biophysics appears to be one of the best examples of an interdisciplinary science.

The delimitation of biophysics from clearly unrelated areas appears to be much easier than its definition. Biophysics, for example, is by no means some sort of conglomeration of various physical methods that can be applied to solving biological problems. The use of a magnifying glass, the most primitive optico-physical instrument, for example, has just as little to do with biophysics as the use of the most up-to-date optical or electronic measuring instruments. Biophysical research, of course, requires modern methods, just as other fields of science do. The nature of biophysics, however, is actually defined by the scientific problems and approaches rather than by the applied methods.

Biophysical chemistry and bioelectrochemistry can be considered as specialized subareas of biophysics. Medical physics, conversely, is an interdisciplinary area which has its roots in biophysics but has ramifications of far-reaching dimensions, even with medical engineering.

Biophysical thought can be traced back to early phases of philosophical speculation on nature, i.e., to antiquity. This applies to the earliest mechanistic theories of processes of life and insights into their dynamics, for example of Heraclitus in fifth century B.C. The promotion of scientific research in the Renaissance also includes biophysical considerations. Leonardo da Vinci (1452–1519), for example, investigated mechanical principles of bird flight, using them as models for engineering design. A remarkably comprehensive biomechanical description of functions, such as the mobility of limbs, bird flight, the movement involved in swimming, etc., was given in the book of Alfonso Borelli (1608–1679) “*De motu animalium*” published in Rome, as early as 1680. The same Borelli founded a school in Pisa of *iatro-mathematics* and *iatro-physics* in which the human body was perceived as a mechanical machine, and where attempts were made to draw medical conclusions from that perception (ἰατρός – Greek term for physician). Iatro-physics has often been considered a mechanistic forerunner of medical biophysics.

Throughout the history of Biophysics there have been attempts to apply the actual state of physics to understand the processes of life. Even though early considerations were based on mechanical models, later other models including electrical, thermodynamic, and quantum mechanical were used. In fact, there is a mutual interaction of physics and biology. Reference can be made, in this context, to the frog experiments undertaken by Luigi Galvani (1737–1798). The physics of electricity was thus studied in direct relationship with phenomena of electrophysiology. Worth mentioning is the famous controversy between Luigi Galvani and Alessandro Volta (1745–1827) about the so-called “*elettricità animale*,” which had serious personal consequences for both.

As soon as electromagnetic fields were discovered and technically generated, the French physicist Jacques Arsène d’Arsonval applied them for therapeutic purposes, and in 1891 he published the paper “*Action physiologique de courants alternatives*.” Similar investigations were performed by Nikola Tesla in the USA. In 1906, the German physicist Karl Franz Nagelschmidt coined the term *diathermy* for this effect.

It is well known that medical observations played a role in the discovery of the first law of thermodynamics by J. R. Mayer (1814–1878). Calorimetric studies of heat generation of mammals were conducted in Paris by A. L. Lavoisier (1743–1794) and P. S. de Laplace (1749–1827) as early as about 1780. Reference should also be made, in this context to the investigations of Thomas Young (1773–1829), and later Hermann von Helmholtz (1821–1894) on the optical aspects of the human eye and on the theory of hearing. These activities added momentum to the development of physiology which thus became the first biological platform for biophysics.

The development of physical chemistry at the beginning of the twentieth century was accompanied by application of these discoveries and knowledge to understanding the various functions of living cells, including osmoses, membrane potential etc. There have also been many instances in which biologically induced problems had stimulating effects upon progress in physics and physical chemistry. Brown’s movement,

discovered in pollen grains and subsequently theoretically explained by A. Einstein, is an example. Research on osmotic processes, as well, was largely stimulated by the botanist W. Pfeffer. The temperature dependence of rate constants of chemical reactions was initially formulated in terms of phenomenology by S. Arrhenius (1859–1927), and has, ever since, been applied for a great number of functions of life, including phenomena as sophisticated as processes of growth. Studies of physico-chemical foundations of cellular processes have continued to be important in biophysical research, especially, after the introduction of the principles of nonequilibrium thermodynamics. In particular, biological membranes, as highly organized anisotropic structures, are always attractive subjects for biophysical investigations.

A decisive impetus has been given to biophysical research through the discovery of X-rays and their application to medicine. It was attributable to close cooperation between physicists, biologists, and medical scientists which paved the way for the emergence of radiation biophysics, which not only opened up possible new approaches to medical diagnosis and therapy but also made substantive contributions to the growth of modern molecular biology.

A cornerstone in the development of biophysics was the lecture of Erwin Schrödinger in 1943, and his subsequently published book “What is Life?” This opened up the discussion of thermodynamics of living systems leading to the expansion of classical thermodynamics to cover nonequilibrium systems with nonlinear behavior. The books of A. Katchalsky, I. Prigogine, H. Haken and many others, introduced elements of thermodynamics and nonlinear systems analysis into biophysics. This led to the development of theoretical biophysics and furthermore to systems theory. As a pioneer Ludwig von Bertalanffy must be mentioned, considering the living organism as an open system and coining the German term *Fließgleichgewicht* (steady state) for this state. The first edition of his book “General System Theory” appeared in 1968. In 1948, Norbert Wiener’s book “Cybernetics: or Control and Communication in the Animal and the Machine” led to the birth of *biocybernetics*, which today is merged with *computational neuroscience*.

This brief review of the history and development of biophysics allows us to draw the following conclusions about its nature and relevance: Biophysics seems to be quite a new branch of interdisciplinary science, but in fact, biophysical questions have always been asked as long as science has speculated about the processes involved in living systems. Biophysics relates to all levels of biological organization, from molecular processes to ecological phenomena. Hence, all the other biological subareas are penetrated by biophysics, including biochemistry, physiology, cytology, morphology, genetics, systematics, and ecology. Many aspects of biophysics are the basis of health protection measures.

Biological processes are among the most intricate phenomena with which scientists find themselves confronted. It is, therefore, not surprising that biologists and other scientists have repeatedly warned against schematism and simplifications. Such warning is justified and is a permanent reminder to the biophysicist of the need for caution. Yet, conversely, there is no reason to conclude that biological

phenomena are too sophisticated for physical calculations. Despite the fact that at present we are not able to explain all biological reactions, no evidence has ever been produced that physical laws are no longer valid when it comes to biological systems.

Chapter 2

Molecular Structure of Biological Systems

It is the intention of this section to familiarize the reader with some specific physical properties of biological systems on the molecular level. The overriding theme of this section is the controversy of thermal fluctuation against the forces of molecular orientation and organization.

Two kinds of physical behavior meet on the molecular level of biological structures: On the one hand, there are the characteristic properties of *microphysical* processes, based on the individual behavior of single small particles like atoms, molecules, or supramolecular structures. These processes are mostly stochastic. On the other hand, there are reactions which resemble *macrophysical* properties, the kind of behavior of “large” bodies. *Macrophysics* is ruled by the laws of classical physics, as for the example of classical mechanics. Our daily experiences with macrophysical systems teach us that their behavior is generally deterministic.

To explain this difference, let us consider a simple mechanical wheelwork. The knowledge of its design and construction allows a precise prediction of the behavior of the system. This prediction is based on the laws of classical mechanics. In contrast to this, a chemical reaction with a small number of molecules in a homogeneous phase depends on stochastic collisions of the individual molecules with each other. Since this process is stochastic, it is predictable only in a statistical way.

This stochastic behavior of molecular systems can be transformed into a deterministic one, if the number of participating stochastic events is large, or if the degrees of freedom of the single reactions are extremely limited. The increase of stochastic events can be realized either by an increasing number of participating molecules, by enlarging the volume for example, where the reaction takes place, or by an increase of the time interval of observation. This consideration indicates an interesting interplay between volume, time constants, and reliability of a biochemical reaction.

The limitation of the degree of freedom of a biochemical reaction is realized by a property of the system which is called *anisotropy*. In contrast to *isotropic* systems, like simple solutions, in anisotropic systems the mobility of molecules in various directions is not identical, but is restricted in some directions, and promoted in

others. This, for example, is the case for enzymatic reactions, where the participating enzymes are oriented in membranes, or if the reactions of charged or polar reactants occur in strong electric fields of electrical double layers.

In many fields the biological organism works as an amplifier of microphysical stochastics. A molecular mutation, for example, leads to a reaction chain, which finally ends with a phenomenological alteration of the organism. Or, as another example: a few molecular events in the pigments of an optical receptor can lead to perception and to reaction in behavior.

During the first steps in considering molecular mechanisms of biological systems, a further aspect is taken into consideration. Unfortunately, biologists often ignore that a qualitative jump has to be considered in the transition from the “visible” macrophysical structures, to the microphysical systems such as atoms or molecules. This includes not only the above-mentioned transition from the deterministic behavior of macroscopic systems to the stochastic behavior of single molecules, but many further aspects as well.

The biologists, for example, must acknowledge that the term “structure” now receives a new meaning. The visible “biological structure,” as known in the fields of anatomy, morphology, and histology, now appears as concentration profiles, or as systems of electric charges, or electromagnetic fields. Instead of visible and measurable lengths, diameters, or distances, as common in the visible world, in the microphysical world so-called *effective parameters* are used. These sorts of parameters are exactly defined, and they can be measured with arbitrary exactness, but they do not correspond to some visible boundaries. A single ion, for example, has no diameter in the sense of the diameter of a cell, or a cell nucleus, which can be measured by a microscopic scale. In the following sections we will define effective parameters like crystal radius, hydration radius, and Debye–Hückel radius, which are very important parameters for functional explanations. We will consider this aspect in detail in Sect. 2.1.3.

It is not the intention of this book to describe the topics of molecular biology. However, the theoretical foundations and principles will be explained to make possible a link between structure and function at the molecular level and current biological thinking in these dimensions.

2.1 Thermal Molecular Movement, Order and Probability

In this section, the biophysics of molecular organization of biological systems will be discussed in the context of processes of thermal movements where a statistical mechanics approach can be applied. Stochastic phenomena are of great importance in molecular biophysics. Here, we are immediately confronted with the dialectics of arbitrary distribution on the one hand, and organized order on the other hand. This also touches upon problems of biological self-organization and stability of the resulting structures.

2.1.1 Thermodynamic Probability and Entropy

In 1854, Rudolf J. E. Clausius introduced the *entropy* (S) as a parameter of phenomenological thermodynamics, and defined it as the heat, added to a system in a reversible way in relation to the temperature (see Eq. 3.10). Later, in 1894, Ludwig Boltzmann used this parameter in the framework of statistical thermodynamics. In these circumstances, the entropy, and in context to this, the second principle of thermodynamics becomes more imaginable. Entropy appears as a kind of measure of disorder, or as a degree of random distribution, i.e., of missing order. The correlation between order and probability and, as will be explained later – information – is of great importance for the understanding of the principles of biological organization.

Let us start with the assumption that the entropy is a measure indicating the degree of randomization of a given distribution. We will consider a system of maximal entropy as a system in maximal disorder. Furthermore, let us demand that the entropy be an *extensive* parameter. Therefore, like volume, or mass, but in contrast to the *intensive* parameters such as for example temperature or density, the entropies S_1 and S_2 of two systems can be added, if these systems come together:

$$S_1 + S_2 = S \quad (2.1)$$

How can we now define a parameter, which indicates the degree of randomization or, on the contrary, a degree of disorder? What does order of organization mean? Of course, our daily experience shows that an ordered system spontaneously transforms into a disordered one, but not vice versa. This, actually, is the consequence of the second principle of thermodynamics.

Let us consider a very simple structure, just the distribution of four distinguishable spheres on two compartments of a box (Fig. 2.1). Let each of these spheres, independently of the three others, just by chance fall into one or the other compartment of the box. All of the 11 possibilities of the distribution, as indicated in Fig. 2.1, therefore, have the same degree of probability, because the probability of each sphere individually, to fall into compartment 1 or into compartment 2 is equal. Summarizing the patterns of distribution shows that there is only one way to

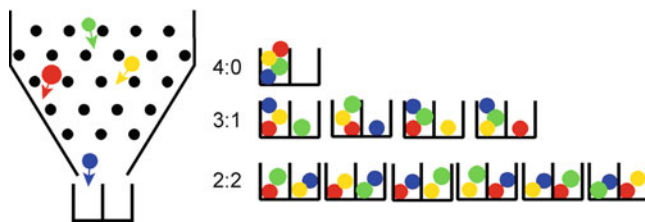


Fig. 2.1 All possibilities of the statistical distribution of four distinguishable spheres in two compartments of a box

realize the distributions 0:4 and 4:0. In contrast, there are four ways to realize the distributions 3:1 and 1:3, and, finally, six ways for equal distribution: 2:2.

Let us now ignore the fact that the spheres are distinguishable. Let us simply ask: How large is the probability that just by stochastic distributions one of the relations 4:0, 3:1, 2:2, 1:3, or 0:4 occur? Apparently, the probability of any kind of distribution will be larger, if it can be realized by a larger number of ways. The distribution mode 2:2, for example, is 6 times more likely, than the distribution 4:0, or 0:4. The number of ways which lead to the realization of a definite situation, in fact, seems to be a measure of the probability of the occurrence of it. We will designate this number of ways by the parameter W which we will call *thermodynamic probability*. The amount of W can be at least 1 and at maximum ∞ , in contrast to the mathematical probability (P), which we will use in Sect. 2.1.2, and which ranges between 0 and 1.

Now, we come to the following sequence of conclusions: If W really is a measure of the probability of getting a definite distribution, and if an increase of the degree of order is the most uncertain result of a stochastic distribution and finally, if the entropy (S) is a parameter, indicating the degree of disorder – than S should be a function of W . If two situations with relative probabilities W_1 and W_2 are connected together, then the probability of this combined situation results from the product ($W_1 \cdot W_2$). Using Eq. 2.1, this means:

$$S = f(W) = S_1 + S_2 = f(W_1) + f(W_2) = f(W_1 \cdot W_2) \quad (2.2)$$

This demand is met by the logarithmic function:

$$\ln A + \ln B = \ln(A \cdot B) \quad (2.3)$$

Hence entropy is proportional to the logarithm of W :

$$S = k \ln W \quad (2.4)$$

This is the Boltzmann equation of entropy. Boltzmann's constant k was defined as a universal constant later by Max Planck. It must have the same unit of measurement as entropy, and is as follows:

$$k = 1.380658 \cdot 10^{-23} \text{ J K}^{-1} = 8.6174 \cdot 10^{-5} \text{ eV K}^{-1}$$

This explanation was just based on the simplest experiment where four spheres were distributed randomly over two compartments. One step toward serious thermodynamics can be taken, considering for example the compartments of this box as molecules of a system, and the spheres, as quanta of energy, distributed among them. This complication, of course, means a transition of handling with larger numbers. If the number of elements and classes are increased, W cannot be

evaluated just by simple trial. It is possible to calculate this value using the following equation:

$$W = \frac{n!}{n_1! \cdot n_2! \cdot n_3! \cdot \dots \cdot n_m!} \quad (2.5)$$

where n is the total number of all elements in the system (in case of Fig. 2.1 – the total number of spheres: $n = 4$); n_i (for $i = 1 \dots m$) is the number of elements in each class of state (this means, the number n_1 in compartment 1 and n_2 in compartment 2); and m is the number of classes of state (namely: number of compartments in the box).

2.1.2 Information and Entropy

In 1948, C. E. Shannon introduced a parameter which in technical information theory has been proved as a useful measure of information content of a message. The information (I) of a message depends on the effort required to guess it by a highly systematic system of questions. Hence, information is some sort of degree of the actuality of a message.

It is not difficult to guess the result of the toss of a coin, since there are only two possibilities of equal probability. To guess a certain card in a full deck of playing cards is much more difficult. In this case, a much greater uncertainty factor has to be taken into account. Using a more systematic approach, a large number of yes-no questions have to be answered. Hence, the information content of a particular playing card is higher than that of a tossed coin. Should a deck consist of cards which are all the same, and should this be known to the challenged person, guessing will not make sense at all. The information content of each of these cards is zero. The probability by which possibilities are turned into reality, consequently, seems to become a measure of information.

In contrast to thermodynamics, in the information theory the mathematical term of probability (P) is used which is defined as follows:

$$P = \frac{\text{number of favorable cases}}{\text{greatest possible number of cases}}$$

On average, coins tossed a 100 times will land with heads up in 50 instances. Hence, the probability of heads facing up may be expressed by:

$$P = \frac{50}{100} = \frac{1}{2}$$

Conversely, the probability of throwing a “six” with some dice is only $P = 1/6$, whereas the probability of throwing one of the three even numbers would be higher: $P = 3/6 = 1/2$.

Whereas the thermodynamic probability (W) is always larger than 1, (cf. Sect. 2.1.1), the value of the mathematical probability lies between 0 and 1 ($0 \leq P \leq 1$). In this context, $P = 0$ means an impossibility, while $P = 1$ expresses absolute certainty.

The logical conclusions which led to the derivation of the Boltzmann equation (cf. Sect. 2.1.1) are the same as those on which the Shannon relation is based. Information (I) is a function of mathematical probability (P):

$$I = f(P)$$

The condition for the function f again, is satisfied by the logarithmic function, since here too, the multiplication rule for the calculation of probabilities must be valid (cf. Eq. 2.3). To obtain positive values of information, considering that $P \leq 1$, the negative value of the logarithmic function must be used. The information of a single event therefore is:

$$I = -K \ln P \quad (2.6)$$

This is the *Shannon equation* of information theory. The unit of I is determined by the unit of the always positive factor K . The bit (*binary digit*) is most commonly used. It expresses the number of binary yes-no decisions that are needed to determine a given message. For example, the one side of the coin can be guessed by one single decision, its information value, consequently, is 1 bit. Five binary decisions will be sufficient to guess a card from a deck. Hence, the information value of one card is 5 bits. The factor $K = 1/\ln 2 = 1.443$ must be used to calculate I in bits. In the information theory the logarithm to the base 2 (\log_2) is occasionally used:

$$I = -1.443 \ln P = -\log_2 P \quad (I \text{ in bit}) \quad (2.7)$$

A message usually consists of elements like symbols, words, or other structures of a set defined in advance. This concerns not only the letters of a word or the words in a written text, but the same approach can be applied also for various cases of biological information storage and transfer. So, for example, the information content of a DNA molecule is coded by the sequence of nucleic acids; the information content of a protein is given by their amino acids. In a similar way the information content of time-dependent structures like nervous signals can be evaluated too. In this case discrete patterns of neural spike trains are considered as elements of the message.

Assuming for example that a mammalian DNA molecule consists of 15,000 pairs of nucleotides and the four possible types of nucleoside bases have an equal probability of occurrence, then the information content of each single nucleotide

will, consequently, have a value of 2 bits. The information capacity of this DNA molecule therefore amounts to 30,000 bits.

In fact, however, in a DNA molecule the probability of the occurrence is not equal for all nucleotides. In this case, the information content of the whole message can be expressed by the sum of the evaluated probabilities (P_i) of the single elements (i) by the following equation:

$$I = -K \sum_i P_i \ln P_i \quad (2.8)$$

Estimates of information capacity in the whole genome of an organism in this way range up to about 10^{10} bit. The amount of actually stored information is even lower, if the redundancy in the storage of genetic information which is required for the protection of the information is taken into account.

It seems important to emphasize that the Shannon Information (I) as evaluated here, is the so-called *syntactic* information. It allows for example important conclusions about the maximum storage capacity of a DNA molecule of the whole genome. This Shannon Information (I), however, has nothing to do with the “information content” in the everyday use of this word, the so-called *semantic information*.

The difference between the syntactic and the semantic character of information can be illustrated best by the following example: A book, consisting just of stochastically jumbled letters, or even words according to Eq. 2.8 have a larger amount of syntactic information than such with a meaningful text, because in the stochastic sequence the probability of the occurrence of elements is lower than in an organized one. If P_i becomes smaller, its negative logarithm, and therefore I gets larger. Knowing some characters of a reasonable word, for example, the rest of them can usually easily be guessed – the probabilities of their occurrence therefore become larger, its content of syntactic information diminishes. If the same number of characters is mixed stochastically, guessing is more complicated – information is larger.

Transferred to the biological situation: The Shannon, i.e., the syntactic information of a polynucleotide with an arbitrary sequence is larger than that of a DNA with the same molecular length. In contrary to the *syntactic* information, the *semantic* information of a reasonable book, of course, is much larger than a book containing stochastically selected characters or even words. The same refers to the biological DNA in contrast to a polynucleotide with no biological function.

Does this mean that it is impossible to quantify biologically important information? Does it mean that the information concept is not applicable to biological systems at all? In fact, the semantic aspects of communication are irrelevant to most engineering problems and therefore not included in the definition of Shannon’s I . Despite many attempts, quantification of semantic information has not yet been achieved.

Nevertheless, even the use of the syntactic information properties of biomolecules and nervous processes has proved to be a useful tool in modern molecular biology.

The enormous increase of the amount of data on the primary structure of biological macromolecules, especially DNA sequences, as well as data of neurophysiological recordings nowadays required mathematical methods for their evaluation, comparison, and storage. For this, bioinformatics was developed as a special discipline, based on these considerations.

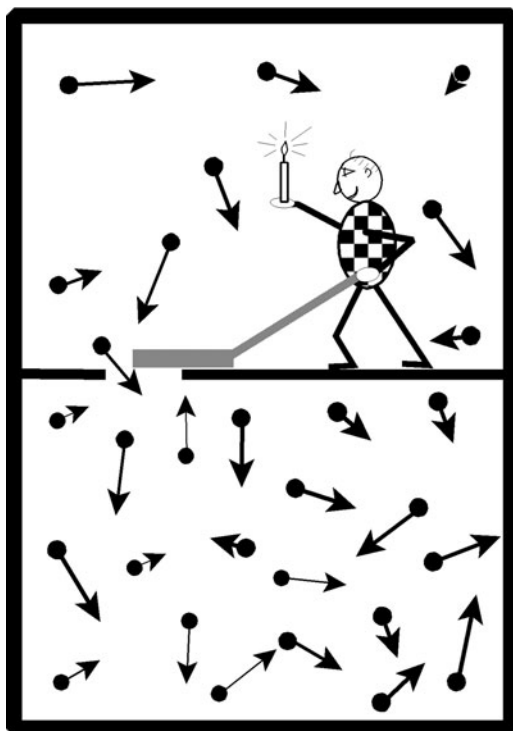
Formally, the Boltzmann's constant (k) can be used in the Shannon equation (Eq. 2.6). In this case information is obtained formally in entropy units (JK^{-1}). Therefore the term "entropy" as Shannon-Entropy in bioinformatics was frequently used in the sense of information. This formal interconnection of Shannon's definition of information with Boltzmann's entropy triggered off a discussion about the relation between these two parameters.

The starting point for this discussion was the second principle of thermodynamics, predicting that isolated systems spontaneously try to reach a state of maximum disorder. The growth and the existence of a living system on the contrary is possible only by decreasing or at least by the conservation of entropy. Erwin Schrödinger (1944) made the frequently quoted statement: "The living system feeds on negative entropy." This is the reason, why sometimes the term "negentropy" was used.

The interconnection of information with entropy in terms of thermodynamics may be illustrated best by a thought experiment conceived by James Clerk Maxwell in 1881 and still discussed today (Fig. 2.2). Maxwell proposed a room, which is connected with another by an opening. This opening can be closed by means of a slide. Both rooms are filled with a gas which is, in the beginning, in equilibrium, for example of equal pressure and temperature. An intelligent creature, later called "Maxwell's demon," is able to handle the slide between the two rooms with ease. This "demon" can observe accurately the direction and the velocity of the molecules in his room. Velocity and direction of these particles in the beginning are statistically distributed. If a particle in the upper room flies accidentally toward the opening, the demon opens the slide to let the particle pass. As a result of such sorting the pressure in the lower room would rise.

The demon could also take another approach. For example, he could separate fast from slow particles. In this case, a difference in the temperature between the two rooms would occur. In both cases, the entropy of the whole system would be reduced and energy might be generated by an appropriate turbine. The result would be a "*perpetuum mobile*," a perpetual motion machine of the second order, as it would contradict the second principle of thermodynamics.

This apparent contradiction subsequently was the subject of a large number of scientific and philosophical publications. The problem was finally resolved by the following consideration: The "demon" requires information to carry out the sorting. He collects this information by "watching" the molecules. In order to "see" he needs light. For this, the demon in Fig. 2.2 symbolically carries a lit candle. Yet, a body will only be able to emit light in a state of nonequilibrium relative to its environment. This, however, contradicts the equilibrium condition at the beginning of the experiment. The same would apply to any other approach to acquisition of information. This resolves the apparent contradiction to the second law of thermodynamics.

Fig. 2.2 Maxwell's demon

Why do we discuss this thought experiment here, if it is clear that it will not work as a perpetual moving machine? What has this to do with biophysics? In fact, independent of this discussion about the virtual violation of the second law of thermodynamics, Maxwell's demon demands particular interest in biophysics because of its analogy to various functions of living systems. The living cell, too, reduces its entropy at the expense of its environment, using information processing. Yet, in the latter instance it is not the energy of statistical fluctuations, which is used in this case. The biological system selects such molecules from its environment, which are rich in free Gibbs energy of formation and correspondingly, with a low content of entropy. It uses this energy and extrudes molecules with lower free energy and larger entropy. The basic information for this process of selection, in other words, the "software" for this process, is stored in the structure information of the proteins, which are responsible for the recognition of these molecules, and eventually for their metabolism. These proteins get this sort of semantic information during the process of synthesis via the RNA, from the DNA of the genome.

This example again raises the question of the relation between semantic and syntactic information: what is the threshold value of information that is required to control the processes of living systems? Or translated into the language of modern computer science: how large must the simplest software for this sort of a biological

Maxwell demon be? What is the threshold information that carries out not only the metabolic function of the primordial organism but additionally its replication? How did the first accumulation of information in a molecule occur? Purely accidental combination of monomers to build their first functional macromolecule must be ruled out. The probability for this occasion is too low by far. Today, so-called *probiotic evolution* is assumed, i.e., chemical selection of developing polymers even before the stage of the biopolymer (see also Sect. 5.2.5).

Further Reading

Shannon and Weaver 1962; Bioinformatics in molecular biology: Kauffman 1993; Strait and Dewey 1996; Yockey 2005; Lesk 2002; Maxwell's demon: Leff and Rex 1990.

2.1.3 Biological Structures: General Aspects

In the previous section we introduced expressions like *order*, *structure*, and *organization* and discussed them in context with entropy and information, as well as with the statements of the second law of thermodynamics. This touches on a set of questions which are of central interest in biophysics and which will be mentioned in many sections of this textbook. Therefore it is necessary at this point, to explain some basic definitions and principal ideas.

What, really, is a *structure*? To the biologist, the term “structure,” usually is related to the macroscopically or microscopically visible organization of an organism. This means, for example, the structure of an animal skeleton, structure of a cell, of a mitochondrion, etc. The term “molecular structure” already lies outside the limits of this view. It refers to a certain arrangement of atoms, without defined contours, which can be described just by means of wave mechanics. The same applies for the concentration profile of an electrical double layer (Fig. 2.43), and particularly for so-called “time structures,” namely special time courses, like oscillations of a biological system (Figs. 5.3 and 5.16), like the shape of an electrocardiogram (Fig. 3.38), or like the sonogram of a bat's cry (Fig. 4.19). This means that the definition of the term “structure,” which is used in biophysics has to be broader than that of the morphologists and cytologists. It must include these structures as well as those of metabolic networks, ecosystems, or others.

The best, and generalized definition of this term is given by the set theory of mathematics. Therefore: *a system is an aggregate of elements with certain interrelations between them*. The totality of these interrelations is called the *structure* of the system. This definition does not prescribe at all, what kind of elements, and what kind of interrelations these are. It is applicable to all kinds of systems including biological systems and structures. In biophysics, we are interested especially in *dynamic systems*, i.e., in such systems, the interrelations between their elements are *interactions*. In contrast to this, in *static systems* the elements have no interaction at all, but are just interrelated by formal *relations*. Examples for static

systems, are the system of natural numbers in mathematics, or the system of animal species or of plants in biology.

The elements of a metabolic network are the metabolites, and the interrelations between them, i.e., their interactions, are the steps of biochemical reactions. Correspondingly, an ecosystem is to be considered an interaction of individuals and populations depending on abiotic conditions.

Biological macromolecules can be considered as systems too. In this case, the following levels or organization occur depending on what will be assumed as their elements.

- *Primary structure* – linear sequence of monomers (= elements) in the linear molecular chain (= system). For example, a polypeptide chain: ...serine-alanine-lysine-arginine-.
- *Secondary structure* – positioning in space of monomers (= elements) in a part of the molecule (= system) relative to each other. For example, the α -helix, or the β -sheet structure of an amino acid sequence in a protein.
- *Tertiary structure* – position in space of molecular regions of homogeneous secondary structures (= elements) in a molecule (= system). For example, intramolecular coordination of the position of several helical regions relative to each other or to a β -sheet.
- *Quaternary structure* – position in space of macromolecules (= elements) in a supramolecular complex (= system). For example, single proteins in a multienzyme complex.

When a salt is crystallized, a *periodic structure* forms, which is characterized by a periodic arrangement of their elements. Erwin Schrödinger (1944) called the biological structure an *aperiodic crystal*. This means a highly organized structure, the elements of which, are not simply repeated periodically. Sometimes one tries to evaluate this structural organization as *structural information*. As we pointed out in the previous Sect. 2.1.2, however, it is hard to quantify this parameter. Probably, the structural information should be measured as the effort which is necessary, to describe such a structure perfectly.

Consequently, the process of structure formation of biological systems, be it the development of life, its reproduction, or simply the biosynthesis of a macromolecule, all are accompanied by reduction of entropy. This appeared to be contrary to the second law of thermodynamics and has given rise to heated philosophical discussions in the past. The second law of thermodynamics actually postulates that in spontaneous processes occurring in isolated systems, the entropy strives towards a maximum. Yet, neither an organism nor its environment, i.e., the earth as a whole can be considered as an isolated system. The earth is constantly absorbing energy from the sun and is emitting this energy again. That continuous flow of energy maintains a permanent nonequilibrium state which manifests itself not only in a direct way in photosynthesis with subsequent metabolism of heterotrophic organisms, but also in the environment of life, for example, in flowing water, in alternation of light and darkness, and in changes in temperature, humidity, etc.

In fact, structures are also formed in inanimate nature under loss of entropy. Basically, a distinction must be made between two kinds of structure formation which at the same time explain the two ways of self-organization in living systems.

- *Equilibrium structures*, for example a salt crystal, formed spontaneously during evaporation of a solvent.
- *Nonequilibrium structures* (or: *dissipative structures*), for example an arrangement of cirrus clouds as the result of processes of air convection under special meteorological conditions.

The genesis of life which is based on the prebiotic formation of a first bio-macromolecule, and the subsequent evolution of all organisms, can be understood as the result of complicated processes that occur far from equilibrium. In this context, a major role is played by dissipative structures of the inanimate environment, such as periodic temperature variations, tides, etc. Substantial importance has to be attributed also to such nonequilibrium structures in the course of life itself. They are frequently represented by time structures, such as for example the heart rate, or other kinds of oscillations, which in some cases are associated with the so-called biological clock (for further explanation see Sects. 3.1.4, 5.2.4, and 5.2.5).

Equilibrium structures, such as inorganic salt crystals with a very primitive geometry, become very complex and variable in shape, when based on the sophisticated pattern of the primary structure of bio-macromolecules rather than on the relatively simple field of interaction of spherico-symmetric ions. The spontaneous folding of proteins and their arrangement to supramolecular structures, such as complex enzyme systems of even ribosomes, must be viewed from this aspect. Such processes are also referred to as *self-assembly*. More specific aspects relating to the formation of equilibrium and nonequilibrium structures will be given in subsequent chapters of this textbook.

Further Reading

Eigen 1971, 1992; Meinhardt 2008, 2009; Kauffman 1993; Strait and Dewey 1996.

2.1.4 *Distribution of Molecular Energy and Velocity at Equilibrium*

The Boltzmann equation of entropy (Eq. 2.4) as derived in Sect. 2.1.1, helps to illustrate the second law of thermodynamics, according to which, isolated systems spontaneously approach a state of maximum entropy. We demonstrated there that one can express this also in the following way: at equilibrium isolated systems reach a state of highest realization probability (maximum of W). Now, we will ask the question: what are the mean properties of the molecules at this equilibrium state? This is of fundamental importance for further considerations.

It is possible to answer this question on the basis of statistical thermodynamics, which helps to make predictions on the probability of energy distribution among the elements of a system in thermodynamic equilibrium. It should, however, be emphasized that a deterministic description of the properties of single particles lies outside the limits of this discipline. Nevertheless, it will be seen that even a statistical statement allows the drawing of important conclusions regarding reaction rates, stability of molecules, and many others.

Let us imagine a space that contains gas molecules of uniform mass (m), which at the beginning all have the same velocity (v), and consequently, the kinetic energy (E) which can be calculated by the following equation:

$$E = \frac{m}{2} v^2 \quad (2.9)$$

This equality of the kinetic energy of all molecules, in fact, is a highly improbable state. In this case, using Fig. 2.1, all molecules would belong to a single box, or a single class of properties n_i . Thus, according to Eq. 2.5: $W = n!/n_i! = 1$. This situation will change instantaneously. The molecules would exchange their energy by elastic collisions with each other, and soon a great number of energy states would be occupied. A simple mathematical example shows that W , and according to Eq. 2.4 also the entropy (S), will increase with a growing number of (m) of state classes, provided the following relation applies:

$$\sum_{i=1}^m n_i = n$$

Because of the law of conservation of energy (first principle of thermodynamics), the following condition must be satisfied at the same time:

$$\sum_{i=1}^m n_i E_i = \text{const} \quad (2.10)$$

The energy of this system, therefore, can be distributed randomly among all of its molecules. The total energy of the system, however, must always remain constant.

Now, let us look for a state with maximum probability, this means, with a maximum of entropy. Corresponding to the second principle of thermodynamics, this is actually the thermodynamic equilibrium. Applied to our example the question arises: How energy and, correspondingly, velocity will be distributed between the n particles, after a sufficiently long period of time if the system is isolated? Even under equilibrium conditions, of course, the energy of individual molecules will change permanently, but nevertheless, the statistical mode of energy distribution, or of the distribution of various degrees of molecular velocity at equilibrium becomes

stationary, i.e., time independent (for a detailed description of various kinds of stationary states see Sect. 3.1.4).

Considerations of this type lead to *Maxwell's equation of velocity distribution*:

$$\frac{dn(v)}{n_0 dv} = \frac{4}{\sqrt{\pi}} \left(\frac{m}{2kT} \right)^{3/2} v^2 e^{-\frac{mv^2}{2kT}} \quad (2.11)$$

The left part of this equation contains a relation which expresses the relative number (dn/n_0) of those molecules which are related to a particular velocity interval (dv). Its unit is $s\ m^{-1}$. This is a function expressing the probability of the distribution of the velocity, where m is the mass of a molecule (not to be confused with the numeral m in Eq. 2.5), and k is the Boltzmann constant. Multiplying in Eq. 2.11 the denominator and numerator of the expression within the brackets as well as that in the exponent with the Avogadro number ($N = 6.023 \cdot 10^{23}\ mol^{-1}$) one introduces molar, instead of molecular parameters:

$$M = N \cdot m \quad (2.12)$$

and

$$R = N \cdot k \quad (2.13)$$

(M – molar mass, $R = 8.314\ J\ K^{-1}\ mol^{-1}$ gas constant)

In Fig. 2.3, as an example, the velocity distribution of molecules of oxygen is depicted ($M = 0.032\ kg\ mol^{-1}$). This curve is not symmetrical. The mean velocity (weighted arithmetic mean value) in general will be higher than the maximum value. The following relation applies:

$$v_{\max} = \sqrt{\frac{2kT}{m}} = \sqrt{\frac{2RT}{M}}; \quad \bar{v} = \sqrt{\frac{8kT}{\pi m}} = \sqrt{\frac{8RT}{\pi M}} \quad (2.14)$$

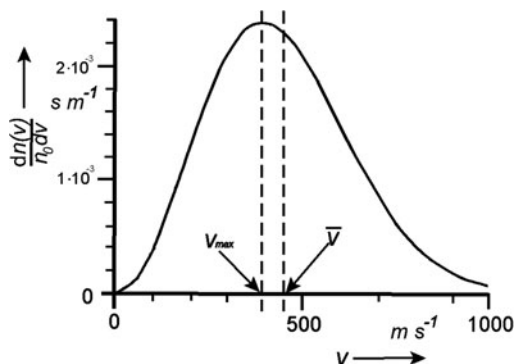


Fig. 2.3 Velocity distribution of O_2 molecules at $37^\circ C$ corresponding to Eq. 2.11

These values increase with increasing temperature (T) and decrease with increasing molar mass (M).

Maxwell's velocity distribution is based on the Boltzmann function of energy distribution. It determines the number (n_i) of those particles which under equilibrium conditions of the system have an energy level of E_i :

$$n_i = C e^{-\frac{E_i}{kT}} \quad (2.15)$$

(C is a factor of standardization). The exponent of this equation as in Eq. 2.11 expresses the relation between energy E_i of the particle and the mean energy of thermal noise (kT). The Boltzmann constant (k) should be replaced by the gas constant (R) in cases where E_i expresses not the energy (in: J) of a single molecule, but the molar parameter (in: J mol⁻¹), as is customary in chemical thermodynamics.

The Boltzmann equation predicts the distribution of any form of energy. It expresses the situation, whereby lower states of energy tend to occur more often than states of higher energetic levels. Related to the total number of particles, it may also be written in the following form:

$$\frac{n_i}{n_{total}} = \frac{e^{-\frac{E_i}{kT}}}{\sum_{i=0}^{\infty} e^{-\frac{E_i}{kT}}} \quad (2.16)$$

This equation provides the basis for many calculations in statistical thermodynamics. For example, it can be used to determine the percentage of molecules in a given system which are able to overcome a certain energy barrier (E_i). This aspect will be discussed in greater detail in the following chapter.

The use of this equation is allowed only in cases of thermodynamic equilibrium. At first glance it seems to limit its applicability to biological systems which are often far from equilibrium. The process of the adjustment of the Boltzmann distribution is, however, so fast that an imbalance of this kind has to be considered only in rare cases. On many occasions in this book we will indicate that in fact subsystems could be in equilibrium even in cases where the whole system is far from it (see especially in Sect. 3.1.4).

2.1.5 Energy of Activation, Theory of Absolute Reaction Rate

The temperature dependence of the rate of a chemical reaction was described by S. Arrhenius by the following equation:

$$k_R = A e^{-\frac{E_A}{RT}} \quad (2.17)$$

Here, k_R is the rate constant. The energy E_A was defined as the *activation energy*; A is an empirical factor.

In fact, this equation describes not only the temperature dependence of chemical reactions. It can also be used to depict corresponding temperature functions of other physicochemical processes, such as diffusion, kinetics of phase transitions, etc. At certain temperature intervals, even complicated biological processes can be fitted in this way, such as for example growth rate, or heart rate of poikilothermic animals.

To evaluate experiments, in which time-dependent parameters, like rate constants, reaction rates, frequencies of oscillations, etc., were measured as functions of temperature, a so-called *Arrhenius plot* is applied (see. Fig. 2.11). In this case the logarithmic form of Eq. 2.17 is used:

$$\ln k_R = \ln A - \frac{E_A}{RT} \quad (2.18)$$

Applying this kind of evaluation, $y = -\ln k_R$ is plotted against $x = 1/T$. If in the investigated process the activation energy itself is constant, then the measured points are located on a straight line:

$$y = ax + b \quad (2.19)$$

The slope of this straight line is, according to Eq. 2.18: $a = -E_A/R$. The relation: $b = \ln A$ and, consequently, the factor of A can be read from the extrapolated value of $x = 1/T = 0$.

Direct derivation of the logarithmic function to dT is more complicated. We need it later:

$$\frac{d \ln k_R}{dT} = \frac{E_A}{RT^2} \quad (2.20)$$

Real biological processes usually consist of a large number of single reactions, each with different activation energies. Hence, a temperature dependence must be expected which is much more complicated than that expressed by the Arrhenius equation (Eq. 2.17). Surprisingly, however, even in these cases functions will still be obtained which can be described by straight lines in the Arrhenius plot. This is caused by the fact that in complicated processes the rate of the overall process is determined by a simple rate-limiting reaction. The Arrhenius plot just reflects the activation energy of this rate-limiting process.

In some physiological papers, a so-called *Van't Hoff rule* of temperature dependence is used. According to this the rate of a reaction increases by 2–3 times, if the temperature rises by 10°C. This factor is referred to as the Q_{10} value. This rule, however, is not supported by thermodynamic considerations. Even in cases where the energy of activation is constant, corresponding to the Arrhenius equation, the Q_{10} value will increase with increasing temperatures.

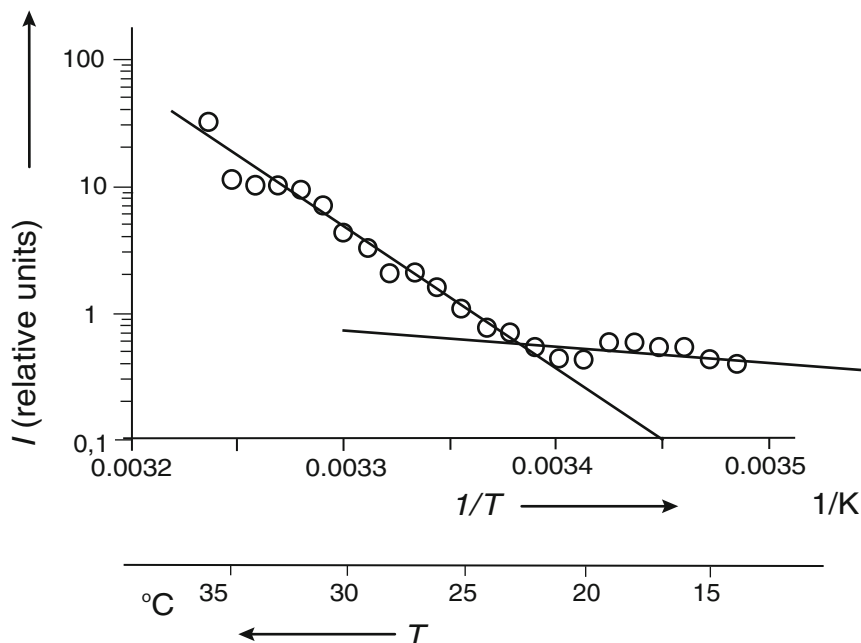


Fig. 2.4 An example of an Arrhenius plot: The logarithm of the heat-activated current (normalized to that at 25°C) in a temperature-sensitive TRPV4-channel is plotted versus the reciprocal value of the temperature. Compare with the plot of the same data directly versus temperature in Fig. 4.1 to calculate the Q_{10} corresponding parameters (Data from Watanabe et al. 2002)

In some reactions the energy of activation (E_A) itself is a function of temperature. In this case the straight line in the Arrhenius plot gets kinks (Fig. 2.4). The reasons for this can vary. In complex processes it is possible that at different temperatures, different reactions with different energies of activation will become rate limiting. In other cases such transitions may occur as the result of conformational changes of one of the components of the reaction, probably, conformational change in the enzyme itself. In reactions of membrane-bound enzymes, sometimes phase transitions of the enzyme near lipids may cause such effects.

The statistical mechanics allows a direct derivation of the Arrhenius equation. This is based on the *theory of absolute reaction rate* (or: *theory of transmission states*) as developed by H. Eyring. Here we can only explain some basic ideas of this approach.

Chemical reactions presuppose splitting of chemical bonds. To analyze such processes it is necessary to know the interaction energy of a given atom in the molecule as a function of the distance to each of its neighbors. This is an intricate function which can be expressed mathematically as an n -dimensional space corresponding to n possible ways of geometrical approach. In this space a line may be found which shows the way from state A to state B over the smallest peaks

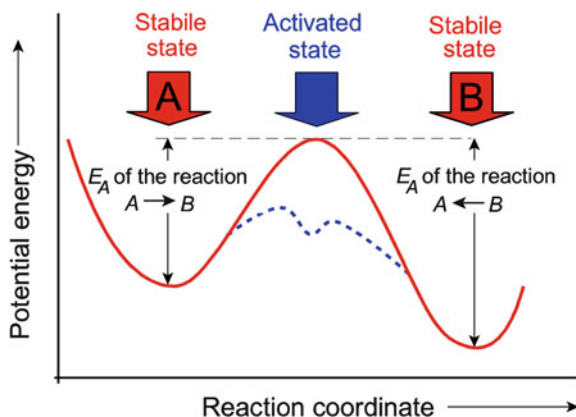


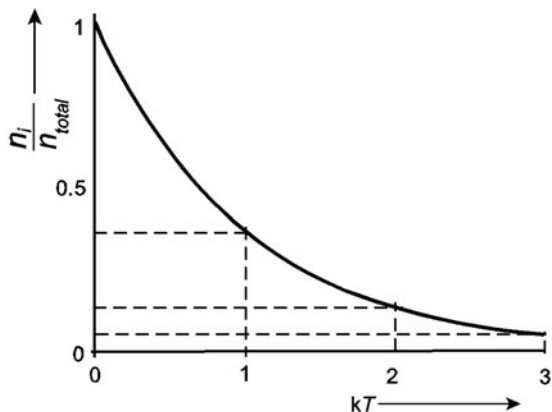
Fig. 2.5 Schematic potential energy diagram of a molecule in states A and B taken along the reaction coordinate. The solid curve represents the reaction without a catalyst; all the explanations correspond to this curve. The *blue dashed line* indicates how the energy of activation is modified in the presence of a catalyst (enzyme). In this case, the curve with a single maximum, dividing states A and B is transformed into another with two maxima, between which the minimal energy of an enzyme-substrate complex is positioned

of the activation energies. This line is called the *reaction coordinate*. It is possible to represent the energy level along this line in a 2-dimensional graph.

In Fig. 2.5 such a function is plotted with two minima, representing the two stationary states A and B. In these positions, the molecule will be in a stable state, which is however continuously deflected in a direction of the reaction coordinate by thermic collisions (the term *stability* will be explained in detail in Sect. 3.1.4). The effectiveness of those collisions depends on their energy, i.e., on the relation of their energy to the activation energy E_A . The energy of collisions, in fact, is thermic energy, which can be expressed by the factor kT (or in molar units, correspondingly in RT). The Boltzmann equation (Eq. 2.16) allows the calculation of the relative number of those molecules (n_i/n_{total}) which are able to overcome the energy barrier (E_i), which in our case is identical with the energy of activation (E_A). The relation E_i/kT (Eq. 2.16), of course, decreases with increasing temperature (T). This is illustrated in Fig. 2.6. While 36.8% of all particles have reached the energy level $E_i/kT = 1$, a value twice as high is only reached by 13.5%, and for $E_i/kT = 3$ there are only 5% reaching this level. The percentages drop fast with rising values of E_i/kT .

The theory of absolute reaction rate is based on the assumption that an activated transition state between the two states A and B exists. Its potential energy is identical with the energy maximum in Fig. 2.5. This activated state, of course, is highly unfavorable and unstable from the energetic point of view. At any point of time it will be assumed only by a very small number of molecules which will drop back from the energy maximum into the stable states of an energy minimum to the right and left of this maximum. An equilibrium constant (K^*) can be defined that characterizes the relation between activated and stable states according to the mass

Fig. 2.6 The relative number (n_i/n_{total}) of molecules as a function of the energy in kT -units. This curve corresponds to Eq. 2.16



action law. The proportionality between the rate constant (k_R) and the number of molecules in the activated state can then be expressed as follows:

$$k_R = q \frac{kT}{h} K^* \quad (2.21)$$

The symbol * indicates the reference to an activated state. The relation between kT and Planck's constant h comes from kinetic quantum theory as derived by Wilson–Sommerfeld. The factor q gives the fraction of activated molecules which leave to the right or left of the energy maximum. For the case of symmetry of the relations, it becomes $q = 1$.

The phenomenological thermodynamics allows calculation of the relation between the molar free Gibbs energy of reaction (ΔG) and the equilibrium constant (K):

$$\Delta G = -RT \ln K \quad (2.22)$$

In the case of a reaction $A \rightarrow B$, i.e., for a reaction in which the total number of molecules remains constant, and under the above-mentioned condition of $q = 1$, it is possible to calculate the conditions for the activated state (using the symbol *):

$$k_R = \frac{kT}{h} e^{-\frac{\Delta G^*}{RT}} \quad (2.23)$$

If one further considers the following relation:

$$\Delta G = \Delta H - T\Delta S \quad (2.24)$$

it follows that:

$$k_R = \frac{kT}{h} e^{\frac{\Delta S^*}{R}} e^{-\frac{\Delta H^*}{RT}} \quad (2.25)$$

The relation of this formula with the Arrhenius equation (Eq. 2.17) can be found by modification of the above equation in the following way:

$$\ln k_R = \ln T + \ln \left(\frac{k}{h} e^{\frac{\Delta S^*}{R}} \right) - \frac{\Delta H^*}{RT} \quad (2.26)$$

and:

$$\frac{d \ln k_R}{dT} = \frac{1}{T} + \frac{\Delta H^*}{RT^2} = \frac{RT + \Delta H^*}{RT^2} \quad (2.27)$$

Comparing Eq. 2.27 with Eq. 2.20 the following correlation is obvious:

$$E_A = RT + \Delta H^* \quad (2.28)$$

For other values of q for reactions of synthesis or lysis, this relation can be modified.

In Eq. 2.25 especially the temperature-independent exponential term is of interest, which contains the *entropy of activation* (S^*). This is an indicator for conformation processes of the molecules participating in the reaction. These molecular conformations are prerequisites especially in cases where macromolecules are included in biochemical reactions. The parameter ΔS^* can become positive as well as negative. This depends on whether the molecule during the reaction gets a larger ($\Delta S^* < 0$) or a lower ($\Delta S^* > 0$) degree of order. Conversely, the entropy changes of the water molecules during this reaction must be considered too. We will discuss the entropy modifications of the water during reactions of hydration later in detail (Sect. 2.2.3, Fig. 2.19).

The acceleration of biochemical reactions by enzymes is based on a change in the reaction coordinate. The energy levels of the starting and the end product of the reaction will remain unaltered by this. However, with the help of an enzyme as a biological catalyst the transition from one state into another is possible with lower barriers of activation energy. This is indicated by the blue dashed curve in Fig. 2.5. First, an energy-substrate complex is formed, which requires less activation energy than that, which is needed to form the activated complex without the enzyme. This enzyme-substrate complex is then broken up again by a small input of activation energy, and substance B will be released. Of course, even with the help of an enzyme, the reaction $A \rightarrow B$ can take place spontaneously only, if the level of B is below that of A.

The concept of the theory of absolute reaction rate is not only applicable to processes of chemical reactions. In a similar manner, processes of diffusion can be described. Diffusion of a substance can be imagined as a reaction consisting entirely of location jumps from one position in the vicinity of a molecule to another. Under this circumstance the reaction coordinate will largely coincide with the direction of diffusion. Investigations of activation energy of diffusion processes

may provide the path of diffusion, for example, the mechanism of permeability of a molecule through a membrane.

Let us now discuss the problem of the life span of a bond and of a molecule. This question is directly connected to the above consideration. The Arrhenius equation also allows one to calculate the stability of a given molecule. In this case, however, not the reaction rate of molecular transformation is of interest, but its reciprocal value, namely the mean time, during which a bond with a given bonding energy will resist the attack of the energy of thermal noise (kT , resp. RT). In this case the energy of activation of a decay reaction (E_D) must be applied. The mean time of stability (t) of a molecule is inversely proportional to the rate of its transformation. Thus:

$$t = \tau e^{\frac{E_D}{kT}} \quad (2.29)$$

This equation contains a time constant (τ), similar to the value A in Eq. 2.17. This factor depends on a number of structural properties of the molecule. It can only roughly be estimated in orders of magnitude. For smaller molecules, τ will lie between 10^{-14} and 10^{-13} s.

The characteristics of function (2.29) are illustrated in Fig. 2.7. Semi-logarithmic plotting ($\ln t$ against E_D) will result in a straight line. It can be clearly seen that even small changes of the activation energy (E_D) of the decomposition reaction will lead to dramatic changes of the life span (t). The range may stretch from 10^{-12} s to geological periods of time amounting 10^5 years.

Considering the energies of covalent bonds with several hundreds of kJ mol^{-1} one understands that they practically cannot be destroyed by thermal noise (RT) which at 300 K amounts to only 2.5 kJ mol^{-1} . However in fact, the word “cannot” does not exist in the language of statistical thermodynamics. As indicated in

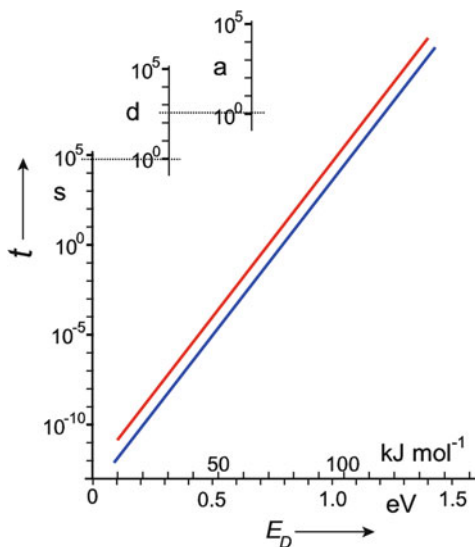


Fig. 2.7 The mean lifetime (t) of a bond dependent on the energy of activation of decay (E_D), corresponding to Eq. 2.29 ($T = 293 \text{ K}$, red line: $\tau = 10^{-13}$, blue line: $\tau = 10^{-14} \text{ s}$)

Fig. 2.7, this just expresses the vanishing small probability of this process, resp. the extremely long life span of the corresponding bond. Even the binding energy of hydrogen bonds, between 13 and 25 kJ mol⁻¹ is larger than RT in this temperature range. Nevertheless, the probability of their destruction by thermal noise is much larger and correspondingly their lifetime extremely short. Discussing the structure of water (Sects. 2.2.2, 2.2.3), we will show that in these cases the hydrogen bonds are destroyed and rebond at high frequency.

Not only the thermal noise (RT) influences the stability of molecular bonds but additionally quanta of various kinds of radiation. This will be discussed in detail in Sects. 4.7, 4.8, and 4.9 (see also Fig. 4.32).

As the mean life span of a molecule is only a question of probability, the concrete destiny of an individual molecule cannot be predicted. The assumption, however, can be made that by biological selection only DNA molecules with monomers of high bonding energy are left. Nevertheless, spontaneous molecular transformations take place which are known as mutations. A mutation results in a new molecule whose stability has not yet been approved by selection. The life span of this molecule can be substantially changed by a slight displacement of its bond energy. This may explain why remutations of mutation of already mutated molecules occur more often than could be expected statistically.

Further Reading

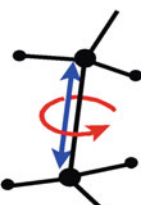
Blumenfeld and Tikhonov 1994; Westerhoff and van Dam 1987. The theory of absolute reaction rate: Eyring and Urry in: Waterman and Morowitz 1965.

2.1.6 Kinds of Thermal Molecular Movement

As a consequence of energy distribution in molecular systems, even under conditions of thermodynamic equilibrium, all of their parts are in vigorous thermal motion driven by the thermal energy kT . The energetic aspects of intermolecular interactions of proteins shall be considered in Sect. 2.2.5. Here we will discuss some general consequences of molecular movements and some methods for their quantification.

In general there exist three forms of molecular movements: vibrations, rotations, and translations: *Vibrations* are oscillations in the binding distances between the atoms in a molecule. The term *rotation* means not only the rotation of the whole molecule, but additionally, the spin of individual atoms or atomic groups around the axis of their bonds. The full translocation of a molecule or a part of it in space is meant by *translation*.

The frequency of these kinds of movements differs in a wide range. As shown in Fig. 2.8, rotation can come in resonance with electromagnetic waves at frequencies between 10¹⁰ and 10¹² Hz. In the case of vibrations, this frequency is higher and reaches about 10¹⁴ Hz. Both are in the range of infrared light (see Fig. 4.32) which is the basis for molecular analysis by infrared spectroscopy (see Sect. 4.8.1).



	FREQUENCY s^{-1}	KORRESPONDING QUANTUM ENERGY	
		J	eV
ROTATION	$10^{10} - 10^{12}$	$\approx 10^{-23}$	$\approx 10^{-4}$
VIBRATION	$< 10^{14}$	$\approx 10^{-20}$	$\approx 10^{-1}$

Fig. 2.8 Magnitudes of vibrational and rotational movement of a covalent C-C bond. The quantum energies are to be considered in relation to the energy of thermal movement which is $4.1 \cdot 10^{-21}$ J, resp. $2.6 \cdot 10^{-2}$ eV at $T = 300$ K

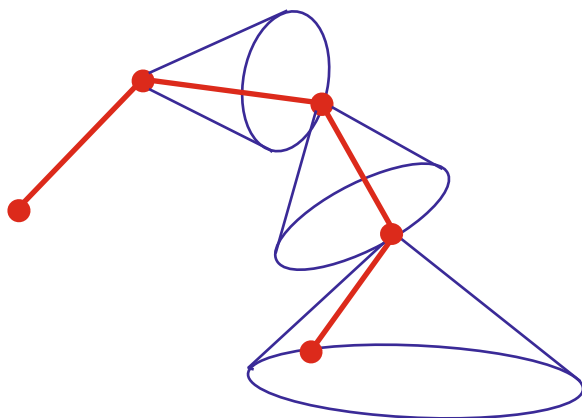


Fig. 2.9 The rotation cones of the atomic groups in a chain molecule

These frequencies allow us to estimate the quantum energies of these processes which appear to be near to that of the thermal energy kT . At a temperature of 300 K, the product kT equals $2.6 \cdot 10^{-2}$ eV. This means that for molecular vibrations and rotations the energy transfer is governed by discrete quantum steps. In contrast to this, translation requires quite small energy levels, of approximately 10^{-35} J or about 10^{-16} eV. Energy transfer of translational movement, therefore, can be considered to be continuous.

The intramolecular rotation of atoms in relation to each other has consequences for the structure of macromolecules. A covalent bond, rotating around its angle relative to another, consequently describes a cone. The next bond rotates, as well, but it does so quasi on the spot moving itself in an orbit of that lateral cone surface (Fig. 2.9). In this way chain molecules can assume a stochastic orientation unless other strong attracting or repelling interactions between the monomers are preventing this. The molecular structures resulting in this way can be calculated by the laws of statistical thermodynamics.

The so-called *random flight model* allows the simplest description of the behavior of a long molecule. In this case, the real molecule is described as a sequence of segments with the individual length l . These segments are connected to each other in such a way that the orientation of each is independent from the orientation of the former segment. This random flight chain is characterized by the length (l) and the

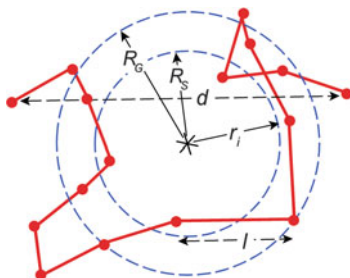


Fig. 2.10 Projection of a random flight chain, composed of 15 segments with a characteristic length l . * center of gravity of the molecule, d distance between the two ends of the chain, r_i distance of the i -th point from the center of gravity, R_G radius of gyration, R_S Stokes radius. This model corresponds to an unfolded polypeptide, composed of 150 amino acids, or a double strand DNA (According to data from G. Damaschun)

number (n) of its segments. The position of the endpoint of the segments can be expressed as vectors in a polar coordinate system which is based on the center of gravity of the molecule (Fig. 2.10). The distance between the endpoints of such a molecule (d_{\max}) at maximal elongation is:

$$d_{\max} = nl \quad (2.30)$$

The real distance of the endpoints of a random flight chain with large enough amounts of n segments, and after a sufficiently long time of observation, however, is distributed according to a Gauss statistic. As shown in Fig. 2.10, the mean of this distance is not identical with its maximum, i.e., the value of maximal probability. The mean value of this distance between the endpoints of the chain can be calculated by the following equation:

$$d = \sqrt{\langle r^2 \rangle} = l\sqrt{n} \quad (2.31)$$

In this case the expression $\langle r^2 \rangle$ is the mean of the squares of the distances as defined as follows:

$$\langle r^2 \rangle = \frac{\sum r_i^2}{n} \quad (2.32)$$

Comparing Eqs. 2.30 and 2.31 it can be seen that the mean distance (d) between the endpoints of the chain is smaller by \sqrt{n} in relation to its maximum (d_{\max}). Considering a chain consisting of 100 segments, the mean distance, therefore, will be only 10% of the maximal length, for chains of 1,000 segments, it will be only 3.2%.

The geometric dimension of such a molecule can be described by the mean of the square distance of all atoms of the molecule from the common center of gravity of

the molecule. This gives the so-called *radius of gyration* (R_G). For a random flight chain with $n \rightarrow \infty$ it gives:

$$R_G = \sqrt{\frac{\langle r^2 \rangle}{6}} = l \sqrt{\frac{n}{6}} \quad (2.33)$$

The radius of gyration, therefore, is proportional to the square root of the length of the chain (\sqrt{n}). For compact molecules, however, the monomers of which are not arranged in this way, one gets $R_G \sim \sqrt[3]{n}$, and for rod-shaped molecules, $R_G \sim n$.

Furthermore, it is possible, to define a so-called *Stokes radius* (R_S) for macromolecules, which is measurable experimentally. This is an example of a typical effective parameter. It predicts that the molecule moves in the water like a macroscopic sphere with a hydrophilic surface. In this case, Stokes' law could be applied. It gives the frictional force (\mathbf{F}) of a sphere with radius r , moving in a fluid with viscosity η by a velocity \mathbf{v} :

$$\mathbf{F} = 6\pi\eta r \mathbf{v} \quad (2.34)$$

(The bold type characterizes the vector character of force and velocity).

The Stokes radius, therefore, is a measurable parameter, which in the same way as the radius of gyration does not correspond to any "visible" radius of the molecule. Both are effective parameters which, however, are well defined (see Fig. 2.10). We will come back to this point in context with the introduction of the hydration radius of ions in Sect. 2.2.2. The relation of the experimentally determined parameters R_G and R_S gives important hints to the configuration of a given macromolecule. If the molecule really resembles a random flight chain, one gets: $R_G/R_S = 1.51$. If the molecules have a spherical, but compact structure, this relation becomes 0.8. In the case of rod-shaped molecules, this parameter depends on the relation of the radius to the length of it.

Real molecular chains, composed of covalently bound atoms (Fig. 2.9) are not usually as flexible as predicted by the random flight model. Nevertheless, this model is also applicable in these cases. For this, such monomeric constituents were connected together as segments, matching the conditions of the random flight model. For stiffer molecules one has to choose segments with a larger characteristic length. In the case of a quite flexible molecule of poly-L-glycine, the characteristic segments are formed by approximately four amino acids with a characteristic segment length of $l = 1.38$ nm. For the stiffer molecule of poly-L-alanine, in contrast, 12 amino acids must be combined in one segment with $l = 4.10$ nm.

A repulsive force between the atoms, for example because of electrical charges of equal sign, results in an additional expansion of coiled molecules. A similar limitation of the random flight model is given by the exclusion principle, i.e., by the fact that two atoms cannot be at the same location at the same time. Therefore, the characteristic length of the segment of a DNA molecule as a strong polyelectrolyte is $l = 90$ nm.

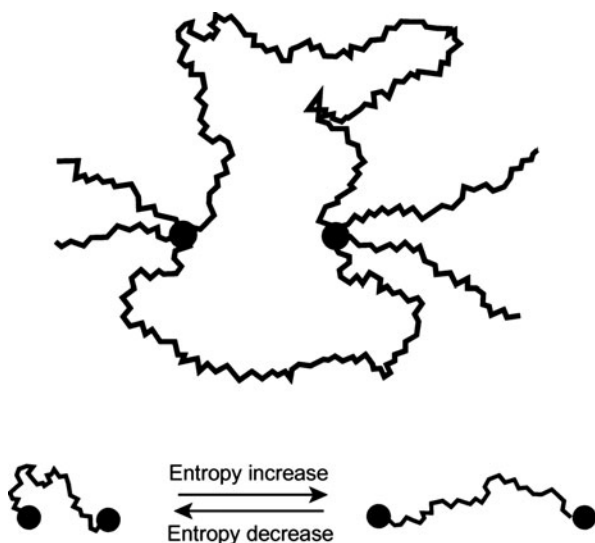


Fig. 2.11 The molecular mechanism of rubber or entropy elasticity. If the distance of the two points of fixation increases, the entropy content of the molecule decreases (schematically illustrated in the elongation of the chain in the lower part of the figure). A force is generated with the tendency to recover the previous situation (the state of maximal entropy) (According to data from G. Damaschun)

If such a chain molecule is fixed at two points (Fig. 2.11) then attractive or repulsive forces are developed. This occurs, if the distance between these points is larger or smaller than the parameter d (Eq. 2.31). The result is the so-called rubber, or *entropy elasticity*. It is caused by the decrease of the entropy of the molecular chain during stretching. In entropy elastic materials, the points of fixation are formed by interconnections of the molecules. This property is the most common type of elasticity of biological materials. We will come back to this in Sect. 3.6.3.

The third kind of molecular movement, mentioned above, *translation*, means a movement in space. This translational movement, driven by thermic hits is important not only for single molecules, but also for microscopically visible particles. This was first observed in 1827 by the Scottish botanist Robert Brown, who observed the stochastic movement of gymnosperm pollen in suspension. This movement was later named *Brownian motion*. The reason for this stochastic translocation of microscopically visible particles is the hits resulting from thermic translations of the surrounding molecules. The movement results from the vector sum of these hits (Fig. 2.12). If the particles are sufficiently small, then a real vector occurs as the sum. For larger particles, this mean vector force becomes zero.

The spatial translocation of the particle, in fact, is stochastic as well as the molecular translocations themselves. Observing the position of the particle over a sufficiently long time, there will be no translocation at all. As a measure of the Brownian motion, therefore, the mean of the squares of the translocations is used. For this, the position of the particle will be registered at definite time intervals

Fig. 2.12 A snapshot indicating the thermic hits by surrounding molecules on a microscopically small particle (*blue arrows*). The individual hits, drawn as fine vectors, induce by vector addition a resting force (*red vector*) which affects the Brownian motion of the particle

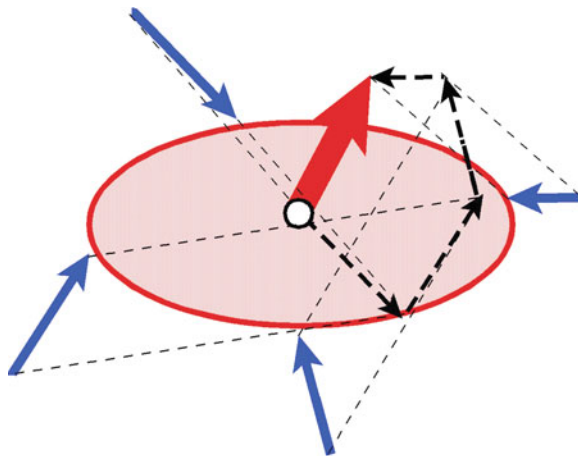
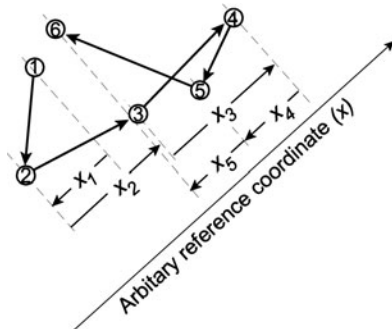


Fig. 2.13 Method to evaluate Brownian motion. The position of a particle was marked at definite intervals of time (Δt). This two-dimensional picture subsequently was projected on an arbitrarily drawn axis, which allows us to measure the distances x_i



Δt (Fig. 2.13). Subsequently, when the distance x_i is measured, the particle has moved in a projection of an arbitrarily drawn axis. If these values are measured n times, the parameters x_i for $i = 1 \dots n$ are obtained. The square of displacement $\langle x^2 \rangle$ can be calculated in analogy to Eq. 2.32 in the following way:

$$\langle x^2 \rangle = \frac{\sum x_i^2}{n} \quad (2.35)$$

This parameter corresponds to the square of the mean path length of the moving particle:

$$\bar{x} = \sqrt{\langle x^2 \rangle} \quad (2.36)$$

The intensity of Brownian motion depends on the shape and size of the particle as well as on the temperature and the viscosity of the medium. Einstein and

Smoluchowski derived an equation for spherical particles, meeting the conditions of the Stokes equation (Eq. 2.34):

$$\langle x^2 \rangle = \frac{RT\Delta t}{3N\pi\eta r} = \frac{kT\Delta t}{3\pi\eta r} \quad (2.37)$$

The Einstein–Smoluchowski Eq. 2.37, therefore connected the measurable square of displacement ($\langle x^2 \rangle$) of Eq. 2.35 in relation to the radius of the particle (r), the viscosity of the fluid (η), and the temperature (T), respectively the energy of thermal movement ($RT/N = kT$). Δt is the time interval in which the values x_i were measured.

This equation indicates that the square of the mean translocation length which was achieved by Brownian motion per time interval was inversely proportional to the particle radius (r) and the viscosity (η). For example: a particle with a diameter $r = 1 \mu\text{m} = 10^{-6} \text{ m}$ moves in water ($\eta = 0.00089 \text{ N s m}^{-2}$) at $T = 298 \text{ K}$ during $\Delta t = 1 \text{ s}$ at a mean distance of

$$\bar{x} = \sqrt{\frac{1.38 \cdot 10^{-23} \cdot 298 \cdot 1}{3 \cdot \pi \cdot 0.00089 \cdot 10^{-6}}} = 0.7 \cdot 10^{-6} \text{ m}$$

The size of the particle (r), as well as the mean translocation (\bar{x}) are just within the resolution of a normal microscope. Measuring (\bar{x}), and knowing (r), it is possible to measure the viscosity (η). This was sometimes used to measure the viscosity of cytoplasm. Conversely it is possible to measure the diameter of very small particles by dark field microscopy in fluids with known viscosity, measuring the parameter \bar{x} .

The Einstein–Smoluchowski equation is not only important for these methods. With the help of this equation it is also possible to calculate the time a particle needs to cross a certain distance by diffusion. If, for example, at one point in the cell a vesicle was created and is bound at another definite location by specific receptors, the mean time this particle needs to cross this distance by stochastic movements can be calculated (considering, however, the cytoplasm as an isotropic and homogeneous viscous medium!). Phenomenologically such an event appears to be the result of directed transport. (This, by the way, is an example of a kind of biological Maxwell demon, as discussed in Sect. 2.1.2!)

This translational movement is the basis of all processes of diffusion. Einstein found the connection between the square of displacement ($\langle x^2 \rangle$) and the diffusion coefficient (D) of a molecule in solution:

$$D = \frac{\langle x^2 \rangle}{2\Delta t} \quad (2.38)$$

Introducing Eq. 2.37, it gives

$$D = \frac{RT}{6N\pi\eta r} = \frac{kT}{6\pi\eta r} \quad (2.39)$$

We will come back to deriving Fick's law of diffusion in Sect. 3.3.1 using the approaches of phenomenological thermodynamics (Eq. 3.130).

From this equation the conclusion can be drawn that the diffusion constants of two molecules with similar molecular shapes are inversely proportional to the square root of their molar masses (M):

$$\frac{D_1}{D_2} = \frac{\sqrt{M_2}}{\sqrt{M_1}} \quad (2.40)$$

Further Reading

Eyal and Bahar 2008; Tinoco et al. 2002.

2.2 Molecular and Ionic Interactions as the Basis for the Formation of Biological Structures

In Sect. 2.1 the relationship between order and life was discussed, and in this context, the molecular basis of biological structures was considered from the point of view of statistical thermodynamics. Emphasis was placed on the energy of thermal movement and thermal noise (kT). These forces of stochastic movement appear to be opposed by intermolecular forces which tend to create order, building up molecular and supramolecular structures. Therefore, the expression E_i/kT (see: Eqs. 2.16, 2.17, 2.29) characterizes the relationship between any sort of structure-forming energy and the destructive energy of thermal noise.

Now, we will consider in particular those forces that contribute to the formation of molecular and supramolecular equilibrium structures. The forces operating in these so-called self-assembling processes are electrostatic interactions and weak bonds, i.e., bonds with energies of the same order as those of the thermal noise (kT). These bonds, therefore, can be split by thermic hits in a realistic period of time. Such weak bonds, for example, are hydrogen bonds as well as the whole complex of van der Waals interactions.

2.2.1 Some Foundations of Electrostatics

Because electrical phenomena are of major importance at all levels of biological organization (see: Sect. 3.5.1, Fig. 3.33) an account will be given of some fundamental parameters, definitions, and implications.

The basis of electrostatic considerations is electric charges. In the CGS-System electric charges are defined by mechanical parameters. So, for example, the electrostatic unit (eu) is an amount of charge which interacts with an opposite polarized charge of the same size at a distance of 1 cm by an attractive force of 1 dyn. This definition eliminates the need for defining specific electrical units. In the *Système International d'Unités* (SI), however, the application of mechanical parameters has been abandoned, and the charge is defined by the unit for current – the ampere (A), as coulomb (C) = ampere (A) · second (s). In this book, we will use this now generally accepted international system of units.

In electrochemical as well as in biophysical calculations the smallest electric charge is the charge of a univalent ion or a corresponding charged group. The *Faraday constant* (F) gives the number of charges per mole of singly charged ions, an amount, which is usually called *val*:

$$F = 9.6485 \cdot 10^4 \text{C val}^{-1}$$

Dividing this molar amount by the Avogadro number (N), gives the absolute value of the charge of a single ion (e), the same value as that of a positron:

$$e = \frac{F}{N} = 1.6021 \cdot 10^{-19} \text{C}$$

Using the units of the SI-system, a conversion factor must be applied, to calculate mechanical parameters resulting from electrical interactions. This conversion factor is the so-called *permittivity of free space* (ϵ_0), which amounts to:

$$\epsilon_0 = 8.854 \cdot 10^{-12} \text{C V}^{-1} \text{m}^{-1}$$

Coulomb's law makes it possible to calculate the repulsion force (**F**) of two charges (q_1, q_2) at distance r from each other:

$$\mathbf{F} = \frac{q_1 q_2}{4\pi\epsilon_0 \epsilon r^2} \quad (2.41)$$

This equation additionally contains the *dielectric constant* (ϵ), which is also called *permittivity*. In contrast to ϵ_0 , which unfortunately is usually written with the same symbol, the dielectric constant ϵ is just a number without dimension. It is a factor indicating, how much the interaction force between two charges will be diminished, if not vacuum, but a real substance is in between. The dielectric constant for pure water at 18°C, for example, amounts to 81.1. For various kinds of glass, it is between 5 and 7, and for arranged layers of lipids about 3.5.

Using Coulomb's law (Eq. 2.41) one can also calculate the energy of electrostatic binding (E) of two charges in a medium with $\epsilon > 1$. With the inverted sign, it equals the work required to move both charges from distance $r = \infty$ to distance $r = r_i$:

$$E = -W = \int_{\infty}^{r_1} \frac{q_1 q_2}{4\pi\epsilon_0\epsilon r^2} dr = -\frac{q_1 q_2}{4\pi\epsilon_0\epsilon r} \quad (2.42)$$

Any charge in space generates an electric field which is characterized by the gradient of the electric potential. The *electric potential* (ψ) is a state parameter of a particular point in space. It is defined as the work that is required to move a positive charge from infinite distance to point r . The unit of the potential is volt (V). In the case of the electric field around a charge q , the electric potential is a radial function of the distance r from this point. According to this definition and Eq. 2.42, the potential ψ_i at distance r_i from charge q is given by:

$$\psi_i = \frac{q_i}{4\pi\epsilon_0\epsilon r_i} \quad (2.43)$$

It is important to stress that the electrical potential is a scalar parameter in space, similar, say, to pressure or temperature. Later we will discuss surface potentials, Donnan potentials, diffusion potentials, etc. These are all just electrical potentials. Their names only indicate the reason for their occurrence. At any given time, at one and the same point in space, there can be only one value of an electrical potential, as there can be only one single temperature at one time at one point!

Points of equal potentials in space can be interconnected by *equipotential lines*.

Perpendicular to these equipotential lines electrostatic forces occur. Correspondingly, an electric field strength (\mathbf{E}) as a vector parameter can be defined as follows:

$$\mathbf{E} = -\text{grad } \psi = -\nabla\psi \quad (2.44)$$

The differential operator “grad” as well as the so-called *Nabla operator* ∇ means: derivation of the potential ψ to all three dimensions of space. It transforms the scalar ψ into vector \mathbf{E} .

Figure 2.14 shows the equipotential lines and the field vectors around a positively charged point. The directions of these vectors correspond with the vectors of the electric field strength (\mathbf{E}). According to convention, these arrows are always directed from the positive toward the negative pole. Similar illustrations can be seen in Figs. 3.35 and 3.37.

Considering the gradient of the electric field in just one direction (x) of space, Eq. 2.44, becomes:

$$\mathbf{E}_x = -\frac{d\psi}{dx}\mathbf{i} \quad (2.45)$$

In this case \mathbf{i} is the unit vector in the direction of the x coordinate.

In Fig. 2.15 the potential and the field strength in the x -direction are depicted in the vicinity of a homogeneous membrane which separates two different electrolyte solutions (phase I and phase II). This demonstrates that points of discontinuity of

Fig. 2.14 Equipotential lines and field vectors around a positively charged point

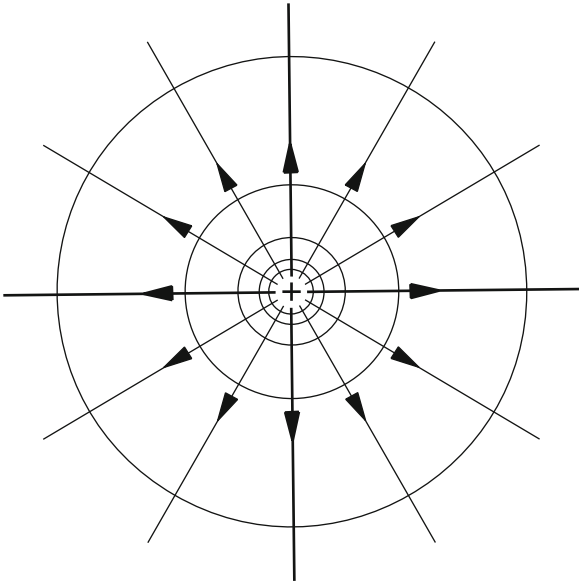
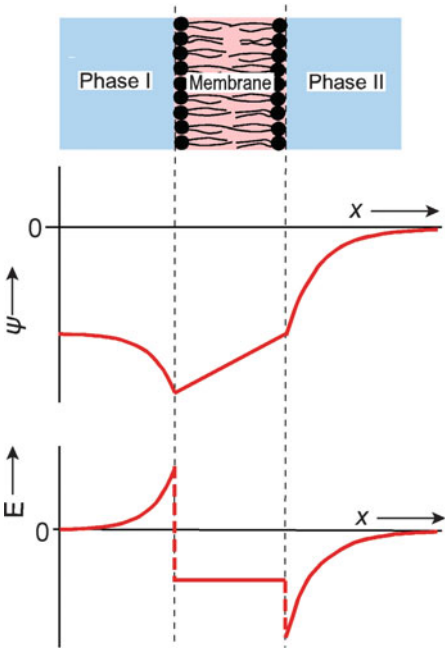


Fig. 2.15 The electrical potential [$\psi(x)$] and the electric field strength [$E(x)$] near a homogeneous membrane which separates two electrolyte solutions with different composition



the field strength can occur, with a reversal of the field direction. In this case, the reason for these discontinuities are surface charges on the phase boundaries. The field vectors and the equipotential lines in this example would indicate a set of parallels, perpendicular to each other. Actually, the biological membrane is not homogeneous, but it is a mosaic of various compounds with different surface charges and various dielectric constants. Even its cross section is not homogeneous. Its electrostatic structure is far more complicated, as we will see in Sect. 2.3.6, Fig. 2.48.

The unit of field strength, according to its definition (Eqs. 2.44, 2.45) is V m^{-1} . As an example, let us calculate the field strength across a cell membrane with a thickness of $7 \text{ nm} = 7 \cdot 10^{-9} \text{ m}$. If the electrical potential difference between the two surfaces amounts to $\Delta\psi = 50 \text{ mV} = 5 \cdot 10^{-2} \text{ V}$, the following field strength results:

$$|E_i| = \frac{5 \cdot 10^{-2}}{7 \cdot 10^{-9}} = 7.14 \cdot 10^6 \text{ Vm}^{-1}$$

The term $|E_i|$ means the magnitude of the field strength in the x -direction. Later we will discuss the biophysical consequence of this large electric field strength (Sects. 2.3.6 and 3.4.3).

Using Eqs. 2.41, 2.43, and 2.44. One can calculate the force (\mathbf{F}), acting on a charge (q) in an electric field (\mathbf{E}):

$$\mathbf{F} = \mathbf{E}q \quad (2.46)$$

Furthermore, dipoles must be considered, which play an important role in electrostatics. A dipole is a structure that contains two equal charges (q) of opposite polarity, positioned at a certain distance (l) from each other. Consequently, an electric dipole, as a whole, is electrically neutral. A dipole moment (μ) is defined as the product of charge (q) and distance (l) between them:

$$\mu = ql \quad (2.47)$$

Dipoles can be formed by ampholytic molecules, this means, in molecules carrying simultaneously positive as well as negative charges. They may also result from polarization phenomena of covalent bonds.

According to Eq. 2.47, the unit of a dipole is C m or, correspondingly: A s m . In molecular studies, sometimes the Debye unit (D) is preferred because its amount is in the order of real molecules. The relation between these two units is:

$$1 \text{ D} = 3.3 \cdot 10^{-30} \text{ Cm}$$

In Table 2.1 dipole moments of several compounds are listed. The dipole moments of chloroform, methanol, and water are caused exclusively by polarizations of covalent bonds. In the case of proteins and nucleic acids, the values

Table 2.1 Dipole moments of various substances (After Netter 1959)

	D	$\cdot 10^{-30}$ C m
Chloroform	1.05	3.47
Methanol	1.68	5.54
Water	1.85	6.11
Urea	8.6	28.4
Glycocoll	15	50
Phosphatidylcholine head group ^a	20	66
α -helix ($l = 1.3$ nm) ^b	63.1	208
Proteins ^c	100–1,000	330–3,300
DNA from calf thymus ^d	32,000	105,600

^aAfter Raudino and Mauzerall 1986^bSee text^cAfter Pethig 1979^dAfter Takashima 1963

listed in this table are to be considered just as orders of magnitude. In these cases charged groups are responsible, whose distance apart, according to the size of these molecules, can be considerable. Single α -helices of proteins have a dipole moment, corresponding to a charge $q = e/2$ on both ends of them. Considering a mean length of such an α -helix of $l = 1.3$ nm corresponding to Eq. 2.47, a value of $\mu = 1.602 \cdot 10^{-19} \cdot 1.3 \cdot 10^{-9} = 2.08 \cdot 10^{-28}$ C m = 63.1 D results.

Ampholytic molecules exhibit electrostatic neutrality in the vicinity of the isoelectric point, since just at this pH-value equal numbers of negative and positive charges are dissociated (for example: $-\text{COO}^-$ and $-\text{NH}_3^+$ groups, see Sect. 2.2.8). Because of the reversibility of this process, however, a rapid alteration of dissociation and association may take place. A molecule of this type can be considered as a *fluctuating dipole* the charges of which oscillate in a frequency range of about 10^7 s⁻¹.

If a noncharged molecule is influenced by a strong electric field, shifts of electrons may take place, and an *induced dipole* occurs. The resulting dipole moment depends on the external field strength (**E**), and on the ability of the molecule to polarize, a property which is described by a factor α :

$$\mu_{\text{ind}} = \alpha E \quad (2.48)$$

Such types of induction are measurable as *protein electric response signals*.

At the beginning of this chapter, the dielectric constant (ϵ) had already been introduced. Now we can characterize this parameter in greater detail on the basis of dipole properties. The electric field strength (**E**) of a plate capacitor is proportional to its surface charge density (σ):

$$\mathbf{E} = \frac{\sigma}{\epsilon_0 \epsilon} \quad (2.49)$$

The surface charge density expresses the charges per area ($\sigma = q/A$). The introduction of a dielectric between the two plates of the capacitor means an increase of the dielectric constant (ϵ). As can easily be seen in Eq. 2.49, an increase of the dielectric constant (ϵ) has the same effect as a decrease of the surface charge density (σ). This is not only a mathematical relation. In fact, it is caused by the polarization of the dielectrics under the influence of the electric field of the capacitor. Therefore, there is a correlation between the dielectric constant and the dipole moment of a substance.

2.2.2 The Structure of Water and Ionic Hydration

The properties of water and its interactions with biological molecules has been a very active area of experimental and theoretical research. Recently Raman, infrared, and NMR spectroscopy, magnetic resonance dispersion, dielectric relaxation, neutron scattering, and various time-resolved fluorescence methods have allowed insights into the molecular properties of water. The liquid water itself, all the more, its interactions with ions and molecules are now known to be so complicated that it stands out as a challenging problem. A complete picture of the complex interactions of hydration water molecules that accounts for all these data is still lacking.

The physicochemical properties of water are quite different from those of chemically similar compounds, such as for example H_2Te , H_2Se , or H_2S . Astonishing values could be predicted extrapolating the properties of these substances to H_2O . This would predict a melting point of -100°C and boiling point of -30°C . Water therefore would be a gas under normal life conditions. The same deviations between predicted and existing values would be found in relation to other physicochemical properties, such as molar heat of evaporation, heat capacity, surface tension, dielectric constant, etc.

Another characteristic feature is the so-called *temperature anomaly* of water, which is of particular relevance for the generation and preservation of life. As illustrated in Fig. 2.16, water has its highest density at a temperature of 4°C . This is the reason why, in contrast to other melting substances, ice swims on the surface, and the temperature of water near the bottom of lakes will not become colder than 4°C . If water had the same properties as other comparable substances, the cold water would always sink and the lakes would freeze downside up. Life in such a lake would not survive the winter.

In general, there are two molecular properties which are responsible for these effects: the dipole moment of water, and the ability of water molecules to build intermolecular hydrogen bonds.

The dipole property of the water molecule results from an asymmetric distribution of the common electron pair in the covalent O–H bonds. In other words, the molecular orbitals of the bonds are shifted toward oxygen, as the larger of the two covalently bonded atoms. Hence, in both H–O bonds of the water molecule a strong polarization occurs. The common valence electrons in these bonds are strongly

Fig. 2.16 Density (ρ) and viscosity (η) of water as a function of temperature (T)

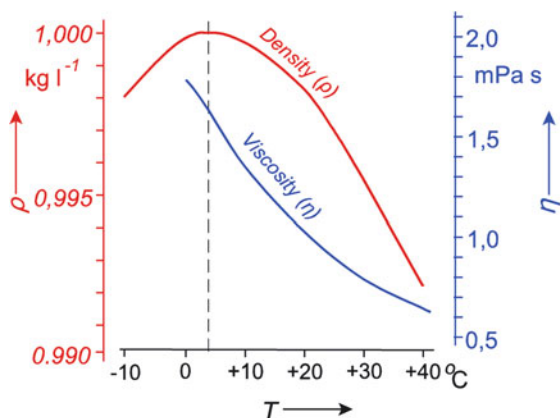
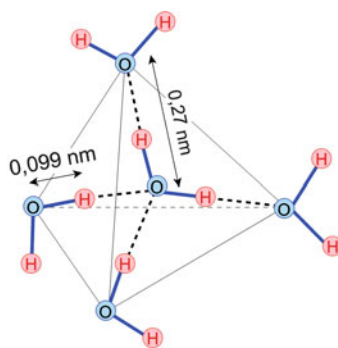


Fig. 2.17 The tetrahedral arrangement of water molecules in ice

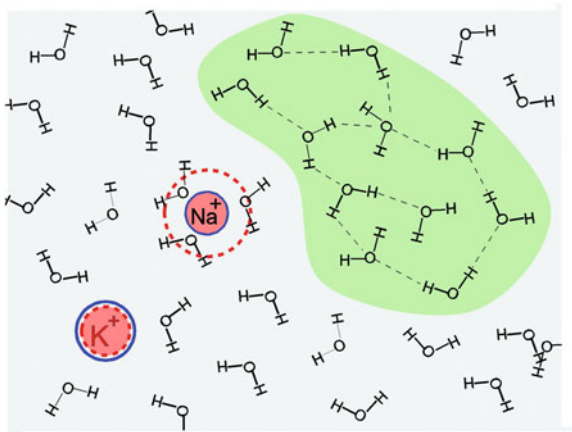


attracted to the oxygen atom, which therefore becomes negative in relation to the protons. The positive charges of the two H atoms lead to a repulsion of each other. This increases the angle between them up to a value of 104.5°. The water molecule as a whole becomes a dipole which is directed towards the bisector of this angle.

This polarization of the H–O bonds in the water molecule not only has consequences for its dielectric behavior but is also responsible for its ability to build hydrogen bonds. They can be established between two water molecules as well as between water and other molecules. Even if the lifetime of these hydrogen bonds is only about 0.5 ps, these kinds of interactions result in quickly fluctuating arrangements, which usually are termed *water structure*.

The structure of ice has received the most attention from investigators. Several ice structure types were found, corresponding to various conditions. The basis of all of them is a tetrahedral structure as shown in Fig. 2.17. It is built by oxygen atoms which are always connected by H-bonds to two hydrogen atoms of other molecules. The continuation of this kind of structure builds up a three-dimensional network which finally leads to the crystalline structure of ice.

Fig. 2.18 Schematic two-dimensional illustration of the arrangement of water molecules which partly are connected to clusters (green region) and partly occur more or less separately (blue region). In the same scale, a Na^+ - and a K^+ -ion is included. The blue circle, surrounded by the full line indicates the crystal radius; the red circles with dashed lines indicate the size of the Stokes radius



These elements of ice structure will not be destroyed fully if the ice melts. Liquid water should be conceived as a rapidly fluctuating mixture of the intermolecular bonding types found in the polymorphism of ice. On average, in this case, one water molecule makes about 3.5 hydrogen bonds. Nemethy and Scheraga proposed already in 1962 that water clusters diminished with increasing temperature. Near the melting point, about 90–650 water molecules are clustered together. Near the boiling temperature, there are only 25–75 molecules connected in such a cluster (the numbers differ according to different authors). In this way one can understand the temperature dependence of the viscosity of water, as depicted in Fig. 2.16.

This property of water molecules is inherent to the mixed cluster model of fluid water as shown in Fig. 2.18. This model suggests that the tetrahedral structure of water, as depicted in Fig. 2.17 is not the only possible geometrical arrangement of the cluster. The strongly oriented crystalline structure of ice is disturbed in floating water. Irregularities as well as supramolecular structures based on nontetrahedral geometry occur.

These structures are to be considered as being in permanent movement. The lifetime of an H-bond is around $0.5 \text{ ps} = 0.5 \cdot 10^{-12} \text{ s}$. The mean lifetime of a single cluster is approximately 10^{-11} – 10^{-10} s . Therefore, during the oscillation, every H-atom will form 100–1,000 times a H-bond with the same oxygen atom before it connects to another one. Because of these dynamics, the name *flickering cluster* is used.

In biological systems the water structure is to a great extent determined by its interactions with ions and organic molecules.

The interaction of ions with water molecules leads to their solvation. Inorganic ions interact with water molecules exclusively electrostatically. In Table 2.1 the dipole moment of water is listed as 1.85 D. The electric field around an ion therefore, will lead to a more or less strong orientation of water molecules. This orientating force is in competition with influences from other water molecules and of course, with the destructive activity of thermic noise. As we discussed in Sect. 2.2.1 (Fig. 2.14), the electric field strength decreases with increasing distance from

Table 2.2 Several characteristic parameters of alkali ions

Ions	Relat. atomic mass	Crystal radius	Hydration radius	Equivalent conductivity at infinite dilution	Relative residence time of a H ₂ O molecule t_i/t
M		r_k (nm)	r_H (nm)	(Sm ² val ⁻¹)	
Li ⁺	6.94	0.069	0.240	0.00387	2.60
Na ⁺	22.99	0.098	0.185	0.00501	1.27
K ⁺	39.10	0.133	0.126	0.00735	0.54
Rb ⁺	85.48	0.148	0.119	0.00778	
Cs ⁺	132.91	0.169	0.120	0.00772	0.59

the center of charge. Correspondingly, two areas at different distances from the ion must be considered: Close to the ion is a region of *primary hydration*. This is a small amount of water molecules that are strongly oriented in the electric field of the ion. The region of *secondary hydration* follows. At this distance the electric field is always too weak to orient the water molecules. Conversely, it is strong enough to disturb the normal water structure. In Fig. 2.18 this is indicated by the destruction of the darkly marked water cluster near the Na⁺-ion. In this, it is like a fault zone with fully disturbed water structure.

To quantify this situation several parameters can be used. They are illustrated in Table 2.2. The amount of water molecules, being more or less affected by the ion is indicated by the *solvation number* and is evaluated by NMR-measurements. Investigations of coefficients of self-diffusion of water, including its activation energy, make it possible to determine the relative residence time of water molecules in the vicinity of ions. *Self-diffusion* means the diffusion of water in water, or in solutions. If t_i stands for the mean time in which the H₂O molecule resists near an ion, and t stands for the time interval, during which such a molecule will permanently stand at any point near another water molecule, the relation t_i/t characterizes the degree of demobilization of these molecules near an ion. If $t_i/t > 1$ then obviously an increase of the structural degree of water near the ion occurs, if $t_i/t < 1$ the ion is a structure-breaking, or *chaotropic* agent.

An important parameter to characterize the hydration shell around an ion is the *Stokes radius*. It corresponds to the effective radius which we introduced in Sect. 2.1.6 for spherical molecules. Measuring the equivalent conductivity of an ion at infinite dilution, it is possible, to calculate its electrophoretic mobility, i.e., its mobility in relation to the applied field strength (see: Eq. 2.84). Equation 2.46 allows one to calculate the force that is applied to an ion in a given electric field (**E**). Under conditions of stationary movement, this driving force is fully compensated by the friction which results from the interaction of the moving ion with the surrounding water. Now let us postulate that this frictional force can be calculated by the Stokes law (Eq. 2.34). This means the ion is assumed to be a macroscopic sphere with a hydrophilic surface. “Macroscopic” means – large enough that water can be considered as a continuum, not mentioning its molecular structure. In this case, knowing the viscosity (η), the driving force (**F**), and the mobility (**v**) of the

ion, the Stokes equation makes it possible to calculate an equivalent radius (r) of it as a sphere.

This *Stokes radius* is a typical example of an *effective parameter*. It is exactly defined by its way of measurement and calculation. Conversely, it does not exist as a visible or at least imaginable border. As indicated in Fig. 2.18, it is in no form identical for example to the thickness of a single layer of water molecules. In fact, the Stokes equation is not the proper law to calculate the real movements of an ion in water as the conditions listed above are not realized in this case. Nevertheless this parameter, because of its exact definition, is an important measure for the hydration properties of ions. However, when using it for example to explain membrane permeability of ions or to discuss other similar mechanisms it is always necessary to be aware of the conditions of its definition.

As illustrated in Table 2.2, the solvation numbers as well as the Stokes radius of the alkali ions decrease with increasing atomic mass. In the case of lithium, the hydration radius is much larger than the crystal radius, whereas conversely for potassium it is already smaller. In this case K^+ and Cs^+ are called *structure-breaking* or *chaotropic* ions. The region of secondary hydration, i.e., the fault zone of destroyed water structure is larger than that of electrically oriented molecules. In contrast to this, in cases of small ions, like Li^+ and Na^+ the region of primary hydration is far larger. These are *structure-building* or *cosmotropic* ions. This category also includes the two-valent cations Mg^{2+} and Ca^{2+} which are not mentioned in the table.

What is the reason for these differences? Both ions, Na^+ as well as K^+ carry the same single positive charge. The Bohr's radius of these ions, i.e., the effective radius of the outer electron shell which is the same as the crystal radius, however, is larger for K^+ than for Na^+ , because of the additional electron orbital in the case of potassium. Where there is a larger distance from the charged nucleus of the ion, the electric field strength will become lower. The ions did not alter the hydrogen bonding network outside their direct vicinity. No long-range structure-making or structure-breaking effects for either cosmotropes or chaotropes take place.

Further Reading

Griffith and Scheraga 2004; Israelachvili 1994; Lynden-Bell et al. 2010; Schmid 2001; Zielkiewicz 2005; Wernet et al. 2004.

2.2.3 Interaction of Water with Macromolecules

Water is an essential compound of all living systems. The specific properties of the water molecule, as listed in the previous chapter, are the preconditions for development and existence of life. We will discuss this in many chapters of this book and consider them over different levels of biological organization.

In general, the chemical activity of the surrounding water influences the functions of proteins at least in the same way as the pH of the medium or some specific ligands. Membrane transport proteins for example, periodically

change their water binding during opening and closing. The hemoglobin molecule adjusts its water content during binding and releasing of oxygen. Similar processes occur in various enzymic reactions. We are only now beginning to understand the various types of water binding reactions and their kinetic parameters. Various theoretical approaches and experimental techniques lead to different models of these kinds of interaction.

One of the phenomenological approaches concerns interactions of water with nonpolar molecules or nonpolar groups of a molecule, which is called *hydrophobic interaction*. This is a dominant factor in the stabilization of biological membranes and other supramolecular structures.

Hydrophobic molecules, in general, behave in relation to the surrounding water like an air bubble, producing a surface tension. From the phenomenological point of view, water molecules surround a nonpolar solute without sacrificing much of the H-bonding; rather, H-bonding networks are stabilized by the presence of other molecules. In view of the tangential orientation of the first water sphere around this solute, a cage is formed which can be compared to an elastic net compressing these hydrophobic structures.

In this way, hydrophobic interactions influence not only the entropy of the macromolecules, but also the entropy of the surrounding water. Because of the structure-building property of the hydrophobic interface, the entropy of the water near the surface is lower than that in the bulk phase. In case of conformational changes of macromolecules, or of processes of supramolecular organization, the surface available for interaction with water molecules may decrease. This happens for example if lipid molecules aggregate to form membranes or vesicles. In this case the water between the molecules will be extruded.

From a thermodynamic point of view, the entropy of the enclosed macromolecule decreases as the result of proper orientation, but simultaneously, the entropy of the surrounding water increases because of the loss of structure-building surfaces. In general, therefore, even if a macromolecular structure with a higher degree of organization results, the entropy of the whole system increases. According to the second principle of thermodynamics therefore, these kinds of processes may occur spontaneously. They are called *entropy-driven*. This is an important mechanism leading to the spontaneous formation of secondary and tertiary structures of macromolecules (Fig. 2.19).

It was found that the desorption of water from surfaces leads to a certain increase in its partial molar volume (Fig. 2.19). Obviously, the more highly organized water molecules at surfaces are packed closer together. In this respect, we should mention the so-called *Le Chatelier principle*. It predicts that if a constraint is applied to a system in equilibrium, the system adjusts to a new equilibrium that tends to counteract the constraint. An increase of the hydrostatic pressure, in this sense, could be counteracted by the decrease of the partial volume of water in the system. In Fig. 2.19 this would mean: an increase of pressure leads to an increase of hydration, and therefore to a decrease in organization of the macromolecular, or the supramolecular structure. Hence, an increase of hydrostatic pressure should destroy structures that are connected by hydrophobic interactions.

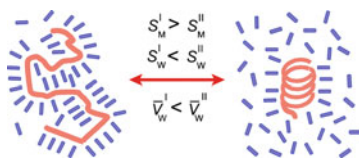


Fig. 2.19 The orientation of water molecules, schematically drawn as bars, near the surfaces of macromolecules. In contrast to the decrease of entropy of the macromolecule during formation of the helix ($S_M^I > S_M^{II}$), the entropy of water during this process increases ($S_W^I < S_W^{II}$). This leads to an increase of the entropy in the whole system. At the same time the partial molar volume of water increases

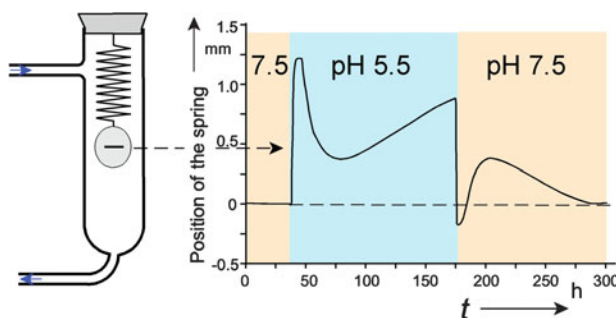


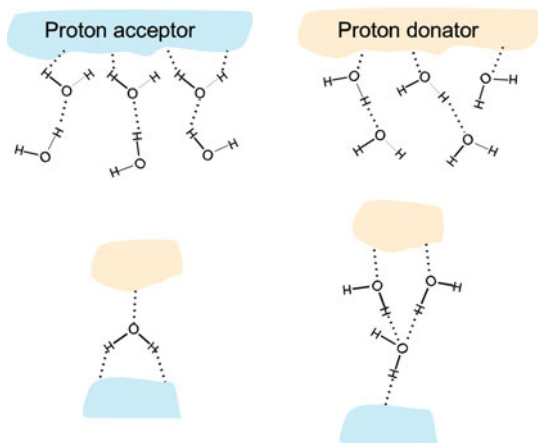
Fig. 2.20 Measurement of the kinetics of a hydration process, using a dialysis sac hanging on a quartz spring. The curve indicates results obtained on the protein of the tobacco mosaic virus (According to data from Lauffer 1975)

In fact, extremely high pressures of the order of 10–100 MPa are necessary for such reactions. Membranes, microtubules, and other supramolecular structures will disintegrate under these conditions. In some cases high hydrostatic pressures were applied to study the physiological properties of membranes. In this context it must be remembered that in the deepest parts of the ocean, i.e., at about 10 km, a hydrostatic pressure of 100 MPa occurs!

Figure 2.20 illustrates how the shift of the partial volume of water can be measured directly during a dissociation–association reaction. For this, in a vessel with exactly controlled temperature the macromolecular sample, enclosed in a dialysis sac, hangs on a very sensitive quartz spring. The surrounding solution of this dialysis sac is exchanged continuously. A modification of the density of the macromolecules will be reflected in a change of the length of the spring. As indicated in the figure, the observable changes are minimal. In this case the change during the pH-dependent polymerization of proteins of the tobacco mosaic virus is demonstrated. The time which is necessary to arrive at equilibrium is quite long.

In fact, there are various kinds of interaction between water and organic molecules. Charged groups of these molecules may interact electrostatically, like ions as described in Sect. 2.2.2, i.e., as a *hydration of the first order*. Attachment of water molecules by hydrogen bonds in contrast is called *hydration of the second order*.

Fig. 2.21 Above: water orientation near proton-accepting and proton-donating surfaces. Below: single and double water bridges



In view of this kind of interaction, two types of molecular surfaces must be considered: proton-donator and proton-acceptor surfaces. As depicted in Fig. 2.21, this difference is responsible for the orientation of the attached water molecules. It may be transmitted by a few layers of water molecules. This kind of hydration can also lead to an interaction between two surfaces. Surfaces with identical orientation of water molecules repel each other, surfaces with opposite orientation of water molecules show attractions at low distances.

In this context Ramachandran already in 1968 coined the term *water bridges* (sometimes even called *Ramachandran bridges*). These connections are important for a number of intermolecular protein interactions forming a ternary structure, as for example the triple helix of collagen. These models proposed two types of water bridges: single water bridges, connecting the molecules over a shorter distance, and double bridges, including three water molecules (Fig. 2.21). Furthermore, dielectric water clusters exist around these bridges, and in some cases inside the protein volume encapsulated bulk water is sequestered.

These circumstances have been established by stoichiometric hydration models, based on predictions from known protein structures and supported by experimental data. A number of methods are available to measure the kind of protein hydration, such as gravimetric measurements (relating the mass of a hydrated protein to one which was dried to equilibrium in vacuum), determination of rehydration rate (indicating various time constants of rehydration), differential scanning calorimetry (direct recording of the energy of the hydration–dehydration process), various proton NMR titration methods, and finally osmotic methods as described in Sect. 3.2.2.

The behavior of water molecules localized in this way near the protein surface, i.e., the time scales of rotational and translation dynamics, as well of the kinetics of its exchange and self-diffusion, shows a wide range of time scales. A complete picture that accounts for all the data is still lacking. In contrast to the hydrogen bond lifetime of 0.5 ps in pure water, NMR and magnetic resonance dispersion show that most water molecules near the protein surface are slowed approximately onefold to

twofold. Quasi-elastic neutron scattering experiments typically show two-step kinetics, with fast (ps) and slow (100 ps–ns) relaxation times. In no case could a memory effect be found, i.e., a conservation of structural properties of water lasting longer times, as argued by homeopaths and other advocates of alternative medicine.

There are also some other peculiarities of the physical properties of water near surfaces in relation to the bulk water. In a layer of 0.1–0.2 nm, for example, the effective dielectric constant of water decreases dramatically up to the value of 10. Furthermore, because of the water structure, this region may become anisotropic. This means, for example, that diffusion rates become higher than the average value in the direction parallel to the surface and lower in the direction normal to it.

Further Reading

Benz and Conti 1986; Fullerton and Cameron 2007; Makarov et al. 1998; Parsegian 2002; Raschke 2006; Takashima 1986.

2.2.4 Ions in Aqueous Solutions, the Debye–Hückel Radius

In the previous sections, we already discussed the electric field that was induced by a single charged point, and which influences molecules in their vicinity. When considering extremely short distances, calculating for example force and energy of ionic bonds, or the influence on water dipoles nearby, it is possible to consider only this single charged point and to use Coulomb's law (Eq. 2.41) for calculation. In the case of larger distances from the source of the field, however, the influence of other charges must also be considered. In a 100-mM solution of NaCl, for example, the mean distance between the ions only amounts to about 2 nm. This means that under physiological conditions the interactions of ions must be considered as a set of charged points.

These kinds of calculations are possible on the basis of the theory of strong electrolytes, the *Debye–Hückel theory*. This theory considers an ionic cloud around an arbitrarily chosen central ion. This central ion is taken as the center of polar coordinates. It attracts ions with opposite charges (counterions) and repels those with identical charges (co-ions). In this way an electric field will build up, generated by the central ion, as well as by the ions surrounding it. This field is spherically symmetric and can be described by the function $\psi(r)$.

At a distance r from the central ion, other ions will be attracted or repelled by the central ion, depending on their charge. We will denote the number of charges of an ion by z_i , for example: $z_{\text{Na}} = +1$, $z_{\text{SO}_4} = -2$, etc. The electrostatic energy of an ion at a particular point in space with an electrical potential ψ amounts to $z_i e \psi$.

Boltzmann's law of energy distribution (Eq. 2.16) makes it possible to calculate the concentration (c_i) of the ions i at a point with the potential ψ . In this case, the energy of an ion in the electric field ($z_i e \psi$) is related to the energy of thermal noise (kT). The orientation of the ions in the electric field is opposed by thermic fluctuations.

$$c_i(\psi) = c_{i0} e^{-\frac{z_i e \psi}{kT}} \quad (2.50)$$

In this equation, c_{i0} is the concentration of the ion i , far away from the influence of the central ion, which means in the bulk solution. To answer our question, we additionally should know the real amount of potential ψ at this point, because it depends on the concentration of the ions in the ionic cloud. Furthermore, we are not interested in function $c_i(\psi)$, but would rather know the concentration and potential distribution in space, i.e., the functions $c_i(r)$ and $\psi(r)$. This is possible using the *Poisson equation*. This is a partial differential equation of the second order, which gives function $\psi(r)$ in a space as a result of a charge density $\rho(r)$:

$$\nabla^2 \psi = -\frac{1}{\epsilon_0 \epsilon} \rho \quad (2.51)$$

$\nabla^2 \psi$ means the second derivation of the potential by the coordinates of the space. This Nabla operator has already been used in Sect. 2.2.1 (Eq. 2.44). For ϵ the dielectric constant of water can be used.

The charge density (ρ , in C m^{-3}) can be calculated from the ion concentration c_i of the ions in the cloud by:

$$\rho = \sum_{i=1}^n c_i z_i N e = F \sum_{i=1}^n c_i z_i \quad (2.52)$$

Introducing Eq. 2.50 in 2.52, and both in 2.51, one obtains the *Poisson–Boltzmann equation*:

$$\nabla^2 \psi = -\frac{F}{\epsilon_0 \epsilon} \sum_{i=1}^n c_{i0} z_i e^{-\frac{z_i e \psi}{kT}} \quad (2.53)$$

The expression (2.53) is a partial differential equation, which can be solved analytically only under simplified conditions. Let us consider the following situation: Let the solution contain only a single kind of ions i , let the potential be very low ($z_i e \psi \ll kT$) and simply a spherical symmetric function from the central ion. For this simplified case, the solution of Eq. 2.53 becomes:

$$\psi_r = \frac{z_i e}{4\pi \epsilon_0 \epsilon r} e^{-\kappa r} \quad (2.54)$$

Comparing this equation with the simple Coulomb's law (Eq. 2.43), it shows that the influence of the potential of the central ion will be diminished by the factor $e^{-\kappa r}$ (considering: $e^{-\kappa r} = 1/e^{\kappa r}$). The extent of this decrease depends on the concentration of the ions in the solution (see Fig. 2.22). This becomes understandable when

considering the definition of the *Debye–Hückel parameter* (κ) in the following equation:

$$\kappa = \sqrt{\frac{e^2 N}{\epsilon_0 \epsilon k T} \sum_{i=1}^n c_{io} z_i^2} = \sqrt{\frac{F^2}{\epsilon_0 \epsilon R T} \sum_{i=1}^n c_{io} z_i^2} = \sqrt{\frac{2 F^2 I}{\epsilon_0 \epsilon R T}} \quad (2.55)$$

The first expression of this equation can easily be transformed into the second one, using the relation between the Faraday constant (F) and the charge of a single electron (e) as: $F = eN$, as well as the relation: $R = kN$. In the last expression of Eq. 2.55, the *ionic strength* (I) was introduced.

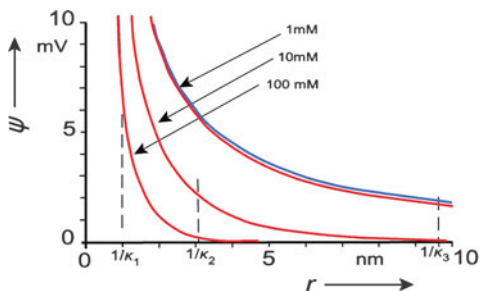
$$I = \frac{1}{2} \sum_{i=1}^n c_{io} z_i^2 \quad (2.56)$$

When calculating the value of κ for real conditions, it can be seen, that this Debye–Hückel parameter has the measure: m^{-1} . (For this of course, the concentration measure: mole per m^3 must be used!). The parameter $1/\kappa$ therefore is a measure of distance. Accordingly, it is called the *Debye–Hückel length* or *Debye–Hückel radius* of the ionic cloud. Considering Fig. 2.22 it is obvious that this value $1/\kappa$ again is not a “visible” radius, but just an effective parameter.

In the following example we will demonstrate the calculation of the ionic strength of a solution: Let us consider a physiological solution with the following composition: 105 mM NaCl, 5 mM KCl, 25 mM Na_2HPO_4 , 2 mM CaCl_2 (mM representing the usual measure of concentration: mmol l^{-1}). Let us assume a full dissociation of the Na_2HPO_4 . Using Eq. 2.56, one obtains:

$$\begin{aligned} I &= \frac{1}{2}(c_{\text{Na}} z_{\text{Na}}^2 + c_{\text{K}} z_{\text{K}}^2 + c_{\text{Cl}} z_{\text{Cl}}^2 + c_{\text{HPO}_4} z_{\text{HPO}_4}^2 + c_{\text{Ca}} z_{\text{Ca}}^2) \\ &= \frac{1}{2}(0.155 \cdot 1^2 + 0.005 \cdot 1^2 + 0.114 \cdot 1^2 + 0.025 \cdot 2^2 + 0.002 \cdot 2^2) \\ &= 0.191 \text{ mol l}^{-1} \end{aligned}$$

Fig. 2.22 The electrical potential (ψ) as a function of the distance (r) from a central ion. The blue line ignores the influence of the ionic cloud according to Coulomb’s law (Eq. 2.43); the red lines show the potential $\psi(r)$ which is diminished by the ionic clouds in solutions of various concentrations of a one-one-valent salt ($z = \pm 1$) according to Eq. 2.54



According to Eq. 2.55, the Debye–Hückel radius (in: nm) at temperature $T = 298$ K can easily be calculated by the following equation, using as concentration measure $\text{mol}\cdot\text{l}^{-1} = \text{M}$:

$$\frac{1}{\kappa} = \frac{0.304}{\sqrt{I}} \quad (\text{in nm}) \quad (2.57)$$

This means for the case of the above considered solution:

$$\frac{1}{\kappa} = \frac{0.304}{\sqrt{0.191}} = 0.696 \quad (\text{in nm}) \quad (2.58)$$

Summarizing these considerations, we conclude that with the Poisson–Boltzmann equation (Eq. 2.53) and its integration, the electrical potential can be calculated dependent on the distance from a central ion. In this case the screening effect of the ionic cloud is considered. The degree of this screening, in fact, increases with increasing ionic strength of the solution. The Debye–Hückel radius ($1/\kappa$) is defined as a measure of the thickness of the ionic cloud, which, in contrast to the crystal radius or the hydration radius, is not a property of a specific ion but just depends on the ionic strength of the solution. The Debye–Hückel radius is an effective parameter in the same way as the others (see Sect. 2.2.2).

It must not be forgotten when using the Debye–Hückel equation that already a number of simplifications were used for its derivation and for integrating the Poisson–Boltzmann equation (Eq. 2.53). Furthermore, several other interactions were ignored. In fact, already in solutions of physiological ionic strengths, a number of experimental data, such as the activity, or osmotic coefficients of ions, show significant deviations from the theoretical predictions. Much effort has gone into extension of this theory, from adding empirical correction factors, to theoretical extensions using advances in the statistical mechanical theory of liquids, or quantum mechanical approaches. The most limiting condition of the Debye–Hückel theory is the ignorance of specific interactions of the ions with other charged groups and molecules. This concerns a number of electrodynamic interactions which will be explained in the next chapter.

Further Reading

Loehe and Donohue 1997.

2.2.5 Intermolecular Interactions

Intermolecular interactions, beginning with various enzyme reactions, receptor properties up to the self-assembly of biological structures represent an intricate process consisting of an interplay between a variety of different intermolecular forces. In Sects. 2.2.2 and 2.2.3 we have already explained the role of hydrogen bonds in building

the structure of water, in interactions among ions and between ions and water dipoles, as well as the role of hydrophobic interactions in the process of self-organization of macromolecules. In the following the spectrum of these interactions will be extended to include additional electrostatic and electrodynamic forces. In fact, any kind of interaction between molecules in an aqueous medium is the sum of several forces which differ by their sign (i.e., repulsive or attractive), and by their distance function.

Let us first consider a dipole interacting with an ion. In this case, the energy of electrostatic binding (E_{ID}), can be calculated using Coulomb's law (Eq. 2.41). In the simplest case it results from the sum of all electrostatic components, namely the interaction of the ion (q_1) with one ($-q_2$), as well as with the other ($+q_2$) charge of the dipole. Considering a linear orientation of the dipole in relation to the ion (Fig. 2.23a), it gives:

$$E_{ID} = \left[-\frac{q_1(-q_2)}{4\pi\epsilon_0\epsilon r} \right] + \left[-\frac{q_1q_2}{4\pi\epsilon_0\epsilon(r+l)} \right] \quad (2.59)$$

We will consider simply that the dipole $\mu = q_2 l$ contains two charges with equal size but opposite signs. A simple algebraic rearrangement of Eq. 2.59, and the introduction of the dipole moment (μ) results in:

$$E_{ID} = \frac{q_1q_2l}{4\pi\epsilon_0\epsilon r(r+l)} = \frac{q_1\mu}{4\pi\epsilon_0\epsilon r(r+l)} \quad (2.60)$$

Assuming: $r > l$, Eq. 2.60 will be simplified to:

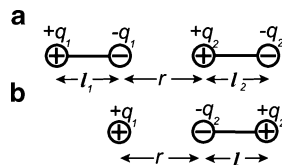
$$E_{ID} \approx \frac{q_1\mu}{4\pi\epsilon_0\epsilon r^2} \quad (2.61)$$

In contrast to the energy of interaction of two ions (Eq. 2.42) as a function of $1/r = r^{-1}$, this equation shows that the interaction energy of an ion with a dipole decreases with r^{-2} .

The mutual interaction between two dipoles (E_{DD}) can be calculated in a similar way. Let us again look at the simplest case of collinear orientation of both dipoles, as depicted in Fig. 2.23b. Using the same approach, one gets:

$$E_{DD} = -\frac{1}{4\pi\epsilon_0\epsilon} \left[\frac{q_1q_2}{l_1+r} + \frac{q_1(-q_2)}{l_1+r+l_2} + \frac{(-q_1)q_2}{r} + \frac{(-q_1)(-q_2)}{r+l_2} \right] \quad (2.62)$$

Fig. 2.23 Ion–dipole (a) and dipole–dipole (b) interaction. q charge, l length of the dipoles, r distance



Again we will rearrange this equation and introduce the dipole moments μ_1 and μ_2 :

$$\begin{aligned} E_{DD} &= \frac{q_1 l_1 q_2 l_2}{4\pi\epsilon_0\epsilon} \cdot \frac{l_1 + l_2 + 2r}{(l_1 + r)(l_2 + r)(l_1 + l_2 + r)r} \\ &= \frac{\mu_1 \mu_2}{4\pi\epsilon_0\epsilon} \cdot \frac{l_1 + l_2 + 2r}{(l_1 + r)(l_2 + r)(l_1 + l_2 + r)r} \end{aligned} \quad (2.63)$$

Using the same approach: $r > l$, gives:

$$E_{DD} \approx \frac{\mu_1 \mu_2}{4\pi\epsilon_0\epsilon} \cdot \frac{2r}{r^4} = \frac{2\mu_1 \mu_2}{4\pi\epsilon_0\epsilon r^3} \quad (2.64)$$

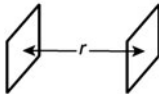
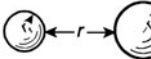
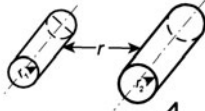

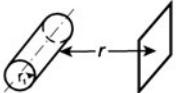
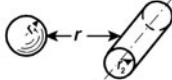
This sequence of possible electrostatic interactions can be extended including interactions with induced dipoles (see Sect. 2.2.1). In this case, a more complicated function of the distance must be expected because the induction of a dipole itself depends on the electric field (Eq. 2.48). For this case, a function can be calculated, including r^{-4} . If an interaction of two dipoles is considered, inducing each other, a function with r^{-6} will be obtained.

This brings us to the wide field of *van der Waals interactions*. The plural of this expression indicates that it includes a number of different intermolecular forces. Van der Waals interactions in general, are attractive forces between molecular components, whole molecules, or supramolecular particles that are not simply based on electrostatic interactions. They are rather attributed to electromagnetic interplay, occurring by fluctuations of charges. The reason for such charge fluctuations in molecules may be different. There can be thermal molecular translations, or oscillations in the thermic level of electrons, as illustrated in the Jablonski diagram (Fig. 4.34, Sect. 4.8.2). Furthermore, fluctuations occur in the electron structure of molecules which can be calculated by quantum mechanics. This is the basis of the so-called *dispersion forces*, or *London-dispersion forces*, sometimes also called *London–van der Waals interactions*. These oscillations occur at frequencies near to that of visible light.

The understanding of van der Waals forces was substantially advanced by F. London, H. C. Hamaker, and E. M. Lifschitz. Nevertheless, many aspects are unclear even today. The central question concerns the character of this interaction, and its energy as a function of interparticle distance (r). Furthermore, it still seems unclear whether in this kind of interactions a molecular specificity can exist which might be related to its frequency.

The energy of van der Waals interactions is governed by the *Hamaker constant*, which results from quantum mechanical calculations. The distance relation of this energy depends strongly on particle geometry. This is illustrated in Fig. 2.24. There is a difference whether it is an interaction between a spherical and/or a cylindrical particle and additionally, there are different functions $f(r^{-n})$ dependent on the total

Fig. 2.24 The distance function [$f(r^{-n})$] of the van der Waals energy of interaction between particles of various shapes (parameters after Israelachvili 1994)

Geometric conditions	Relation of distances	
	$r_1, r_2 \geq r$	$r_1, r_2 \leq r$
	r^{-2}	
	r^{-1}	r^{-6}
	$r^{-3/2}$	r^{-5}
	r^{-1}	r^{-3}
	$r^{-3/2}$	r^{-3}
	r^{-1}	r^{-5}

distance between these particles. This means at the same time that simple hyperbolic functions for this kind of interactions are just rough approximations.

To understand the relevance of van der Waals forces, it is necessary to consider them in context with electrostatic interactions. Here we touch upon a central concept of colloid chemistry, especially the concept which is connected with the names of B. V. Derjaguin, L. D. Landau, E. J. W. Verwey, and J. Th. G. Overbeek. This *DLVO theory* was originally developed for modeling the behavior of lyophobic colloids (see: Verwey and Overbeek 1948). To extend this approach to proteins, they are considered to be hard, polarizable spheres with charges, uniformly distributed over their surface. Water is treated as a structureless, continuum dielectric medium. The basic ideas of this approach, considering the Debye–Hückel theory of ionic clouds, as well as the van der Waals properties, can be listed as follows:

- The interaction force between two particles results from the sum of the particular forces, each one decreasing with increasing distance in different ways.
- In contrast to generally attractive van der Waals forces, electrostatic interactions can also become repulsive.
- The intensity of electrostatic interactions, in contrast to van der Waals forces, depends to a high degree on environmental conditions.

Fig. 2.25 Functions: $y = r^{-n}$ to illustrate the character of short- and long-distance interactions

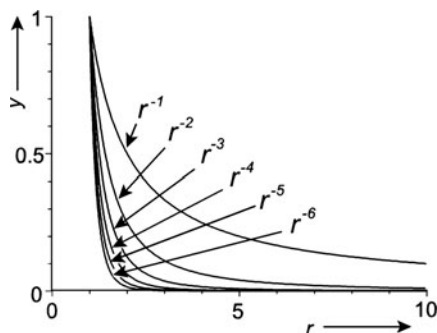
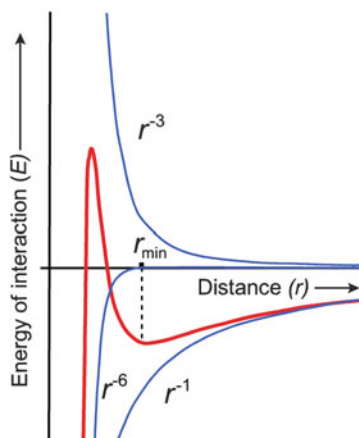


Fig. 2.26 The total energy of interaction (E) (red line) between two particles as a sum of three functions (blue lines): One type of electrostatic repulsion (r^{-3}), and two kinds of van der Waals attractions (r^{-1} and r^{-6}). The point r_{\min} indicates the distance of the secondary minimum



To illustrate the consequences of these postulates, we must imagine that the distance function of the interaction energy [$E(r)$] results as a sum of various functions, shown schematically in Fig. 2.25. There are so-called *short-distance interactions*, decreasing with r^{-5} or r^{-6} , and *long-distance interactions*, the curves of which do not decrease as fast (r^{-1}). Some of these curves can be looked at upside down, representing repulsive forces, which consequently, must be subtracted by the others. Additionally, the dependence of the electrostatic forces on the ionic strength of the solution is to be considered, as illustrated in Fig. 2.22, as well as the various distance functions of the van der Waals forces, as depicted in Fig. 2.24. Hence, the shape of the resulting sum can be quite complicated.

Figure 2.26 illustrates this situation for a more or less simple case of the interaction of two particles in a solution. In this case the interaction is supposed to be the sum of an electrostatic repulsion (r^{-3}) and two kinds of van der Waals attraction forces, one as a long (r^{-1}), and another as a short-distance interaction (r^{-6}). This curve demonstrates at least qualitatively some basic properties of macromolecular interactions. It shows that two particles may attract each other at larger distances, even if they carry charges of the same sign, if additionally a long-ranging van der

Waals attraction occurs. This is demonstrated in Fig. 2.26 at distance $r > r_{\min}$. The point r_{\min} marks the position of minimum interaction energy. The attractive force ($F = dE/dr$) at this point becomes zero. If this distance is decreased a repulsion force appears. The particles therefore, will rest at this point, which is called the distance of the *second minimum*. If in any way this repulsion force were to be overcome, attractive interactions again would take place and the particles would aggregate at the distance of the *primary minimum*. In Fig. 2.26 this primary minimum is not shown. This would predict an inclusion of further components of interactions at very low distances. So for example, the influence of Bohr's radii must have been considered, the extremely strong repulsion forces in the case of interactions of the electron shells.

As already mentioned, the electrostatic interaction strongly depends on the ionic conditions of the medium. This is understandable in view of Eq. 2.54 (Sect. 2.2.4), as well as Fig. 2.22. Increasing the ionic strength, for example, the energy barrier in the region $r < r_{\min}$ will become diminished. Finally, thermic hits could be sufficient to let the particles cross this barrier and to agglutinate. This effect is experimentally verified. An increase of the salt concentration of a colloid solution leads to the agglutination of the particles. This is known as the *salt-out effect*, in contrast to the *salt-in* process, leading to a re-stabilization of a colloidal system.

Despite the general success of DLVO theory, there are many situations where it is unable to provide even a proper qualitative description of macromolecular interaction. The reason for this, particularly considering protein–protein interactions, is the ignorance of a number of circumstances, such as for example:

- Protein–ion interactions cannot be simply expressed by ionic strength. The limitations of the Debye–Hückel theory as mentioned in the previous section must be considered. This particularly concerns specific properties of anions. In this respect, the DLVO theory is unable to predict any dependence on salt type.
- The proteins cannot be taken simply as charged spheres. The charged groups are rather distributed not homogeneous. This results in an anisotropic character of protein–protein interactions. In fact, weak protein–protein interactions should be treated more like molecular recognition processes.
- Hydrophobic interactions must be considered. Protein–protein interactions are very sensitive to surface properties, such as protonation states of the charged residues on the surfaces, or of surface hydrophobicity.

In 1888, the German pharmacologist Franz Hofmeister, investigating the colligative properties of colloids, had already determined that the ability of an ion to induce protein precipitation strongly depends on its type. A *Hofmeister series* was developed as a ranking of the salting-out effectiveness of various ions for globular proteins. The effectiveness is much greater for anions than for cations in the decreasing order: $\text{SO}_4^{-2} > \text{HPO}_4^{-2} > \text{CH}_3\text{COO}^- > \text{Cl}^- > \text{Br}^- > \text{I}^- > \text{SCN}^-$, and correspondingly: $\text{Li}^+ > \text{Na}^+ \sim \text{K}^+ > \text{NH}_4^+ > \text{Mg}^{2+}$. This sequence refers to properties of anions and cations, on a wide range of phenomena. Particularly the phosphate is of physiological interest, with its position at one end of the Hofmeister sequence, and its strong adsorption at biological surfaces.

The Hofmeister series cannot be explained by the classical DLVO theory. One reason for this is that van der Waals interactions and other electrodynamic interactions until recently were not taken into account for the description of ion–ion interactions. Already at physiological concentrations the Coulomb interactions are more or less screened. This is not the case for the dispersion forces which react under these conditions. In the case of physiological solutions, where the Debye–Hückel length amounts to only about 0.6 nm (see Eq. 2.58 in Sect. 2.2.4), dispersion forces may become important even for ions.

In molecular biology these kinds of interplay between various forces of intermolecular interactions are very important. This is one way to control supramolecular processes. The influence is possible either by control of the dielectric constant, or by changing the ionic conditions. The local dielectric constant depends on the arrangement of molecules, on their polarization and on the water structure in their vicinity. Mostly, these processes, however, are controlled by changing local ionic compositions.

In this context the influence of multivalent ions is of particular interest. Let us mention the role of calcium in muscle contraction, nerve excitation, exocytosis, cell adhesion, and many other processes. In colloid chemistry the so-called *Schulze–Hardy rule* is an empirically determined relation which later was also verified theoretically. This rule predicts that the concentration of a salt, which would be necessary to overcome the energy barrier between the second and the first minimum, is inversely proportional to the 6th degree of its valency. This means that the critical concentration produced by a univalent ion relates to this concentration of a bivalent one, like 1^6-2^6 , which means: $1-1/64$. Consequently, 2-mM Ca^{++} would be as effective as 128-mM K^+ or Na^+ . This is a relation that reminds us of the real concentrations of in vivo conditions. It must however be considered that this would not explain the difference in the behavior of Ca^{++} in contrast to Mg^{++} . Other reasons must also be taken into consideration.

Further Reading

For van der Waals interactions in molecular biology: Israelachvili 1994; Curtis and Lue 2006; Hofmeister effect: Kunz et al. 2004; Zhang and Cremer 2006.

2.2.6 Structure of Proteins

In the preceding sections problems relating to the information content of biomacromolecules were discussed, the idea of random coil formation by molecular chains as a state of maximum entropy was introduced, as well as the role of water in the entropy balance of macromolecular systems, and the broad spectrum of intermolecular interactions was outlined. Knowledge of these aspects is an important prerequisite to understand the dynamic structure of proteins as a polypeptide chain with a unique spatial structure guaranteeing its specific function in the living cell.

The first X-ray crystallographic structure of a protein in 1959 (M. F. Perutz and J. Kendrew – Nobel Prize 1962) revealed a surprising amount of irregularities, and an apparent lack of symmetry as compared to the simple double-stranded DNA structure observed 5 years earlier. The process by which a polypeptide chain acquires its stable three-dimensional structure to achieve its specific biological activity, the so-called *native state*, i.e., the process of *protein folding*, remains a key problem of molecular biophysics (see Sect. 2.2.7). As a central dogma of molecular biology, all the necessary information for the protein to achieve the native secondary and tertiary structures is buried in the amino acid sequence per se.

Proteins are composed of α -L-amino acids which are covalently linked to each other by peptide bonds, eventually forming long chains (see Fig. 2.27). The peptide bond is formed by the carboxyl group of the α -C-atom of one, and the amino group of the following monomer. By a particular type of resonance interaction the corresponding C–N bond is shorter than a usual C–N bond, and resembles the resonance of the C=O double bond. As a result of this partial enol character of the peptide bond the group of six atoms – C_α –CONH– C_α – is planar. Therefore, the polypeptide appears as a sequence of planar structures that are twisted only by the rotating bonds with angles φ and ψ (see Fig. 2.27). Nevertheless, depending on the specificity of amino acids, steric constraints only allow distinct angles of rotation. The allowed angles φ and ψ determined by the particular van der Waals distances are usually plotted in the so-called *Ramachandran diagram*.

The linear sequence of amino acids as demonstrated in Fig. 2.27 is called the *primary structure* of the protein. The symbols **R** in this figure stand for the side chains, responsible for the specific properties of the 20 naturally occurring amino acids. By interactions of various parts of these chains, a three-dimensional arrangement is established – the *secondary structure* of the molecule.

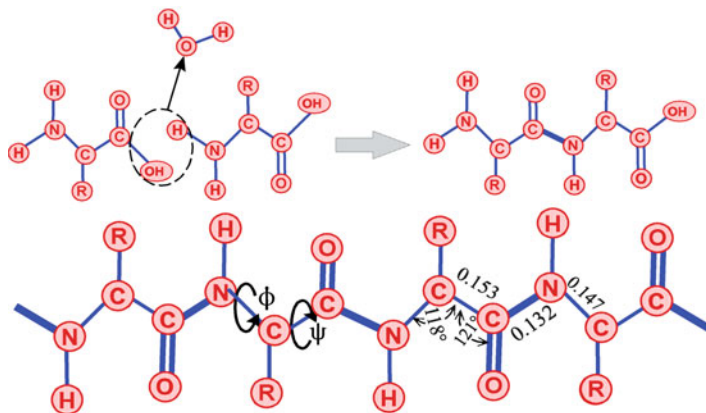


Fig. 2.27 Above: formation of the peptide bond between the amino acids, as a reaction of condensation, releasing a molecule of water (see text). Below: parameters of a polypeptide chain: R – side chains characterizing the particular amino acids, φ and ψ – angles of rotating bonds

Depending on the parameters of angles φ and ψ , two principal types of secondary structures are possible: the α -helix and the β -pleated sheet. By repeating angles of φ and ψ the β -pleated sheet is formed. The stability of these sheets increases through parallel or antiparallel arrangements of several sheets that are connected noncovalently by intermolecular hydrogen bonds. The side chains **R** in this case are oriented perpendicular to the plane of the monomers. Mostly the folded elements themselves form a right-hand spiral.

The α -helix is a rigid arrangement of the polypeptide chain that is twisted by right-hand torsion of the angles $\varphi = -57^\circ$ and $\psi = -47^\circ$ (see Fig. 2.28). Therefore it is a right-handed spiral, where each amino acid corresponds to a 100° turn. Hence, there are 3.6 amino acids per turn. The side chains (**R**) are oriented perpendicular to the helical axis, avoiding steric interference with the polypeptide backbone. In contrast to the β -pleated sheet, the α -helix is stabilized not by *intermolecular*, but by *intramolecular* hydrogen bonds in such a way that the peptide N–H bond of the n -th residue is connected with the C=O group of the $(n-4)$ -th residue.

In general there are also arrangements of other polypeptide helices possible. They are usually described in an n_m -nomenclature. In this case n means the number of residues per helical turn, and m is the number of atoms, including H, in the ring that is closed by the hydrogen bond. The α -helix as the most abundant type in this nomenclature is a 3.6_{13} -helix, corresponding to 3.6 amino acids per turn. Approximately 15–20% of all helical structures are so-called 3_{10} -helices, which also are right-handed containing however only three residues per turn. They are typically only three to five residues long compared with a mean of 10–12 residues for 3.6_{13} -helices. These 3_{10} -helices have been proposed to be intermediates in the folding/unfolding dynamics of α -helices. In principle, also a left-handed 4.4_{16} -, a so-called

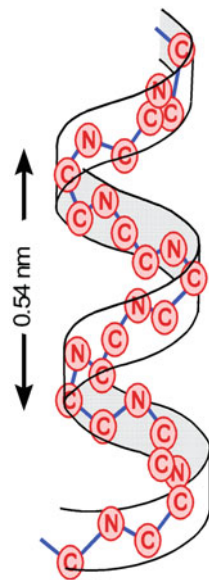


Fig. 2.28 The right-handed α -helix (for more clarity, only the backbone-forming atoms of the polypeptide chain are depicted)

π -helix is possible. It is, however, unlikely to be observed in proteins because only glycine is likely to adopt this chirality.

The α -helix is a typical secondary structure of the membrane-spanning part of most integral membrane proteins (see also Figs. 2.39 and 3.32). Essentially, amino acids with hydrophobic side chains stabilize this secondary structure in the lipid bilayer. Sometimes these proteins span the membrane by several α -helices stabilizing each other (for more detail see Sect. 2.3.3).

One has always tried to find correlations between amino acid sequences and their secondary structures, to derive rules for structure prediction. In fact, the secondary structure of a protein is determined through the collective interactions between the elements of the molecule: the specific properties of a single amino acid alone do not by themselves exclusively determine which type of secondary structure will be seen. For example, the five negatively charged glutamic acids in five places in the chain A subunit of bovine pancreatic ribonuclease are organized in different secondary structures: GLU 2 and GLU 49 are in a coil configuration, GLU9 is in the α -helix, and GLU111 is in a β -sheet.

There are, however, specific sequences of amino acids that tend to form helical structures, while others form β -sheets or unstructured regions connecting different secondary structures. For example, collagen as a typical structural protein forms a triplet-helix as a tertiary structure. This is already encoded in the primary structure by a repeating sequence $(\text{GLY-X-Y})_n$ where GLY stands for glycine, and X and Y resemble any other amino acids. Mostly these are proline and hydroxyproline. This repeated occurrence of glycine in every third position stabilizes the triplet-helix via hydrogen bonds. In some cases amino acids with the ability to form disulfide bridges provide a kind of starting point for the formation of super helices. In this case, as in the formation of α -helices, the process resembles a zip fastener. In a similar way myosin forms a left-hand double super-helix. The turn of this helix is predicted by the arrangement of amino acids in the corresponding α -helices of the secondary structure.

Further Reading

Voet and Voet 2011.

2.2.7 Protein Folding and Protein Dynamics

A central question in molecular biophysics is the mechanism of protein folding, i.e., the formation of its tertiary structure. Already in 1968 Cyrus Levinthal showed that protein folding cannot be considered simply as a stochastic process of reversibly connecting and disconnecting bridges between monomers to eventually find a structure of minimal free energy. This can easily be reduced to absurdity, considering the time which would be necessary for this process: Let us take a protein with 300 amino acids, each of which should have eight rotational positions. In this case $8^{300} = 10^{270}$ positions of the whole molecule are possible. Even in a case where, by

reason of steric hindrance the number of positions would be reduced, and even if the conformational changes were extremely fast, the time needed to arrive at the correct structure would be more than astronomically long. Even a polypeptide chain of only 150 residues could adopt 10^{68} possible conformations. Effectively folding into the native structure would require about 10^{52} years if we assume a mean time of 1 ns for each conformation. Actually, the folding time of a protein is between 0.1 and 1,000 s.

To solve this “*Levinthal paradox*,” a minimal set of parameters must be found directing the process of protein folding to a suitable reaction coordinate. First the two principal pathways of protein synthesis must be considered: on one hand the *cotranslational* folding, where the native conformation of protein emerges directly during the biosynthesis on the ribosome, and on the other hand the *post-translational* folding, where the polypeptide chain starts to fold only after its full synthesis by the ribosome. Many experiments *in vivo* and on model systems have demonstrated that almost all proteins fold cotranslationally. The possibility of post-translational folding is supported by denaturation–renaturation experiments because many denatured small globular proteins restore their initial state quickly after removal of the denaturing agent. However, refolding of an unfolded protein probably is not an optimal model to study the protein folding *in vivo*, because the protein is not completely unfolded even under denaturing conditions.

Furthermore, it should be considered that proteins fold in different compartments of the cell. Some of them fold in the cytosol near the ribosomes, others are transported to different cellular locations, such as for example to the lumen of the endoplasmic reticulum across membranes. The cell compartments where proteins do fold are either aqueous or even hydrophobic such as the interior of the membrane. The targeting to these compartments is directed by the amino-terminal signal sequences.

In the last decades a number of fast spectroscopic techniques with a dramatically improved time resolution have been developed, such as photochemical triggering, temperature and pressure jump, ultrarapid mixing methods, etc. They allowed insights into the fast processes of protein folding. The first observable event in the folding pathway of a number of proteins, lasting a few milliseconds, is a collapse of the flexible disordered polypeptide chain into a partly organized globular state, which is called the *molten globule*. In this structure, hydrophilic parts of the chain are exposed externally to the aqueous environment, whereas other parts build hydrophobic centers inside the globule. Hence, a kind of microphase separation occurs. The generic collapse increases the packing density, and results in an extensive amount of entropy reduction of the entire molecule. Nevertheless, this is a typical entropy-driven process because the entropy changes of hydration of polar and nonpolar groups are both positive. The resulting conformational entropy loss caused by backbone packing, as well as side-chain interaction will be compensated by an increase in entropy of the surrounding water molecules (see Fig. 2.19).

To solve the Levinthal paradox, an energy landscape was defined for the system of folding as a mapping of the chain conformation to its internal energy, along with particular rules defining what transient intermediates are accessible from a given configuration, and how the protein can switch between them. Since the native

structure of a protein is the lowest in energy, one should expect the shape of the landscape for a protein to have a funnel-like topography, at least in the vicinity of the native structure. The extremely high degrees of freedom of a polypeptide chain make the funnel multidimensional. It was proposed to reduce this multidimensional picture into one with few relevant coordinates (see Fig. 2.29).

In fact, protein folding is a hierarchic and multiphasic process consisting of a number of particular reactions with quite different time constants. There is no singular way through the multidimensional energy funnel, but rather a large number of bifurcations and bypasses. This is an explanation for differences in the time course to eventually acquiring a native protein.

Various cellular mechanisms have been found to ensure a faster and proper protein folding. In the living cell this process is mostly assisted by the so-called *chaperones*. This is an ancient French word, standing for an older woman accompanying a young girl to guarantee her virtue. Chaperones, such as for example heat shock, or chaperonin proteins facilitate protein folding, and recognize and repair misfolded proteins. By reversible binding to unfolded polypeptides and separating them from bulk cytosol they dramatically accelerate folding and prevent misfolding and precipitation. Folding of about one third of all proteins in living cells is coordinated by such cofactors.

Because chaperones simply act by binding reversibly to specific sequences of the polypeptide, not modifying their chemical structure, it is not correct to use the term “structure-forming enzymes” for them. Such are for example the enzymes protein-disulfide-isomerases and poly-*cis/trans*-isomerases which also promote the process of folding. In contrast to chaperones, they modify the polypeptides directly. They are necessary, because about 93% of the peptide bonds are primarily synthesized in *trans*-form. To generate a native protein, a *trans-cis* transformation is necessary, which generally proceeds quite slowly but can be significantly accelerated by these enzymes.

Most investigations into protein folding are performed by in vitro studies. As already mentioned above, the process of refolding of denatured proteins is not directly comparable with in vivo protein folding, because a number of structural elements remain preserved even after denaturation. A second difference between studies in vitro versus such in vivo are the specific conditions in the cell. Most in vitro studies are performed in buffered solution with <1% protein concentration. In contrast, concentrations of these solutes in the living cell can reach hundreds of grams per liter. This crowding environment can have a significant impact on the stability and the native structure of a protein. Hence, it can significantly change the energy landscape of protein folding.

The difference of the free molar Gibbs energy (ΔG) between the native versus the denaturated state of a protein, as depicted for example in Fig. 2.29, is sometimes below 50 kJ mol^{-1} . Considering that about 100 noncovalent interactions between amino acids of neighboring chains must be disrupted during denaturation, their interaction energies must be lower than the energy of thermal motion ($RT = 2.5 \text{ kJ mol}^{-1}$ for $T = 300 \text{ K}$). The stability of the protein structure, therefore, can be explained only as a phenomenon of cooperativity.

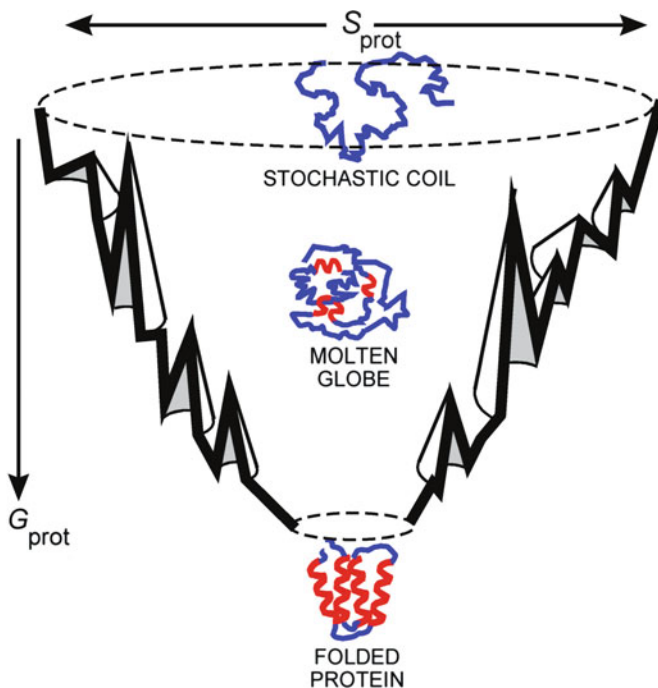


Fig. 2.29 Three steps of protein folding on a schematic energy landscape which has an overall funnel topography. The funnel is a three-dimensional projection of a many-dimensional space corresponding to the intricate connectivities between configurational states. G_{prot} the free energy of the conformational state, S_{prot} a measure of the configurational entropy of the protein (Corresponding to Plotkin and Onuchic 2002)

Conversely, the protein as a functional unit in the biological system not only needs sufficient structural stability, but must rather be considered as a highly dynamic structure. Obviously, in native proteins small conformational transitions in the form of stochastic fluctuations are the precondition for their function as enzymes or transport proteins. Further modifications may be controlled by ligands, by changes of the local ionic conditions, by temperature, or in the case of membrane proteins by the electric fields. In particular, the electric membrane voltage regulates ion channels, causing the action potentials in neurons (see Sect. 3.4.4).

These fluctuations can be expressed as the average square of displacement of various parameters such as the molecular internal energy ($\langle \Delta U^2 \rangle$), or volume ($\langle \Delta V^2 \rangle$) or even as amplitude of motion of an individual part of the molecule ($\langle \Delta x^2 \rangle$). They are functions of the energy of thermal noise (kT), the molecular mass (m), the molecular volume (V), the isochoric heat capacity (C_V), as well as the isothermic volume compressibility (β_T):

$$\langle \Delta U^2 \rangle = kT^2 m C_V \quad (2.65)$$

$$\langle \Delta V^2 \rangle = kTV\beta_T \quad (2.66)$$

Let us consider a globular protein with a molecular mass of $M = 25,000$, the mass (m) of an individual molecule therefore is:

$$m = \frac{M}{N} = 4.15 \cdot 10^{-20} \text{ g} = 4.15 \cdot 10^{-23} \text{ kg}$$

Assuming further the following case: $V = 3.2 \cdot 10^{-26} \text{ m}^3$, $C_V = 1.34 \text{ kJ kg}^{-1}$, $T = 298 \text{ K}$, as well as $\beta_T = 2 \cdot 10^{-11} \text{ Pa}^{-1}$, from Eq. 2.65 results $\Delta U = 2.61 \cdot 10^{-19} \text{ J}$ per molecule. Multiplying this by the Avogadro number ($N = 6.022 \cdot 10^{23} \text{ mol}^{-1}$), a molar energy of 157 kJ mol^{-1} results for the case that all molecules would fluctuate synchronously. This is already close to the energy that would be sufficient for the denaturation of the protein. The deviations in the volume of this molecule are given by Eq. 2.66 for: $\Delta V = 5.1 \cdot 10^{-29} \text{ m}^3$, i.e., about 0.16% of its total volume.

This rather simple approach, however, does not reflect the true detail of the functional dynamics of the protein. Recently, methods have become available that allow us to determine fluctuations of selected parts of the molecule. For example, elastic incoherent neutron scattering (EINS), provides information about the motion of individual nuclei, especially by replacing selected hydrogen atoms by deuterium. These investigations show that the mean-square amplitudes of motion ($\langle \Delta x^2 \rangle$) in particular segments of the molecule, as for example the active center of myoglobin, are smaller than those of the rest of the protein. These parts of lower motion can be considered as an aperiodic solid, whereas the outside of the protein has semiliquid-like properties.

It is a question of biological optimization: why do proteins not have a higher degree of stability? In fact, proteins must have enough structural stability for maintenance of their specific three-dimensional conformations which is required for their function. But this stability must not prevent rapid and precise alterations of protein subdomains as required for function. Furthermore, a protein with a short half-life, especially, if it is needed at high concentrations, must be expeditiously synthesized demanding large energy costs. Conversely, a protein with an abnormally long half-life may be difficult to remove from the cell when its activity is no longer required.

Further Reading

Chen et al. 2008; Frauenfelder and McMahon 1998; Kloss et al. 2008; Muñoz 2007; Pain 2000; Plotkin and Onuchic 2002.

2.2.8 Ampholytes in Solution, the Acid–Base Equilibrium

We already noted that charges in biomolecules significantly influence the formation of their molecular and supramolecular structures. This aspect will be emphasized in

the next section when we discuss properties of interfaces and biological membranes. Furthermore, in Sects. 3.2.4 and 3.2.5 we will discuss the role of fixed charges in the cytoplasm to explain the occurrence of Donnan potentials. These aspects demand knowledge about the pH dependence of these sorts of charges.

An acid, per definition, is a proton donor, i.e., a compound that is dissociating hydrogen ions from particular groups. A base, conversely, is a proton acceptor and thus is able to associate them. According to this definition, water is a weak acid, dissociating with the following scheme:

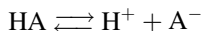


This equilibrium leads to the definition of the ionic product of water and, finally, to the pH-value:

$$\text{pH} = -\log c_{\text{H}} \quad (2.67)$$

Strictly of course, not the concentration (c_{H}), but rather the activity of the protons must be considered but in view of the extremely small concentration of these protons, this makes no difference.

In the same way, the dissociation of any other acid can be written:



The equilibrium constant of this reaction may be written as:

$$K = \frac{c_{\text{A}}c_{\text{H}}}{c_{\text{HA}}} \quad (2.68)$$

In the same way, as in the case of pH, it is possible to define a pK-value:

$$\text{pK} = -\log K = -\left(\log c_{\text{H}} + \log \frac{c_{\text{A}}}{c_{\text{HA}}}\right) \quad (2.69)$$

or using Eq. 2.67:

$$\text{pK} = \text{pH} - \log \frac{c_{\text{A}}}{c_{\text{HA}}} \quad (2.70)$$

This is the *Henderson–Hasselbalch equation* which makes it possible to understand the buffer properties of various substances and their titration curves.

Substances that contain acid, as well as basic groups, are called *Ampholytes*; the first example we will mention is the amino acids. The buffer capacity, as well as the dissociation property of these substances can be investigated by titration experiments. This furthermore allows us to calculate the dynamics of molecular charges of these molecules. In Fig. 2.30 the charge of glycine is demonstrated as a function of the pH

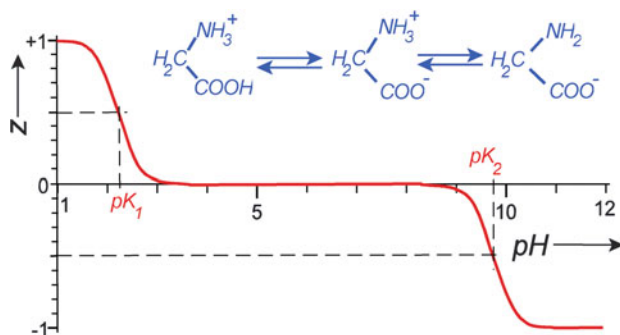


Fig. 2.30 The average number of charges (z) of glycine plotted against pH in the solution ($pK_1 = 2.35$, $pK_2 = 9.78$)

in the solution. For this, the mean number of elementary charges (z) per molecule is used. In the preceding sections we learned that dissociation and association are stochastic processes. Hence these amounts of z can be considered as an average in time or as a mean value of a large number of molecules.

Glycine is a cation at extremely low pH, and on the contrary an anion at extremely high pH. In the pH region in between, both groups of this molecule are dissociated. In this case the term zwitterion is used. This expression comes from German “Zwitter” or hermaphrodite, and means a molecule that is electrically neutral but carries positive as well as negative charges. In this case it has a considerable dipole moment and a maximal number of charges. This curve allows us to determine two pK-values which always correspond to the pH of the medium, where 50% of the charges are dissociated (in Fig. 2.30 corresponding to $z = +1/2$, and $z = -1/2$), as well as an isoelectric point which is positioned exactly in the middle between these two pH-values.

The pH-dependent charge of ampholytes of course also strongly influences the structure of the surrounding water. This is the reason why the effective volume of these molecules was often modified by the pH of the medium (see Sect. 2.2.3, Fig. 2.21).

Ampholytes are typical constituents of buffer solutions. Buffers are substances that are able to bind protons, or lose them, stabilizing in this way the pH of the solution. A buffer capacity can be defined, which corresponds to the slope of the titration curve. It is maximal near the pK-values.

Whereas the Henderson–Hasselbalch equation can easily be applied to binary ampholytes with two pK-values at considerable distance from each other, this becomes impossible in the case of polyampholytes, such as proteins with a large amount of different amino acids. This is illustrated in Fig. 2.31, showing the titration curve of hemoglobin. This curve can be considered as a sum of many sigmoidal curves with different properties. Near the isoelectric point, this curve can be fitted by the simple function:

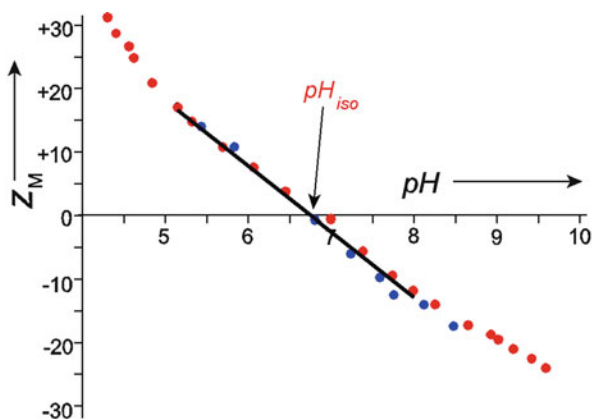


Fig. 2.31 The mean number of charges of horse hemoglobin as a function of the pH (blue points – oxidized, red points – deoxidized hemoglobin). The line between pH 5 and pH 8 corresponds to Eq. 2.71 with a buffer capacity of $z_{M0} = 10.2 \text{ val mol}^{-1}$, and an isoelectric point $\text{pH}_{\text{iso}} = 6.68$ (According to measurements of German and Wyman 1937)

$$z_M = -z_{M0}(\text{pH} - \text{pH}_{\text{iso}}) \quad (2.71)$$

(z_{M0} – buffer capacity, pH_{iso} – isoelectric point)

The titration curve of course indicates only the effective charges of molecules as a function of the pH in the solution. In the case of polymer molecules, the interaction of their monomers must be considered, depending on their mutual distance in the molecule. Furthermore, near the place of dissociation a local pH may exist, differing from that of the bulk solution. In this case, the term *intrinsic* pK is used, meaning the true pK of specific dissociable groups in a polymer, considering the real local proton activity. This, in fact, is possible to measure. Using nuclear magnetic spin resonance (NMR) for example, the dissociation of specific carboxyl groups can be studied, traced by the isotope ^{13}C . Investigations of polypeptides indicated that considerable differences may exist between the effective and the intrinsic pK of some groups even if the influences of neighboring amino acids are just slow. The local pH near functional groups in a protein can deviate significantly from that of the bulk solution. This is caused by fixed charges and therefore can be predicted by local electric potentials in the sense of the Boltzmann- or Nernst-equations (Eq. 3.112).

The different content of amino acids with acidic and basic residues in proteins is used to separate them by isoelectric focusing. For this a gel, composed of polyacrylamide, starch, or agarose is used with an immobilized pH gradient. In this gradient an electrophoresis is performed. If the protein is in a pH region below its isoelectric point, where it is positively charged it migrates towards the cathode. Arriving in this way in regions of higher pH value, the charge will decrease until the protein reaches its isoelectric point where the migration ceases. Isoelectric focusing can resolve proteins that differ in isoelectric points by as little as 0.01pH-numbers.

In the case of two-dimensional gel electrophoresis, proteins are first separated in one dimension by isoelectric focusing in a pH-gradient and subsequently in a second dimension by their electrophoretic mobility at constant pH, separating them according to their molecular weight or the length of their polypeptide chain.

Further Reading

On electrostatic properties of proteins: Matthew 1985; Pethig and Kell 1987.

2.3 Interfacial Phenomena and Membranes

After explaining various kinds of molecular interactions in the previous section, we will now explore the next level of biological organization, considering self-assembly and molecular dynamics of supramolecular structures, in particular of membranes.

Biological membranes do not only surround the living cells but lead also to separation of numerous intracellular compartments. It is a well-organized structure fulfilling a broad spectrum of physiological functions:

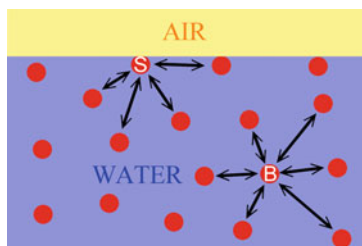
- As a surface, it forms a dynamic matrix for enzymatic reactions, receptor processes, and immunological recognition;
- As a barrier of diffusion it controls the ionic composition of the compartments by highly specific transporters and channels;
- As an electrically isolating leaflet it contains various passive and active electric devices, controlling membrane potential as well as membrane near electrodynamic conditions;
- As a mechanical structure it maintains the integrity of the compartments, and is a determinant of cell shape and cell movement as well as the displacement of organelles.

We will discuss the underlying biophysical mechanisms of membrane functions in several sections of this book in detail. Here, we consider its molecular structure and dynamics as well as the processes of self-organization. In some cases I cannot avoid mentioning topics that will only be explained in detail in later sections.

2.3.1 *Surface and Interfacial Tensions*

Each molecule of a homogeneous isotropic phase interacts energetically with all of its surrounding neighbors. If these forces are averaged for a sufficiently long period of time, all the moments of interaction compensate each other. Considering however a molecule that is not in the bulk phase, but at the phase boundary, or at the liquid–air surface, the situation is quite different (Fig. 2.32). In this case these forces of interaction with molecules in one phase differ to those in the other phase. A molecule of a liquid for example at the liquid–gas surface senses a stronger

Fig. 2.32 Interaction forces of molecules at the air–water surface (S), and in the bulk solution (B)



attraction to its neighbor molecules in the liquid than to molecules of the gas phase. Thus, the energy of the surface molecules is higher than that of the molecules in the bulk phase. This is the reason why water droplets in air spontaneously form a sphere as the shape of minimal relative surface and therefore minimal surface energy.

To increase the surface of a liquid against air or against a gas that is saturated by the vapor of this liquid, means to increase the number of molecules with a higher level of energy. The energy that is required to enlarge the surface of a liquid phase in this way by 1 m^2 is called the *specific surface energy*, which from the physical point of view, is the same as the *surface tension* γ ($\text{J} \cdot \text{m}^{-2} = \text{N} \cdot \text{m} \cdot \text{m}^{-2} = \text{N} \cdot \text{m}^{-1}$).

Water as a polar liquid with strong intermolecular interactions (see Sect. 2.2.2) has a high surface tension which decreases with increasing temperature. At 25°C it amounts to 0.0728 N m^{-1} . The surface tension of organic liquids, for example, of benzol is only 0.0282 N m^{-1} , and of ethanol only 0.0223 N m^{-1} at the same temperature.

For biological reactions the surface tension at the water–air interface is interesting from two aspects: On one hand it is important for organisms living directly at the surface of a lake or the sea. These are microorganisms and small algae, forming the so-called *neuston*, as well as various higher plants, and finally various insects using hydrophobic properties of their legs to run on the water surface. On the other hand surface tension plays an important role in the biomechanics of the lung, where a highly solvated surface is in contact with the air.

To explain this situation, the behavior of bubbles or vesicles formed by a liquid should be considered in detail: In the same way as the surface tension tends to decrease the surface of a droplet, leading to a spherical shape, it tends to decrease the volume of a gas bubble, generating an internal pressure (Fig. 2.33a). The pressure difference $\Delta p = p_i - p_e$ can be calculated by the Laplace equation:

$$\Delta p = \frac{2\gamma}{r} \quad (2.72)$$

This means that the internal pressure of a vesicle is proportional to the surface tension of the liquid (γ) which forms it, and inversely proportional to its radius (r).

In most cases it is not the *surface tension* liquid–air, but rather the *interfacial tensions* between two liquid phases that are important, or, in other words, the

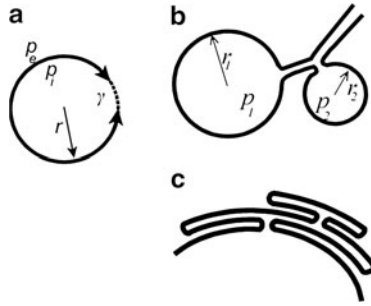


Fig. 2.33 (a) A vesicle with a surface tension γ tends to decrease the internal volume and generates an internal pressure $\Delta p = p_i - p_e$; (b) two vesicles of different size, built at the same interface generate different internal pressures which are inversely proportional to their radius. In the case of interconnection, the smaller vesicle should collapse in favor of the larger one. (c) The change in the surface of the surfactant layer by folding

specific energy of interfacial molecules. In the ideal case the interfacial tension of two liquids is equal to the difference between their surface tensions.

Earlier, these considerations sometimes uncritically were transformed to understand various cell physiological properties such as phagocytosis, exocytosis, cell division, amoeboid movement, cell shape, etc. Ludwig Rhumbler for example already in 1898 demonstrated that a drop of chloroform in water could “phagocytose” a thin, shellac-covered glass fiber in a manner similar to that of an amoeba, ingesting a filamentous alga. D’Arcy Thompson, in his classic book “On Growth and Form,” first published in 1917, discussed in detail the problem of cell shape in relation to water droplets or even cellular ensembles in relation to the structure of foam.

Those comparisons, however, as visually impressive as they may be, are misleading in explaining cellular processes. In the case of a liquid droplet, surface molecules can always be drawn back into the bulk, or extruded again in the surface layer very quickly. This process is reversible; the mechanically deformed drop will re-assume its spherical shape as soon as the external mechanical disturbance is discontinued. In contrast, a living cell is surrounded by a membrane. An increase of the membrane area by inclination of additional molecules from the cytoplasm is a process which is much slower.

Conversely, the problem of surface tension is crucial, to understand the function of the alveoli in the mammalian lung. These are spherical outcroppings of the respiratory bronchioles covered by an epithelial layer, the diameter of them periodically varying between 0.05 and 0.1 mm. They are responsible for gas exchange with the blood.

The challenge of the system is to ensure their periodic change of size, and the stability of coexistence of vesicles of different diameter, connected to each other. As indicated in Fig. 2.33b, according to Eq. 2.72, the pressure in the smaller vesicles would be higher than in the larger vesicles. This would lead to a collapse of smaller vesicles in favor of the larger ones if they were connected to each other.

To avoid this, the surface tension of the water–air interface in the alveoli of the lung is strongly reduced by a surface-active material known as *pulmonary surfactant*, permanently secreted by Type II cells of the alveolar epithelium. This is a complex mixture of lipids and proteins that forms a monolayer at the alveolar liquid–water interface. It decreases the surface tension water–air from approximately 0.007 N m^{-1} to nearly zero. The periodic change of the alveolar volume and their surface during inhalation requires the property of this layer to be expanded quickly and reversibly. For this, a reservoir was formed, reversibly folding and unfolding these layers (Fig. 2.33c). Despite intensive research in this field, this process is still unclear in many aspects.

Further Reading

Historical papers: Rhumbler 1898; Thompson 1966. For the surfactant of the lung: Lalchev et al. 2008; Veldhuizen and Haagsman 2000.

2.3.2 Orientation of Molecules at Phase Boundaries; Self-Assembly of Membranes

The spontaneous orientations of molecules at phase boundaries are of particular interest for questions of self-organization and stability of biological membranes. Let us first consider the behavior of molecules at the boundary between two phases, one being hydrophilic (water), the other hydrophobic (for example oil). A protein tends to orient its polar groups as much as possible toward the aqueous phase, and its nonpolar, i.e., hydrophobic groups toward the oil (Fig. 2.35a). This leads to an orientation of the molecule at the phase boundary. If the polar groups of the molecule are distributed homogeneously over the whole molecule, then it could become unfolded by orientation in such a boundary. The protein, now being a fibril will spread over this interface. Some of these proteins form amphipathic helices or even stretches of α -helix, one side of which is polar (charged) and the other hydrophobic. In this case it orients parallel to the boundary layer.

The most important constituents of biological membranes are phospholipids. This is a group of compounds that consist of two fatty acids and a phosphate group, all bound to a polyhydric alcohol such as glycerol. The phosphate group is usually esterified with a nitrogen-containing alcohol. Structure and formula of a simple phospholipid is depicted schematically in Fig. 2.34. Two long hydrophobic fatty acid chains and a single phosphate group esterified with an ethanolamine residue are linked to a central glycerol molecule. The phosphate and the amino group of the ethanolamine residue represent the hydrophilic part of the molecule.

The large variety of biological lipids is realized by the diversity of the fatty acids as well as by their different head groups. In the case of phosphatidylcholine, the head group is polar but does not carry net charges at physiological conditions. The negative charge of the phosphate is compensated by a positive charge of the amino group. Phosphatidylserine, in contrast, contains an additional negative charge,

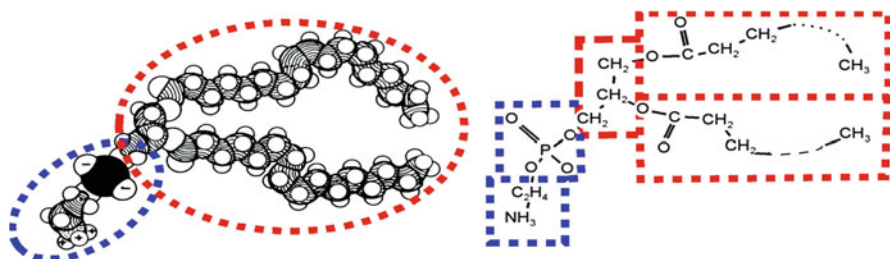


Fig. 2.34 Model and formula of a simple phospholipid (phosphatidylethanol amine with olein- and palmitin acid-rests). The hydrophilic part of the molecule is surrounded with *blue*, the hydrophobic with *red lines* (after Haggis 1965 redrawn)

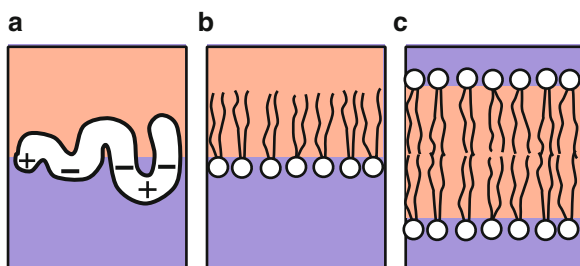


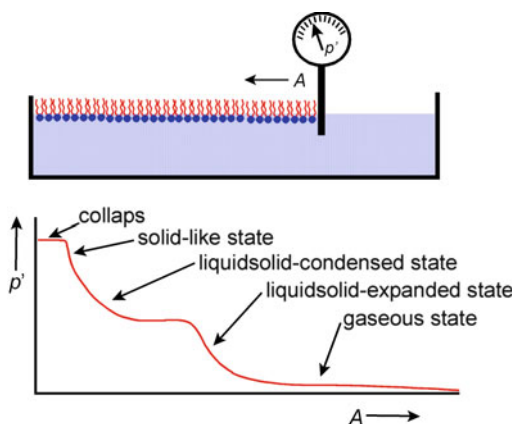
Fig. 2.35 Arrangement of molecules in phase boundaries (*red area* – hydrophobic phase, *blue area* – hydrophilic phase): (a) orientation of a protein with hydrophobic and hydrophilic regions; (b) orientation of a monolayer of phospholipids (this corresponds also to the lipid layer at a water–air surface); (c) a lipid bilayer in a homogeneous aqueous phase. In the region of the fatty-acid chains, oriented against each other, a hydrophobic microphase is formed

namely that of a carboxyl group. This is the reason why it significantly influences the surface charges of biological membranes.

Phospholipids strongly orient themselves at interfaces. In this case the hydrophobic chains of fatty acids are directed parallel to each other and perpendicular to the interface (Fig. 2.35b). In this way, monomolecular layers, so called *monolayers*, are formed. The same occurs at the surface of an aqueous solution, i.e., at the water–air interface. Phospholipids that are oriented in such a way can be considered as a kind of two-dimensional phase.

Depending on the surface concentration of these molecules, it resembles gaseous, liquid, or solid states. Like in three-dimensional phases, phase diagrams as a function of pressure versus concentration, for example intermolecular distances, can be obtained. For this, a so-called *Langmuir trough* is used (Fig. 2.36). It consists of a temperature-controlled flat rectangular trough with a mobile barrier, as a part of a film balance. Phase diagrams can be obtained, measuring the resulting force resp. the lateral pressure (p' as force per length of the barrier) by shifting this barrier, i.e., increasing the lateral concentration of the molecules in the layer.

Fig. 2.36 Langmuir trough and a corresponding phase diagram of a layer of surfactant molecules



Whereas at low lateral concentration the molecules like in a gaseous state are moving more or less freely, they are always touching one another being concentrated to the liquid state. Depending on the degree of hydration of their head groups, an expanded liquid state can be differentiated from a condensed one, which finally is transformed to a solid-like state, without hydration of the head groups. A collapse finally occurs if several lipid layers slip one upon the others, forming multilayers. The shape of these phase diagrams reflects the properties of the flexible hydrocarbon chains as well as that of the polar head groups of the molecules. These kinds of investigations allow important inferences to be drawn on the properties of biological membranes.

In the case of homogeneous aqueous phases, phospholipids reach a state of minimal surface tension if they orient their hydrophobic portions toward one another (Fig. 2.37a). There is a large diversity in lipid aggregates in aqueous solutions. Using special methods, for example sonication of lipid suspensions with ultrasound, it is possible to produce structures with lipid double layers (*bilayers*) and vesicles, containing inside aqueous solutions (*liposomes*). There are two types of liposomes: unilamellar liposomes covered only by a single bilayer (Fig. 2.37b), and multilamellar liposomes (Fig. 2.37c) containing numerous lipid bilayers, separated from each other by a thin water layer.

In this context it must be mentioned that surface tension at least partly, is the phenomenological expression of processes which on the molecular level are the result of the dynamics of hydration, i.e., of water structure and hydrophobic bonds. All the lipid structures mentioned here can be destroyed by extremely high pressure in the sense of the Le Chatelier principle, as discussed in Sect. 2.2.3.

It is also possible to prepare in aqueous phases planar lipid membranes (Fig. 2.35c). If such a membrane is formed, the fatty acid chains of the lipids have built a hydrophobic microphase. During the last decades ingenious methods have been developed to produce bimolecular membranes with high structural and chemical specificity. These are the so-called BLMs, which is an abbreviation for:

Fig. 2.37 Self-assembly of lipids in aqueous solutions: (a) micelle; (b) unilamellar liposome; (c) sector of a multilamellar liposome. The blue areas indicate local microphases of water

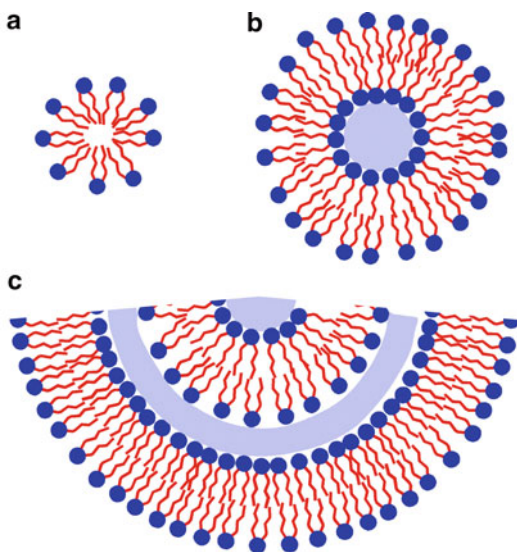
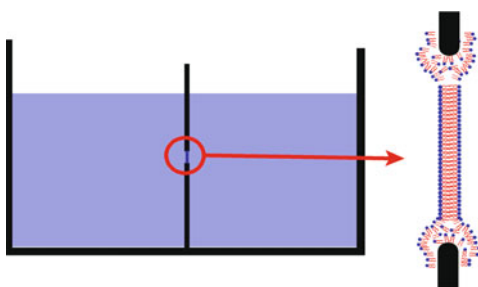


Fig. 2.38 BLM in a hole of a barrier between two aqueous phases



Bimolecular Lipid Membrane. By various techniques it is possible to expand such membranes in a small hole of a barrier, dividing two aqueous phases (Fig. 2.38). In this case various electrical and electrochemical measurements are possible. Usually, these membranes are at first multilamellar, showing interference of visible light. Later the excessive lipids concentrate on the edges of the hole, and the membranes become bilayers with a thickness of less than 10 nm. In this case no interference of visible light is possible; they become optically nondetectable. For this reason the abbreviation BLM is also interpreted as “Black Lipid Membrane.” The real existence of a barrier in the hole in this case can be proved only by its electrical resistance.

These artificial membranes are very important tools to investigate properties of biological membranes and their constituents. It is possible to introduce specific molecules into these membranes, such as for example transport and signal proteins, and investigate their specific characteristics under various conditions. Their lipid

composition can be varied as well as the ionic composition of the solution on both sides of the membrane. Using specific techniques it is even possible to mimic the lipid asymmetry of biological membranes.

Further Reading

Butt et al. 2006; Hianik and Passechnik 1995.

2.3.3 The Molecular Structure of Biological Membranes

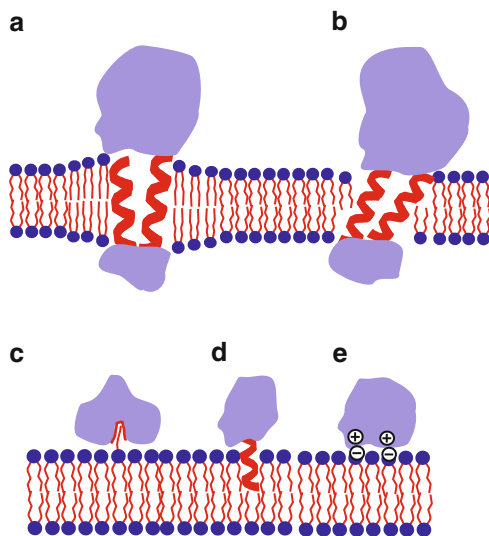
Measurements of electric conductivity of cell suspensions already indicated a long time ago that the cell must be surrounded by an electrically isolating layer with a specific capacity of about 10 mF m^{-2} (for more detail, see Sects. 2.3.6 and 3.5.3). It corresponds to a double layer of lipids with a thickness of ca. 8 nm. This led to a number of ideas and models of membrane composition, and its molecular structure. In 1972 Singer and Nicolson proposed a *fluid-mosaic model* of the cell membrane which in fact reflects a number of experimental properties. It considers a lipid bilayer as a two-dimensional solvent with integral membrane proteins, embedded in it, that are free to move. This model has been modified by recent experimental results. In fact the real concentration of proteins in the membrane is larger as suggested by Singer and Nicolson. Furthermore, many membrane compounds are not freely mobile, but rather connected to the cytoskeleton, and the plasma membrane of eukaryotic cells contains a variety of inhomogeneities such as protein and lipid domains.

The lipid composition of the inner and outer membrane leaflet of biological membranes is highly specific. This concerns the particularity of their fatty acids, as well as their polar head groups. There is also an asymmetric distribution of lipids between the inner and the outer layer, maintained by specific transport proteins, so-called *flippases*.

The fluidity of the membrane, i.e., the lateral movement of its constituents is strongly determined by the length of the aliphatic fatty acid chains, as well as by their degree of saturation (see also Sect. 3.6.3). The property of the polar head groups determines their mutual distance in the membrane surface, i.e., their density of packing.

The proteins are organized in the membrane according to their hydrophobic (apolar), and hydrophilic (polar) regions (Fig. 2.39). Most of the functional proteins, such as transporters, penetrate the hydrophobic core of the double layer of the membrane by helical parts. The number of membrane-spanning parts of a protein may range from one to more than 12. This is dictated by the dimension of the hydrophobic part of the proteins and has to be matched to the hydrophobic thickness of the lipid bilayer. In order to compensate for the mismatch, the lipid molecules closest to the protein stretch out or compress in order to cover the hydrophobic core of the protein (Fig. 2.39a). This leads to a perturbed region around the protein. Another possibility is an inclination of the helices in the

Fig. 2.39 Various kinds of protein interaction with the lipid double layer: (a, b) integral membrane proteins that span the bilayer; (c) protein, anchored via a lipid extended conformation; (d) amphiphilic protein partially penetrating the bilayer; (e) electrostatically bound protein



membrane (Fig. 2.39b). Nonpenetrating proteins can be fixed to the membrane in various ways (see Fig. 2.39c–e).

Many proteins are exposed to the external surface of the cell membrane as *glycoproteins*. Typically, they are located with their C-terminal end in the cytoplasm. Their N-terminal end, which is modified by various carbohydrates, extends several nanometers into the outer environment of the cell. These carbohydrates in some cases represent the receptor-sides of the proteins, and are the location of immunological reactions. Preferentially, at the end of the protein chains, monomers of the *N-acetylneuraminic acid* (also called: *sialic acid*) are located which carry a dissociable carboxyl group. In this way, these are the most important carriers of fixed negative surface charges of the cell (see Fig. 2.48). Depending on the type of cell, there are between 1 and 10 groups of sialic acid per 1 nm² membrane area. Hence, the glycoproteins form a loose external layer of the cell, the so-called *glycocalyx* or *surface coat*. The charges, located at the ends of these filaments, modify the thickness of this structure electrostatically (for detail, see Sect. 2.3.6).

The lateral distribution of membrane constituents according to the Singer–Nicolson model could be random only if all pairwise interaction energies are within the range of the thermal energy kT . Considering the diversity of lipid and protein species and the variety of intermolecular interactions such as hydrogen bonds, electrostatics, and hydrophobic effects, this condition is hardly fulfilled. In contrast, it should have been expected that regions of biased composition would exist.

Recent investigations indicate that membrane proteins may be organized in large functional complexes. In common with proteins, lipids also tend to group together by specific lipid–lipid and lipid–protein interactions. Some lipids form specific complexes with proteins. As indicated in Fig. 2.39a, at the rim of proteins and protein complexes, different thicknesses of the lipid bilayer may exist.

Lateral movement of proteins has been recently investigated by single-particle tracking experiments. For this gold-labeled molecules were moved in the membrane surface until a barrier was encountered. Some receptor proteins in this way show a lateral compartmentalization in 300–600-nm diameter domains. These results led to the *membrane-skeleton fence* model, including the possibility of successive movements (*hops*) to adjacent compartments.

From the truncation experiments it was estimated that the barriers to this lateral mobility were located 2–3 nm below the cytoplasmic leaflet. Obviously it consists of a network of cytoplasmatic proteins as part of the cytoskeleton with connections to the inner side of the membrane. It is best investigated in human erythrocytes, where it consists mostly of spectrin. This is a heterodimeric protein (molecular mass: 260 and 220 kDa) that forms rods of approximately 100 nm length, aggregating head-to-head, to form tetramers of double length. These spectrin rods are connected to each other by the globular actin molecules (molecular mass: 100 kDa). This leads to an *anchored-protein picket* model assuming that close to the cytoplasmic leaflet the actin membrane skeleton meshwork anchors various transmembrane proteins.

Another important class of membrane inhomogeneities is *caveolae*. These are flask-shaped membrane invaginations of about 60 nm diameter containing mainly *caveolin*, a protein that binds cholesterol. Caveolae have been implicated in processes like cholesterol transport and endocytosis.

Studies on lipid marker molecules indicated lipid microdomains in the exoplasmic leaflet, so-called *lipid rafts*. These are areas ranging from several tens of nanometers in biological membranes to almost a micrometer in model membranes that are enriched in cholesterol and lipids with saturated acyl chains, such as sphingolipids. A tight packing of these constituents results in a liquid-ordered phase, which separates the lipid rafts from the surrounding. The biological functions of lipid rafts range from membrane trafficking and sorting to a dynamic role in signal transduction.

The membrane therefore, can be considered as a two-dimensional compartmentalized structure, which is more mosaic-like than fluid.

Further Reading

Engelman 2005; Lommerse et al. 2004.

2.3.4 Mechanical Properties of Biological Membranes

The mechanical properties of biological membranes are very important to understand a number of physiological cell functions such as cell movement, cell division, volume alterations, vesiculation, membrane fusion, etc. In the same way, the mechanical properties of tissues, as well as the streaming properties of the blood are controlled, which we will discuss in Sect. 3.6.

It must be emphasized that parameters, for example viscosity, elasticity, strain, etc., in fact, are macrophysical quantities, which are defined for homogeneous phases and materials (see Sect. 3.6). Hence, they are not proper parameters for

biological membranes at the submicroscopic scale, and also not for highly organized supramolecular structures. Nevertheless, in some cases it may be convenient to use these parameters, but then they must be considered *effective parameters*, which we explained in the introduction to Sect. 2, and which we have already used in many other cases.

The effective membrane viscosity, for example, is a parameter that can be measured by monitoring the rotation or translocation of a marker molecule. This is possible through the use of special fluorescence methods or electron spin resonance techniques (ESR). Considering the membrane as a homogeneous phase, and the marker molecule as a macroscopic body with a particular shape, one can formally calculate the viscosity, applying the usual phenomenological equations for movement. This measurement is reproducible, and the parameter is exactly defined by the experimental method, and by the applied theoretical approach. The problem which arises through the use of equations that are not adequate for the movement of a molecule in a heterogeneous phase is not relevant here, because it is included in the definition. It is an “effective” parameter, since it is defined by observing a particular effect. Conversely, this means that measuring these effective viscosities by different methods, one gets different parameters, but not a general viscosity according to the original physical definition.

This is the general difference between molecular and macroscopic types of measurement. In the case of homogeneous liquids, the resulting viscosity is independent of the method used to obtain it. In fact: a “viscosity” in the sense of macroscopic materials does not exist in supramolecular structures such as biological membranes. The same applies to the other parameters mentioned above.

Applying different markers, the parameters measured in this way depend on the regions where the marker was located, for example, at different locations in the membrane depth, i.e., at different positions in the fatty acid of a lipid, or even at different lateral positions of the membrane. These differences indicate the inhomogeneity perpendicular to the plane of the membrane, as well as the anisotropy of the mechanical properties of the biological membrane in general.

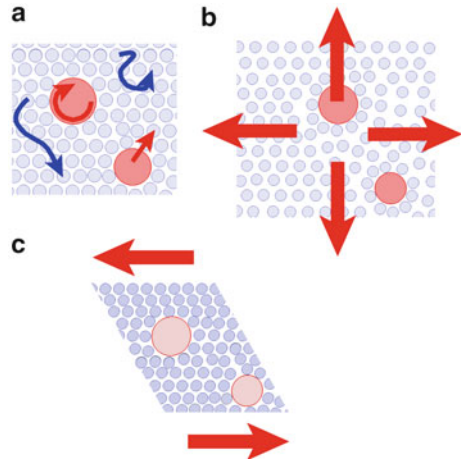
For some biomechanical questions, such as for example deformation of erythrocytes in small capillaries (Sect. 3.7.2), osmotic swelling of cells, etc., some general properties of the membrane are of interest, independently of its inhomogeneity.

In general, the cell membrane is easily deformable by shearing forces without increasing the area (Fig. 2.40c). In contrast to this it is nearly impossible to extend the area of the membrane (Fig. 2.40b). The stress (σ'), which must be applied, to extend the area A of a membrane by the amount ΔA can be calculated by

$$\sigma' = Y' \frac{\Delta A}{A} \quad (2.73)$$

Y' is the specific modulus of elasticity, a coefficient which is related to the thickness of the membrane. The symbol Y stands for *Young's modulus*, another name for the same parameter (see Sect. 3.6.3). For the membrane of erythrocytes a

Fig. 2.40 Possible kinds of passive molecular displacement and mechanical deformations of biological membranes: (a) translation and rotation of membrane components; (b) planar extension of the membrane area; (c) shear deformation



value of $Y' = 0.45 \text{ Nm}^{-1}$ was estimated, for lymphocytes 0.64 Nm^{-1} . To find a relationship between membrane properties and properties of macroscopic materials, the parameter Y' must be divided by the thickness of the membrane (d), which is approximately $8 \cdot 10^{-9} \text{ m}$. In this case, for erythrocytes it results in:

$$Y = Y'/d = 5.6 \cdot 10^7 \text{ Nm}^{-2} = 56 \text{ MPa}$$

The membrane will rupture if it expands more than 1–2% of its area.

In Sect. 3.6.3 we will differentiate between two kinds of elastic materials, resembling steel, in contrast to rubber elasticity. Steel elasticity occurs in crystalline materials and is caused by modifications of the crystal structure. It is characterized by a large Young's modulus (for steel about $2 \cdot 10^5 \text{ MPa}$) and a low expandability. In contrast, rubber elasticity is a typical behavior of macromolecular structures and based on the stretching of molecules, as illustrated in Fig. 2.11. It indicates a Young's modulus of only around several MPa.

In fact, the mechanical behavior of biological membranes already tends toward the values of steel-elastic materials. This is understandable considering that any increase of the membrane area leads to an increase of the distance between the lipid head groups, analogous to the ion atoms in steel. Properties of rubber elasticity are given only by the intrinsic proteins.

The membrane therefore resembles a material with a high degree of flexibility in plan, but with minimal ability for area extension (Fig. 2.40). Hence, it has properties very unlike a rubber sheet, which can easily be expanded on account of its thickness. Living cells, therefore, can swell only by spherulation, by unfolding of the membrane, or by inclusion of new membrane material. Each swelling of the cell, which extends the surface of a sphere, or increases the capacity of unfolding, immediately leads to cytolysis.

As we will see later, using the elastic modulus, it is also possible to calculate forces that oppose the bending of a material (Sect. 3.6.3). Applying this to the biological membrane, again, the inhomogeneity of the membrane must be considered. According to the heterogeneity of the lipid composition and the differences in the distribution of the proteins, the bending resistance of the membrane differs from place to place. Recently, atomic force microscopy (AFM) has been used for such investigations. AFM consists of a microscale cantilever with a sharp tip at its end. Using this, periodic indents of the membrane at particular locations can be produced. Scanning the membrane in pixels approx. 70 nm apart, and by using a 20-nm tip radius cantilever deflected by 50-nm a relative pattern of regions with different stiffness can be obtained.

In addition to these passive mechanical properties, biological membranes per se are bent. As illustrated in Fig. 2.41, this active bending is realized in different ways. One possibility is the bilayer couple mechanism. In this case bending occurs by inclusion of additional molecules in one of the membrane leaflets. Another possibility is the introduction of conically shaped molecules. A factor (f) characterizes this property. In the case of lipids, it relates the area of the head group (A_H) to the mean area (A_F) which is required for the region of fatty acids (Fig. 2.42). A_F can be

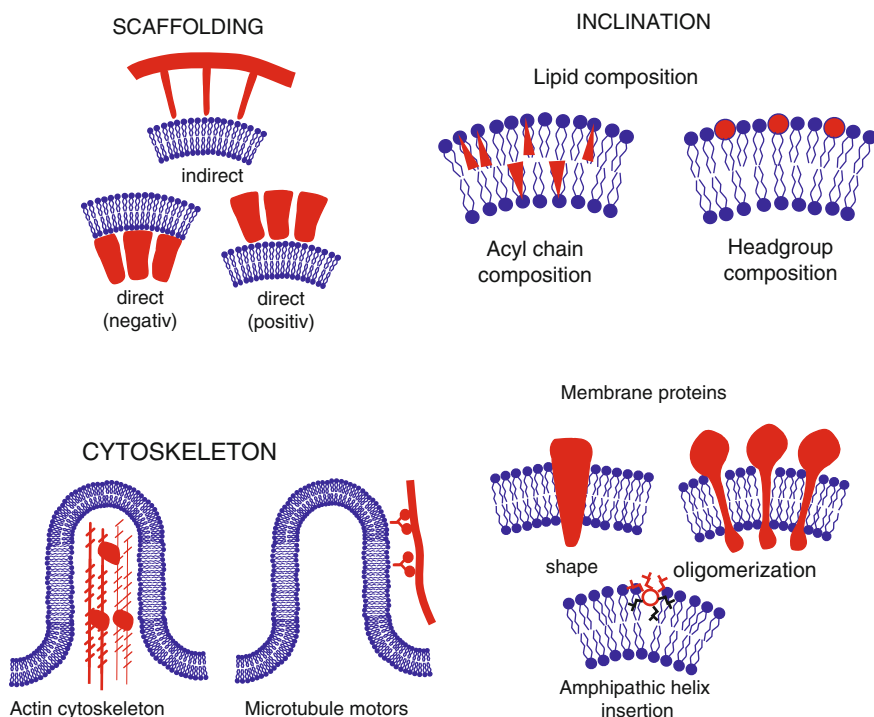
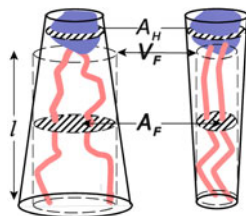


Fig. 2.41 Possible kinds of active membrane bending (*blue* – phospholipids, *red* – proteins) (Redrawn after McMahon and Gallop 2005)

Fig. 2.42 The effective shape of a phospholipid (explanations in text)



calculated by the mean length of the fatty acids (l) and the effective volume of the region of fatty acids (V_F) using: $A_F = V_F/l$. The shape factor f , therefore, is defined as

$$f = \frac{V_F}{lA_H} \quad (2.74)$$

The effective volume of the fatty acid chains (V_F) strongly depends on their phase conditions, i.e., on their degree of order. Dynamic variations are possible as well as phase transitions through temperature modification. Other parameters of the molecule directly depend on their composition. So, for example, lysophospholipids, containing only one single fatty acid chain, are to a large degree conical with a dominant A_H area; in this case, $f < 1/3$. In contrast, phosphatidylethanolamine containing usually a large amount of unsaturated fatty acids indicates $f > 1$, with a dominant A_F . Mostly lipids are more or less cylindrical, with $1/2 < f < 1$. The area of the head group is in fact a parameter which includes the mean distance to the neighboring molecule, which can also be influenced by electrostatic interactions.

Figure 2.41 shows further possible types of active membrane bending. This concerns not only inclusions of other molecules in the leaflet, but also bending by external or internal scaffolding proteins. Also very important is the assembly and disassembly of the cytoskeleton of the membrane. In many areas of high membrane curvature, such as filopodia, pseudopodia, phagocytic cups, etc., actin filaments are involved in the generation and remodeling. In some cases kinesin motors attached to Golgi membranes are involved in processes of fenestration or generation of some transport processes.

Further Reading

Devaux and Herrmann 2011; McLaughlin and Murray 2005; McMahon and Gallop 2005; Roduit et al. 2008; Schmid-Schönbein et al. 1986.

2.3.5 Electrical Double Layers and Electrokinetic Phenomena

An interface carrying fixed charges induces in its vicinity an electric field, and accordingly, modifies the local ionic conditions, analogous to a point charge, as described in Sect. 2.2.4. In the simplest case, a double layer occurs with the fixed

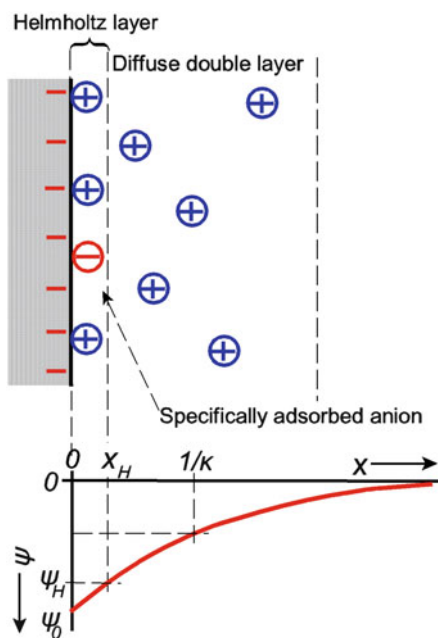
charges on one side, and the electrostatically attracted mobile counterions on the opposite side. This concept was originally put forward by Herrmann von Helmholtz. It only applies for extremely high concentrations of mobile ions in solution. In general, the thermal movement repels a great part of the ions from their positions near the fixed charges. This leads to the formation of a *diffuse double layer*, the electrical potential of which is declining exponentially with the distance from the surface.

Figure 2.43 illustrates the generalized model of Stern, which includes the Helmholtz double layer as well as its diffuse part. To avoid overloading of the picture, only the fixed charges (in this case proposed as negative), and the compensating mobile counterions are depicted. Of course, there are also mobile anions present, and there are a great number of both cations and anions that should be included to give a true picture. This picture indicates an enlarged concentration of cations near the surface for the case of negative surface charges, and the electroneutrality of the total system, including the bulk solution.

Despite the thermal noise, some of the mobile ions directly oppose the fixed charges, forming the so-called *Helmholtz layer*. Between the surface and the end of the Helmholtz layer, the potential is proposed to decline linearly from ψ_0 to ψ_H . Further, the potential declines according to the model of the diffuse double layer which was calculated by Gouy and Chapman.

The theory of the diffuse double layer again is based on the Poisson–Boltzmann equation, which we have already explained in context with the Debye–Hückel theory of ion clouds (Sect. 2.2.4.). This equation makes it possible to calculate the distribution of mobile charges in a given electric field:

Fig. 2.43 Schematic illustration of an electrical double layer. *Above:* fixed charges and mobile counterions near a surface; *beneath:* the function of the electrical potential, according to the model of Stern. The charge of the surface in this particular case is enforced additionally by a chemically adsorbed negatively charged ion. ψ_0 surface potential, ψ_H potential at the end of the Helmholtz layer, x_H thickness of the Helmholtz layer, $1/\kappa$ Debye–Hückel length as a measure of the effective thickness of the diffuse double layer



$$\nabla^2\psi = -\frac{F}{\varepsilon_0\varepsilon} \sum_{i=1}^n c_{io}z_i e^{-\frac{z_ie\psi}{kT}} \quad (2.75)$$

In this case the amount of fixed charges is compensated by the excess of charges in the whole double layer.

Applying this equation for one-one-valent electrolytes, like KCl or NaCl, it reduces to:

$$\nabla^2\psi = -\frac{Fc_0}{\varepsilon_0\varepsilon} \left(e^{-\frac{e\psi}{kT}} - e^{-\frac{e\psi}{kT}} \right) = \frac{2Fc_0}{\varepsilon_0\varepsilon} \sinh \frac{e\psi}{kT} \quad (2.76)$$

In this case, a rearrangement was taken, using the function: $\sinh x = (e^x - e^{-x})/2$.

To solve the Poisson–Boltzmann equation, i.e., to calculate the function $\psi(x)$, one needs to define two initial conditions. For our case this means:

$$\begin{aligned} \psi(\infty) &= 0 \\ \psi(0) &= \psi_H \end{aligned}$$

whereas the end of the Helmholtz layer (ψ_H) is taken as the starting point of the function $[\psi(0)]$. The algebraic integration of this equation again, is possible only after linearization. For $x > x_H$ one gets the following function $\psi(x)$:

$$\psi(x) = \psi_H e^{-\kappa x} \quad (2.77)$$

In this equation, κ again is the Debye–Hückel parameter, which was already introduced in Sect. 2.2.4 (Eq. 2.55). The Debye–Hückel length ($1/\kappa$) is the distance, where the potential ψ_H is dropped by the factor e :

$$\frac{\psi(1/\kappa)}{\psi_H} = \frac{1}{e} = 0.368 \quad (2.78)$$

The amount of this parameter, as a function of the ionic strength (I) is given in Eqs. 2.55 and 2.57.

What does “linearization” of Eq. 2.76 mean, and what are the limitations of this simplification? In fact, the basis for such approaches is the expansion of these functions in series, using the following equations:

$$e^x = 1 + x + \frac{x^2}{2} + \frac{x^3}{3} + \frac{x^4}{4} + \cdots \quad (2.79)$$

and:

$$\sinh x = x + \frac{x^3}{3} + \frac{x^5}{5} + \frac{x^7}{7} + \cdots \quad (2.80)$$

For the case:

$$\frac{ze\psi}{kT} \ll 1 \quad (2.81)$$

one can apply these equations, using only the first terms of the series:

$$e^{-\frac{ze\psi}{kT}} \approx 1 - \frac{ze\psi}{kT}, \quad \text{and :} \quad \sinh \frac{e\psi}{kT} \approx \frac{e\psi}{kT}. \quad (2.82)$$

Let us investigate the scope where this simplification is possible. For $T = 298 \text{ K}$ one can calculate:

$$z\psi \frac{e}{kT} = z\psi \frac{1.60218 \cdot 10^{-19}}{1.3807 \cdot 10^{-23} \cdot 298} = 38.94z\psi \quad (2.83)$$

Using, for example, a potential $\psi = 0.01 \text{ V}$ and a charge number $z = 1$, one gets: $e^{-0.3894} = 0.6774$, and according to Eq. 2.79: $1 - 0.3894 = 0.6106$. Correspondingly: $\sinh 0.3894 = 0.3993$. Both calculations, really, indicate some differences, which will become larger, if the electrical potential exceeds the value of 10 mV . This is important also for the application of this linearization in the previous Sect. 2.2.4., as well as in the following considerations.

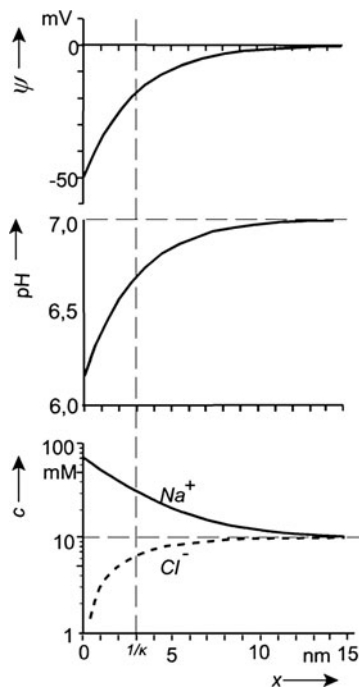
Beside the calculation of the diffusion layer it is necessary to know the relation between the potentials ψ_o and ψ_H . In general, the difference between these two parameters will become smaller, as the ionic strength of the solution becomes lower. In solutions of extremely low ionic strength, this difference can be neglected.

In fact, besides the screening of fixed charges by mobile ions, other processes must be considered, such as dipole orientations, interactions with water molecules or other processes of adsorption, based on van der Waals interactions. In this way, it is even possible that against electrostatic repulsion, co-ions were absorbed at the interface, increasing its charge density. In this case the Helmholtz potential (ψ_H) can even become larger than the surface potential (ψ_o).

We will discuss the relevance of electrical double layers for functions of biological membranes in detail in Sect. 2.3.6. The most important consequence are local ionic concentrations, including, local pH values (Fig. 2.44). This can be calculated according to the Nernst equilibrium (Eq. 3.112, Sect. 3.2.2). The relevance of these local conditions is understandable, considering the function of membrane-bound enzymes.

A number of electromechanical interactions are caused by these electrical double layers near charged surfaces which are summarized by the term: *electrokinetic phenomena*. On one hand, an externally applied electric field may induce mechanical effects, like movement or streaming; on the other hand, a mechanical

Fig. 2.44 Electrical potential (ψ), pH, as well as Na^+ - and Cl^- -concentrations (c) as a function of distance (x), near a negatively charged surface in a diffuse double layer. The curves consider the following parameters: $\psi_0 = -50$ mV, $\text{pH}_\infty = 7.0$, $c_\infty = 10$ mM, $T = 298$ K (the index ∞ means: at $x \rightarrow \infty$). In this case the Debye–Hückel-length ($1/\kappa$) amounts to 3 nm (dissociation constants of salt and water are not taken into account)



movement may lead to the occurrence of electrical potential differences. In Fig. 2.45 these phenomena are depicted schematically.

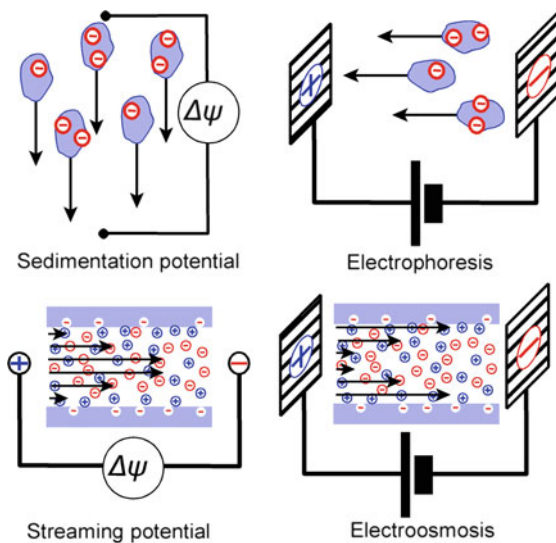
Electrophoresis means the movement of charged molecules, particles, or cells in an externally applied electric field. The displacement of particles and cells can be measured by microscopic observation. Their mobility (b) is defined as the relation between their velocity (v) and the electric field strength (E):

$$b = \frac{v}{E} \quad (2.84)$$

In an electric field the particles themselves move in the opposite direction to the ions of their double layers, caused by their opposite polarization. The movement of ions furthermore includes a movement of bound water. This leads to an increased frictional force, diminishing the resulting electrophoretic mobility.

The classical theory of cell electrophoresis is based on the model which was already proposed by Smoluchowski, considering the movement of large spherical particles. “Large” in this sense means that the diameter of the particle, and the corresponding radius of bending, must be large in relation to the Debye–Hückel parameter ($1/\kappa$). Smoluchowski postulated that a shearing plane exists, caused by the electrokinetically induced streaming. At this plane, the electroneutrality of the system is disturbed. The electrical potential at this distance from the surface is

Fig. 2.45 Schematic representation of electrokinetic phenomena of particles (*above*), and surfaces of tubes or capillaries (*beneath*); *left*: the occurrence of electrical potentials as a result of mechanical movement (*arrows*); *right*: the induction of movement (*above*) or streaming (*beneath*) as a result of an applied electric field (*arrows*)



called *electrokinetic potential*, or ζ -*potential*. It can be calculated by the measured electrophoretic mobility (b) using the following equation:

$$\zeta = \frac{\eta V}{\varepsilon_0 \varepsilon \mathbf{E}} = \frac{\eta b}{\varepsilon_0 \varepsilon} \quad (2.85)$$

where η is the viscosity of the medium. Surprisingly, the radius (r) of the particle is not included in this equation. Sometimes, this equation contains an additional factor, considering stronger deviations from the shape of a sphere.

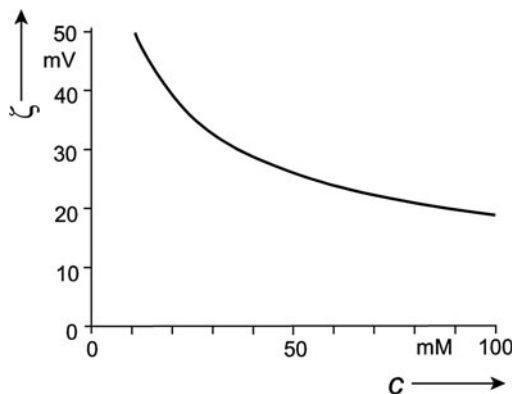
Of course, the above-introduced ζ -potential is not identical with the surface potential (ψ_0). With some approximation, however, it can be considered as equal to the Helmholtz potential (ψ_H) (see Fig. 2.43). In the same way, as discussed above, the ζ -potential increases, by constant surface charge density (σ_0), if the ionic strength in the solution is decreasing (Fig. 2.46).

It is possible to calculate the surface potential (ψ_0) and the surface charge density (σ_0) from the ζ -potential only with certain approximations. For this, the equation is used, derived from the Gouy–Chapman theory of electrical double layers:

$$\sigma_0 = \sqrt{8\varepsilon_0 \varepsilon RT} \sqrt{I} \sinh \frac{F\zeta}{2RT} \quad (2.86)$$

In Fig. 2.46 the dependence of the ζ -potential is depicted as a function of the ionic strength of a one-one-valent electrolyte, which in this case is equal to its concentration. In this case a surface charged density of $\sigma_0 = 0.014 \text{ C m}^{-2}$ was chosen which corresponds to that of a human erythrocyte.

Fig. 2.46 The ζ -potential (depicted as a negative value) as a function of the concentration of an one-one-valent electrolyte for a surface with a constant surface charge density ($\sigma_0 = 0.014 \text{ C m}^{-2}$) according to Eq. 2.86



Equation 2.86 can be simplified expanding it as a series, corresponding to Eq. 2.80. If the Debye–Hückel parameter (κ) was used according to Eq. 2.55, we have

$$\sigma_0 = \zeta \kappa \epsilon \epsilon_0 \quad (2.87)$$

which, when combined with Eq. 2.85, results in

$$\sigma_0 = \eta \kappa b \quad (2.88)$$

Introducing this kind of linearization we found that this simplification can be accepted as a reasonable approximation only for potentials $\leq 10 \text{ mV}$. As indicated, however, in Fig. 2.46, this limit is exceeded for most cases of biological membranes.

Furthermore, all equations were derived considering the fixed charges as located on a planar boundary. The thickness of the Helmholtz layer, which amounts only to a few tenths of a nanometer explains what the word “planar” means. The roughness in the order of a molecular diameter is always enough to exceed this assumption.

To measure the electrophoresis a piece of equipment is used, which allows observation of the movement of cells in a homogeneous dc-field microscopically (Fig. 2.47). Usually, for electrophoretic measurements, a field strength of the order of 1 kV m^{-1} is used. The chamber needs a sufficient temperature control, because according to Eq. 2.85 the electrophoretic movement depends on the viscosity of the solution. Furthermore, it must be considered that an electro-osmotic flow occurs, caused by the surface charges of the glass, which mostly are negative. Hence the chamber is closed, this flux must return through the center of the chamber. Between these two flows of opposite directions, neutral planes must exist, where the streaming velocity becomes zero. This is the location, where it is possible to measure a correct electrophoretic mobility of cells. These planes are located at a distance, 21.1% of the total thickness of the chamber on both sides of it (see Fig. 2.47).

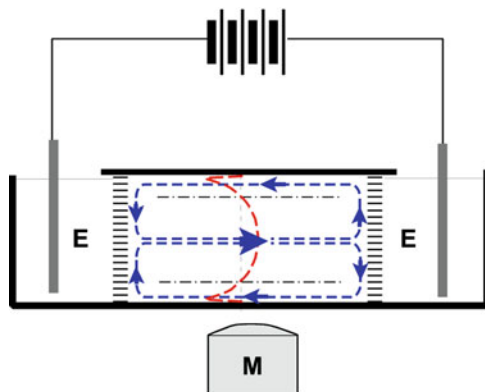


Fig. 2.47 Schematic illustration of equipment to measure cell electrophoresis; **M** microscope; **E** chamber for electrodes, which are connected to the measuring chamber by a conducting medium. The *blue arrows* indicate the direction of the electro-osmotic flow. The *red line* shows the flow profile in the chamber. *-----* the neutral planes

Since the development of the method of cell electrophoresis many cells and organisms have been investigated. Under physiological conditions, nearly all biological objects indicate a ζ -potential between -10 and -70 mV. Only at $\text{pH} < 4$ in some cases, do these potentials become positive. This holds also for bacteria, yeast, and plant cells, which are surrounded by a cell wall. The cell wall is a thick (in relation to the membrane) porous ion exchange layer with fixed negative charges.

In contrast to cell electrophoresis, which is used routinely to measure surface charges, the reversible effect, i.e., the *sedimentation potential*, sometimes also called the *Dorn effect*, seems not to be so important. In this case, a small electrical potential difference occurs in suspensions of quickly sedimenting charged particles.

An electrical potential difference which occurs if a solution is streaming along a charged surface, or inside a capillary with surface charges, is called the *streaming potential*. The reason for this is the superposition of the streaming profile and the profile of mobile charges in an electrical double layer. In detail, the theory of streaming potential is quite complex. In the center of a capillary, for example, the streaming velocity is maximum, and the concentration of the co-ions of the surface charges – minimum. Therefore, these co-ions will flow faster than the counterions near the surface. This leads to a certain charge differentiation at the ends of the capillary. The resulting electric field, however, enforces the movement of counterions and lowers the movement of the co-ions. The effect, therefore, is lowered by this reaction. In capillaries with an extremely small diameter, additionally, a Donnan effect must be considered. This means, that the amount of surface charges always shifts the total electrolyte compositions in this fine capillary toward a Donnan equilibrium (see Sect. 3.2.4). The total amount of counterions, therefore, is larger than that of the co-ions. This can even lead to a streaming potential with opposite polarity. It is possible to measure the streaming potential in biological

tissue, if a flow of solutions is induced by osmotic gradients. The potential difference, measured in such experiments is only of the order of a few millivolts.

The reversible effect of streaming potential is *electro-osmosis*. In this case, based on the same considerations, near a charged surface, or in a capillary, a flow of solution occurs if a dc-field is applied.

Recently, the combination of electrophoresis and electro-osmosis has been used to measure the ζ -potential of the intercellular space in tissue. For this, small spots of fluorescent probes were injected into tissue slides and their movement under the influence of an applied electric field was observed microscopically. This movement is the result of the electrophoresis of the marker molecules driven by their own charges, plus their transport by the electro-osmotic flow, generated by the surface charges of the intercellular cleft. The electrophoretic mobility of the applied molecules can be measured separately, which allows us to evaluate the electro-osmotic effect separately. A neutral compound without an electrophoretic mobility in a cleft with negative charges would move toward the cathode (see Fig. 2.45). One difficulty is the tortuosity of the path in the intercellular space. This problem can be overcome by simultaneously using two fluorophores of different sizes and each differently charged. As the result of such experiments, the ζ -potential of brain tissue has been found to be in the same range as that of isolated cells, i.e., around -25 mV.

Furthermore, these kinds of investigations are important to give a quantitative insight into the interplay between different phenomena and permanent properties influencing the process of *iontophoresis*. This is a noninvasive technology for the delivery and monitoring of drugs through the skin epidermis under the influence of an applied electric field to increase and modulate the rate of transport.

The physiological role of electrokinetic phenomena is not very clear. It cannot be ruled out that electrophoresis of charged components in the membrane, or of proteins in intercellular spaces occurs under the influence of in vivo generated potentials (see Sect. 3.5.2). A hypothesis exists, whereby microorganisms can move electrophoretically in their own generated field. Streaming potentials occur in bone during mechanical deformations. In this case, it is induced by the flow in the small canals of the bone, the so-called canaliculi (see Sect. 3.5.2).

Further Reading

Theory of electrical double layers: Donath and Voigt 1986; Glaser 1996; Voigt and Donath 1989; cell electrophoresis: Bauer 1994; streaming potential in bones: Rubinacci et al. 2002; self-electrophoresis of microorganisms: Lammert et al. 1996; electrokinetic phenomena in tissue: Guy et al. 2008; iontophoresis: Imanidis and Luetolf 2006.

2.3.6 The Electrostatic Structure of the Membrane

After explaining some structural and mechanical properties of biological membranes in previous Sections, and after having discussed ionic properties near

fixed charges, we will now introduce the membrane as an electrical structure. This is a crucial point for many biophysical considerations, and it will pervade this whole book. Electrical properties of the biological membrane will be mentioned in the context of many topics. Whereas we introduce here, first of all the electrostatic properties of the membrane, we will continue this discussion later in Sects. 3.4.3 and 3.5.2, including discussion of electrodynamic processes, such as generation of electrical currents by processes coupled with cell metabolism. In Sect. 3.5.3 the explanation of passive membrane properties will be continued, including its behavior in electromagnetic fields. This knowledge will be applied in Sect. 3.5.5 to introduce techniques of cell manipulation with electric fields. Finally, in Sects. 4.5, 4.6, and 4.7 we will use this model to discuss the influence of weak electric and electromagnetic fields on cells and organisms.

In contrast to the external medium and to the cell plasma, the cell membrane has a high electrical resistance and a low dielectric constant. The membrane is an extremely thin hydrophobic electrically isolating interface between two aqueous phases. It can be considered to be like an electrical RC-system with certain capacity (C), and an electric resistance (R) (see Figs. 3.27 and 3.39).

The specific capacity (C_{sp}) can be calculated from the membrane thickness (Δx) and the dielectric constant (ϵ):

$$C_{sp} = \frac{\epsilon_0 \epsilon}{\Delta x} \quad (2.89)$$

The specific capacity of the cell membrane is relatively constant, because both parameters, the dielectric constant (ϵ), as well as the membrane thickness (Δx) cannot vary to a great extent. For the normal cell membrane it amounts approximately to 10 mF m^{-2} (for methods used to measure this parameter, see Sects. 3.5.3 and 3.5.4). Using this value, the rearranged equation (Eq. 2.89) allows calculation of the mean dielectric constant of the membrane, assuming: $\Delta x = 8 \cdot 10^{-9} \text{ m}$:

$$\epsilon = \frac{C_{sp} \Delta x}{\epsilon_0} = \frac{10^{-2} \cdot 8 \cdot 10^{-9}}{8.854 \cdot 10^{-12}} = 9.0$$

This value is surprisingly large, because the dielectric constant of lipids amounts only to $\epsilon = 3.5$. The reason for this difference is the heterogeneity of the membrane, in particular, its content of proteins.

The specific capacity of the biological membrane is a very important parameter, indicating the relation between the amount of charges, resp. their distribution density (σ , in C m^{-2}), which are required to generate a membrane potential difference ($\Delta\psi$, in V):

$$C_{sp} = \frac{\sigma}{\Delta\psi} \quad (2.90)$$

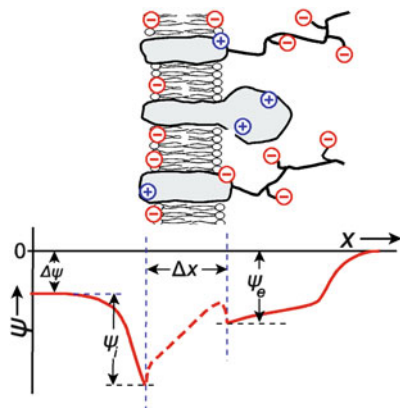


Fig. 2.48 The fixed charges of a cell membrane and the corresponding electrical potential (ψ). The transmembrane potential ($\Delta\psi$) is indicated as the difference between the potentials of the internal and external phases. The parameters ψ_i and ψ_e indicate the internal and external surface potentials; Δx means the thickness of the membrane. Inside the membrane itself, the function $\psi(x)$ depends on the positions at which the membrane is cut. Therefore, the function in this region is marked only by a simple connection between the surface potentials, considering only the dipole effect of the polar lipid groups

In Sect. 3.4.3 we will use this relation to explain the role of the membrane as an accumulator of electrochemical energy.

Despite the usefulness of these rough estimations it must be mentioned that the real molecular structure of the membrane and its surface, and therefore the real dielectric properties of it, in fact, are more complicated. In Sect. 2.3.3 we already described the membrane as a dynamic structure containing fixed charges in a two-dimensional pattern of the membrane area and in a charge profile perpendicular to this. This particular profile along an x -coordinate, cutting the membrane area perpendicularly, is depicted schematically in Fig. 2.48. This picture does not contain the mobile ions in the double layers. It assumes the existence of a transmembrane potential ($\Delta\psi$) as a potential difference between the inner and the outer phase of the cell. This could be a result of a Donnan equilibrium, of a diffusion potential, or finally directly generated by electrogenic transporters. These circumstances will be discussed later in Sect. 3.4 in detail.

The fixed surface charges of the outer part of the cell membrane mainly are the result of dissociations of carboxyl groups of neuraminic acids (also called: sialic acids), positioned at the ends of the glycoproteins. The isoelectric point of these neuraminic acids is near pH 2.6. For most membranes the surface density of negative charges is between -0.01 and -0.02 C m^{-2} . Human erythrocytes carry on their surface, which is about $140 \text{ } \mu\text{m}^2$, a number of approximately 10^7 dissociated neuraminic acid groups.

Additional carriers of electric charges are the polar groups of several phospholipids. The negatively charged phosphatidylserine is of particular importance.

In most cells it is located exclusively at the inner leaflet of the membrane. In human erythrocytes its surface charge density amounts approximately to -0.09 C m^{-2} .

As already mentioned in Sect. 2.3.3, proteins are located in the membrane in such a way that their polar groups are oriented towards the water phase (Fig. 2.39). Hence, they also will contribute to the surface charge of the membrane by charges of both sign. These charges of membrane proteins are very important controlling various protein functions, such as for example the ion transport through the membrane.

These considerations indicate that the models of electrical double layers, resulting from charges of a planar surface (Fig. 2.43) are just rough approximations for the case of the real cell membrane. In fact the glycocalyx forms an external layer covering the surface of most cells. The fixed charges of it are to be considered as charges in a definite space, but not directly on the plane of the surface. It must be therefore introduced into the Poisson–Boltzmann equation (see Eq. 2.75) as space density (ρ , in C m^{-3}), instead of the surface charge density (σ , in C m^{-2}):

$$\nabla^2 \psi = -\frac{1}{\epsilon_0 \epsilon} \left(\rho + F \sum_{i=1}^n c_{i0} z_i e^{-\frac{z_i e \psi}{kT}} \right) \quad (2.91)$$

Even for the simplest functions $\rho(x)$, which are to be introduced into Eq. 2.91, an integration of this equation is possible only by iteration. The resulting function $\psi(x)$ (Fig. 2.48), of course differs from that of a classical double layer (Fig. 2.43). The effective range of the electrical potential in this case is predicted first by the thickness of the glycocalyx, and only indirectly, by the Debye–Hückel length ($1/\kappa$). This thickness, however, itself is determined by electrostatic interactions of its charges. In solutions of high ionic strengths, these charges are more or less shielded, and the glycoproteins are more closely attached to the membrane. In solutions with lower ionic strength, however, the fixed charges repulse each other, the glycoprotein filaments, therefore are packed more loosely, and the glycocalyx becomes thicker. Hence, the function $\rho(x)$, i.e., the distribution of the charges of the neuraminic acids, itself depends on $\psi(x)$. The surface structure of the cell therefore is controlled by the cell itself, as well as by the conditions of its surroundings.

In the case of cells in tissue these glycoproteins are packed in the intercellular cleft. As pointed out in the previous Sect. 2.3.5 it is possible to obtain a picture of this situation, measuring the electrokinetic movement of fluorescent dyes in this region. The Donnan osmotic properties of this layer control its dimension. In Sect. 3.5.2 the intercellular cleft will be considered as a region for extracellular ionic currents (see Fig. 3.34).

Not only the external, but also the internal membrane surface carries fixed charges. These are in the first line the charges of phosphatidylserine, and additionally, a number of surface-associated proteins, which are not depicted in Fig. 2.48. In contrast to the surface charges at the outside of the cell, which can be measured by

cell electrophoresis, the charges of the cytoplasmatic side cannot be measured easily.

In Fig. 2.48 the function $\psi(x)$ inside the membrane, the *inner membrane potential*, is depicted only schematically. In fact, this function strongly depends on the place, where the cut was drawn. If it is taken through the lipid phases, the influence of the lipid head groups as dipoles must be considered, as schematically depicted in this figure. The profile through a protein region, of course is quite different from this.

Beside this electrical profile in a direction perpendicular to the membrane surface, a lateral mosaic of charges exists, which is partly controlled by the cytoskeleton. This charge distribution has been investigated by electron microscopy, and recently also by atomic force microscopy. It is also possible to measure the lateral electrophoretic movement of charged membrane proteins, marked by fluorescence labels if the cells are observed in strong external fields. The investigation of the dynamics of this lateral electrostatic structure of biological membranes and its regulatory function is, however, only at the beginning.

Although we will discuss the role of the membrane electric field, as well as the application of external fields to cell membranes in detail later in Sects. 3.4, 3.5, as well as 4.5–4.7, we will nevertheless already mention here the biological relevance of this electrostatic structure, and its role in several cell physiological processes. In general, on one hand, the profile of the electrical potential near the membrane, directly creates a profile of specific ion conditions. On the other hand, its first derivative, the electric field strength, affects some processes directly.

Inside the membrane, and in its vicinity, the electric field strength is of the order of 10^7 V m^{-1} . In Fig. 2.48 the course of the potential inside the membrane is marked only by a simple connection between the surface potentials, considering just the dipole effect of the polar lipid groups. It is a dotted line, because the real function $\psi(x)$ in this region strongly depends on the location, where the membrane was cut. This function, representing the intermembrane potential, would be quite different, if the x -axis were to cut a protein. The local charge densities, as well as the local dielectric constants in these cases would differ strongly. Furthermore, it must be considered that the function $\psi(x)$ will change dramatically, if the transmembrane potential $\Delta\psi$ is depolarized during the action potential in a time course of about one millisecond (see Fig. 3.29). Hence, we come to the conclusion that inside and near the membrane, an extremely strong electric field occurs which is not only a one-dimensional function $\psi(x)$, as depicted in Fig. 2.48, but which rather must be considered as a three-dimensional time-dependent parameter.

It is clear that this field must influence all polar or polarizable molecules in this region. This particularly concerns membrane proteins containing special segments acting as voltage sensors. Nerve cells are the classical example of a feedback circle, generating the action potential: An alteration of the membrane permeability, caused by any influences, changes the diffusion potential $\Delta\psi$. The resulting modification of the field strength in the membrane changes the permeability of the Na^+ , and subsequently K^+ -channels, which again influences the membrane potential. We will explain these reactions in detail in Sect. 3.4.4. Potential sensitive transporters occur in all cells.

Not only proteins, but also lipids are influenced by the field strength inside the membrane. The field influences the orientation of the polar head groups of the lipids, and in this way modifies the effective area which they occupy in the membrane. We already discussed the consequence of this modification in Sect. 2.3.2. The field may also induce an intrinsic membrane pressure, which may alter the phase transition in the hydrophobic region of the membrane. Such processes are the basis of mechanisms of electromechanical coupling.

The electrical potential near the membrane generates particular microconditions near charged groups. Dependent on the ionic strength, and therefore on the Debye–Hückel length ($1/\kappa$), a lateral mosaic of local concentrations, and local pH values can occur. In Fig. 2.44 the local ionic concentrations and local pH values are illustrated as a function of distance from a charged surface. It must be emphasized that the field changed the concentration of multivalent ions, as for example Ca^{++} to a larger degree than univalent ions like K^+ or Na^+ . In this context, the role of local charges must be considered for the control of membrane-bound enzymes and transporters.

Recently it was shown that this extremely high electric field strength near the membrane also influences the dissociation constant of molecules and molecular complexes. This process is applicable to a wide range of biological complexes including small molecules (norepinephrine-morphine sulfate), protein-protein complexes (insulin-glucagon), and small molecule-protein complexes (epinephrine-bovine serum albumin).

Further Reading

Electrostatic structure: Cevc 1990; Glaser 1996; Honig et al. 1986; McLaughlin 1989; role of membrane lipids: Starke-Peterkovic et al. 2005; surface mapping with atomic force microscopy: Heinz and Hoh 1999; effects of membrane near electric field: Dillon et al. 2006; lateral distribution of charges: McLaughlin and Murray 2005.

Chapter 3

Energetics and Dynamics of Biological Systems

While in the previous part of this book basic physical principles are explained governing the formation of molecular and supramolecular biological structures, we will come now to various functions of cells, tissues, organs, and organisms. For this, of course, molecular considerations form an important fundament, but at the same time, phenomenological parameters, like concentration, volume, viscosity, dielectric constants, conductivity, etc., are used which in fact are defined for large and homogeneous systems. In this way, we begin to enter the field of the so-called continuum physics.

At the same time, the approaches of quantum mechanics and statistical thermodynamics will now be replaced by those of phenomenological thermodynamics. These approaches also are defined for sufficiently large homogeneous phases. We should always be aware that in some cases, this in fact, does not meet the particular conditions of the biological system. We mentioned this problem in the previous part of this book, considering for example, mechanical or electrical properties of biological membranes. The step from a molecular to a phenomenological approach, nevertheless, is inevitable when considering larger systems like cells and organisms.

We will come back to this point in general in context with the electrical structure of organisms, discussing levels of biological organization in Sect. 3.5.1 (Fig. 3.33).

3.1 Some Fundamental Concepts of Thermodynamics

One of the first important treatise on the problems of thermodynamics was the famous monograph by Sadi Carnot, entitled “Réflexions sur la puissance motrice du feu et sur les machines propres à développer cette puissance” which was published in 1824. In this book, thermodynamics was introduced as the theory of heat engines. Julius Robert Mayer (1814–1878) was a physician and scientist to whom we owe

the first numerically defined correlation between heat and work (1842). He therefore can be considered as the discoverer of the first law of thermodynamics. As a result of physiological observations he had already discussed how the functioning of the human body related to energy transformation in a heat engine. Meanwhile, thermodynamics has become the theoretical fundament of all kinds of energy transformation, and consequently, of all kinds of movement.

Applying thermodynamic considerations to open biological systems, however, requires an extension towards explanations of irreversible processes, i.e., towards nonequilibrium thermodynamics. This extension is made in two steps: Firstly, only small deviations away from the equilibrium are considered. In this case, linear relations between forces and rates can be assumed. In contrast to these *linear approaches*, the thermodynamics of *nonlinear* processes is applied to systems far from equilibrium. In this case, so-called *dissipative structures* appear, which are stationary states with completely new qualities.

It seems important to emphasize here that although the far-from-equilibrium condition of an organism represents an absolute precondition of life, nevertheless there exist many subsystems, which can be properly calculated using equilibrium thermodynamics, or thermodynamics of linear approaches. This means that biophysics must concern the whole scale of thermodynamic approaches.

3.1.1 Systems, Parameters, and State Functions

In Sect. 2.1.3, the term “system” was introduced in context with an explanation of the term “structure.” We defined the system as an aggregate of elements with certain interrelations. Furthermore, we explained that systems, the interrelations of which are not simply relations, but interactions, are so-called dynamic systems. These are the kinds of system to which thermodynamic approaches are to be applied.

The question of what kind of model we should use, what we should consider as a system, and what are its elements, depends exclusively on the particular problem, and the corresponding point of view. An element of one kind of system can become a system in itself when calculating another problem. An organism, for example, as an element in an ecological system can become a system itself, if we ask a physiological question, such as for example the interactions between its organs. The organ can be considered as a system of cells, the cell as a system of organelles, and so on.

A dynamic system can be analyzed in different ways. In contrast to system theory, which calculates the kinetic interplay of individual elements (see the Introduction to Sect. 5), thermodynamics considers a system simply as a continuum which stands in a defined interrelation with its environment. The limit of this continuum does not have to be a wall or a membrane. It can also be a process that changes the quality of the subject of study. Chemical reactions as well as processes of proliferation and evolution are examples of this.

In thermodynamics systems are classified as follows according to the nature of their boundary against their environment:

- The *isolated system*: this is an idealized system that does not exchange any kind of energy or matter with its environment;
- The *closed system*: this system can exchange all kinds of energy with its environment, but not matter;
- The *open system*: it can exchange both energy and matter with its environment.

The closed system can be influenced by its environment, and can cause changes in its environment. However, it cannot be involved in an exchange of matter.

The state of a system can be described by a number of *state variables*. These are either extensive or intensive parameters. *Intensive* parameters are nonadditive and independent of the size of the system (e.g., temperature, concentration, pressure, density). Conversely *extensive* parameters are additive when two systems are combined (e.g., mass, volume).

Changes in a system are often characterized by differentials of its state variables. A *differential* describes a very small change of a dependent variable (dy), if in a function $y = f(x)$, a small change in the variable (dx) occurs. It can be calculated from the product of the first derivative of the function $f(x)$, multiplied by dx :

$$dy = f'(x)dx \quad (3.1)$$

Most thermodynamic equations are functions with several variables. Hence, the derivatives can be obtained with respect to one variable if the others are kept constant.

This procedure is called *partial differentiation*. It has a special notation with the parameters that are to be kept constant put as subscript to the symbols in parentheses. The following example is quoted out of context, to demonstrate this.

$$\left(\frac{\partial G}{\partial n_i} \right)_{p,T,n_j} = \mu_i \quad \text{for } j \neq i \quad (3.2)$$

The partial derivative of the Gibbs free energy G with respect to the molar number of substance i , when pressure (p), temperature (T), and the molar number of all the other substances (n_j) are kept constant, gives per definition, the chemical potential (μ_i) of the substance i . In general, this is the same procedure as is used when a function with several dependent variables is represented graphically in a two-dimensional plot against one selected variable, keeping all other variables constant.

Small changes in a state function with several variables can be represented by a so-called *total differential*. For this, all partial differentials of this function must be summarized. These partial differentials are calculated as shown in Eq. 3.1, using,

however, partial derivatives. The following equation for example, would apply to the Gibbs free energy $[G(p, T, n_i)]$:

$$dG = \left(\frac{\partial G}{\partial p}\right)_{T, n_i} dp + \left(\frac{\partial G}{\partial T}\right)_{p, n_i} dT + \sum_{i=1}^m \left(\frac{\partial G}{\partial n_i}\right)_{p, T, n_j} dn_i \quad (3.3)$$

The mathematical definition of the total differential is of very great physical importance to thermodynamics. This will be indicated by the following chain of statements with reversible logical connections:

dG is a total differential	\leftrightarrow	G is a state function	\leftrightarrow	G depends only on the state of the system, and not on the way in which that state was achieved
---------------------------------	-------------------	----------------------------	-------------------	---

For this reason it is important that this property of a function is able to be mathematically proven.

A differential equation is not always written in the easily followed way shown in Eq. 3.3.

Often it is presented as a *Pfaffian differential equation*:

$$dn = Ldx + Mdy + Ndz \quad (3.4)$$

The capital letters here represent any variable. There is a mathematical indicator which allows one to determine whether dn is a total differential. This is the so-called *Cauchy condition*, stressing that dn is a total differential if the following conditions are fulfilled:

$$\frac{\partial L}{\partial y} \stackrel{!}{=} \frac{\partial M}{\partial x}; \quad \frac{\partial M}{\partial z} \stackrel{!}{=} \frac{\partial N}{\partial y}; \quad \frac{\partial L}{\partial z} \stackrel{!}{=} \frac{\partial N}{\partial x} \quad (3.5)$$

If this is applied to Eq. 3.3, this means:

$$\frac{\partial \left(\frac{\partial G}{\partial p}\right)}{\partial T} = \frac{\partial \left(\frac{\partial G}{\partial T}\right)}{\partial p} = \frac{\partial^2 G}{\partial T \partial p} \quad (3.6)$$

From Eq. 3.6 it follows additionally that for such functions if they are differentiated several times, the sequence of differentiations is unimportant. We will use this property in a later derivation (Eq. 3.82).

Total differentials not only result from energetic parameters. This formalism of course can be applied to any state function. Volume changes in mixed phases for example, can be described by the following total differential equation:

$$dV = \bar{V}_1 dn_1 + \bar{V}_2 dn_2 + \bar{V}_3 dn_3 + \dots + \bar{V}_m dn_m \quad (3.7)$$

where dV is the shift of the volume that occurs when the molar number of one or more components of the system is changed. Furthermore, Eq. 3.7 allows one to define the partial molar volume of a given substance i :

$$\bar{V}_i = \left(\frac{\partial V}{\partial n_i} \right) \quad \text{for } j \neq i \quad (3.8)$$

The partial molar volume has the inverse unit as the concentration, namely: $\text{m}^3 \text{mol}^{-1}$.

3.1.2 Gibbs Fundamental Equation

The scientific basis of thermodynamics is its three principles which are founded on experimentally verifiable, empirical facts. Upon this solid foundation a framework of definitions and relations has been built up which enables far-reaching postulations on all kinds of energy transformations to be made.

The principle of conservation of energy, the so-called *first law of thermodynamics* states that there must exist a physical parameter having the property of a state function, which includes the consequences discussed in Sect. 3.1.1. Work (W), as a physical parameter does not comply with this condition. General experience shows that a change of a system from state A to state B can be achieved in many ways that differ greatly from one another in the amount of work that is required. Therefore, work cannot be a state function that could be used to characterize the energy state of a system independently of the way in which it was achieved.

Let us introduce a parameter called *internal energy* (U). It shall be defined as a state function which, as such, has a total differential dU . Let furthermore the internal energy of a system be increased by dU if a certain amount of heat (dQ) is introduced into the system, and/or if certain work (dW) is done in the system:

$$dU = dQ + dW \quad (3.9)$$

This equation contains the essence of the first principle of thermodynamics. Usually, it is derived in detail with the help of the so-called *Carnot cycle* reflecting processes in steam engines. For simplification, not losing any detail of biophysical relevance, we choose here the simpler way of defining this.

Both, the differentials dQ , as well as dW , are reflections of a change of energy. However, according to the *second principle of thermodynamics*, the heat (Q) differs from all other forms of energy because it possesses a specific property: Any form of energy can be completely transformed into heat but heat itself can only partly be transformed into work.

Here again the entropy (S) has to be inserted. It is the same parameter as introduced in Sect. 2.1.1 in its statistical character (Eq. 2.4). In phenomenological

thermodynamics entropy appears as a kind of measure of heat quality. For a quasi-reversible process it is defined as follows:

$$dS = \frac{dQ_{\text{rev}}}{T} \quad (3.10)$$

This equation offers an expression for dQ which can be introduced into Eq. 3.9.

Let us now consider in more detail the differential of work (dW) in Eq. 3.9. This can be a sum of different kinds of work. Each being a product of a work coefficient multiplied by a work coordinate. *Work coefficients* are intensive parameters indicating a kind of measure of the performed work. In contrast, *work coordinates* are extensive parameters, reflecting the effort of the performed work. For example, the work (dW_p) which is done when a gas is compressed by a pressure p resulting in a volume alteration dV will be:

$$dW_p = -p dV \quad (3.11)$$

where p represents the work coefficient and dV the work coordinate. The sign of this equation depends on the definition of the work differential. A positive dW means that there is an increase in the work done in favor of the system. In this case work is achieved through the compression, i.e., a negative differential of the volume.

Work can be done in many different ways. An expansion of a material, for example, means elongation (dl) in response to the application of a force (\mathbf{F}):

$$dW_l = \mathbf{F} dl \quad (3.12)$$

A cell can do work by transporting a certain number of atoms or molecules (dn) against a concentration gradient. At this point, the chemical potential (μ) must be introduced as the work coefficient. We will come back to this parameter in detail later (Eq. 3.33). In this case the work differential can be written as follows.

$$dW_n = \mu dn \quad (3.13)$$

Let us finally, from among the many other possible examples, consider charge transport. If a particular amount of charge (dq) is transported against an electric potential (ψ), then the electrical work done will be:

$$dW_q = \psi dq \quad (3.14)$$

Equations 3.11, 3.12, 3.13, and 3.14 can be combined:

$$dW = -p dV + \mathbf{F} dl + \mu dn + \psi dq \quad (3.15)$$

Considering that usually in the system a number of m substances are transported, then this equation can be expanded as follows:

$$dW = -p \, dV + \mathbf{F} \, dl + \sum_{i=1}^m \mu_i \, dn_i + \psi \, dq \quad (3.16)$$

This is a more detailed form of the work differential which together with Eq. 3.10, can be introduced into Eq. 3.9:

$$dU = T \, dS - p \, dV + \mathbf{F} \, dl + \sum_{i=1}^m \mu_i \, dn_i + \psi \, dq \quad (3.17)$$

Equation 3.17 is a differential form of the so-called *Gibbs fundamental equation*. Of course it can be expanded optionally by adding more kinds of work differentials, for example for magnetic field influences (see: Eq. 4.19, Sect. 4.4). Alternatively, this equation will be automatically reduced if certain changes become irrelevant. For example, suppose a defined transformation within a system is not accompanied by a mechanical strain. Then l remains constant, and consequently, $dl = 0$. As a consequence, the corresponding term disappears from the equation.

In Eq. 3.17 the Gibbs fundamental equation appears in the form of a *Pfaffian differential*. Such expressions can be integrated under certain conditions, which apply in this case. This gives:

$$U = T \, S - p \, V + \mathbf{F} \, l + \sum_{i=1}^m \mu_i \, n_i + \psi \, q \quad (3.18)$$

It must be noted that the transition from Eq. 3.17 to 3.18 does not mean a simple elimination of the differential operators; it is the result of a proper integration which is not described here!

Using the rule given in Eq. 3.3, this process of integration can be reversed. It gives:

$$\begin{aligned} dU = & T \, dS + S \, dT - p \, dV - V \, dp + \mathbf{F} \, dl + l \, d\mathbf{F} \\ & + \sum_{i=1}^m \mu_i \, dn_i + \sum_{i=1}^m n_i \, d\mu_i + \psi \, dq + q \, d\psi \end{aligned} \quad (3.19)$$

A comparison of this result with the initial equation (3.17) shows that the following condition must be satisfied:

$$S \, dT - V \, dp + l \, d\mathbf{F} + \sum_{i=1}^m n_i \, d\mu_i + q \, d\psi = 0 \quad (3.20)$$

This is the so-called *Gibbs–Duham equation*. It is useful for some calculations because it allows one to reduce the degree of freedom of a system by one variable.

It has proved useful to define not only the internal energy (U), but also three further energy functions. In some books the introduction of these parameters is explained in a physical way, discussing processes of vapor compression, etc., but it seems to be simpler just to accept the definitions of these parameters, and subsequently substantiate their usefulness.

The definitions are:

$$\text{enthalpy: } H = U + pV \quad (3.21)$$

$$\text{Helmholtz free energy } F = U - TS \quad (3.22)$$

$$\text{Gibbs free energy } G = H - TS \quad (3.23)$$

The Gibbs fundamental equation (Eq. 3.17) can now easily be written down for these new defined functions. Let us first transform Eq. 3.21 into a total differential, according to Eq. 3.3. Using the definition (Eq. 3.21), the enthalpy (H) is a function of: U , p , and V . This gives:

$$dH = \left(\frac{\partial H}{\partial U} \right)_{p,V} dU + \left(\frac{\partial H}{\partial p} \right)_{U,V} dp + \left(\frac{\partial H}{\partial V} \right)_{U,p} dV \quad (3.24)$$

From Eq. 3.21 follows directly:

$$\left(\frac{\partial H}{\partial U} \right)_{p,V} = 1; \quad \left(\frac{\partial H}{\partial p} \right)_{U,V} = V; \quad \left(\frac{\partial H}{\partial V} \right)_{U,p} = p \quad (3.25)$$

which, when combined with Eq. 3.24, results in

$$dH = dU + Vdp + pdV \quad (3.26)$$

Combining this with Eq. 3.17 gives the Gibbs fundamental equation for dH :

$$dH = TdS + V dp + \mathbf{F} dl + \sum_{i=1}^m \mu_i dn_i + \psi dq \quad (3.27)$$

In the same way it is possible to derive from Eq. 3.22 the relations:

$$dF = S dT - p dV + \mathbf{F} dl + \sum_{i=1}^m \mu_i dn_i + \psi dq \quad (3.28)$$

and:

$$dG = S dT + V dp + \mathbf{F} dl + \sum_{i=1}^m \mu_i dn_i + \psi dq \quad (3.29)$$

The choice whether Eq. 3.17, 3.27, 3.28, or 3.29 should be used to calculate a particular system depends on the external conditions and the question which is being asked.

Investigating a system under isobaric conditions ($p = \text{const.}$, i.e., $dp = 0$), it is useful to apply the equation for dH (Eq. 3.27), or for dG (Eq. 3.29), because in this case the term of the volume expansion work (Vdp) vanishes. This corresponds to the situation of most biological investigations. Therefore we will use mostly the enthalpy (H) and the Gibbs free energy (G) in all further biophysical calculations instead of the inner energy (U) and the Helmholtz free energy (F).

If the conditions are isothermal ($dT = 0$), as well as isobaric ($dp = 0$), then in Eq. 3.29 the term, connected with heat, as well as that for volume work will vanish. Hence, dG expresses directly the deviation of the energy content, as a result of work which was done. The gradient of free Gibbs energy therefore indicates the direction of a spontaneous process in the same way as a gradient of the potential energy indicates the path of a rolling sphere on an inclined surface.

All these forms of the Gibbs fundamental function (Eqs. 3.27, 3.28, 3.29), as well as Eq. 3.7 for partial volume, can be integrated according to Eq. 3.17 for dU .

In this way, a particular chemical reaction can be characterized by differences of these parameters:

$$\begin{aligned}\Delta_R G &= G_{\text{product}} - G_{\text{substrat}} \\ \Delta_R H &= H_{\text{product}} - H_{\text{substrat}} \\ \Delta_R S &= S_{\text{product}} - S_{\text{substrat}}\end{aligned}\tag{3.30}$$

The parameter $\Delta_R G$ indicates whether this reaction occurs spontaneously, i.e., whether $G_{\text{product}} < G_{\text{substrate}}$, i.e., $\Delta_R G < 0$. Conversely the reaction enthalpy $\Delta_R H$ is a measure of the thermal characteristics of isobaric processes. This means:

$\Delta_R H > 0$ – endothermic reaction

$\Delta_R H < 0$ – exothermic reaction

How does one obtain the parameters as used in Eq. 3.30? Only in the case of entropy are absolute values available. This is a result of the *third principle of thermodynamics*, based on the heat theorem of Walther Nernst, stating that at a temperature of absolute zero the entropy becomes zero.

In contrast to entropy, no absolute values exist for the energetic parameters. They always need a defined reference value. Therefore, *standard energies of formation* ($\Delta_F U$, $\Delta_F H$, $\Delta_F F$, and $\Delta_F G$) are defined as energetic changes that occur when a substance is formed from its elements under standard conditions ($T = 297$ K), or better: *would occur*, because in most cases a direct synthesis of the substance from its elements is impossible. In this case, they are estimated from particular sets of chemical reactions starting from substances with known energy of formation. This is possible using the definition of these parameters as state functions, their amount being independent of the way the state was achieved (see definition in Sect. 3.1.1).

According to definition (2.23), for isothermal systems this is:

$$\Delta_R G = \Delta_R H - T \Delta_R S \quad (3.31)$$

A spontaneous reaction ($\Delta_R G < 0$) therefore requires an exothermic reaction ($\Delta_R H < 0$) or the condition $\Delta_R H < T \Delta_R S$. In this case the direction of the reaction is determined by the rise in entropy. This type of process is called an *entropy-driven reaction*. The classical example of such a reaction is the melting of ice. In Sect. 2.2.2 we already discussed various biomolecular reactions of this type in context with the structure of water.

Let us now consider the chemical potential which we need for numerous considerations in the following sections. The chemical potential (μ_i) of the substance i is particularly important for the following calculations. It can be easily defined using Eqs. 3.17, 3.27, 3.28, or 3.29:

$$\mu_i = \left(\frac{\partial U}{\partial n_i} \right)_{S,V,l,n_j,q} = \left(\frac{\partial H}{\partial n_i} \right)_{S,p,l,n_j,q} = \left(\frac{\partial F}{\partial n_i} \right)_{T,V,l,n_j,q} = \left(\frac{\partial G}{\partial n_i} \right)_{T,p,l,n_j,q} \quad (3.32)$$

The chemical potential therefore, is a partial expression having the dimensions J mol^{-1} . The chemical potential of the substance i (μ_i) can be calculated from the concentration (c_i), resp. the chemical activity (a_i) of this substance using:

$$\mu_i = \mu_i^0 + RT \ln a_i \quad (3.33)$$

The chemical activity (a_i) is a kind of effective concentration. Its relation to concentration is given by:

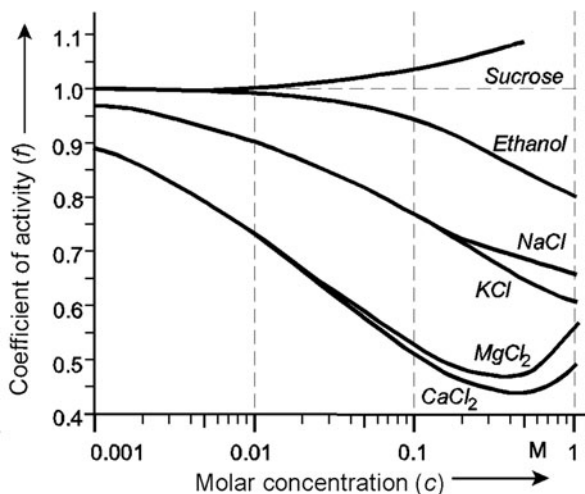
$$a_i = f_i c_i \quad (3.34)$$

In this equation, f_i is the *coefficient of activity*. In ideal solutions, $f_i = 1$, that is, the activity of the dissolved substance is equal to its concentration c_i . Usually f_i decreases as the concentration in the solution increases (see Fig. 3.1). For dissociating salts, f represents an average activity coefficient for the ions. For example, the ions in a 100-mM solution of NaCl show an average activity coefficient of 0.8. The chemical activity of this solution therefore is equal to an ideal solution with a concentration of only 80 mM. In contrast to the coefficient of activity, which is a dimensionless number, the chemical activity has the same units as the concentration.

In some cases it may be useful to employ the mole fraction as a measure of concentration. The *mole fraction* of a substance i is defined as the number of moles of that substance (n_i), divided by the total number of moles of all substances present:

$$x_i = \frac{n_i}{\sum_{i=1}^m n_i} \quad (3.35)$$

Fig. 3.1 Coefficients of activity (f_i) of various substances as functions of their concentrations (c_i) in aqueous solutions under standard conditions ($T = 297\text{ K}$)



where:

$$\sum_{i=1}^m x_i = 1 \quad (3.36)$$

According to Eq. 3.34, the mole fraction of a substance can also be expressed as the mole fraction activity (a_{xi}).

The standard potential (μ_i^0) can be easily defined by means of Eq. 3.33. It follows from this equation that when $a_i = 1$, $\mu_i = \mu_i^0$. The standard potential therefore is the chemical potential of a substance i in a solution, with an activity of 1 M, if this measure of concentration is applied. Using in Eq. 3.33 as concentration measure the mol fraction (x_i) of a substance, or its mol fraction activity (a_{xi}), then the chemical standard potential (μ_i^0) is determined by the chemical potential of the pure substance ($a_{ix} = 1$).

A further extension of the Gibbs fundamental equation concerns the term ψ/dq . The charge of a single charged ion is determined by the Faraday constant (F). The charge on n moles of a z -fold charged ion is obtained as follows:

$$q = znF \quad (3.37)$$

This is a function with one independent variable (n). Therefore, it is easily transformed into a differential according to Eq. 3.1:

$$dq = \left(\frac{dq}{dn} \right) dn = zF dn \quad (3.38)$$

If more than one ion is in the solution, the charges can be summarized:

$$dq = \sum_{i=1}^m z_i F dn_i \quad (3.39)$$

Introducing this expression into the Gibbs fundamental equation, the equation then gets two terms with the differential dn_i . It is obviously useful to combine these terms. First, consider these two terms of the Gibbs equation in isolation from the other terms of this equation:

$$\psi \sum_{i=1}^m z_i F \, dn_i + \sum_{i=1}^m \mu_i \, dn_i = \sum_{i=1}^m (\mu_i + z_i F \psi) dn_i = \sum_{i=1}^m \tilde{\mu}_i \, dn_i \quad (3.40)$$

Here, the expression inside the brackets taken together is described as the *electrochemical potential* ($\tilde{\mu}_i$) of the substance i .

$$\tilde{\mu}_i = \mu_i + z_i F \psi \quad (3.41)$$

The electrochemical potential is the basis for most electrochemical calculations, and thus forms an important starting point for further considerations.

3.1.3 Force and Motion

After introduction of the generalized functions for the energetic state of a system in the previous section we will now consider their application in determining forces leading to any sort of motion.

A sphere is rolling downhill. It is moving spontaneously from a position with a higher potential energy to one with a lower potential. The direction of this movement follows a force vector (\mathbf{X}) and is, consequently, determined by the negative gradient of the energy U .

$$\mathbf{X} = -\text{grad } U \quad (3.42)$$

If consideration of the energy gradient is confined to the direction of the x -coordinate, this equation can be simplified to give:

$$\mathbf{X}_x = -\frac{dU}{dx} \mathbf{i} \quad (3.43)$$

where \mathbf{i} is simply a unit vector, i.e., a vector with the amount of 1, and an arrow, directing toward the x -coordinate. For dU , any appropriate energy state function should be substituted as shown in Sect. 3.1.2. Let us consider for example the force acting on a charge (q) in an electric field in the x -direction ($\mathbf{E} = -d\psi/dx \cdot \mathbf{i}$). Substituting for dU in Eq. 3.43, the expression, according to Eq. 3.17, gives:

$$\mathbf{X}_q = -\frac{dU}{dx} \mathbf{i} = -\frac{d\psi}{dx} \mathbf{i} q = \mathbf{E} q \quad (3.44)$$

This assumes that there is no other gradient in the x -direction, i.e., that neither p , nor T , nor μ are functions of x . The introduction of the field strength (\mathbf{E}) is in accordance with the definition given in Eq. 2.45. Equation 3.44 is identical with Eq. 2.46, which was derived by other approaches.

Equation 3.44 cannot be applied to practical calculations of ion transport because only the transport of an electrical charge is considered. In contrast to the movement of an electron, the transport of ions always means that there is an additional change in concentration. For transport of a noncharged substance (i), the negative gradient of its chemical potential (μ_i) is the driving force:

$$\mathbf{X}_i = -\text{grad } \mu_i \quad (3.45)$$

Ions, in contrast, are driven by the gradient of the electrochemical potential ($\tilde{\mu}$), which according to Eq. 3.41 includes the electric potential. Applying the differential operator, the electric potential (ψ) transforms into an electric field strength (\mathbf{E}):

$$\mathbf{X}_i = -\text{grad } \tilde{\mu}_i = -(\text{grad } \mu_i - z_i \mathbf{E}) \quad (3.46)$$

There are many kinds of movement in biological systems that are calculated by biophysical approaches. Their range covers electron transfer, structural changes of molecules, chemical reactions, fluxes of molecules and ions, streaming of liquids in the body, and finally, mechanical movements of limbs and whole bodies. Fluxes occupy a central position in the study of movements in biology, and therefore, these will be considered here for the sake of simplicity, as a sort of generalized movement.

The flux (\mathbf{J}_i) is defined as the amount of a substance (i) that passes perpendicularly through a unit of surface per unit of time. This definition shows that the flux in general is a vector. Often the flux through a membrane is considered, the direction of which is always predicted.

The relation between flux \mathbf{J}_i of component i , and its velocity \mathbf{v}_i is:

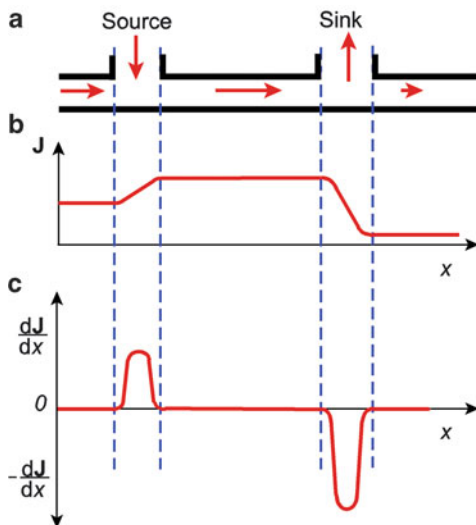
$$\mathbf{J}_i = c_i \mathbf{v}_i \quad (3.47)$$

A system which is traversed by a flux, and where no substance is added or removed to this flow, is called a *conservative* one. In this case the following conditions apply:

$$\text{div } \mathbf{J}_i = 0 \quad (3.48)$$

The differential operator div (“divergence”) can be replaced by the Nabla operator (∇), as explained in Sect. 2.2.1. In contrast to the operator “grad,” which, applied to a scalar results in a vector, the “div” operator is to be applied to a vector, producing a scalar. The Nabla operator (∇) is applicable to differentiate vectors as well as scalars.

Fig. 3.2 (a) graphic representation of a linear flow with a source, and a sink; (b) the flux (J) as a function of x ; (c) the change of the flow (dJ/dx) as a function of x



If $\text{div} \mathbf{J}_i > 0$ in such a system which is being traversed by a flux, it indicates that there is a source that is adding substance to the flux. However, if $\text{div} \mathbf{J}_i < 0$, then there will be a removal of some of the substance from the flux like a sink. Figure 3.2 illustrates this situation using an example where the flow is simply in the x -direction.

This formalism is applied, describing for example fluxes through biological tissue. For the transport of ions, this system is conservative, because no accumulation or depletion of ions occurs. In contrast, for the transport of oxygen the condition $\text{div} \mathbf{J}_o < 0$ holds, because the tissue uses oxygen for its own respiration.

If a constant force acts on a body, then the latter accelerates. However, as the velocity of this body increases, friction is likely to increase too. When both the driving force and the frictional force become the same amount, then the body will move with a constant velocity. This is a special case of a stationary state, the so-called *stationary motion*.

It is a fundamental experience in physics that the particular relation between force and movement characterizes regions with different qualities. If, for example, a comparatively minor force acts on a body, then the body will attain a velocity that is simply proportional to the force; this is a linear force-movement relationship and therefore the linear approaches of irreversible thermodynamics are applicable. If the same body is more forcibly moved, then the frictional force will increase in a stronger way, the frictional coefficient is no longer constant, and a nonlinear force-movement approach is necessary. Therefore, nonlinear thermodynamics, far from equilibrium, must be applied including some new qualitative properties.

This concept, illustrated here by a mechanical example, has a more general application. The linear approach can be formulated in a generalized form as an equation of motion in the following way:

$$\mathbf{J}_i = L_i \mathbf{X}_i \quad (3.49)$$

where the coefficient L_i is a kind of *generalized conductance*. In the same way, the following equation can be written:

$$\mathbf{X}_i = R_i \mathbf{J}_i \quad (3.50)$$

In this case a *resistance factor* is applied: $R_i = 1/L_i$. Ohm's law is a special form of this general expression:

$$U = RI \quad (3.51)$$

where U in this case is the electromotive force, depending on the potential gradient ($\text{grad}\psi$), and I is the electric current.

In a similar way force and velocity can be related in a linear way. For this it is necessary to introduce a mobility factor (ω), and its corresponding *coefficient of friction* (f):

$$\mathbf{v} = \omega \mathbf{X} = \frac{\mathbf{X}}{f} \quad (3.52)$$

The flux equation (3.47) can now be written as follows:

$$\mathbf{J}_i = c_i \omega_i \mathbf{X}_i \quad (3.53)$$

We introduced the Stoke's equation (2.34) for the discussion of molecular movement in Sect. 2.1.6. This is also a particular expression of this linear approach. It indicates the frictional force (\mathbf{F}) of a sphere with radius (r), moving in a liquid with viscosity (η) at velocity (\mathbf{v}). Under conditions of stationary motion, this friction force is equal to the driving force:

$$\mathbf{F} = 6\pi r \eta \mathbf{v} \quad (3.54)$$

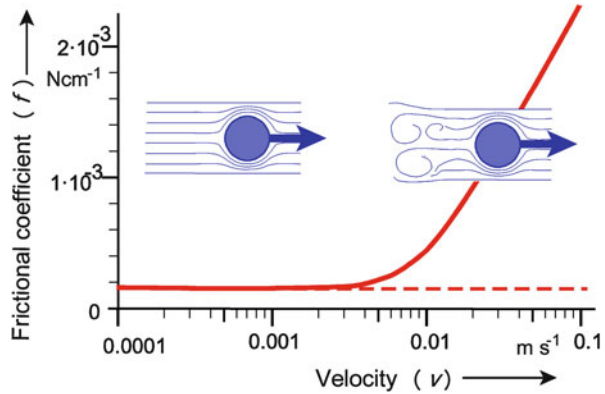
In view of Eq. 3.52, the following analogy is obvious:

$$f = \frac{1}{\omega} = 6\pi r \eta \quad (3.55)$$

The definitive mechanical behavior of bodies moving in water will be discussed in detail in Sect. 3.7. Here it is only being used as an example for the consideration of general principles.

Figure 3.3 shows the frictional coefficient (f) for a sphere with a diameter of 1 cm in water as a function of its velocity (\mathbf{v}). The frictional coefficient only remains constant up to a velocity of about 1 mm s^{-1} , corresponding to Eq. 3.55. As the velocity increases above this, the frictional factor deviates, at first slightly, and then greatly from the horizontal line. This illustrates the transition of the system from linear to nonlinear behavior. This is the transition to far-equilibrium conditions, where nonlinear thermodynamic approaches must be applied.

Fig. 3.3 The coefficient of friction ($f = F/v$) of a sphere with a diameter of 0.01 m in water as a function of its velocity (v)



Let us explain the genesis of such a nonlinear relation by a simple example. Let a phenomenological coefficient be a sum of a constant term (L_i), and a variable term L'_i , which is proportional to the flux (J_i). In this case, the linear approach of Eq. 3.49 transforms into:

$$J_i = (L_i + L'_i J_i) X_i \quad (3.56)$$

Solving this equation for J_i , gives:

$$J_i = \frac{L_i X_i}{1 - L'_i X_i} \quad (3.57)$$

Hence we have a nonlinear function $J_i(X_i)$.

The qualitative consequences of these nonlinear approaches will be considered further in the next section. Let us at present remain in the field of linear thermodynamics.

One of the fundamental statements of linear thermodynamics concerns the coupling of forces and movements in a system. If different kinds of motions and forces occur simultaneously in a system, they will influence each other.

Let us consider again the flux (J_i) as a generalized kind of motion. Nonequilibrium thermodynamics in its scope of linear approaches allows us to formulate a set of phenomenological equations forming a flux matrix. This is the mathematical expression of the general statement whereas all fluxes in a system in principle are coupled with each other.

The simple Eq. 3.49, therefore, will be expanded to the following set of equations:

$$\begin{aligned} J_1 &= L_{11}X_1 + L_{12}X_2 + L_{13}X_3 + \dots + L_{1n}X_n \\ J_2 &= L_{21}X_1 + L_{22}X_2 + L_{23}X_3 + \dots + L_{2n}X_n \\ J_3 &= L_{31}X_1 + L_{32}X_2 + L_{33}X_3 + \dots + L_{3n}X_n \\ &\dots \\ J_n &= L_{n1}X_1 + L_{n2}X_2 + L_{n3}X_3 + \dots + L_{nn}X_n \end{aligned} \quad (3.58)$$

In these equations the vector notation of fluxes and forces is still retained (bold letters), regardless of whether they may in fact, in rare cases, be scalars, which we will discuss in the following.

The parameters L_{mn} are phenomenological coefficients, also called *coupling coefficients*, *cross coefficients*, or *Onsager coefficients* (this approach was introduced first by Lars Onsager in 1931). In reality, this general set of equations may be reduced, because a flux \mathbf{J}_m is coupled with a force \mathbf{X}_m only when $L_{mn} \neq 0$.

Equation 3.58 shows that n forces with their corresponding fluxes require a set of equations with n^2 coupling coefficients. Onsager, however, was able to show that this matrix is symmetric. This means that near equilibrium the following relation holds:

$$L_{mn} = L_{nm} \quad \text{for : } n \neq m \quad (3.59)$$

This is *Onsager's law on the reciprocal relation*. It leads to a significant reduction of the coefficients in the matrix from n^2 , down to $n(n+1)/2$.

Directly linked pairs of forces and fluxes, as for example \mathbf{J}_1 and \mathbf{X}_1 , \mathbf{J}_2 and \mathbf{X}_2 , ..., \mathbf{J}_n and \mathbf{X}_n , are called *conjugated*. The coefficients (L_{nn}), linking these pairs, are always positive. If two fluxes really are coupled, the following condition must hold:

$$L_{mm} \cdot L_{nn} \geq L_{mn}^2 \quad (3.60)$$

From this, a *degree of coupling* (q_{mn}) can be defined:

$$q_{mn} = \frac{L_{mn}}{\sqrt{L_{mm}L_{nn}}} \quad (3.61)$$

This degree of coupling can vary as follows: $1 \geq q_{mn} \geq 0$. When $q_{mn} = 0$, the fluxes are completely independent of each other, when $q_{mn} = 1$, there is maximal coupling.

As mentioned earlier in Eq. 3.58, all fluxes and forces are printed in bold letters as vector parameters. At the same time we mentioned that we will consider here fluxes in a very general sense, symbolizing all kinds of motion. This means that not only true fluxes of matter which really are vectors going in a particular direction, but for example, also chemical reactions will be considered. The flux, as mentioned in this equation, therefore, can also mean a production of a substance, or the removal of it, by a chemical reaction. In this case however the flux does not remain a vector, but becomes a scalar. How can we include these scalar fluxes into a matrix together with vectors and not violate mathematical rules?

Let us consider a simple set of flux equations, including a transport of a substance (\mathbf{J}_i) and a chemical reaction, the rate of which we will denote by the scalar flux J_R . Formal mathematics appears to allow only the following possibility:

$$\begin{aligned} J_R &= L_{RR}X_R + \mathbf{L}_{Ri}\mathbf{X}_i \\ \mathbf{J}_i &= \mathbf{L}_{iR}X_R + L_{ii}\mathbf{X}_i \end{aligned} \quad (3.62)$$

In the first equation, the product of two vectors ($\mathbf{L}_{Ri}\mathbf{X}_i$) gives a scalar, as well as the product of scalars ($L_{RR}X_R$), and so, this equation is a sum of scalars resulting in a scalar flux (J_R). In contrast, in the second equation all the terms of the sum are vectors. In both equations therefore, the mathematical requirement for homogeneity has been satisfied.

But: what is the meaning of a vectorial coupling coefficient? Introducing this parameter, we declared it as a sort of conductivity. What does a conductivity vector mean? In fact, in so-called *anisotropic systems* vectorial coefficients can appear. In an isotropic system, for example, in an aqueous solution, the mobility of an ion in all directions is equal. The parameter L , therefore is a scalar. Considering, however, the same ion in a pore of a membrane, its movement is possible only in a predetermined direction, and its conductivity consequently becomes a vector.

These circumstances are considered in the so-called *Curie–Prigogine principle*. It states that the direct coupling of scalar and vectorial fluxes is possible only in anisotropic systems. This principle, for example, is important in biophysical considerations of active ion transport. In this case a hydrolysis of ATP is coupled to the transport of an ion against an electrochemical gradient.

This example already indicates that the concept of coupled movements is not limited to mechanical frictional interactions. This in fact is the case in some electro-osmotic processes as described in Sect. 2.3.5 (Fig. 2.45). Furthermore, in Sect. 3.3.1 we will use this approach to introduce Staverman's reflection coefficient, governing the coupling of the osmotic fluxes of water and solute (Eq. 3.150). In general, however, the concept of Onsager coefficients is applied in various networks of biochemical processes.

Further Reading

Katchalsky and Curran 1965; Prigogine 1967; Schnakenberg 1981; Kjelstrup and Bedeaux 2008.

3.1.4 Entropy, Stability, and Stationary States

The second principle of thermodynamics states that an isolated system moves spontaneously towards a maximum in its entropy. When this state is achieved, then the system is in thermodynamic equilibrium. In the same way, the decrease of the free energy down to a minimum can be considered as the way towards the equilibrium in the sense of the second principle.

Any movement as a result of energy transformation leads to an increase in the entropy of the system or its environment. The term *entropy production* ($\sigma = dS/dt$) has been introduced to characterize this process. The entropy production is always positive, but can approach zero. The condition: $\sigma = 0$ would mean an idealized reversible process. Thermodynamically, a process is defined as being reversible if it can be repeated an arbitrary number of times without requiring the supply of additional energy.

To prevent misunderstanding, the different meanings of the term “reversible” in physics, chemistry, and biology must be pointed out. In physics, the term “reversible” is used according to the thermodynamic definition, i.e., connected with the above-mentioned condition: $\sigma = 0$. When a chemist speaks about a “reversible reaction,” or a “reversible electrode,” he only means processes that in principle could run in both directions, independently of the required energy. Finally, the biologist states that a change in a biological system is “reversible” when it is able to reverse an induced change so that no irreparable damage is caused (for example: reversible inhibition of metabolic processes).

Let us consider the total entropy balance of a system. In closed systems, as a result of energy transformations, only an entropy increase is possible up to the point where the thermodynamic equilibrium is established without any further entropy-producing processes. In open systems, however, which are able to exchange not only energy but additionally matter, the entropy may change in both directions. An entropy decrease can occur if substances with low entropy content are incorporated, in exchange for entropy-rich substances that are being extruded. To characterize this process, an entropy flux (\mathbf{J}_S) is formulated which penetrates the whole system. Hence, the total entropy balance of the system can be written as follows:

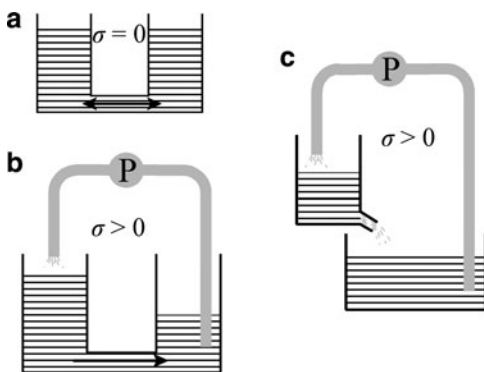
$$\frac{\partial S}{\partial t} = -\nabla \mathbf{J}_S + \sigma \quad (3.63)$$

The overall change of the entropy of an open system ($\partial S/\partial t$), therefore, results as the sum of the always positive entropy production (σ), and the divergence of the entropy flux ($\nabla \mathbf{J}_S \equiv \text{div } \mathbf{J}_S$), penetrating the system. In reference to the definition in Sect. 3.1.3 (see Fig. 3.2), the system in fact is not conservative in relation to this entropy flux. Depending on the relation of the two terms in the sum (Eq. 3.63), the total entropy change ($\partial S/\partial t$) can become positive as well as negative. The control of the term $\nabla \mathbf{J}_S$ can be considered as the work of a Maxwell demon, as described in Sect. 2.1.2 (Fig. 2.2).

For further considerations it may be useful to introduce a thermodynamically based classification of the various kinds of stationary states. We define a *stationary state* as a state where the structure and parameters are time independent. The reasons, leading to this quality can be quite different. The water level of a lake, for example, can be time independent, i.e., constant, either because there is no inflow into the lake, and no outflow, or because inflow and outflow are equal. These two kinds of stationary states can be distinguished by their entropy production. In the first case no energy is required to maintain this state, therefore there is no entropy production, the system is in *thermodynamic equilibrium* ($\sigma \stackrel{!}{=} 0$). In contrast, the lake with exactly the same in- and outflow is in *steady state*. This is a stationary state with entropy production ($\sigma > 0$). The thermodynamic definition of the steady state is the only possible one. It seems important to emphasize that a steady state cannot be defined by its kinetic properties.

Let us illustrate this statement with an example: Using radionuclides it is possible to demonstrate that human erythrocytes exchange chloride as well as

Fig. 3.4 Stationary states in hydraulic models: (a) thermodynamic equilibrium ($\sigma = 0$); (b) a steady-state system ($\sigma > 0$), which becomes an equilibrium (A), if the pump (P) is stopped; (c) a steady-state system ($\sigma > 0$), where stopping the pump (P) would lead to a complete outflow of the liquid from the upper vessel

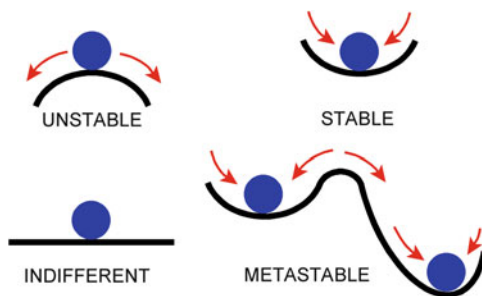


potassium ions with their environment. With this method, it is possible to measure directly the corresponding exchange rates. This kinetic method of analysis may give the impression that both ions, Cl^- , as well as K^+ , are in a steady state because in both cases the unidirectional fluxes, outside \rightarrow in and inside \rightarrow out, are equal. This, however, is an incorrect conclusion. The chloride ions in fact are distributed passively between the external medium and the cytoplasm, according to their thermodynamic equilibrium. The observed exchange of radioactive chloride results from their self-diffusion by thermal motion. This process is not accomplished by entropy production because no energy is converted. It is like a stochastic exchange of water molecules between the two vessels in Fig. 3.4a. In contrast to this, potassium is pumped actively into the cell by use of metabolic energy against its electrochemical gradient, and diffuses passively back, driven by this gradient (Fig. 3.4b). Both, the active transport as well as the passive diffusion, are true fluxes in the thermodynamic sense, producing entropy. Potassium, therefore, in contrast to chloride, really, is in a stationary state. This example indicates that the above-described kinetic experiment of compartment analyses is unsuitable to distinguish between an equilibrium state and a steady state. We will come back to this problem later in detail (Sects. 3.4.1 and 5.1.1).

As we will see later (Sect. 3.3.3), the steady state of potassium can be transformed into an equilibrium state if the pump is inhibited. In this case, a Donnan equilibrium will be established which means an equilibration of all electrochemical potentials. The steady state of sodium and potassium in the cell therefore resembles case b in Fig. 3.4, passing into the equilibrium state (Fig. 3.4a), when the active transport is inhibited. In contrast to this, various substances do not show the possibility of an equilibrium distribution. If the influx is stopped they disappear completely (Fig. 3.4c).

An important property of all stationary states is their kind of stability. Let us illustrate this by a mechanical example of a sphere on a surface (Fig. 3.5). The requirement for a stationary state, in this case this simply means an equilibrium, is the sphere coming to rest on any small, but horizontal part of the surface. In the case of an *indifferent* state, the surface is horizontal in general. In this case the energy of the sphere will not be changed by alteration of its position. In the case of a *stable*

Fig. 3.5 Various kinds of stability conditions



state, every change of the position leads to an increase of the energy of the sphere, and generates a force driving the sphere back to its original state. In contrast an *unstable state* is characterized by a situation where even small changes of the position release forces that cause the system to be deflected even more. Additionally sometimes so-called metastable states are considered. As *metastable*, a stable state can be considered which is delimited from another one by a small barrier which can easily be overcome.

In the mechanical examples of Fig. 3.5 the shape of the surface automatically indicates the function of its potential energy. In general, however, these surfaces have to be replaced by true functions of the free energy like those of Figs. 2.5 or 2.26.

Figure 3.6 indicates all possible kinds of stationary states. First of all the presence, or the absence of entropy production indicates whether the given stationary state is a *thermodynamic equilibrium* ($\sigma = 0$), or whether it is a *steady state* ($\sigma > 0$). In the case of thermodynamic equilibrium one must distinguish between global and local equilibria. In the case of a *global equilibrium*, the function of free energy indicates only one minimum. This means that no alteration, however strong it may be, can bring the system into another equilibrium state. An example is the equilibrium distribution of an ion between the cell and its environment. In contrast, in the case of the *local equilibrium*, the energetic function indicates two or more minima which are separated by more or less high energy barriers. As an example, isotherms of biochemical reactions can be considered, as illustrated in Fig. 2.5. The stability of such local equilibria is determined by the energy barrier between them. If this barrier is so low that thermal fluctuations can lead to quick transitions, the state is called *metastable*. This for example is the typical situation for enzyme-substrate complexes. For the schemes in Fig. 3.6 it must be considered that in reality G is not simply a function of a single reaction coordinate (x), but rather a hyperplane in n -dimensional space.

In contrast to equilibrium states, the stability conditions of the steady states are not simply dictated by their energy functions, but by their kind of entropy production. For this we must learn something more about this parameter.

In concordance with its definition the entropy production has the measure: $\text{J K}^{-1} \text{s}^{-1}$. In some cases additionally this parameter is related to the mass in kg or to the molar mass.

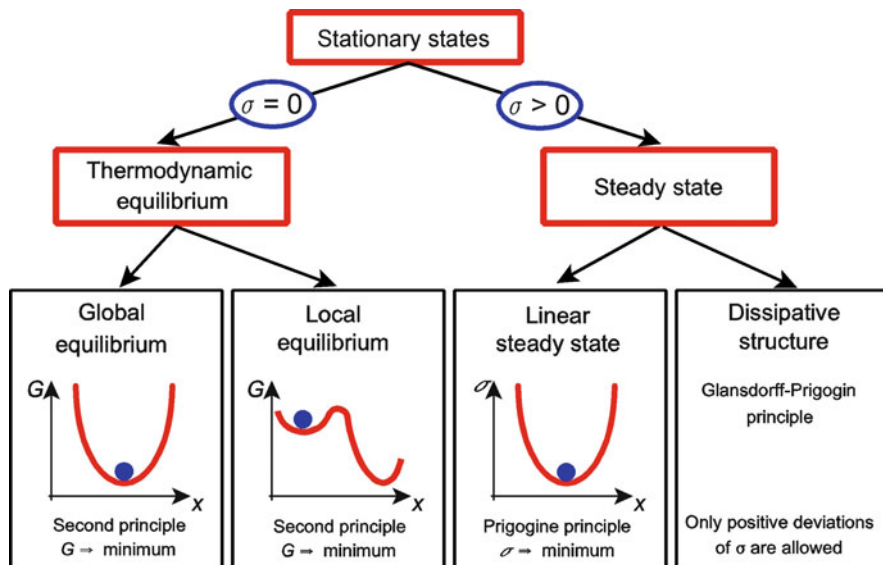


Fig. 3.6 Possible types of stationary states and conditions of their stability

Multiplying entropy production by temperature, one gets *Rayleigh's dissipation function* (Φ), which can be calculated from fluxes and forces as follows:

$$\Phi = \sigma T = \sum_{i=1}^m \mathbf{J}_i \mathbf{X}_i \quad (3.64)$$

This equation holds for the region of linear, as well as of nonlinear flux-force relations. Particularly for the linear region, Ilya Prigogine was able to show that systems tend to develop towards a reduced entropy production. This is the *Prigogine principle of minimal entropy production*. Systems, which are not far from thermodynamic equilibrium, and which are kept in imbalance by continuously acting forces consequently may move towards a steady state, the stability of which is included in this criterion (see Fig. 3.6).

The dissipation function represents the specific heat generation of a system. In living organisms it reflects the metabolic rate which in the case of aerobic organisms is equivalent to the oxygen consumption. Therefore, it can be measured directly by calorimetry or even calculated from parameters of respiration (see Sect. 3.8).

Some authors discuss this metabolic rate in context with this Prigogine principle of minimal entropy production. Figure 3.7 shows for example a mean scheme of the characteristic situation in mice. While the *total* heat production (\dot{Q}) increases in accordance with their age and mass (m), the *specific* heat production, that is the heat production relative to the body weight, i.e., the dissipation function (Φ), reaches a maximum and then decreases. It seems therefore that an animal in the state of

development increases entropy production, whereas an adult organism tends to arrive at a minimum. Disturbances in life lead to various deflections from this curve. If an organism for example is injured, if it is stressed by some environmental conditions, or in the case of tumor growth, the corresponding disturbances of the metabolism lead to a temporary increase in the slope of the curve.

It is still controversial, however, as to whether the Prigogine principle can be applied to large systems that include a great number of different subsystems, particularly those which are far from equilibrium. If a system deviates from the region of linear approaches, then the Prigogine principle is no longer valid. In contrast to steady states in the scope of linear thermodynamic approaches which are always stable and do not show any kind of metastability, systems in the region of nonlinear approaches show more complicated behavior.

Considering nonlinear systems we must first generalize the term “stationary.” We already mentioned in context with Stoke’s law (Eq. 3.54) that the so-called stationary movement is a movement with constant velocity where the frictional force is equal to the driving force. This sort of movement is also a kind of stationary, i.e., time-independent state. The term “stationary state” can also be applied to states that are not at rest, but show repetitive periodic movements in a stationary manner. For example, cardiac function can be called “stationary,” if the frequency and amplitude of the heart beat does not change during the period of observation.

In the case of a linear steady state, fluctuations of the system parameters produce only positive deviations of entropy production, bringing the system back to the stationary state, which is therefore stable in any case. In contrast to this, in the region of nonlinear approaches, far from equilibrium, fluctuations of a stationary state can lead also to negative deviations of entropy production, leading to a destabilization of the system. The system may jump from one into another

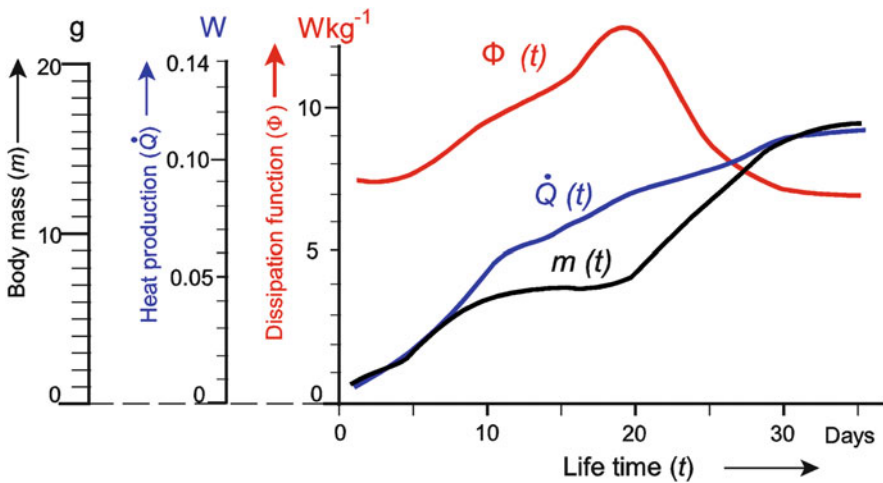


Fig. 3.7 Dissipation function (Φ), heat production (\dot{Q}), and mass (m) of mice as a function of their life span from the time of birth, up to the 35th day of life (Data from Prat and Roberge 1960)

stationary state. The steady states in the region of nonlinear approaches therefore are mostly metastable (see Fig. 3.5). Their stability condition requires the occurrence of only positive deviation of entropy production. This is the so-called *Glansdorff–Prigogine principle*.

The transition from linear, to nonlinear approaches of thermodynamics is not only associated with an increase in the coefficient of friction, or with the occurrence of several stationary states, but also with the spontaneous development of so-called *dissipative structures*. The general concept of the diversity of structures has already been discussed in Sect. 2.1.3. We postulated that there are basically, two different types of structures: equilibrium structures ($\sigma = 0$), and dissipative structures ($\sigma > 0$). In contrast to equilibrium structures which are always structures in space, dissipative structures can also be structures in time, or in a space–time continuum.

In order to illustrate this, Fig. 3.3 shows the function of the frictional coefficient $f(\mathbf{v})$ of a sphere, and the pattern of flow close to it. At a low velocity a laminar flow around the sphere occurs, changing into a turbulent flow, when the nonlinear region of this function is attained. When the flux-force relation changes from the laminar to the nonlaminar region, then the unstructured laminar flow becomes unstable and vortices appear which in the terminology of thermodynamics, are dissipative structures (for details of streaming behavior, see Sect. 3.7.1).

There exists an extensive literature on the theory of dissipative structures and on their occurrence in nature. Most of these dissipative structures are periodic structures in space, such as cloud patterns, flow phenomena in liquids with an applied temperature gradient, so-called Benard cells, plasma waves in electron beam tubes, etc. In addition, there are many time structures, including for example all kinds of sound production, from the electronic organ to vibrating strings, and to wind instruments.

In biological systems, in spite of many speculations dissipative structures in space have not been unequivocally demonstrated. Conversely, time patterns showing this property have been found to occur quite often. Examples of this include oscillations in metabolism, periodic changes in the membrane potential of certain cells, for example in the cells of the cardiac pacemaker in the sino-auricular node, and finally more complex, oscillatory movements in ecosystems. Sometimes, such oscillations in local concentrations of certain substances become “frozen,” so that structures arise which do not require entropy-producing processes to sustain them, but which originally had been built as dissipative structures. One can conceive, for example, the genesis of the first viable biomacromolecule occurring in this way. Similar processes could form the basis of morphogenesis (for further detail, see Sects. 5.2.3, 5.2.4, and 5.2.5).

Let us come now to a further aspect of stationary states. If a system is considered as stationary, this is always in relation to a defined period of time. An organism could be said to be in a stationary state for a number of hours, or for days. That is, its volume, the content of certain substances, its shape, temperature, etc., are constant within defined tolerance for this period of time. Nevertheless, the organism in fact is ageing or growing. If a longer time period is considered, then the changes in these parameters will exceed the defined tolerance limits.

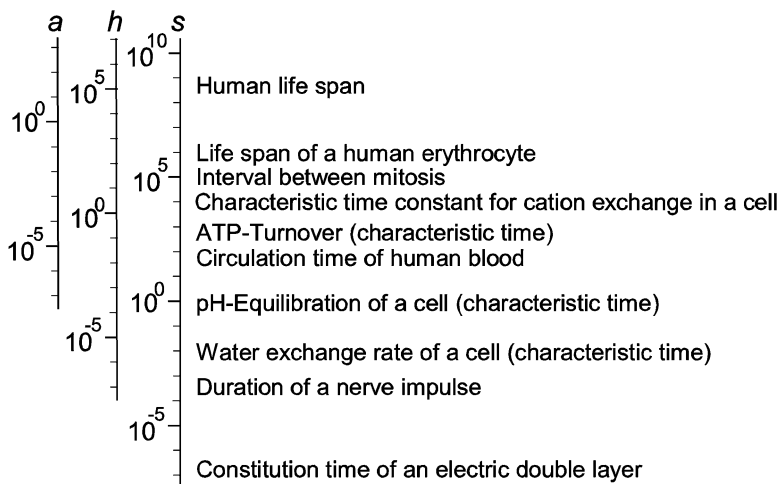


Fig. 3.8 Characteristic time constants (s – seconds, h – hours, a – years) of various processes on a logarithmic scale, to illustrate the time hierarchy in biological systems

A biological system consists of subsystems, each of which is associated with a different time constant. For example, the rate of aging in man is slow, in comparison with the rate of mitosis of a hemopoietic cell. Thus, the conditions in bone marrow can be regarded as stationary during the course of several cycles of mitosis in spite of the ageing processes affecting the organism as a whole. In vivo, the life span of a human erythrocyte is about 100 days. The ionic composition of juvenile erythrocytes differs somewhat from that of mature cells. If one is interested in the ionic regulation of these cells, then, because of the temporal characteristics of such processes, experiments lasting a few hours will be sufficient to provide the required information. Within such short time periods, the ionic concentration can be regarded as stationary. Figure 3.8 shows some characteristic times on a logarithmic scale that will serve to extend this list of examples.

These considerations hint at the existence of a *time hierarchy* of stationary states, related to their time constants which range over several orders of magnitude. The kinetic consequences of this circumstance will be considered in Sect. 3.2.5. The following concept, however, is important for our further thermodynamic considerations:

The living organism as a whole when considered within a limited period of time, is in a stationary state with entropy production, i.e., in a steady state. It is made up of a great number of subsystems which are ordered in a defined time hierarchy. The steady state of the system as a whole does not imply that all of the subsystems are also in a steady state. A large proportion of them, particularly those with a short time constant, are in thermodynamic equilibrium. If the system as a whole changes its parameters slowly, then these subsystems are capable of following such changes quickly, so that they almost completely adapt within their characteristic time, and thus are always in a stationary state. This is sometimes called a *quasi-stationary*, or *quasi-equilibrium* state.

The following example will illustrate this: The water content of a tissue depends on the ionic composition of its cells. Sodium and potassium ions are being actively transported against passive fluxes, giving rise to a steady state. In this way the active transport and the corresponding passive fluxes regulate the osmotic properties of the cells. The characteristic time of the water flux is much shorter than that of the cations (see Fig. 3.8). As a result, the water in the interior of the cells is always in osmotic equilibrium with the surrounding medium. In Sect. 3.2.3 we will discuss this example in detail in context with the Donnan-osmotic quasi-equilibrium.

Further Reading

Feistel and Ebeling 2011; Glansdorff and Prigogine 1985; Haken 2010; Zotin 1990.

3.1.5 Stochastic Resonance and Noise-Enhanced Processes

Noise-enhanced processes have been widely observed throughout nature. It is now broadly applied to describe any phenomenon where the presence of noise is better for output signal quality than its absence. It spans the field from diverse climate models, social and financial problems, via technical systems like electronic circuits, SQUIDS and lasers, to various biological systems such as ecological and neural models, ion channels, networks of biochemical reactions, etc.

Stochastic resonance (SR) as a particular property of dynamic systems is one of the mechanisms that makes a nonlinearity less detrimental to the noise of a signal. Quite unexpectedly, in this case noise can even lead to the formation of more regular temporal and spatial structures, and cause recognition and amplification of weak signals accompanied by growth of their signal-to-noise ratio (SNR). In fact, noise in this case can play a constructive or beneficial role in nonlinear dynamics.

The term stochastic resonance was first used in the context of noise-enhanced signal processing in 1980 to explain the almost periodic recurrence of the primary cycle of ice ages every 1,00,000 years. The application of SR in biology started in the early 1990s wherein SR was discovered in sensory neurons that had been subjected to external noise.

SR can be considered as a phenomenon occurring in nonlinear systems far from equilibrium, where the presence of a particular quantity of noise, superimposing a signal input is better for its output quality than its absence. The word “resonance” in this term on one hand was used because the plot of output performance, resp. the signal-to-noise ratio – against the noise intensity resembles a bell-shaped function with a single maximum; a similar appearance to frequency-dependent systems for some resonant frequency. On the other hand a kind of periodic resonance between properties of the system and some noise inherent frequencies occurs.

The basic mechanism of SR is schematically illustrated in Fig. 3.9. This example stands for a kind of excitation model of a sensory system. An external signal (red line) does not attain the threshold value (green line), therefore has no influence on the system. If however, the superimposed noise (blue) becomes large enough to arrive at the threshold value in cases of the maxima of the signal intensity, the

periodicity of the signal will be reflected in the system (black bars). In this case the noise becomes helpful for recognition of the signal by the system. If, however, the noise intensity increases further, the character of the signal will be fully masked and the SNR again tends to zero. This bell-shaped function of SNR in dependence of the noise intensity is depicted in the curve below.

This particular system of SR in the mechanism of nerve excitation is investigated quantitatively using the basic equations of the Hodgkin–Huxley model (Eqs. 3.195–3.200 in Sect. 3.4.4). It helps to understand the role of SR in many animal sensory systems.

To explain mechanisms of stochastic resonance in chemical reactions and various transport processes, the double-well potential profile of potential energy as a function of the reaction coordinate must be mentioned (cf. Fig. 2.5). In Sect. 2.1.5 we considered this system to deviate the Arrhenius equation. We mentioned there that even processes of diffusion and membrane transport in principle are based on the same kind of multiwell potential profiles.

In contrast to the equilibrium considerations of Sect. 2.1.5, let us now focus on the transition process between the minima of this function, including fluctuations induced by internal and external noise. For illustration one can apply a more mechanistic view, considering the state of a system like a position of a particle in a mechanical well (Fig. 3.10). The motion of such a particle qualitatively can be characterized by two time scales: The first defines relaxations of fluctuations in the linear regime around the stable fixed points (*intrawell dynamics*), the second concerns the mean time of barrier crossings (*global dynamics*), as a result of a nonlinear process.

The transformation of an external signal into an equivalent system behavior can be considered as a periodic transition from one state to another, or in a mechanistic view, as the movement of the particle from one well into the neighboring well. This is only possible in a reasonable time if the external signal transforms the energy function in such a way that the barrier between the two wells becomes small enough. Otherwise (Fig. 3.10a), this transition does not occur, in other words, the external signal will not be reflected by system behavior. In the case of an additional noise (Fig. 3.10b), however, this becomes possible. In the same way as depicted in Fig. 3.9, the external signal at optimal noise intensity can be received by the system.

In this context an interesting case of biological optimization should be noted. If a particular noise intensity in some sensory systems is necessary to achieve optimal SR, and if this noise has its origin in the stochastic properties of ion channels, optimization can occur as a cooperation of several channels. The noise of a single channel can be reduced by averaging the effect of an assembly of other channels. Conversely, the SR effect vanishes if the assembly becomes too large. Apparently, an optimal size of cooperating channel assemblies exists in such systems.

Thus, stochastic resonance allows us to realize the input of a signal at an optimal noise level by switching events in a bistable system. Similarly, noise can enhance oscillations in a dynamic system even without external periodic signals. This phenomenon is called *coherence resonance*.

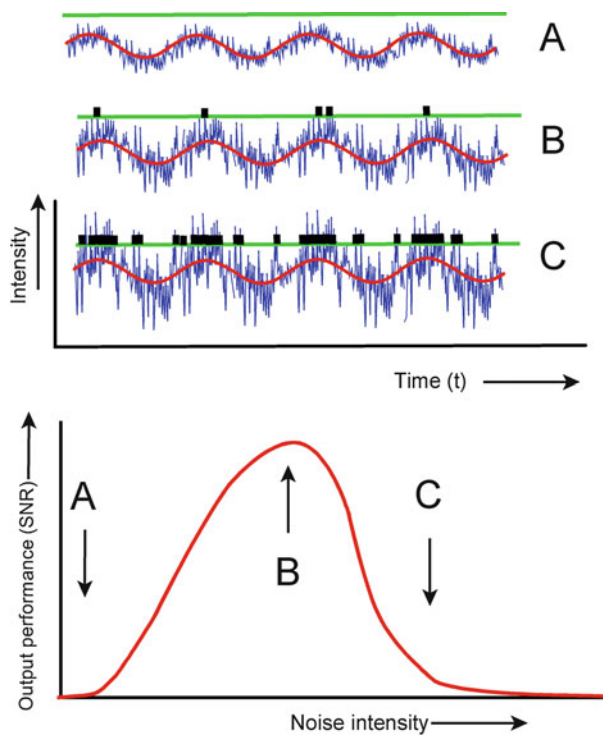


Fig. 3.9 A scheme explaining the mechanism of stochastic resonance in a system with a definite threshold like a neuron. *Above* – A, B, C: a subliminal sinusoidal signal (red line), superimposed by noise (blue) with increasing intensity. A – weak noise, not attaining the threshold intensity (green line). B – optimal noise intensity, leading to excitations (black bars), the time course of which reflect the periodic character of the signal. C – extensive noise intensity, smearing the character of the signal. *Below*: The output performance as a bell-shaped function of the noise intensity

This situation is modeled for example in nerve cells using a simplified version of the Hodgkin–Huxley equations, which accounts for the essentials of regenerative firing mechanisms. In fact, neurons are permanently affected by various kinds of noise, such as for example the fluctuating opening and closing of ion channels, the noise of presynaptic currents, fluctuations of conductivities, etc. As the neuron is unable to fire during the recovery state, it becomes excited at a particular time scale by sufficient intense noise. In this way an oscillatory behavior occurs. Since the excitation time depends strongly on the noise, an optimal noise level exists like in SR systems. However, unlike stochastic resonance, in the case of coherence resonance no particular input signal exists.

There are also several other mechanisms, distinct from SR, where noise has a constructive role. An example is the so-called *Brownian ratchet*, a mechanism, which is used in technics, for example in some battery-less wristwatches that are wound up by random movement. In this case a stochastically oriented force

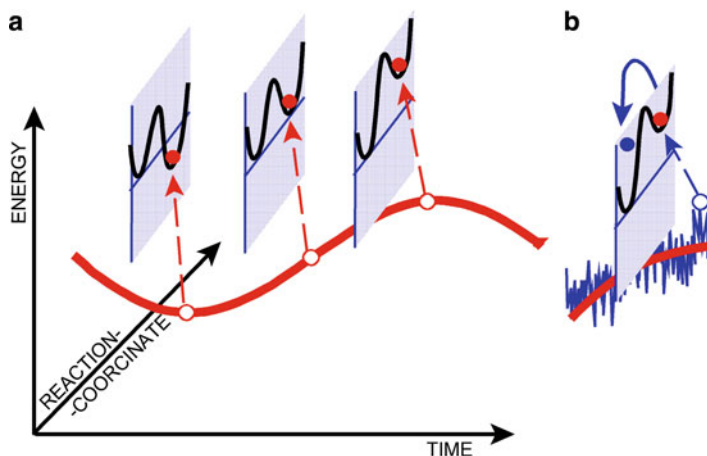


Fig. 3.10 Stochastic resonance in terms of the reaction coordinate. (a) A periodic signal (*red line*) modifies periodically one part (the backmost) of the double well potential of a reaction, but not sufficiently to lead the system (illustrated as the position of the *red particle*) in a reasonable time to transit into the neighboring (in front) well (A). (b) This however occurs under the influence of a sufficiently intense noise. The effective transition of the particle between these two states in this case may reflect the time scale of the periodic signal (*red line*) in the same way as in the example of Fig. 3.9

generates a directed movement. The physical basis of this phenomenon is an oscillating force, affecting an anisotropic system, where friction in one direction differs from that in the opposite.

This mechanism led to speculations that such a Brownian ratchet, like a Maxwell demon could be driven by thermal fluctuations, and in this way would form a perpetuum mobile of the second order (see Sect. 2.1.2). A Brownian particle was considered to interact with a ratchet exhibiting a sawtooth profile. In this case it could move forward, driven simply by thermal noise. This would represent a microscopic version of the winding mechanisms that are realized in wristwatches. In accordance with the second law of thermodynamics, however, it is impossible to produce oriented movement driven by thermal oscillations. Thus, as long as the pawl has the same temperature as the ratchet it is subjected to the thermal noise itself, i.e., its particular structure will be disturbed in molecular dimensions. In fact the self-winding mechanism of the wristwatch is driven not by thermal noise but by stochastically oriented acceleration, the energy consuming movement of the arm. Its mechanism is based not on a rectifying of thermal noise, but on an input of energy consuming stochastic force.

In biological systems various types of Brownian ratchets, as mechanisms rectifying stochastically oriented forces, are identified, on the molecular level as well as on the level of movement of cells and organisms. So for example, there exist complex patterns of directed molecular movement that are based on filaments and motor molecules, performing mechanical work on the nanometer scale driven by ATP hydrolysis. These molecular motors and motor particles bind to cytoskeletal

filaments and walk along these filaments in a directed fashion. This can be understood in terms of Brownian ratchets driven by nonequilibrium processes. The motor molecules can be considered as Brownian particles which can attain several internal states, and experiences a certain molecular force potential in relation to the cytoskeletal filament.

Further Reading

Anishchenko et al. 2006; Hänggi 2000; Lipowsky and Klumpp 2005; Mahmud et al. 2009; McDonnell and Abbott 2009.

3.1.6 Thermodynamic Basis of Biochemical Reactions

Chemical processes in biological systems are very complex and specific. In Sect. 2.1.5, we already mentioned the ability of enzymes to overcome energy maxima of activation energy. Some aspects of chemical thermodynamics are explained in Sect. 3.2.1. In Sect. 5.2.1, we will discuss the particularities of the kinetics of biochemical reactions. Now, some general thermodynamic aspects of chemical and biochemical reactions will be introduced to complete the approaches of equilibrium as well as nonequilibrium thermodynamics.

In the previous formulation of the Gibbs equation (Eqs. 3.17, 3.27, 3.28, 3.29), the chemical reaction was not explicitly enclosed. In fact, a chemical reaction can be analyzed by referring to the chemical potential of its components. The reaction:



can be considered as a replacement of the substances A and B by the substances C and D.

Considering this process for isobaric ($dp = 0$) and isothermic ($dT = 0$) conditions, where only concentration changes take place ($dn_i \neq 0$, $dq = 0$, $dl = 0$), the equation for Gibbs free energy (Eq. 3.29) reduces to:

$$dG = \sum_{i=1}^m \mu_i dn_i \quad (3.66)$$

The change in the Gibbs free energy ($\Delta_R G$) according to the definition (3.31) is then given by:

$$\Delta_R G = v_c \mu_c + v_d \mu_d - v_a \mu_a - v_b \mu_b \quad (3.67)$$

Substituting in this equation the expression for the chemical potential (Eq. 3.33), we obtain:

$$\Delta_R G^0 = v_c \mu_c^0 + v_d \mu_d^0 - v_a \mu_a^0 - v_b \mu_b^0 + RT(v_c \ln a_c + v_d \ln a_d - v_a \ln a_a - v_b \ln a_b) \quad (3.68)$$

Now, we can summarize these standard potential expressions and define a molar free standard reaction energy:

$$\Delta_R G^0 = v_c \mu_c^0 + v_d \mu_d^0 - v_a \mu_a^0 - v_b \mu_b^0 \quad (3.69)$$

Substituting this in Eq. 3.68 and combining the logarithmic terms, we obtain the *van't Hoff equation*:

$$\Delta_R G = \Delta_R G^0 + RT \ln \frac{a_c^{v_c} a_d^{v_d}}{a_a^{v_a} a_b^{v_b}} \quad (3.70)$$

In the case of thermodynamic equilibrium, $\Delta_R G = 0$. This allows to define the equilibrium constant (K_p) for reactions under isobaric conditions, whereas the symbols a_i^0 stand for activities of the system in equilibrium:

$$K_p \equiv \frac{a_c^{0v_c} a_d^{0v_d}}{a_a^{0v_a} a_b^{0v_b}} = e^{-\frac{\Delta_R G^0}{RT}} \quad (3.71)$$

Substituting this expression into Eq. 3.70, we obtain for the case of equilibrium ($\Delta_R G = 0$):

$$\Delta_R G^0 = -RT \ln K_p \quad (3.72)$$

The molar free standard Gibbs energy of a chemical reaction ($\Delta_R G^0$) can be calculated using the standard energies of formation ($\Delta_f G^0$) obtained from tables (see Eq. 3.31). In fact, we use here the property of Gibbs energy as being a state function, that is, a parameter which is independent of the way in which it is obtained (see Sect. 3.1.1).

If a reaction is not in equilibrium, then the direction which it will take can be determined by calculating $\Delta_R G$ (Eq. 3.70) from the given activities of the components. Spontaneously the reaction can run in the direction indicated in Eq. 3.65 only in the case where $\Delta_R G < 0$, resulting in a reduction of the free energy during this process.

In this approach, we considered chemical reactions simply by changing the concentrations of their compounds and not considering whether the alterations of these concentrations were realized by transport processes or actually by a chemical reaction. This is, in fact, more appropriate and useful in most cases. In some cases,

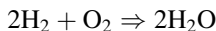
however, chemical reactions take place combined with the processes of transport of matter through biological membranes. In such cases, this approach can be misleading. In this case it is useful to introduce a special term into the Gibbs equation differentiating between transport processes and the chemical reaction.

In fact, it is possible to measure the progress of a chemical reaction using the definition of the *degree of advancement* ($d\zeta$):

$$d\zeta = \frac{1}{\nu_i} dn_i \quad (3.73)$$

A positive value of $d\zeta$ represents a particular step in the reaction from left to right, whereas $dn_i > 0$ indicates an increase and $dn_i < 0$ a decrease of the molar concentration of the substance. In order to maintain these sign conventions, the stoichiometric coefficients (ν_i) of the initial substrates must become negative and those of the end products, in contrast, positive.

Taking, for example, the reaction



the following applies:

$$d\zeta = -\frac{1}{2} dn_{\text{H}_2} = -dn_{\text{O}_2} = \frac{1}{2} dn_{\text{H}_2\text{O}} \quad (3.74)$$

Now, as the work coefficient according to our assumption in Sect. 3.1.2, the *chemical affinity* (A) is defined as:

$$A = - \sum_{i=1}^m \nu_i \mu_i \quad (3.75)$$

where $i = 1 \dots m$ are the components of the given reaction. Using Eq. 3.73, the following relation is obtained:

$$\sum_{i=1}^m \mu_i dn_i = -A d\zeta \quad (3.76)$$

To differentiate real chemical reactions ($d\zeta$) and transport processes (dn_i), it is useful to substitute both expressions into the Gibbs equation simultaneously. This may be important when determining the processes of active ion transport through membranes.

Affinity, from its definition, proves to be an energy difference and, as such, the driving force of a chemical reaction. Consequently, it can be directly inserted into the flux matrix (Eq. 3.58) instead of a generalized force (\mathbf{X}).

3.2 The Aqueous and Ionic Equilibrium of the Living Cell

From the point of view of thermodynamics, a living cell can be considered in general as a system in nonequilibrium. This nonequilibrium state is maintained by permanent processes of energy transformation. In spite of the state of the whole system, however, some of these subsystems nevertheless may be in thermodynamic equilibrium. The occurrence of such equilibria, or quasi-equilibria have already been discussed in Sect. 3.1.4, in connection with the time hierarchy of biological systems. In this section we will direct our attention to these equilibrium states, especially equilibrium distributions of charged and uncharged components.

3.2.1 The Van't Hoff Equation for Osmotic Pressure

It is very important to understand the concept of osmotic pressure for a number of cell physiological processes. Unfortunately, osmotic pressure is often confused with hydrostatic pressure, especially with the turgor pressure in plant cells. In this section it will be shown that osmotic pressure is just a property of a solution or a suspension and that although it can be the origin of the hydrostatic pressure in a cell, it is not at all identical to it.

The effect of osmotic pressure can be demonstrated best by a *Pfeffer cell* (see Fig. 3.22). This consists of a glass bell with a vertical tube or a manometer at the top. The mouth of the bell is closed by a semipermeable membrane. The bell is filled with a solution and submerged in a vessel with pure water. The membrane allows the water, but not the molecules of the solute, to pass through. In his experiments in 1877, the botanist Wilhelm Pfeffer used a sheet of porous pottery to close the bell. The pores of this “membrane” were covered by a precipitated layer of cupric(II)-hexacyanoferrate(II). In this experiment the water penetrates the membrane, passes into the glass bell, and increases the internal hydrostatic pressure. Eventually an equilibrium will be reached when the internal pressure balances the osmotic force which drives the water into the Pfeffer cell. If the membrane is truly semipermeable, if pure solvent is outside the glass bell, and if the system is really in equilibrium, then, and only then, does the hydrostatic pressure difference between the internal and the external phase in equilibrium state equal the osmotic pressure of the solution inside the Pfeffer cell.

In order to analyze this situation let us assume two phases (I and II) separated from each other by a membrane (see Fig. 3.11). Let the membrane be semipermeable thus allowing the solvent (w), but not the solute (s), to pass through. Because of solvent flux through the membrane, the volumes of both phases and their concentrations will change. This will also alter the chemical potentials (μ_i) of the components (i).

The concentration dependence of the chemical potential, using the mole fraction (x_i) (see Eq. 3.33) as a measure of concentration is:

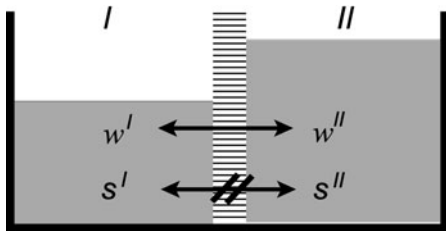


Fig. 3.11 Derivation of osmotic pressure: Phases I and II are separated from each other by a semipermeable membrane. Only the solvent (w), but not the solute (s), can penetrate this membrane. The hydrostatic pressure of both sides of the membrane is different

$$\mu_i = \mu_{ix}^0 + RT \ln x_i \quad (3.77)$$

To achieve thermodynamic equilibrium under isothermal conditions, the chemical potential of the exchangeable components of both phases I and II must be equal. In the example considered here, this will only apply to the solvent (w). This requires

$$\mu_w^I = \mu_w^{II} \quad (3.78)$$

which means that

$$\mu_{wx}^{0I} + RT \ln x_w^I = \mu_{wx}^{0II} + RT \ln x_w^{II} \quad (3.79)$$

or

$$RT \ln \frac{x_w^I}{x_w^{II}} = \mu_{wx}^{0II} - \mu_{wx}^{0I} \quad (3.80)$$

Now, let us first direct our attention to the difference in the standard chemical potentials on the right-hand-side of Eq. 3.80. Because of the pressure difference induced by the flow of the solvent, $\mu_{wx}^{0I} \neq \mu_{wx}^{0II}$. The pressure dependence of the standard chemical potentials therefore must be studied in more detail.

The following definition of the standard chemical potential can be derived from Eq. 3.32:

$$\mu_i^0 = \left(\frac{\partial G^0}{\partial n_i} \right)_{T,p,n_j} \quad (3.81)$$

The pressure dependence of μ_i^0 can therefore be expressed as

$$\frac{\partial \mu_i^0}{\partial p} = \frac{\partial \left(\frac{\partial G^0}{\partial n_i} \right)}{\partial p} = \frac{\partial \left(\frac{\partial G^0}{\partial p} \right)}{\partial n_i} \quad (3.82)$$

The way in which the sequence of the derivation steps in a total differential can be changed was discussed in Sect. 3.1.1 (see also Eq. 3.6).

From the equation of Gibb's free energy (Eq. 3.29) we obtain:

$$dG^0 = V dp \quad \text{for :} \quad dT = 0, \quad dl = 0, \quad dq = 0, \quad dn = 0 \quad (3.83)$$

and

$$\frac{\partial G_i^0}{\partial p} = V_i \quad (3.84)$$

If this is substituted into Eq. 3.82, then

$$\frac{\partial \mu_i^0}{\partial p} = \frac{\partial V_i}{\partial n_i} = \bar{V}_i \quad (3.85)$$

We are already acquainted with the variable \bar{V} which is the partial molar volume of the substance i (Eq. 3.8). Now, the differential $d\mu_i^0$ can be calculated. This is a small change in μ_i^0 that occurs when there is a small change in the pressure dp . Following the general rule (3.1), we obtain:

$$d\mu_i^0 = \left(\frac{\partial \mu_i^0}{\partial p} \right) dp = \bar{V}_i dp \quad (3.86)$$

To obtain an expression for the difference in the standard potentials due to alterations in pressure, the equation must be integrated between the corresponding limits:

$$\int_{\mu_i^{0I}}^{\mu_i^{0II}} d\mu_i^0 = \int_{p^I}^{p^{II}} \bar{V}_i dp \quad (3.87)$$

This leads to:

$$\mu_i^{0II} - \mu_i^{0I} = \bar{V}_i(p^{II} - p^I) \quad (3.88)$$

(This manner of integration only applies when $\bar{V} = \text{const}$, that is, when the partial molar volume is independent of the pressure. This is only the case for ideal solutions.)

Now, it is possible to substitute Eq. 3.88, related to the solvent (w) into Eq. 3.80:

$$p^{II} - p^I = \frac{RT}{\bar{V}_w} \ln \frac{x_w^I}{x_w^{II}} \quad (3.89)$$

This equation indicates the hydrostatic pressure difference ($p^{\text{II}} - p^{\text{I}}$) induced in our system at thermodynamic equilibrium for different values of the mole fractions in the two phases $x_w^{\text{I}} \neq x_w^{\text{II}}$. If phase I contains only pure solvent ($x_w^{\text{I}} = 1$), this pressure difference is called osmotic pressure (π), therefore

$$\pi = \frac{RT}{\bar{V}_w} \ln \frac{1}{x_w^{\text{II}}} = -\frac{RT}{\bar{V}_w} \ln x_w^{\text{II}} \quad (3.90)$$

This equation and its derivation allow us to describe osmotic pressure as a parameter reflecting a property of a solution. Under isothermal conditions the osmotic pressure of a solution is equal to the hydrostatic pressure, which is required to alter the chemical potential of the pure solvent in such a way that it will be equal to the chemical potential of the solvent in this solution.

Equation 3.90 is a precise expression to calculate the osmotic pressure of a solution, provided the mole fraction activity of the solvent x_w^{II} is used to express the concentration of the solvent (w) in the solution. This is indeed a quite unusual and rather cumbersome form for the equation. How can this equation be transformed into a more simplified form?

Some simplifications can be made for dilute solutions. First, the *activity*, which as a matter of fact is what x_w in Eq. 2.14 means, can be replaced by mole fraction *concentration*. Using Eq. 3.35, this means:

$$x_w = \frac{n_w}{n_s + n_w} \quad (3.91)$$

If the sum of the mole fractions of all components of a solution equals 1 (Eq. 3.36), then:

$$x_w = 1 - x_s \quad (3.92)$$

The number of molecules of solvent in diluted solutions is much larger than the number of molecules of the solute ($n_w \gg n_s$). A 0.1-molar solution of the substance (s) in water (w), for example, contains $n_s = 0.1$ mol of solute, but $n_w = 55.6$ mol of water (n_w means moles of water per 1,000 g solution, therefore, $1,000/18 = 55.6$). In this case, using the definition of x_s (Eq. 3.35) it becomes:

$$x_w = 1 - \frac{n_s}{n_s + n_w} \approx 1 - \frac{n_s}{n_w} \quad (3.93)$$

In addition, the following holds for dilute solutions:

$$\bar{V}_w = \frac{\partial V_w}{\partial n_w} \approx \frac{V_w}{n_w} \quad (3.94)$$

Thus, $n_w = V_w/\bar{V}$. One can also introduce the molar concentration of the solute (s), using $n_s/V = c_s$, and the total volume of the solution ($V \approx V_w$). Substituting this in Eq. 3.93, we obtain:

$$x_w = 1 - \frac{n_s \bar{V}_w}{V} = 1 - c_s \bar{V}_w \quad (3.95)$$

Let us substitute this expression in Eq. 3.90:

$$\pi = -\frac{RT}{\bar{V}_w} \ln(1 - c_s^H \bar{V}_w) \quad (3.96)$$

Now, we use the following rule to expand logarithmic expressions in series. For any number $|q| < 1$, it holds:

$$\ln(1 - q) = -q - \frac{q^2}{2} - \frac{q^3}{3} - \frac{q^4}{4} - \dots \quad (3.97)$$

For q we use the expression $c_s^H \bar{V}_w$, which is much smaller than 1. In this case we are justified in retaining only the first term of this series. This gives

$$\pi = -\frac{RT}{\bar{V}_w} (-c_s^H \bar{V}_w) \quad (3.98)$$

and

$$\pi = RT c_s^H \quad (3.99)$$

This is the *Van't Hoff equation* for osmotic pressure.

It is interesting to note that, in 1877, Van't Hoff derived this relation not in the way shown here, but by considering the analogy with the state equation for an ideal gas. He assumed that 1 mol of an ideal gas, when confined to a volume of 1 L, would exert a pressure of 2.27 MPa on the walls of the vessel. He reasoned that the molecules of a solute, when dissolved at a concentration of 1 mol/l should behave in the same way as the particles of the gas. This pressure he called “osmotic.”

The equation of state for an ideal gas is:

$$p = \frac{n}{V} RT \quad (3.100)$$

This can be transformed into:

$$\pi = \frac{n}{V} RT = cRT \quad (3.101)$$

Both the thermodynamic derivation as well as the analogy used by Van't Hoff show that Eq. 3.99 is only applicable for ideal solutions, or with some approximation for solutions that are very diluted. This restriction can be overcome by using a correction factor. This is the so-called *osmotic coefficient* (g):

$$\pi = gcRT \quad (3.102)$$

This factor must not be confused with the activity coefficient (f), which was introduced in Sect. 3.1.2 (Eq. 3.34). This will be easily understood if f is directly introduced into the thermodynamic derivation of the Van't Hoff equation (e.g. in Eq. 3.79). In this case, it would reflect the activity coefficient of water, not that of the solute. The successive transformation of this equation up to Eq. 3.25, introducing the concentration of the solute (c_s) instead of the mole fraction activity of water (x_w) in fact used other assumptions. The relation between f and g is somewhat complicated and will not be discussed here.

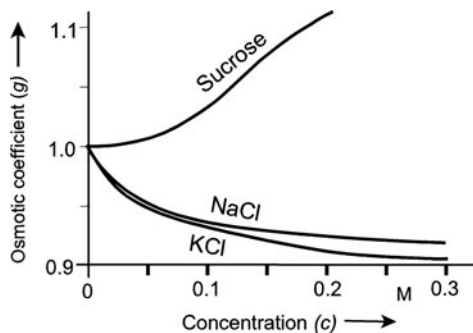
In addition to the nonideal behavior of a solution, it must be considered furthermore that the osmotic pressure exerted by dissociable substances in fact results from the sum of the osmotic pressures of all the ions formed by dissociation. Thus, if a mono-monovalent salt (such as NaCl) is assumed to be completely dissociated into Na^+ and Cl^- then the osmotically active concentration is twice as great as the salt concentration.

Therefore, one must distinguish between the *osmolarity* and the *molarity* of a solution. The osmolarity of a 0.1-M solution of NaCl (at 25°C), is therefore: $2 \times 0.1 \times g = 2 \times 0.1 \times 0.932 = 0.1864$ osM. The osmolarity can be pH-dependent for polyelectrolyte solutions because it will change according to the degree of dissociation. In contrast to the *osmolarity*, expressed as osmol/l, sometimes the term *osmolality* is used as a measure expressed as osmol/kg. Considering the temperature dependence of the volume, osmolality, i.e., relation to the mass in osmometry is preferred in relation to osmolarity.

The osmolarity of a solution can be measured by various methods. The direct method is the *membrane osmometer* where the hydrostatic pressure is indicated, resulting in a solution separated from pure solvent by a semipermeable membrane, like in a Pfeffer cell (schematically shown in Fig. 3.22). In this equipment, however, sensitive pressure sensors are used instead of the vertical tube with a water column. This method is applicable if membranes are available that are actually semipermeable to the particular solutes.

Mostly, the osmotic pressure of a solution is measured in an indirect way. In fact, the reduction of the vapor pressure, or the freezing point of a solution are basically determined in the same way as their osmolarity. Therefore, instruments are used to measure these properties. In the case of clear solutions the most accurate and comfortable technique is *cryoscopy*, i.e., the measurement of the freezing point. For solutions of macromolecules, and for polydisperse suspensions, however, this method cannot be used. In this case *vapor pressure* osmometers are preferred. They indicate small stationary temperature differences between a drop of the fluid, and a

Fig. 3.12 The osmotic coefficient (g) as a function of the concentration (c) in aqueous solutions



drop of a control solution with known osmolality, caused by differences of evaporation in a definite atmosphere.

Relating the osmolality of solutions given theoretically by Eq. 3.25, with the measured values, the osmotic coefficient (g) can be determined directly. Figure 3.12 indicates that for solutions of NaCl and KCl, starting from the value 1 in extremely dilute, i.e., “ideal” solutions, it drops to values near to 0.9 with increasing concentrations. The opposite behavior of sucrose is caused by its influence on the water structure. We will come back to this point, as well as to the relation between osmotic (g) and activity (f) coefficients in the next section.

3.2.2 Osmotic Pressure in Cells and Biologically Relevant Fluids

In the previous chapter basic equations were derived which help to understand osmotic pressure as a property of a solution, and allow exact (Eq. 3.90) – or with good approximation (van’t Hoff Eq. 3.102) – calculation of this quantity. In order to apply the van’t Hoff equation to nonideal solutions, a correction term was included, the osmotic coefficient (g). It usually becomes <1 if the concentration of the solute increases (see Fig. 3.12), in a similar way to the coefficient of activity (f) as shown in Fig. 3.1. The function $g(c)$, however, does not decline as strongly as the activity coefficient $f(c)$. But in both figures it is exceptional that the values for sucrose are rising. What could be the reason for this? What happens with other organic molecules, especially with the highly concentrated macromolecules in biological compartments? This question is important as the osmotic pressure is not only a property of true solutions, but also of colloidal solutions and to some extent also of suspensions. In this context it is referred to as *colloid-osmotic pressure*.

To solve this problem, the real properties of water in these solutions must be considered in detail. We already discussed the interaction of water dipoles with polar and apolar molecular compounds as well as with various kinds of surfaces (see Sect. 2.2.3). We asserted that several kinds of fixed, or even encapsulated water molecules in these kinds of solutions exist, which certainly contribute to the osmotic property to different degrees.

In fact, even in sucrose solutions, more so in suspensions, and in the cytoplasm of living cells, water exists both in osmotically active and in osmotically inactive phases, which may be even caused by these fixed, or “bound” kinds of water molecules. This is obvious, when analyzing the real osmotic behavior of these liquids.

To show this, let us introduce a molal *concentration* c'_s , representing the true concentration of the solvent (s) in regard to the amount of osmotically active water. Therefore it is defined as:

$$c'_s = \frac{m_s}{m_{fw}} = \frac{m_s}{m_{tw} - m_{bw}} \quad (3.103)$$

where m are mol masses of s = solute, fw = “free” water, tw = “total water,” bw = “bound” water. Now, we can define W as a measure of “bound” water as:

$$W = \frac{m_{bw}}{m_s} \quad (3.104)$$

Introducing m_{bw} from Eq. 3.104 into Eq. 3.103, rearranging, and replacing $m_s/m_{tw} = c_s$, results in:

$$c'_s = \frac{c_s}{1 - Wc_s} \quad (3.105)$$

The van't Hoff equation (Eq. 3.102) can now be used to introduce the “true” osmotic active concentration c'_s , corrected with a “true” osmotic coefficient g' . For simplification we are using the osmotic pressure in the unit “osmolal,” not in “Pa” as in Eq. 3.102. Therefore, we do not need the RT -term. It becomes:

$$\pi = g' c'_s \quad (3.106)$$

Introducing this into Eq. 3.105 and rearranging, results in:

$$\frac{1}{c_s} = W + \frac{g'}{\pi} \quad (3.107)$$

If now the value $1/c_s$ is plotted against $1/\pi$, a linear function appears, indicating W as the point at which the ordinate is crossed at $1/\pi \rightarrow 0$, and g' as the slope. This is done in Fig. 3.13 for sucrose, polyethylene glycole ($M = 400$), and bovine serum albumin (BSA) (to fit in the same figure, the molal and osmolal values for BSA were multiplied by 100!). The dotted line shows the slope for the case of $g' = 1$, which corresponds to all of the three lines. The intercepts (W) are obtained at 0.0956 for sucrose, 0.4094 for PEG400, and 189.48 for BSA (considering the factor of 100!). Converting this in terms of mol H_2O /mol solute, they are five for sucrose, 22.7 for PEG400, and 10,500 for BSA. Sucrose therefore is hydrated with one water

per OH-group, whereas BSA is surrounded by multilayers of water that do not participate in osmotic activity (according to Garlid 2000).

The osmotic pressure of the surrounding medium is an important parameter for the survival of cells, especially for their volume regulation. In classical physiology it is postulated that cells shrink in *hypertonic solutions* ($\pi_{\text{sol}} > \pi_{\text{cell}}$), attain their volume in *isotonic solutions* ($\pi_{\text{sol}} = \pi_{\text{cell}}$), and swell in *hypotonic solutions* ($\pi_{\text{sol}} < \pi_{\text{cell}}$). This is based on the assumption that cells behave like osmometers, the membrane of which is freely permeable to water but impermeable for all solutes. Furthermore, all the cellular water is considered as totally osmotically active.

There are two reasons why this assumption can be considered as strongly simplified. First, cells even more so than macromolecules in solutions contain water which is osmotically inactive. This can be demonstrated by considering the simplest animal cell: the mammalian erythrocyte, which can be considered as just a membrane-enclosed volume, packed with a large concentration of hemoglobin. Usually, the volume of these erythrocytes is measured as centrifuged hematocrit values, or by coulter counter measurements. Both methods may lead to artefacts, especially if solutions of different compositions are used. Therefore in Fig. 3.14 the relative volume is expressed as the inverse of protein concentration ($m_{\text{water}}/m_{\text{protein}}$ in g/g).

Similar to Fig. 3.13, the volume as an inverse of a concentration is plotted against the reciprocal osmolality of the external medium. The resulting curve in the same way as in Fig. 3.13 does not attain the zero point of the ordinate, clearly indicating that erythrocytes contain osmotically inactive water. Furthermore, the measured points seem to show three different linear segments. In the region of isotonic osmolality, this reflects in fact the behavior of native hemoglobin with a

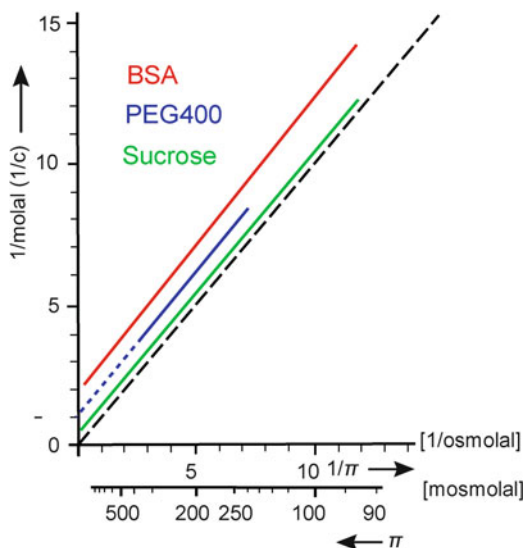


Fig. 3.13 The plot of measured osmotic activity of bovine serum albumin (BSA), polyethylene glycole (PEG400), and sucrose according to Eq. 3.107. Molal and osmolal values for BSA were multiplied by 100 to fit in the same figure (After Garlid 2000)

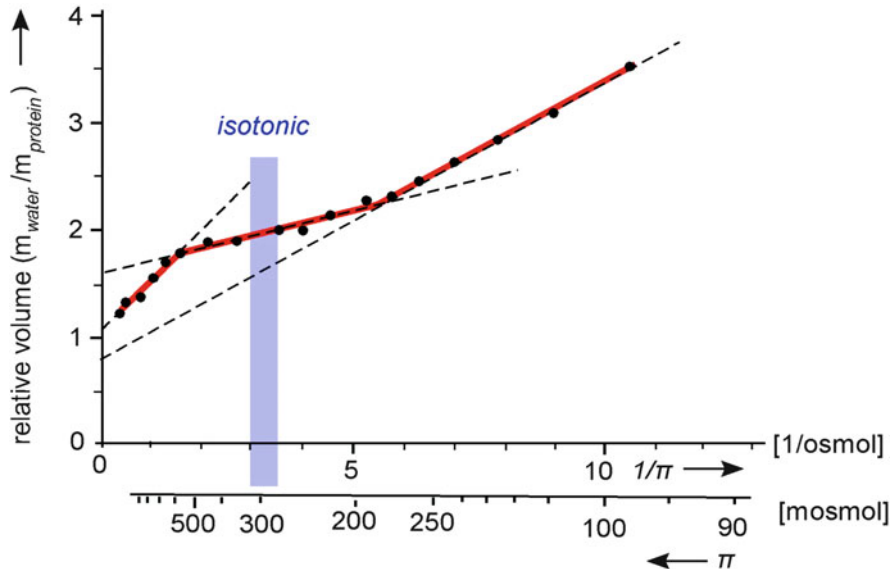


Fig. 3.14 Osmotic behavior of human erythrocytes in NaCl-solutions of various osmolarity. The relative volume is measured in mean water content ($m_{\text{water}}/m_{\text{protein}}$) (After Fullerton et al. 2006, redrawn)

content of fixed water of about 1.6 g per gram protein. Also shown are deviations away from isotonicity. In the hypotonic region, obviously a number of osmotically active small-sized molecules are released, and in the shrunken cells in hypotonic media hemoglobin aggregates. Similar experiments with corresponding results have been performed also in mitochondria.

The second, nonrealistic approach in the oversimplified consideration of the cell as a simple osmometer, concerns the question of semipermeability. Investigating for example erythrocytes in sucrose solutions, the permeability of water is in fact much greater than that of the external solutes. Using, however, smaller molecules in the external solution, like glucose, urea, etc., the cell membrane cannot be considered semipermeable and furthermore, the conditions used for the derivation of the osmotic equations do not hold. The cell will not arrive at a thermodynamic equilibrium as long as the solute is penetrating the membrane.

This question ultimately concerns the relation between the difference between osmotic ($\Delta\pi$) and hydrostatic pressure (Δp). Only in the case of thermodynamic equilibrium of water, and only if the membrane is actually semipermeable to all components of the solution, does the difference in osmotic pressure equal the difference in the generated hydrostatic pressure.

In the case of solutions with several components indicating various degrees of permeability, the following relation between osmotic ($\Delta\pi$) and hydrostatic (Δp) differences can be applied:

$$\Delta p = \sum_{i=1}^n \sigma_i \Delta \pi_i \quad (3.108)$$

This equation takes into account that in the system n substances, each with an osmotic difference $\Delta \pi_i = \pi_{i(\text{internal})} - \pi_{i(\text{external})}$, determine the osmotic conditions. Their effectiveness, with regard to the development of hydrostatic pressure, depends on the value of a factor (σ_i) which is known as *Staverman's reflection coefficient*. In contrast to the “classical” approach, this model takes into account that the membrane is not semipermeable, but permselective. This means that all components of the solution can more or less penetrate the membrane. We will analyze this situation in detail later using the approaches of nonequilibrium thermodynamics (Sect. 3.3.1). Upon consideration of the corresponding flux matrix (Eq. 3.147) the following relation is derived:

$$\sigma_s = \frac{v_w - v_s}{v_w} \quad (3.109)$$

Using the indices of the Van't Hoff's equation, v_w and v_s represent the rate of movement of the solvent (water) and the solute in the membrane. In the case of $v_s \rightarrow 0$, the reflection coefficient becomes $\sigma_s \rightarrow 1$. This is the “classical” situation of semipermeability, and therefore $\Delta p \rightarrow \Delta \pi$. However, when $v_s \rightarrow v_w$, then $\sigma_i \rightarrow 0$ and hence $\Delta p \rightarrow 0$. This occurs in the case of a membrane which allows the osmotically active substance, as well as the solvent to pass through. In this case, no hydrostatic pressure can develop even at initial differences in osmotic pressure. In such a system a thermodynamic equilibrium would finally lead to a homogeneous distribution of substance i (see Fig. 3.22).

In general, the reflection coefficient for disaccharides, such as sucrose, and for larger molecules equals nearly one. Smaller molecules, especially those which can directly penetrate the lipid layer of a biological membrane, show lower values (see Table 3.1).

Table 3.1 Typical values of reflection coefficients of nonelectrolytes for human erythrocytes (Giebisch et al. 1979) and for *Nitella flexilis* (Zimmermann and Steudle 1978). Values in parentheses: Levitt and Mlekoday 1983

	Erythrocytes	<i>Nitella</i>
Urea	0.79 (0.95)	0.91
Thiourea 0.91		
Ethylene glycol	0.86 (1.0)	0.94
Glycerol 0.88	0.80	
Acetamide 0.80	0.91	
Propionamide	0.84	
Malonamide	1.00	
Sucrose	0.97	
Glucose	0.96	
Methanol	0.31	
Ethanol	0.34	
Isopropanol		0.35
n-Propanol	0.17	

In plant cells, osmotic differences generate the so-called *turgor pressure*, which can be as high as several hundred kPa. This intracellular pressure plays a large role in the mechanical stabilization of plants. It forces the cell membrane of the plant cell against the mechanically stabilizing cell wall, which itself is freely permeable to ions and small nonelectrolytes. The turgor pressure can be directly measured by means of special pressure probes that can be introduced into the cell (Zimmermann and Neil 1989) or even noninvasively by a leaf patch-clamp pressure probe (Zimmermann et al. 2010).

In contrast to plant cells, animal cells (and also plant protoplasts) do not have the ability to resist an internal hydrostatic pressure. Within certain limits, osmotic changes in the environment can be compensated by alterations in cellular volume, thus maintaining a constant membrane area. Cell swelling, for example, can cause the forming of a sphere or smoothing of the cell surface. As described in Sect. 2.3.4, a mechanical expansion of the cell membrane, without introduction of additional molecules, however, is nearly impossible. An osmotic difference of only 1 mosmol can generate a maximum internal pressure of 2.27 kPa. Measurements made on erythrocyte membranes have shown that their tension can, at the most, only withstand an internal pressure of 0.1 kPa. For this reason, these cells have complicated mechanisms for osmoregulation, including a number of coupled receptors and transport systems.

Consequently, it is not possible to explain the volume regulation of a cell without considering the complete ionic metabolism. We will come back to this problem in connection with the Donnan equilibrium (Sect. 3.2.5).

Further Reading

Osmotically inactive water: Fullerton et al. 2006; Fullerton and Cameron 2007; Garlid 2000; Osmoregulation of animal cells: Okada 2004; Osmotic pressure in plants: Martin et al. 2001; Zimmermann et al. 2004.

3.2.3 Electrochemical Equilibrium: The Nernst Equation

We will now consider a system consisting of two phases, each of which contains a solution with a salt AB, but with different concentrations in phases I and II (Fig. 3.15). Let the salt be completely dissociated into its ions A and B. The membrane is semipermeable in that it allows ion A, but not ion B, to pass through.

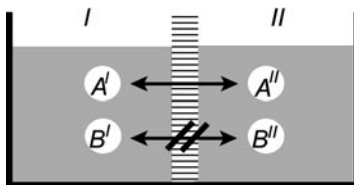


Fig. 3.15 Diagram illustrating the derivation of the Nernst equation. Phases I and II are separated from each other by a membrane which is permeable only for the ion A, but not B, of the salt AB

Before analyzing conditions of thermodynamic equilibrium of this system, we will consider the situation qualitatively. Let us suppose that there is an osmotic equilibration between both phases, compensated by electroneutral components. Thus, alteration of concentrations, induced by volume changes, can be neglected. In this system the electrostatic equilibrium will be disturbed because only ion A, driven by its concentration gradient, but not its counterpart, ion B can penetrate the membrane. This leads to an increase in the electrical potential difference across the membrane. Eventually, a strong electric field hinders a further diffusion of ion A. Ion A, consequently, will be subject to two opposing forces: on the one hand, the driving force, induced by the concentration gradient, i.e., the gradient of its chemical potential and, on the other hand, an opposing electrostatic force which only arises as a result of its own diffusion. The following equilibrium will be established: a few ions cross the membrane, inducing an electric field which stops further diffusion.

The basic condition to calculate this equilibrium is the equality of the electrochemical potentials of ion A between phases I and II:

$$\tilde{\mu}_A^I = \tilde{\mu}_A^{II}$$

Substituting the expressions for the electrochemical potentials according to Eq. 3.41, one obtains

$$\mu_A^{0I} + RT \ln a_A^I + z_A F \psi^I = \mu_A^{0II} + RT \ln a_A^{II} + z_A F \psi^{II} \quad (3.110)$$

(We will assume isothermal conditions, i.e., $T^I = T^{II} = T$.)

In contrast to the derivation of the equation for osmotic pressure, in the present case, the standard potentials of these components of phases I and II are equal, because there is no pressure difference ($\mu_A^{0I} = \mu_A^{0II}$).

Taking this into account, and re-arranging Eq. 3.110, gives

$$z_A F (\psi^I - \psi^{II}) = RT (\ln a_A^{II} - \ln a_A^I) \quad (3.111)$$

and therefore

$$\Delta\psi \equiv (\psi^I - \psi^{II}) = \frac{RT}{z_A F} \ln \frac{a_A^{II}}{a_A^I} \quad (3.112)$$

This is the *Nernst-Equation*. It gives the electrical potential difference ($\Delta\psi$) across the membrane as a function of the chemical activities of the permeating ion in both phases at thermodynamic equilibrium.

Equation 3.112 can be re-arranged to show the difference of ion activities (a^I and a^{II}) which builds up in two phases, with a given potential difference ($\Delta\psi$) in between:

$$a_A^I = a_A^{II} e^{-\frac{z_A F \Delta\psi}{RT}} \quad (3.113)$$

Such a relation has already been derived and employed using the Boltzmann equation (Sect. 2.1.4), and applied to calculate local ion concentrations near charged particles (Eq. 2.50), or in electric double layers (Eq. 2.77). In these cases, however, the concentrations (c_i) were used instead of the chemical activities (a_i). This is allowed only for ideal, i.e., diluted, solutions, or when the activity coefficients (f_i) are equal in both phases.

The Nernst equation, therefore, permits the calculation on the one hand, of the distribution of ions as a function of the electrical potential (Eq. 3.113) and, on the other hand, the electrical potential, which is induced by an unequal distribution of ions (Eq. 3.112). For both cases, however, thermodynamic equilibrium is required!

The separation of the two phases of the system by a semipermeable membrane, as discussed here, reflects just one special example. In the case of ion distribution in electric double layers, or ionic clouds, as described in Sects. 2.3.5 and 2.3.6, the electric potential gradient is predicted by the fixed charges, and the distributions of the ions are not limited by a membrane. Here, both the anions and the cations in the case of equilibrium are distributed according to this equation.

All equations, derived in the present section, are applicable only for thermodynamic equilibria. This means that the Nernst equation cannot be used to calculate the membrane potential of a living cell. Actually, the membrane potential of a living cell is either a diffusion potential (liquid junction potential), or it is generated by electrogenic ion pumps (see Sects. 3.4.2 and 3.4.3). Conversely, despite the nonequilibrium distribution of the ions in the cell in general, it is quite possible that some types of ions may be distributed passively and then they are actually in equilibrium.

An example of this is the chloride concentration of most animal cells. Because its membrane permeability is rather fast, and because no chloride pumps in the membrane exist, its distribution is mostly passive, and predicted by the existing transmembrane potential. The Nernst equation, therefore, allows one to calculate the internal chloride concentration, if the external chloride concentration and the transmembrane potential are known. Conversely, knowing, for example, the distribution of chloride inside and outside the cells, one can calculate the transmembrane potential of the cell. In this context, however, it is important to underline that in this case the chloride distribution is just an *indicator* for the membrane potential, but not the *reason* for it!

This consideration allows the establishment of a method to measure the transmembrane potential of cells without microelectrodes. The chloride distribution can be easily determined using the radioisotope ^{36}Cl . Sometimes one can use other small charged organic molecules, which penetrate the membrane quickly, and are labeled by ^3H or ^{14}C . Knowing the distribution of these ions, the membrane potential ($\Delta\psi$) can be calculated according to Eq. 3.112.

Furthermore, the Nernst equation allows the calculation of electrode potentials. If a metal is dipped into an electrolyte solution, then cations are detached from the

metal lattice and an electrochemical equilibrium is obtained. In general, this is the same situation as discussed above. The resulting electrical potential difference between the metal and the solution is called *electrode potential*. If the cation from the metal of the electrode forms a poorly soluble salt with an anion of the solution, and if the concentration of this anion is much greater than that of the cation, then the electrode potential, linked through the soluble product of the salt, is directly determined by the chemical activity of this anion. In this way, for example, a silver electrode covered by AgCl can be used to measure the Cl^- activity of a solution (see Fig. 3.16). The voltage of this electrode with respect to a reference electrode, according to the Nernst equation (Eq. 3.112), is proportional to the logarithm of the Cl^- activity in the solution. It is important to emphasize that these kinds of electrochemical methods allow the measurement of ion activities (a_i), in contrast to most chemical methods, indicating its concentrations (c_i).

The electrochemical measurement of ion activity and of many other chemical substances has become an important and universal technique. Using special semi-permeable membranes, or water impermeable layers containing special ionophores, electrodes can be made which are highly selective to measure the activity of special chemical components. In this case, these membranes or microphases separate a reference solution of known composition from the test solution. The potential difference between the reference solution and the test solution is measured as the voltage between two reference electrodes dipped in both phases.

A typical example for this kind of measurement is the usual pH electrode (see Fig. 3.16). In this case, a thin membrane of special glass or another material allows the protons to equilibrate between both phases. If, as a reference phase, the pH electrode is filled by a buffer holding the pH constant, then the voltage between the pH electrode and the reference electrode indicates the pH of the test solution. Usually, combinations between pH and reference electrodes are used, consisting of one single glass body.

For electrochemical methods of analysis a pair of electrodes is always necessary. The electromotive force will be measured between a selective measuring electrode and a so-called reference electrode. A *reference electrode* is an electrode that exhibits a constant electrode potential that is independent of the composition of the solution into which it is immersed. If a pair of identical reference electrodes is used, then the electrode potentials will mutually oppose each other and therefore the direct potential difference, i.e., the actual electromotive force between the two phases can be measured. If two different reference electrodes are used, a constant $\Delta\psi$ is superimposed, which can be subtracted from the measured value.

To make a reference electrode, a metal electrode is covered with a slightly soluble salt of this metal. This is immersed into a reference solution of constant composition, which is connected to the test solution by means of a pathway, where little, if any, convection can take place (liquid junction). This liquid junction is formed in different ways. It can be a tube, filled by an agar gel. However, in industrially produced electrodes, it is mostly just an opening in the glass wall, covered by a small filter made from sintered glass. Sometimes a ground glass stopper is used with a liquid film in between.

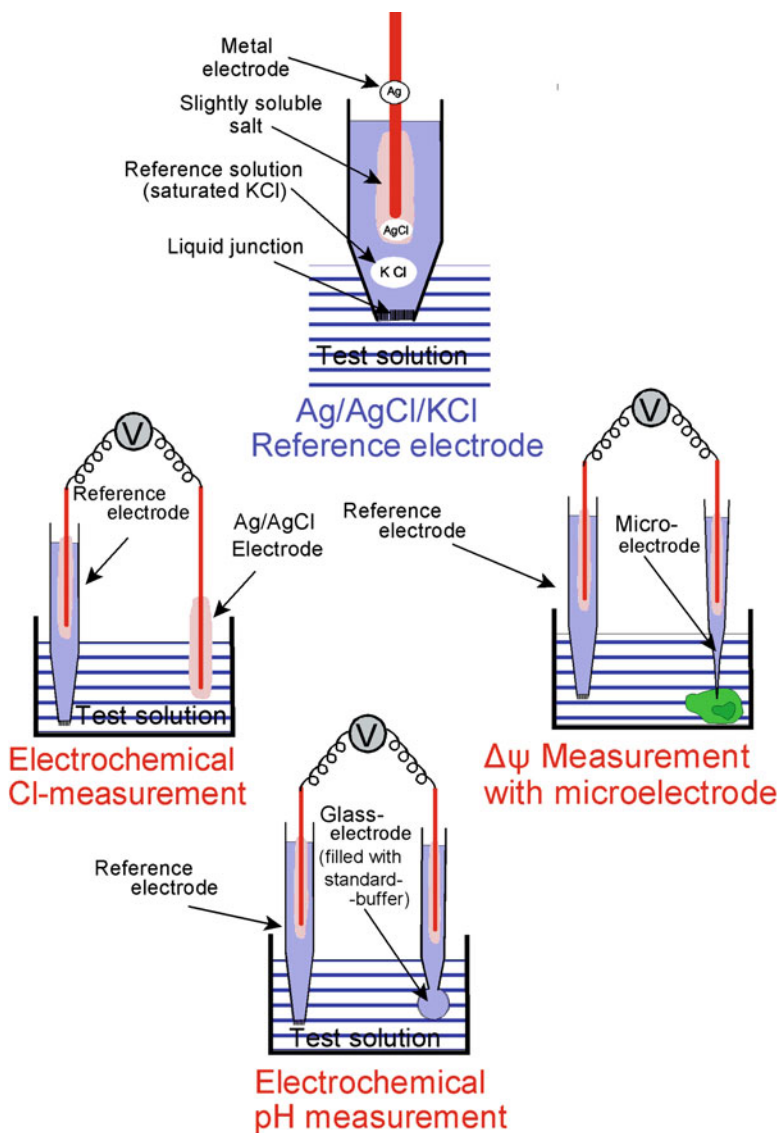


Fig. 3.16 The construction of an Ag/AgCl/KCl-reference electrode, as well as its application for electrochemical measurements of Cl^- activity and pH, and as a reference electrode in electrophysiology

The most common types of reference electrodes are Ag/AgCl/KCl electrodes (see Fig. 3.16). In this case, a saturated solution containing 4.6 M KCl as a reference solution is used. The reason for choosing KCl is that in any case, a concentration gradient exists through the connection between the reference solution and the test solution. This could become a source of an electrical potential difference (diffusion

potential, see Sect. 3.3.3). This would contradict the conditions for reference electrodes, i.e., the constant potential, independent of the test solution. This diffusion potential, however, depends not only on the concentration gradient, but also on the mobility of both ions (see Eq. 3.190). The mobility of the K^+ and Cl^- ions in aqueous solutions, however, is almost equal. This means that even for strong concentration gradients between reference solution and test solution, no diffusion potential can occur. This would not be the case, if for example NaCl was used as the reference solution.

Unfortunately, this useful property of K^+ and Cl^- ions is valid only in pure solutions. If a reference electrode is immersed for example in suspensions of charged particles, a so-called *suspension effect* occurs. Under the influence of the electric double layers of these particles, the mobility of the ions may change. Therefore, small diffusion potentials can occur. This effect, for example is important in the case of pH measurement in blood. Moreover, it can be the source of errors in measurements with microelectrodes in cells.

Microelectrodes, as applied in electrophysiology, are usually Ag/AgCl/KCl electrodes. They are constructed from glass tubes that are drawn out at one end to a very fine capillary (diameter $< 1 \mu m$). In this case, no further diffusion limitation in the tip is necessary. Because of the electric charge of the glass and the extremely thin glass wall at the tip, so-called *tip potentials* of microelectrodes occur which in worst cases can be as large as several millivolts. Microelectrodes can also be sealed with ion-selective materials so that intracellular ionic activity can be measured directly.

Further Reading

Electrodes in general: Fry and Langley 2005; Varma and Selman 1991; microelectrodes: Amman 1986.

3.2.4 The Donnan Equilibrium: Basic Properties

The Donnan state represents an equilibrium between two phases, containing not only small anions (A) and cations (C), both of which are freely exchangeable between the phases (i.e., can penetrate the membrane), but also charged molecules or particles (M) that are fixed in one phase (i.e., cannot penetrate the membrane; see Fig. 3.17). This type of equilibrium was investigated by F. G. Donnan in 1911.

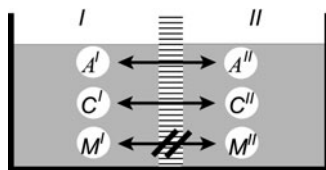


Fig. 3.17 The derivation of the Donnan equilibrium. Phases I and II are separated from each other by a membrane, that is permeable for the anions (A) and cations (C), but not for the charged molecules M

These considerations are of particular importance to understand the properties of various colloidal as well as biological systems where the phases are separated by membranes with particular conditions or permeability.

To analyze this situation, let us denote the exchangeable ions by the index i , and the fixed charge components by m ; the concentration, or the activity of the components of one phase, resp. inside the cell with c^I or a^I , and in the second phase, or external solution with c^{II} or a^{II} , respectively. The parameter z_m denotes the number and the sign of charges of the nonexchangeable molecules.

The Donnan equilibrium is defined by the following three conditions:

- All permeable ions (i), being in equilibrium, are distributed according to the Nernst equation (Eq. 3.113):

$$a_i^I = a_i^{II} e^{-\frac{z_i F \Delta\psi}{RT}} \quad (3.114)$$

- In both phases the sum of charges must be zero (electroneutrality condition):

$$\sum z_i c_i + \sum z_m c_m = 0 \quad (3.115)$$

- Water between both phases is distributed according to its thermodynamic equilibrium:

$$\Delta\mu_W = 0 \quad (3.116)$$

This last condition is easily fulfilled in isobaric systems ($\Delta p = 0$) and in the case of free water movement. This applies for animal cells, which to some extent are able to swell or shrink. The volume of plant cells is limited by the rigid cell wall. In this case a so-called *Donnan-osmotic pressure* may occur.

These basic equations can be combined and solved for particular parameters. Equation 3.114 allows us to calculate the relation of the activities of the exchangeable ions inside and outside the cell as a function of the phase potential difference $\Delta\psi$. If the activity of a univalent anion is denoted by a_A ($z_A = -1$), and that of the corresponding cation by a_C ($z_C = +1$), it follows:

$$\frac{a_A^I}{a_A^{II}} = \frac{a_C^I}{a_C^{II}} = e^{\frac{F\Delta\psi}{RT}} \equiv r \quad (3.117)$$

The so-defined parameter r is known as the *Donnan ratio*. According to Eq. 3.117, it is related to the Donnan potential in the following way:

$$\Delta\psi = \frac{RT}{F} \ln r \quad (3.118)$$

The Donnan potential ($\Delta\psi$) is substantially determined by the amount of nonexchangeable charged components in the phases, as reflected in condition (3.115).

Let us now consider the situation with just a single kind of cation (C, with $z_C = +1$), and a single kind of anion (A, with $z_A = -1$), and only one kind of charged component (M, with z_M) inside the cell. In this case, the equation of electroneutrality of both phases, according to Eq. 3.115 can be written easily:

$$\begin{cases} c_C^I - c_A^I + z_M^I c_M^I = 0 \\ c_C^{II} - c_A^{II} = 0 \end{cases} \quad (3.119)$$

This system of equations can be solved by reorganizing it and dividing one by the other:

$$\frac{c_A^I}{c_A^{II}} = \frac{c_C^I + z_M^I c_M^I}{c_C^{II}} \quad (3.120)$$

If the activity coefficients of the ions in both phases are equal (i.e., $a_i = c_i$), one can substitute the Donnan ratio (Eq. 3.117) in this equation:

$$r = \frac{1}{r} + \frac{z_M^I c_M^I}{c_C^{II}} \quad (3.121)$$

Re-arranging this, a simple quadratic equation for r is obtained:

$$r^2 - \frac{z_M^I c_M^I}{c_C^{II}} r - 1 = 0 \quad (3.122)$$

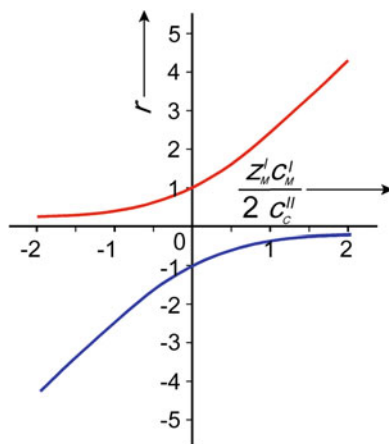
which can be solved in the usual way:

$$r = \frac{z_M^I c_M^I}{2c_C^{II}} \pm \sqrt{\left(\frac{z_M^I c_M^I}{2c_C^{II}}\right)^2 + 1} \quad (3.123)$$

The dependence of the Donnan ratio (r) for this simplified case on the expression $\frac{z_M^I c_M^I}{2c_C^{II}}$ is shown in Fig. 3.18. Considering the corresponding definition (3.117), it is clear that the negative values of the Donnan ratio (r) (blue curve) have no real meaning. Therefore, only the positive sign of the root in Eq. 3.123 is of interest. For the case $z_M^I c_M^I = 0$ it becomes $r = 1$ and Eq. 3.118 gives $\Delta\psi = 0$. If the nonexchangeable components are negatively charged ($z_M^I < 0$), then $r < 1$. This gives a negative Donnan potential ($\Delta\psi < 0$). Conversely, the value $z_M^I > 0$ means $r > 1$ and therefore $\Delta\psi > 0$.

A reduction of the ionic strength in the external solution, i.e., a reduction of c_K^{II} , leads to an increase in the absolute value of the ratio $\frac{z_M^I c_M^I}{2c_C^{II}}$. If $z_M^I < 0$, this means a

Fig. 3.18 Plot of Eq. 3.123.
 Red curve – positive, blue
 curve – negative values of the
 root expression



shift of $\Delta\psi$ in a negative direction, and vice versa, if $z_M^I > 0$, a reduction of the ionic strength causes an increase in the Donnan potential. The absolute value of the Donnan potential, therefore, increases when the charge concentration ($z_M^I c_M^I$) is increased, as well as when the ionic strength in the external solution (c_C^{II}) is reduced.

It should be noted, however, that the Donnan potential can only be measured using reference electrodes with salt bridges (see Fig. 3.16), but not with simple Ag/AgCl electrodes. In fact, if the whole electrochemical system, consisting of reversible electrodes and both electrolyte phases, is in thermodynamic equilibrium, no electromotoric force (emf) can occur. Using electrodes with salt bridges, however, an emf emerges as the difference in the liquid junction potentials at the tops of the reference electrodes that are not in equilibrium. This is called the “indirect method” for determining the Donnan potential. The same is possible by measuring the pH difference between both systems.

Donnan equilibrium not only occurs in phases which are bounded by a membrane but is also of particular importance in various colloidal systems and matrices consisting of charged molecules. It determines the swelling and shrinking of these phases as a result of Donnan-osmotic processes.

Further Reading

Overbeek 1956; Dukhin et al. 2004.

3.2.5 The Donnan Equilibrium in Biological Systems

Although the living system in general and particularly living cells are not in thermodynamic equilibrium, we already pointed out that in fact a number of subsystems nevertheless fulfill equilibrium conditions. Therefore, there exist a number of Donnan systems which are worthy of consideration.

One example concerns the distribution of ions near the fixed charges of the membrane surface coat. This means the extension of the theory of electric double layers (Sect. 2.3.5, Fig. 2.43), to real conditions in the layer of glycoprotein and glycolipid molecules at the surface of most cells (Sect. 2.3.6, Fig. 2.48). Furthermore, this is quite important in order to calculate the conditions of intercellular clefts, and accordingly the Donnan-osmotic behavior of the extracellular space in tissue (Fig. 3.34). In the same way, despite the fact that in most animal cells Na^+ and K^+ ions are actively pumped, others, like Cl^- , as well as water, could be freely exchangeable. Therefore, a Donnan equilibrium occurs as a result of these permeable ions. This also concerns the equilibration of water, resp. volume regulation. In this case, however, not only the intracellular proteins are to be considered as carriers of nonexchangeable charges, but additionally the charges of Na^+ and K^+ are quasi “fixed.” Even if the transmembrane potential is determined by diffusion processes (see Sect. 3.3.3), or electrogenic pumps (Sect. 3.4.1), the relation between $\Delta\psi$, the distribution of freely exchangeable ions like Cl^- or pH, and subsequently the volume (V) can be calculated in a similar way.

Furthermore, Donnan equilibrium may occur if the cell membrane is occasionally opened to ions by an ionophore, by a toxin, or by any other influences. A living cell will also shift slowly towards a Donnan equilibrium, if, as a result of low temperature or a chemical blocker, the ATP-driven ion pumps are inhibited. A typical example of this is the distribution of ions in erythrocytes of blood preserves, stored over a longer period of time. These situations occur not only under experimental conditions, but sometimes also in vivo.

Similar Donnan potentials are established in biocolloids like the eye lense or articular cartilage. The fixed charges of the proteoglycans in this case effect the mechanical behavior of articular cartilage by Donnan-osmotic pressure.

In most of these cases the Donnan equilibrium can be considered as a sort of *quasi-equilibrium state* (see Sect. 3.1.4). Figure 3.8 demonstrates the time hierarchy of characteristic rate constants of various biological processes. In human erythrocytes, and in most animal cells, as a rule, the distribution of Cl^- ions between internal and external media is passive. The same is true for the equilibration of pH. These processes of equilibration are very fast, and their stationary state can be calculated according to the Donnan equilibrium. This situation is called a “*quasi*” equilibrium state, because over a longer time interval the concentrations of potassium and sodium may shift. The equilibration of Cl^- concentration, and the pH, therefore follows the slow shift in Na^+ and K^+ concentration, and can be considered as in equilibrium only for a limited time of observation (see Sect. 3.1.4, Fig. 3.8).

Equation 3.120 in the previous chapter allows us to understand the behavior of a Donnan system in general, but in fact, it reflects an extremely simplified situation. It has not been taken into account that according to the condition of iso-osmolarity (Eq. 3.116), changes in cell volume could appear. Volume changes, however, would lead to changes in intracellular concentrations, and consequently in activity and osmotic coefficients of all participants.

Furthermore, according to the Donnan condition, a pH equilibrium between the two phases in this system occurs. A pH change, however, influence the charges of organic molecules. In its simplest form, this dependence can be expressed as:

$$z_M = -z_{M0}(pH - pH_{iso}) \quad (3.124)$$

Near the isoelectric point ($pH = pH_{iso}$) the total molecule is uncharged ($z_M = 0$). Below this point ($pH < pH_{iso}$), z_M becomes positive, and above it ($pH > pH_{iso}$), z_M becomes negative (see Sect. 2.2.8, Figs. 2.30 and 2.31).

Considering all of these relations, one obtains a system of nonlinear equations that can be solved by iteration. Some basic stationary properties of such a feedback system will be demonstrated for the case of Donnan equilibrium in human erythrocytes. As already mentioned, this situation can occur under conditions of blood storage or as a result of other treatment.

The most important nonpenetrating charge component in erythrocytes is hemoglobin. In vivo, it has a concentration of 7 mM. Its isoelectric point is $pH_{iso} = 6.8$ (at 25°C) and $z_{M0} = 10.5$ eq/mol. The Donnan potential ($\Delta\psi$) of erythrocytes in solutions of various NaCl concentrations is shown in Fig. 3.19. The constant osmotic pressure of 300 mosmol of the external solution is maintained by various concentrations of sucrose. The intracellular pH of the erythrocytes depends on the pH value of the external solution which is indicated on the abscissa, as well as on the Donnan potential itself. At the isoelectric point of hemoglobin, the Donnan potential becomes zero ($\Delta\psi = 0$). The highest potentials, both negative and positive, are obtained at the greatest distances from the isoelectric point, and in solutions with the lowest NaCl concentrations.

The volume changes in these cells are shown in Fig. 3.20. The volume, expressed here as the percentage of the in vivo volume of erythrocytes, is indicated as a function of the NaCl concentration of the external solution at constant $pH = 7.4$. It is important to note that independently of the external NaCl concentration, the osmotic pressure of the external medium is always adjusted to the isotonic osmotic pressure:

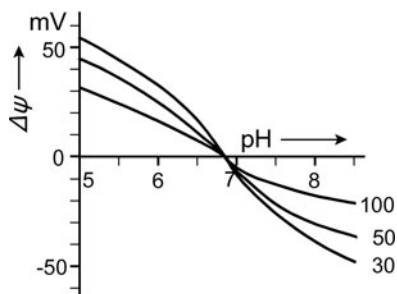


Fig. 3.19 The Donnan potential ($\Delta\psi$) of human erythrocytes in isotonic NaCl-sucrose solutions dependent on external pH. The curves represent situations for NaCl solutions of 30, 50, and 100 mM ($\pi = 300$ mosmol, $T = 25^\circ\text{C}$) (After Glaser et al. 1980)

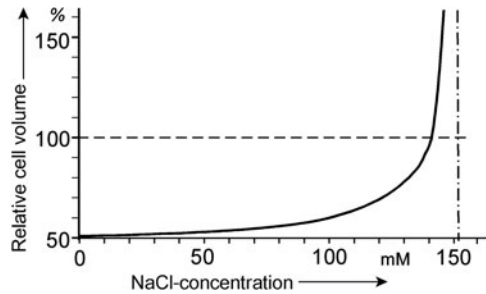


Fig. 3.20 Donnan-osmotic alterations of the relative volume (V) of human erythrocytes as a function of external NaCl concentration in isotonic NaCl-sucrose solutions ($\text{pH} = 7.4$, $T = 25^\circ\text{C}$). The osmotic pressure of the solutions with different NaCl-concentrations is always balanced by sucrose, adjusting to $\pi = 300$ mosmol. At $c = 152$ mM, the solution contains only NaCl, without sucrose. The volume is given as the percentage of the in vivo volume of erythrocytes (after Glaser et al. 1980, redrawn)

$\pi = 300$ mosmol. In spite of this, the cells shrink in solutions of low ionic strength. An isotonic condition of the incubation medium, therefore, is no guarantee for the maintenance of a normal cell volume! A normal volume is achieved only in a solution containing about 20 mosmol sucrose and 140 mM NaCl. In this case, the sucrose just compensates the osmotic pressure of the hemoglobin, which has a comparatively high osmotic coefficient (see Fig. 3.13). Furthermore, it is interesting that the volume curve rises steeply at high ionic strengths. Mathematically, no results can be obtained from the equations assuming conditions of pure (isotonic!) 152 mM NaCl solutions. In this case, the volume would become infinitely large. The experiments indicate that erythrocytes in solutions of pure electrolytes undergo hemolysis, if the membrane becomes permeable for these ions. This is known as *Donnan-osmotic hemolysis*. As indicated in Fig. 3.20, this can occur even in isotonic solutions.

To determine experimentally whether a cell is in a state of Donnan equilibrium, the relation between the internal and external ionic activities must be determined. For a Donnan state to be present, the Donnan ratios (r) from Eq. 3.117 must correspond. For most cells in vivo, the ratio obtained using the sum of the sodium and potassium ions $[(a'_{\text{K}} + a'_{\text{Na}})/(a''_{\text{K}} + a''_{\text{Na}})]$ may be close to that for a Donnan equilibrium, but this would not be the case for either Na^+ or K^+ alone. Active transport changes their relative concentrations in opposing directions.

Similar calculations to those demonstrated here for the case of erythrocytes can be applied to other cells, taking into account the membrane potential. In these cases, however, the situation will be complicated by the charges and structure of organelles and the endoplasmatic reticulum net.

Further Reading

Glaser and Donath 1984; Sun et al. 2004; Fraser and Huang 2007.

3.3 Phenomenological Analysis of Fluxes

As mentioned repeatedly, biological functions result in particular molecular processes and appear finally in the form of visible and measurable phenomena. This circumstance becomes apparent especially in various membrane functions as will be described in the next chapters. In a strict sense, the transport of ions and molecules through biological membranes must be considered as a highly specific process of molecular interaction of these species with the structure of a particular transport protein (see Sect. 3.4.5). Nevertheless, the classical approaches based on phenomenological thermodynamic considerations have been used extensively and with considerable success to investigate various processes of molecular and ion transport, and their consequences in volume regulation and various electric phenomena.

A flux, as defined phenomenologically in Sect. 3.1.3, is the amount of a substance, which passes in a perpendicular direction through a definite area of a surface in a unit of time. Its basic measure therefore, is: $\text{kg s}^{-1} \text{m}^{-2}$, or better: $\text{mol s}^{-1} \text{m}^{-2}$. In the following discussion, the molar flux of a substance will be denoted by the symbol \mathbf{J} .

When considering fluxes through cell membranes difficulties may arise in some cases if the exact surface area of a cell is not known. In these cases modified units are used, such as: “ mol s^{-1} per cell,” “ mol s^{-1} per liter of cells,” “ mol s^{-1} per liter of cell-water,” etc. As the unit of time, instead of seconds, minutes, hours, or even days are frequently used. Such units are quite convenient for some physiological investigations. However, it is not possible to substitute such units directly into the thermodynamic equations.

3.3.1 The Flux of Uncharged Substances

Let us first consider the diffusion of an uncharged substance (i), not considering any coupling with other fluxes. Using Eqs. 3.45, 3.49, and 3.33, gives

$$\mathbf{J}_i = L_i \mathbf{X}_i = -L_i \text{grad } \mu_i = -L_i \text{grad } (\mu_i^0 + RT \ln a_i) \quad (3.125)$$

In order to calculate the gradient of the chemical potential the dependence of the parameters on their position in space must be considered. Assuming constant temperature ($\text{grad } T = 0$), and constant pressure ($\text{grad } p = 0$), the gradient of the standard potential ($\text{grad } \mu_i^0$) also becomes zero. Equation 3.125 therefore can be written as:

$$\mathbf{J}_i = -L_i RT \text{grad } \ln a_i = -\frac{L_i RT}{a_i} \text{grad } a_i \quad (3.126)$$

(To explain this rearrangement: the differential operator “grad” can be handled like a deviation d/dx . Then the rules of sequential differentiation are applied).

As has already been discussed in Sect. 3.1.3 a flux can also be expressed by the parameters concentration (c_i) and velocity (v_i) (Eq. 3.47). From this we came to an equation which includes the mobility (ω_i) (Eq. 3.53). If we consider, furthermore, that \mathbf{X}_i is not a simple mechanical force, but results from the gradient of a chemical potential determined by molar concentrations, the Avogadro number (N) must be included.

$$\mathbf{J}_i = \frac{c_i \omega_i}{N} \mathbf{X}_i \quad (3.127)$$

Comparing Eq. 3.125 and 3.127 results in:

$$L_i = \frac{c_i \omega_i}{N} \quad (3.128)$$

Suppose $c_i \approx a_i$, then the combination of Eq. 3.128 with 3.126 gives:

$$\mathbf{J}_i = -\frac{\omega_i}{N} RT \text{ grad } c_i = -\omega_i kT \text{ grad } c_i \quad (3.129)$$

Introducing the diffusion coefficient one gets *Fick's first law of diffusion*:

$$\mathbf{J}_i = -D_i \text{ grad } c_i \quad (3.130)$$

If there is only a one-dimensional concentration gradient in the x -direction this equation simplifies to:

$$\mathbf{J}_{ix} = -D_i \frac{dc_i}{dx} \quad (3.131)$$

The mobility (ω_i) of an almost spherical molecule can be interpreted mechanically by referring to Stoke's law (Eq. 3.54). By definition it has the unit: $\text{m s}^{-1} \text{N}^{-1}$. In Sect. 2.1.6 we introduced the diffusion coefficient (D_i) in the context of the translational movement of molecules. This led to relations between the diffusion coefficient, mobility (ω_i), and the molar mass (M_i) (see Eqs. 2.38–2.40).

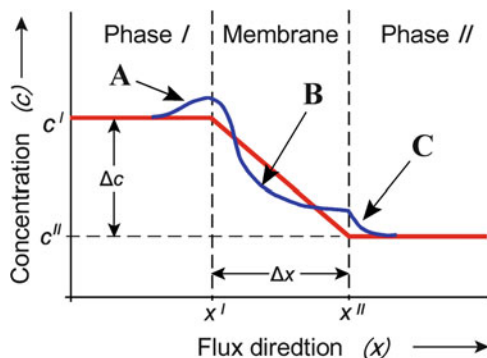
We introduced the diffusion coefficient as $D_i = \omega_i kT$. Its unit therefore is $\text{m}^2 \text{s}^{-1}$. Using the measure $\text{mol m}^{-3} \text{m}^{-1} = \text{mol m}^{-4}$ for the concentration gradient in Eq. 3.130 or Eq. 3.131, one obtains the unit $\text{mol s}^{-1} \text{m}^{-2}$ for the flux, according to its definition.

In addition to Fick's first law of diffusion, which gives the flux as a function of the concentration gradient, *Fick's second law* allows one to determine the establishment of a concentration gradient of a given substance (c_i) as a function of time (t).

Fig. 3.21 Possible functions $c(x)$ in a membrane system.

Red – ideal case, blue – disturbed case, caused by the following reasons:

- (a) Adsorption of substance at the membrane surface;
- (b) differences in the mobility of the substance inside the membrane;
- (c) diffusion layer caused by imperfect stirring at the surface



It is a partial differential equation of the second order which for diffusion in one direction is:

$$\left(\frac{\partial c_i}{\partial t}\right)_x = D \left(\frac{\partial^2 c_i}{\partial x^2}\right)_t \quad (3.132)$$

This equation is used to calculate a concentration gradient which occurs when diffusion takes place in a homogeneous phase.

In contrast to this case where a continuous gradient of concentration occurs, in membrane systems various discontinuities in concentration are to be expected. This means that the function $c_i(x)$ could become rather complicated. Schematically, this is illustrated in Fig. 3.21. The simplest case is shown by the red line where phase I contains a solution with concentration c^I and correspondingly, c^{II} is the concentration in phase II. The concentration in the membranes falls linearly with a slope of $\Delta c/\Delta x$.

The blue line in Fig. 3.21 shows an irregular concentration pattern. In this case effects are considered which can occur at the membrane surface as well as in its interior. The deviation A marks an adsorption of the substance at the membrane surface, or a change in concentration of ions in an electric double layer. The effective thickness of this layer is very small, being less than 10 nm. Inside the membrane (Fig. 3.21B), deviations of the linear concentration profile can occur by differences of the mobility of the substance in the x -direction or, in the case of ion transport, even by dielectric inhomogeneities.

In special cases *diffusion layers*, also called *unstirred* or *boundary layers*, may occur near the surface of membranes (Fig. 3.21C). These are near membrane regions without streaming and convections. In this region substances transported through the membrane, or involved in a chemical reaction, can move only by diffusion. An increase or decrease of local concentration can occur, which depends on the relationship between the transport or reaction rate, introducing substance into the region, and the rate of its removal, i.e., the rate of diffusion. In this case a stationary concentration gradient is built up.

In systems with artificial ion exchange membranes, large diffusion layers can be indicated by special interference methods or by microelectrodes. In contrast to

speculation in earlier papers, in vivo such layers are mostly a lot less than 1 μm . These layers may significantly affect biochemical reactions or transport processes of biological membranes. They can become important especially in cases where the cell surface is covered by microvilli or special caverns, or where reactions take place in the intercellular space (Fig. 3.34). Even the occurrence of diffusional layers of protons near membranes is discussed.

Let us now consider the simplest case, represented by the solid line in Fig. 3.21 in more detail. Let the concentration gradient (dc_i/dx) inside the membrane be constant and equal $\Delta c_i/\Delta x$, whereas: $\Delta c_i = c_i^I - c_i^I$. In this case Eq. 3.131 becomes

$$\mathbf{J}_i = -D_i \frac{\Delta c_i}{\Delta x} \equiv -P_i \Delta c_i \quad (3.133)$$

The parameter $P_i = D_i/\Delta x$ is the *permeability coefficient* measured in m s^{-1} . The same parameter will be used in Sect. 3.3.3 to calculate fluxes of ions. It is important to stress that this is the same parameter with an identical definition.

Let us now consider a system with flux interactions. In this case the flux coupling must be taken into account as discussed in Sect. 3.1.3. To demonstrate these approaches, we will consider only the simplest case of a binary system, represented for example by a flux of an uncharged substance (\mathbf{J}_s) and its solvent, the flux of water (\mathbf{J}_w). The driving forces of both fluxes are the negative gradients of their chemical potentials. To simplify the derivation we will use simply their differences $\Delta\mu_s$ and $\Delta\mu_w$ as driving forces.

In a first step we must write down the equation for the dissipation function according to the rule of Eq. 3.64:

$$\Phi = \mathbf{J}_w \Delta\mu_w + \mathbf{J}_s \Delta\mu_s \quad (3.134)$$

In this case Φ is an integral parameter for the whole membrane thickness Δx .

In the next step we will modify this equation in such a way that instead of the parameters $\Delta\mu_s$ and $\Delta\mu_w$, forces are introduced that are directly measurable.

Let us first consider the difference of chemical potential of the water ($\Delta\mu_w$).

$$\Delta\mu_w = \Delta\mu_w^0 + RT \ln \frac{x_w^I}{x_w^{II}} \quad (3.135)$$

In Sect. 3.2.1 we discussed the chemical potential of water in detail and considered in particular its dependence on mole fraction (x_w) as well as on pressure (p). Now we will make use of these derivations. Using Eqs. 3.88 and 3.89, as well as the definition of osmotic pressure, we come to:

$$\Delta\mu_w^0 = \bar{V}_w \Delta p \quad \text{and :} \quad RT \ln \frac{x_w^I}{x_w^{II}} = -\bar{V}_w \Delta\pi \quad (3.136)$$

where \bar{V} is the partial volume of water, Δp – the difference of hydrostatic pressure, and $\Delta\pi$ – the difference of osmotic pressure. Equation 3.135 therefore can be rewritten as

$$\Delta\mu_w = \bar{V}_w(\Delta p - \Delta\pi) \quad (3.137)$$

Let us now consider the other chemical potential difference in Eq. 3.134, namely $\Delta\mu_s$. This parameter depends on pressure difference in the same way as the difference of the chemical potential of water. According to Eqs. 3.135 and 3.136 one can write:

$$\Delta\mu_s = \bar{V}_s \Delta p + RT \ln \frac{c_s^I}{c_s^{II}} \quad (3.138)$$

The second term of this equation can be expanded as a series using the common rule for parameters $x > 0$:

$$\ln x = 2 \left[\left(\frac{x-1}{x+1} \right) + \frac{1}{3} \left(\frac{x-1}{x+1} \right)^3 + \frac{1}{5} \left(\frac{x-1}{x+1} \right)^5 + \dots \right] \quad (3.139)$$

Now we substitute for x the value c^I/c^{II} . With good approximation we are justified in retaining only the first term of this series:

$$\ln \frac{c^I}{c^{II}} = 2 \left(\frac{\frac{c^I}{c^{II}} - 1}{\frac{c^I}{c^{II}} + 1} \right) = \frac{\frac{c^I}{c^{II}} - c^{II}}{2} = \frac{\Delta c}{\bar{c}} \quad (3.140)$$

This shows that the logarithm of the ratio of the two concentrations, or even activities, can be replaced by the difference in concentration (Δc) divided by the arithmetic mean (\bar{c}) of the concentrations of the two phases. This connection is not only true approximately, as seen by this expansion of the series, but it can be proved mathematically that it is exactly equal.

Relation (3.140), applied to Eq. 3.138, together with the Van't Hoff equation for osmotic pressure: $\Delta\pi = RT\Delta c$ (Eq. 3.99), gives

$$\Delta\mu_s = \bar{V}_s \Delta p + \frac{1}{\bar{c}_s} \Delta\pi \quad (3.141)$$

Now we have really found reasonable expressions for the differences of the chemical potentials, and using them we can rearrange the equation for the dissipation function (Eq. 3.134). Introducing Eqs. 3.138 and 3.141 into Eq. 3.134 we get:

$$\Phi = J_w \bar{V}_w (\Delta p - \Delta\pi) + J_s \left(\bar{V}_s \Delta p + \frac{1}{\bar{c}_s} \Delta\pi \right) \quad (3.142)$$

and after rearrangement:

$$\Phi = \Delta p(J_w \bar{V}_w + J_s \bar{V}_s) + \Delta \pi \left(\frac{J_s}{\bar{c}_s} - J_w \bar{V}_w \right) \quad (3.143)$$

The expressions in parentheses can be regarded as new flux variables. Flux was defined as the amount of a substance in moles that traverses a surface in a unit of time. Multiplying this parameter by the partial molar volume (\bar{V}), one obtains a volume flow. Let us define a total volume flux (\mathbf{J}_V) as the sum of individual volume fluxes of all components:

$$\mathbf{J}_V = \mathbf{J}_w \bar{V}_w + \mathbf{J}_s \bar{V}_s \quad (3.144)$$

To illustrate the meaning of the second term of Eq. 3.143, let us remember the relation $\mathbf{J}_i = c_i \mathbf{v}_i$ (Eq. 3.47). The term: \mathbf{J}_s/\bar{c}_s in Eq. 3.143 therefore is an expression for the velocity of the substance (\mathbf{v}_s). The volume flow of the solvent ($\mathbf{J}_w \bar{V}_w$) can also be considered as the velocity of it (\mathbf{v}_w). From this it follows:

$$\frac{\mathbf{J}_s}{\bar{c}_s} - \mathbf{J}_w \bar{V}_w = \mathbf{v}_s - \mathbf{v}_w \equiv \mathbf{J}_D \quad (3.145)$$

The parameter \mathbf{J}_D , called *exchange flux* is therefore the difference between the velocity of solute relative to the solvent.

As a result of these transformations we finally obtain a dissipation function which contains measurable parameters:

$$\Phi = \mathbf{J}_V \Delta p + \mathbf{J}_D \Delta \pi \quad (3.146)$$

This makes it possible to write down a flux matrix following the pattern of Eq. 3.58:

$$\begin{aligned} \mathbf{J}_V &= L_V \Delta p + L_{VD} \Delta \pi \\ \mathbf{J}_D &= L_{DV} \Delta p + L_D \Delta \pi \end{aligned} \quad (3.147)$$

This equation shows that in the case of the transport of a solution with a single uncharged substance, the membrane permeability is determined by four coefficients: L_V , L_{VD} , L_{DV} , and L_D . The driving forces are the osmotic ($\Delta \pi$) and hydrostatic (Δp) pressures.

Some of the coefficients used in these phenomenological equations can easily be illustrated: The parameter L_V shows for example how fast the solution passes through the membrane in response to a hydrostatic pressure difference (Δp). Introducing the conditions $\Delta \pi = 0$ and $\Delta p > 0$ in the upper part of Eq. 3.147 it becomes: $L_V = \mathbf{J}_V/\Delta p$. This means that L_V is a mechanical *filtration coefficient* or a kind of *hydraulic conductivity*.

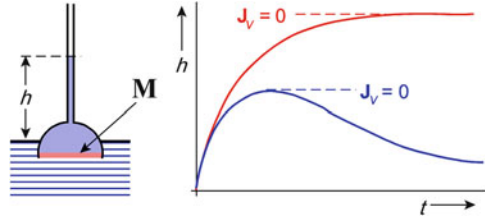


Fig. 3.22 The height of the water column in a Pfeffer osmometer dependent on time in two experiments under different conditions: Pfeffer's cell is closed by a semipermeable membrane (**M**) (*red line*), and: solvent as well as solute can penetrate the membrane (*blue line*). The conditions with stationary pressure in this manometer ($dh/dt = 0$, i.e., $dp/dt = 0$) mean at the same time stationary volume ($dV/dt = J_V = 0$)

Under the same conditions the substance can also be forced through a filter. In this case the second line in Eq. 3.147 gives $L_{DV} = J_D/\Delta p$. Because of this, L_{DV} is called the *ultrafiltration coefficient*.

The flux matrix (3.147) makes it possible to describe the time dependence of an osmotic system. Figure 3.22 shows how the pressure in a Pfeffer's cell, measured by the height of the water column in the vertical tube changes with time. Only in the case of a semipermeable membrane (red line), will a thermodynamic equilibrium with a constant pressure difference be achieved. If in addition to the solvent the membrane allows some of the solute to pass through then there will be a decline in the pressure after the initial rise. The osmotic pressure inside the osmometer will continuously fall (blue line). In this case a state without any volume flux ($dV/dt = J_V = 0$) is achieved only for a short time. For this situation it follows from the first line of the flux matrix (Eq. 3.147):

$$L_V \Delta p + L_{VD} \Delta \pi = 0 \quad (3.148)$$

and furthermore:

$$(\Delta p)_{J_V=0} = -\frac{L_{VD}}{L_V} \Delta \pi \quad (3.149)$$

In Sect. 3.2.2 we introduced Staverman's reflection coefficient (σ). Relating Eqs. 3.108 to 3.149, the following connection between the reflection coefficient and the coupling coefficients results:

$$\sigma = -\frac{L_{VD}}{L_V} \quad (3.150)$$

For a semipermeable membrane with: $\sigma = 1$, it therefore holds that:

$$L_{VD} = -L_V$$

A better illustration of this parameter allows the following consideration: A solution will be forced through a membrane by a hydrostatic pressure ($\Delta p > 0$) without any osmotic difference ($\Delta \pi = 0$). Using these conditions and dividing the second equation of the flux matrix (3.147) by the first one, we obtain:

$$\frac{J_D}{J_V} = \frac{L_{DV}}{L_V} = -\sigma \quad (3.151)$$

According to Eq. 3.145, J_D can be replaced by the difference ($\mathbf{v}_s - \mathbf{v}_w$). For very dilute solutions it holds: $\mathbf{J}_w \bar{\mathbf{V}}_w \gg \mathbf{J}_s \bar{\mathbf{V}}_s$. In this case Eq. 3.151 can be written as:

$$\sigma = \frac{\mathbf{v}_w - \mathbf{v}_s}{\mathbf{v}_w} \quad (3.152)$$

This equation has already been discussed in Sect. 3.2.2 (Eq. 3.109).

The deviations shown here clearly demonstrate that the consideration of more complex systems, for example solutions with more components, would lead to a huge number of parameters and relations.

This theory of coupled fluxes has been widely used to explain the stationary volume and stationary pressure observed in experiments with cells under nonequilibrium conditions. Such experiments have been carried out mostly on red blood cells and on plant cells. Human red blood cells, placed initially in an isotonic solution, shrink immediately when the osmotic pressure of the external solution is increased by adding a permeable substance i ($\sigma_i < 1$). Then however, if the substance passes through the membrane, the cell usually reverts to its initial volume, as illustrated in Fig. 3.22 by the blue line. Such experiments with fast volume changes are undertaken to determine the reflection coefficient using stop-flow techniques.

Further Reading

Katchalsky and Curran 1965; Stein 1990; Zimmermann and Neil 1989. Papers on unstirred diffusion layers: Barry 1998; Evtodienko et al. 1998; Pohl et al. 1998.

3.3.2 Fluxes of Electrolytes

The diffusion of ions is governed by the same fundamental laws as fluxes of uncharged substances (see Eq. 3.125).

$$\mathbf{J}_i = L_i \mathbf{X}_i = \frac{c_i \omega_i}{N} \mathbf{X}_i = -\frac{c_i \omega_i}{N} \text{grad } \tilde{\mu}_i \quad (3.153)$$

In the case of electrolyte fluxes the driving force is induced by the negative gradient of the electrochemical potential ($\tilde{\mu}_i$). The coupling coefficient (L_i) again can be considered as a product of concentration (c_i) and mobility (ω_i/N).

Let us first consider the expression $\text{grad } \tilde{\mu}_i$. For a concentration gradient only in the x -direction, instead of the differential operator *grad*, the differential quotient can be applied:

$$\frac{d\tilde{\mu}_i}{dx} = \frac{d}{dx}(\mu_i^0 + RT \ln a_i + z_i F \psi) \quad (3.154)$$

In order to simplify the equation, let the solution be close to ideal, i.e., let the activity coefficient be: $f_i \approx 1$, and therefore $a_i \approx c_i$. Furthermore, let the system be under isobaric ($\text{grad } p = 0$), and isothermic ($\text{grad } T = 0$) conditions. In this case Eq. 3.154 becomes

$$\frac{d\tilde{\mu}_i}{dx} = \frac{RT}{c_i} \frac{dc_i}{dx} + z_i F \frac{d\psi}{dx} \quad (3.155)$$

which when combined with Eq. 3.153, results in

$$\mathbf{J}_i = -\frac{c_i \omega_i}{N} \left(\frac{RT}{c_i} \frac{dc_i}{dx} + z_i F \frac{d\psi}{dx} \right) \quad (3.156)$$

or modified using the diffusion coefficient according to Eqs. 3.129, 3.130

$$\mathbf{J}_i = -D \left(\frac{dc_i}{dx} + \frac{z_i F c_i}{RT} \frac{d\psi}{dx} \right) \quad (3.157)$$

This is the *Nernst–Planck equation*. It contains the differential quotients of the concentration $[c_i(x)]$, and of the electrical potential $[\psi(x)]$. These differential quotients can be integrated only if the corresponding functions are known. This problem has already been discussed with regard to the concentration $[c_i(x)]$ in Sect. 3.3.1 (see Fig. 3.21). In contrast to the case of Fick's equation, here the function $\psi(x)$ must be known in addition (see Fig. 2.48).

The simplest approach is the consideration of linear gradients. This corresponds for example to a membrane with large, water-filled, and noncharged pores where the ions can move freely as in bulk water. Integrating Eq. 3.157 for these conditions, instead of the differential quotients, simple ratios of differences appear, and instead of concentration c_i the mean concentration of both phases $[\bar{c}_i = (c_i^I + c_i^{II})/2]$ appears.

$$\mathbf{J}_{ix} = -D \left(\frac{\Delta c_i}{\Delta x} + \frac{z_i F \bar{c}_i}{RT} \frac{\Delta \psi}{\Delta x} \right) \quad (3.158)$$

Or, using the permeability coefficient: $P_i = D/\Delta x$ (see Eq. 3.133):

$$\mathbf{J}_{ix} = -P_i \left(\Delta c_i + \frac{z_i F \bar{c}_i}{RT} \Delta \psi \right) \quad (3.159)$$

In 1943, D. E. Goldman integrated the Nernst–Planck equation, supposing only the so-called *constant field conditions*, i.e., assuming: $\mathbf{E} = -\text{grad } \psi = \text{const.}$

The concentration profile results from the bulk concentrations in both phases, and from its passive distribution in the electric field. Integrating the Nernst–Planck equation (3.157) with these conditions, one gets the following expression:

$$\mathbf{J}_i = -P_i \beta \frac{c_i^I - c_i^{II} e^\beta}{1 - e^\beta} \quad \text{with:} \quad \beta = \frac{z_i F}{RT} \Delta\psi \quad (3.160)$$

The function $\mathbf{J} = f(\Delta\psi)$ is illustrated in Fig. 3.23. It considers the flux of a monovalent cation ($z_i = +1$) penetrating the membrane with a permeability $P_i = 10^{-7} \text{ ms}^{-1}$. Let the flux (\mathbf{J}_i) be positive if it is directed phase I \Rightarrow phase II. A negative \mathbf{J}_i therefore, means that the flux is in the opposite direction. The potential gradient is negative per definition, if ψ decreases from phase I to phase II.

The red line in Fig. 3.23 represents a cation flux which is driven only by the electric field without a concentration difference ($c_i^I = c_i^{II} = 100 \text{ mM}$). In this case, the flux vanishes ($\mathbf{J}_i = 0$) if there is no potential difference ($\Delta\psi = 0$).

If there is an additional driving force resulting from a concentration gradient, the curve for example is displaced towards one of the two blue lines. In these cases an ion flux exists even if $\Delta\psi = 0$. If $c_i^I = 100 \text{ mM}$, $c_i^{II} = 10 \text{ mM}$, and $\Delta\psi = 0$, there is a flux from I to II (i. e. $\mathbf{J}_i > 0$). When the concentrations are reversed, \mathbf{J}_i becomes negative according to the given definition. The blue lines cut the abscissa at $+60 \text{ mV}$ and -60 mV , respectively. These points mark where the driving force due to the electric field exactly balances the driving force due to the concentration gradient. This corresponds to the equilibrium situation with $\mathbf{J}_i = 0$, described by the Nernst equation (Eq. 3.113).

In fact, the Goldman equation is a good approach even in cases, where the linearity of the function $\psi(x)$ is not exactly realized. This equation is used frequently in various physiological calculations.

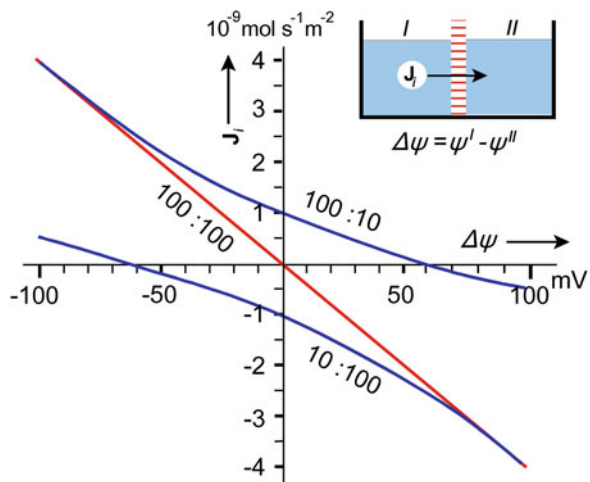


Fig. 3.23 The flux (\mathbf{J}_i) of a univalent cation ($z_i = +1$, $P_i = 10^{-7} \text{ ms}^{-1}$) driven by a concentration, as well as an electrical potential gradient according to the Goldman equation (3.160). The concentrations in phases I and II are listed on the curves. Description in the text

Furthermore, in recent calculations electrostatic interactions of the ions being transported have been taken into account, as well as the fixed charges of the pores. For this, the Nernst–Planck approach is combined with the Poisson equation, predicting the field divergency ($\nabla^2\psi$) as a function of the three-dimensional charge distribution $\rho(x,y,z)$. In this case the generalized Poisson equation (see Sect. 2.2.4, Eqs. 2.51–2.53) must be written as:

$$\nabla^2\psi = \frac{1}{\varepsilon\varepsilon_0} \left(F \sum_{i=1}^n c_i z_i + \rho \right) \quad (3.161)$$

where the sum term gives the charge density of mobile ions in the electrolyte, and ρ represents the fixed charges on the boundary of the pore. This expression can be combined with Eq. 3.157, written also in the general form:

$$\mathbf{J}_i = -D \left(\nabla C_i + \frac{z_i F c_i}{RT} \nabla \psi \right) \quad (3.162)$$

This approach, known as *Poisson–Nernst–Planck theory* (PNP-Theory), can be solved analytically only in some very specialized cases. Using it to calculate transport in discrete ion channels, numerical computer simulations are required.

In any case, the following limitations of these approaches must be taken into account:

- In fact knowledge of the appropriate dielectric constant (ε) (see Eq. 3.161) for channels has not as yet been established. For calculation of the molecular structures of proteins and lipids, usually the dielectric constant of $\varepsilon \sim 2$ is applied, in electrolyte medium that of water: $\varepsilon \sim 80$.
- Special dielectric effects, such as dehydration of the ions inside the channels, specific short-range interactions of ions with polar amino-acid side chains at these locations, etc., are not taken into account.
- The problem is reduced to a steady-state situation. Ion permeation in channels, however, is basically a time-dependent process. As an ion moves through the channel, the forces acting on it change.

In addition to these new approaches, a “classical” one should not be forgotten: In 1949 Hans Ussing introduced an equation which has been used even in recent physiological publications. It makes it possible to calculate the ratios of ionic fluxes independently of the functions $c(x)$ and $\psi(x)$.

In this context a term must be defined that we will need in further considerations: The flux that is measurable directly through changes of chemical concentrations is called the *net flux* (\mathbf{J}). Using radioactive isotopes it is possible to indicate that this net flux in fact results from the difference of two opposite *unidirectional fluxes* \mathbf{J}_{12} and \mathbf{J}_{21} :

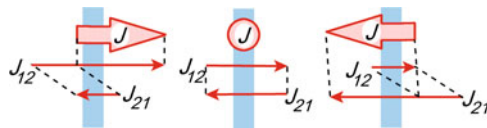


Fig. 3.24 The net flux (\mathbf{J}), as the sum of the unidirectional fluxes \mathbf{J}_{12} and \mathbf{J}_{21}

$$\mathbf{J} = \mathbf{J}_{12} - \mathbf{J}_{21} \quad (3.163)$$

In contrast to the unidirectional fluxes, the direction of which is indicated by the subscripts (for example: \mathbf{J}_{12} , \mathbf{J}_{21}), the net flux will be defined as positive, as in the example of Fig. 3.24, if it takes place in the direction corresponding to Eq. 3.163.

It must be noted that these are examples of simple diffusion processes based on thermodynamic fluctuations. Later in Sect. 3.4.1 (see Fig. 3.25) we will consider systems of co-transport through biological membranes, which shows the same kinetics but which must be treated thermodynamically with completely different approaches. These two kinds of transport processes therefore must not be mixed.

Considering Eq. 3.163, one can write Fick's equation (Eq. 3.133) in the following way:

$$\mathbf{J} = \mathbf{J}_{12} - \mathbf{J}_{21} = -P \Delta c = P(c^I - c^{II}) = P c^I - P c^{II} \quad (3.164)$$

From this we can conclude with certain justification:

$$\mathbf{J}_{12} = P c^I; \quad \mathbf{J}_{21} = P c^{II} \quad (3.165)$$

This, by the way, resembles the approach of compartment analysis and reactions of the first order as will be used in Sect. 5.1.1.

To calculate unidirectional fluxes of ions one could use the same approaches as in Eq. 3.165, but just introducing a kind of electrochemical concentration (\tilde{c}). This parameter, as we will indicate later, has no further relevance. The definition results from the following equation:

$$\tilde{\mu} = \mu^0 + RT \ln c + zF\psi \stackrel{!}{=} \mu^0 + RT \ln \tilde{c} + zF\psi^0 \quad (3.166)$$

Therefore: $\tilde{c} = c$, if $\psi^0 = \psi$. This means that a kind of zero-potential will be established, whatever it is. From this definition follows:

$$\ln \tilde{c} = \ln c + \frac{zF}{RT}(\psi - \psi^0) \quad (3.167)$$

and:

$$\tilde{c} = c e^{\frac{zF}{RT}(\psi - \psi^0)} \quad (3.168)$$

This equation is of little use since the reference potential ψ^0 is unknown. However, if this relation is substituted in Eq. 3.165, and if the ratio of the unidirectional fluxes are calculated, all unknown parameters cancel:

$$\frac{J_{12}}{J_{21}} = \frac{c^I}{c^{II}} e^{\frac{zF}{RT}(\psi^I - \psi^{II})} \quad (3.169)$$

(To avoid the accumulation of subscripts we ignored in these derivations the index for the specific substance i).

This formula is known as *Ussing's equation* or *Ussing's flux ratio criterion* which relates unidirectional fluxes to concentrations and electrical potential differences. All parameters of this equation can be measured and the validity of the equation therefore, can be checked experimentally. As preconditions for this equation only differences of concentrations and electrical potentials are considered. If the relation between two measured unidirectional fluxes does not agree with the results calculated by these gradients, then additional driving forces are involved. This could be a hint at coupling processes or of any kind of active transport processes.

Further Reading

Goldman 1943; Kontturi et al. 2008; Roux et al. 2004; Syganov and von Klitzing 1999; Ussing 1949; Zhou and Uesaka 2009.

3.3.3 The Diffusion Potential

The diffusion of an electrolyte, in general, can be considered as a separate movement of the dissociated ions along their particular electrochemical gradient, coupled, however, with the electric field resp. the potential difference $\Delta\psi$. Cations and anions, however, may have different mobilities (w_i). The slower diffusion types of ions will lag behind the faster ones. This produces a local potential difference called the *diffusion potential*, retarding the faster ions and speeding up the slower ones. Diffusion potentials can occur in homogenous solutions as well as between two phases separated by a membrane which is permeable for the ions.

An equation for the diffusion potential can be derived, postulating the electroneutrality of the sum of all ion fluxes. In this case the cation flux (J_c) induces a flow of charge (J_{cz_c}). In the case of electroneutrality, it must be compensated by the charge transport of anions (J_{Az_A}):

$$J_{cz_c} + J_{Az_A} = 0 \quad (3.170)$$

Now, we can use the flux equations derived in the previous Sect. 3.3.2. If linear functions $c(x)$ and $\psi(x)$ are proposed, the simplest flux equations (Eq. 3.159) can be used. Inserting them into Eq. 3.170, one obtains:

$$z_C P_C \Delta c_C + \frac{z_C^2 F \bar{c}_C P_C}{RT} \Delta \psi + z_A P_A \Delta c_A + \frac{z_A^2 F \bar{c}_A P_A}{RT} \Delta \psi = 0 \quad (3.171)$$

Let us take into account that the concentration of the ions (c_C , c_A) depends on the concentration of the salt (\bar{c}), whereas $c_C = v_C \bar{c}$, and $c_A = v_A \bar{c}$. Introducing this, one can rearrange Eq. 3.171 and resolve it for $\Delta \psi$ in the following way:

$$\Delta \psi = -\frac{RT}{F} \left(\frac{z_C P_C v_C + z_A P_A v_A}{z_C^2 P_C v_C + z_A^2 P_A v_A} \right) \frac{\Delta c}{\bar{c}} \quad (3.172)$$

or:

$$\Delta \psi = \frac{RT}{F} \left(\frac{z_C P_C v_C + z_A P_A v_A}{z_C^2 P_C v_C + z_A^2 P_A v_A} \right) \ln \frac{c^I}{c^II} \quad (3.173)$$

The relation: $\Delta c/\bar{c} = \ln(c^I/c^{II})$ has already been introduced in Sect. 3.3.1 in context with Eq. 3.140.

It is easy to understand that Eq. 3.173 will be transformed into the Nernst Equation (Eq. 3.112), if the membrane becomes semipermeable, i.e., if $P_C = 0$, or if $P_A = 0$.

A better approach for the conditions of membranes will be the flux equation, derived by Goldman for conditions of constant electric field (see Sect. 3.3.2). Introducing this Goldman equation (Eq. 3.160) into the equation for electroneutrality of fluxes (Eq. 3.170), one gets the following expression:

$$P_C \frac{F \Delta \psi}{RT} \left(\frac{c_C^I - c_C^{II} e^{\frac{F \Delta \psi}{RT}}}{1 - e^{\frac{F \Delta \psi}{RT}}} \right) + P_A \frac{F \Delta \psi}{RT} \left(\frac{c_A^I - c_A^{II} e^{-\frac{F \Delta \psi}{RT}}}{1 - e^{-\frac{F \Delta \psi}{RT}}} \right) = 0 \quad (3.174)$$

This equation can also be rearranged and solved for $\Delta \psi$. For this we first transform the denominator of the fractions, using the expression:

$$1 - e^{-x} = -e^{-x}(1 - e^x)$$

This leads to:

$$\frac{F \Delta \psi}{RT \left(1 - e^{\frac{F \Delta \psi}{RT}} \right)} \left[P_C \left(c_C^I - c_C^{II} e^{\frac{F \Delta \psi}{RT}} \right) - P_A e^{\frac{F \Delta \psi}{RT}} \left(c_A^I - c_A^{II} e^{-\frac{F \Delta \psi}{RT}} \right) \right] = 0 \quad (3.175)$$

When $\Delta \psi \Rightarrow 0$ the expression in front of the square parentheses will not approach zero. Therefore, the sum inside the parentheses must be equal to zero:

$$P_C c_C^I - P_C c_C^{II} e^{\frac{F \Delta \psi}{RT}} - P_A c_A^I e^{\frac{F \Delta \psi}{RT}} + P_A c_A^{II} = 0 \quad (3.176)$$

Which gives after some rearrangements:

$$\frac{F\Delta\psi}{eRT} = \frac{P_A c_A^I + P_C c_C^I}{P_A c_A^J + P_C c_C^J} \quad (3.177)$$

and:

$$\Delta\psi = \frac{RT}{F} \ln \frac{P_A c_A^I + P_C c_C^I}{P_A c_A^J + P_C c_C^J} \quad (3.178)$$

This is the *Goldman–Hodgkin–Katz equation* which is commonly used in electrophysiology to calculate diffusion potentials in living cells (mostly it is just named the *Goldman equation*). This expression also becomes a Nernst equation (Eq. 3.112), introducing the conditions of a semipermeable membrane ($P_A = 0$, or $P_C = 0$).

It is possible to extend this equation also for systems containing more than one salt, i.e., various monovalent ions. To take into account the nonideal character of solutions, chemical activities (a_i) instead of concentrations (c_i) can be used. In this case the *Goldman–Hodgkin–Katz equation* can be written as follows:

$$\Delta\psi = \frac{RT}{F} \ln \frac{\sum_{\text{Anions}} P_A a_A^i + \sum_{\text{Cations}} P_C a_C^e}{\sum_{\text{Anions}} P_A a_A^e + \sum_{\text{Cations}} P_C a_C^i} \quad (3.179)$$

When considering cells, the superscript i in this formula means “internal,” the superscript e – “external” concentrations. Correspondingly: $\Delta\psi = \psi^i - \psi^e$.

In a large number of papers experiments are described indicating the applicability of the Goldman equation for various cells and membranes. Some limitations are obvious, however, when this equation is applied to real cellular conditions. Mostly these are already included in the conditions of the applied flux equation as discussed previously.

The most important limitation comes from the assumption considering a free diffusion of ions in a homogeneous and constant electric field. Even in the case of large pores in the membrane this assumption is valid only with some approximations. It must be taken into account that the thickness of a biological membrane is just as large as the Debye–Hückel radius of an ion. Additionally, the coefficients of permeability are defined for large systems and can be used for considerations of molecular systems only approximately.

As mentioned in Sect. 3.3.2, there exist a number of approaches that consider local concentrations of ions directly at the membrane boundary, to calculate fluxes and transmembrane potentials. This takes into account the influence of surface potentials and electric double layers (see Sect. 2.3.5). It is possible to introduce surface concentrations into the Goldman equation, instead of the concentration of ions in the bulk phase, using Eq. 2.10 or Eq. 3.113, if the surface potential ψ_0 is known. In this case, however, the value $\Delta\psi$ of Eq. 3.179 no longer means the

potential difference between the two bulk phases, as measured in electrophysiology by microelectrodes, but just between the inner and outer surface (see Fig. 2.48). Thus, this is an indication that the diffusion potential of a membrane is in fact controlled by local surface charges.

Further Reading

Goldman 1943; Katchalsky and Curran 1965; Syganov and von Klitzing 1999; Fraser and Huang 2007.

3.4 Membrane Transport and Membrane Potential

The living cell and its internal compartments maintain a particular electrolyte state which on the one hand guarantees constant conditions for all enzymatic processes, whilst on the other hand, it acts as an accumulator of electrochemical energy. This requires a complicated system of transporters in the membranes, which are particularly specialized and precisely controlled by a system of interactions. Furthermore, a number of receptor proteins transform external signals to internal information by modification of their transport properties. In a highly specialized way, nerve and muscle cells used the accumulated electrochemical energy for processes of excitation. In this chapter we will concentrate just on transport of ions with regard to transport of metabolites.

Numerous experimental investigations indicate an enormous variety of transporters even in a single cell. This research during the second half of the last century was supported strongly by the introduction of radioactive isotopes, later on by fluorometric methods and especially by the use of patch-clamp measurements. In the last decades, our knowledge on the molecular mechanisms of these transporters and their control has developed rapidly thanks to studies involving X-ray crystal analyses, and various high-resolution functional measurements.

3.4.1 *Channels and Pumps: The Variety of Cellular Transport Mechanisms*

Figure 3.25 illustrates a functional classification of various types of ion transport mechanisms in the membrane of cells and organelles. The differentiation between *pores* and *channels* is rather unclear. Mostly the term “pore” is used to denote larger membrane openings with low selectivity, which are for example produced by electric pulses (electric break down, see Sect. 3.5.5) or by other influences. In contrast “channels” are protein transporters, which are characterized by a certain selectivity. In any case, fluxes through these kinds of transporters are governed by the laws of electrodiffusion.

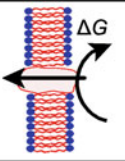
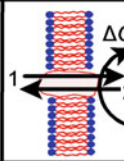
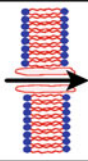
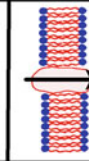
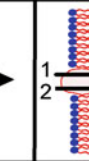
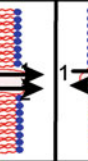
	Passive transport				Primary active transport	
	Diffusion		Cotransport			
	PORE	CHANNEL	SYMPORT	ANTI-PORT		
						
Electro-neutral $v_1 z_1 = v_2 z_2$			$\text{Na}^+, \text{K}^+ - 2\text{Cl}^-$ $\text{K}^+ - \text{Cl}^-$ $\text{Na}^+ - \text{Cl}^-$	$\text{Cl}^- - \text{HCO}_3^-$ $\text{Cl}^- - \text{Cl}^-$ $\text{Na}^+ - \text{H}^+$ $\text{K}^+ - \text{H}^+$		$\text{Na}^+ - \text{H}^+$ -ATPase
Rheogen $v_1 z_1 \neq v_2 z_2$	$\text{Na}^+, \text{K}^+, \text{Cl}^-$	$\text{Na}^+, \text{K}^+, \text{Cl}^-$	$\text{Na}^+ -$ - Glucose	$3\text{Na}^+ - \text{Ca}^{++}$ $\text{H}^+ - \text{Ca}^{++}$	$\text{H}^+ - \text{ATPase}$ $\text{Ca}^{++} - \text{ATPase}$	$3\text{Na}^+ - 2\text{K}^+$ -ATPase

Fig. 3.25 Classification of various systems of ion transporters in biological membranes, including particular examples

Another type of passive transporters is so-called *carriers* or *porters* transporting simultaneously two or more ions in a well-defined stoichiometric relation. Such stoichiometrically coupled fluxes are called *co-transporters*. There are two kinds of co-transport systems: In the case of the *symport*, a strongly coupled flux of two species in the same direction occurs. An example of this could be a complex that simultaneously transfers one Cl^- and one K^+ ion through the membrane in the same direction. In the same way, transport of an ion could be coupled to an uncharged molecule, like glucose. An *antiport*, in contrast to this, is a system simultaneously transporting two ions with identical charges in opposite directions, for example one K^+ , against one H^+ .

Co-transport systems are electroneutral, if an equal number of charges is transported, either of opposite sign in the case of a symport, or of the same sign in antiports. In this case the flux does not depend directly on electric field conditions. It is electrically silent, i.e., it cannot be identified by electrophysiological methods. In cases of unequal charge transporters, an electrical current will be the result of the transport. We will call this type of process *rheogenic*, i.e., “current producing.” Rheogenic co-transport processes can be recognized by their electrical conductivity, a property which they have in common with simple diffusion processes. They can be controlled by electric fields, especially by the transmembrane potential.

An *active transport* is a sort of pump, transporting ions or even uncharged molecules against their own chemical or electrochemical gradients. Therefore, it is an “uphill transport,” using metabolic energy (ΔG , in Fig. 3.25). In most cases these are so-called transport ATPases, using the energy of the hydrolytic reaction: $\text{ATP} \Rightarrow \text{ADP}$. Furthermore, ionic pumps are known which are driven by other

sources of energy, such as for example decarboxylation, oxyreduction, or even the quantum energy of light. Some of these mechanisms can also run in the opposite direction. In chloroplasts and mitochondria, for example, ATP is synthesized by a “downhill” proton flux (see Sect. 4.8.3, Fig. 4.36).

Active transport can also be rheogenic. In this case the transport directly induces electric currents, like for example the Na-K-ATPase, transporting three charges from the inside out, but only two in the opposite direction, or a Ca^{++} -ATPase (see Sect. 3.5.2, Fig. 3.35). Frequently such transports are also called *electrogenic* which means: “generating an electrical membrane potential.” Looking at the terms “rheogenic” and “electrogenic” accurately, they are however not identical. Even an electro-neutral pump can be “electrogenic” if it produces a concentration gradient of ions which subsequently generates a diffusion potential. Conversely, a rheogenic pump may influence the transmembrane potential only if there is a sufficiently high membrane resistance.

This leads to the differentiation between *primary* and *secondary* active transporters. An example of a primary active transporter is the Na-K-ATPase, where the uphill flux of ions is directly driven by a biochemical process. In contrast, secondary active transporters exploit the energy already stored in the electrochemical gradient of one species to drive the uphill transport of another substrate. This can be realized by various kinds of symporters or antiporters. As an example in Figs. 3.25 and 3.26 the co-transport of Na^+ with glucose is shown. It is “secondary active,” because in fact the uphill glucose uptake is driven by the downhill Na^+ -flux in a gradient, produced by the Na-K-ATPase. In a similar way fluxes of amino acids are coupled with transport of Na^+ or H^+ ions.

The number of different transport paths in a single membrane can be rather high. In Fig. 3.26 this is illustrated for the case of cells of a renal proximal tubule. It is

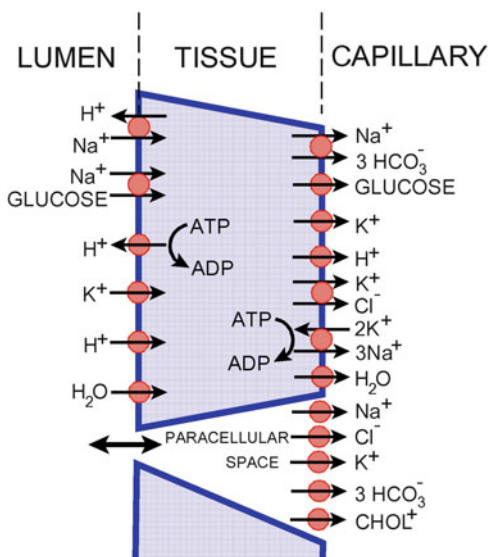


Fig. 3.26 Various types of ion transporters in cells of a renal proximal tubule and in the paracellular space (After Verkman and Alpern 1987)

obvious that the fluxes are coupled with each other by the transmembrane potential as well as by the concentrations of their common ions. Additionally, changes of fixed charges inside the cell induced by internal pH changes need to be taken into account.

The existence of transporters where the participants have strongly fixed stoichiometry forces us to rethink the requirement of flux electroneutrality which we postulated in Sect. 3.3.3 (Eq. 3.170). Considering rheogenic symports, it is not the electroneutrality of a single flux that is required, but rather the electroneutrality of all fluxes in the membrane of a single cell together. The calculation of the balance of charges and ions in a cell is therefore only possible by considering all fluxes. This type of coupling can formally be calculated using the flux matrix as discussed in Sect. 3.1.3.

The existence of co-transporters in a cell rather than simple diffusion processes can be regarded as a form of optimization. Ionic transport, based on electrodiffusion, strongly depends on the transmembrane potential. An alteration of the transmembrane potential would cause an immediate change of electrolyte fluxes in the whole cell, and subsequently a shift in the internal concentration of all ions. In contrast, the system of electroneutral co-transporters is independent of the transmembrane potential and will protect the cell against such disturbances.

Further Reading

Läuger 1991; Luckey 2008.

3.4.2 *The Network of Cellular Transporters*

If a cell were only a poly-electrolyte system without metabolically driven ion pumps it would remain in a state of Donnan equilibrium. This means that there would be a Donnan distribution of all mobile ions according to fixed charges, and as a result, a Donnan osmotic pressure (see Sects. 3.2.4, 3.2.5). In the living cell however, active transport systems driven by metabolic energy (Fig. 3.25) modify this ionic composition, as shown schematically in the model of Fig. 3.4b. The living cell therefore reaches a steady state, i.e., a stationary state of nonequilibrium in general (see Fig. 3.6), and for particular ionic species.

This nonequilibrium state has manifold functions. In general the cell can be regarded as a kind of electrochemical energy store which may be easily tapped. This, for example, is the case in electrical membrane de- and repolarizations (see Sect. 3.4.4). Furthermore, the nonequilibrium state of a system is the precondition for its homeostatic regulation. This, by the way, is also the reason for the increased temperature in homeothermic animals. The setting up of a concentration gradient of ions across the membrane makes the cells able to control and regulate an intracellular environment, which is the precondition of various cellular processes. In the case of Ca-ATPase an effective signal system is established. This pump creates an extremely low calcium level in the cytoplasm which is of the order of 10^4 times lower than the concentration in the extracellular fluid. In this way an important

signal transduction pathway is established, which can be triggered even by a minimal increase in the Ca-permeability of the membrane. The cytoplasmic Ca^{++} -ions act as a second messenger, in a number of cellular functions.

What therefore are the immediate effects of ionic pumps on the cell?

- They control and regulate the internal ionic milieu. In this way, steep gradients of the electrochemical potentials of particular ions are built up, essentially without changing the total internal ionic concentration. The internal potassium concentration of animal cells, for example, is usually much higher than the external one. Simultaneously however, the sodium concentration is lower to the same degree. The sum of both of these ions in the cytoplasm, taken together, is nearly the same as in the external medium.
- In the case of rheogenic pumps, they directly induce transmembrane potentials. In this case the pumps are called electrogenic.
- They can produce a direct osmotic effect changing the concentration of osmotically active substances.
- They can establish particular internal ionic conditions, controlling, for example, the extremely low intracellular calcium concentration.

Some direct consequences of the active transport processes can be demonstrated by the effects of stopping the pump through the use of specific inhibitors. In this case effects can be observed like Donnan-osmotic swelling, internal pH shifts, an increase in the internal calcium concentration, a change of transmembrane potential, etc. Mostly, using such inhibitors, the overall internal ionic conditions are altered.

As an example, the system of transport processes in kidney tubule cells is illustrated in Fig. 3.24. There are 13 different transport systems shown which determine the cellular milieu and additionally five other fluxes between the luminal and serosal surfaces of the epithelium across the paracellular gap. This picture in fact is incomplete as, for example, Ca^{++} fluxes are not shown, and the diagram does not include the intracellular organelles with their own transporters.

Using this example we will illustrate the interconnections of these transport properties qualitatively, following for example one particular path: Transport ATPases pump protons out of the cell, others decrease the internal sodium content, and in the same way enrich the cytoplasm with potassium. Extruding positive charges, both primary active transporters induce an inside negative transmembrane potential. Simultaneously, an electrochemical sodium gradient was generated which drives a sodium flux outside-in. This influx, however, is realized by a glucose-sodium co-transporter and acts therefore as a secondary active transporter for glucose entry. The glucose finally diffuses via its own concentration gradient on the opposite side of the cell layer from the cytoplasm into the capillary.

All these manifold transporters occurring in a single cell respond to different stimulants. Some of them become active only if a particular pH exists, others if the internal calcium concentration was increased. There are voltage-sensitive transporters responding to particular transmembrane potentials, or others that respond to mechanical stress of the membrane or to minimal temperature changes

(see Sect. 4.1). The electroneutral Na^+H^+ antiporter, which is present in most animal cells, merits particular attention. Under physiological conditions, at neutral pH_i it is inactive. However, if the internal pH increases, it becomes activated. This property qualifies it to be a volume-regulating system. This mechanism was demonstrated in the case of lymphocytes. It has also been shown that this Na^+H^+ antiporter can be activated by a multitude of substances including hormones, growth factors, lectins, etc. These substances alter the above-mentioned pH threshold. This seems to be an important control mechanism for the regulation of complex biological phenomena.

Beside direct calculations of flux coupling, the equations of nonequilibrium thermodynamics can be applied to describe the energy balance of primary and secondary active transport. As an example the energy balance at steady state of the above-mentioned Na^+ -Glucose symport will be evaluated. This is an example of a steady-state system like that of Fig. 3.4b which is determined by the active transport (\mathbf{J}_A) as well as by the passive flux (\mathbf{J}_i). In Sect. 3.1.4 we introduced the dissipation function $\Phi = \sigma T$ (Eq. 3.64), which must be larger than 0. According to Eq. 3.64 for our system it amounts to

$$\Phi = \mathbf{J}_A \mathbf{X}_A + \mathbf{J}_i \mathbf{X}_i \quad \text{for: } \Phi > 0 \quad (3.180)$$

In our particular case the glucose uptake (\mathbf{J}_G) is driven by the passive influx of sodium (\mathbf{J}_{Na}), driven by its electrochemical gradient. Corresponding to Eq. 3.180 this results in:

$$\mathbf{J}_G \mathbf{X}_G + \mathbf{J}_{\text{Na}} \mathbf{X}_{\text{Na}} > 0 \quad (3.181)$$

If v equivalents of sodium ions are transported for each mole of glucose then:

$$\mathbf{J}_G = v \mathbf{J}_{\text{Na}} \quad (3.182)$$

Introducing this in Eq. 3.181 and considering that both fluxes are not equal to zero, it follows that:

$$v \mathbf{X}_G + \mathbf{X}_{\text{Na}} > 0 \quad (3.183)$$

respectively:

$$\mathbf{X}_{\text{Na}} > -v \mathbf{X}_G \quad (3.184)$$

Let us now replace the forces (\mathbf{X}) by the differences of the corresponding chemical, resp. electrochemical potential (see Sect. 3.3.1), we obtain:

$$-v \Delta \mu_G < \Delta \tilde{\mu}_{\text{Na}} \quad (3.185)$$

Using Eqs. 3.33 and 3.41, and the conditions: $\Delta T = 0$ and $\Delta p = 0$, we get:

$$vRT \ln \frac{a_G^i}{a_G^e} < - \left(RT \ln \frac{a_{Na}^i}{a_{Na}^e} + F \Delta\psi \right) \quad (3.186)$$

(where $\Delta\psi = \psi_i - \psi_e$) and after rearrangement:

$$\left(\frac{a_G^i}{a_G^e} \right)^v < \frac{a_{Na}^e}{a_{Na}^i} e^{-\frac{F \Delta\psi}{RT}} \quad (3.187)$$

This equation allows us to calculate the maximal rate of enrichment of glucose in the cell that can be achieved for a given electrochemical gradient of sodium ions. Assuming that the membrane potential of the cell is: $\Delta\psi = -50$ mV, and the relation of sodium ions: $a_{Na}^i/a_{Na}^e = 10$ ($T = 300$ K), it follows:

$$\left(\frac{a_G^i}{a_G^e} \right)^v < 69 \quad (3.188)$$

If the fluxes are coupled 1:1 ($v = 1$), this process gives a maximum enrichment of glucose by a factor of 69, when the pump is performing optimally.

Similar calculations can be applied to primary active transports, i.e., those that are driven by chemical reactions, for example transport ATPases. In this case in the equation of the dissipation function (Eq. 3.180), the reaction rate (as a type of scalar flux), and the chemical affinity of the energy supplying reaction (Eq. 3.75) must be included.

The calculation of the intensity of a pump which is necessary to build up a certain concentration gradient depends both on the coupling stoichiometry of the fluxes and on the passive back flow. This means that not only the power of the pump is responsible for the steady-state level achieved, but also the conductivity, resp. the permeability of the considered substance, leading it to flow backwards. This is illustrated in the scheme shown in Fig. 3.4b: the power of the pump must be higher if a greater difference in the levels of the vessels is reached, and if the outflow becomes faster.

Further Reading

Luckey 2008.

3.4.3 The Membrane Potential

As outlined in the previous section, the pumps lead to gradients of ion concentrations and therefore accumulate electrochemical energy. Now we will discuss how the cell generates an electrical membrane potential, using this accumulated energy.

First it is necessary to remember the general definition of electrical potential as defined in Sect. 2.2.1. According to this, the electrical potential $[\psi(x,y,z)]$ is a scalar state parameter in three-dimensional space, similar to temperature (T) or pressure (p). Mostly as a simplification the function $\psi(x)$ is used to characterize the potential along a line that runs perpendicularly through the membrane (Figs. 2.15, 2.48). As the *transmembrane potential* ($\Delta\psi$) the potential difference is defined between two points, one on the inside, the other on the outside of the membrane, each at a suitable distance from it (Fig. 2.48). The sign of this difference results from its definition:

$$\Delta\psi = \psi_i - \psi_e \quad (3.189)$$

Note that terms such as Donnan potential, diffusion potential, Nernst potential, are just expressions describing the *mechanisms* which can give rise to the electrical transmembrane potential and do not refer in any way to different kinds of electrical potentials that might exist simultaneously. In fact there is only one electrical potential $\psi(x,y,z, t)$ at a given point in the space (x, y, z), and at a given time (t). In Fig. 2.48, the function $\psi(x)$ illustrates this in a very simplified way. It includes the transmembrane potential and the two surface potentials at both boundaries.

We have already learned that processes of active transport can be rheogenic (Fig. 3.25). If the so-far transported charges can be rapidly neutralized by other fluxes, for example by Cl^- exchange in the membrane of human erythrocytes, then a rheogenic pump has no direct electrical consequences for the cell. If however, no such short-circuit flux exists, the transported net charges build up a transmembrane potential, and the *rheogenic* pump becomes *electrogenic*.

In any case, the Na-K-ATPase, occurring in nearly all cell membranes, generates an electrochemical gradient of sodium and potassium. For most animal cells a relation near 1:10 occurs for $a_K^i > a_K^e$ and $a_{\text{Na}}^i < a_{\text{Na}}^e$. Chloride ions are distributed mostly passively, according to the Nernst equation. This nonequilibrium distribution of the cations can lead to a diffusion potential which can be calculated by the Goldman equation (Eq. 3.179) as follows:

$$\Delta\psi = \frac{RT}{F} \ln \frac{P_{\text{Cl}} a_{\text{Cl}}^i + P_{\text{K}} a_{\text{K}}^e + P_{\text{Na}} a_{\text{Na}}^e}{P_{\text{Cl}} a_{\text{Cl}}^e + P_{\text{K}} a_{\text{K}}^i + P_{\text{Na}} a_{\text{Na}}^i} \quad (3.190)$$

Even if the internal ion activities a_K^i and a_{Na}^i remain constant, the diffusion potential ($\Delta\psi$) can vary widely because of changing permeabilities (P_i). The limits of such variations can be easily obtained from Eq. 3.190:

For $P_{\text{K}} \gg P_{\text{Na}}, P_{\text{Cl}}$ Eq. 3.190 reduces to:

$$\Delta\psi_{\text{K}} = \frac{RT}{F} \ln \frac{a_{\text{K}}^e}{a_{\text{K}}^i} \quad (3.191)$$

and for $P_{\text{Na}} \gg P_{\text{K}}, P_{\text{Cl}}$ it follows:

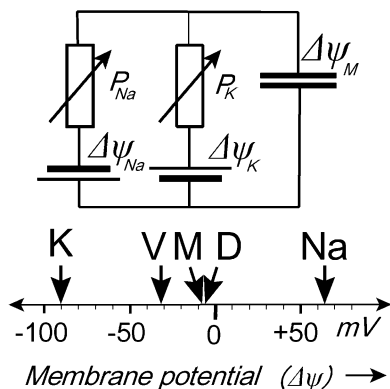


Fig. 3.27 An electrical circuit as a model illustrating the $\text{Na}^+ \text{-} \text{K}^+$ diffusion potential of a cell as the result of a sodium ($\Delta\psi_{Na}$), and a potassium ($\Delta\psi_K$) battery. In the lower part of the figure, possible potential alterations are illustrated for the case of human erythrocytes in a solution containing 145 mM NaCl and 5 mM KCL. **K** – $\Delta\psi_K$, **V** – valinomycin-induced diffusion potential, **M** – potential of untreated erythrocytes corresponding to $\Delta\psi_M$, **D** – position of the Donnan potential, **Na** – $\Delta\psi_{Na}$

$$\Delta\psi_{Na} = \frac{RT}{F} \ln \frac{a_{Na}^e}{a_{Na}^i} \quad (3.192)$$

For these particular cases the Goldman equation (Eq. 3.190), therefore, reduces to a Nernst equation (Eq. 3.112) which was derived for such kinds of semipermeable membranes. If the typical relations of activities for sodium and potassium, as mentioned before, are inserted into Eqs. 3.191 and 3.192, then it is easy to understand that $\Delta\psi_K < 0$ and $\Delta\psi_{Na} > 0$.

This situation is illustrated in Fig. 3.27. The electrochemical gradients of potassium and sodium which are generated using metabolic energy can be considered as storage batteries, or electrical accumulators having opposite polarities. The permeability characteristics of the ions are expressed in this model as conductivities of the variable resistors, or potentiometers through which these accumulators are discharged. If the resistance is low, then a large discharge current would flow, and if the accumulator is not recharged continuously, it would soon be empty. In fact, the permeabilities P_{Na} and P_K are usually so low that the electrochemical gradient of the living cell persists for hours or even days. The effective membrane potential in this model is represented by the voltage difference across the capacitor $\Delta\psi_M$. This capacitor represents the capacity of the membrane (see Sect. 2.3.6). If P_{Na} and P_K have about the same value, then $\Delta\psi_M$ will be very small. If they differ, a membrane potential will be established according to Eqs. 3.191 and 3.192.

Figure 3.27 demonstrates membrane potentials that can be induced in human erythrocytes. In this case the Nernst potentials for potassium and sodium give the limits of these possible shifts. They range approximately between -95 mV and

+65 mV. The actual membrane potential of human erythrocytes in vivo is found to be -9 mV (**M**), and is only a little greater than the Donnan potential (**D**) which would result if the cell achieved a thermodynamic equilibrium (see Fig. 3.19). If the cells are treated with valinomycin, the membrane potential falls to about -35 mV (**V**). Valinomycin is an ionophore that is rapidly incorporated into the membrane causing a highly selective increase of potassium permeability. It will not reach the limiting value of the Nernst potential of potassium, because the values of P_{Cl} and P_{Na} are not negligible, as was assumed for Eq. 3.191. However, it is shifted in this direction.

Even if these types of potential alterations are possible without a significant change of concentration profiles, they must in fact be accompanied by a certain transmembrane shift of charges. It is easy to show that this charge flux is extremely small. For this we calculate the charge transfer across the membrane capacitor, which is required to adjust these potential differences ($\Delta\psi_M$ in Fig. 3.27). Let us ask the question: how many charges must be displaced in the cell membrane with a specific capacity of 10^{-2} F m^{-2} (see Sect. 2.3.6) in order to generate a transmembrane potential $\Delta\psi_M = 0.1$ V?

Equation 2.90 gives the corresponding relation for a capacitor. This enables us to calculate the surface charge density (σ) as a function of the transmembrane potential ($\Delta\psi$) and specific capacity (C_{sp}):

$$\sigma = C_{sp}\Delta\psi = 10^{-3} \text{ C m}^{-2} \quad (3.193)$$

This value can be converted into charge equivalents of ions, using the Faraday constant (F):

$$\frac{\sigma}{F} = \frac{10^{-3}}{9.65 \cdot 10^4} \approx 10^{-8} \quad \text{charge equivalents} \cdot \text{m}^{-2}$$

The resulting charge density, so far, is very small. Considering a certain geometry of the cell, for example a sphere, or in the case of a neuron, a cylinder, one can easily transform this number into a concentration shift. The result will be a fully negligible proportion of the amount of internal ions.

This example demonstrates a most important element in the functional arrangement of the living cell: An ion pump driven by metabolic energy, accumulates electrochemical energy by generating a concentration gradient of sodium and potassium. This electrochemical energy can be converted into electrical energy altering the membrane permeabilities (for example: P_K and P_{Na}). In this way a wide-ranging control of the electric field in the cell membrane is possible. Even if the shift of the membrane potential amounts to only some tenths of a millivolt, the resulting variations of the field strength, sensed by the membrane proteins, are of the order of millions of volts per meter (see Sect. 2.2.1)! It must be emphasized that this control is possible without any sizeable input of energy and can be realized in milliseconds. Such permeability changes can be induced by the cell itself as well as by external influences.

As mentioned before, there are many ion-selective transporters in the cell which are controlled by internal calcium concentration, by internal pH, by mechanical tension of the membrane, or by modifications of other parameters. Diffusion potentials may also result from an interaction between the cell and specific drugs, or may be triggered locally through mechanical contacts with surfaces or particles, such as for example viruses. These alterations of membrane potentials caused by local permeability changes can induce electric potential differences and therefore electric fields not only in the x -direction, perpendicular to the membrane surface, but also in the y -, z -direction, i.e., in the plane of the membrane (see Sect. 3.5.2).

In the next section we will consider the action potential of nerve cells as a classical example of the feedback loop between an electric field and ionic permeability in more detail. Recently, the interest in the transmembrane potential of the cell as a regulator of cellular events has greatly increased. This concerns the size of the membrane potential in various cells, as well as its time dependence. Although action potentials have a special significance in signal transfer of neurons, they occur also in many other cells.

Although opening to particular transporters or integration of specific channels in the membrane may always modify the membrane potential by generating diffusion potentials, the resting potential of many cells is exclusively generated by electrogenic pumps. In this case transmembrane potentials appear to be independent of external potassium concentrations. Inhibition of the pumps in this case immediately leads to changes of $\Delta\psi$ (see Bashford and Pasternak 1986).

In Fig. 3.28 correlations of membrane potential and the state of various animal cells are illustrated. In contrast to cells with active proliferation like cancer cells or cells of embryos, indicating a transmembrane potential between -10 and -30 mV, nondividing cells, like neurons or skeletal muscle cells show membrane potentials between -70 and 90 mV. The transmembrane potential of cells which pass through a state of proliferation falls before mitosis takes place. It is not yet clear whether this reflects a regulatory mechanism of the cell, or whether it is only a phenomenon that accompanies such a mechanism.

In fact, in many cases alterations in the electrical field of a membrane seem to be of functional importance. The following mechanisms may cause this:

- The transverse component of an electrical field in the membrane may affect the functional state of intrinsic molecules. Dipole orientations for example, may modify the function of transport or other functional proteins, phase transitions in the lipid components of the membrane can be influenced by the field, or a transversal shift of small charged molecules can occur.
- The lateral component of the field can cause a displacement in its mosaic structure. This could lead to a local change in the mechanical properties of the membrane causing vesiculation, spike formation, etc.
- The electrical field can influence local ionic concentrations, as well as local pH values in close proximity to the membrane which, in turn, could affect transport processes, biochemical reactions at the membrane surface as well as receptor properties.

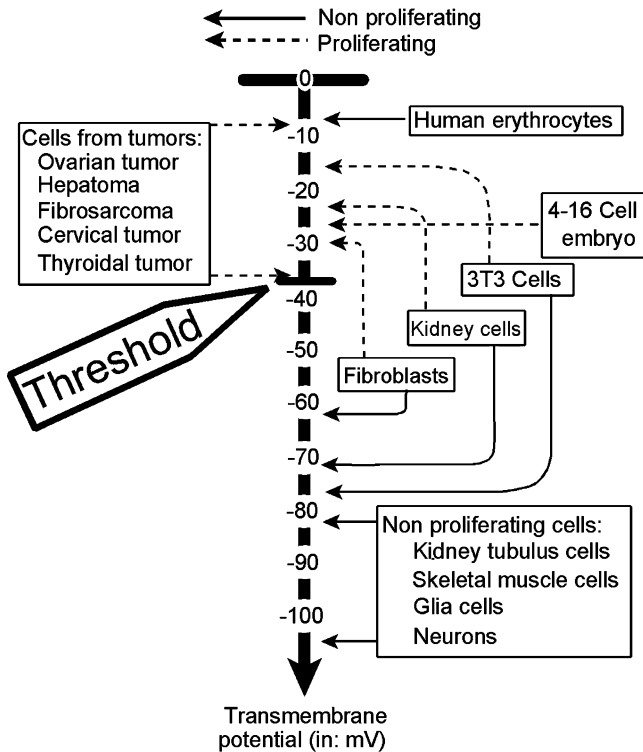


Fig. 3.28 The transmembrane potential of normal animal cells (*right*) and transformed tumor cells (*left*). It can be seen that proliferating cells indicate a membrane potential which is above the threshold value of -37 mV. Cells transiently arriving at the proliferating state lower their absolute potential. The human erythrocyte, as a non-nucleated cell with special physiological functions appears to be an exception (Drawn according to values from Bingeli and Weinstein 1986)

Further Reading

Glaser 1996; Starke-Peterkovic et al. 2005; Wang et al. 2003.

3.4.4 The Action Potential

In the previous section we described the possibility of cells to use the electrochemical gradient of potassium and sodium ions which is built up by active transport, to trigger various amounts of membrane potential simply by changing their permeabilities. This mechanism is expressed most efficiently in nerve and muscle cells. This was the reason why excitation phenomena were detected first in these cells.

Particular progress was achieved following the rediscovery of the giant axon of the squid in 1937 by John Zachary Young, and its subsequent introduction for biophysical measurements by Kenneth Stewart Cole. The use of these giant axons

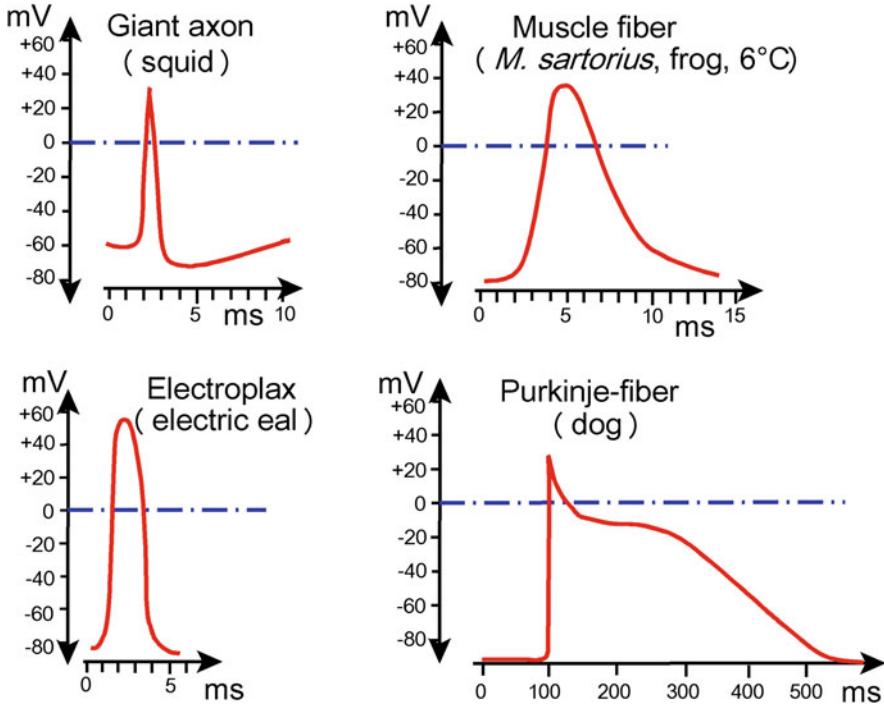


Fig. 3.29 Examples of various action potentials (After Penzlin 1991)

with a diameter up to 1 mm, have made it possible to apply the voltage-clamp technique to determine the ionic currents during the nerve impulse in extensive experiments by Alan Lloyd Hodgkin and Sir Andrew Fielding Huxley. In this technique, the electrical conductivity of the membrane is determined at various fixed transmembrane potentials, generated by microelectrodes. Recently, using patch-clamp techniques it has been possible to investigate the kinetics of these permeability alterations in extremely small membrane areas.

The action potentials of various nerve and muscle cells as illustrated in Fig. 3.29, can be qualitatively explained using the electrical scheme of Fig. 3.27 which was discussed in the previous section. The nonexcited nerve shows a very low sodium permeability (P_{Na}), its resting potential therefore, was determined chiefly by the diffusion potential of potassium which is negative inside-out. After excitation the membrane permeability for ions increased abruptly, whereas the sodium permeability rose quicker than that of potassium. For a short time therefore, the diffusion potential of sodium becomes dominant. This has the opposite polarity to the potassium potential which explains the spike of the action potentials.

As we demonstrated in the previous section the amount of charges that are needed for this kind of depolarization is extremely low. This was checked by flux measurements in excited nerves. During the generation of an action potential, therefore, no significant alterations of the internal ion concentration occur.

A nerve can generate action potentials for a long time after the ion pumps have been blocked. Only after hours does the electrochemical battery of the cell become empty.

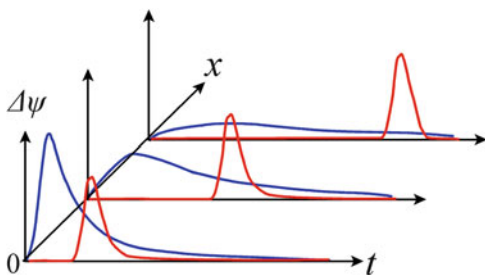
Beside the entire mechanism of membrane excitation, the translation of the action potential along the axon of a nerve cell is of particular interest. In unmyelinated axons the process of pulse transmission is based on a lateral spreading of excitability by the electric field, generated by the excitation itself (see Fig. 3.37). The action potential generated by excited proteins in the membrane triggers the behavior of neighboring proteins. The impulse can proceed only in one direction, because of the refractory period of several milliseconds which the proteins need after an excitation to become excitable again.

Figure 3.30 illustrates the advantage of this kind of impulse propagation in relation to the transmission of a voltage pulse in an electric cable. In contrast to the cable, the time characteristics of the nerve pulse remains more or less constant, even after a certain distance of transmission. Conversely of course, the absolute velocity of pulse transmission in a cable is much faster than in an axon of a nerve.

The advantage of simple electrical conductivity is used in many vertebrate, and in a few invertebrate axons. In this case the axons are surrounded by *Schwann cells* forming the myelin sheath as an electrically isolating layer. Such nerves are called *myelinated*. This sheath is interrupted at intervals of some millimeters by so-called *nodes of Ranvier*, i.e., unmyelinated regions. In the myelinated regions simple electric conductivity of the pulse occurs, as in a cable. The nodes of Ranvier represent membrane areas which are excitable in a normal way. If a certain node of Ranvier is excited, then the pulse propagates by simple electric conduction along the myelinated length and excites the subsequent node. This so-called *saltatory conduction* is a form of pulse amplification leading to a faster transport of information. In contrast to about 1 m/s in unmyelinated nerves, the pulse propagation in fast myelinated nerves is up to 100 m/s.

In 1952 Hodgkin and Huxley, based on intensive experimental investigations on squid axons, proposed a theoretical model of membrane excitation in nerves (Nobel Prize 1963). Its form is of a purely kinetic nature and does not contain information about concrete molecular mechanisms taking place in the membrane.

Fig. 3.30 The time course of a voltage pulse which is set at time $t = 0$ at point $x = 0$, transmitted in an isolated cable (blue lines) and in an unmyelinated nerve (red lines)



The basic equation describes the kinetics of the current in an electrical circuit, similar to the scheme in Fig. 3.27. The current density (j) in such a system can be described by the following equation:

$$j = C' \frac{d(\Delta\psi_M)}{dt} + (\Delta\psi_M - \Delta\psi_K)G'_K + (\Delta\psi_M - \Delta\psi_{Na})G'_{Na} \quad (3.194)$$

$\Delta\psi_M$ is the electrical membrane potential, whereas the symbols $\Delta\psi_K$ and $\Delta\psi_{Na}$ indicate the Nernst potentials of potassium and sodium according to Eqs. 3.191 and 3.192. C' is the capacity of the membrane, and G'_K and G'_{Na} the potassium and sodium conductivities, always corresponding to a unit of area in the membrane. The conductivity of the membrane for individual ions cannot be measured electrically but can be obtained from experiments in which the kinetics of radioactive tracer ions is measured.

The first term of Eq. 3.194 gives the current density which leads to the charge of the membrane capacitor (Fig. 3.27). The following terms represent the current densities associated with potassium and sodium fluxes.

The conductivities G'_K and G'_{Na} are not constant, but functions of the electric field in the membrane, resp. of the membrane potential. The potentiometers in Fig. 3.27, therefore, are controlled directly by $\Delta\psi_M$. From the molecular point of view this means that these conductivities are the result of voltage-dependent channels. It is therefore necessary to make statements about field dependents of these conductivities, i.e., the functions $G'_K(\Delta\psi_M)$ and $G'_{Na}(\Delta\psi_M)$.

To describe the behavior of these channels, Hodgkin and Huxley used a statistical approach. They assumed that the channels can obtain only two discrete states: "open," or "closed." The phenomenological conductivities (G'_K , G'_{Na}) then represent the average of the functional states of a large number of such channels. If all of the channels are open then the maximal conductivities $G'_{K \max}$ and $G'_{Na \max}$ are established.

Furthermore, it is assumed that the potassium channel will be open when exactly four events take place simultaneously, all having the same probability of occurrence (n). The real nature of these events is not explained. It could be, for example, the presence of four potassium ions near the entrance of the channel.

This assumption leads to the following equation:

$$G'_K = G_{K \max} n^4 \quad (3.195)$$

The probability n is a function of time and can be characterized by rate constants α_n and β_n as follows:

$$\frac{dn}{dt} = \alpha_n(1 - n) - \beta_n n \quad (3.196)$$

Concerning the sodium permeability, it is assumed that the channel will be open when three events, each having the probability m occur simultaneously, and if

another inhibitory event having the probability h has not taken place. This leads to the expression

$$G'_{\text{Na}} = G_{\text{Na max}} m^3 h \quad (3.197)$$

For the parameters m and h also kinetic equations can be written:

$$\frac{dm}{dt} = \alpha_m(1 - m) - \beta_m m \quad (3.198)$$

$$\frac{dh}{dt} = \alpha_h(1 - h) - \beta_h h \quad (3.199)$$

The voltage dependence of the channels is proposed to be the result of influences on the rate constants α and β :

$$\begin{aligned} \alpha_n &= \frac{0.01(\Delta\psi+10)}{e^{\frac{\Delta\psi+10}{10}} - 1} & \beta_n &= 0,125 e^{\frac{\Delta\psi}{80}} \\ \alpha_m &= \frac{0.1(\Delta\psi+25)}{e^{\frac{\Delta\psi+25}{10}} - 1} & \beta_m &= 4 e^{\frac{\Delta\psi}{18}} \\ \alpha_h &= 0,7e^{\frac{\Delta\psi}{20}} & \beta_h &= \frac{1}{e^{\frac{\Delta\psi+30}{10}} + 1} \end{aligned} \quad (3.200)$$

(In these equations, the potentials are in mV!)

These equations were obtained from a purely empirical approach, analyzing measured parameters.

It is easy to see that if the relations given in Eq. 3.200 are substituted into Eqs. 3.196, 3.198, and 3.199, a system of nonlinear differential equations will be obtained. The solution of these equations can be substituted into Eqs. 3.195 and 3.197, and eventually, into the basic Eq. 3.194. An analytical solution of this system of differential equations is not possible. Computer simulations of these equations, however, indicate a good accordance with experimental results.

Figure 3.31 shows the calculated time courses for the changes in sodium and potassium conductivities at different membrane potentials. This also corresponds well with experimental findings. These curves illustrate the mechanism described above for the generation of an action potential. The conductivities from Fig. 3.31 illustrate the time-dependent changes of the potentiometers shown in Fig. 3.27, whereas the conductivities are directly proportional to the permeabilities. Within the first millisecond following the stimulus, the sodium potential is dominant because of the rapid increase in G_{Na}' (and thus P_{Na}). This will then be counteracted, by the increasing potassium potential.

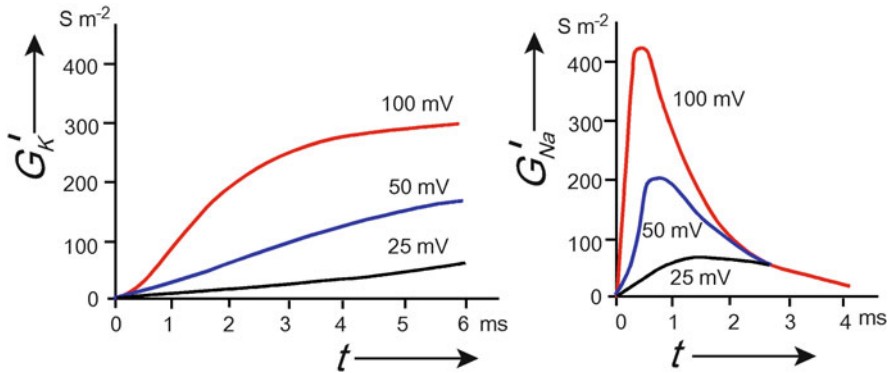


Fig. 3.31 The time dependence of the conductivities G'_K and G'_{Na} for various membrane potentials, corresponding to the theory of Hodgkin and Huxley

The Hodgkin–Huxley model and the corresponding measurements have provided a benchmark in our understanding of cellular excitability. New experimental techniques leading to more precise data nevertheless require some revisions of these approaches. So for example the mechanisms for the voltage-gated potassium and sodium ion currents have been superseded by more recent formulations that more accurately describe voltage-clamp measurements of these components. Especially its current–voltage relation has a nonlinear dependence upon driving force, corresponding to the Goldman–Hodgkin–Katz relation, rather than the linear approach used by Hodgkin and Huxley.

The original formulations of G'_{Na} and G'_K by Hodgkin and Huxley nevertheless continue to be used even though they do not adequately fit voltage-clamp measurements. The deviations between the m^3h and n^4 models (Eqs. 3.195 and 3.197), and the corresponding sodium and potassium currents do not appear to be eminently significant. Models that do describe these circumstances more precisely are more complex, which limits their practical utility in computational neuroscience.

Further Reading

Clay 2005; Hodgkin and Huxley 1952; Huxley 2002.

3.4.5 Molecular Aspects of Membrane Transport

In Sect. 3.4.1 various types of membrane transporters were characterized only in a phenomenological way. Now we will direct our attention to their structure and function. In fact, charged hydrophilic ions and molecules can penetrate the lipid membrane of cells and organelles only with the help of these mediators, usually proteins, the polypeptide chains of which span the lipid bilayer several times. In the last decades the molecular structure of a large number of these proteins has been revealed thanks to X-ray crystallography. In this way, the former more or less

mechanistic models of transport processes were replaced by more realistic molecular mechanisms.

In general, the following properties of transporters require an answer from these molecular considerations:

- Their extremely high selectivity including the phenomena of dehydration and rehydration of hydrophilic species in the process of membrane permeation.
- The mechanism of coupling between transport and the energy supporting biochemical reactions.
- The mechanisms of transport regulation by ligands and the transmembrane potential.

In 1998 MacKinnon unlocked the three-dimensional molecular structure of a potassium channel, a success which was awarded with the Nobel Prize in 2003. Such K^+ channels are found in bacterial as well as in eukaryotic cells of plants and animals, which are related members of a single protein family. Their amino acid sequences are easy to recognize because they contain a highly conserved segment called the K^+ channel signature sequence.

Let us answer some of the questions noted above using this extensively investigated example. The pore of this transporter is comprised of four identical subunits that encircle the central ion conduction pathway (two of them are depicted in Fig. 3.32). Each subunit contains two fully transmembrane α -helices, and a tilted pore helix that runs half way through the membrane. The hydrated K^+ ion, entering this channel from the cytoplasmatic side, first remains in the hydration state in a water-filled cavity with a diameter of 1 nm near the midpoint of the membrane. This cavity helps the K^+ -ion to overcome the electrostatic repulsion that it would

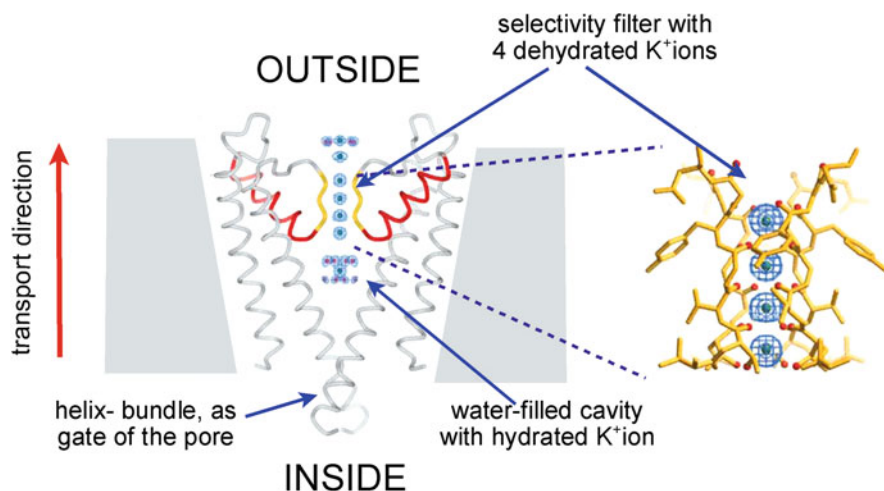


Fig. 3.32 The molecular structure of the KcsA channel. Only two subunits of this tetrameric molecule are shown. According to the position of the intracellular ends of the inner helices forming the gate, it is shown in a closed state (From MacKinnon 2003, modified)

normally experience when moving from the cytoplasmatic water phase into the low dielectric membrane environment. By allowing it to remain hydrated at the membrane center, and by directing the C-terminal negative ends of the protein helices toward the ion pathway, it becomes stabilized at the membrane interior. After this it enters the selectivity filter which contains four evenly spaced layers of carbonyl oxygen atoms, and a single layer of threonine hydroxyl oxygen atoms, which create four K^+ binding sites. In fact, on average only two K^+ ions are present at a given time in these four positions, always separated by one water molecule. It is very important that the arrangement of these protein oxygen atoms is very similar to that of water molecules around the hydrated K^+ ion. In this way the energetic cost of dehydration is minimized. Furthermore, a part of the binding energy is used for conformational changes of the proteins, which also is a prerequisite for the high conduction. In fact, the flux achieves up to 10^8 ions per second. This rate is large enough for sensitive amplifiers to record the electric current of a single channel. Na^+ ions cannot enter this filter because of their different crystal structure.

The gate of the channel is represented by a helix bundle near the intracellular membrane surface. In the closed position, as depicted in Fig. 3.32, the pore narrows to about 0.35 nm and is lined with hydrophobic amino acids, creating an effective barrier to the hydrophilic ions. This structure seems to be representative for many different potassium channels, irrespective of the stimulus that causes the pore to be in closed or open state. The conformational changes of these polypeptide chains that open and close the channel gate occur on the order of 10^2 times per second.

As discussed in previous chapters the membrane potential, and consequently the membrane electric field and its modification forms not only the basic principle of nerve and muscle excitation but regulates various functions in nearly all cells. This requires proteins, especially transporters, embedded in the membrane that sense alterations of this field and transform them into various cellular signals.

It is easy to imagine how an electric charge or an electric dipole can be reorientated within a protein when the field is changed. This can produce a conformational change in the protein that may regulate its function. The movement of the charge or the dipole induces a transient current (*gating current*) that can be measured experimentally and provides direct information about such conformational changes. The extent of the charge movement depends on the magnitude of the charge and the strength of the electric field in the region where the charge moves. In Sect. 2.2.1 (Fig. 2.15) as a crude estimation, this field strength was indicated to be of the order of 10^7 V m⁻¹. In fact, the exact value of this parameter near the corresponding charges or dipoles is unknown. In some cases the field can be concentrated to a narrow region around this location. Furthermore, the dielectric constant of this region inside the molecular structure is unknown.

The most extensively investigated voltage-gated channel is the so-called Shaker K^+ channel which can be expressed at a high density in *Xenopus* oocytes. It was isolated from *Drosophila melanogaster* and was named after the shaking that the fly

undergoes under anesthesia in its absence. Measurement of the gating current by patch-clamp techniques indicates that 13 electron equivalent charges per molecule are moving in this case. On the basis of the crystal structure of this protein, the so-called *paddle model* was proposed. It is assumed that voltage-gating charges are located on a hydrophobic helix-turn-helix structure, the so-called S4-segment, which can move within the membrane near the protein-lipid interface according to the direction of the electric field. Recently an S4-type sensor has been found in a voltage-dependent phosphatase, suggesting that this type of sensor may be modular and might have been incorporated into other types of proteins.

The kinetic model of nerve excitation as discussed in the previous section requires a particular sequence of opening and closing of potassium and sodium channels controlled by the membrane potential. Probably the four voltage-sensor domains of these channels react with individual time courses.

Although K^+ channels are excellent prototypes for voltage-gated channels, there are several other types of membrane proteins that differ in function, selectivity, regulation, kinetics, and voltage dependence. So for example a G-protein coupled muscarinic receptor has been found, in which a voltage-sensor is an integral part of the structure. It is expected that many other sensors will be discovered in the near future. More structures and biophysical analyses are still needed for a full molecular understanding of the function of these voltage sensors.

In contrast to the relatively simple mechanisms of channels, the pumps, and the co-transport systems require more functional elements, and the transport mechanisms demand more conformational changes in the corresponding transport protein. Especially the energy release by hydrolyzing ATP, and its coupling to ion movement needs a series of protein conformations. The first atomic-resolution structure of an ion pump was published in 2000 for the Ca-ATPase by Toyoshima et al. It shows an integral membrane protein with a large extended cytosolic part. In spite of the enormous progress of research in this field, a number of questions, especially concerning the Na-K-ATPase, are still open. The required conformational changes that accompany these transport processes mean that their speed is much slower than processes of channel transport.

The progress in determining the molecular structures of these channels has greatly facilitated the theoretical modeling and numerical simulation of the ion transport process itself. The most detailed description is based on the concept of *molecular dynamics* (MD). In this case microscopic forces of interactions between the penetrating ion and all atoms of the channel are calculated based on the classical Newton's equation of motion. This leads to trajectories of all the atoms in the system. In recent years, this approach has been used to simulate an increasing number of channels. Although this is the most detailed and accurate approach, it is limited by its shortcomings in quantitatively characterizing a large system, and its application depends considerably on advanced computational techniques.

Simpler and computationally less expensive of course are continuum models based on macroscopic or semimicroscopic continuum calculations like the Poisson-Nernst-Planck (PNP) approach. They, however, include a number of

limitations that have already been discussed in Sect. 3.3.2. A more realistic approach, situated between MD- and PNP-models, is based on *Brownian dynamics* (BD). In this case the trajectories of the ions in the system are followed using the Langevin equation. This is an approach, based on the theory of Brownian movement that considers the force not as the result of interactions of all atoms in the system, but rather of a subset of relevant coordinates. BD simulations have been applied to a variety of ion channels in recent years and the agreement with experimental work has been quite encouraging.

Further Reading

Bezanilla 2008; Faller 2008; Gadsby 2009; Kuyucak and Bastug 2003; Luckey 2008; MacKinnon 2003; Toyoshima et al. 2000; Zhou and Uesaka 2009.

3.5 Electric Fields in Cells and Organisms

3.5.1 The Electric Structure of the Living Organism

In Sect. 2.1.3 we pointed out that the term “structure” must be used in a very broad sense, not limiting it to visible details of a biological system. In fact, structures can be defined, based on the distribution pattern of various physical properties, such as concentrations, temperatures, streaming vectors, etc. Using the definition of the electric potential as a scalar state parameter $\psi(x, y, z, t)$ (Sect. 2.2.1), an *electric structure* can be established too. This, in fact exists on all levels of biological organization as a hierarchic structure, fully corresponding to the usual classification in biology, but governed by particular physical approaches (Fig. 3.33).

At the atomic and molecular level, the interactions can be explained using the approaches of wave mechanics. Particularly, the Schrödinger equation allows us to calculate the electric parameters at atomic dimensions, determining the energies of chemical bonds and molecular interactions.

Considering supramolecular structures like membranes, the electrical structure is determined by fixed and mobile charges and dipoles, forming electric double layers, and governing intermolecular interactions. Statistical thermodynamics can be used to consider these systems, leading to the Poisson–Boltzmann equation (Eq. 2.53). We already discussed electrical structures at this level of organization in Sects. 2.3.5 and 2.3.6.

For consideration of the cell as a thermodynamic system, the approaches of phenomenological thermodynamics were used. The interior and exterior medium of cells and organelles can be considered as phases with particular properties. Differences in the electrical potential between these phases, such as transmembrane potentials, can be described by the Nernst–Planck equation (Eqs. 3.156, 3.157). Their properties, dynamics, and biological relevance were discussed in Sects. 3.4.3 and 3.4.4. However, we remarked at that point that these properties cannot be

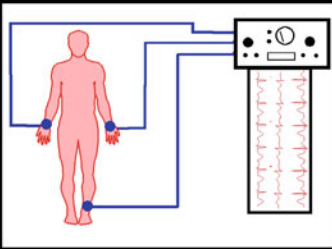
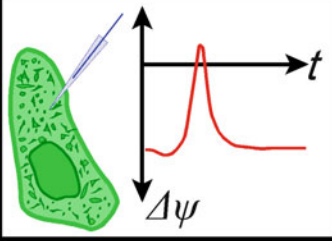
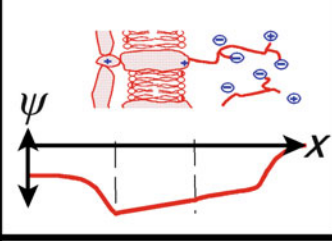
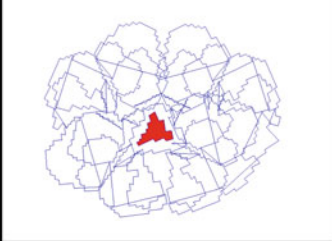
Organism		Electrodynamics <i>Maxwell equations</i>
Cell		Phenomenological thermodynamics <i>Nernst-Planck equation</i>
Supramolecular structures		Statistical thermodynamics <i>Poisson-Boltzmann equation</i>
Atomar, and molecular structures		Wave, and quantum mechanics <i>Schrödinger equation</i>

Fig. 3.33 The hierarchic system of the electric structure of the living organism and the corresponding physical approaches

calculated by phenomenological approaches alone. We mentioned the Poisson–Nernst–Planck theory (PNP-Theory, Sect. 3.3.2) as necessary for the completion of this approach.

In the following text we enter a further region of the hierarchic structure. We will consider electric fields in the extracellular space, in tissues and organs. This already extends into the area of classical electrodynamics where the Maxwell equations allow us to calculate electric fields in inhomogeneous dielectrics. The question arises: how does the field distribute inside the body, which consists of organs with different conductivities, like bones, soft tissue, air-filled cavities, etc.? We will

come back to this question again in Sects. 4.5, 4.6, and 4.7 where the influence of electromagnetic fields on biological systems is considered.

In spite of these hierarchic structure levels, it should be pointed out again that the electric potential is a physically defined parameter, independent of its source. Therefore, there exists only *one* electrical potential at *one* moment in time at *one* point in space. This statement is important with respect to the circumstances which will be discussed in the following text, namely the influence of externally applied electric fields on the intrinsic ones.

3.5.2 Extracellular Electric Fields and Currents

It has long been known that electric fields and currents exist not only across the membranes of cells and organelles but also in tissue and the whole body. Such fields are measured as electrocardiograms (ECG), electromyograms (EMG), and electroencephalograms (EEG) in medical diagnosis for example. EKG and EMG potential differences are of the order of several millivolts. EEG potentials are much lower because of the isolating role of the skull. Beside these oscillating potentials, it is also possible to detect DC potentials between various parts of the animal body and even in plants. Many measurements of these potentials, however, have suffered because of inaccurate use of electrodes.

The origin of extracellular electric fields and currents in biological organisms can be different. They can be generated directly as a result of ion transport in living cells or indirectly as streaming potentials or even by piezoelectric effects.

Let us first consider the electric conditions in the intercellular space (Fig. 3.34). As already mentioned (Sect. 2.3.6), there are strong electric fields around the fixed surface charges, not only perpendicular to the membrane plane, as considered in the electric double layer theory, but because of the lateral mosaic of surface charges also in the tangential direction. These fields, however, like the strong electric fields

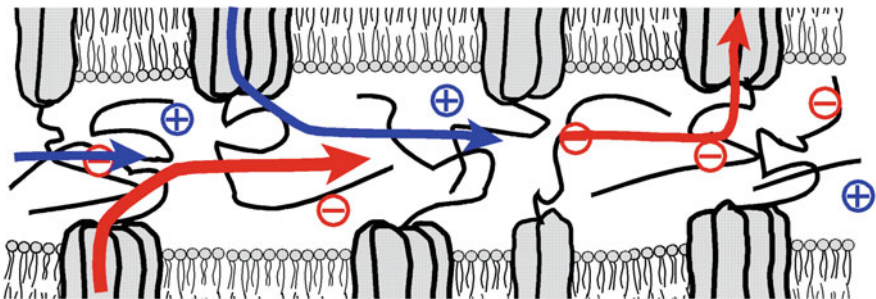


Fig. 3.34 The intercellular cleft can be considered as a Donnan system, formed by the mostly negative fixed charges of the surface coats of adjacent cells. This equilibrium, however, may be deflected by fluxes of positive and negative ions, extruded or taken in by transporters located in the membrane mosaic. Presently this picture remains speculative as there is no method to measure these parameters at the moment

in a double layer, are part of an electrochemical equilibrium, and do not generate electric currents. This Donnan system, however, is deflected by ion transport processes through the adjacent cell membranes. As illustrated schematically in Fig. 3.26, a large variety of rheogenic transporters permanently move ions across the membrane at various locations. This results in a complicated current pattern through the system of fixed charges around the cells, and in the intercellular clefts. Unfortunately, this situation is just speculative because to date no method exists to analyze this situation experimentally in intercellular clefts, or in the dimensions of surface coats of cells.

In some cases the distances of the individual electrogenic transporters are large enough to produce currents which are measurable in their vicinity. One example is a rheogenic calcium pump which is localized at a particular point in the membrane of eggs of the fucoid seaweed *Pelvetia*. This pump induces a current density of up to 0.3 A m^{-2} , and an electric field around the cell. It is possible to measure this field using the vibrating probe technique (see Fig. 3.35). For this a glass microelectrode is used, the top of which consists of a metallic sphere with a diameter of a few micrometers. This electrode vibrates at a frequency of several hundreds of Hz causing this sphere to be displaced periodically at an amplitude of 10–30 μm . Using a very sensitive low noise amplifier it is possible to measure voltage differences near $1 \text{ nV} = 10^{-9} \text{ V}$ between these two reversal points.

Interestingly, this field seems to play a role in growth regulation of these eggs, and determines the direction in which the first rhizoid of the cell will grow. Experiments with external fields confirmed that in fact, the field determines the direction of growth.

Meanwhile, a large number of biological objects have been investigated with this method, such as single cells, as well as various growing parts of plants, follicles of insect eggs, growing embryos, muscle fibers, and other tissues and organs. Investigation of embryos of the clawed frog *Xenopus* showed particular currents

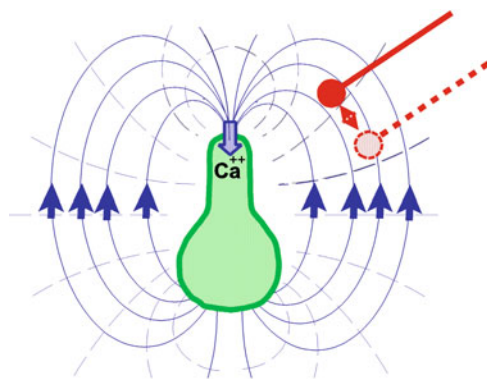


Fig. 3.35 Current- (—), and equipotential lines (---) near a growing egg of the brown algae *Pelvetia*, caused by local rheogenic calcium transport, which determines the location where the rhizoid is formed. The external field was measured using the vibrating probe method. The corresponding electrode is shown in two extreme positions during vibration. Relations of electrode and cell sizes are approximately correct (According to data from Jaffe 1979)

and fields that change direction and intensity at different stages of development. Obviously, this reflects a particular control function, the mechanism of which is unclear. It is possible that in this way the diffusion of charged growth factors is directed from one group of cells to another. In this case the vector orientation of the electric field is imposed on the chemical gradient.

A special source of electric currents and fields are the so-called *wound potentials*. The reason for them is permanent electrostatic potential differences between different cavities in the body, caused for example by ion pumps, differences in electrolyte composition and others. This leads to various transepithelial potentials. An intact mammalian corneal epithelium for example maintains a potential difference of about 40 mV. It results from net inward transport of K^+ and Na^+ , and the net outward transport of Cl^- to the tear fluid. In mammalian skin the inward transport of Na^+ leads to a potential difference of about 70 mV between dermis and the outer layer of the epidermis. The maintained potential difference is possible due to tight junctions, which are closely associated areas of the cells forming a barrier with high electrical resistance.

If the isolating properties of an epithelium are disturbed by a wound, the potential collapses at this point because short circuit currents occur. This leads to an electric field with a lateral orientation to the plane of the epithelium (see Fig. 3.36). In most cases the wound has a positive polarity in relation to the surrounding surface. As a noninvasive method the above-mentioned vibrating electrodes are also used for mapping the electric field near wounds. For this a small metal vibrating probe with a displacement of 0.18 mm in air above the skin measures the surface potential of the epidermis through capacitive coupling. In Sect. 4.4.2 we will indicate how cells may use this field by way of galvanotaxis to close a wound.

Another source of extracellular currents and fields are various membrane excitations. The axon of a nerve cell may be the best example. If an action potential propagates along the axon, small areas of this membrane will become depolarized. In contrast to the parts of the axon with resting potential, where the external membrane side is positively charged in relation to the inner one, the polarity of membrane sides with action potential is opposite. The result is a lateral electrical current from one point of the membrane to another (Fig. 3.37). Such local depolarizations not only occur in nerves, but also in muscle and other cells.

In the case of a synchronized excitation of a bundle of nerve or muscle cells, the currents and fields of the single cells are superimposed, and can be measured as electrocardiograms (ECG), electromyograms (EMG), and electroencephalograms (EEG) on the periphery of the body. The use of surface electric potential differences in medical diagnosis raises the question: how do electric fields, generated for example by the beating heart, spread in the body? Modern techniques allow a reasonably good reconstruction of the electric field distribution, using parallel recordings of the ECG simultaneously in different parts of the body. The measurable ECG results from excitation of definite parts of the cardiac muscle in the rhythm of the beating heart. Consequently, the heart becomes an oscillating dipole, the orientation of which changes according to the orientation of the actually excited parts of the muscle.

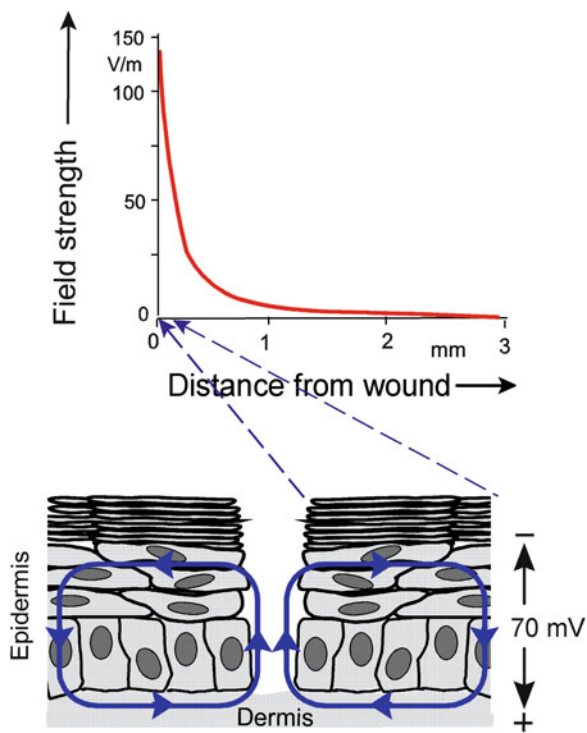


Fig. 3.36 Schematic representation of the occurrence of a wound potential in the mammalian epidermis. The field strength near the wound edge reaches 140 V m^{-1} and declines strongly with distance (Data from McCaig et al. 2005)

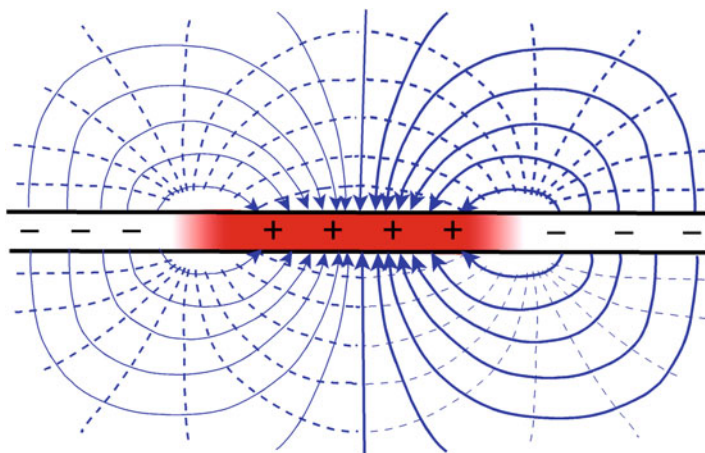


Fig. 3.37 Schematic illustration of a snapshot of an excited nerve. The *red areas* represent the actual location of the depolarized membrane (see also Fig. 3.29)

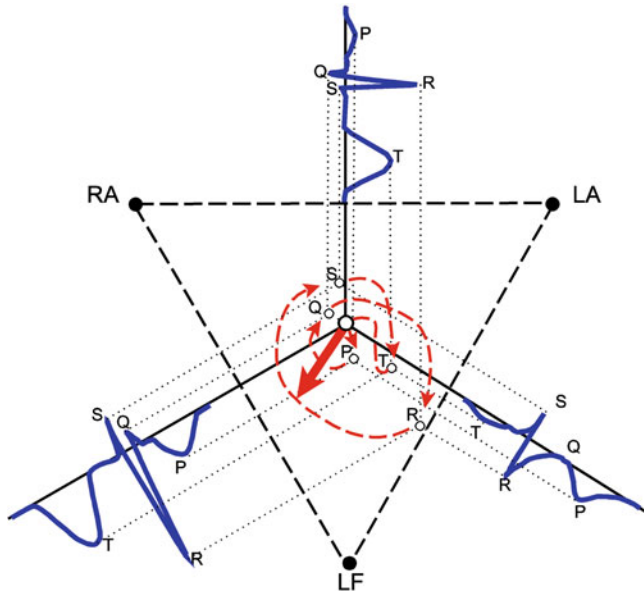


Fig. 3.38 Construction of a vector cardiogram according to the Einthoven triangle. Using the time course of the potential curves (blue), orientated in an equilateral triangle, a rotating dipole (red arrow) can be constructed in the center. P, Q, R, S, T – are the corresponding waves of the ECG. RA right arm, LA left arm, and LF left foot

The first person to propose a method to evaluate ECGs was the Dutch physiologist Willem Einthoven, who was awarded the Nobel Prize for his work in 1924. He proposed that it should be possible to localize the excited parts of the heart by detecting the potentials at three points on the body, which are more or less equidistant from the heart. This so-called *Einthoven triangle* is illustrated schematically in Fig. 3.38. The three cardiograms, derived from the corresponding points, indicate a periodic sequence of P, Q, R, S, and T waves. These waves represent the sequence of excitation of different parts of the heart muscle, starting with the excitation of the atria (P wave). If all the atrial fibers are in the plateau phase, the PQ segment is reached. Subsequently, the excitation spreads over the ventricles, beginning on the left side of the ventricular septum, spreading toward the apex (QRS waves), and finally reaching the ventricular recovery phase (T wave). As a result of the projection of these curves corresponding to the geometry of an equilateral triangle, a rotating vector appears, the origin of which lies in the crossing point of the three axes. The arrowhead moves periodically along the dashed line. Because of the dielectric heterogeneity of the body, an exact correlation between the resulting dipole vectors of the field with the anatomical position of various parts of the heart muscle is impossible.

In a special adaption, electric fishes may use extracellular fields generated by excitable cells. In the electric organs specialized muscle cells are organized in so-called *electroplaques*, where the voltage of several cells adds up to

approximately 800 V. Such high voltage is only possible, however, in freshwater fishes like the electric eels. Marine fishes generate smaller voltages because of the high conductivity of seawater. In the electric organs of these fishes the cellular elements are not arranged in series, leading to higher voltages, but rather in parallel to increase the current. So-called *strong electric fishes* use the induced electric fields to catch their prey, or to defend themselves, whilst *weak electric fishes* use electric fields only for orientation. Meanwhile, the amazing capacity of this sensory system has been determined. Although it works only over small distances, which amounts approximately to half of the body length of the fish, shapes and dielectric properties of subjects in muddy water are detected in this way (for electroreception, see Sect. 4.5, Fig. 4.25).

Beside these kinds of currents and fields generated by living cells, other sources in the body occur as a result of electromechanical transformations, for example by piezoelectric and electrokinetic effects (Sect. 2.3.5). Both effects occur in bones and cartilage during in vivo deformations.

Piezoelectricity is comprised of translocations of charges in crystals or crystalloid macromolecular structures resulting from mechanical loading. In bones piezoelectric potentials are mostly the result of deformations of collagens. Furthermore, under mechanical stress, such deformations cause a flow in the narrow channels of the bone with negative surface charges. This additionally leads to streaming potentials. In contrast to the electric fields in the whole body induced by nerves and muscles, these fields are rather low at larger distances. Conversely, they seem to be important locally in the process of bone growth and bone remodeling in vivo. This circumstance has been used in efforts to stimulate bone repair by applying electric field pulses or ultrasound.

Further Reading

Electric fields near cells: McCaig et al. 2005; Nuccitelli et al. 2008; Shi and Borgens 1995; EEG: Nunez and Srinivasan 2006; Reilly 1998; electric fields in bone: MacGinitie 1995; electric fishes: Fortune 2006; Peters et al. 2007.

3.5.3 *Passive Electrical Properties of Tissue and Cell Suspensions*

In contrast to the membrane potential and the electrical currents driven by rheogenic pumps, all requiring *active* biological processes, other qualities like electrical resistance, conductivity, or membrane capacity can be summarized as *passive electric properties* of biological systems. These parameters are essential not only to calculate the distribution of the actively generated fields and currents in the organism, as described in the previous section, but also the penetration and induction of currents in the body caused by external influences (Sect. 4.6.1). This includes AC fields over a broad frequency range.

To derive the basic parameters and equations, let us first consider the electrical properties of a plate capacitor filled with material of certain conductivity and

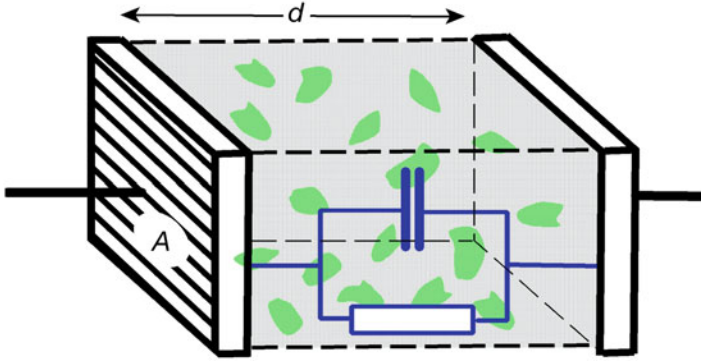


Fig. 3.39 A capacitor consisting of two parallel plates of area A , and a mutual distance d , filled with an inhomogeneous medium (e.g., cell suspension or biological tissue). Neglecting the real dielectric heterogeneity, it can be described formally by a RC-circuit (blue)

dielectric constant, such as for example biological tissue. Irrespective of its dielectrical heterogeneity, this system formally can be described by an equivalent RC circuit consisting of a conventional capacitor and a resistor in parallel (Fig. 3.39).

An AC voltage applied to such a system generates a current which will flow through the resistor and periodically recharge the capacitor. This results in an effective AC resistance which is called *impedance*, whereas its reciprocal is the AC conductance or *admittance* (Y^*). In fact, this admittance is determined by the static conductance (G) of the resistor, and the frequency (ω) dependent displacement current passing through capacitor. This behavior can be summarized as follows:

$$Y^* = G + j\omega C \quad (3.201)$$

This equation includes the imaginary number $j = \sqrt{-1}$ transforming the admittance (Y^*) into a complex parameter, marked by the superscript *. The reason for this is the response of the system to the particular time function of the AC-current. For the behavior of an RC circuit in an AC field not only the amplitude of a sine current must be taken into account, but also the occurring phase shift which is coded by the term j .

To understand this, let us consider an AC voltage having the following time function:

$$U = U_{\max} \sin \omega t \quad (3.202)$$

An applied voltage (U) therefore oscillates with an angular frequency ($\omega = 2\pi\nu$) and a peak amplitude of U_{\max} . According to Ohm's law (Eq. 3.51), for a circuit containing only a resistance the current (I) possesses the same time behavior:

$$I = I_{\max} \sin \omega t \quad (3.203)$$

In an RC circuit, however, additionally the displacement current of the capacitor must be taken into account. The charge (q) of a capacitor with the capacity (C) is determined by the following equation:

$$q = U C \quad (3.204)$$

Since the current is defined as the time derivative of charge, for the time-independent capacity (C), the displacement current (I_C) is:

$$I_C = \frac{dq}{dt} = \frac{d(U C)}{dt} = C \frac{dU}{dt} \quad (3.205)$$

Introducing Eq. 3.202 into Eq. 3.205 for $C = \text{const}$, one gets:

$$I_C = C \frac{d(U_{\max} \sin \omega t)}{dt} = C U_{\max} \omega \cos \omega t = C U_{\max} \omega \sin\left(\frac{\pi}{2} + \omega t\right) \quad (3.206)$$

Further, defining:

$$I_{C \max} = C U_{\max} \omega \quad (3.207)$$

it follows:

$$I_C = I_{C \max} \sin\left(\frac{\pi}{2} + \omega t\right) \quad (3.208)$$

Comparison of Eq. 3.202 and 3.208 clearly indicates for a simple capacitor a phase shift of $\pi/2$ in between current and voltage.

Consequently, for the AC current in Eq. 3.201 this phase shift, for example the time course of the capacitive component of the sum, must be considered. This circumstance is taken into account using a Gaussian plane containing imaginary numbers. For this in Eq. 3.201 the term ωC is multiplied by the imaginary number $j = \sqrt{-1}$ and plotted on the ordinate in these Gauss coordinates.

Equation 3.201 can be modified using the geometrical parameters of the electrode: area (A) and mutual distance (d), together with the material constants: specific conductivity (g), and permittivity ($\varepsilon \varepsilon_0$), whereas:

$$G = \frac{A}{d} g \quad \text{and} \quad C = \frac{A}{d} \varepsilon \varepsilon_0 \quad (3.209)$$

Introducing these relations into Eq. 3.201, one gets:

$$g^* = g + j \varepsilon \varepsilon_0 \omega \quad \text{whereas :} \quad Y^* = \frac{A}{d} g^* \quad (3.210)$$

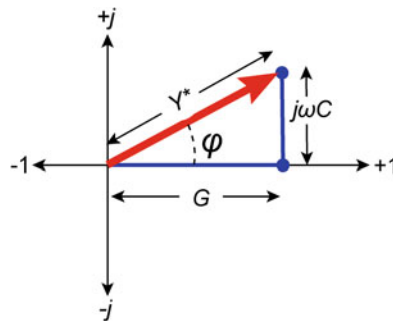


Fig. 3.40 Representation of the conductance (G) of the resistor and the admittance of the capacitor ($j\omega C$) of the analog circuit from Fig. 3.39 in the Gaussian plane of complex numbers. The admittance of the system (Y^*) corresponds to the length of the resulting vector. The angle φ represents the resulting phase shift

In this equation g^* is called the *complex specific admittance*.

In the same way as the complex specific admittance (g^*), a complex dielectric constant (or *relative permittivity*) (ε^*) of the system can be formulated. Starting with the equation for the effective AC resistance of a capacitor ($R = 1/\omega C$), one can introduce a complex capacitance (C^*) as follows:

$$C^* = \frac{Y^*}{j\omega} \quad (3.211)$$

Inserting the parameters of Eqs. 3.201 and 3.209 into this equation, and considering $C^* = \varepsilon^* \varepsilon_0 A/d$, and $1/j = -j$, one gets:

$$\varepsilon^* = \varepsilon - j \frac{g}{\omega \varepsilon_0} \quad (3.212)$$

The derivation of these basic equations (Eqs. 3.210 and 3.212) allows us to understand the properties of complex dielectrics in AC fields.

In fact, the electrical properties of biological tissues containing a multitude of various cells with various size and intercellular clefts are too complicated to be described by a simple RC circuit as shown in Fig. 3.39. An enormous network of elementary RC elements with different parameters would be required to build an appropriate model. This, of course, is impossible for practical use. For this reason one uses the simplified scheme as shown in Fig. 3.39 with the knowledge that the properties of the resistor as well as the capacitor are now frequency dependent.

Figure 3.41 shows mean values of the complex dielectric constants (ε^*) and the specific admittance (g^*) of various tissues over a broad frequency range. It can be seen that the specific admittance increases with frequency. The main reason for this is the increase of the membrane admittance. Conversely, the dielectric constants are

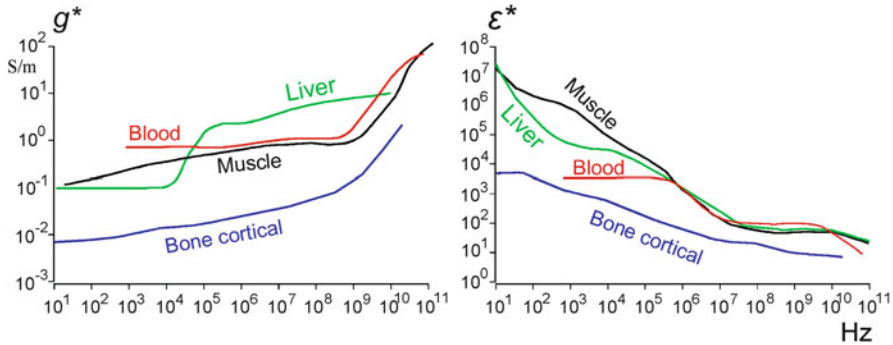
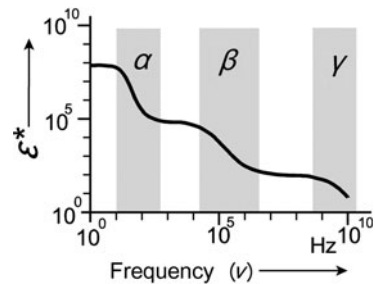


Fig. 3.41 Complex specific admittances and complex dielectric constants of various tissues as functions of frequency (According to averaged data from Gabriel et al. 1996)

Fig. 3.42 Schematic illustration of the frequency regions of the α -, β -, and γ -dispersions of the complex dielectric constant of a biological tissue



extremely high for tissues at frequencies below the MHz range. They drop to the standard value of the dielectric constant of water ($\epsilon \approx 80$) only at microwave frequencies.

Typically, the dielectric constant of tissues or cell suspensions decreases in a characteristic step-wise fashion. The reasons for this are the properties of various kinds of RC-circuits in the system, each with different time constants of relaxation. These frequency regions, in the first instance have formally been designated as α -, β -, and γ -dispersions (see Fig. 3.42).

Various phenomena are summarized as α -dispersion, which for cell-sized objects occurs in the frequency range below 10 kHz. Mostly, reactions in the electric double layer of the cell membranes are responsible for this, such as various electrokinetic phenomena (see Sect. 2.3.5), for example the deformation of ionic clouds. Because of various artefacts, related to polarization phenomena of the electrodes, and electro-osmotically induced convections, it is difficult to measure the real physical parameters in this frequency region. It is difficult to determine the individual components of an inhomogeneous dielectric system.

The β -dispersion, also designated as *Maxwell–Wagner dispersion*, is mostly based on processes connected to membranes as dielectric barriers. Consequently these dispersions are caused by structural inhomogeneities of the material, such as

cellular and organelle structures. Some authors subdivide the β -region into β_1 -, and β_2 -ranges which refer to the dispersion of the cell membrane (β_1), and cytoplasm (β_2) polarization dispersions, respectively.

The γ -dispersion at higher frequencies is caused by so-called *Debye relaxations* of various molecular dipoles. At frequencies of the γ -dispersion region, even the resistivity of the internal and external milieu of the cell cannot simply be described by ohmic resistances (see the small RC-circuits in the resistors of Fig. 3.43).

The dispersion of water dipoles occurs at 18.7 GHz. This in fact, is true only for free water. Recent measurements indicate that the dispersion of bound water may occur at frequency regions even below 1 GHz. This is important because of possible effects of high-frequency electromagnetic fields on biological tissue (Sect. 4.7.1).

The measurement of dielectric properties of solutions and heterogeneous media, the so-called *impedance spectroscopy* is used in various applications. Furthermore, electrical impedance tomography (EIT) has been developed for imaging of particular dielectric properties of parts of the body. For measurements a multitude of electrodes is placed on the skin. Proposed applications include the monitoring of lung function, detection of cancer in the skin or breast, and the location of epileptic foci. This method, however, is still at an experimental stage and not used in routine diagnostics.

Further Reading

Barnes and Greenbaum 2006, Gabriel et al. 1996; Holder 2005; Orazem and Tribollet 2008; Pethig and Kell 1987; Riu et al. 1999; Schwan 1957.

3.5.4 Single Cells in External Electric Fields

Cell suspensions or single cells are exposed to electric fields of various frequencies in many biotechnological approaches (see Sect. 3.5.5). Therefore, it is important to analyze field distribution and currents through and around these cells.

Figure 3.43 indicates a simplified electronic circuit, describing the passive electric properties of a spherical cell in an electrolyte solution. As in Fig. 3.27, the membrane is represented by a capacitor (C_m) in parallel with a resistor (R_m) simulating the membrane resistance. At DC, and low-frequency AC fields, the conductivities of the external and internal media can be described by simple resistors (R_{e1} , R_{e2} , R_i).

The high specific conductivity of the cytoplasm and the external medium on the one hand, and the extremely low conductivity of the membrane on the other results in R_m being more than 7 orders of magnitude higher than R_e or R_i . Applying Kirchhoff's law, a low frequency current therefore does not flow through the cell, but around it.

This is demonstrated schematically in Fig. 3.44 where a spherical cell is shown with the electrical properties as described above. In case of DC fields and extremely

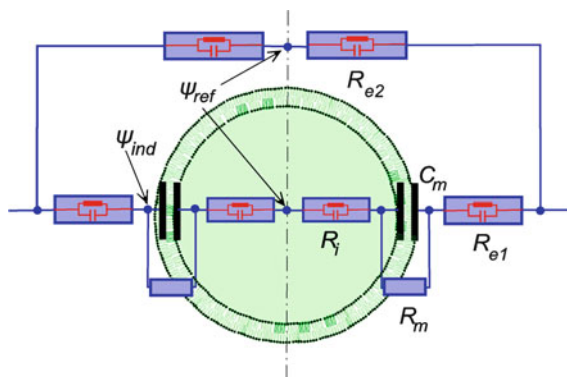


Fig. 3.43 Simplified analog circuit demonstrating current paths through and around a spherical cell. R_{e1} and R_{e2} – resistances of the external medium, R_i – resistance of the cytoplasm, R_m – membrane resistance, C_m – membrane capacity, ψ_{ref} – reference potential at the symmetry plane of the system, ψ_{ind} – induced potential at the membrane surface. At higher frequencies RC properties must be attributed to the resistances R_{e1} , R_{e2} , and R_i too (Internal and external resistors were split to obtain a reference potential ψ_{ref})

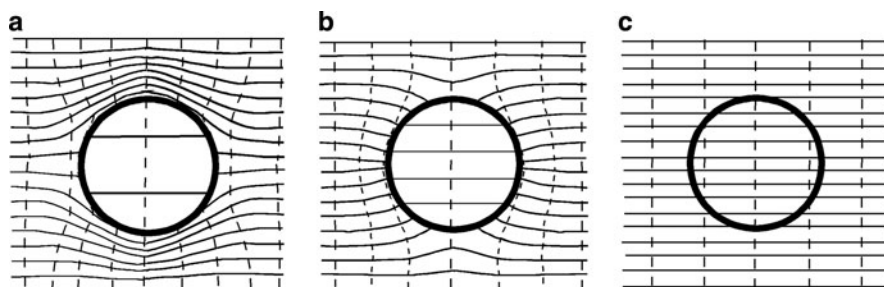


Fig. 3.44 Current lines (—) and equipotential lines (- -) in, and around a spherical cell in a homogeneous electric AC field. In contrast to Fig. 3.45 the polarization charges at the membrane are not depicted. (a) The cell in a low frequency, (b) and (c) – in a high frequency AC field. In case **B** the conductivity of the external medium is lower than that of the internal one, in case **C** the cell is surrounded by a physiological medium where the permittivities of the internal and external milieus are the same

low-frequency AC fields, because of the high membrane resistance, the field inside the cell is negligible, and the external field becomes deformed (Fig. 3.44a). The membrane capacitor will be increasingly bridged with increasing frequency. This leads to a change of the field distribution in, and around the cell (Fig. 3.44b, c). The field penetration increases with increasing frequencies. In parallel, the membrane polarization decreases, and the polarization of the cytoplasm increases. Taking into account that there is no large difference in the permittivities between medium and cell plasma, the degree of field penetration will be high and constant at frequencies above some 100 MHz.

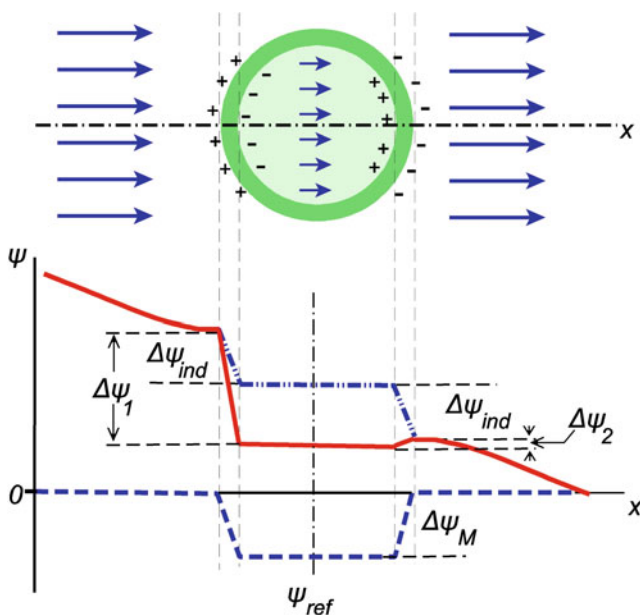


Fig. 3.45 The distribution of charges and potentials at a spherical cell in an external DC field. In the upper part the arrows indicate direction and strengths of the field vectors. The charges near the membrane are the result of polarization by the external field. In the lower part of the figure, the potential profile is depicted over the x -axis through the center of the cell: - - - membrane potential without external field influence, including the undisturbed in vivo transmembrane potential ($\Delta\psi_M$); - . - . potential induced by the external field, without the in vivo potential (induced potential difference: $\Delta\psi_{ind}$); — actual potential function, as the sum of both functions and the resulting membrane potential differences at both poles of the cell ($\Delta\psi_1$, $\Delta\psi_2$). ψ_{ref} is the reference potential according to Fig. 3.43. In contrast to Fig. 2.48, the electric double layer is not taken into account for simplicity

Let us consider the case of Fig. 3.44a in detail in order to discuss the influence of external fields on the membrane potential of a spherical cell in DC or in extremely low-frequency AC fields (Fig. 3.45). Neglecting the conductivity of the membrane, it can be understood that the external field leads to an accumulation of charges, especially in membrane areas that are orientated perpendicular to the undisturbed current lines. This leads to a deformation of the external electric field, and to the charging of the membrane capacitor. The polarization of the membrane induces an intracellular counter-field which significantly mitigates the field influence from outside (short arrows inside the cell). Because of the low field strength inside the cell, a possible polarization of the membranes of cell organelles can be neglected, at least at DC, or low-frequency AC fields.

The transmembrane potential difference $\Delta\psi_{ind}$ induced by an external field, corresponds to the difference $\psi_{ref} - \psi_{ind}$ in Fig. 3.43. There is no potential difference across the resistor R_i and no current flows inside the cell.

Because of the deformation of the external field by cell polarization (Fig. 3.44a), the potential ψ_{ind} (Fig. 3.43) is somewhat different from the potential at an identical x -coordinate away from the cell. This was taken into consideration introducing the resistor R_{e1} and a bending of the extracellular potential function near the membranes in the lower part of Fig. 3.45. The extent of this deviation depends on the shape and the radius (r) of the cell. In the case of a spherical cell, the characteristic distance from the center of the spherical cell is $1.5r$.

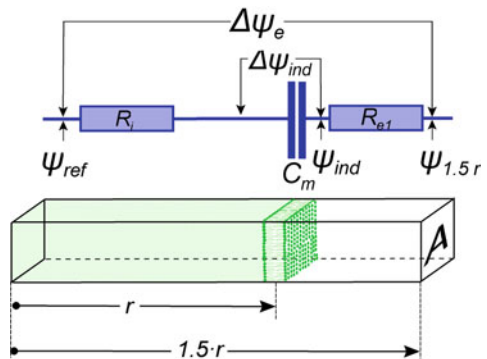
The polarity of the induced potential difference ($\Delta\psi_{\text{ind}}$) at both sides of the cell corresponds to the external field, whereas the undisturbed in vivo transmembrane potential $\Delta\psi_M$ is always oriented inside-out. Therefore, the cell is polarized in the same direction as the induced potential difference ($\Delta\psi_{\text{ind}}$) on one side, and oriented in an opposite direction on the other side. This means that the resulting potential differences ($\Delta\psi_1, \Delta\psi_2$) differ from one another on both sides of the cell. The in vivo potential at locations oriented perpendicular to the field lines will not be influenced by the external field at all. For this reason, the reference level for the induced potential (ψ_{ref}) is identical inside and outside the cell in Fig. 3.45.

To calculate the induced membrane potential (ψ_{ind}), a small column can be considered. It is oriented in the field direction and cut out along a line through the center of the sphere with the cross-sectional area A . The characteristic length up to which cell polarization may enhance the external medium potential is $r_e = 1.5r$. Furthermore, we shall consider that the Ohmic current flow through the membrane resistor R_m (Fig. 3.43) can be neglected. In this case the following relations apply:

$$R_i = \frac{r}{g_i A}; \quad R_{e1} = \frac{0.5r}{g_e A} = \frac{r}{2g_e A}; \quad C_m = \frac{C}{A} \quad (3.213)$$

Furthermore, the time constant (τ) can be considered reflecting the characteristic time to charge the membrane capacitor:

Fig. 3.46 A column with a constant cross-sectional area A , cut out from the cell (Fig. 3.43), and the electric scheme for the corresponding circuit. The resistor R_m and the capacitors for the cytoplasmic and external media are neglected (Corresponding to the approach of Gimsa and Wachner 1999)



$$\tau = C(R_i + R_{e1}) = r C_m \left(\frac{1}{g_i} + \frac{1}{2g_e} \right) \quad (3.214)$$

Considering an external field of strength \mathbf{E} , the potential difference outside the cell, from point ψ_{ref} to $\psi_{1.5r}$ is: $1.5r\mathbf{E}$. The same potential must drop over the column. Using the proportionality between the impedance, as effective AC resistance, and the potential drop, and considering the impedance of the membrane capacitor: $Z^* = 1/Y^* = -j/\omega C$, we get:

$$\frac{\Delta\psi_{ind}}{1.5Er} = \frac{-j/\omega C}{-j/\omega C + R_1 + R_{e1}} = \frac{-j}{-j + \omega C(R_1 + R_{e1})} \quad (3.215)$$

Introducing the time constant of Eq. 3.214 and rearranging, leads to:

$$\frac{\Delta\psi_{ind}}{1.5Er} = \frac{-j}{-j + \omega\tau} \quad (3.216)$$

Noticing that:

$$|-j + a| = \sqrt{1 + a^2}$$

We obtain for the absolute value of the induced transmembrane potential:

$$\frac{\Delta\psi'_{ind}}{1.5Er} = \frac{1}{\sqrt{1 + \omega^2\tau^2}} \quad (3.217)$$

After rearrangement and introduction of the expression τ from Eq. 3.214 one gets:

$$\Delta\psi_{ind} = \frac{1.5 r E}{\sqrt{1 + \left[r\omega C_m \left(\frac{1}{g_i} + \frac{1}{2g_e} \right) \right]^2}} \quad (3.218)$$

This is the maximum of the membrane potential induced in the direction of the external field. At an angle perpendicular to the field (Fig. 3.45) the induced transmembrane potential vanishes. To consider the induced potential at all points of the sphere, the radial coordinate α can be introduced, and Eq. 3.218 must be multiplied by $\cos\alpha$.

Using common cell parameters, like for example: $r = 10^{-5}$ m, $C_{sp} = 10^{-2}$ F m⁻², $g_i = 0.5$ S m⁻¹, and $g_e = 1$ Sm⁻¹ in Eq. 3.217 a relaxation time of $\tau = 2.5 \cdot 10^{-7}$ s is obtained. Introducing this parameter into Eq. 3.18 it is easy to demonstrate that for low-frequency AC fields ($\nu < 10^5$ Hz), the denominator of this equation will approach 1.

For DC fields and low-frequency AC fields, including the dependence of the vector angle α , the equation reduces to:

$$\Delta\psi_{\text{ind}} = 1,5 E r \cos \alpha \quad (3.219)$$

As already mentioned, this, as well as Eq. 3.218, is correct only for a very high membrane resistance which is justified in most cases. Using this equation one can for example calculate that in a spherical cell with a diameter $r = 10 \mu\text{m}$ a superimposed transmembrane potential of $\Delta\psi_{\text{rel}} = 1.5 \text{ V}$ will be induced at position $\alpha = 0^\circ$ by an external low frequency, or DC field of approximately $E = 100 \text{ kV m}^{-1}$.

As will be considered in the next section, this equation is useful to calculate the field strength which is necessary to manipulate cells by electric membrane break down or cell-cell fusion. Conversely, it must be noted that it is not applicable to the calculation of stimulus thresholds for muscle and nerve tissues (see Sect. 4.6.2). In this case, there are complicated situations of field distribution in the intercellular space. Furthermore, these cells are extremely elongated. Finally, in some cases they are electrically connected to one another by gap junctions. In this case, not the length of the individual cell in the field direction is representative of the induced membrane potential, but the length of the whole electrically connected system.

Further Reading

Gimsa and Wachner 1999; Grosse and Schwan 1992.

3.5.5 Manipulation of Cells by Electric Fields

The interaction of cells with external electric fields as discussed in the previous section has led to various applications in biotechnology. In contrast to the effects of weak electric and electromagnetic fields on biological systems, which we will discuss later (Sects. 4.6, 4.7), these applications require rather strong fields. Field-induced charges that accumulate at cell-medium interfaces on the one hand influence the transmembrane potential ($\Delta\psi$), and on the other directly induce mechanical forces which are able to move or deform the cells.

The application of strong electric fields in physiological solutions with considerable conductivities will, of course, induce significant Joule-heating. The electrical power, absorbed in a medium per volume equals $E^2 g$ (see Eq. 4.23). For a specific conductivity in a physiological milieu of approximately $g = 1 \text{ S m}^{-1}$, the applied field strength of $E = 10^5 \text{ V m}^{-1}$, as required for these applications, results in an absorbed power density of 10^{10} W m^{-3} , i.e., 10^7 W kg^{-1} . This leads to an enormous local heating of the system which can only be avoided by the use of short field pulses, an artificial medium of low conductivity, or microscopic electrode systems with relatively large heat-conducting surfaces.

Let us first consider the influence of strong electric fields on the hyper-, or hypopolarization of the membrane. As explained in the previous section, an externally applied DC field increases the membrane potential ($\Delta\psi_I$ in Fig. 3.45) on one side of the cell, and consequently, raises the internal electric field strength in the membrane at this point. As mentioned in Sect. 2.2.1, the electric transmembrane field in vivo is nearly 10^7 V m^{-1} . If the transmembrane potential is artificially elevated, reaching an amount of approximately one volt, the internal field strength becomes overcritical, and the membrane will be destabilized. In this so-called *electric break down* the membrane loses its property as a diffusion barrier. Its electrical resistance breaks down, which may even lead to complete destruction of the cell by subsequent lysis. However, by selecting the parameters of treatment properly, i.e., using moderate shape, amplitude, and duration of the pulse at optimal temperature and electrolyte conditions, a reversible electric break down can be induced. This additionally requires use of the proper composition of the external medium to avoid Donnan-osmotic cytolysis of the cells (see Sect. 3.2.5). In this case, the membrane may stabilize again and the cell may survive. The duration of the applied field pulses is of the order of 1–500 μs .

The electric break down of cell membranes is used for so-called *electroinjection* or *electroporabilization*. When cells are suspended in a solution containing macromolecules or small particles, the short moment of the induced membrane destabilization may be sufficient to allow the penetration of these molecules or particles into the cell with sufficient probability. Additionally, such disturbances of the membrane structure can also increase the ability of the cell to absorb particles by phagocytosis. Electroinjection is used for example to introduce drugs, genetic materials, or even small organelles into cells.

The basic equation for the external field which is necessary to reach the required membrane field for these manipulations was introduced in the previous section (Eqs. 3.218, 3.219). It indicates that the external field strength (E) which is required to induce an electric break down is inversely proportional to the cell radius (r). Stronger external fields therefore are needed for smaller cells. Recently, conditions were investigated to achieve break down also in intercellular organelles, such as mitochondria or chloroplasts. For this, nanosecond pulses are used with field strengths up to 10^7 V m^{-1} .

If two cells are attached to each other during application of an electric pulse, or a short time after the pulse, the points of induced membrane instability may interact with each other, leading to a fusion of the lipid phases of the membranes of the neighboring cells. Subsequently membrane fusion, an osmotically governed process of fusion of the whole cell may follow. This process of *electrofusion* of cells has recently been explained by several mechanisms which cannot be discussed here in detail. Membrane and subsequently cell fusion are in fact triggered by electric pulses, but need a time longer than the pulse duration to be accomplished.

Electrofusion of cells is broadly applied in biotechnology. In contrast to the induction of cell fusion by some chemical agents, electrofusion has the advantage of being much more selective. Using sufficiently small electrodes, it is possible to fuse even two particular cells. Meanwhile, methods have been developed to bring

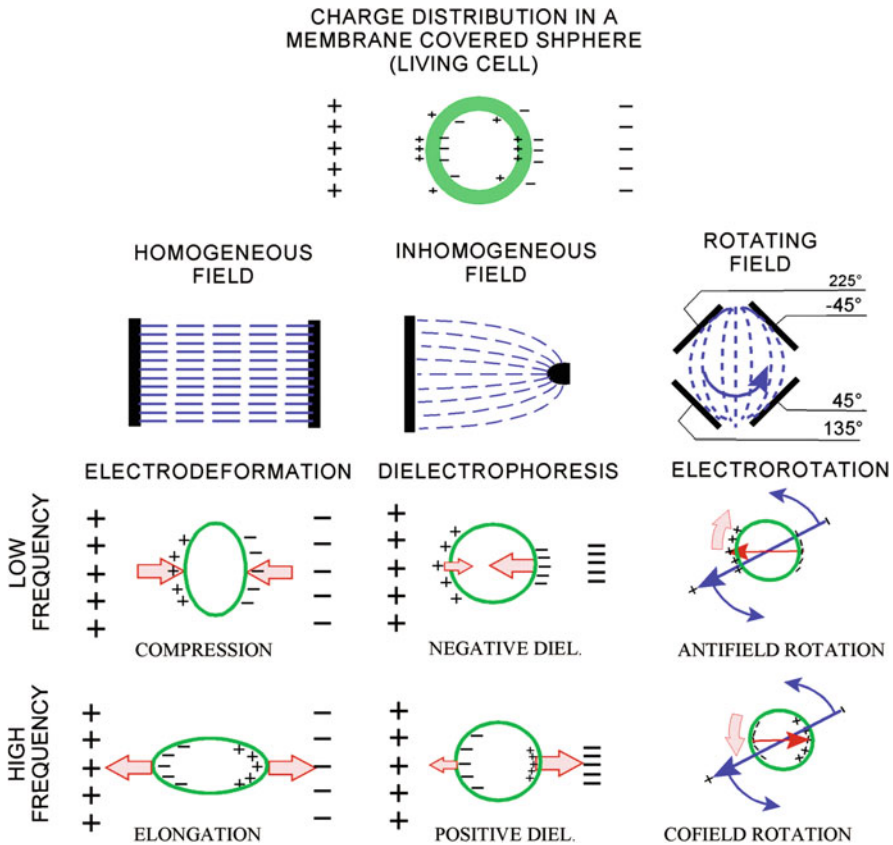


Fig. 3.47 Various electromechanical effects on a spherical cell model. Above: general distributions of charges near the membrane in relation to electrode charges. Below: snapshots of charge distributions in homogeneous, inhomogeneous, and rotating AC-fields at different frequencies

specific cells into contact with each other. Usually, an aggregation by dielectrophoresis is used for this purpose.

Charge separation, i.e., the induction of cell dipoles of various quality, leads to a number of electromechanical phenomena. These can be used either to separate and mechanically manipulate cells, or to directly measure dielectric properties of individual cells (Fig. 3.47).

In general, two qualitative cases of cell polarization exist. At frequencies below membrane dispersion the effective charge distribution is opposite to the case at higher frequencies, where the membrane is fully capacitively bridged. Corresponding to this, at low frequencies cells will be compressed by the field, whereas they will be stretched in high frequency fields.

In inhomogeneous fields, the cell hemisphere in the higher field area experiences a higher polarization, i.e., a higher force than its opposite. This imbalance leads to a

Fig. 3.48 The dielectrophoresis of yeast cells in an inhomogeneous field which spans between the peak of a needle, shown as a small hemispheric electrode, and a flat electrode opposite. The figure shows the accumulation of the cells near this electrode as a result of positive dielectrophoresis. The polarization of the cells additionally leads to their mutual attraction and the *pearl chain formation*

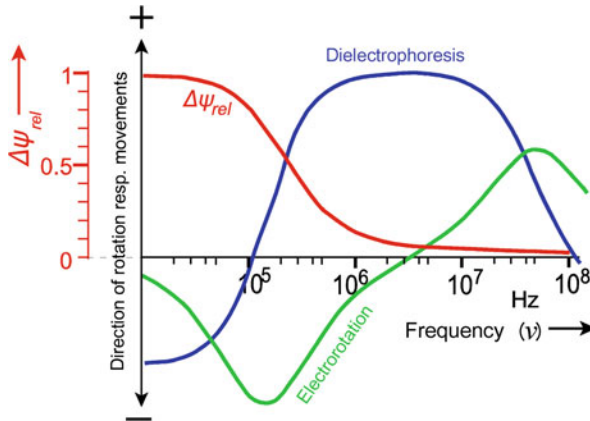
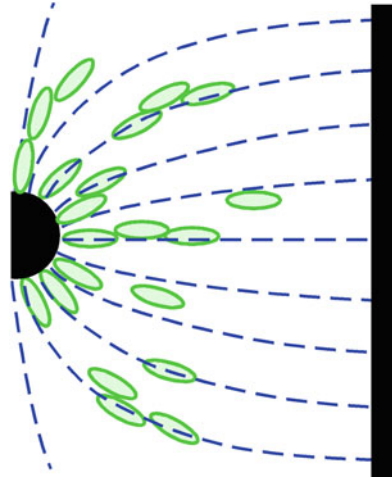


Fig. 3.49 An example of electrorotation (green line) and dielectrophoresis (blue line) of a cell in a solution of low conductivity, dependent on the applied frequency. Positive values of the electrorotation spectrum mean spinning in the direction of field rotation (cofield rotation), or in the case of dielectrophoresis a movement into the direction of higher field intensity, and vice versa. Additionally, the parameter $\Delta\psi_{rel} = [1 + (2\pi\nu)^2]^{-1/2}$ is depicted as a function of frequency (red line) according to Eq. 3.218. It describes the influence of the induced membrane potential (with the help of Wachner and Gimsa)

translocation of the cell which is called *dielectrophoresis* (Fig. 3.48). The direction of the dielectrophoretic translation depends on the orientation of the effective polarization of the cell and on the orientation of the field gradient. According to the polarization effects there are frequency regions of *negative* as well as of *positive* dielectrophoresis (see Fig. 3.49). The driving forces in all electromechanical processes are proportional to the square of the field strength (E^2). Because of the deformation of the field around the individual cells (see Fig. 3.44), the gradient of

the fields directly near the cells results in mutual dielectrophoresis, i.e., in the attraction of cells to each other. This leads to the formation of pearl chain-like structures. This effect can of course occur also in homogeneous external fields.

As depicted in Fig. 3.47 in rotating fields the dipole induction results in *electrorotation*. Rotating fields can be applied using multielectrode systems. In the case of a four-electrode system, as shown in this scheme, a generator produces an AC signal with adjustable frequency, which is split into four signals of particular phase relations. In this case, the field rotates at the frequency of the generator. Correspondingly, the induced dipole rotates too. Because of the time constant of polarization, an angular difference however may occur in between the induced dipole, and the external field vector. This leads to a permanent force of interaction between them, either in the form of repulsion, resulting in an *antifield rotation* at low frequencies, or attraction at higher frequencies, the so-called *cofield rotation*. The resulting rotation frequency of the cells in the chamber is much smaller than that of the field, and again is proportional to the square of the applied field strength. In contrast to dielectrophoresis, which can be described by the real part of the induced dipole, electrorotation is related to its imaginary component.

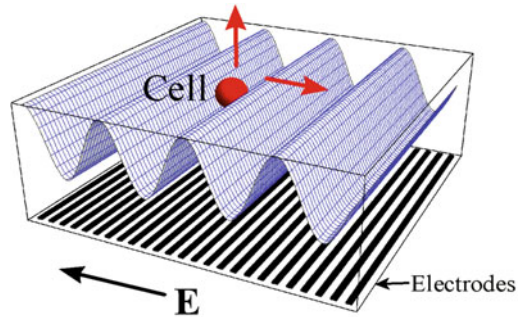
In Fig. 3.49, as an example, the frequency dependence of electrorotation of a cell in a low conductivity medium is depicted. It shows that the antifield rotation occurs at frequencies of about 100 kHz, followed by a point of zero rotation and finally, in the MHz region a maximum of cofield rotation occurs at about 50 MHz.

Figure 3.47 represents the simplest case, a single shell model, which considers a dielectrically homogeneous sphere, covered by a membrane. It represents a spherical cell without electrically significant organelles (Fig. 3.43). Non-nucleated swollen erythrocytes can be described by this model. In this case, the radius of the cell and the membrane thickness are included as geometrical parameters in the corresponding equations, as well as the conductivities and permittivities of the membrane, the cytoplasm, and the external medium. These model considerations indicate that the peak of the antifield rotation in its frequency position, and its amplitude, reflects the membrane properties of the cell. This first characteristic frequency is determined by the time constant given by Eq. 3.214. Changing the permeability of the membrane as a diffusion barrier, this peak vanishes. The maximum of the cofield rotation, i.e., the second characteristic frequency, indicates the conductivity and permittivity of the cytoplasm in relation to that of the external milieu.

Electrorotation and dielectrophoresis are approved methods to measure dielectric properties of individual living cells. There are a number of investigations indicating that electrorotation can measure alterations of cell properties which are induced by drugs, toxic agents, virus attacks, cell activations, and other events. Under special conditions even properties of organelles like nuclei or vacuoles can be measured. Automatic video systems and methods of dynamic light scattering are applied to register the induced movement.

In contrast to electrorotation, which is mostly used for cell analyses, dielectrophoresis can also be applied for preparative cell separation and other

Fig. 3.50 Cells moving in a system of interdigitated travelling-wave electrodes. Note that the direction of field propagation is opposite to the direction of cell motion according to negative dielectrophoresis (After Fuhr et al. 1996)



biotechnologically interesting techniques. As already mentioned, the formation of cell chains, or cell-cell attachment by dielectrophoresis is used to perform selective electrofusion of cells by applying short pulses of high field intensity. Recently, special microdevices have been constructed to investigate single cells by electrorotation and to manipulate them by dielectrophoretic translations. According to the large relative surface of such chambers in relation to their volume, heating of the samples was minimized. This allows one to apply dielectrophoresis and electrorotation also in physiological solutions of relatively high conductivities. It is also possible to produce microscopic field traps, in which cells, lifted by dielectrophoretic force can be held in a stable position without any surface contact. Using electrode arrangements inducing traveling waves (Fig. 3.50), cells can be moved on microchips. This new technique opens enormous possibilities for biotechnological applications.

Further Reading

For electrorotation and dielectrophoresis: Georgiewa et al. 1998, Gimsa and Wachner 1998; Fuhr et al. 1996; Fuhr and Hagedorn 1996; electromanipulation and electrofusion: Cevc 1990; Kolb et al. 2006; Lynch and Davey 1996, Zimmermann 1996; dielectric cell deformation: Sukhorukov et al. 1998; Ziemann et al. 1994.

3.6 Mechanical Properties of Biological Materials

Biomechanics is a branch of biophysics that examines mechanical properties of biological materials and forces acting upon and within the biological structure, as well as effects produced by such forces. It explains the anatomical stability of plants and animals, the movements of limbs, and in this way the mechanics of walking, flying, and swimming. A special branch of biomechanics, biorheology concerns the mechanics of blood flow and other kinds of fluid movement inside the organism. Hearing and other kinds of acoustic organs can be explained on the basis of various kinds of cellular mechanoreceptors.

Biomechanics is in fact one of the oldest branches of biophysics. Its development followed closely that of physical mechanics itself. In the period of renaissance the pioneers of mechanics, such as Gallileo Galilei, René Descartes, Isaac Newton, and many others were always also interested in the mechanics of animals. The first classical book on biomechanics, Alfonso Borelli's "De motu animalium" was printed in Rome in 1680. It already contained basic physical considerations on swimming, flying, and movement of animals as well as various calculations of moments of human limbs and the spine under conditions of loads. As we already mentioned in Sect. 1, this book marked the beginning of medical physics, which was called at that time "iatro-physics." Furthermore, D'Arcy Thompson's book "On Growth and Form," published first in 1917 must be mentioned. This book analyzed for the first time basic processes of cell mechanics, shape formations, and many other biomechanical fields.

In recent decades biomechanics has become increasingly important in diagnostics and therapeutics, especially the biophysics of blood circulation (hemorheology), including the properties of blood vessels and the pumping properties of the heart, the biomechanics of the skeleton, as well as of limbs and joints, all of which form the basis for medical physics and medical engineering.

A further interest in special questions of biomechanics comes from sport. Here various kinds of complex body motions are of interest to optimize outcome. There is also interest in biomechanics from ecology, considering life in moving fluids, wind resistance of plants, etc.

Further Reading

On biomechanics in general, Alexander 2003; Bels et al. 2003; Chein et al. 2008; Fung 1993; Nigg and Herzog 2007; Niklas 1992; Oomens et al. 2009; Özkaya and Nordin 1999; Skalak and Chien 1987; Vogel 1994.

3.6.1 Some Basic Properties of Fluids

Many parameters characterizing the physical property of a fluid are defined for the case of laminar flow. A flow is called *laminar* if particles of the fluid move in parallel layers to each other (for more on the properties of laminar flow see Sect. 3.7.1). Figure 3.51 illustrates examples of three kinds of laminar flow. In contrast to turbulent flow, processes in laminar flows can be calculated by linear thermodynamic approaches (see Sect. 3.1.3, Fig. 3.3).

To understand the approaches of fluid mechanics, and to characterize laminar fluid profiles, we first need to introduce the *velocity gradient* (γ), also known as *shear rate*. It is the derivation of the streaming velocity (\mathbf{v}) with respect to a coordinate (z), perpendicular to the direction of the flow.

$$\gamma = \frac{d\mathbf{v}}{dz} \quad (3.220)$$

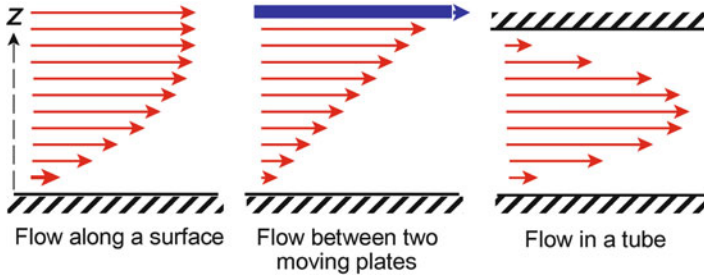


Fig. 3.51 Various examples of laminar velocity profiles

The measure of this velocity gradient is therefore s^{-1} .

The shear rate of a moving fluid is at a maximum near a solid surface, or in the case of a tube, near the wall. Far away from the surface this gradient becomes zero.

If two parallel plates slowly move one in relation to the other, they generate a laminar flow in between, the velocity gradient of which is always constant, and is equal to the relative velocity of the plates, divided by their mutual distance ($\gamma = \Delta v / \Delta z$).

The force (**F**) driving a plate with a surface *A* in the case of laminar flow is proportional to the velocity gradient (γ) between the two plates:

$$\mathbf{F} = \eta \gamma A \quad (3.221)$$

In this equation a friction coefficient (η) is introduced which is called *viscosity*. The viscosity therefore determines the force, which is required to move a plate with an area of 1 m^2 , at 1 m distance from a parallel surface at a velocity of 1 ms^{-1} , if a laminar flow of a liquid between these surfaces is induced. Thus, the measuring unit for the viscosity is: Nsm^{-2} , or: $\text{Pa} \cdot \text{s}$. Sometimes an older unit *P* (Poise) is used, whereas: $1 \text{ P} = 0.1 \text{ Nsm}^{-2}$.

Parallel to the viscosity (η), another parameter, the *fluidity* ($\varphi = 1/\eta$) is used, as well as the *kinematic viscosity* ($\nu = \eta/\rho$), a parameter, containing the density (ρ) of the fluid. The force deforming a body in a streaming fluid with a velocity gradient (γ) is given by the *shear stress* (τ):

$$\tau = \eta \gamma \quad (3.222)$$

The viscosity depends to a high degree on the temperature. Especially for aqueous solutions this is caused by the cluster structure of the water (see Sect. 2.2.2). In contrast to the viscosity of pure water at $T = 0^\circ\text{C}$, which is 1.79 mPa s , it amounts at 25°C to only 0.89 mPa s , and at 100°C , finally to 0.28 mPa s (see Fig. 2.16 in Sect. 2.2.2).

In order to characterize the influence of dissolved or suspended substances on the viscosity of a fluid, the following derived parameters are used:

Relative viscosity: $\eta_{\text{rel}} = \frac{\eta}{\eta_w}$

Specific viscosity: $\eta_{\text{sp}} = \eta_{\text{rel}} - 1$

Reduced viscosity: $\eta_{\text{red}} = \frac{\eta_{\text{sp}}}{c}$ (in: 1 mol^{-1})

Intrinsic viscosity: $[\eta] = \lim_{c \rightarrow 0} \eta_{\text{red}}$ (in: 1 mol^{-1})

where η is the viscosity of the solution or suspension, η_w is the viscosity of the pure solvent, and c is the molar concentration of the solute.

The viscosity increases with increasing concentration of the dissolved or suspended substances. As already pointed out in the definition of the intrinsic viscosity $[\eta]$, even the reduced viscosity of a solution is a function of the concentration. The intrinsic viscosity contains information on the structure and the molecular mass of a substance.

For diluted suspensions of small rigid spherical particles the *Einstein relation* can be applied:

$$\eta_{\text{sp}} = 2.5 V_{\text{rel}} \quad (\text{for } V_{\text{rel}} < 0.1) \quad (3.223)$$

The relative volume (V_{rel}) is the volume of all particles in the suspension together in relation to the volume of the suspension. For suspension of cells (sperms, erythrocytes, etc.) the term *cytocrit* (or specifically *spermatocrit*, *hematocrit*) is used. It should be emphasized that neither the absolute size of an individual particle, nor the homogeneity of the diameters of all particles in the suspension are of importance for this relation. The Einstein equation, however, is correct only for very diluted suspensions.

Fluids, the viscosity of which is independent of the velocity gradient ($\dot{\gamma}$) are called *Newtonian fluids*. In contrast, *non-Newtonian fluids* alter their viscosity depending on this parameter.

In Fig. 3.52 the behavior of various kinds of non-Newtonian fluids are demonstrated. *Dilatant fluids* are mostly suspensions of solids, like quartz particles. They produce entropy by mutual friction as much as the shear rate of fluid increases. The *Bingham-plastic* behavior occurs, for example, in a suspension of

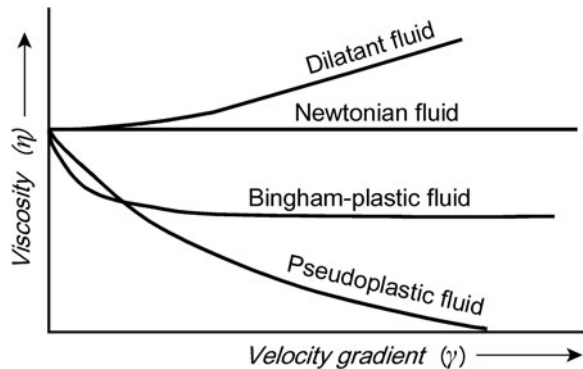


Fig. 3.52 The dependence of the viscosity (η) on the velocity gradient ($\dot{\gamma}$) for a Newtonian fluid (—), and for various types of non-Newtonian fluids (—) (After Glaser 1989)

nonspherical particles. In this case velocity gradients lead to their orientation, which decreases the viscosity of the suspension. At certain points, if the particles are oriented at maximum these suspensions behave like Newtonian fluids. The same behavior is to be expected if the particles tend to aggregate in the resting fluid, but disaggregate at low shear stress.

The most common property of biological fluids is the *pseudoplastic* behavior. It occurs for example in blood (Sect. 3.6.2, Fig. 3.54), and many other biological fluids with heterogeneous composition. Different components of these fluids, such as for example blood cells, proteins and other macromolecules, aggregate, orientate, and deform at various shear gradients. The resulting function, therefore, does not come to a saturation level at reasonable values of γ .

These shear-induced processes of course need some time to become established. In the case of spontaneous induction of a shear gradient, or of a sudden change of it, a time dependence of the viscosity occurs. This behavior was first observed in 1923 in gels, which could be transformed by shaking into a liquid sol. The same property was also found in the viscous behavior of cell protoplasm. The term “thixotropy” was introduced as a combination of the Greek words *thixis* (stirring, shaking) and *trepo* (turning or changing). According to the IUPAC terminology: *thixotropy* is defined as the continuous decrease of viscosity with time when flow is applied to a sample that has been previously at rest and the subsequent recovery of viscosity in time when the flow is discontinued. Confusion between thixotropy and shear thinning still persists in some cases in the literature. Therefore, it should be emphasized that thixotropy applies to the time dependence of non-Newtonian behavior of a liquid.

The opposite phenomenon also exists, i.e., a reversible, time-dependent increase in viscosity. This could be the result of flow-induced aggregation of particles. This is called *antithixotropy*, earlier known as *rheopexy*. In the next section we will demonstrate these properties for the case of the behavior of blood.

These properties of fluids must be taken into account, when choosing instruments to measure the viscosity. Some of the most common methods are depicted in Fig. 3.53. In the case of a capillary viscosimeter (Fig. 3.53a) the time is measured which a given fluid needs to pass a capillary under a known pressure. For this, a certain volume of the fluid is placed in the left part of the U-tube and the time is measured for the fluid to pass the two marks M_1 and M_2 . This time is proportional to the viscosity. In this way the viscosity can be measured after calibration of the setup using a fluid with known viscosity. This so-called Ostwald viscosimeter was later modified by Ubbelohde in such a way that a further vertical tube arranged at the end of the capillary leads to an interruption of the fluid. Another way, to measure the viscosity of a fluid is the use of falling bodies (Fig. 3.53b). Both methods have the advantage of being simple and in relation to other equipment, inexpensive. However, conversely they are only suited to measuring the viscosity of Newtonian fluids because the velocity gradient of the streaming fluid in the capillary, as well as between the falling sphere and the wall of the tubes, are not at all constant (see Fig. 3.51).

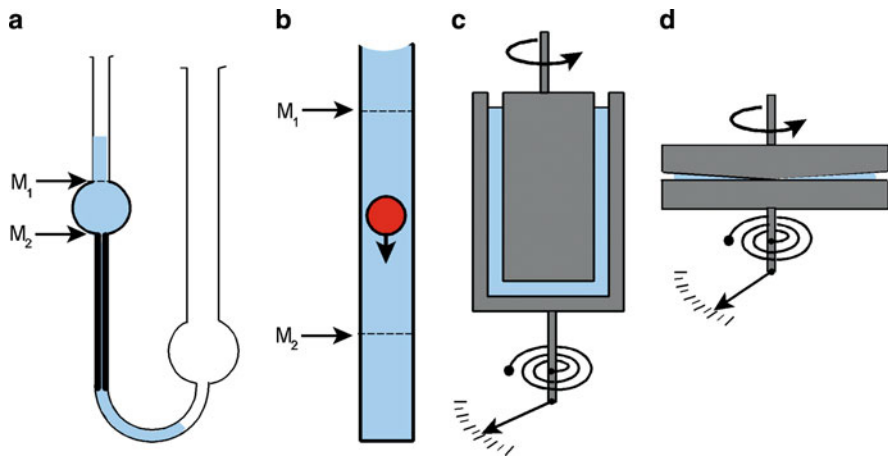


Fig. 3.53 Various setups to measure the viscosity of fluids. (a) Capillary viscosimeter after Ostwald, (b) viscosimeter with falling sphere, (c) coaxial-type rotational viscosimeter, (d) cone-type rotational viscosimeter

To investigate the viscosity of non-Newtonian fluids measuring instruments are required which allow one to establish well-defined shear gradients. As demonstrated in Fig. 3.51, this is possible between plates moving parallel to each other. This principle is used in so-called *Couette*, or *rotational viscosimeters*. In this case the fluid is placed in the gap between two coaxial cylinders (Fig. 3.53c), or between a flat plate and a truncated cone (Fig. 3.53d). By moving one part of this equipment, the viscosity of the fluid transmits a torque which can be measured. Usually, one part is rotated with adjustable speed, and the torque of the opposite part is measured by a sensitive instrument. By changing the speed and the thickness of the cleft, one can produce various velocity gradients. In the case of the rotating cone the tangential velocity is compensated by the increasing broadness of the cleft. Therefore, in this case a constant velocity gradient is also established. Curves like those shown in Figs. 3.52 and 3.54 are produced using such instruments.

Recently, the *Stabinger viscosimeter* has been used as a modification of the classic Couette rotational viscosimeter. In this case, the internal cylinder is hollow and thus floats freely in the sample, centered by centrifugal forces. This avoids any bearing friction. The torque of this cylinder is implemented by a rotating magnetic field.

Further Reading

On thixotropy, Mewis and Wagner (2009).

3.6.2 The Viscosity of Biological Fluids

Through the development of sensible rotational viscosimeters it was possible to investigate viscous properties of a large number of biological materials, such as

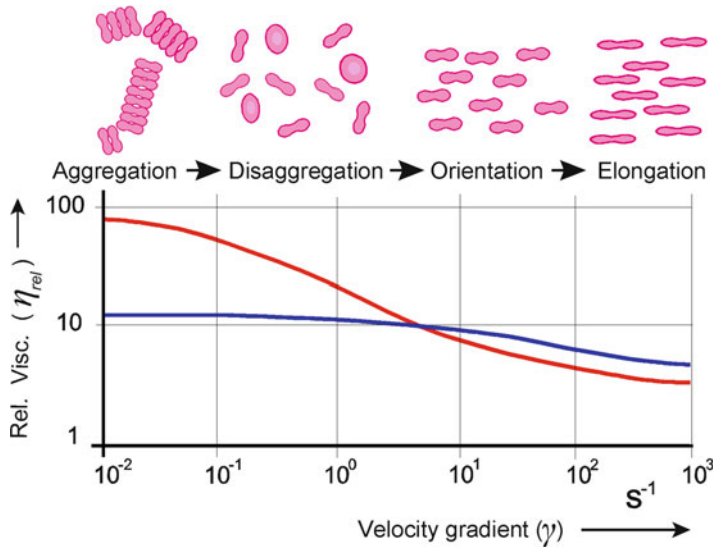


Fig. 3.54 The relative viscosity (η_{rel}) dependent on the velocity gradient ($\dot{\gamma}$) of human blood (—), and heat-hardened erythrocytes, which were resuspended in plasma (—). The difference of both curves in the region of low shear rates is achieved by aggregation, disaggregation and elongation of the native red blood cells at increased shear rates. (Data from Lerche and Bäumler 1984)

blood, lymph, various secretions, synovial fluids, and many others. This has led on one hand to a better understanding of the processes of blood flow, of the mechanics of joints, etc., and on the other hand it has become a useful indicator in diagnostics.

In Fig. 3.54 the viscosity of a suspension of red blood cells is depicted as a function of the velocity gradient. In contrast to blood plasma which appears to be a Newtonian fluid, these suspensions, as well as the whole blood, indicate pseudoplastic thixotropic behavior. The reason for this is complex: Native erythrocytes, suspended in blood plasma aggregate at low shear rates. These aggregates are not very stable and disaggregate at somewhat higher shear rates. This disaggregation lowers the viscosity of the suspension. A further increase of the shear rate leads to orientation, and finally to a deformation of the erythrocytes, further decreasing the viscosity. Looking at the behavior of erythrocytes which were hardened by fixation, neither the aggregation, nor deformation caused by shear stress occurs. Living erythrocytes in the shear gradient become elongated ellipsoids that are oriented along the streaming vectors. With increasing shear stress they become more and more elongated. Whereas slow elongations are reversible up to a certain degree of elongation, irreversible alterations in the membrane occur. Eventually hemolysis occurs as a result of maximal shear stress.

The energy of the streaming fluid which leads to deformation of the cells corresponds to the shear stress (τ) according to Eq. 3.222. This means that it depends not only on the velocity gradient ($\dot{\gamma}$), but additionally on the viscosity (η)

of the fluid. To investigate shear-induced shape deformations therefore, high viscosity solutions are usually used, for example, solutions of high molecular weight dextrans of various concentrations.

For biomechanical problems in orthopedics the properties of the synovial fluid which is located in the joint cavity and surrounded by the joint capsule are of particular interest. In the language of tribology, which represents the effects of friction in moving machines, the mechanism of joints represents a kind of *depot lubrication* with porous surfaces. The expression “depot lubrication” points to the synovial fluid which is accumulated in the bursa of the joints. “Porous surface” refers to the articular cartilage which covers the cortical bone in the joint capsule by a 0.3–0.5-mm thick layer. The joints are not only burdened by movement, i.e., by a shear stress of the synovial fluid, but additionally in some cases by a considerable static pressure. It must be guaranteed that these loads do not press the synovial fluid out of the cleft. In fact the synovial fluid has thixotropic pseudoplastic properties. A large viscosity of this fluid in the joints (η between 1 and 40 Pa s) prevents its exclusion from the cleft by hydrostatic pressure. If the joint is moving however, shear rates up to 10^5 s^{-1} appear. In this case, the viscosity of the synovial fluid decreases to 10^{-2} Pa s, leading to a highly effective lubrication of the joint. This particular property of synovial fluid is caused by a special structure of proteoglycans, which are high molecular weight glycoproteins with an intriguing structure.

This leads us to considerations of the viscosity of microscopic, or even supra-molecular structures. It must be pointed out that the definition of the viscosity, as given in the previous Sect. 3.6.1, again comes from continuum physics. It does not take into account the behavior of molecules and supramolecular composition of the liquids. It just considers a homogeneous, continuous fluid without any structures. Already the non-Newtonian behavior discussed above indicates some limitations of this assumption. Moreover, problems arise, if we consider viscoelastic properties of cells and their constituents. At the beginning of the twentieth century, at a time when the physical properties of cells were first being discussed, many investigators measured the viscosity of the cytoplasm in relation to various biological functions. This question from a modern point of view is outdated because of our knowledge of the highly organized structure of this region. Even in cells with pronounced cytoplasmic streaming as in amoebas or in plant cells the question of the origin of these movements and of the molecular mechanisms driving these flows, is of more central interest than phenomenological models using plasma viscosity in general.

The problem of viscosity in molecular and supramolecular dimensions is important to a high degree in relation to the “viscosity” of the cell membrane (see also Sect. 2.3.4). Various molecular probes allow us to obtain physical parameters of the membrane, which in any case are functions of the viscosity. So, for example, a lateral diffusion constant of fluorescent labels in the membrane can be measured. Using the Einstein equation (Eq. 2.39) some conclusions can be made on the viscosity of their surrounding medium. In the same way the rotation diffusion constant of specific spin-probes bound to particular membrane molecules, measured by electron-spin resonance techniques (ESR), allows us to determine the viscosity. Furthermore, microscopic stress deformations can be applied.

In all cases the viscosity which is determined by these methods is a typical effective parameter. In the same way as for example the hydration radius of an ion (see Sect. 2.2.2), this parameter depends on the techniques and the physical phenomenon used for its determination. One should not consider these “quasi-viscosities” in the same sense as the phenomenological viscosity of a fluid which we discussed before. Moreover, the mechanical anisotropy of the membrane must be taken into account, which means that the mobility of particles is not equal in all directions of space.

Further Reading

Leyton 1975; Owens 2006; Skalak and Chien 1987.

3.6.3 Viscoelastic Properties of Biomaterials

The simplest kind of deformation of a body is its stretching, hence the basic parameters of viscoelasticity are defined for this case. In Fig. 3.55 this is illustrated by a stress–strain diagram, the coordinates of which are defined as follows:

$$\text{Stress: } \sigma = \frac{F}{A} \quad (\text{in : } \text{N m}^{-2}) \quad (3.224)$$

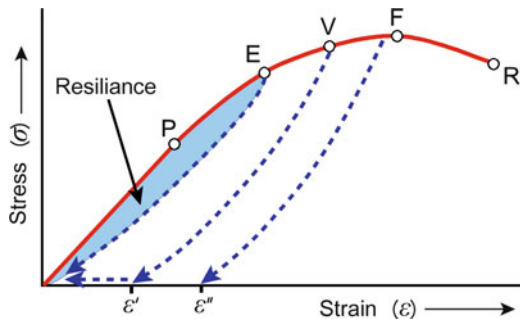
$$\text{Strain : } \varepsilon = \frac{\Delta l}{l} \quad (3.225)$$

where: F is the force, A is the cross-sectional area of the body, l its length, and Δl is the difference between the resting and the extended material. The strain (ε) is just a relation and therefore does not have a measuring unit.

In general this diagram indicates some regions of different behavior: In the region of minimal strain up to the limit of proportionality (P) *Hooke's law* is valid stating that the strain (ε) is directly proportional to the stress (σ). The ratio of these two properties is called the *modulus of elasticity*, or *Young's modulus* (Y).

$$Y = \frac{\sigma}{\varepsilon} \quad (3.226)$$

Fig. 3.55 A generalized stress–strain diagram. P – limit of proportionality, E – limit of elasticity, V – limit of reversible viscoelastic deformation, F – point of floating deformation, R – rupture point, ε' , ε'' – residual strains, blue area – resilience



Young's modulus (Y) has a measuring unit N m^{-2} or Pa. Because of the large amount of this parameter in common materials, mostly the units $1 \text{ GPa} = 10^3 \text{ MPa} = 10^9 \text{ Pa}$ are used.

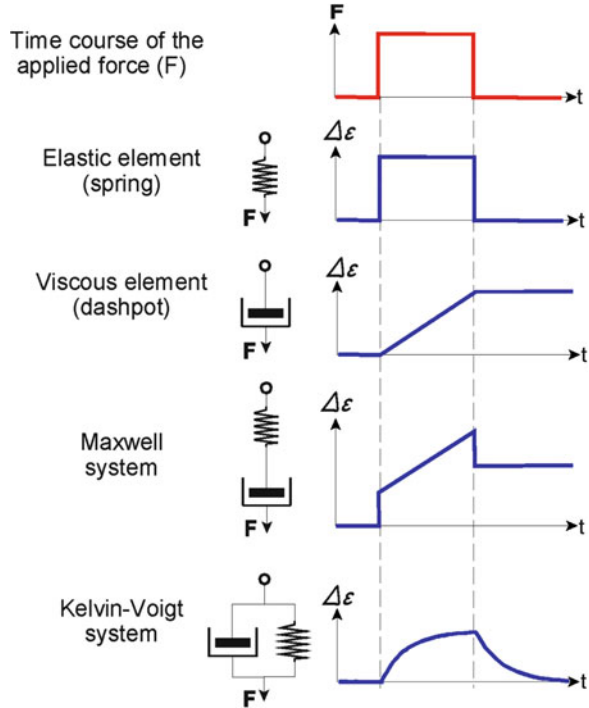
The linear relationship between stress and strain does not hold for large stress behind point P. The deformation however, is reversible up to the elastic limit (E). This means that the body will spontaneously and quickly return to its original length when the deforming force is removed. In this case, however, the relaxation curve (blue dotted line) did not follow the extension curve (red line). The area between these two lines is called *resilience*. In general the area in this plot, i.e., the product of stress and strain has the unit of energy. The resilience therefore represents the thermal energy which is dissipated during the process of extension and relaxation. In terms of irreversible thermodynamics it is the dissipation function (Φ) of this process (see: Eq. 3.64, Sect. 3.1.4). It results from the viscous friction within the body and is an expression of the nonideal behavior of the system.

If the strain is taken beyond the elastic limit (E), the body begins to deform irreversibly. Up to the limit of viscoelastic deformation (point V), the body quickly relaxes to some residual strain (ϵ') which eventually may slowly vanish. Overcoming this point, an irreversible deformation (ϵ'') persists after the stress has been removed. A further extension leads to the point where the body begins to show spontaneous flowing elongation, in spite of further increase in stress until it rips up at point R.

This stress-strain diagram indicates that the deformation of a body outside the limit of proportionality depends not only on elastic, but also on viscose properties of the material. For this the term *viscoelasticity* is used. Biological tissue shows viscoelasticity to a great extent. In this case not only the stationary stress-strain function is important, as demonstrated in Fig. 3.55, but also the kinetics of deformation. It is possible to simulate the strain behavior of elastic and viscoelastic materials by mechanical systems made up of elasticity elements, as well as of viscosity, or damping elements (see Fig. 3.56). In contrast to the spring, which, if it is massless, elongates immediately if a rectangular force function is applied, the viscous damping element (e.g., a dashpot) elongates with a constant velocity and remains in position if the force vanishes. A damping element such as a dashpot combined in series with a spring is called a *Maxwell element*. In the case of an applied rectangular function of force, the spring elongates immediately, then a further continuous elongation of the system proceeds. If the force vanishes, the spring contracts but the viscous part of the Maxwell element remains elongated. If a spring and a damping element such as a dashpot are connected in a parallel arrangement we obtain a *Kelvin-Voigt element*. In this way a sudden increase of the force (F), leads to an exponential elongation of the system which will contract in the same way if the force vanishes.

Real systems must be regarded as being made up of both Maxwell and Voigt elements in various combinations and with different properties (e.g., *Maxwell-Weichert models*). Correspondingly, complicated strain-relaxation graphs and complex kinetic behavior is to be expected. If the time constants of the viscous elements are large

Fig. 3.56 Mechanical models to demonstrate the kinetics of elongation [$\Delta\epsilon(t)$] of elastic, viscous, and viscoelastic elements after applying a rectangular function of force [$F(t)$](—)



in comparison to the length of the mechanical impulse, i.e., in the case of a short-term mechanical stress, viscoelastic systems can respond elastically.

Parallel to the strain function [$\epsilon(t)$] as a result of an applied function of force [$F(t)$] (*isotonic tension*) as illustrated in Fig. 3.56, an examination of materials is also possible by applying a definite strain function and measuring the resulting stress in time dependence [$\sigma(t)$] (*isometric tension*). In this case the Maxwell element shows an exponential decline of stress as a result of viscous elongation of the dashpot relaxing the spring.

The schematic graphs of Fig. 3.56 with their different combined mechanical elements are similar to the electrical RC circuits in Figs. 3.39 and 3.43. In fact there are some similarities in their kinetic treatment. In analogy to the electrical impedance which we discussed in Sect. 3.5.3, a kind of mechanical impedance can be formulated as the response of a viscoelastic system to time-varying mechanical forces. Measurements of this kind give information on basic mechanical properties of biological tissue.

The modulus of elasticity (Y in Eq. 3.226) of various materials differs in several degrees of magnitude. In contrast to the value for steel, which is about $2 \cdot 10^5$ MPa, there are only some tenth to hundreds of MPa for various kinds of rubber. This marks also the range for different elastic properties of biological materials. Resilin, the most elastic animal protein from locust tendons shows a Young's modulus even

below one MPa. For bone, values of several hundreds of MPa are determined. Wood, stressed along the grain arrives at an elasticity of up to 10^4 MPa.

For the molecular basis of deformations, two mechanisms must be considered: In the case of *steel elasticity*, which occurs in most crystalline materials, the elongation leads to changes of the atomic spacings in the crystal lattice. The atoms are forced to move from their equilibrium position of minimal internal energy to a higher energy level. The free energy that is stored up in the strained state is released during relaxation. This is the reason for the high elasticity modules and also for the only short amount of elastic strain of these materials near to the rupture point.

In the case of so-called *rubber elasticity*, which is typical for macromolecular materials, the deformation of the body is the result of molecular deformations. We discussed this type of elasticity in Sect. 2.1.6 (Fig. 2.11). The applied strain leads to a partial deconvolution and orientation of the molecules and as a result, to a reduction of their entropy. The subsequent relaxation is the result of the increase in entropy back to its maximum according to the second law of thermodynamics. This is the reason why this type of stretching mechanism is also called *entropy elasticity*. The characteristic differences to steel elasticity are the large amount of possible elongation, and the strong temperature dependence of rubber elastic materials.

A number of structure proteins like for example collagen, elastin, or resilin are responsible for the rubber elastic properties of tendons and ligaments. These substances also play an important role in the storage of mechanical energy in some periodic or suddenly occurring movements. Although these processes also occur in mammals, they are best investigated in insect jumping and flying. In the case of locust jumping for example, the muscles apply a tension to a system of tendons which takes up the energy and subsequently releases it, by means of a trigger mechanism to achieve an increase in power. In this way a power, i.e., a transformed energy per time can be generated, which is impossible to obtain directly from the muscle. In fact, this is the same mechanism that we use for archery. To carry out a jump, a grasshopper for example, needs a specific power of about 5 kW per kg muscle. This would exceed the maximal output of a muscle by ten times. Similarly, during flight a grasshopper stores about 20–30% of the energy of the oscillating movements of the wings using passive elastic elements.

The following characteristics of the viscoelastic behavior of biological systems, such as cells, tissues, and organs should be noted:

The occurrence of different regions of elasticity in the stress–strain diagram: Frequently cells and elastic fibers in a tissue are interconnected to each other forming a network. In this case the viscoelastic properties of this system result not only in the intrinsic properties of these fibers, but first of all in the construction of this network, and in the viscosity of the fluid within it. Such a network can be stretched easily as long as it is deformable as a whole. This is the first region in the stress–strain diagram with low Young's modulus. If the components of the network eventually are fully oriented by the applied stress, a further strain is possible only by stretching the fibers themselves. This means that the Young's modulus increases suddenly. The resulting stress–strain diagram therefore, does not indicate a

flattening of the curve as in Fig. 3.55, but on the contrary it becomes steeper at a certain degree of strain. This sort of behavior for example, occurs in the walls of blood vessels, where the situation becomes even more complicated due to the activity of the smooth muscles.

The regulation of the elasticity behavior: The viscoelastic properties of the network as described above can be controlled by biological processes. This for example is possible by loosening and fastening the points of connections between the components. The result would be a sort of viscous prolongation of the body. Another way to control the viscoelastic behavior is the alteration of the water content of the tissue. Changing the cell volume or the intercellular space in the network would change its elasticity modulus. This sort of control is best investigated for the case of viscoelastic properties of the uterine cervix at various periods of pregnancy.

Mechanical anisotropy: In many biological systems the elasticity modulus depends on the orientation of the applied stress. This property is best investigated in bones. Depending on the angle of measurement, Young's modulus in bones can vary by a factor of two. The reason for this is the particular structure of the bone. The struts of the cancellous bone, the so-called *trabeculae* are oriented according to its loading in vivo. They are oriented according to the trajectories of pressure and tension of the bone in the skeleton. This is the result of self-orientation and adaption, which was first investigated in 1892 by the German anatomist Julius Wolff formulating the law of bone remodeling (*Gesetz der Transformation der Knochen*).

The special mechanical properties of the cell membrane have been discussed in detail in Sect. 2.3.4 (Fig. 2.40). A lipid membrane can be considered as a two-dimensional crystal. The head groups of the phospholipids show similar behavior to the atoms in the three-dimensional crystal of steel. Therefore, they are showing a kind of steel elastic behavior in the plane. The elasticity modulus of these membranes is high and their limit of rupture low. The membrane proteins with their rubber elastic properties have no significant influence on this behavior. In contrast to technical materials like rubber sheets, the biological membrane can easily be deformed in an isoplanar fashion, but cannot resist expansion. This property is important for the dynamics of cell shape and for cell swelling. So for example swelling of human erythrocytes can easily occur only by deformation up to the volume of a perfect sphere. The swelling of lymphocytes or other cells is possible thanks to smoothing of the membrane.

Further Reading

Viscoelasticity in general: Hosford 2009, viscoelastic properties of biomaterials: Oomens 2009, Skalak and Chien 1987, the historical paper: Wolff 1986.

3.6.4 The Biomechanics of the Human Body

Investigations of biomechanical conditions of the human body, its movement, its carriage under conditions of loading, etc. are important tasks in orthopedics, surgery, sports, and occupational safety. What kinds of loads of the muscles and joints result from a particular loading and a particular movement? How can pathologic deviations be cured by surgical operations? How can diseased bones and joints be replaced by artificial materials? In this case immunological tolerance must be realized as well as the adaption of the applied material to the viscoelastic properties of the living bone the artificial joints will be connected to.

Recently, greater effort has been directed toward the construction of mathematical models of the locomotor system including human walking movements. With computer simulation programs, hopefully surgical corrections can be performed with optimal success. Of course, in this context sports efforts should also be mentioned which optimize various techniques of jumping, running, etc.

In contrast to the structure of plants which can be modeled statically by systems of flexionally and torsionally loaded bonded structures, for animals and human systems dynamic body models are required. They are composed of combinations of elements that are stable towards pressure and bending with tendons and muscles as elements of tensile stability and contraction. This is illustrated in Fig. 3.57. The head of the horse for example, is held by a system of elastin tendons, the

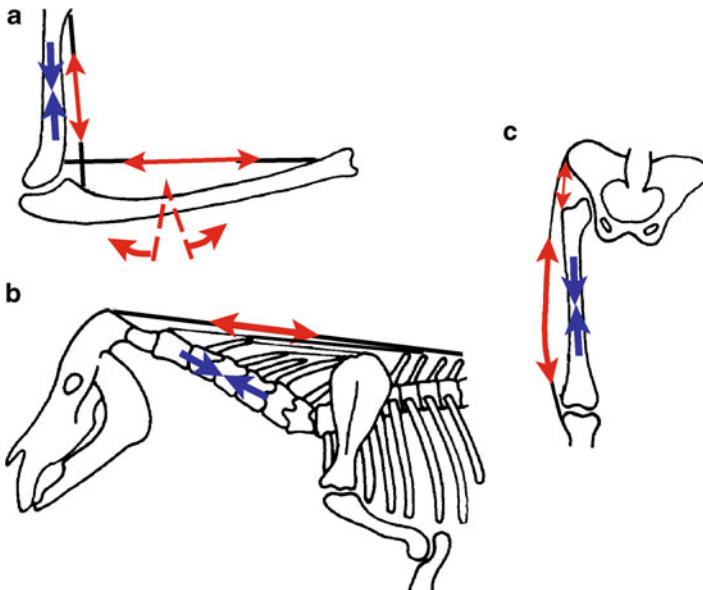


Fig. 3.57 Stabilization of the human forearm (a) and thigh (b), as well as the head of a horse (c) by the combination of bending and compressing stable bones, and tensioning stable muscles and tendons (After Glaser 1989)

ligamentum nuchae, as the tensile element and the cervical part of the spinal column as the component in compression. In the abdominal region of quadrupeds the compression element is located dorsally and the tension element is located ventrally. The carriage of the body is stabilized by tendons and it is maintained in an upright position without an additional supply of energy.

During evolution, the systems of combined elements which are stable against compression, together with those for tension, have been developed toward optimal use of muscle force, as well as toward maximal stability of the supporting bones. The attachment point of the muscles and tendons determines the vectors of tension and compression in the bone. We already mentioned in the previous section that this induces the oriented growth of the struts in the cancellous bone (Wolff's law).

Figure 3.58 shows an example of optimization of the shape of a human forearm. The bending of the bone and the shift of the muscle attachment have led to a significant reduction of the bending force. Similar principles of optimization can also be found in other parts of the muscle-skeleton apparatus.

The bending or the torque of a body can be attributed to stretching and compression of the material. As a measure of the bending, the *radius of bending* (R), or its reciprocal, the *curvature of bending* ($K = 1/R$) is used. If a homogeneous beam or bar bends, a compression of the concave side and a stretching of the convex side occurs. Between these two regions there must be a neutral plane that is neither compressed nor stretched.

Let us consider a section of a bar that is bent between two points that subtend an angle α at the center of curvature (Fig. 3.59). Let R be the radius of the curvature

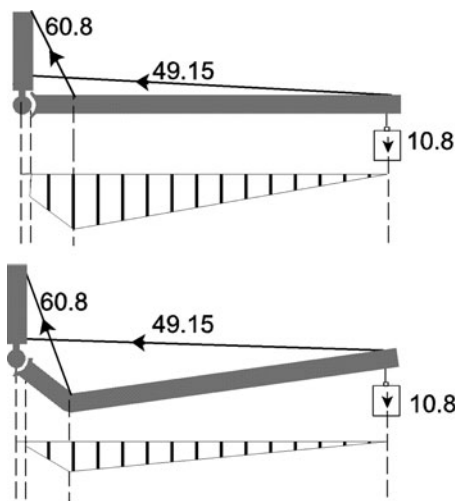


Fig. 3.58 Two steps for the optimization of the bending load of the forearm. The two muscles which held the forearm bone must generate forces of 60.8 N and 49.15 N, respectively, to compensate the loading force of 10.8 N at its end. These muscle forces are identical in both cases. The bending force of the bone, expressed by the graphs below, however, are quite different. Compare this scheme with the real situation in Fig. 3.57a (Redrawn after Pauwels 1980)

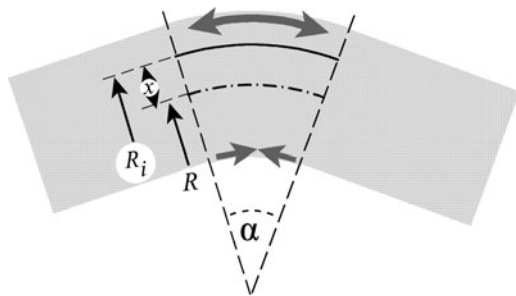


Fig. 3.59 The bending of a bar

measured to the neutral plane. If the angle is expressed in radians then the length $[l(x)]$ of the section at the distance (x) from the neutral plane is:

$$l(x) = \alpha R_i = \alpha(R + x) \quad (3.227)$$

If the length of the section along the neutral plane (at: $x_n = 0$; $R_n = R$) is l_n then the strain (ε) of any plane, parallel to the neutral plane can be calculated by Eq. 3.225:

$$\varepsilon(x) = \frac{\Delta l}{l} = \frac{l(x) - l_n}{l_n} = \frac{\alpha R_1 - \alpha R}{\alpha R} = \frac{x}{R} \quad (3.228)$$

As the distance x is measured from the neutral plane, ε can have both positive and negative values. Negative strain in this context means compression of the material and occurs when $x < 0$.

The combination of Hooke's law (Eq. 3.226) and the above relation enables the stress to be calculated:

$$\sigma(x) = Y \frac{x}{R} \quad (3.229)$$

The differential of moment of force (dM) is used to find the force which is necessary to bend the bar. It is calculated from the product of the force (\mathbf{F}) and the leverage distance from the neutral plane (x), and it is also related to the differential of an area (dA):

$$dM = x \quad d\mathbf{F} = x \sigma \quad dA = \frac{x^2 Y}{R} dA \quad (3.230)$$

The bending moment of the bar is obtained by integration of this equation. If the bar is homogeneous, the modulus of elasticity (Y) is not a function of x .

$$M = \frac{Y}{R} \int x^2 dA = \frac{Y}{R} I_A \quad (3.231)$$

In this equation the integral expression has been replaced by the *area moment of inertia* (I_A). This is a measure of the bending resistance of a bar made out of a material having a modulus of elasticity (Y). It can be seen that for a given bending moment, the smaller the value of I_A the more the bar bends. This means that if the bending moment does not change, and the modulus of elasticity remains constant, then the second moment of area is directly proportional to the radius of curvature.

It is easily understood from everyday experience that the area moment of inertia of a bar depends on the plane in which it is bent. A bar having a flat section bends more easily across the flat than it does on edge. This is determined by the position of the neutral plane. It passes through the center of gravity of a cross-section of the bar, and is always perpendicular to the radius of curvature.

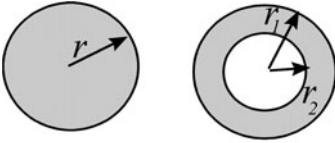
For bars with particular geometrical profiles analytical expressions for I_A are derived. Comparing these values for a compact cylindrical rod with a radius r , with that of a tube with an inner radius r_1 , and an outer radius r_2 , the advantage of a structure made up of hollow tubes, is evident (Fig. 3.60). This optimizing principle has occurred in the construction of bones and some plant stems.

In the case of geometrically nondefinable structures the area moment of inertia can be determined by iterative methods. First it is necessary to find the position of the neutral plane. This is perpendicular to the bending radius and located at the center of gravity of the profile. For structures of homogeneous density the center of gravity can easily be found using a cardboard model of the cross-section. The center of gravity is the crossover point of all lines which can be drawn in a vertical direction from various points of suspension (Fig. 3.61).

After the position of the neutral plane has been fixed, the cross-section can be subdivided into rectangular areas (Fig. 3.62). The following approximation can be obtained from the definition of the second moment of area:

$$I_A \approx \sum_{i=1}^n x_i^2 \Delta A_i \quad (3.232)$$

The individual rectangular areas (ΔA_i) are calculated, the distance of their centers from the neutral plane (x_i) are measured, and the products of these two



$$\begin{aligned}
 I_A &= \frac{\pi}{4} r^4 & I_A &= \frac{\pi}{4} (r_1^4 - r_2^4) \\
 I_P &= \frac{\pi}{2} r^4 & I_P &= \frac{\pi}{2} (r_1^4 - r_2^4)
 \end{aligned}$$

Fig. 3.60 Area moment of inertia (I_A) and polar moment of inertia (I_P) for a solid shaft and a tube

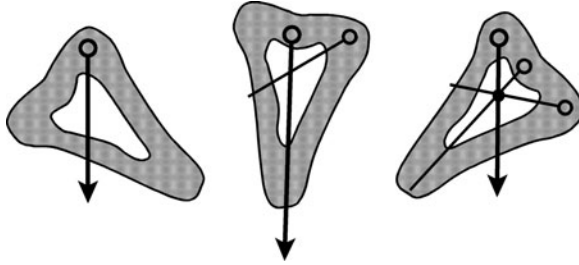


Fig. 3.61 Determination of the center of gravity of a tube of arbitrary cross-section as the crossover point of perpendicular lines drawn from the points of suspension

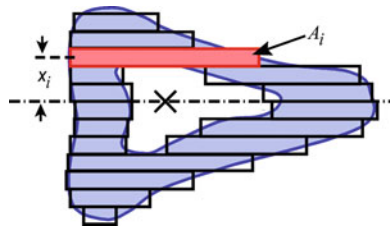


Fig. 3.62 Determination of the second moment of area of a tubular bone by a graphical method. The profile of the bone is approximated by many rectangles with individual areas A_i and distances from the neutral plane (---) x_i ; X – center of gravity

measurements are summed as required by Eq. 3.232. The accuracy of this method will be increased if the areas are made as small as possible. The units of the second moment of area are m^4 .

In a similar way to bending, the torsional deformation can be calculated. In this case not a neutral plane exists between the stretched and the compressed areas, but all regions are stressed as much as they are away from a central axis of gravity. A *polar moment of inertia* (I_P) can be formulated where the area elements (dA) are multiplied by the square of the radial distance (r) from the center of gravity, and subsequently, the products are summed. In an integral formulation this means:

$$I_P = \int r^2 dA \quad (3.233)$$

In this way the polar moment of inertia of bars and tubes of definite geometrical profile can be calculated (Fig. 3.60).

These statements form the basis for calculating the stability of structural elements of plants, animals, and humans. However, it can be seen from the above equations that some simplifying assumptions have to be made. In particular it must be noted that the modulus of elasticity (Y) may vary with the position and also, as

illustrated by the properties of bone, with the direction of the applied force. If the viscoelastic properties are also taken into account, the calculations become even more complicated.

Further Reading

Biostatistics of humans and animals: Fung 1993; Skalak and Chien 1987; biostatistics of plants: Niklas 1992; Bone modeling: Carter and Beaupré 2001; Martínez-Reina et al. 2009.

3.7 Biomechanics of Fluid Behavior

Streaming of viscous fluids occurs at all levels of biological structure. There is water flow through pores of membranes, streaming of cytoplasm in plants, streaming of blood in the vessels, and finally flow of water and air around animals, i.e. the problems of flying and swimming. In this section we will concentrate on some medically important problems of hemodynamics, or hemorheology. But as this is a classical problem of biomechanics, some basic aspects of flying and swimming will be included.

3.7.1 Laminar and Turbulent Flows

When a liquid flows along the surface of a thin plate, a flow profile is formed over this boundary which changes its character as the distance (l) from the leading edge of the plate (at: $l = 0$) increases. This effect is illustrated in Fig. 3.63.

Directly near the surface of the plate there is a trapped layer, i.e., a fixed, nonmoving film of air or liquid (not marked in Fig. 3.63). At increasing distance from the plate (z), the velocity of the flow increases. Near the edge of the plate, up to a particular critical length (l) there is a region of laminar flow, where fluid layers (lat.: *lamina*) slide over one another in a parallel direction. This boundary layer is characterized by a velocity gradient ($dv/dz \neq 0$) perpendicular to the surface of the plate (see Fig. 3.64). There is no sharp transition between this layer and the region of unaffected bulk flow ($dv/dz = 0$, $v = v_\infty$). Usually the thickness of this laminar boundary layer is defined by the distance $z = \delta_L$, when $v = 0.99v_\infty$. As can be seen in Figs. 3.63, and 3.64, δ_L gets larger as the distance from the leading edge of the plate increases.

The region of laminar flow in relation to the distance from the edge of the plate decreases at increasing bulk velocity (v_∞). With increasing distance (l) from the edge, the boundary layer gets thicker and becomes increasingly unstable. At a critical point, this leads to spontaneous appearance of turbulences. As already explained (Sect. 3.1.3, Fig. 3.3) turbulent flow occurs if the streaming processes becomes nonlinear, i.e., if the linear flux-force relation (Eq. 3.49) is no longer

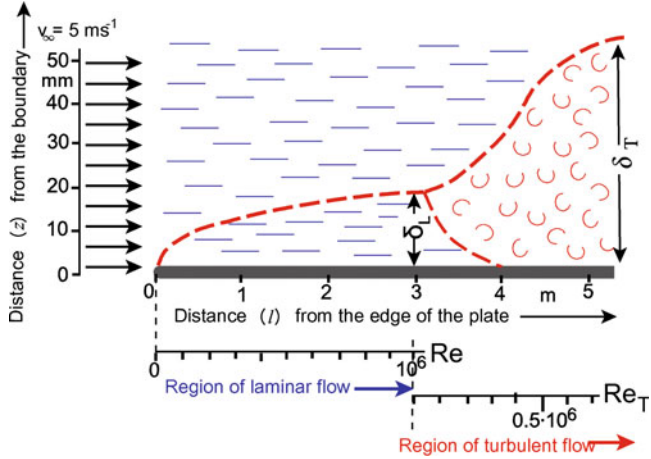
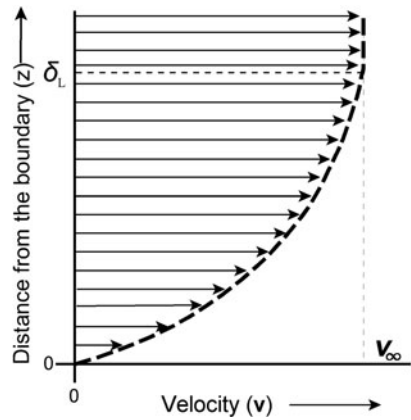


Fig. 3.63 Formation of laminar and turbulent flow profiles near a planar plate. As an example, the numbers correspond to a flow of air ($\nu = 1.5 \cdot 10^{-5} \text{ m}^2 \text{ s}^{-1}$) with a velocity $v_\infty = 5 \text{ ms}^{-1}$. The thickness of the laminar (δ_L) and the turbulent boundary layer (δ_T) are calculated according to the equations in Table 3.2

Fig. 3.64 Laminar velocity profile near a surface. The velocity vectors are depicted at various distances (z) from the surface; at $z = 0$, $v = 0$; δ_L is the thickness of the laminar boundary layer



applicable. The transition from laminar to turbulent flow therefore is accompanied by an increase in friction and a substantial increase in the thickness of the boundary layer ($\delta_L < \delta_T$). The resulting whirls can be considered as a kind of dissipative structure.

In Table 3.2 some basic equations are listed to calculate parameters of laminar and turbulent flow. Partly, these equations are based on empirical observations, particularly those for turbulent flow.

It is not possible to determine exactly the critical point at which the transition from laminar to turbulent flow takes place. In fact, this is a stochastic process of

Table 3.2 Equations for parameters of laminar and turbulent flow near boundaries. Symbols (see also Fig. 3.63): \mathbf{v} – velocity, z – distance from the boundary, ρ – density of the medium, η – viscosity, Re – Reynolds number; subscripts: 0 – at the boundary, ∞ – in the bulk phase

	Laminar flow (subscript L)	Turbulent flow (subscript T)
Sheer stress $\tau(z)$	$\tau_0(1-z/\delta_L)$	
τ_0	$0,332 \rho v_\infty^2 (\text{Re}_L)^{-1/2}$	$0,023 \rho v_\infty^2 (\text{Re}_T)^{-1/5}$
Velocity $\mathbf{v}(z)$	$2 v_\infty / \delta_L (z - z^2 / 2 \delta_L)$	$\mathbf{v}_\infty (z / \delta_T)^{1/7}$
Thickness δ	$5 l_L (\text{Re}_L)^{-1/2}$	$0,376 l_T (\text{Re}_T)^{-1/5}$
Force of surface friction $\mathbf{F}_0(l)$	$0,664 \rho v_\infty^{2l} (\text{Re}_L)^{-1/2}$	$0,0366 \rho v_\infty^{2l} (\text{Re}_T)^{-1/5}$

state transition of the system that occurs as the result of increasing destabilization. The position of the critical point where the laminar flow abruptly transforms into a turbulent one can only be calculated as a matter of probability. It depends on the flow velocity (\mathbf{v}), the viscosity (η), the density of the medium (ρ), and a characteristic streaming distance (l). These parameters are connected in the so-called *Reynolds number* (Re) which plays a crucial role in rheology:

$$\text{Re} = \frac{l \mathbf{v} \rho}{\eta} = \frac{l \mathbf{v}}{\nu} \quad (3.234)$$

In this equation the kinematic viscosity ($\nu = \eta/\rho$) is used which has already been introduced in Sect. 3.6.1.

The Reynolds number is a typical parameter of the theory of similarity. Bodies of identical shape show identical flow behavior, for flow conditions with the same Reynolds number. This is independent of whether the body is large or small, or whether it is moving in water or air.

The critical Reynolds number characterizing the transition from laminar to turbulent streaming of a flow parallel to a flat surface, as illustrated in Fig. 3.63, is about $\text{Re} = 10^6$. For a sphere with a flow around it, this transition already occurs at $\text{Re} \approx 10^3$ (see Fig. 3.3). The critical Reynolds numbers of streamlined bodies with so-called laminar profiles are somewhere between these limits, depending on their exact shape.

Flow inside a cylindrical tube can be characterized in a similar way. In this case in Eq. 3.234 the radius of the tube is taken for the characteristic length l . The critical value of the Reynolds number for a flow in a tube is about 10^3 . Turbulent flow means that the entire flow has become destabilized.

The following values for the kinematic viscosity can be used to calculate the Reynolds number for $T = 291 \text{ K}$:

$$\begin{aligned} \nu_{\text{Water}} &= 1.06 \cdot 10^{-6} \text{ m}^2 \text{ s}^{-1} \\ \nu_{\text{Air}} &= 14.9 \cdot 10^{-6} \text{ m}^2 \text{ s}^{-1} \end{aligned}$$

Table 3.3 Characteristic Reynolds numbers of various moving organisms

	Characteristic length (l) m	Characteristic velocity (v) (m.s ⁻¹)	Reynolds number (Re)
<i>Paramecium caudatum</i>	$2.1 \cdot 10^{-4}$	$1.1 \cdot 10^{-3}$	$1.8 \cdot 10^{-1}$
Mosquito (<i>Ceratopogonidea</i>)	$0.9 \cdot 10^{-2}$	$2.5 \cdot 10^{-1}$	$1.5 \cdot 10^2$
Chaffinch	$3.6 \cdot 10^{-2}$	$2.1 \cdot 10^1$	$5.4 \cdot 10^4$
Crane	$2.6 \cdot 10^{-1}$	$2.8 \cdot 10^1$	$5.0 \cdot 10^5$
Water bug (<i>Dytiscus</i>)	$3.0 \cdot 10^{-2}$	$3.0 \cdot 10^{-1}$	$8.4 \cdot 10^3$
Stickleback (marine)	$1.0 \cdot 10^{-1}$	$7.2 \cdot 10^{-1}$	$5.5 \cdot 10^4$
Shark	$1.5 \cdot 10^0$	$5.2 \cdot 10^0$	$6.1 \cdot 10^6$
Dolphin (<i>Stenella spec.</i>)	$2.1 \cdot 10^0$	$9.3 \cdot 10^0$	$1.5 \cdot 10^7$
Blue whale	$3.3 \cdot 10^1$	$1.0 \cdot 10^1$	$2.6 \cdot 10^8$

If these values are substituted into Eq. 3.234 it is seen that, in contrast to streaming air, as shown in Fig. 3.63, turbulence in the case of water would occur not at $l = 3$ m, but already at $l = 0.2$ m from the leading edge. A comparison of flying and swimming objects can only be made under conditions of equal Reynolds numbers. It is quite senseless simply to relate velocities to the length of an object (which unhappily is often done in popular scientific publications, much to the amazement of the readers!).

Table 3.3 shows some typical Reynolds numbers for the movement of quite different organisms. In Sect. 3.7.3 we will show that, although they may possibly have optimal laminar flow shapes, large, fast-swimming fishes and aquatic animals exceed the critical Reynolds number for laminar flow.

3.7.2 Biomechanics of Blood Circulation

The biophysical properties of blood flow are quite complicated, even if all the complex physiological and biochemical control mechanisms are at first neglected. There are at least two features in which blood flow in general differs from the movement of a normal fluid through a tube, even having non-Newtonian properties. On the one hand, the elasticity of the blood vessels must be considered, changing their diameter as a function of the internal pressure, and on the other, it must be taken into account that blood is a suspension of flexible cells, the mean diameter of which may be of the same order as the smallest capillaries through which they must flow. The viscosity of the blood is a function of the cell concentration, the so-called *hematocrit*, which itself appears to be a function of the shear conditions in the vessel (see also Sect. 3.6.2). Furthermore, the pulsed character of blood flow must be considered and the particular aspects of bifurcations of the vessels.

Let us first consider some basic equations of fluid mechanics. Laminar flow through a tube may be thought of as the mutual movement of concentric hollow

cylinders. In this case each of these cylinders with a radius r , and a thickness dr experience a frictional force (\mathbf{F}_F) which is proportional to the velocity gradient (dv/dr), to their surface area ($2\pi r l$), and to the viscosity of the fluid (η).

$$\mathbf{F}_F = 2 \pi r l \eta \frac{dv}{dr} \quad (3.235)$$

The driving force (\mathbf{F}_D) behind such a flow can be obtained from the pressure difference (Δp) and the cross-sectional area of the cylinder (πr^2):

$$\mathbf{F}_D = \pi r^2 \Delta p \quad (3.236)$$

In the case of stationary movement, both forces are balanced: $\mathbf{F}_F = \mathbf{F}_D$. Let the radius of the tube be r' , and let us assume that there is a trapped boundary layer of liquid [$v(r') = 0$] at the inner surface of the tube. Connecting Eqs. 3.235 and 3.236, and integrating them according to dv one obtains an equation for the velocity profile of this flow:

$$v(r) = \frac{\Delta p}{4 l \eta} (r'^2 - r^2) \quad (3.237)$$

Thus, $v(r)$ is a parabolic function, whereas v_{\max} is the velocity at the center of the tube (see Figs. 3.51, and 3.66). In order to calculate the total volume flux (\mathbf{J}_V in: $\text{m}^3 \text{s}^{-1}$) through the tube, the function $v(r)$ must be integrated over the entire profile. This leads finally to the *Hagen–Poiseuille equation*:

$$\mathbf{J}_v = \frac{\pi \Delta p r'^4}{8 l \eta} \quad (3.238)$$

Thus, the flow through a tube is proportional to the fourth power of its radius. This is a very important aspect for the physiological control of local circulation. A slight widening or narrowing of the blood vessels causes large changes in the blood flow.

These fundamental laws of physical rheology must be considered just as a first approximation for what really happens in blood flow through the vessels. The following particularities must be taken into account, leading to an extension of these approaches:

- *Blood is a non-Newtonian fluid*, i.e., its viscosity (η) depends on the shear rate ($\dot{\gamma}$) of the flow (see Sect. 3.6.2, Fig. 3.54). Integrating Eq. 3.235 in order to derive Eq. 3.237, we considered the viscosity as constant. This, obviously, is not acceptable in the case of blood. If, however, a particular function $\eta(\dot{\gamma})$ was applied, the integration would be more complicated, and the resulting velocity profile would not show a simple parabolic function.

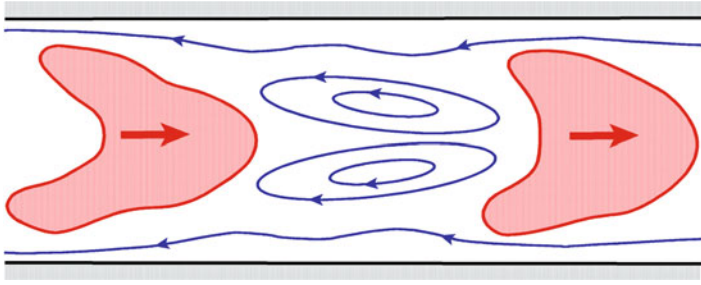


Fig. 3.65 Movement of deformed erythrocytes through a narrow capillary. The plasma, trapped in vortices between the cells is transported in the same direction together with them. Near the wall of the capillary plasma may move in the opposite direction (Modified after Talbot and Berger 1974)

- *Blood is not a homogeneous liquid, but a suspension of cells.* In capillaries, the diameters of which are of the same order of magnitude or even lower than the diameter of erythrocytes, the velocity profile of the plasma is determined by the moving cells which become strongly deformed in these narrow and branched vessels. This is a problem of *microrheology* of circulation (see Fig. 3.65). This particular streaming profile, especially the microscopic vortices in fact optimize gas exchange between tissue and erythrocytes.

In large vessels the so-called *Fahraeus–Lindqvist effect* occurs. This leads the erythrocytes to concentrate in regions of minimal shear stress, namely in the center of the vessel. This means that the viscosity of the blood which we found to depend on the hematocrit (Sect. 3.6.1) increases in this region, but decreases near the wall of the vessel. This leads to a lowering of the streaming resistance of the total blood flow. Conversely of course, the streaming profile is changed dramatically. The parabola becomes flattened at the center of the vessel and steeper near the walls. Furthermore, this effect leads to a redistribution of different sorts of blood cells. In fact, the intensity of the force, shifting the cells by the Fahraeus–Lindqvist effect into regions of lower shear stress, depends on their size. As a result smaller cells like blood platelets are not influenced as much by this effect as erythrocytes with a larger diameter. This leads to concentration of the platelets near the walls of the vessel, which appears to be helpful in the case of injury.

The Fahraeus–Lindqvist effect can be understood as the result of the Prigogine principle of minimal entropy production, as described in Sect. 3.1.4. It is valid for linear approaches, thus also for laminar flow (see Sect. 3.6.1). Minimal entropy production, for the case of blood flow means that the cells should concentrate at locations of minimal frictional energy dissipation, namely at locations of minimal shear rate.

- *The diameter of the blood vessels differs along the system of circulation.* If a tube suddenly becomes narrow a so-called *entrance effect* occurs (Fig. 3.66). This means that first the velocity profile of the narrow part of the tube corresponds to that of the central part of the broad tube. Only after a certain

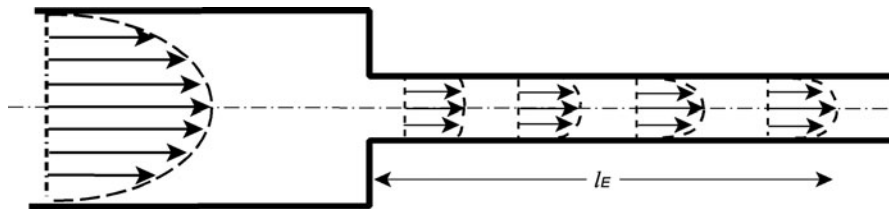


Fig. 3.66 Laminar flow in a tube: the parabolic velocity profile changes during a sudden narrowing of the tube's radius (entrance effect). Only at a distance l_E is a new parabolic profile established again

distance from the place of narrowing (l_E in Fig. 3.60), will a new profile be established. Usually this occurs at: $l_E = 0.06 \cdot r \cdot \text{Re}$, where r is the radius of the narrow tube, and Re is the Reynolds number. This effect becomes important at the entrance of blood in the aorta. Furthermore, it occurs in the case of air flow in the lungs.

- *Blood flow is not stationary, but pulsed.* Therefore, the condition $\mathbf{F}_F = \mathbf{F}_D$ is no longer valid (see Eqs. 3.235 and 3.236). This fact is only important for arterial flow. The pulse waves of the heart partly become damped by the elasticity of the walls of the vessels, but they nevertheless proceed as oscillations of pressure and velocity of the blood flow up to the arterioles. One must differentiate between the true velocity of the blood streaming (\mathbf{v}) on one hand, and the velocity of pulse propagation (\mathbf{v}_p) on the other. The pulse propagation can be calculated using the *Moens–Korteweg equation*, which is derived for cylindrical tubes with thin walls:

$$\mathbf{v}_p = \sqrt{\frac{Y d}{2r \rho}} \quad (3.239)$$

In contrast to the flow rate (\mathbf{v}) (Eq. 3.237), the pulse propagation rate (\mathbf{v}_p) depends on the density (ρ) of the medium, but not on the viscosity (η). Conversely, the thickness (d), and the elasticity modulus (Y) of the wall of the vessels are included. This equation does not say anything about the damping of pulse waves.

To relate various pulsing flows, similar to the Reynolds number, another parameter of similarity (α) is introduced:

$$\alpha = r \sqrt{\frac{\omega \rho}{\eta}} \quad (3.240)$$

In this equation beside the density (ρ) and the viscosity (η), additionally the pulse frequency (ω) is used. This parameter α allows us to evaluate the relation between \mathbf{v}_p and \mathbf{v} under various streaming conditions. For small values of α the pulse propagation is faster than the velocity of the blood stream. At $\alpha \geq 3$, the pulse

propagation velocity becomes equal to the streaming velocity ($\mathbf{v}_p = \mathbf{v}$). In the aorta this parameter is higher than that limit ($\alpha = 15$). In the arteria femoralis this limit is reached ($\alpha = 3$). In these vessels therefore the pulses are propagated with the same velocity as the total flow rate of the blood. This consideration, however, does not include the fact that at points of branching of the vessels, reflections of the waves can occur.

To evaluate the elasticity of the blood vessels, let us start again with simple physical approaches. How is the radius of a tube changed by the strain of the tube wall? For the simplest case, Hooke's law (Eq. 3.226) can be applied, considering an elastic tube with a radius r , and a circumference of $2\pi r$. In this case the stress (σ) of the wall of the tube can be written as follows:

$$\sigma = Y \frac{\Delta r}{r} \quad (3.241)$$

We defined the stress (σ) as stretching force (\mathbf{F}) per area (A) of the stretched material (Eq. 3.224). Let the thickness of the wall be d , and its length l , the cross-sectional area of the stretched wall becomes $A = d \cdot l$. Therefore:

$$\sigma = \frac{\mathbf{F}}{ld} \quad (3.242)$$

Defining σ' as stretching force per length, gives:

$$\sigma' = \frac{\mathbf{F}}{l} = \sigma d \quad (3.243)$$

Now, we must know how large the stress (σ) would be in the wall of a tube with an internal pressure p . This can be calculated using the *Laplace equation* for a tube with radius r :

$$\sigma' = p r \quad (3.244)$$

Equations 3.241–3.244 enable us to derive a formula to calculate the increase of the radius of a tube (Δr) as a function of the internal pressure (p):

$$\Delta r = \frac{p r^2}{Y d} \quad (3.245)$$

Let us remember that we started with the assumption that the vessel wall shows a linear elastic behavior of the material according to Hooke's law. The relation between stress and strain therefore would correspond to Young's modulus (Eqs. 3.226 and 3.241). In fact the viscoelastic behavior of the wall of vessels is much more complicated and not at all linear. Furthermore, its strain is not only controlled by an interplay between various elastic materials, but is additionally

Table 3.4 Some rheological parameters of human blood circulation (Data from Talbot and Berger 1974)

Vessel	Average velocity $\text{m (s}^{-1}\text{)}$	Diameter (m)	Average wall shear rate (s^{-1})	Reynolds number (Re)
Aorta	$4.8 \cdot 10^{-1}$	$2.5 \cdot 10^{-2}$	155	$3.4 \cdot 10^3$
Artery	$4.5 \cdot 10^{-1}$	$4 \cdot 10^{-3}$	900	$5 \cdot 10^2$
Arteriole	$5 \cdot 10^{-2}$	$5 \cdot 10^{-5}$	8,000	$7 \cdot 10^{-1}$
Capillary	$1 \cdot 10^{-3}$	$8 \cdot 10^{-6}$	1,000	$2 \cdot 10^{-3}$
Venule	$2 \cdot 10^{-3}$	$2 \cdot 10^{-5}$	800	$1 \cdot 10^{-2}$
Vein	$1 \cdot 10^{-1}$	$5 \cdot 10^{-3}$	160	$1.4 \cdot 10^3$
Vena cava	$3.8 \cdot 10^{-1}$	$3 \cdot 10^{-2}$	100	$3.3 \cdot 10^3$

regulated actively by smooth muscles. This interaction of passive and active properties of the vessel wall is of great importance for the regulation of blood flow.

All these calculations consider the blood flow in the system of circulation as laminar. Is this correct or are there turbulences in human blood flow? The numbers of Table 3.4 indicate that in fact, at least in large arteries critical Reynolds numbers can occur. We mentioned in Sect. 3.7.1 that the laminar flow in tubes becomes unstable in the case of Reynolds numbers near 1,000. This limit, however, is only correct for smooth, rigid and absolutely cylindrical tubes. In blood vessels, some factors help to stabilize the laminarity even beyond this limit, while others induce turbulent flow already at lower Reynolds numbers. Factors inducing turbulences are for example the branching of the vessels and inhomogeneities of their walls. These include arteriosclerotic alterations, or consequences of various surgical operations. In general, however, the system of blood circulation can be considered as biomechanically optimized.

Various techniques have been applied to investigate the properties of streaming blood. In medical diagnosis various ultrasound techniques are used to analyze flow properties of blood as well as viscoelastic properties of the vessels. Especially, the Doppler effect of ultrasound allows us to investigate blood flow, and even the flow profiles in large vessels. Special computer programs help to visualize these processes. This also allows us to check the blood flow and functions of the heart valves. Furthermore, investigations are undertaken analytically in tubes made of transparent materials which copy particular regions of human vessels (Fig. 3.67).

The model system of Fig. 3.67 shows special properties of flow behavior near the branching of vessels. The critical Reynolds numbers in these regions are lower than in unbranched regions. This depends on the angle of bifurcation (α in Fig. 3.67). For $\alpha = 180^\circ$ the laminar flow already becomes critical at $\text{Re} = 350$. If the angle is only 165° , the critical point is at $\text{Re} = 1,500$. Additionally, the critical Reynolds number depends on the relation of the radius of these branches.

These few aspects of biophysics of blood rheology already show us how complicated this situation in vivo in fact is. This branch of biomechanics is developing quickly. This tendency is promoted by the fast progress in surgery of blood vessels on the one hand, and on the construction of artificial cardiac valves, of artificial hearts as well as of various systems of extra-corporal circulation on the other.

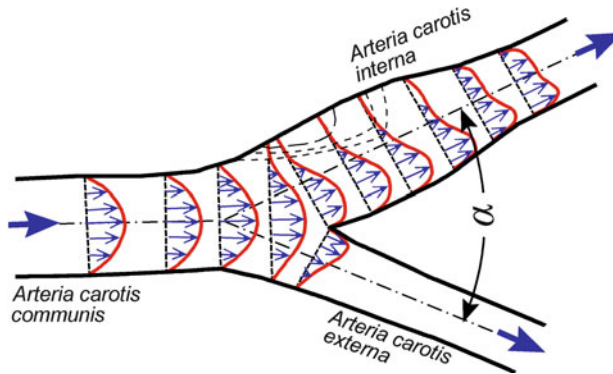


Fig. 3.67 Velocity profile in a model of the human carotis with 70% flow through the Arteria carotis interna. α – angle of bifurcation (After Schneck 1980 modified)

Further Reading

Bejan and Lorente 2008; Fung 1984, 1993, Skalak 1987, Waite and Fine 2007.

3.7.3 Swimming and Flying

The Reynolds number enables us to relate flow properties of moving objects in water with those of air. This allows us to relate the mechanisms of swimming to that of flying.

When a body moves relative to its surrounding medium then a *surface friction*, or *skin friction* occurs, and furthermore a drag on the surface of the body which is caused by its shape, the so-called *profile*, or *form drag*. The surface friction arises from phenomena which occur in the boundary layer and which have already been discussed in Sect. 3.7.1. Conversely, form drag arises because of the form (shape) of the object and is related to the volume of the surrounding medium that is displaced by the moving body. Objects with a larger cross-section perpendicular to their direction of movement will have a higher drag than thinner bodies.

Regions around the body therefore can become influenced to a much larger extent than those at the boundary layer in a fluid flowing parallel to a flat plate (Fig. 3.63). The critical Reynolds number, indicating that the laminar boundary layer becomes destabilized, therefore, is determined, to a great extent by the shape of the body. This explains why these two components of drag generation, namely the surface friction on the one hand, and the form drag on the other, cannot be treated separately from one another.

The boundary layer around a moving body is to a great extent influenced by local pressure differences. This arises from differences of the velocity (\mathbf{v}) at different locations. Because of the law of conservation of energy, the sum of kinetic energy

of the moving medium ($\frac{1}{2}\rho\mathbf{v}^2V$), and the static energy of compression (pV) must be constant at all points of the space:

$$p + \frac{1}{2}\rho \mathbf{v}^2 = \text{const} \quad (3.246)$$

This constant is determined by the following conditions: at $\mathbf{v} = 0$ the hydrostatic pressure must be: $p = p_0$, whereas p_0 is the static pressure of the environment. This leads to the *Bernoulli equation*:

$$p = p_0 - \frac{1}{2}\rho \mathbf{v}^2 \quad (3.247)$$

The local pressure (p) acting on the surface of a body which moves in relation to a fluid, therefore results from the hydrostatic pressure (p_0) of the environment, lowered by the parameter $\frac{1}{2}\rho\mathbf{v}^2$. In addition to this pressure, however, an impact pressure must be considered affecting the points where the velocity vector of the streaming fluid is perpendicular to the body surface.

Figure 3.68 indicates the flow profile and the pressure distribution around a streamlined body, i.e., a body with minimal form drag. At points where the velocity of the flow is greatest, which in the diagram is indicated by maximal density of the lines of flow, i.e., near the middle of the body, there will be a maximum of negative pressure. At both ends the form drags dominate. These local pressures lead to forces which are directed always perpendicular to the surface of the body. At locations where this pressure becomes negative, i.e., at points where the forces are directed away from the surface, a disruption of the flow from the surface of the body can occur. This results in a wake of large eddies that is a region of turbulent flow which occupies a space much larger than that corresponding to the thickness of the turbulent boundary layers given in Fig. 3.63 and Table 3.2.

Considering this situation, there are three discrete qualities of flow pattern, obtained by increasing velocities or better, increasing Reynolds numbers:

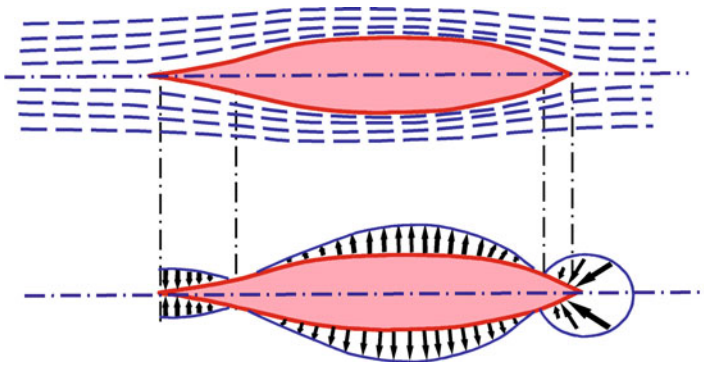


Fig. 3.68 Flow profile and pressure distribution around a moving streamlined body (Redrawn after Hertel 1963)

- Laminar flow around the body,
- Turbulent flow in a boundary layer,
- Turbulent, disrupted flow forming a wake.

The resistance to streaming around the body in these situations always increases stepwise.

In Table 3.3 Reynolds numbers are indicated for some swimming animals. The small animals, in spite of having unfavorable shapes, are in the region where the flow is always laminar, but for animals where the Reynolds number is in the order of 100 and higher, a hydrodynamic optimization of the shape of the body is required. In contrast to the sphere which allows laminar flow up to $Re \approx 1,000$, the critical Reynolds number for a streamlined body, of course, is higher. The body shape of fast swimming animals really does appear to be optimal in this respect. Nevertheless, the Reynolds numbers of fast swimming fishes and dolphins, and also of many birds, lie in the supercritical range. In such cases various adaptations can be found that, at least, impede the disruption of the turbulent flow. This, for example, is achieved by the early induction of microturbulences at the surface of the body by particular surface structures such as feathers or scales.

Much has been written in papers on bionics about the specific adaptations of the dolphin which enable it to swim extremely fast. Apparently there is a viscoelastic damping layer of the skin and furthermore, the ability to induce folds at the body surface by muscular activity, both of which prevent the occurrence of latent instabilities in the flow.

The real friction of a swimming fish is difficult to measure. The simplest way would be to pull dead or anesthetized fish at a given velocity through water. The frictional resistance, obtained in this way, however, is so high that it would be impossible for the musculature of an actively swimming animal to overcome them. This leads to the conclusion: the fish is unable to swim! This circumstance was postulated in 1936 by Sir James Gray, and calculated particularly for the case of dolphins. The solution of *Gray's paradox* is as follows: In contrast to technical constructions, for example vessels containing two distinct elements, the driving component (screw), and the frictional component (body), in the case of living fishes or even dolphins, both elements are combined. A fish diminishes its friction during active swimming. The actively swimming animal has a lower drag than one that is towed passively through the water.

The part of the body that is involved in propulsion of fishes can be very different. Depending on the relative length and flexibility of the tail, three main types can be differentiated. The *anguilliform type* of movement, for example the eel, involves the whole body for propulsion. Most fishes show the *canrangiform type* of propulsion (Fig. 3.69). In this case tapering tails of medium length allow fast and dexterous swimming. They are able to accelerate quickly reaching a high speed after a short time. In the case of *ostraciiform type* of movement, named after the trunk of coffer fish, only the fins, like propellers propagate the fish. These fishes are only able to swim slowly.

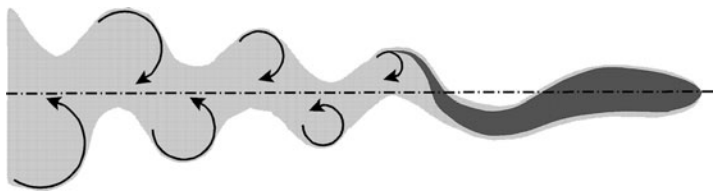


Fig. 3.69 Carangiform mechanism of propulsion of a fish with the vortex street behind

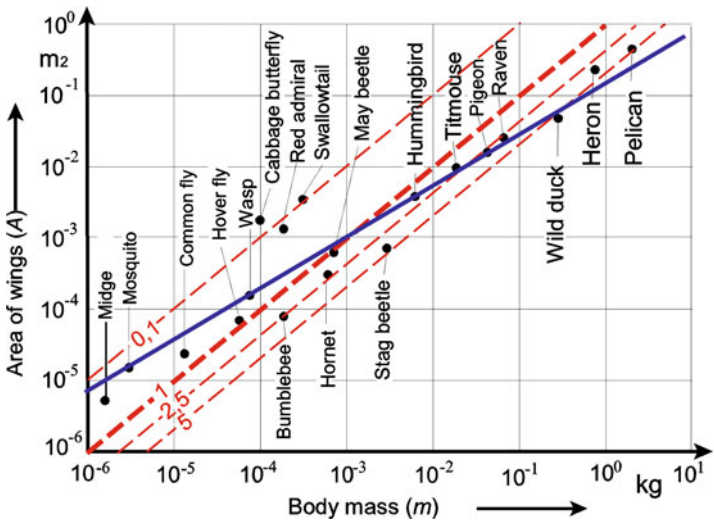
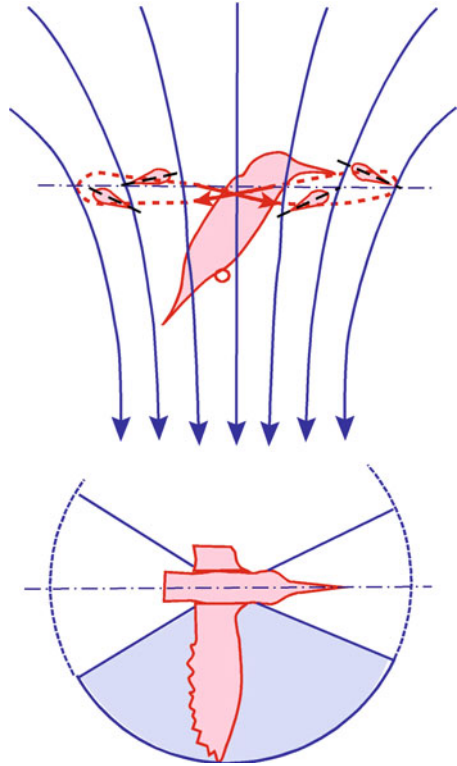


Fig. 3.70 The area of wings (A , in m^2) of various insects and birds as a function of body mass (m , in kg), follows approximately an allometric equation: $A = 0.11 m^{2/3}$ (blue line). The red lines demonstrate the resulting wing loading (in $kg \cdot m^{-2}$) (Data from Hertel 1963)

In contrast to swimming, where we considered mechanisms of drag reduction, in the biomechanics of flying, lift is of central interest. First, we must distinguish between passive gliding and active flying. There are not only various insects, especially butterflies, birds, and mammals that can glide, but also a number of plant seeds and fruits. Many types of aerodynamic mechanisms are employed by these vegetable objects ranging from simple adaptations to reduce the rate of descent by means of rotating wings, for example the propeller seed of the maple tree, up to the stable gliding flying-wing of the seeds of the climbing pumpkin species *Zanonia macrocarpa*.

Active flying by animals is achieved through a most varied assortment of mechanisms. Those animals which can rise into the air without some initial help and then move forwards or backwards from this position can be considered as ideal fliers. This hover flight, however, requires power which dramatically increases with the wing loading, i.e., with the relation between body weight and wing area. Figure 3.70 shows that in fact the area of wings in various flying animals does

Fig. 3.71 Wing position and air flow of a hovering hummingbird (Redrawn after Hertel 1963)



not increase in direct proportion to their body mass, but rather to the power of $\frac{2}{3}$. This is a common allometric relation which we will discuss in the next section (Sect. 3.8). The isometric lines (in Fig. 3.70: red lines, all with a power of 1) allow us to evaluate the real wing loading. The upper limit of the specific muscular performance is already reached in the case of humming birds. Larger birds and even large insects, like the stag beetle (see Fig. 3.70) are not able to hover on their own although some of them, like the kestrel can do so with the help of the wind.

Figure 3.71 shows schematically the wing movement of a hovering hummingbird. The wings move horizontally, and each covers about a third of the horizontal area around the bird. The angle of attack is continuously adjusted so that lift is generated as the wing moves forwards, as well as backwards. In contrast to humming birds, the wings of flies and bees move in a vertical direction. The wings are also twisted but only the downstroke is used to generate lift. In the case of birds flying forward, the wings are moved up and down. The angle of attack is also regulated, but additionally the wings are bent or partly folded on the upstroke.

In contrast to the hummingbird, all other birds require some assistance to get off the ground. Just as an aircraft must reach a take-off speed to ensure that the wings are generating sufficient lift, so must birds. This is achieved by running, jumping, or dropping. Nevertheless, the take-off speed of birds is remarkable low. This means

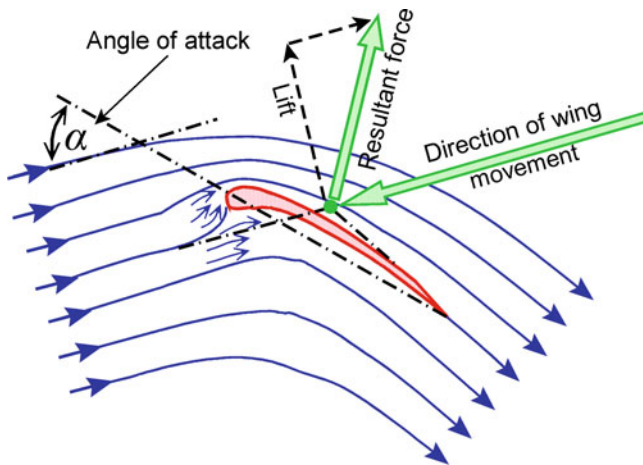


Fig. 3.72 Air stream and forces of bird flight during a stroke of the wing forwards and downwards (Redrawn after Hertel 1963)

that maximum lift at minimal Reynolds number is required. Figure 3.72 shows the force diagram for a wing that is being thrust forwards and downwards. If the incident flow remains constant (v_0) then the lift generated by the wing increases with the degree of curvature of the wing section as well as with the angle of attack (α). The amount of increase of both these parameters however, is limited. If they exceed a critical value, the flow is disrupted, and turbulence develops. This will not only cause a loss of thrust but will at the same time destabilize the entire system. Such an occurrence would cause an airplane to crash. There are a number of biological adaptations preventing such a disaster and at the same time ensuring maximum lift at low speeds.

Further Reading

Ahlborn 2005; Alexander 2003; Azuma 1992; Leyton 1975; Videler 1993; Vogel 1994; Webb and Weihs 1983.

3.8 Allometric Considerations of Structure and Function

We already touched upon the problem of scaling of various processes in some earlier sections. So, the Reynolds number (Sects. 3.7.1, 3.7.3), and the α -parameter (Sect. 3.7.2) enable the mutual compatibility of streaming behavior of structures with various sizes. In fact, there exists a long list of dimensionless numbers of this kind which are used in technical engineering, allowing us to scale streaming, convection, heat conduction, and many other processes. But it is not only in engineering that it is necessary to compare structures and functions of systems

with different size, in fact, it is attracting more attention also in regard to various biological systems.

From the smallest organism to the largest, the size ranges through many orders of magnitude. Additionally, there are considerable size differences in the individual development of a single animal, as well as in different animals of the same species. To relate various biological functions with size, i.e., mass, is not only a general question of theoretical understanding of biological organization, but in some cases it is quite important for practical problems in medicine, sports, agriculture, etc.

For many biological variables a relationship can be written in the form of a so-called *allometric function*:

$$y = \alpha x^\beta \quad (3.248)$$

where α is the *allometric coefficient* and β the *allometric exponent*. The word *allometric* comes from the Greek and means “different measure.” As a matter of fact, the relation is said to be allometric only if $\beta \neq 1$. Otherwise the relationship is linear and it is called not allometric, but *isometric*.

In general, we find three levels of scaling in biology: scaling relations within an individual organism during its growth (*ontogenetic allometry*), scaling relations among individuals of one species (*intraspecific allometry*), and finally scaling relations for individuals of different taxonomic or functional groups (*interspecific*, or *phylogenetic allometry*).

This problem in general involves the question of similarity, which was formulated already in Euclidean geometry. Archimedes, more than 2,000 years ago formulated that the surface of bodies with similar shapes grows in proportion to the square of their size, whereas their volume, correspondingly, with the cube. Galileo Galilei in the seventeenth century already speculated about the similarities of animals. He noted that a dog may be able to carry two or even three similar dogs on his back in contrast to a horse which hardly can do the same. He furthermore emphasized that the thickness of the bones in animals of different sizes are not simply proportional to the linear dimensions of their body.

Physiologists have for a long time formulated allometric relationships between body mass and various structural and functional properties of animals of various size. In the center of interest is metabolic rate as a function of size. This is an important parameter because it limits almost all biological processes at different levels of organization. In aerobic organisms, it is equivalent to oxygen consumption, and can be determined in this way. From a thermodynamic point of view, neglecting the storage of chemical energy, it must finally result in heat production, and in fact, it represents the entropy production which is expressed by the dissipation function Φ (see Sect. 3.1.4, Eq. 3.64).

In 1883 the German physiologist Max Rubner formulated a *surface law of metabolism*. He reasoned: if an animal is n times as big as another, then its surface (A) should increase by n^2 and its volume or mass (m) by n^3 , provided that the density (ρ) of all organisms is more or less the same. The relations $A \sim n^2$ and

$m \sim n^3$, can be transformed into $A \sim m^{2/3}$. Rubner considered that the metabolism of an animal will finally produce heat (dQ_M/dt). He supported these considerations with experimental data, derived from measurements of heat production and oxygen consumption in animals of the gamut of sizes. To keep the temperature of the organism constant, there must be a steady state, i.e., the heat must continuously be dissipated from the skin's surface. This leads to $\Phi \sim A$, or $\Phi \sim m^{2/3}$.

To determine these allometric parameters usually a log-log plot of the data is used. Taking the logarithms of both sides the above-mentioned generalized allometric equation will be transformed into:

$$\log y = \log \alpha + \beta \log x \quad (3.249)$$

Plotting ($\log y$) against ($\log x$) one obtains a straight line with the slope β . The parameters α and β therefore are the results of a linear fit to the log-transformed data.

In contrast to Rubner's data, another physiologist Max Kleiber in 1932 stressed that β was larger than Rubner postulated, namely $\beta = 3/4$. In this way the discussion in animal physiology about this allometric function ($\Phi = \alpha m^\beta$) began and proceeds even today. The following questions arise: what is the real value of the allometric coefficient α and the allometric exponent β ? Are these parameters really constant for mammals of different size, from the mouse to the elephant, or even in other animals like reptilians, or birds? Considering the large amount of papers published in recent years, and the hundreds of measurements on various animals, some researchers maintain Kleiber's proposition, while others support the $\beta = 2/3$ parameter of Max Rubner.

What may be the reason for this controversy? This should be clarified first, before further theoretical explanations are helpful. The parameter Φ , representing the metabolic rate of an organism, of course varies depending on its physiological condition. A so-called *basal metabolic rate* (BMR) is formulated, representing the minimum power which is required to maintain the tissues and essential life functions in nontorpid animals. Alternatively, the metabolic rate arrives at a maximum (MMR) when the animal is running at top speed. This can exceed the BMR by a factor of ten, and the allometric exponents for MMR data have been found to increase up to nearly 0.9. In fact, the measured metabolic rate is in any case between these two extremes and therefore variable. Furthermore, the metabolic rate strongly depends on the temperature of the environment and the corresponding adaption. Therefore, the allometric exponent may also depend on geographical regions and habitats.

In general, all biological variables have a measurement error and therefore deviate from expected curves. Moreover, it is difficult to standardize the measuring conditions for different animals. In general, the value of the exponent depends on the conditions under which the data are recorded. Finally, the parameters of the allometric functions to some extent depend also on the applied fitting procedure.

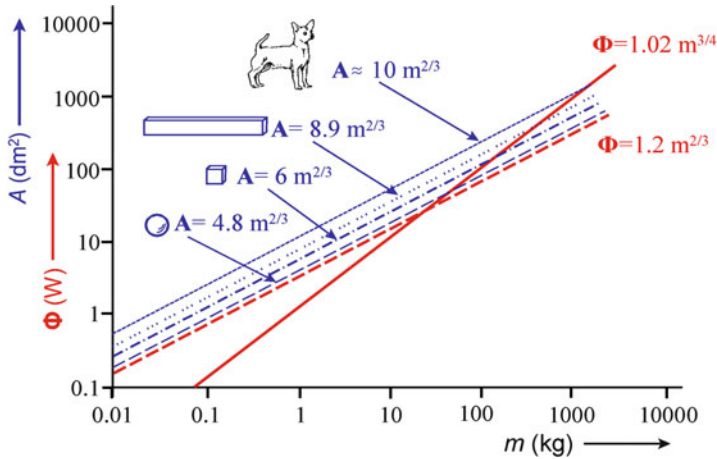


Fig. 3.73 Allometric functions of metabolism (red) according to Rubner's surface rule (broken red line: $\Phi = 1.2 \text{ m}^{2/3}$), and Kleiber's rule (full red line: $\Phi = 1.02 \text{ m}^{3/4}$), as well as simple surface mass relations (suppose: $\rho = 10^3 \text{ kg m}^{-3}$) of a sphere, a cube, a 1:10 prolonged cube, and that for a dog as a representative quadrupedal animal (blue lines)

Furthermore, the problem of geometrical and of course functional similarity of the animals should be considered. How can ectothermic animals relate to endothermic, herbivores to carnivores, birds to quadrupeds, etc. In fact, differences of the allometric parameters have been found in these special cases.

What are the real surface–volume relations of various bodies, and how do they influence the allometric function (Fig. 3.73)? Simple geometrical considerations show that the surface of a sphere grows with a function $4.84 V^{2/3}$, that of a cube by $6 V^{2/3}$. If one dimension of the cube is prolonged 10 times, the result would be $8.89 V^{2/3}$. For mammals, factors between 9 and 11 have been estimated. This means that even for not fully geometrically similar animals, the surface-to-volume relation, or if the density (ρ) is constant, even the surface–mass relation may only differ in terms of the allometric coefficient, but not the allometric exponent.

Recently, a number of theories have been proposed to explain the allometric exponent of metabolism and its variation according to size and physiological condition. Instead of the body surface as considered by Rubner, mostly the transportation network of blood circulation was taken into account. So, for example a quarter-power law was derived, based mostly on geometry, particularly the hierarchical nature of circulatory networks. It seems however, that none of these models is complete, because the metabolic rate must account for all irreversible losses.

Allometric considerations are in fact not only centered on problems of metabolism. In Fig. 3.70 the area of wings of various insects and birds is plotted as a function of body mass. This also shows an allometric exponent of around $2/3$. In fact, similar functions are considered in relation to a large number of structural and functional parameters such as size of legs, thickness of bones, dimensions of eggs, lifetime, heart frequency, speed of running, swimming, etc.

These parameters are always first estimated experimentally, followed by attempts at theoretical explanation. Sometimes simple relations are combined with measured allometric exponents. For example, the number of steps (n), a running animal must take per unit distance, inversely related to the length of the steps, and therefore inversely proportional to its length (l). This in turn can be related to the body mass:

$$n \sim \frac{1}{l} \sim m^{-1/3} \quad (3.250)$$

Furthermore, flat running can be related to the corresponding friction, whereas the climbing involved in running uphill means it must additionally account for the increase in potential energy. These types of speculation are usually connected again to metabolic and structural factors, deriving various dimensionless parameters.

Further Reading

Ahlborn 2005; Bejan and Marden 2009; Da Silva et al. 2006; Schmidt-Nielsen 1999; West and Brown 2004.

Chapter 4

Physical Factors of the Environment

Physical influences from the environment on biological systems are considered an energetic, as well as an informational input. This chapter therefore on the one hand refers to the biophysical basis of sensory systems, and on the other hand to the input of energy. This energy input can occur either by a specific mechanism of absorption, for example in the case of photosynthesis, where the energy of photons is transformed into chemical energy, or it can be absorbed in an unspecific way, randomly modifying molecules or directly dissipating in the form of heat.

Beside aspects of sensory physiology, these interactions therefore concern the fields of environmental biophysics especially problems of radiation protection, as well as medical physics in respect of therapeutic and diagnostic use of physical influences.

In order to affect the molecular structures of a biological organism, the input energy must overcome the activation energy of a particular biochemical reaction or molecular displacement. Regardless of whether this energy input is sufficient for this or not, in any case this interaction finally ends in thermal motion, or using the language of phenomenological thermodynamics: in entropy dissipation. This eventually leads to a greater or smaller increase of the temperature in the system, and possibly to thermal effects.

Already in 1817 Theodor Grotthuss and John W. Draper stated that only that light which really is absorbed by a system can bring about a photochemical change. This circumstance has been generalized as the so-called *Grotthuss–Draper principle*, whereby it is not the energy penetrating the organism which is effective, but that part of the energy that is actually absorbed by the system. Investigations into the mechanism of molecular interaction, which forms the basis of all questions of the influence of any physical input into a biological system, therefore start with the question: how and where is this kind of energy absorbed? Eventually it must be clarified whether particular changes are induced by the absorbed energy, and how these influence the biological system.

In any case the relation must be considered between the absorbed energy, and the energy of thermic noise (kT , resp. RT , see Sect. 2.1.5). This question, however, cannot be answered in a general way. The prerequisite is knowledge of the particular molecular absorption mechanism. In many sensory systems special mechanisms are

developed to overcome thermal noise. Conversely, a particular noise can even increase the sensibility of a sensory system through stochastic resonance (see Sect. 3.1.5).

As a peculiarity of biological systems particular cascades of amplification exist. Very small amounts of energy, occasionally being absorbed at particular targets, can significantly influence biological mechanisms of control, and eventually will be expressed in a macroscopically visible way. The place of interaction in this case does not need to be a specific receptor, but possibly it could be a crucial point in the network of reactions. An example of this is the induction of mutations by ionizing radiation. In this case a single modification of a DNA molecule could lead to dramatic consequences in the system. Even if the theory of biological amplification was first developed in radiation biology, it is crucial also for a number of other physical influences.

In this context it should be noted that organisms have adapted through evolution to all physical parameters of their environment. In this sense, for example, the development of double-stranded DNA could be considered as a precaution against single-strand breaks caused by ionizing radiation from the natural environment. Additionally, a large number of mechanisms have been developed to repair these kinds of damage. In this context the problem is relevant, how does the range of completely new environmental influences caused by man affect life today. This is a topical problem of environmental biophysics concerning especially electromagnetic influences of power lines and various radio frequency-emitting devices.

Further Reading

On environmental biophysics in general: Campbell and Norman 1998.

4.1 Temperature

In Sect. 2.1.5 the Arrhenius equation was introduced to describe the temperature dependence of chemical reactions. We learned that all elementary reaction steps will be accelerated by an increase in temperature helping to overcome the corresponding activation energies. In this way, even concentrations of metabolites in a steady-state system can be changed. Degree and direction of these changes depends on the extent to which two opposing fluxes are affected by temperature alterations. If for example, in an open system the temperature dependence of the rate of decomposition resp. of the efflux of a certain component is predominant then with rising temperature the corresponding steady-state concentration decreases, and vice versa. It follows that in complex systems, an increase in temperature can lead to an increase, as well as to a decrease in the overall activity. An optimum of temperature therefore exists for all biological functions.

There are two possible strategies that the biological system may use to handle temperature changes in the environment. On the one hand there exist various mechanisms of protection against the results of dangerous heating or cooling, and on the other hand some organisms achieve a constant temperature in their body, which is more or less independent of the temperature of the environment. Both of

these mechanisms require a precise system of thermoreceptors. In homeothermic animals, such as birds and mammals, but also in some poikilothermic organisms a particular mechanism of thermoregulation exists.

Let us first consider the systems of *temperature sensation*. Some animals possess a particular receptor system for infrared light, such as for example some snakes (*Boa constrictor*, *Crotalus atrox*) and insects (fire beetles, mosquitoes) which aid them in hunting, feeding, and overall survival. As much as is known to date is that these organs do not contain specific quantum receptors for infrared, but in these sensors the IR radiation is transformed into temperature changes of particular thermosensitive membranes. It has been calculated that for example the thermosensitive membrane in the pit organ of snakes already activates at temperature alterations of 0.003–0.01 K. The threshold for temperature perception in the receptors of human skin is only in the range of about 0.06 K.

A particular kind of temperature response is *thermotaxis*, the oriented movement of organisms in a thermal gradient. This kind of behavior can be seen in various organizational levels from bacteria (*Escherichia coli*) to more highly organized organisms (*Caenorhabditis elegans*); its molecular mechanisms, however, remain largely unknown.

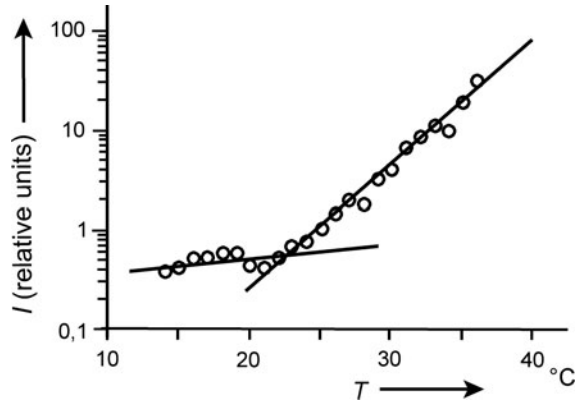
Recent research, however, allows some insights into the molecular basis of thermoreception itself. In general two different types of thermosensitive molecules have been found. The members of one family are temperature-dependent RNA molecules, forming a special kind of *riboswitch*, which have been found in bacteria and in cells of many other organisms. In *E. coli* for example, these molecules are responsible for the activation of enzymes for the synthesis of new phospholipids, for regulation of the expression of heatshock proteins, etc. In the particular case of *E. coli* the metabolism is activated if the bacterium is transferred from a cold environment into a warm homeothermic organism.

The other type of thermoreceptor molecules are membrane-located proteins, such as for example some of the so-called TRPV-transport proteins, which usually are cation channels, located in the membrane of various cells everywhere in the living world (TRP stands for “transient receptor potential,” V indicates a vallinoid-sensitive subfamily). These molecules are not only distributed in skin and various endothelial cells, but they are especially responsible for thermoreception in neurons. Recently they have also been found in the thermosensitive membrane of the pit organs of rattle snakes.

A characteristic of both types of thermosensitive molecules is their property to achieve particular structural conformations always in a quite close temperature range. In this case the RNA riboswitches will be activated, resp. the gating mechanisms of the protein channels change their transport properties. Recently it was shown that the pathway for temperature-dependent activation of TRP channels is distinct from those for ligand- and voltage-dependent activation and involves the pore turret. The pore turret therefore, is an important part of the heat activation machinery of this molecule.

As an example, Fig. 4.1 shows the abrupt change of the permeability of a thermosensitive cation channel TRPV4 at 24°C, which is characterized by the average current in a patch-clamp experiment. As explained in Sect. 2.1.5, the Q_{10} -parameter is

Fig. 4.1 Heat-activated average current (normalized to that at 25°C) in a temperature-sensitive TRPV4-channel. The lower slope exhibited $Q_{10} = 1.6$; the higher slope $Q_{10} = 19.1$. From the intersection of these two lines, the threshold temperature was estimated to be 24°C. In Fig. 2.4 the same data are demonstrated in an Arrhenius plot (Data from Watanabe et al. 2002)



an indicator for the increase of a reaction rate in the temperature range of 10 K. For biochemical reactions this factor usually amounts to around 2. This is in fact also the case for this transport protein in the temperature range below 24°C. At higher temperatures, however, this factor abruptly increases up to 19.2. Small temperature elevations therefore dramatically increase the transport rate of this channel.

Thermosensitive cells usually contain a large amount and diversity of such molecules. Each of these kinds of receptor molecules operates over a specific temperature range and also in a specific way. In some cases biochemical reactions as a whole exhibit a so-called *non-Arrhenius* behavior. This means that an increasing activity is seen in one, and a decreasing activity in another temperature region, caused by the combined action of cold and warm receptors. Of course, this contradiction to the Arrhenius rule is only apparent, because in fact each class of these molecules behaves exactly according to the rule, but one changes the process in one direction, while the other changes it in the opposite direction.

Thermoregulation itself is controlled by these thermoreceptors which are distributed everywhere in the body. In Fig. 4.2 the system of thermoreception in men is illustrated schematically, starting at the molecular thermosensors in thermoreceptor cells and finally leading to psycho-physiological reactions in the brain. This scheme illustrates the noise reduction of this system by averaging over a large number of receptors, the local consequences of thermoreception in each position (the system of thermosensitive riboswitches is not included), and finally the information processing in the hypothalamus. For minimal temperature changes, only the local reactions will be activated without sending the signal to the cortex. This means that various physiological regulations are possible even without a conscious feeling of warmth or cold.

The regulation of temperature in the body in general is considered a steady-state system composed of heat production through metabolism on the one hand, and heat dissipation into the environment on the other hand. In Fig. 4.3 different forms of heat transport between an organism and its environment are illustrated. In general, a distinction has to be made between conduction, convection, radiation, and heat loss due to evaporation of water. It is possible to define a thermal flux (J_Q) which is the thermal energy that is exchanged per unit of time across a particular area of the surface.

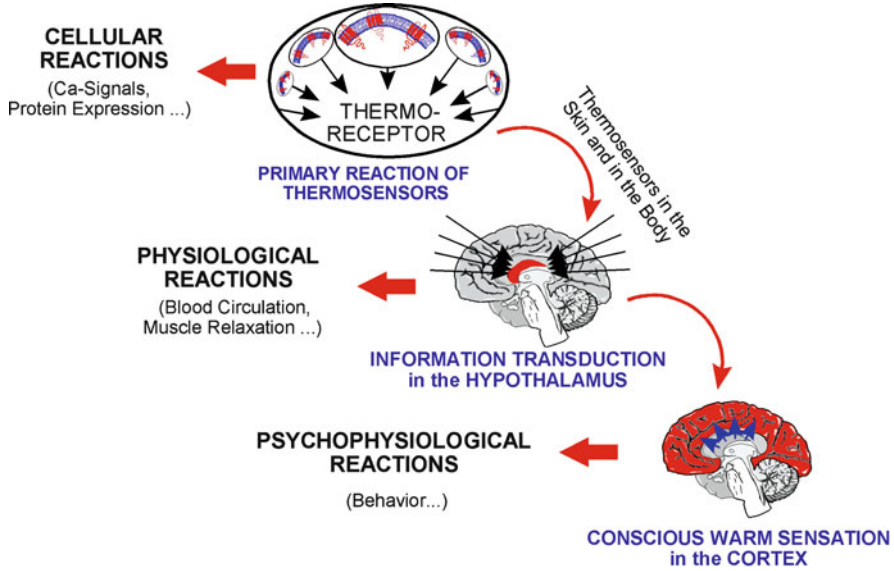


Fig. 4.2 A schematic illustration of the steps in the hierarchic system of neuronal thermosensation, starting from molecular reactions in thermoreceptor cells, up to the information center in the hypothalamus, leading under some conditions to a conscious warm sensation. In each of these steps, filtering is accomplished to reduce noise, and each step is sensitive to physiological influences. Other thermally sensitive effects, such as variation in expression of heat shock protein or effects of thermoregulatory changes in blood flow, are possible at each stage of thermoregulation (After Foster and Glaser 2007 modified)

Thermal conductivity is governed by *Fourier's law*, an equation which corresponds to the first Fick's law, which was used to calculate the diffusion of a substance (see Sect. 3.3.1, Eq. 3.131). For a temperature gradient in the x -direction, the heat flux \mathbf{J}_{Qc} (in J m^{-2}) can be calculated by the following equation:

$$\mathbf{J}_{Qc} = -\lambda \frac{dT}{dx} \quad (4.1)$$

The negative sign indicates the direction of the thermal flux from higher to lower temperature. The factor λ in this equation is the *thermal conductivity*, also called *thermal coefficient*. It has the measuring unit $\text{J m}^{-1} \text{s}^{-1} \text{K}^{-1}$, accordingly: $\text{Wm}^{-1} \text{K}^{-1}$. In Table 4.1 some examples of thermal conductivity are listed. The data shows the good thermal insulating properties of a layer of fat.

Relating to the diffusion of substances the heat flux can also be included into the flux matrix (Sect. 3.1.3, Eq. 3.58), and consequently it can be coupled with the transport of matter, the so-called *thermodiffusion*.

If the stationary heat flux (Eq. 4.1) is not of interest, but the time dependence of the temperature at a particular point in a heat-conducting medium, than in

Table 4.1 Thermal conductivity for various tissues and other biological materials (After Precht et al. 1955)

	Thermal conductivity [J m ⁻¹ s ⁻¹ K ⁻¹]
Tissue in vitro	
Human skin: epidermis and corium	0.34
Epidermis (pig)	0.21
Fat (pig)	0.17
Muscle (pig)	0.46
Living tissues in situ	
Human skin, poor blood flow	0.31...0.33
Human skin, strong blood flow	1.5
Human muscle, no blood flow	0.46
Human muscle, normal blood flow	0.53
Human muscle, strong blood flow	0.63
Hair and feathers	
Wool (loose)	0.024
Feathers	0.024
Rabbit fur	0.025
Other substances	
Air	0.023
Water (20°C)	0.59
Silver	420

correspondence to the second Fick’s law (Eq. 3.132), the second deviation of the temperature is included:

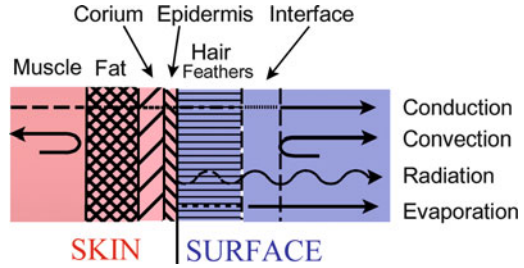
$$\frac{\partial T}{\partial t} = \frac{\lambda}{\rho C} \nabla^2 T \tag{4.2}$$

In contrast to Eq. 4.1, here the three-dimensional temperature gradient is considered. The Nabla operator (∇) stands for grad T , i.e., for deviation of T with respect to x , y , and z . The operator ∇^2 means the second derivation, the same as: div grad T . This equation contains the specific heat capacity (C) which in biological tissue amounts to approximately 3,600 Jkg⁻¹ K⁻¹, and the density (ρ in kg m⁻³).

Convection allows a very intensive heat transport at physiological temperatures. As shown in Fig. 4.3, convection in homoiothermic animals is mediated internally through blood circulation, and externally through the flow of air or water. At constant flow, the amount of heat transported by convection is proportional to the temperature differences between the two phases. The flow rate on the one hand, and the geometric conditions of the boundary layer on the other are crucial for the effectiveness of this kind of heat transfer. The level of convection depends on the thickness of unstirred layers of air or water near the body, and on the flow behavior at the boundary layer (see Sect. 3.7.1). This insulating layer in some cases is improved artificially by hairs or feathers.

Thermal radiation is an electromagnetic radiation in the Terra-Hertz or IR frequency range (see the spectrum in Fig. 4.32). To investigate thermal radiation, a so-called *black body* is used. This is a physical body that will completely absorb

Fig. 4.3 Various forms of heat transport between the body and its environment



all of the infrared radiant energy that falls on it, and correspondingly changed its temperature. The black body is also suitable for measuring the radiation which is emitted by an object at a given temperature. In contrast to the other kinds of heat transfer, thermal radiation does not depend on the temperature difference between the body and the environment but rather on the absolute temperature of the emitting body itself, whereas the temperature (T) occurs in the fourth power. The amount of heat emitted by a black body (J_{QR}) obeys the *Stefan–Boltzmann law*.

$$J_{QR} = \sigma T^4 \quad (4.3)$$

The proportionality constant is $\sigma = 5.67 \cdot 10^{-8} \text{ Wm}^{-2} \text{ K}^{-4}$. Additionally, a correction factor is used if the surface of real bodies does not behave like a black body. For the surface of biological objects it lies between 0.9 and 1. The value depends on the pigmentation of the organism.

Evaporation of water is an important mechanism of temperature regulation of animals living in air. The evaporation of 1 g of water will cause the body to lose about 2.4 kJ. The rate of evaporation depends on the surface structure, on the wind velocity, and on the difference in the vapor pressure between the surface of the body and the surroundings. Because the vapor pressure is temperature dependent, evaporation can still take place when the air is saturated with moisture provided that the ambient temperature is below the body temperature.

For a number of problems in the field of occupational health and safety, as well as in various kinds of therapeutic treatment the overall temperature balance in the human body is of interest. What is the temperature increase in a particular part of the body if it is exposed for example to ultrasound, or to high-frequency electromagnetic fields, or other kinds of energy input (see Sects. 4.3.5, 4.7)? In 1948 Harry H. Pennes developed the so-called *bioheat equation*, which is an expression of the heat balance in the body as the result of internal and external heat exchange.

In its general form this equation can be written as follows:

$$C \frac{\partial T}{\partial t} = \frac{\lambda}{\rho} \nabla^2 T + SAR + \Phi - B(T - T_B) \quad (4.4)$$

The first term in the sum corresponds to the heat conduction (Eq. 4.2). The second summand reflects the heat input from outside, whereas SAR means *specific*

absorption rate (for more detail see Sect. 4.7.1). Φ is the Rayleigh dissipation function (see Sect. 3.1.4, Eq. 3.64), and represents the heat production by metabolism. The last negative term concerns the heat acquisition by blood circulation. This in fact is a quite complex expression, including various anatomical and hemorheological properties. All the summands in Eq. 4.4 have the measure Wkg^{-1} .

This equation can be solved only using particular assumptions of the included parameters. It becomes even more complex if it is considered that the parameters C , λ , and ρ in fact are space-dependent functions.

Further Reading

General aspects: Ahlborn 2004; Foster and Glaser 2007; Hirata et al. 2009; molecular aspects: Digel et al. 2008; Yang et al. 2010; thermoreception in snakes: De Cock Buning 1983; Gracheva et al. 2010.

4.2 Pressure

Pressure, in addition to other influences on systems in equilibrium, is governed by the *Le Chatelier principle*. It states that if a system at equilibrium experiences a change in various conditions (e.g., concentration, temperature, volume, or pressure), then it shifts to counteract the imposed change. This reaction scheme was mentioned in Sects. 2.2.3 and 2.3.2 in context with processes of self-assembly of supramolecular structures. The thermodynamic explanation of Le Chatelier's principle is provided by the Gibbs equation (see Sect. 3.1.2, e.g., Eq. 3.29). Amongst other factors, this equation indicates the shift of the energy of the system as the result of pressure alterations.

Simply put, pressure will change the volume of gas-filled cavities in biological systems. Another mechanism involves the pressure dependence of the solubility of gases in the various aqueous phases such as blood and tissue. Le Chatelier's principle in this case is reflected in the circumstance that the volume of the mixed phases liquid + gas is larger than the gas dissolved in the liquid. The reversible effect is well known as the exsolution of carbon dioxide when opening a bottle of sparkling mineral water.

These kinds of pressure effects occur already at moderate degrees of pressure alterations. They are important for problems in occupational medicine and sports. The reduced oxygen solubility in the blood which occurs for example at great heights, is particularly important for alpinists and pilots. This oxygen deficit can be compensated for by an increase in the partial pressure of oxygen in the inspired air. If humans remain under an increased pressure, for example in the case of divers, the reverse effect occurs, namely an increase in the amount of dissolved gas in the blood. At greater depth this may lead to a dangerous shift in the proportions of dissolved gases. In the case of a fast return to normal atmospheric pressure, especially following a long stay at great depth, gas bubbles are formed in the blood (mainly N_2). This can cause *aeroembolism*, which is also called the *bends*

or *Caisson disease*. In general, the term *barotrauma* is used to denote various kinds of physiological damage caused by pressure alterations.

Another mechanism of pressure interaction involves different changes in the partial mole volume of various reaction components. This also leads to a shift of the equilibrium distribution according to Le Chatelier's principle. The range of pressure changes that affect this mechanism without participating gas phases is much higher than that for the cases mentioned earlier. This mechanism forms the basis of the so-called *high-pressure biotechnology*, which has been developed in recent decades. It is an interdisciplinary field of science in which pressure is applied to biological systems in order to address a wide range of issues of relevance to biochemistry, biophysics, biotechnology, and the environment.

The ecological aspect of pressure influences arises from the fact that 70% of the surface of the earth is covered by oceans with an average depth of 3,800 m. Fifty percent of the surface of our planet is covered by water at least 4,000 m deep. In fact this is the depth with the greatest number of different species. Considering that at an increase in depth of 10 m the pressure increases by about 0.1 MPa, deep-sea organisms live under conditions where the hydrostatic pressure can be as high as 100 MPa. Some organisms are called *piezophiles*, i.e., they are especially adapted for life under conditions of high pressure. So for example deep-sea microbes have been isolated from both cold and hot deep-sea environments which for their growth indicate an upper pressure limit of 130 MPa. In processes of high-pressure biotechnology the applications of pressure extend even up to 600–800 MPa, and sometimes higher. The word *piezophysiology* was proposed to describe the unique cellular responses to high pressure in living cells.

Since the early 1990s high pressure has been used also as a nonthermal method in the commercial processing of foods. The term *pascalization* was introduced for a method of pressure preservation in contrast to heat treatment of food by *pasteurization*. It turned out to be an effective approach for inactivating microbes and viruses. In some cases high-pressure treatments may cause changes which are not observed under other conditions.

What are the molecular mechanisms of these processes? As already mentioned in Sects. 2.2.3 and 2.3.2, mainly the water structure and the corresponding hydrophobic reactions are responsible for these processes. As illustrated in Fig. 2.19 the degree of orientation of the water molecules strongly influences its partial mole volume. In the case of highly condensed or coiled macromolecules the amount of bound, and therefore structured water molecules is lower than in the case of loosely packed or uncoiled macromolecules. The same concerns supramolecular structures like membranes, cytoskeleton, microtubules, etc. According to Le Chatelier's principle the application of pressure shifts the equilibrium toward higher structured water, i.e., to the condition of lower organized macromolecules. Thus, all these macromolecular and supramolecular structures that are stuck together by hydrophobic interactions will be destabilized under extremely high pressure conditions. Life in the deep sea is possible only if this risk is compensated for by special adaptations. For example, in this regard, it was found that the actin and other

proteins from deep-sea fishes have unique amino acid substitutions in comparison to those of non deep-sea fishes.

High-pressure biotechnology exploits the possibility to optimize biochemical reactions using structural peculiarities of these molecules. In this way some enzyme activities can be substantially elevated by increases in pressure.

Further Reading

Bartlett 2010; Péqueux and Gilles 1985; Winter 2003.

4.3 Mechanical Oscillations

Mechanical oscillations, on one hand, can influence biological systems directly by an input of considerable energy; on the other hand, they can be recorded by animals with the help of highly specified receptors and serve as physical carriers of information. Oscillations in the air at frequencies between 16 Hz and 20 kHz are perceived by humans as sound. Low-frequency oscillations (lower than 20 Hz) are called *infrasound*, while high-frequency oscillations (higher than 20 kHz) are called *ultrasound*. Besides this, vibrations can also influence biological systems by direct contact with vibrating materials at various frequencies. Because of these peculiarities, these kinds of vibrations will be considered in a separate chapter.

4.3.1 Vibration

Biological effects of vibration, on one hand, are of particular interest in occupational medicine; on the other hand, they are used in various kinds of medical treatments and for training aspects in sports. Vibrating seats of tractors, trucks, cars, etc., which create whole body oscillations, as well as the use of various vibrating handheld machineries, which create vibrations of the arms and hands, are considered as possible sources of occupational diseases. Whole body vibrating devices are used for therapeutic purposes to treat particular body parts affected by sports injuries.

If the body is treated by harmonic oscillation with the angular frequency ω , the displacement $x(t)$ occurs in the form:

$$x = x_{\max} \sin \omega t \quad (4.5)$$

where x_{\max} is the amplitude of this oscillation. Deriving this equation in respect to the time (t), the function for the vibration velocity $v(t)$ is obtained by:

$$v = \frac{dx}{dt} = \omega x_{\max} \cos \omega t = v_{\max} \cos \omega t \quad (4.6)$$

These equations, though, represent only a simplified case. The vibrations usually occur in all directions, i.e., the derivatives of x , y , and z should also be considered.

In order to evaluate its influence on living systems, the energy input of this vibration is required. This is a function of the acceleration (a), resulting from the second derivation of Eq. 4.5:

$$a = \ddot{x} = -x_{\max}\omega^2 \sin \omega t \quad (4.7)$$

with a peak acceleration:

$$a_{\max} = x_{\max}\omega^2 \quad (4.8)$$

Usually, the total vibration exposure for sinusoidal oscillations is expressed by RMS (root-mean-square) given by:

$$a_{\text{RMS}} = \frac{a_{\max}}{\sqrt{2}} \quad (4.9)$$

In the case of a time variable intensity of the exposition, the root-mean-square of the frequency weighted, a_{RMS} , is evaluated, where t_D is the duration of the exposition:

$$a_W = \sqrt{\frac{1}{t_D} \int_0^{t_D} a_{\text{RMS}}^2(t) dt} \quad (4.10)$$

The use of frequency weighted exposition results from differences in the biological effectiveness of vibrations. For this, a frequency-dependent evaluation factor is introduced, based on empirical values (Fig. 4.4). There are only slight variations when applied to the whole body or only to parts of the body, like the arms and hands. This curve shows that for biological systems, the most effective frequencies are in the region of about 10 Hz.

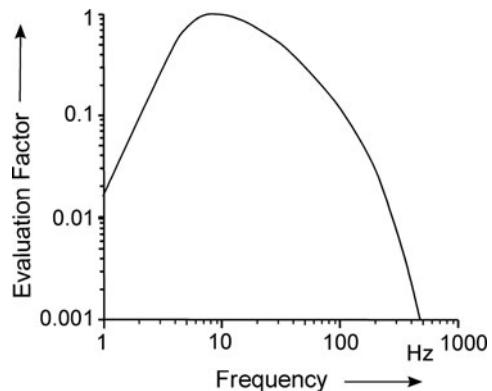


Fig. 4.4 Evaluation factor indicating the relative sensibility of human body to vibrations in dependence of the frequency

Another problem consists in determining the kind of mechanical coupling used to link the body to the oscillating materials and transmitting the vibration to its different parts. As an example, Fig. 4.5 shows the movement of the head, shoulders, and hips of a person sitting on a vertically vibrating seat. The relation between the amplitude of vibration of the specific points of the body and the amplitude of vibration of the support is depicted on the ordinate. This demonstrates that the amplitudes of vibration in different parts of the body are different and a function of frequency. At frequencies below 2 Hz, the body simply follows the movement of the seat, i.e., the degree of amplification is about 1. When the applied frequency is approximate to the resonance frequency or equals it, the amplitude of the enforced oscillation is greater than that of the seat. Even at the fundamental resonance frequency of a human body of about 5 Hz, differences are observed between vibrations of head, shoulder, and hip. This is the frequency at the maximal displacement of all three reference points. The head, furthermore, shows a second maximum of vibration amplitude at about 20 Hz. A further increase in the external frequency leads to a decrease in the coupling efficiency up to 0, at $\omega \Rightarrow \infty$.

The biophysical evaluation of possible mechanical influences on cells and organs, like muscles, joints, and bones, requires analytical models based on data of viscoelastic properties of the tissue. The oscillation of a mechanical system can be treated analog to the electrical oscillations of an RCL circuit (see also Sect. 3.5.3). Similar to these approaches, complex parameters are used to describe damping and elastic parameters of these mechanical systems. In analogy to electrical systems, a *mechanical impedance* of the system can be defined. It is the relation between vibrational force and vibrational velocity.

In practice, the quantitative evaluation of possible exposition of humans in vibrating devices is quite complicated. In most cases, the vibration does not consist of only simple sinus waves, but rather a sum of many frequencies differing in

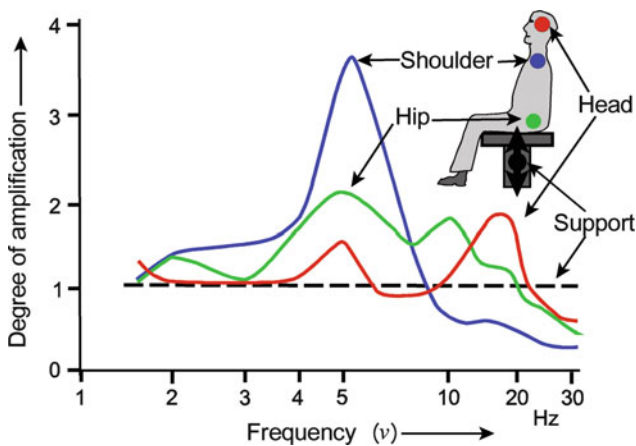


Fig. 4.5 The amplification factor for vertical oscillations at various points of the body of a man sitting on a vibrating support as a function of frequency (Redrawn from Dieckmann 1984)

amplitude as well as in phase. Using Fourier analysis, these can be regarded as a sum of many sinusoidal oscillations. On the other hand, the biological influence of vibration cannot be calculated simply by physical models, using passive mechanical properties of anatomical constituents of the body. It must be taken into account that forces are generated by the muscles actively trying to compensate the external stress. This, however, in some cases, can lead to an opposite effect, namely, to an additional strain on the joints. In general, one must take into consideration that the mechanical impedances of parts of the body are influenced by neurophysiological processes.

As already mentioned, the biological effects of vibration find applications not only in occupational medicine but, to some extent, also in therapeutics, rehabilitation, and preventive medicine, as well as in sports and in prevention of degenerative effects due to loss of weight in men. Vibration exercises activate the muscles, induce warm-up effects, and help to improve flexibility. In general, as an effect of vibration, muscles and tendons tend to elongate initially (stretch phase), followed by a period of shortening (shortening phase). Vibration, therefore, can be characterized by a cyclic transition between eccentric and concentric muscle contractions. In the same manner, it acts on joints and bones, while strain energy density or fluid flow may serve as osteogenic stimuli. Vibration can affect muscle perfusion and increase the metabolic demands of the contracting musculature. These reactions are partly stimulated by mechanoreceptors, such as *Meissner corpuscles*, which are most responsive around 40 Hz, and the *Vater-Pacini corpuscles*, in the frequency region of 100 Hz and above.

The question whether these frequencies are detrimental in the sense of occupational protection or helpful regarding therapy or sports apparently depends on the intensity, duration, and the kind of exposure. The long-term biological effects of whole body vibrations consist mostly of wear of the joints. Continuous use of vibrating handheld machinery or exposure to other kinds of local vibrations may cause the hand-arm vibration syndrome or the *Raynaud's phenomenon*, a vasospastic disorder causing discoloration of the fingers, toes, and occasionally other extremities.

The frequency range below 2 Hz shows some neurophysiological peculiarities. At these extremely low frequencies, the effects of vibration cannot be simply attributed to mechanical processes. This applies, for example, to the occurrence of *kinetosis*. Particularly, a frequency of 0.3 Hz causes seasickness. All these conditions are caused by neurophysiological influences and are induced by the vestibular system that detects accelerations of the head. This can lead to variations in blood circulation, along with all subsequent physiological consequences.

The mechanisms by which humans and animals can perceive vibrations are a special field of research. Humans, for example, can detect a frequency of 200 Hz with an amplitude of only 0.1 μm ! The sensibility of insects and spiders can be several orders of amplitude higher.

For water-gliding insects that hunt for objects fallen on the surface of relatively still ponds, the surface waves act as important signals. The propagation of these waves, its velocity and frequency are functions of surface tension and density of water. The typical frequencies are in the range of 10–100 Hz and the propagation speed between 0.3 and 0.5 m/s.

Further Reading

Arnason et al. 2002; Fritz 1998, HVDI 2057 2002; Mansfield 2005; Rittweger 2010; Rossing and Fletcher 2004; Wilson 1994.

4.3.2 Sound

Sound consists of longitudinal, i.e., compression, waves transmitted through solids, liquids, or gases. This leads to periodic deviations in pressure from that at the equilibrium. As a result of this pressure oscillation (p in $\text{N m}^{-2} = \text{Pa}$), a certain *particle displacement* (Δx) occurs, with a *particle velocity* v_p , and, therefore, an energy dissipation. The power density of sound is termed *sound intensity* (I in W/m^2). The physical link between these parameters is given by the *acoustic impedance* (Z in $\text{N s m}^{-3} = \text{Pa s m}^{-1}$), which itself appears as a function of frequency. These parameters are linked by the following relations:

$$p = \rho v_s \omega \Delta x = Z \omega \Delta x = Z v_p \quad (4.11)$$

and

$$I = Z v_p^2 = \frac{p^2}{Z} \quad (4.12)$$

where ρ is the density of the medium (in kg m^{-3}) and v_s is the speed of sound, distinct from the particle velocity v_p . The values of these parameters for air and water at 20°C are listed in Table 4.2. The parameters of soft tissue are similar to that of water, whereas the speed of sound in bone is much greater.

To characterize the sound intensity, a unit that was first applied in the Bell Telephone Laboratories to characterize the quality of telephone transmission is used. A *Bell* was defined as the base-10 logarithm of the ratio of the measured power to a reference power. Later, the *decibel* (dB) was used, which was ten times the base-10 logarithm of the ratio of measured power to reference.

This parameter, which is used widely in engineering techniques, coincides with experiences in the physiology of sensory reception. At the end of the nineteenth century, the physiologist Ernst Heinrich Weber found that alterations of external signals are perceived more sensitively the lower the absolute intensity of the signal is. Later, the psychologist Gustav Fechner elaborated the theoretical interpretation of this connection and formulated the logarithmic relation between stimulus and perception. This is the so-called Weber–Fechner law, which is valid for all kinds of sense organs.

Table 4.2 Acoustically relevant parameters of air and water

Medium	Density (ρ) [kg m^{-3}]	Speed of sound (v_s) [m s^{-1}]	Acoustic impedance (Z) [N s m^{-3}]
Air	1.204	343	413.5
Water	$1 \cdot 10^3$	1 440	$1.44 \cdot 10^6$

According to these considerations, the *sound level* (L), which is measured in dB according to the intensity relation (I/I_0), is defined as follows:

$$L = 10 \log \frac{I}{I_0} = 10 \log \frac{p^2}{p_0^2} = 20 \log \frac{p}{p_0} \quad (4.13)$$

The introduction of p^2 instead of I corresponds to Eq. 4.12.

This relation requires the reference parameters I_0 and p_0 . For this, the human hearing threshold level of 1,000 Hz was used as the reference frequency (see Fig. 4.6). The values for these parameters are $I_0 = 10^{-12} \text{ Wm}^{-2}$ and $p_0 = 2 \cdot 10^{-5} \text{ Pa}$.

Figure 4.6 shows a frequency diagram of human auditory sense. The sound level (L in dB) and the sound pressure (p in Pa) are plotted against the frequency. To characterize the sensibility of the human ear at all frequencies, furthermore, the unit *phon* is introduced. By definition, one phon is the apparent loudness of a sound equal to 1 dB of a sound having a frequency of 1,000 Hz. As an example, in Fig. 4.6, the 60-phon curve is shown crossing the frequency of 1,000 Hz at the level of 60 dB. Similarly, the 0-phon curve is plotted, indicating the auditory threshold at all frequencies.

This indicates that a 50 Hz sound, for example, with an objective sound intensity level of 120 dB, subjectively seems to have the same loudness as a 1 kHz sound of only 60 dB. The phon unit, therefore, is a subjective measure which is related to a normal human ear.

It is, however, useful to evaluate not only various *pure tones* (sinus waves) but also *sounds* (true musical tones with overtones) and *noise* (random mixture of various frequencies). This evaluation depends not only on the loudness but also on the duration of the sound. Sounds with a duration less than 1 s will be subjectively underestimated.

Furthermore, a simple measurement of an overall sound intensity in dB is too crude to be used for practical environmental protection because of the broad

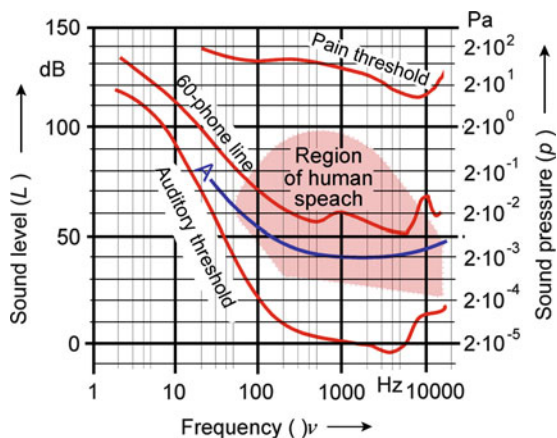


Fig. 4.6 Response of human ear to sound intensity. The curves are extended up to the frequencies of infrasound ($\nu < 20 \text{ Hz}$) according to data from Møller 1984. A-40-phon weighting scale as in Fig. 4.7

variations of the character of noise. Usually, weighing scales are used to measure a particular noise (Fig. 4.7). These scales are determined empirically and fixed by standardization. They are always normalized for 1 kHz and take into account the frequency dependence of the sensibility of the human ear. With the help of these scales, it is possible to take into account the lower sensibility of the human ear to frequencies that are lower or higher than 1 kHz. Of course, such scales are dependent on the real character of the measured noise and on the question being asked. There are differences in the frequency mixture, depending on whether the noise comes from engines, from a discotheque, from traffic, or from other sources. It also makes a difference as to whether the noise limits of a workplace in a factory, in an apartment, or in a hospital are to be determined. Mostly, the weighing scale A (Fig. 4.7) is used. It particularly takes into account the low sensibility of the human ear to lower frequencies (Fig. 4.6), and not so much to the higher ones. A very flat scale B (not drawn) is not used anymore. Instead, the C-scale is used for noise in a very broad frequency range. The SI-scale considers particularly the disturbances in human communication. In Fig. 4.6, the parameters are shown according to the scale A to indicate its similarity to the sensibility of the human ear.

Figure 4.8 shows some examples of noise production from various sources. It has been known for a long time that in case of permanent exposure, vegetative diseases may possibly occur at 60 dB. The level of dangerous noise at low

Fig. 4.7 Weighting scales for practical noise measurement (description in the text)

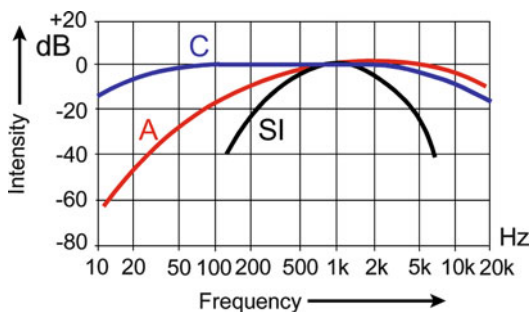
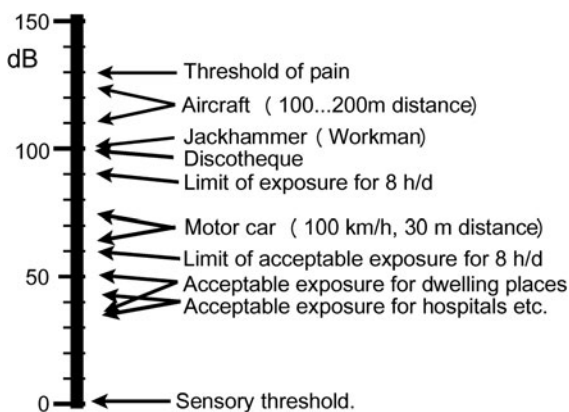


Fig. 4.8 Examples of noise production from various sources



frequencies is 80 dB and at high frequencies is 90 dB. Sound intensities that can induce acute injury lie above this region.

From the legal point of view, noise is a sound which is able to disturb, endanger, or significantly discriminate or cause inconvenience to neighbors or anybody else.

Further Reading

Everest and Pohlmann 2009; Rossing and Fletcher 2004; Wilson 1994

4.3.3 The Biophysics of Hearing

Anatomically, the ear can be divided into three parts: outer, middle, and inner (see Fig. 4.9). The eardrum, or the tympanic membrane, separates the outer ear from the inner ear, and both are air filled. The middle ear contains three ear ossicles (*malleus*, *incus*, *stapes*), which transmit the vibration from the eardrum to the oval window. The cochlea is a part of the inner ear, which also includes the vestibular apparatus, a sense organ for gravity and motion (not included in Fig. 4.9). The cochlea is a spiraled, hollow, conical chamber of bone, containing the *scala vestibuli* and the *scala tympani*, which are connected together at the top by the *helicotrema*. Both are filled with a sodium-rich perilymph. The *scala tympani* terminates at the round window. Between these two scales is located the *scala media* (see also Fig. 4.15), which contains the *organ of Corti* having hair cells that are vibration-sensitive elements. The *scala media* as well as the vestibular apparatus is filled with potassium-rich endolymph.

The human ear as well as the ears of other vertebrates are amazingly optimized organs. This concerns not only the astonishing sensibility, the noise suppression, and the property to analyze the sound, but also its angular localization. All these properties are the result of an interplay of at least four components, each of them with particular mechanisms requiring different biophysical approaches, not including the neuronal sound analysis in the brain:

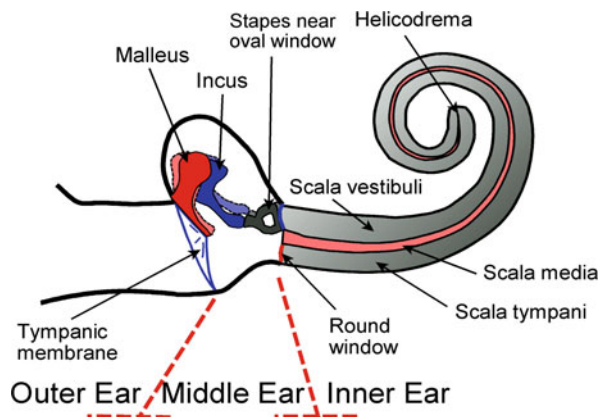


Fig. 4.9 Schematic illustration of the human ear consisting of the air-filled outer and middle ear, with the eardrum in between, and the inner ear, filled with perilymph (*scala vestibuli* and *scala tympani*)

- The outer ear, as an amplifier with particular frequency and angular characteristics
- The middle ear, as an adjustable impedance transducer, transforming vibrations of air in vibrations of a kind of an aqueous medium
- The cochlea, where traveling waves appear with frequency-specific maxima at particular locations
- The organ of Corti with hair cells as sensible noise-reducing elements, transducing vibrations into coded nerve impulses.

In this way, biophysical approaches are essentially applied to various different levels of organization, from laws of phenomenological acoustics up to various considerations at the cellular and the molecular level.

The *outer ear* consists of the pinna, the concha, and the auditory meatus. The human outer ear selectively concentrates the sound pressure by 30- to 100-fold at frequencies between 2 and 3 kHz. This corresponds to the frequency range of the human voice (see Fig. 4.6). At the same time, this amplification reduces the noise of other frequency bands. These characteristic frequencies of sound acceleration and filtering, and correspondingly also the anatomical structure of the outer ear, are optimized in various animals in different ways corresponding to conditions of their life.

Furthermore, the sensibility of the human outer ear, and even to a larger extent that of other animals, depends on the direction from where the sound comes. In general, however, the process of sound localization is based on the analysis of the time delay in which the sound reaches the two ears and on the evaluation of the differences in the corresponding intensities. Humans are able to localize a source of sound with an angular resolution of 3° . This means that the minimal time resolution of the differences of arriving auditory signals must be about 35 μ s. In fact, the conversion of a mechanical stimulus into an electrical potential in the hair cells of the inner ear lasted as little as 10 μ s. In fact, the entire process of angular sound localization occurs in the temporal lobe of the brain.

To differentiate between neurobiological and purely biomechanical mechanisms of sound analysis, the following experiments can be instructive. Two pure tones are transmitted by a headphone. In the first case, both of them are simultaneously transmitted into both earpieces, whereas in the second case, one tone is transmitted into the right and the other into the left earpiece. If both tones are mistuned, having nearly the same frequency, first-order beats occur, i.e., the amplitude of the resulting tone becomes low frequency modulated. This sort of mistuned sound is recognized if both tones are fed together into both earphones. In this way, a so-called first-order beat results, which is sensed in the inner ear. In contrast to this, tones with nearly harmonic frequencies, i.e., mistuned fourth or third, etc., produce *beats of the second order*, namely, a sort of frequency modulation. In this case, recognition is also possible if each tone is fed only into one of the two earphones, while no physical interference of the tones can occur. They come together only in the brain.

The *middle ear* is considered to be an impedance transformer, transmitting the air sound of the outer ear into sound of the aqueous medium of the inner ear. The acoustic impedance (Z) of the lymph in the inner ear, resembling that of water, is nearly four orders of magnitude larger than that of air. Equations 4.11 and 4.12 (Sect. 4.3.2) show the interdependence of sound intensity (I), pressure (p), and

particle displacement (Δx) with this parameter (Z). The amplitude of low pressure and large displacement in air, therefore, should be transformed into vibrations of an aqueous medium with high pressure and low displacement amplitude.

As the simplest approach to calculate this efficiency of the impedance transformation, a mechanical lever arm system can be considered (Fig. 4.10). The force from the vibrating eardrum that results from the product of pressure and area ($F_T = p_T A_T$) is boosted by the ossicles represented by the malleus lever (l_1) and the incus–stapes (l_2) lever arms in the relation l_1/l_2 and transmitted via stapes to the oval window. This produces a pressure: $p_o = F_T(l_1/l_2)/A_o$.

These considerations lead to the following relation of the pressure amplification in the middle ear:

$$\frac{p_o}{p_t} = \frac{A_t}{A_o} \frac{l_1}{l_2}$$

(4.14)

In Table 4.3, characteristic parameters of this system, as depicted in Fig. 4.10, are listed for humans and various animals of different sizes. An analysis of these data indicates that the absolute size of the auditory ossicles as well of the tympanic membrane and the oval windows corresponds in an isometric relation to the size of the animals (see Sect. 3.8). The transformer ratio (p_o/p_t), however, seems to be more or less independent of the size and ranges between 30 and 80.

This system of the auditory ossicles clearly indicates specific frequency characteristics of transmission. Their dimensions and mass as well as their mechanical inertia allows to calculate the maximal frequency limits of transmission (v_{lim}) (see Table 4.3), which correspond to the physiologically and behaviorally measured values.

Fig. 4.10 Schematic illustration of the force acceleration in the middle ear between the tympanic membrane (A_t) and the oval window (A_o) by the malleus (blue, l_1) and the incus–stape (green, l_2) lever arms as a mechanical system (for corresponding values, see Table 4.3)

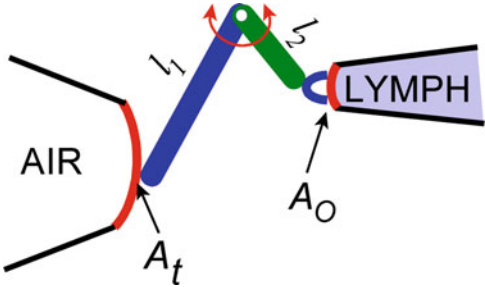


Table 4.3 Mechanical parameters of the middle ear of human and various animals corresponding to Fig. 4.10 (Data from Hemilä et al. 1995)

	A_t [mm ²]	A_o [mm ²]	l_1 [mm]	l_2 [mm]	p_o/p_t	v_{lim} [kHz]
Human	633	298	624	446	3,209	20
Indian elephant	454	136	163	85	6,409	10
Cattle	511	266	897	331	5,206	35
Dog	633	196	816	256	1,030	45
Rabbit	282	134	500	208	5,051	69
Mouse	422	15	200	52	108	92

As a result, our knowledge about the anatomy and physiology of the middle ear is improved, based inter alia on realistic 3D anatomical data from high-resolution CT scan. Obviously, the lever arm ratio of the middle ear ossicles is variable and, therefore, indicates a frequency-dependent transmission quality. Furthermore, the efficiency of this system is regulated by two small muscles that trigger the transmission by loud noise to protect the inner ear.

These circumstances are taken into account in new models, using the finite element method. The most significant development in these models is the inclusion of the function of ligaments and tendons in the middle ear as well as the properties of the external ear canal. These models allow to evaluate the static and dynamic behavior of middle ear components and to simulate various pathologic alterations.

The actual sound detection takes place in the *inner ear*. Early on, it had been found that the auditory nerve transmits so-called microphonic potentials, i.e., electrical vibrations analog to the received sound. These effects, however, are generated passively as a result of vibrations of the inner ear as inhomogeneous dielectrics. It remains unclear as to whether this has any physiological relevance.

In 1863, Hermann von Helmholtz proposed a resonance model to explain the frequency analysis of the inner ear. He postulated that the basilar membrane at particular locations along its extension shows resonance properties at various frequencies. For this kind of resonance, however, there is no experimental evidence. Nevertheless, his single-point theory of auditory perception, i.e., the idea that various positions in the basilar membrane would be sensible to particular frequencies, has been confirmed by George von Békésy (Nobel prize winner 1961), even though in quite a different way.

The sound, transmitted by the ossicles of the middle ear to the oval window, spreads out very fast through the cochlea and is finally damped at its end at the round window. In fact, the basilar membrane exhibits very specific viscoelastic properties along its extension. In humans, it is 35-mm long and gets tenfold wider over this diameter away from the oval window. Likewise, its elastic properties change. As shown in Fig. 4.11, the stiffness of the basilar membrane is reduced by about 100-fold with increasing distance from the oval window.

Investigations into the vibration characteristics of this system are very complicated. Békésy's theory can be understood using the mechanistic model shown in Fig. 4.12. To represent the sound in the cochlea, a rigidly coupled pendulum transmits its movement to a rotatable rod to which other pendulums are fixed. They are made of balls fixed on rods of various lengths, reflecting the above-mentioned viscoelastic differences along the basilar membrane. The oscillating movements of the axle cause all of these pendulums to move simultaneously at a given frequency, which, in most cases, does not correspond to the resonance frequency of the individual pendulums. Additionally, the pendulums are coupled to one another in series by threads weighted with small ball bearings. The degree of coupling depends on the point of attachment of these threads.

This model demonstrates the following properties of the inner ear. The sound waves spread so rapidly in the perilymph fluid that, similar to the effect of the axle in the model, there is a simultaneous stimulation along the entire length of the membrane.

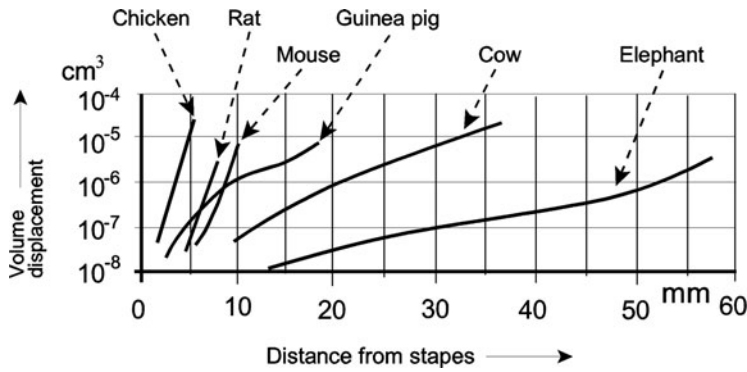


Fig. 4.11 The elasticity of the cochlear partition of various animals as a function of the distance from the stapes. As a measure, the volume displacement is used, produced in a 1-mm segment by a pressure of 1 cm of water (After Békésy, in Waterman and Morowitz 1965)

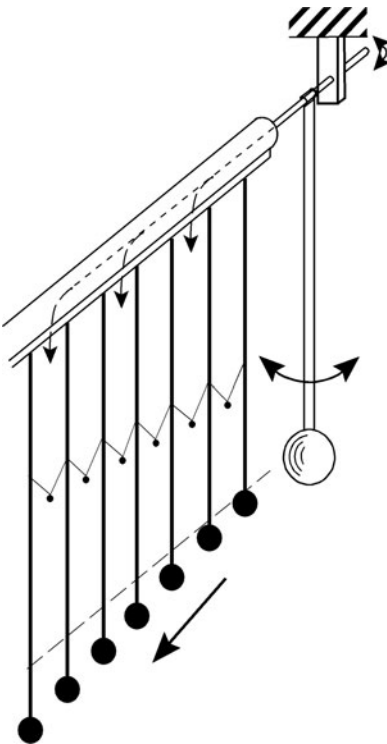


Fig. 4.12 Pendulum resonator model to demonstrate traveling waves in the inner ear (After Békésy, in Waterman and Morowitz 1965 modified)

The length and coupling of the pendulums correspond to the position-dependent viscoelastic properties of the basilar membrane. If this system is activated by a simple harmonic oscillation, traveling waves along the series of pendulums are generated, which always move from the stiffer to the more compliantly coupled pendulums. The

Fig. 4.13 Amplitude of cochlear vibration at two instances within a cycle and the corresponding envelope curve at a frequency of 200 Hz (After Békésy, in Waterman and Morowitz 1965)

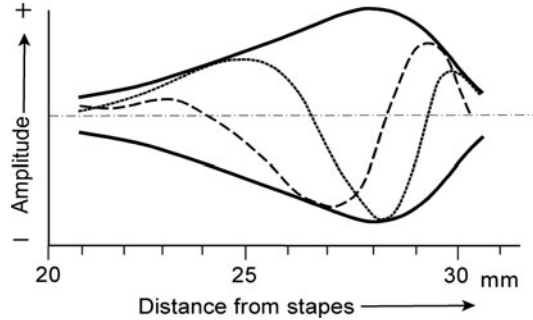
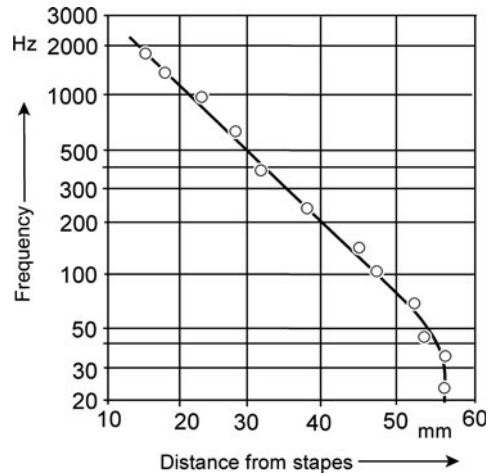


Fig. 4.14 Position of maximum stimulation along the cochlear partition in the inner ear of an elephant (After Békésy, in Waterman and Morowitz 1965)



amplitudes of these waves, however, are not constant. They reach their maxima at certain points of their movement in the direction of the arrow in Fig. 4.13, depending on the frequency of activation. Connecting these maxima, an *envelope curve* is obtained, the position of its maximum depending on the frequency of activation.

In fact, this model takes into consideration the plasto-elastic properties of the basilar membrane and the hydromechanics of the endolymph and the perilymph. It shows that the higher the frequency, the closer the maximum of this envelope curve to the oval window. Figure 4.14 indicates the positions of maximum stimulation along the cochlear partition as a function of frequency in the ear of an elephant. In the case of complicated tones, envelope curves have several maxima. At the points of maximal amplitude, the basilar membrane is deformed, which eventually leads to stimulations of the sensory cells. This mechanism can be considered as one particular step of frequency analysis.

The entire reception of the sound takes place in the *organ of Corti* (Fig. 4.15). In contrast to the anatomical-hydrodynamic and anatomical-mechanical approaches, which are used to explain the function of the middle ear and the traveling waves in

Fig. 4.15 Section of the human cochlea

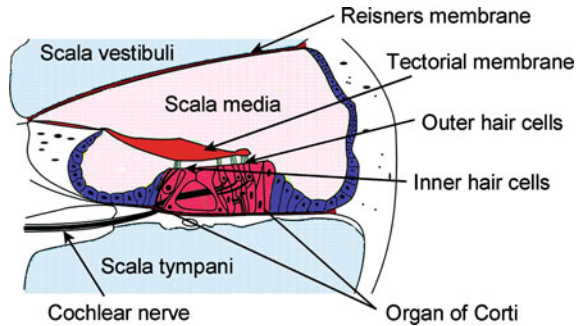
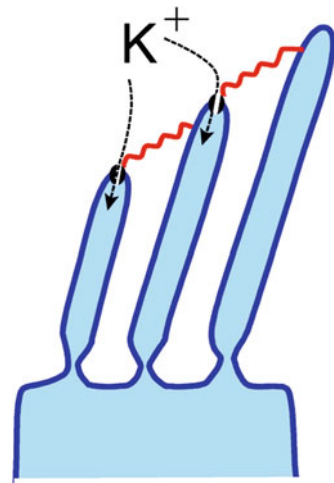


Fig. 4.16 Schematic illustration of the apical part of a hair cell with stereocilia, connected at the top by filaments (red). In the case of deflecting, these tip links mechanically open potassium-selective channels, which leads to depolarization of the corresponding cells



the cochlea, the processes in the organ of Corti have to be discussed on the cellular and the molecular level. The entire transformation of mechanical vibrations into coded nerve impulses is performed by the so-called hair cells. About 16,000 of them are located at the basilar membrane in the human ear. Hair cells are flask-shaped epithelial cells, containing 30 to hundreds of hair-like extensions, so-called *stereocilia*, in a hexagonal arrangement. These stereocilia differ in length and thickness. Some of them reach to the tectorial membrane on the opposite side of the cleft.

The basic mechanism to transform the vibration of the basilar and tectorial membranes into a neuronal signal is illustrated schematically in Fig. 4.16. Parallel to the plane of bilateral symmetry of the cochlea, the stereocilia are connected to each other at the tip by filamentous structures. In the case of bending, these tip links mechanically open cation-selective channels, leading to a depolarization of the cell. An erection of the stereocilia closes these channels again. In this way, a mechanical vibration is transformed into an electrochemical one with respect to electrical signals.

This simplified description indicates particular noise-reducing properties. The orientation of the stereocilia and their interconnection by their tip links show that

not all kinds of displacements of the stereocilia will cause depolarization, but only those that are directed in line with the filaments.

But this is only one aspect of the highly optimized system of sound reception. In fact, there are a number of other peculiarities. One of them is the enormous spread of the range of sensibility. At 1 kHz, the auditory threshold for pain (130 dB) corresponds to a sound pressure that is about 10^7 times larger than that of the sensory threshold (0 dB) (see Fig. 4.6). According to Eq. 4.11 (Sect. 4.3.2), the sound pressure should be proportional to the amplitude of the displacement (Δx). It was observed that at a threshold stimulus of 0 dB, the basilar membrane vibrates at an amplitude around $0.1 \text{ nm} = 10^{-10} \text{ m}$. Applying this proportionality, this would be $1 \text{ mm} = 10^{-3} \text{ m}$ near the pain threshold! Vibrations of this intensity, however, would definitely destroy the organ of Corti. In fact, measurements show that even at this maximal intensity, the basilar membrane vibrates only around 10 nm.

One reason for this strongly nonlinear behavior is the regulatory role of the ossicle system in the middle ear as well as the wave propagation in the cochlea is a result of its particular viscoelastic structure. Furthermore, a nonlinear behavior occurs also in the hair cells. So, a negative hair-bundle stiffness at a particular area of bending has been observed. This means that in particular situations, an additional force will be activated, supporting the bending of the stereocilia. In fact, the hair bundles can perform work and, in this way, amplify their input. Experiments on isolated outer hair cells indicate that not only the hair bundles but, after electrical excitation, the cell body itself undergoes a striking change in its shape. Furthermore, active movement of the basilar membrane has been observed.

This indicates that the inner ear is not only a simple passive receptor organ. In fact, it becomes an energy-consuming generator of sound. It was found that the ear of humans and that of many animals can emanate sound continuously at one or more frequencies. This is called *otoacoustic emission*. There are two types of otoacoustic emissions: *spontaneous* and *evoked*. Evoked otoacoustic emissions occur after the application of a pure-tone stimulus. This is used as a simple, noninvasive test for hearing defects. Otoacoustic emission seems to be the result of a frequency-specific neuronal feedback system, which leads to an increase in the distinguishability of signal variations in the inner ear.

Further Reading

Middle ear: Zhao et al. 2009; inner ear: Bekesy 1960; Ramamoorthy et al. 2010; hair cells: Hudspeth 2008; otoacoustic emission: Lilaonitkul and Guinan 2009; biophysics of music: Roederer 2008; neurophysiological background: Purves et al. 2008.

4.3.4 Infrasound

Infrasound is considered an acoustic oscillation of frequencies below 20 Hz. Its definition, as being below the low-frequency limit of audible sound, is in fact misleading, as sound remains audible at frequencies down to 4 Hz for exposure

in an acoustic chamber and down to 1.5 Hz for earphone listening. Sometimes, infrasound leads to unspecific auditory sensations. Even clinking and overtones can be heard. In general, infrasound interacts with air-filled cavities in the human body, like lungs, nasal cavities, frontal sinus, middle ear, and gut. Therefore, a range of vibrational receptors exist in the body, of which the ear is the most sensitive for higher frequencies (see Fig. 4.6), whereas vibration and contact detectors dominate at lower frequencies (see Sect. 4.3.1).

Because of its large wavelength, infrasound has some special properties in relation to sounds of higher frequencies. For example, it is only poorly damped by stone walls, buildings, etc. It is, therefore, difficult to shield from this kind of noise. Sometimes, even in spaces with corresponding resonance frequencies, an amplification occurs. These, for example, are cars or housing spaces with a typical resonance frequency of 2–8 Hz.

In general, we are surrounded by a large scale of naturally occurring infrasound in the range of about 0.01–2 Hz. These frequencies eventually merge into fast fluctuations of barometric pressure. There are various natural sources of infrasound, like wind, waterfalls, volcanic eruptions, and ocean waves, i.e., any effects that lead to slow oscillations of the air. Some speculations assume that animals may perceive infrasonic waves passing through the earth indicating the onset of natural disasters like earthquakes or tsunamis. Some large animals such as elephants, hippopotamuses, rhinoceroses, and whales use low-frequency sound and infrasound for long-range communication.

In a technical context, infrasound is produced by vibrations of various engines. There is a considerable exposition to infrasound in cars, if the air stream induces internal resonances. Sometimes the 100-dB limit is exceeded. In recent years, special infrasound loudspeakers are used in discotheques in order to cause psychosomatic resonance in the participants. Because of the above-mentioned problems of shielding, they may considerably disturb the neighborhood.

Two types of biological influences of infrasound must be considered: direct physical injuries on one hand and psychophysical influences on the other. At high intensities, damage to the middle ear is to be expected. For a 24-h exposure of sound below 20 Hz, levels of 120–130 dB seem to be tolerable. These limits, however, prevent only direct physiological damage. Unfortunately, there are no clear national and international protection standards for infrasound including their psychophysical aspects. Usually, this is managed as a subset of general protection from noise.

Further Reading

Garstang 2004; Le Pichon et al. 2010; Leventhall 2007; Møller 1984.

4.3.5 Ultrasound

Ultrasound per definition is a mechanical wave at a frequency range between 16 kHz and 1 GHz. Sometimes, for frequencies higher than 1.6 GHz, the term

hypersound is used. Technically, ultrasound can be generated by electroacoustic converters. Because of its short wavelength, it can be focused by appropriate reflectors and diffraction lenses.

In gases, fluids, and correspondingly in most tissues, ultrasound consists of longitudinal waves, i.e., of periodic alterations of local pressure, where the particles horizontally move back and forth relative to the direction of the wave energy. In harder biological materials such as bones, the particles move at right angles to the direction of the wave propagation. In these cases, shear waves occur. Diagnostic use of ultrasound is based on its interaction with the tissue in the form of absorption and scattering. Absorption is due to translational and rotational vibration of biological macromolecules even if their frequency of ultrasound is much lower than their thermal vibration modes. Scattering is related to small inhomogeneities of viscoelastic properties. Strong scattering contributions are produced by small arterioles and the collagen meshwork of parenchymal tissues. The propagation speed of ultrasound in a tissue is typically assumed to be constant at 1,540 m/s, similar to that of water.

Figure 4.17 shows the relation between the frequency (ν) and the wavelength (λ) of ultrasound in air and aqueous media. So, for example, ultrasound of $\nu = 10$ MHz in tissues has a wavelength of 0.15 mm. This ensures a satisfactory resolution for sonographic images. It can be increased further by using ultrasound of higher frequencies; however, this will decrease its penetration into the body. In fatty tissues, for ultrasound of 0.8 MHz, a half-value thickness of penetration of 3.3 cm has been measured; in muscles, it amounts to only 2.1 cm. This parameter decreases with increasing frequencies.

For diagnostic sonography (ultrasonography) as an image-generating technique, ultrasound pulses of various bandwidths are used. In general, these applications can be divided into pulse-echo imaging techniques and Doppler techniques for studying blood flow or tissue movement. 2D images are obtained by sweeping the ultrasonic beam. This results in a 2D representation of a slice in the body. A series of adjacent 2D images can be integrated into a 3D image. In this way, even a live 3D image of the beating heart is possible to obtain.

In order to achieve a good resolution of the images, an optimum of frequency and bandwidth of the applied pulses must be determined as a typical compromise: To resolve targets lying close together, side by side at the same range, a narrow beam is necessary, which requires a wide pulse. Both are possible using higher frequencies. This, however, limits the depth of penetration. This is the reason why for prenatal observations and mammographic diagnostics, frequencies of only 1–5 MHz are used. Around 3 MHz is typical of abdominal applications in adults and around 5 MHz in children. In ophthalmologic diagnosis for the anterior chamber of the eye, 10–50 MHz are used, or even 100 MHz in very superficial applications, such as imaging the cornea.

Ultrasound-induced bioeffects are generally separated into thermal and nonthermal mechanisms. The absorption of ultrasonic energy means a reduction in its amplitude as it propagates through a medium. This represents its dissipation, i.e., the conversion of mechanical energy into heat. This energy dissipation is proportional to the ultrasound

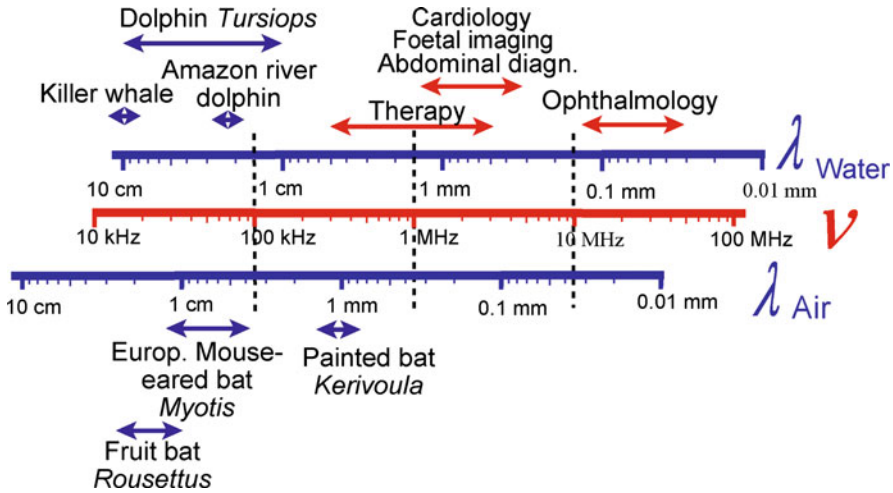


Fig. 4.17 Frequencies (ν), wavelengths in water (λ_{Water}) equivalent to that of soft tissue, and air (λ_{Air}) of ultrasound as well as regions of medical applications (red arrows) and animal sonar systems (blue arrows) (Data for animal frequencies from Withers 1992)

intensity, with respect to the square of its sound pressure (see Eq. 4.12), and to the absorption coefficient of the material, which increases with increasing frequency.

To characterize the temperature increase in the body resulting from ultrasound absorption, the term *diathermy* is used. This temperature increase takes place if the rate of heat production along the penetration of ultrasound is greater than the rate of heat removal. In general, the heat removal is governed by heat conduction. In a living body, the complex system of biorheological thermoregulation must be considered, including the acquisition of heat by blood circulation. This is governed by the bioheat equation of Harry H. Pennes, as discussed in Sect. 4.1 (Eq. 4.4).

Caused by diffraction, or reflection of ultrasound in the tissue, as well as by differences in blood supply, the amount of thermal effects can be heterogeneous. Hot spots may occur in the tissue, i.e., small locations of increased temperature (see a detailed discussion of possible sizes of them in Sect. 4.7.2).

The nonthermal mechanism of ultrasound interaction is mainly based on the effect of *cavitation*. Cavitation is the formation of gas-filled tiny bubbles as a result of alternating positive and negative pressures at particular spots in the medium. Homogeneous liquids have a considerable resistance to such an effect of disruption. In order to pull pure water apart in such a way that two parallel surfaces are formed, separated one from another, a negative pressure of about 1,500 MPa is required. However, only 150 MPa is sufficient to form a spherical cavity. Even lower local negative-pressure differences are required to induce cavitations in an inhomogeneous medium, such as a solution, suspension, or even a biological tissue.

These bubbles originate within materials at particular locations, so-called nucleation sites. In a complex medium such as a tissue, the exact nature of these nuclei of cavitation is not well understood. The occurrence of cavitation depends not only on the

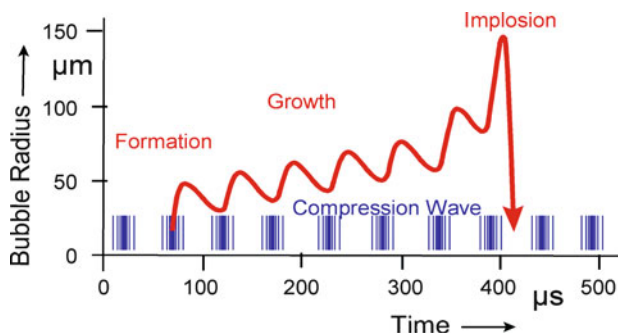


Fig. 4.18 Bubble growth and collapse during transient cavitation at a frequency of about 16 kHz (Redrawn from Suslick and Doktycz 1991)

concentration of these nuclei and on the intensity of ultrasound, but further more on some qualitative factors, such as frequency, field homogeneity (focused or unfocused ultrasound), and pulse parameters. For short pulses of ultrasound, cavitations in tissue have been observed at local pressure differences of above 10 MPa. To enhance the echogenicity of ultrasound in medical diagnostics, contrast agents are often used, indicating the preexistence of microbubbles. This, however, increases the possibility for ultrasound-induced cavitation and the potential for bioeffects. Cavitation can produce audible noise, leading finally to “boiling” of the liquid.

These cavities are filled by the gas which is dissolved in the liquid and by the vapor of the liquid itself. They may persist for several acoustic cycles. In this way, they expand to the sizes of their resonance radii, before collapsing violently during a single compression half cycle (Fig. 4.18). The sizes of these resonance radii depend on the medium and the frequency. So, for example, for argon bubbles in water, it amounts approximately to 4 μm at a frequency of 1 MHz and 80 μm at 50 kHz. The violent and rapid compression of the collapsing bubbles leads to an extremely large increase in the local temperature. The collapse of cavitation bubbles at 20 kHz is not completely adiabatic, and heat is lost from the bubbles to the surrounding liquid by means of conduction. At acoustic frequencies in the MHz region, however, the compression is so rapid that there is insufficient time for heat transfer to the liquid. The resulting temperature in the imploding cavity locally may increase to values of the order of several thousand degrees Kelvin, with a heating rate of about 10^{10} K/s and the corresponding pressure of hundreds of atmospheres.

These extreme conditions lead to the formation of hydroxyl radicals and hydrogen atoms. The formation of $\cdot\text{H}$ and $\cdot\text{OH}$ radicals as the primary products of sonolysis of water has been confirmed by electron spin resonance. These radicals either combine to form H_2 , H_2O_2 , and water or attack solute molecules, which are reduced or oxidized. In contrast to the radiolysis, where the primary products are hydrated electrons (see Sect. 4.9.2), in the course of sonolysis of water, these radicals are produced directly by thermal decomposition. In this respect, the definitions of “thermal” and “nonthermal” effects must be reconsidered. If an effect is called “thermal,” then it is usually concerned with the consequences of

diathermal heating of macroscopic regions, and not the microscopic processes of thermal decomposition.

Besides this thermal theory of sonolysis, the possibility of an electrochemical effect also must be considered. This theory recently had been supported by investigations of long-living microbubbles, the collapse of which is slower and does not show a dramatic temperature increase. The microbubbles, in fact, contain surface charges and, correspondingly, show a ζ -potential at the gas–water interface (for ζ -potential, see Sect. 2.3.5). This ζ -potential is inversely proportional to the bubble size and increases with respect to the rate of shrinkage. The rate of movement of electrolyte ions of the corresponding electric double layer in water is not sufficiently high to counteract this increasing rate. Consequently, it is likely that some excess ions become trapped at the gas–water interface, which leads to an extreme accumulation of ions during the final stage of the collapse process. This might trigger generation of radicals via dispersion of the elevated chemical potential that has accumulated around the interface.

In biological systems, cavitation therefore leads to two kinds of effects: On one hand, the cavitation simply generates mechanical destructions of cell organelles and various supramolecular structures, and on the other hand, sonochemical effects need to be considered, i.e., consequences of the formation of reactive radicals.

Ultrasound is widely used in biotechnology and medicine. The range of frequency for therapeutic treatment is between 0.7 and 3.3 MHz (see Fig. 4.17). In contrast to diagnostic use, the optimization in therapeutics is oriented to maximal penetration depths. This diathermy treatment of ultrasound is used with the same intensity as diathermy produced by high-frequency electromagnetic fields (see Sect. 4.7.2). It is mainly applied in the treatment of soft tissue injuries, the acceleration of wound healing, the resolution of edema, softening of scar tissue, bone injuries, and circulatory disorders. Raising the temperature by a few degrees may have a number of beneficial physiological effects, for example, an increase in blood circulation.

Recently, high-intensity focused ultrasound (HIFU) is used for *lithotripsy*, a noninvasive treatment of stones in the kidney and the gallbladder. It is possible to enhance the uptake of pharmacologically active drugs through the skin by application of ultrasound (*sonophoresis*). Parallel to the electroporation of cells (see Sect. 3.5.5), a temporary opening of the cell membrane is possible also by pulses of ultrasound (*sonoporation*). A number of experiments to stimulate healing of bone fracture are still not conclusively proven.

Although ultrasound, by definition, is a nonaudible sound, it was found that the human ear may sense tones even up to 40 kHz, although with very low sensibility. Nevertheless, it is argued whether the weighing scales, discussed in Sect. 4.3.2 (Fig. 4.7), must be extended up to this frequency.

To evaluate the intensity of the ultrasound radiation, usually the power density in W m^{-2} is used. This, however, does not account for the real energy absorption in the body or in the tissue. The local sound pressure, for example, would be more convenient, but this is a not directly measurable parameter. The only measurable reference value is the temperature elevation during exposition. This safety parameter depends on the time of exposure. From nonhuman experiments, for example, a

threshold for tissue damage was found: if, at a duration of 0.1 s, the temperature was increased up to 18°C, for 5 min, an increase of only 4°C is possible.

The American Institute of Ultrasound in Medicine (AIUM) periodically examines the literature on ultrasound bioeffects and develops conclusions and recommendations related to diagnostic ultrasound. In general, diagnostic ultrasound has been in use since the late 1950s, and this has indicated that no independently confirmed adverse effects caused by exposure from present diagnostic ultrasound instruments have been reported in human patients. There has been no experimental evidence to suggest that inertial cavitation events occur from exposure-like conditions employed with diagnostic ultrasound equipment. In some cases, localized pulmonary bleeding has been reported, but the clinical significance of such effects is not yet known. Nevertheless, the ALARA principle (“as low as reasonably achievable”) is recommended to obtain the necessary diagnostic information. The use of ALARA is an important consideration in all fetal examinations, and the dwell time may be overlooked as a means to reduce exposure.

WHO recommends a mean intensity limit of $3 \cdot 10^4 \text{ Wm}^{-2}$. The real values applied in diagnostic methods are far below this limit. Therefore, no risk is involved in the common ultrasound diagnostics for health, especially regarding replacement of X-ray methods by ultrasound.

Further Reading

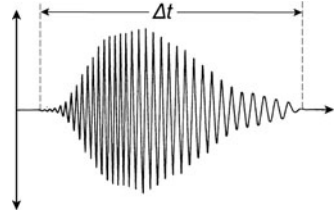
Primary mechanisms: Riesz and Kondo 1992; Takahashi et al. 2007; therapeutic applications: Frenkel 2010; ter Haar 2007; sonoporation: Ohta et al. 2008; safety recommendations of AIUM: Fowlkes 2008.

4.3.6 *Biophysics of Sonar Systems*

It has been known since the experiments of the Italian monk and scientist Lazzaro Spallanzani in 1794 that bats have the ability to orientate in complete darkness and to avoid obstacles. It had been demonstrated at that time that a bat could still fly safely when blinded but was quite helpless when its ears were plugged. Only in 1938 was D. R. Griffin able to show by direct measurements that bats are able to emit ultrasound and detect the echoes reflected from objects in their surroundings to determine their location. In the following years, many investigations had been carried out and it was found that not only bats but also rodents, aquatic mammals, birds, fish, and insects use echolocation by high-frequency sound, or ultrasound.

Because bats have been investigated in this respect more than other species, we will start with their example to explain the biophysical basis of echolocation. Figure 4.19 shows an oscillogram of a typical orientation signal of a bat. It is characterized by the following physical parameters: frequency (ν), amplitude, duration (Δt), and, of course, the distance between two signals.

Fig. 4.19 Example of a typical FM-modulated signal of a bat



The frequency of the emitted sound depends on the size of the object which is to be localized. An effective echo can be expected only from targets that are equal or even larger in size than the wavelength of the sound. The relation between wavelength (λ) and frequency (ν) is given by:

$$\lambda = \frac{v}{\nu} \quad (4.15)$$

where v is the velocity of sound waves. In air under atmospheric conditions, $v = 331 \text{ ms}^{-1}$. A sound of 25 kHz, therefore, has a wavelength $\lambda = 331/25,000 = 0.01324 \text{ m}$, i.e., 1.324 cm.

The frequencies emitted by bats differ from species to species (see Fig. 4.17). Usually, these amount to between 25 and 150 kHz. Bats that use their sonar system in order to catch insects emit signals of higher frequencies than those simply orientating in dark caves.

The amplitude of the signal, i.e., its intensity, is determined by the required range of the sonar system. Naturally, the auditory system of the bats must be sufficiently sensitive to receive the corresponding echo. The power of the emitted sound of the big brown bat (*Eptesicus fuscus*), for example, is about 10^{-4} W , whereas, as a minimum, a power of 10^{-16} – 10^{-14} W corresponds to the auditory threshold. At a distance of 5–10 cm from the head of this bat, the sound pressure is 2–6 Pa. According to Eq. 4.13, this amounts to an intensity of 100–110 dB. We learned in Sect. 4.3.2, Fig. 4.8, that if transformed into audible sound, this would be a considerable noise.

This large difference between the emitted and the accepted sound causes problems for the animal in listening while crying. The maximal duration of a signal (Δt), therefore, is more or less limited by the time the sound needs to arrive at the target and come back, i.e., by the distance of the target.

Frequency (ν), amplitude, and pulse duration (Δt), therefore, are the basic parameters of an echolocation signal. Looking at the sonogram in Fig. 4.19 carefully, however, more details can be noticed. The frequency, for example, as well as the amplitude of the signal is time dependent. It is a *frequency-modulated* sound, a so-called FM signal. In the case of the little brown bat (*Myotis lucifugus*), the signal starts with a frequency of 100 kHz and goes down by an octave at the end of the 2-ms-long-lasting signal. In contrast, the horseshoe bats use signals of 50–65 ms with constant frequencies (*CF signals*). Mostly, however, these CF signals have a

short final frequency-modulated part. The character of the emitted signals depends on the species, the situations, and even on the individual animal.

Frequency modulation seems to be an important property of the signal, especially for the measurement of distances. In some cases, it appears that by superimposition of the FM part at the end of the signal with the initial CF part of the returning echo, specific beat frequencies occur, which are evaluated by the animal. Furthermore, a principle known as the *pulse compression technique* in radar technology has been discussed. For this, short frequency components of the echo are delayed by a special acoustic filter, a so-called optimal filter, with respect to their frequencies. The higher frequency parts at the beginning of the signal are delayed to a larger extent than the lower frequency parts at the end of the signal. This process of impulse compression provides an instantaneous cross-correlation between the emitted and the received signals. Pulse depression explains the ability of the animal to measure the distance even in cases when the echo time is less than the duration of the emitted sound. Owing to this principle, disturbances by the Doppler effect are also minimized.

On one hand, the Doppler effect disturbs the bat when measuring distances, while on the other hand, it is used to measure relative velocities between the animal and its surroundings. For this, a Doppler-shift compensation occurs. If a flying bat approaches an object at a certain speed, the echo from that object will be Doppler-shifted upward to a higher frequency than the emitted signal. Subsequently, the bat lowers the frequency of the next emitted CF component by an amount that is nearly equivalent to the Doppler shift in the preceding echo. The amount of the required frequency is used by the bat to calculate its relative velocity.

Many questions on the information processing of the echolocation behavior in bats are still open. These concern mechanisms to determine the direction of the object as well as those regarding information about size, shape, and material of the target. Recently, it was found that a bat-pollinated rainforest vine lures the bats with specific disk-shaped leaves located near the ring of flowers which act as reflectors of ultrasound (Simon et al. 2011).

Echolocation is also seen in some birds. Alexander von Humboldt has described a South American cave-dwelling bird, the oilbird or guacharo (*Steatornis caripensis*), which flew around the head of the explorer uttering “screams of fear.” It is now known that these cries serve as signals in an echolocation mechanism. The sound is pitched at frequencies of 6–10 kHz, which means that it can be heard by humans. These birds are herbivores, but they require their sonar system in order to orient in large dark caves. The frequency employed corresponds to wavelengths between 3 and 5 cm, which means that the sound can be reflected even from small rock projections.

Sonar systems are found not only in airborne species but also in several aquatic animals. There are, however, vast differences between these two kinds of echolocation. Due to the differences in the velocity of sound, the wavelength of a sound in water ($v = 1,500 \text{ ms}^{-1}$) is about 4.5 times larger than that of the same frequency in air (see Fig. 4.17). This diminishes the resolution of the aquatic sonar system at the corresponding frequency. On the other hand, the distance of the applied echolocation differs. Bats use their sonar system at short ranges up to approximately 3–4 m, whereas dolphins can detect small targets at ranges varying from a few tens of

meters, for the harbor porpoise, to over a 100 m, for the bottlenose and other large dolphins. Dolphins, for example, emit short click signals with a duration mostly of only 40–50 ms, having a frequency up to 130 kHz, with a repetition after 20–40 ms.

The mechanisms of sound production between bats and whales are also different. While bats use the larynx as the traditional mammalian kind for sound generation, dolphins have usurped and completely revamped the nasal apparatus for producing sonar sounds. Both power their sound-generation system pneumatically and produce their signals by pushing air past tissue gates or valves. Bats use subglottal pressure, while dolphins use intranasal pressure. The mechanisms in dolphins may be similar to the behavior of a human trumpet player's lips. Some animals are able to focus the beam of high-frequency clicks. This is the case of toothed whales, in which it is done by reflection at a dense concave bone of the cranium and by a large fatty organ known as the *melon*. This is a large forepart of the head containing triglycerol and wax esters, i.e., lipids of differing densities. It acts like an acoustic lens, focusing the beam in the direction in which their head is pointing.

In the case of aquatic sonar systems, the specific sound distribution in water must be taken into consideration. The transmission loss for a point source of sound in free space in general is:

$$\text{Transmission loss} = 10 \log \frac{I_0}{I_r} = 10 \log r^2 + \alpha r \quad (4.16)$$

where I_0 is the intensity at $r = 0$, I_r is the intensity at distance r , and α is the coefficient of absorption.

The first term of the sum describes the simple geometrical distribution of the sound, whereas the second term concerns the absorption of sound in the medium. Especially in seawater, because of its content of magnesium sulfate, the coefficient α is frequency dependent (Fig. 4.20). This means that in contrast to airborne animals, the frequency of the emitted signal for seawater animals is additionally a function of the distance. This coefficient also shows a dependence on pressure and temperature. So, for example, the 20°C value of 38.6 dB/km for 100 kHz diminishes to 28.9 dB/km for 5°C. The values in Fig. 4.20 correspond to surface

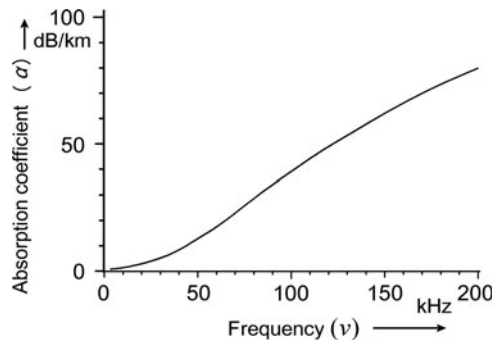


Fig. 4.20 The absorption coefficient (α) of seawater at 20°C as a function of frequency (ν) (Data from Ainslie and McCole 1998)

water. At 1,000-m depths, for example, the 100 kHz value for 5°C would shift from 28.9 to 24.8 dB/km.

Further Reading

Au 1994; Griffin 1958; Pollak und Casseda 1989; Thomas et al. 2004.

4.4 The Static Magnetic Field

The influence of magnetic fields on biological systems and human health is an ancient and contentious subject. It became topical with the introduction of clinical magnetic resonance imaging (MRI), where patients are exposed to an intense magnetic field of a strength not previously encountered by humans. On the other end of the scale (Fig. 4.21), the orientation of animals in the weak geomagnetic field is a permanent matter of discussion. In between these two regions of intensity, the field of magnetic healing quackery is positioned with its own history, starting with Anton Mesmer at the end of the eighteenth century.

The interaction of magnetic fields with biological materials is to be considered from two aspects: on the one hand a direct magnetic influence on molecules and supramolecular structures is possible, on the other hand, pulsed or alternating magnetic fields induce eddy-currents which finally may interact electrically. This leads to the topic of electromagnetic field effects on biological systems which will be discussed later (Sect. 4.6.1). Here, however, beside direct magnetic effects, we

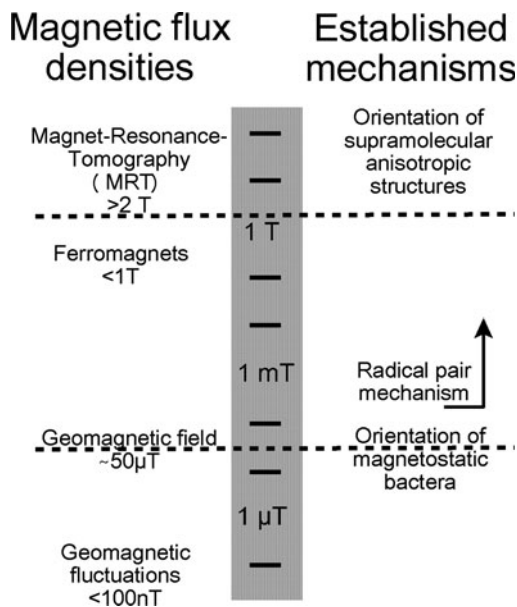


Fig. 4.21 The logarithmic scale of intensity of homogeneous static magnetic fields (magnetic flux intensity in Tesla), versus established mechanisms of possible biological interaction

Table 4.4 The magnetic susceptibility (χ) of air, water, and various biological materials (all: $\cdot 10^{-6}$)

	After Maniewski (1991)	After Khenia et al. in: Maret et al. (1986)
Air	+0.34	+0.264
Water	−9.05	−9.04
Arterial blood	−9.1	−9.3
Oxygenized erythrocytes		−9.03
Venous blood	−8.4	−7.8
Deoxygenized erythrocytes		+3.88
Lungs (breathed in)	−3.9	
Lungs (breathed out)	−4.1	
Lungs (15% air content)		−4.2
Muscle	−9.0	−9.03
Liver	−8.8	−8.26
Bone	−10	−10

must consider induction processes, generated in biological systems which are moving in a static magnetic field, for example in the case of blood flow, or animal locomotion.

Let's first explain some basic principles. A magnetic field with the *magnetic field strength* \mathbf{H} (in A m^{-1}) causes in a body a *magnetic induction*, or *magnetic flux density* \mathbf{B} (in Tesla: $\text{T} = \text{V s m}^{-2}$) of:

$$\mathbf{B} = \mu\mu_0\mathbf{H} \quad (4.17)$$

where $\mu_0 = 1.257 \cdot 10^{-6} \text{ V s A}^{-1} \text{ m}^{-1}$ is the *magnetic permeability of vacuum*, and μ is the *magnetic permeability* – a dimensionless number in relation to vacuum, where $\mu = 1$. The deviations of the magnetic permeability for nonferromagnetic materials from 1 are extremely small. For practical use therefore another parameter is defined: the *magnetic susceptibility* $\chi = \mu - 1$.

Magnetic properties of various materials are classified according to their properties as being:

$$\begin{aligned}
 \text{diamagnetic} : & \quad \mu < 1; \chi < 0 \\
 \text{paramagnetic} : & \quad \mu > 1; \chi > 0 \\
 \text{ferromagnetic} : & \quad \mu \gg 1; \chi \gg 0
 \end{aligned} \quad (4.18)$$

As shown in Table 4.4, cells and tissue in general have diamagnetic properties. An exception is deoxygenized erythrocytes, which are paramagnetic due to the property of the central Fe-atom in the hemoglobin molecule. If it binds oxygen, it changes from paramagnetic to a diamagnetic state.

In contrast to the strong effects of electric fields, very strong magnetic fields show only subtle and hard to observe effects. The reason for this is the different deviations of the dimensionless dielectric constant (ε) on the one hand (see Sect. 3.5.3), and the magnetic permeability (μ) on the other hand, of the materials, from

the value 1 of the vacuum. For typical, nonferromagnetic materials, the electric susceptibility ($\epsilon - 1$) is of the order of $10^5 - 10^6$ times larger than the magnetic susceptibility ($\chi = \mu - 1$).

This of course does not apply to ferromagnetic interactions, where the magnetic permeability (μ) may achieve the order of 10^4 . Ferromagnetic materials form domains that exhibit a long-range ordering phenomenon at the atomic level which causes the unpaired electron spins to line up parallel with each other. Ferromagnetic materials, mostly consisting of Fe, Co, or Ni, remain magnetized after the external field is removed.

For some biotechnological and medical applications nanoparticles are used with *superparamagnetic* properties. These particles are magnetized by an external magnetic field like ferromagnetic materials, but unlike them they do not retain any significant amount of magnetization in the absence of the field. When the external magnetic field is applied all the domains are oriented in the same direction. If the field is removed, then they are mixed by Brownian motion. In contrast to ferromagnetic particles, these superparamagnetic nanoparticles, therefore, do not aggregate in the absence of an external magnetic field. This is the particular property that makes them applicable for biotechnological use. Coupling these particles with drugs, monoclonal antibodies, particular cells, or including them in liposomes opens a large field of biotechnological and medical application. These particles can easily be separated in moderate magnetic fields (*magnetoseparation*, *immunomagnetic cell separation*). Furthermore, magnetic drug targeting employing superparamagnetic nanoparticles as carriers is a promising cancer treatment.

In biological tissues various kinds of iron-containing ferromagnetic particles have been found; some of them concentrated in particular regions, others dispersed at very low concentrations. In some cases these particles are just absorbed from polluted external medium or they are the product of destruction of iron-containing proteins like hemoglobin or ferritin. In some cases larger ferromagnetic inclusions were observed. It is under discussion whether they play a role in magnetic field reception.

In many kinds of bacteria, ferromagnetic particles exist, so-called *magnetites*, which are synthesized by a particular mechanism from chelate bound iron. These are cuboidal crystals of Fe_3O_4 (a mixture of FeO and Fe_2O_3) with a size of 50–200 nm. The largest are surrounded by a membrane forming *magnetosomes*. The density of these particles is 5.1 g cm^{-3} . Because of their extremely small size, they consist of only one single magnetic domain which means that they represent a single elementary magnet.

These particles are arranged in a chain, several micrometers long, which is located along the axis of motility of the bacterium. The magnetic momentum of a single magnetosome is not large enough to resist the Brownian movement in the geomagnetic field of about 50 μT . The chain, however, forms a magnetic dipole which is sufficiently strong to orient the whole bacterium in the geomagnetic field.

Magnetostatic bacteria are usually anaerobic inhabitants of the sediment. Cells with appropriate polarization swim along the geomagnetic vectors, which in regions away from the equator are directed into the anaerobic depths (see Figs. 4.22 and 4.23). This oriented movement is called *magnetotaxis*. Bacteria

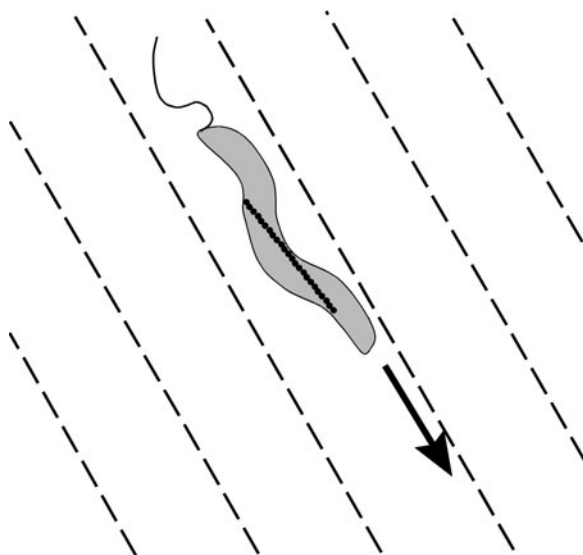


Fig. 4.22 Magnetospirillum moving down, oriented by its chain of magnetosomes along the geomagnetic field lines

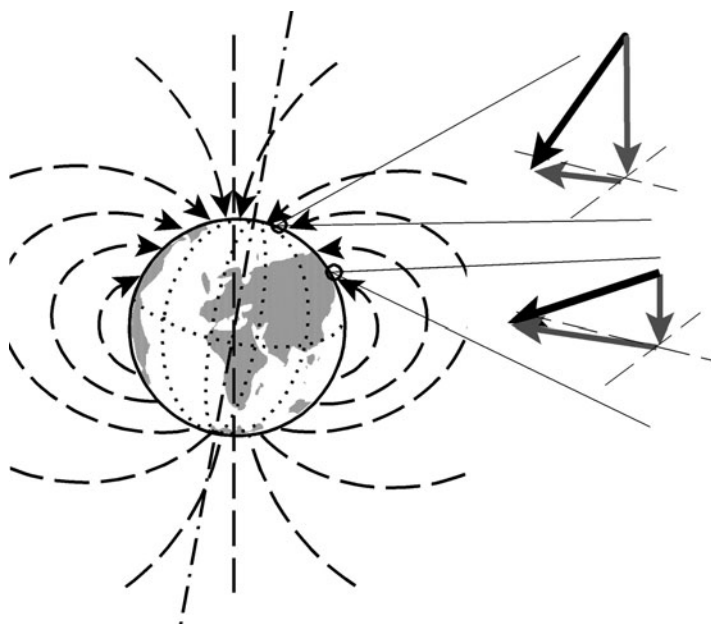


Fig. 4.23 Geomagnetic field lines, and their inclinations at two particular locations

with incorrect magnetic polarizations swim in the wrong direction and die in the oxygen-containing surface water. During division of the cells by mitosis, the magnetic chains will also be divided into two parts, and the daughter cells get a part of the correctly polarized chain. Such newly synthesized particles automatically become polarized according to the inherited part of the chain. It is possible to reverse the polarity of these chains in artificial magnetic fields of at least 10 mT.

The influence of magnetic fields on diamagnetic or paramagnetic molecules or structures requires much stronger field strengths. If during a chemical reaction a transformation occurs from diamagnetic to paramagnetic states or vice versa than according to Le Châtelier's principle (see Sect. 2.2.3) an influence of magnetic fields is possible (*magnetokinetics*). The magnetic contribution to the free enthalpy (ΔG) of reaction in a field with a magnetic flux density \mathbf{B} can be considered as a further summand in the Gibbs equation (see Sect. 3.1.2, Eq. 3.29):

$$\Delta G = \frac{1}{2} \Delta \chi_m B^2 \quad (4.19)$$

where $\Delta \chi_m$ is the change of magnetic susceptibility during the reaction of one molar unit. Even assuming a rather high value of $\Delta \chi_m$, even at $\mathbf{B} = 1$ T, the change of the equilibrium constant will be negligible and amounts only to a factor of about 10^{-5} . Thus, even in very strong magnetic fields these kinds of effects should be difficult to detect.

In contrast to this, another possible mechanism of magnetic effect on chemical reactions is well established theoretically and experimentally, which is more sensible to magnetic field strength. It bases on the fact that in some cases in the course of chemical reactions pairs of radicals occur as short-living intermediates that quickly recombine. In the same way after photoexcitation, a donor molecule D^* may transfer an electron to an acceptor molecule A, resulting also in a radical pair 1D and 1A . Usually, the unpaired electrons of these radicals will be in singlet states. Under the influence of a magnetic field, however, a singlet-triplet interconversion is possible. As the result, singlet-singlet and singlet-triplet reactions may occur to give distinct products, with respective rate constants k_S and k_T .

The sensibility of this *radical pair reaction* in respect of magnetic field intensity depends on the reaction time, i.e., on the probability that the singlet-triplet interconversion occurs faster than the recombination of the radicals. In fact, the energy of magnetic interaction involved in this radical pair process may be much smaller than the average thermal energy under normal temperature conditions because the spin of electrons bound to biomolecules is not coupled strongly to the thermal bath. Thermal motion, however, is important for the diffusion of the radicals from the region of generation. Clear effects of this mechanism are established at magnetic fields $\mathbf{B} > 1$ mT. The possibility of the occurrence of this effect at intensities of the geomagnetic field of 50 μ T has been extensively discussed, and requires some hypothetical conditions.

The orientation of objects is a further possible use of magnetic field effects. It is based on the *magnetic anisotropy* of some materials, a direction dependence of their magnetic properties. This anisotropy $\Delta\chi$ is defined as follows:

$$\Delta\chi = \chi_{\parallel} - \chi_{\perp} \quad (4.20)$$

whereas χ_{\parallel} and χ_{\perp} are the corresponding susceptibilities parallel and perpendicular to a characteristic direction of the structure. For biological structures the magnitude of $\Delta\chi$ can be of the order of 1–10% of the mean susceptibility. Macromolecules with differing susceptibilities in different directions will experience a torque that will attempt to minimize the magnetic energy. This orienting torque is proportional to \mathbf{B}^2 . For single molecules even in strong magnetic fields these torques are, however, too small with respect to the thermal energy. However, in supramolecular structures, such as membranes or protein filaments the alignment torque is proportional to the total number of molecules in the structure and to the volume.

This kind of orientation effect has been demonstrated in red blood cells in the case of sickle cell anemia in vitro already at 0.5 T. In this case the anisotropy is amplified by abnormal aggregates of oriented hemoglobin molecules. At a field strength of 8 T it was possible to orient the cleavage planes of the developing *Xenopus* embryo. This distortion of the third cleavage furrow of the developing egg, however, did not affect the postcleavage development.

In the case of gradients of magnetic fields, paramagnetic materials experience an attraction, and vice versa diamagnetic materials repulsion. In contrast to ferromagnetic effects, as noted earlier, these forces however, are very weak. It has been evaluated for example, that for a large diamagnetic protein even in strong field gradients, this force is about 10^8 times lower than the force of thermal movement. Again, only large organized supramolecular structures can be moved in extremely strong field gradients. So, for example, in strong field gradients *magnetophoretic movement* of paramagnetic deoxygenized erythrocytes occurs (see Table 4.4). Diamagnetic repulsion could be demonstrated for macroscopic objects. So, for example, a living frog with a magnetic susceptibility of the order of 10^{-5} could be levitated near a strong magnet of 16 T in a field gradient of $-1,400.9 \text{ T}^2 \text{ m}^{-1}$. Interestingly, no harm to the frog resulted from this exposure.

A highly controversial problem is whether, and how animals can sense the geomagnetic field. As demonstrated in Fig. 4.23, the vector of this field can be subdivided into a vertical (*inclination*), and a horizontal (*declination*) component. The inclination is at its maximum at the poles, achieving a magnetic flux density of 60–70 μT . The *declination*, has its maximum near the equator, with 34–40 μT . The intensity of the geomagnetic field shows various variations, which are caused chiefly by changes in sunspot activity. Furthermore, small variations during the day, during the month, and during the year occur. These astronomically caused fluctuations, however, amount to less than a tenth of a percent of the absolute value. Throughout history the magnetic pole of the earth has shifted a considerable

distance. Significant inhomogeneities of the geomagnetic field furthermore, are caused by geological factors.

Despite research over many decades and hundreds of papers, our knowledge of the reception of the geomagnetic field in animals is still controversial and fragmentary. We only have hints derived from behavioral experiments that show effects of magnetic orientation in nearly all classes of the animal kingdom. With the exception of magnetobacteria, where this effect is well established and the mechanism has been clarified, experiments with mollusks, insects and various other invertebrates, with fish, birds, and mammals suggest the existence of a magnetic sense without any reliable indications as to the responsible organs, before even beginning to think of reasonable biophysical mechanisms. Moreover, the results of these papers show a large diversity of phenomena suggesting that probably quite different mechanisms in different kinds of animals should exist. Some of them, like birds, do not estimate the declination, but rather the inclination of the field, and additionally require for this particular frequencies of light. Others, like turtles or subterranean rodents respond to declination without the necessity for light. In some cases, like in the movement of salamanders, light of different frequencies has been determined to change the angle of the field-orientated movement. Some authors even suspect that animals, like turtles or pigs, are able to achieve magnetic localization, i.e., to determine their position from the declination of the field and the local field gradient.

Unfortunately, in no case have reliable physiological or biophysical experiments led to a clarification of this issue. Nevertheless, a number of mechanisms have been proposed to explain these reactions. In the case of large quickly swimming fish like sharks the possibility of induction phenomena has been discussed. For birds the hypothesis has been formulated that the retinal rhodopsin or cryptochrome could be responsible using the radical pair mechanism. Other authors suggest magnetomechanical reactions caused by magnetosomes. Despite the lack of corresponding *in vitro* experiments, all of these mechanisms from a physical point of view are unrealistic in response to the required sensibility. Moreover, not only the high sensitivity of this hypothetical organ for the magnetic field strength must be explained by these mechanisms, but additionally the exact direction of the field lines must be known. In addition, the hypotheses of magnetic localization, or speculation about sensing of astronomically caused variations of the magnetic field by various animals lack any physical fundament.

We will now discuss the problem of induction processes for the strong magnetic field of MRI. During these diagnostic examinations, the patient is placed in a magnetic field with a flux density of more than 2 T. Even in a static magnetic field, it is necessary to consider the effects of patient motion within the field including the motion of the blood and other tissues within the patient. These kinds of induction processes in fact may provide the ultimate limitation in regard to the maximum magnetic strengths that can be tolerated by human beings. The largest magnetically induced current is associated with pulsating blood flow into the aorta during each cardiac cycle. Calculations have shown that at a field of 45 T the blood flow in the aorta will induce a current density of the order of 1.5 A m^{-2}

at the cardiac pacemaker site in the sinoatrial node, a 50% response level for ventricular fibrillation. Even at flux densities of about 1 T it is possible to measure these inductions in small deviations of the ECG.

During recent decades the method of *transcranial magnetic stimulation* has become available. In this case, stimulations occur by eddy currents, induced by magnetic millisecond pulses of an intensity of 1.5–2 T. In this way the primary motor cortex and the motor roots in conscious patients can be activated. This technique proved sensitive enough to illustrate early abnormalities of central motor conduction in various neurological diseases such as multiple sclerosis, amyotrophic lateral sclerosis, cervical spondylotic myelopathy, degenerative ataxias, or hereditary spastic paraplegias. At magnetic field pulses >10 mT an activation of the optical system is possible which results in so-called *magnetophosphenes*, i.e., visual sensations.

On the one hand eddy currents are induced in bodies moving in magnetic fields, on the other hand, currents in the body (see Sect. 3.5.2) induce magnetic fields themselves. To detect these fields sensitive measuring instruments using superconducting materials (SQUIDs = Superconductive Quantum Interference Devices) are used. To record magneto-cardiograms, -encephalograms, or -myograms, magnetic fields weaker than 1 pT must be measured. This is seven orders of magnitude lower than the geomagnetic field (not included in Fig. 4.21)! These methods are more costly than the corresponding ECG or EEG records, but they provide are much more instructive. As we already pointed out, the magnetic permeability (μ) of various materials does not deviate very much from $\mu = 1$. This means that the magnetic field in the body will not become significantly disordered, in contrast to the electric field. Therefore, it is much easier to localize the oscillating electric dipole in nerves and muscles by its induced magnetic field (see also the explanation of the ECG in Sect. 3.5.2, Fig. 3.38).

The application of highly sensitive measuring instruments for magnetic fields can also be used in occupational medicine to analyze the accumulation of iron dust in the lungs. Furthermore, the activity of the ciliated epithelium of the lung can be checked. For this purpose, a small amount of ferromagnetic labels are inhaled, and the rate of reorganization of these particles is recorded magnetically.

In conclusion: no dramatic influences of static magnetic fields are known, at least at flux densities $B < 10$ T. This is underscored by a large number of experiments with cells in vitro as well as with animals.

On the basis of established results on the influence of static magnetic fields, safety standards have been recommended by the International Commission on Non-Ionizing Radiation Protection (ICNIRP). According to this direct exposure of the general public to static fields should not exceed 400 mT. Because of potential indirect effects, for people with implanted electronic medical devices and implants containing ferromagnetic materials, a lower restriction level of 0.5 mT is recommended. Concerning MRI-diagnostics, a short-lasting exposition of humans without ferromagnetic implants and without cardiac pacemakers up to 8 T can be permitted if the environment is controlled, and if appropriate work practices are implemented to control movement-induced effects.

Further Reading

Influence of magnetic fields in general: Barnes and Greenbaum 2006; Glaser 2008; Miyakoshi 2005; Schenck 2005; magnetic applications in diagnostics and therapy: Andrä and Nowak 2006; magnetic orientation of animals: Rozhok 2008; Wiltshko and Wiltshko 1995; radical-pair reaction: Solov'yov and Schulten 2009; McLauchlan and Steiner 1991; magnetosomes: Bazylnski and Frankel 2004; magnetophoresis: Zborowski et al. 2003; immunomagnetic separation: Moore et al. 1998; Olsvik et al. 1994; magnetic stimulation: George et al. 1998; deNoordhout 1998; Wassermann 1998; Guidelines: ICNIRP 2009.

4.5 The Electrostatic Field

Aspects of bioelectricity have already been discussed in context with various other circumstances in previous sections of this book: After some basic parameters, and laws of electrostatics (Sect. 2.2.1), the principles of membrane electrostatics (Sect. 2.3.6) were explained. Subsequently, we came to transmembrane potentials (Sects. 3.4.3, 3.4.4) and to discussions of electric fields in cells and organisms (Sects. 3.5.1, 3.5.2). Finally, active, and passive electric properties of cells and tissues were considered (Sect. 3.5), as well as various techniques of cell manipulation using electric fields (Sect. 3.5.5). Now the question arises: how do electric fields in the environment influence cells and organisms? In this aspect one must distinguish between life in low conductivity air, and life in water or aqueous mediums with more or less high electrical conductivity.

In this section the effects of static, resp. DC-fields will be considered. However, it must be pointed out that a strong division between electrostatic fields and oscillating electromagnetic fields is artificial. In fact, there is a continuous transition, starting with static fields changing for example in a circadian rhythm or in periods of hours or minutes, up to extremely low-frequency electromagnetic fields (ELF-fields). The same holds for static fields switched on and off after some time, and pulsed forms. Using Fourier analysis, every kind of pulse can be considered as a sum of sine fields of various frequencies. Some aspects discussed in this chapter therefore, are important also for low-frequency electromagnetic fields.

Let us start with the conditions of organisms in air. Terrestrial animals live in the electrostatic field of the earth which is caused by astrophysical, as well as meteorological processes. Our globe can be considered as a negatively charged sphere. This results in a field strength near the earth's surface at sea level of approximately $100\text{--}200\text{ V m}^{-1}$. For astrophysical reasons, the electric field of the earth changes during day and night and also through the year. Additionally, various meteorological processes lead to charge separations, for example in between clouds, as well as between clouds and the earth's surface. Below a thundercloud the electric field strength on the surface can rise up to $20,000\text{ V m}^{-1}$.

These electrostatic fields, as well as those in the vicinity of various technical devices, are not very important from the point of view of health protection. They could, however, lead to a considerable charge of objects acting as capacitors. This, for example, includes car bodies or other isolated metal surfaces. Touching such electrically charged surfaces can produce painful discharges.

The specific conductivity (g) of dry air near the earth's surface amounts to about $2.5 \cdot 10^{-14} \text{ S m}^{-1}$. According to Ohm's law ($j = g \cdot E$), at a field strength of $E = 200 \text{ V m}^{-1}$ a current density of $j = 5 \cdot 10^{-12} \text{ A m}^{-2}$ results. This conductivity depends on the presence of charged particles. These particles have the misleading name *air ions*. In fact, these are particles, charged with both polarities, which have quite heterogeneous nature and size, beginning with true ions of nitrogen or sulfate compounds, up to various microscopic dust particles containing surface charges. The concentration of such "ions" in air at high altitudes amounts to approximately 10^9 ions per m^3 . This varies across a broad range depending on various meteorological situations. It increases near cascades or breakers. To measure this parameter, so-called *Gerdien tubes* are used. These are coaxially arranged capacitors, and their discharge over time is recorded during a particular air flow.

There has been much discussion of a possible biological role of these charges. A number of attempts have been made to construct various artificial sources of air ions for therapeutic reasons. *Electro-aerosol therapy* (EAT) was proposed to cure various bronchial afflictions. However, in fact no correlation with any kind of disorder, and no true indications of its therapeutic efficiency have been established yet.

The biological influence of environment fields in air depends on their penetration into the body. The electric conductivity of the human body is about 10^{14} times larger than that of air. The dielectric constant exceeds that of air by a factor of 50. This leads to considerable distortion of the field lines around the body (Fig. 4.24). The electric field strength near the surface of the body can significantly increase, especially at the head or at the end of a raised hand.

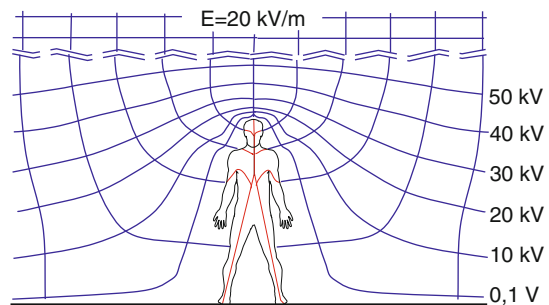


Fig. 4.24 Distortion of an electrostatic field near a standing man (After Leitgeb 1990 redrawn)

The external electric field leads to a displacement of internal charges in the body. An external electrical field \mathbf{E}_e in air ($\varepsilon_e = 1$) induces in a spherical, nonconducting body ($g_i = 0$) the following internal field \mathbf{E}_i :

$$\mathbf{E}_i = \frac{3\mathbf{E}_e}{2 + \varepsilon_i} \quad \text{for: } g_i = 0 \quad (4.21)$$

If the body is conducting ($g_i > 0$), this internal field immediately will be neutralized by an internal current. Taking this process into account, one gets the real field strength \mathbf{E}_i as a function of time (t):

$$\mathbf{E}_i = \frac{3\mathbf{E}_e}{2 + \varepsilon_i} e^{-\frac{g_i t}{\varepsilon_i \varepsilon_0}} = \frac{3\mathbf{E}_e}{2 + \varepsilon_i} e^{-kt} \quad \text{where: } k \equiv \frac{g_i}{\varepsilon_i \varepsilon_0} \quad (4.22)$$

Using appropriate parameters ($g_i = 0.6 \text{ S m}^{-1}$, $\varepsilon_i = 50$, $\varepsilon_0 = 8.84 \cdot 10^{-12} \text{ C V}^{-1} \text{ m}^{-1}$), a corresponding rate constant $k = 1.35 \cdot 10^9 \text{ s}^{-1}$ can be obtained. This means that an internal electric field which is induced by a single rectangular pulse will vanish by a half life time of $(\ln 2)/(1.35 \cdot 10^9) = 5.13 \cdot 10^{-10} \text{ s}$. A significant field in the body therefore, can be induced only in the case of higher frequency AC, or frequently pulsed fields. We will come back to this point in Sect. 4.6.

This consideration indicates that an electrostatic field in the air environment may influence only the surface of the body. This concerns, for example, electrostatic erection of hairs. In the case of low-frequency AC fields, a vibration of hairs can be sensed. The threshold where humans sense electrostatic fields in the environment is near 10 kV m^{-1} .

In contrast to the conditions in air, the field interaction in aqueous media between organisms and the environment is much stronger, leading to the electric field becoming an important element in the behavior of some aquatic organisms. In this respect, animals that live in water developed electroreceptors with an enormous sensibility. Behavioral experiments suggest that the lowest field perceived by freshwater fish is of the order of 0.1 V m^{-1} , and $2 \mu\text{V m}^{-1}$ by marine fish. These sense organs are used for active, as well as for passive electrolocation. "Active" means the measurement of distortion of the electric field pulses which are generated by the fish itself, caused by dielectric bodies in the neighborhood. This is a mechanism of orientation in weak electric fish (see Sect. 3.5.2). "Passive" electrolocation denotes the reception of electric fields from extrinsic sources and is used, for example, for the localization of living prey, or for communication with other individuals in a shoal. Furthermore, it is possible that some fishes are able to sense inorganic sources of electric fields, like electric potentials generated by streaming water, or even eddy currents induced in the body of large, fast-swimming fish by the geomagnetic field.

There is a large diversity in organs of electroreception in various aquatic animals. In general, two types of organs have been found (Fig. 4.25). *Ampullar organs*, which are cavities more or less deeply embedded in the body, connected with the surrounding water by a canal, filled with a highly conducting mucous jelly. The cavities contain the sensory epithelium. The wall of the canal is composed of

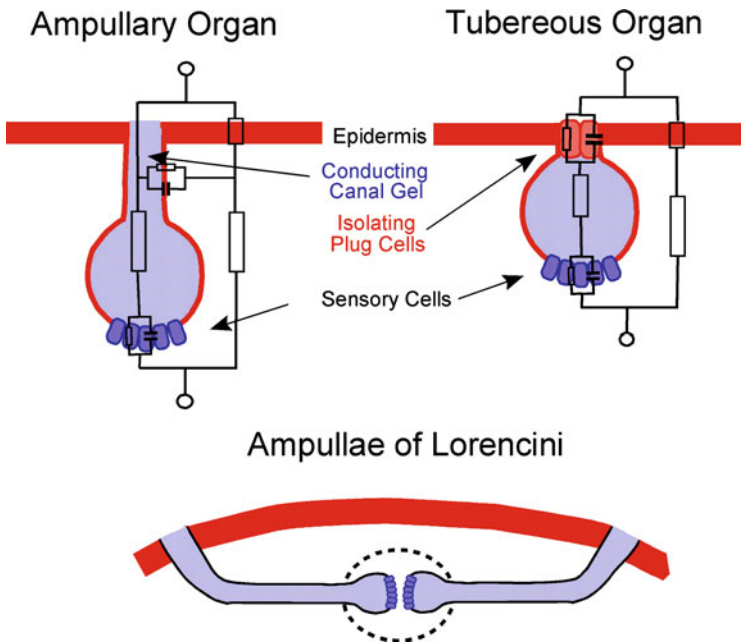


Fig. 4.25 Schematic illustration of electroreceptors in the skin of fishes. *Ampullary organs* are coupled by conducting gel to the surrounding water, whereas *tuberculous organs* are connected just capacitively to the environment. The *ampullae of Lorenzini* are centimeter-long canals, open to the surface at distant locations via a pore in the skin, and ending with many sense organs in a common capsule

flattened cells, connected by tight junctions and therefore indicating a high electrical resistance. *Tuberculous organs* are similarly constructed organs, but without an electrically conducting connection to the surrounding water. In this case the pore is covered by a plug of epidermal cells. These tuberculous organs therefore are not able to measure static potential differences, but as they are coupled capacitively to the environment, they therefore respond to time-dependent voltage changes. Tuberculous organs have so far been demonstrated only in teleost fishes.

In the same way as for field-generating electrical organs (Sect. 3.5.2), also in the case of electroreceptors the different conductivity of seawater and freshwater is very important. Furthermore, the high skin resistances in freshwater fishes effectively exclude external fields, and the receptors only need to measure the potential drop across the skin, in relation to the nearly isopotential body interior. In marine fishes external fields can pervade the internal tissues. In this case centimeter-long canals are required to permit sampling of a significant portion of shallow uniform voltage gradients (Fig. 4.25). Such structures were first described by Stefano Lorenzini in 1678, and later called *ampullae of Lorenzini*. Dielectric measurements of the glycoprotein gel in these canals show that they are too sluggish for frequencies above 1 kHz, and therefore may act as antennas for extremely low frequencies. These long canals permit the clustering of the ampullae from many different receptors, with skin pores

widely distributed across the body surface. Recently it has been found that the glycoprotein gel in the ampullae of Lorenzini has electrical characteristics similar to those of semiconductors, exhibiting strong thermoelectric effects. This therefore also qualifies them for temperature sensibility.

In fact, the electroreception in fish emerged as a highly sophisticated system which is still far from being decoded. On the one hand the unbelievable sensitivity, at high noise–signal ratio must be explained, while on the other hand its ability for detailed analyses of signal information must be elucidated. Some of these phenomena can be explained as the result of averaging the response of a large number of sense organs. Additionally, mechanisms of stochastic resonance seem to optimize the noise–signal relation (see Sect. 3.1.5). The different properties of ampullar and tuberous organs allow particular responses to DC and AC fields. Furthermore, it has been found that the tuberous organs contain P (probability)- and T (time)-receptors; the first respond to amplitude characteristics, the second to the time of zero-point crossing of the potential. This complex net of different primary sensors is part of a sophisticated system of neurological data processing.

In context with biological influences of electric fields in aqueous media galvanotaxis and galvanotropism must be mentioned. *Galvanotaxis* represents the movement of cells directed by a static electric field with field strengths of about $100\text{--}500\text{ V m}^{-1}$. These kinds of orientation have been observed not only in various protozoans, sperms, and zoospores, but also in many other kinds of cells, moving slowly on the substratum, like granulocytes, fibroblasts, osteoblasts, and nerve cells, etc. This effect must not be confused with electrophoresis, which means a passive transportation of charged particles in an electric field (see Sect. 2.3.5). In contrast to electrophoresis, galvanotaxis predicts an active movement of the cell. In the case of galvanotaxis, the electric field does not act as a driving force, but just as a parameter for orientation of the actual movement. In contrast to electrophoresis, galvanotaxis, being an active biological process, needs a much longer lag phase. Galvanotactically moving cells may still drive in one direction for many seconds or even minutes, even if the polarity of the applied field has already been changed. Most cells, even if negatively charged, move by galvanotaxis toward the cathode, i.e., in the direction opposite to that electrophoresis would move them.

The mechanism leading to galvanotaxis of tissue cells is rather unclear and probably not consistent for all kinds of cells. Whereas in ciliata the orientation of their motion is apparently controlled by the modified transmembrane potential, for other cells other mechanisms are proposed. Using fluorescent labels it is possible to indicate that membrane-bound receptors shift under the influence of an external field laterally in the cell membrane, and concentrate on one side of the cell. This does not seem to be the result of individual electrophoresis of these molecules but rather a reaction to local electro-osmotic flow, induced by the interaction of the external field with all charges of the membrane surface (for electro-osmosis, see Sect. 2.3.5, Fig. 2.45).

A further property of cells in electrostatic fields is *galvanotropism*. This means the influence of an external DC-field on the direction of cell growth. So, for example, nerve cells in electric fields of $0.1\text{--}1\text{ kV m}^{-1}$ form dendrites, preferentially

at the side of the cell which is directed toward the cathode. In this case, local modifications of the membrane potential are probably responsible for these reactions, but the above-mentioned lateral translation of membrane proteins can also be responsible for this. This property has been applied in therapy to stimulate nerve regeneration.

We already pointed out that there are various biogenic electrical fields in the human body which probably are not simple passive concomitant phenomena of some electrophysiological processes, but which may have functional relevance. In this context we already mentioned wound potentials, which can create local electric fields of hundreds of V m^{-1} , or electric fields in loaded bones (Sect. 3.5.2, Fig. 3.36). This formed the basis for various attempts at using electrotherapy to control biological processes like cell motion, cell differentiation, or cell growth, etc., through the application of applied electric fields.

Further Reading

Electric fields in the environment: Barnes and Greenbaum 2006; Glaser 2008; Reiter 1992; air ions: Charry and Kavet 1987; electroreception: Brown 2003; Bullock et al. 2005; Fortune 2006; Kalmijn et al. 2002; Peters et al. 2007; galvanotaxis: Mycielska and Djamgoz 2004; Ogawa et al. 2006; wound healing: Zhao et al. 2006.

4.6 Low-Frequency Electromagnetic Fields

In this section we begin consideration of the biophysical aspects of electromagnetic fields starting with the low-frequency range (ELF – “extremely low,” ULF – “ultra low,” VLF – “very low,” LF – “low” frequency). The classification of the frequency regions is arbitrary and defined by engineers just for technical use (see Fig. 4.26). By chance, however, this division between “low” (L), and “high” (H) frequency in a positive way also matches differences in the dielectric behavior of living cells. As explained in Sect. 3.5.3, the cell, surrounded by a membrane, can be described as an electrical RC-circuit (Fig. 3.43), whereas at frequencies above a level of 10^5 – 10^6 Hz the membrane will be capacitively short-circuited. This means that biophysically the “low” frequency can be defined as the frequency region where the externally applied fields more or less are able to influence the membrane potential of living cells.

4.6.1 Physical Background and Dosimetry

In the low-frequency region particular physical properties of electromagnetic fields must be considered. According to Maxwell’s equations, electromagnetic fields have a magnetic and an electric vector. With increasing frequencies, it is more and more difficult to separate experimentally the magnetic from the electric component. At ELF

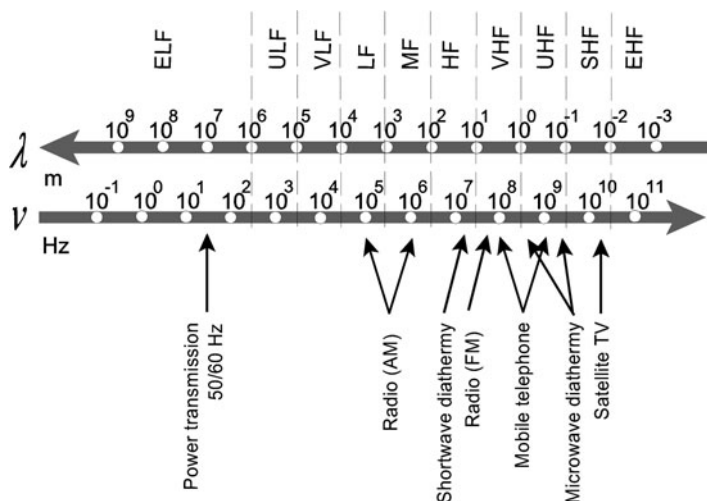


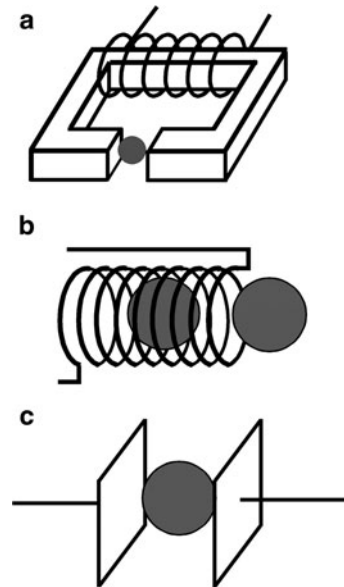
Fig. 4.26 The spectrum of technically used frequencies of electromagnetic fields (HF = high frequency, LF = low frequency, E = extremely, S = super, U = ultra, V = very, M = mean)

and ULF fields however, it depends on the techniques of field application, whether mostly an electric, or preferentially a magnetic interaction will occur. To illustrate this, three methods are shown in Fig. 4.27 to apply low frequency fields as examples. If, for example, the object is positioned between the two poles of an iron core, it is influenced mainly by the magnetic component of the field, induced in the core by the current in the coil (Fig. 4.27a). If the object is between two plates of a capacitor in air, not being in touch with them, almost only electric influences occur (Fig. 4.27c). In the same way, alternating current (AC) can be applied by electrodes connected directly with the biological object. Using coils (Fig. 4.27b), the magnetic, as well as the electric component of the field interacts with the object. There is no difference, whether the object is positioned inside the coil or in its vicinity. This situation resembles the general kind of exposition of humans near technical devices.

If humans are exposed to technical facilities, additional circumstances have to be considered which may lead to a dominance of the magnetic or the electric vector of the field. The electric field of a particular position near a power line, for example, depends on the construction of the line, i.e., on the voltage of the transported energy, which is constant for a given construction. The magnetic field, emitted by this line however, varies depending on the current flowing through it, which is always changing during the day. Furthermore, usually several wires are combined in one line, carrying different currents with phase differences. The magnetic field of a power line therefore can vary considerably. Similar aspects must also be taken into account in cases of other electric devices.

Furthermore, it must be considered that electric fields are shielded by buildings, trees, etc. Magnetic fields, in contrast to this penetrate the walls of houses and

Fig. 4.27 Examples of three possibilities to apply low-frequency electromagnetic fields: (a) Iron core techniques with dominating magnetic field influences; (b) influences of magnetic and electric component; (c) capacitive arrangement with dominating electric field component



diminish only with increasing distance from the source. This leads to the dominance of the magnetic field in buildings and at places shielded in other ways.

Before discussing possible biological effects of electromagnetic fields, its penetration into the body must be explained, i.e., the particular conditions near the real site of interaction. From the physical point of view, low-frequency electromagnetic fields can generate electric current in the human body in two ways: by *influence* and by *induction*. The influence has already been explained in the previous Sect. 4.5 (Fig. 4.24). It is governed by Eq. 4.22, whereas the rate constant (k) of the neutralization phenomenon is compensated for by the frequency of electric field variation. Therefore, the current density increases with the frequency.

The second way to generate currents and electric fields in the body, induction, profits from the fact that the magnetic component freely penetrates the body. In this way it induces eddy currents in the tissues. The current density produced in this way depends on the frequency of the field, on its intensity, and on the geometrical parameters of the induction loop. The latter is determined by the size or diameter of the body in a projection plane perpendicular to the magnetic field vector, but also by anatomical factors and the impedance of the corresponding tissue.

Differences in the conductivity of the organs in the body can lead to preferential conductivity loops. Such loops may occur via special arrangements of blood vessels or even between electrically communicating cells. The cell itself as a space of high conductivity, surrounded by an isolating membrane, can act as a conducting loop, whose diameter however, is very small. All these circumstances in fact complicate the calculation of the real field conditions, and the local current densities in the body even if the environmental field parameters are known.

In the case of an undisturbed electromagnetic field, for example in a human under, or near a high-voltage power line, the currents generated in the body by processes of influence or of induction are quantitatively similar to each other. Qualitatively, however, the directions of these currents are different, and therefore also the appearance of current densities. Various model calculations and measurements on equivalent phantom objects are used to evaluate these parameters.

To assess safety standards, two types of parameters are used: *basic restrictions*, which refer to the current density at the site of the biological interaction in the body, and *reference levels* describing the corresponding electric or magnetic field strength in the environment which probably will induce these conditions. In contrast to the parameters of basic restrictions, these reference levels are easily measurable, and therefore used in safety control. They are evaluated from the basic restrictions by the above-mentioned models. If a measured value in the human environment exceeds the reference level, it does not necessarily follow that the basic restriction will also be exceeded. However, in such cases, model calculations are necessary to evaluate the actual reference level.

Besides technically induced fields, humans are also exposed to natural electromagnetic fields in the environment. These fields have wide-spread frequencies, starting with weak variations of the electrostatic field of the earth, as discussed in Sect. 4.5, in a region far below 1 Hz. Furthermore, pulses induced by lightning produce frequencies above 10 Hz. These world-wide discharges induce permanent low-frequency electromagnetic oscillation in the atmosphere, which can be measured even in clear weather conditions. In this case the so-called *Schumann resonance* is important, an electrical resonance frequency of the globe of 7.8 Hz with corresponding harmonics at 13.8, 19.7, 25.7, and 31.7 Hz. At higher frequencies, so-called *atmospherics* (or *spherics*) can be measured. These are pulses induced by various discharges in the atmosphere, containing electromagnetic frequencies up to tenths of kHz. The intensity of these *atmospherics* decreases with increasing frequency, and depends on various atmospheric conditions. Their magnetic flux density does not exceed 20 nT. It is speculated that the weather sensitivity of humans could be based on this influence, but this hypothesis has no scientific support.

Further Reading

Dosimetry and safety standards: ICNIRP 1998; Barnes and Greenbaum 2006; natural fields and spherics: Glaser 2008; Reiter 1992; Schienle et al. 1997.

4.6.2 Biological Effects and Biophysical Background

The question as to how electromagnetic fields may affect biological systems is of particular interest for problems of health protection, and also it opens new possibilities of therapeutic use. Basically, the quantum energy of electromagnetic fields in the frequency region below UV-light is insufficient to break chemical

bonds (see Sect. 4.8, Fig. 4.32). Each kind of absorbed energy is finally transformed into heat. Therefore, at higher field intensities, thermal effects occur, but the real threshold at which low frequency fields affect various membrane processes, like transport and reaction rates, and finally cause excitations of corresponding cells, is much lower.

As already mentioned in the previous section and explained in detail in Sects. 3.5.3 and 3.5.4, low-frequency electromagnetic fields and field pulses modify the transmembrane potential proportional to the external field strength, and to the diameter of the cell in the direction of the field vectors (Eq. 3.219). As shown in Fig. 3.45, caused by its low conductivity, the external field increases in the membrane. This can be considered as a considerable effect of field amplification in the membrane (cf. Fig. 3.49).

Assuming for example a cell with an elongation in the field direction of $10\ \mu\text{m} = 10^{-5}\ \text{m}$, and a membrane thickness of $10\ \text{nm} = 10^{-8}\ \text{m}$, then the field strength in the membrane, in relation to a low-frequency external field, will be enlarged by a factor of 10^3 . In fact the effective extension of the cell in the direction of the electric field vector can be much larger. For neurons elongations in field direction of up to $1\ \text{mm}$ are assumed. Furthermore, the word “effective” means that not only the length of the single cell, but also the extension of the conductive area surrounded by the isolating membrane must be considered. If two cells, for example, are connected by gap junctions with a high electric conductivity, then the “effective” length means the extension of this whole system.

The real effect of low-frequency electromagnetic fields, i.e., their interaction with essential molecules, therefore occurs preferentially in the membrane. As already explained in Sects. 3.4 and 3.5 there are various membrane properties which are governed by the internal field strength. This of course concerns predominantly the excitable membranes of nerves and muscles.

Experimentally, neuronal sensitivity has been detected for field strengths in the tissue as low as $0.1\ \text{V m}^{-1}$. To explain this high sensibility, in addition to the above-mentioned amplification effect of the field in the membrane, other circumstances must also be considered. Firstly, it is assumed that the field-sensitive channels, mainly Ca-channels, are positioned at both ends of these long cells. Furthermore, these channels as a kind of electric rectifier can accumulate the gating charges during several cycles of alternating field. Finally, the mean response of a large number of field-sensitive channels in a single cell is essential, as well as processes of stochastic resonance (see Sect. 3.1.5).

Figure 4.29 shows simplified functions of threshold values for various effects based on experimental data. The minimal range of observed biological effects of $0.01\ \text{A m}^{-2}$ corresponds to the above-mentioned field strength of $0.1\ \text{V m}^{-1}$, taking into account a conductivity of tissue of $0.1\ \text{S m}^{-1}$ (see Fig. 3.41).

In this figure an increase of the threshold is shown at frequencies $>300\ \text{Hz}$. This is caused simultaneously for two reasons. First, as explained in Sect. 3.5.5 (see also Fig. 3.49), the influence of low frequency fields on the membrane potential decreases with increasing frequency, and nearly vanishes at frequencies above $1\ \text{MHz}$. This is caused by the increasing capacitive conductivity of the membrane,

i.e., by its RC behavior. The second reason concerns the mechanism of membrane excitation itself. As shown in Fig. 3.29, the action potential of nerve or muscle cells lasted several milliseconds. This means that an excitation with a frequency of several 100 Hz, which is faster than 10 ms, becomes ineffective because of the refractory period of the cell after an actual excitation.

At the frequency range of electromagnetic fields, and of alternating currents up to about 100 kHz, the dominant mechanisms of interaction therefore, are excitations of nerves and muscles. These excitation properties, shown in Fig. 4.29, led to the recommendation that in the frequency range between 4 Hz and 1 kHz the basic restriction level of occupational exposure should be limited to current densities less than 10 mA m^{-2} in the body. For the general public an additional factor of 5 is applied. This corresponds to the reference level as listed in Fig. 4.28.

Comparing Figs. 4.28 and 4.29 it strikes one that in one case the functions step up with increasing frequency, while in the other case, they decline. The reason for this is that even if the sensibility of nerve and muscles to electric excitations decreases with increasing frequency (restriction level) (Fig. 4.29), the current density of induction, i.e., the efficiency of the induction process, increases at the same time, and will eventually dominate the function of the reference levels (Fig. 4.28).

To evaluate thermal effects of low frequency fields the resulting power density (P) must be calculated. This parameter in W m^{-3} can easily be translated into the *specific absorption rate* (SAR) in W kg^{-1} using a mean density of tissue of about $1,000 \text{ kg m}^{-3}$. For low frequency fields, using Ohm's law, this means:

$$P = j\mathbf{E} = \frac{j^2}{g} = \mathbf{E}^2 g \quad (4.23)$$

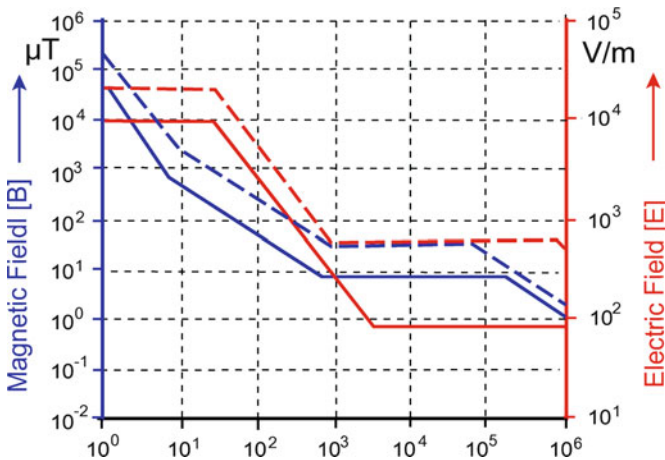
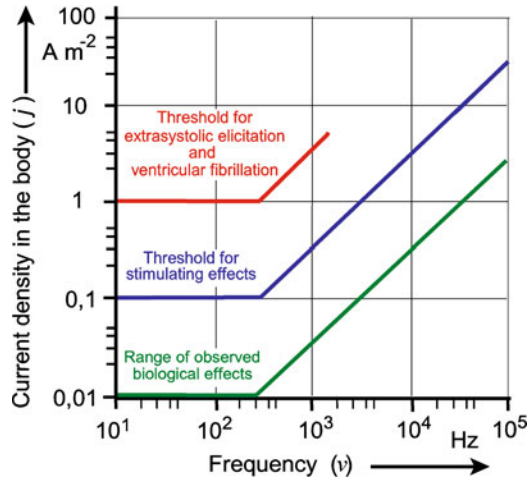


Fig. 4.28 Reference levels for magnetic and electric field for the general public (solid lines) and occupational exposure (dashed lines) (After ICNIRP 1998)

Fig. 4.29 Threshold values of the electric current density in the human body in the low frequency region (After Bernhardt 1988, simplified)



where j is the current density in A m^{-2} , g is the specific conductivity in S m^{-1} , and E the field strength in V m^{-1} .

Using the parameters of Fig. 4.29 and inserting them into Eq. 4.23 one can easily see that for low frequency fields even at the highest threshold, namely absolute danger to life, the SAR amounts to only about 0.001 W kg^{-1} . This is negligibly small in relation to the basic metabolic rate, i.e., the heat production of the resting human body which amounts to 1 W kg^{-1} , and which furthermore often increases in the case of active movement by a factor of ten (see Sect. 3.8). This can be directly related to the dissipation function (Φ) (see Sect. 3.1.4, Eq. 3.64) which for small mammals is even larger (see Fig. 3.7).

Because of its relevance to safety standards for the population, the biological effects of low-frequency electromagnetic fields, even below the threshold of excitation, have been the subject of many investigations in recent decades. Experiments have been carried out on all levels of biological organization, and completed by epidemiological studies. Consistent with biophysical considerations, however, no reproducible effects below the level of excitation could be found. Therefore, there is no good reason to decrease the safety levels as discussed above.

Low frequency currents and electromagnetic fields have also been applied for various therapeutic uses. This includes implanted stimulating electrodes, like cardiac pacemakers, and various kinds of externally applied electrodes for electrostimulation of nerves (TENS = Transcutaneous Electric Nerve Stimulation), as well as attempts to influence biological processes using eddy currents, induced by external coils. As already mentioned in Sect. 4.4, neuronal electric stimulation in this way is possible, using very strong millisecond magnetic pulses (1.5–2 T). In some cases implanted electrodes are used, combined with small coils, which receive their current by induction processes in external magnetic fields. Furthermore, various PEMF-techniques (Pulsed Electro-Magnetic Field) with external coils have been proposed for the treatment of various diseases.

As a criterion for the real efficiency of these methods a calculation must be made as to whether the induced current density is really large enough to affect the biological system. This not only depends on the intensity and frequency of the applied field, but additionally on the edge steepness, if rectangular pulses are applied. Not only for public health protection, but moreover for possible therapeutic effects the values depicted in Fig. 4.29 are important. As no biological effects have been found below these levels, no therapeutic results would be expected below these intensities. In fact, many of the advertized therapeutic machines must be classed as placebo therapy.

Further Reading

General aspects of low-frequency field effects: Barnes and Greenbaum 2006; biological sensibility: Francis et al. 2003; Weaver et al. 1998; aspects of health protection: ICNIRP 1998; history of medical applications: Rowbottom and Susskind 1984.

4.7 Radio- and Microwave Electromagnetic Fields

As already mentioned in Sect. 4.6, the classification of the frequency regions of electromagnetic fields is rather arbitrary (see Fig. 4.26), but there are a number of physical and biophysical peculiarities which makes it reasonable to distinguish between “low” and “high” frequency. The frequency bands which in Fig. 4.26 contain the Letter “H,” in their lower part, are mostly called “radio frequency.” In the region above about 3 GHz on the other hand the term “microwaves” is used.

4.7.1 Physical Background and Dosimetry

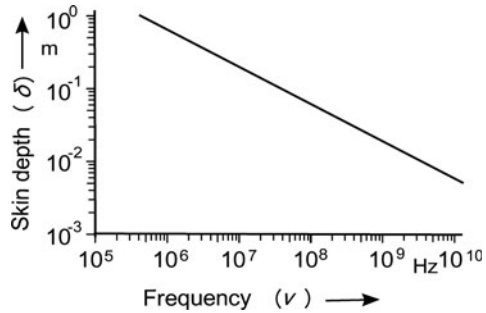
In contrast to low frequency fields, in the HF region a splitting into the magnetic and electric field vector is impossible. Although up to UHF frequencies it is technically possible to apply fields at least with pronounced electrical or magnetic components, in the practical use of technical devices, both components are connected.

In this frequency region the so-called *skin effect* becomes important. This is the result of the eddy currents, induced in the conductor, cancelling the current flow in the center, and reinforcing it at the surface. With increasing frequency the tendency for an alternating electric current to distribute itself within a conductor increases so that the current density near the surface of the conductor is larger than that at its core.

This effect also determines the depth of penetration of high-frequency electromagnetic fields in a homogeneous body. It is characterized by the *skin depth* (δ) which can be calculated for practical use by:

$$\delta = \sqrt{\frac{1}{\pi \mu \mu_0 g \nu}} \quad (4.24)$$

Fig. 4.30 Skin depth (δ) according to Eq. 4.24 for a homogeneous body with a specific conductivity $g = 0.6 \text{ S m}^{-1}$ as a function of frequency (ν)



This skin depth (δ) is the distance over which the field intensity decreases to a factor of $1/e = 0.368$. Figure 4.30 indicates the function $\delta(\nu)$ for the following conditions: magnetic permeability of vacuum $\mu_0 = 1.256 \cdot 10^{-6} \text{ V s A}^{-1} \text{ m}^{-1}$, the magnetic permeability number $\mu = 1$, and the mean conductivity of the tissue: $g = 0.6 \text{ S m}^{-1}$ (see Fig. 3.41). It shows that this effect will become important only at high frequencies ($\nu > 10^6 \text{ Hz}$).

The distribution and penetration of high-frequency electromagnetic fields in the living body is an important factor in therapeutic applications, as well as in all safety considerations. However, calculations using Eq. 4.24, and correspondingly Fig. 4.30 only represent a rough estimation of field penetration because they assume a body with homogeneous dielectric properties. After all, it shows that in the frequency range from about 20 to 300 MHz a relatively high absorption can occur in the whole body. At higher frequencies, the field does not penetrate deeply, but will be absorbed mainly at the surface. It is possible, for example, to estimate in this way that the electromagnetic field of a mobile telephone with a frequency of around 1 GHz fully penetrates the body of small laboratory animals, but affects in humans only a certain part of the surface. At frequencies above about 10 GHz the absorption occurs primarily at the skin.

For detailed dosimetric considerations, particular anatomic and dielectric peculiarities of the body must be taken into account. As already shown in Sect. 3.5.3 (Fig. 3.41), the dielectric parameters of different tissues vary to a large extent. Furthermore, in contrast to the assumptions leading to the graph in Fig. 4.30, the conductivities themselves are functions of the frequency. To establish a realistic dosimetry for high-frequency field exposition, various model calculations have been proposed, using the finite-difference time-domain (FDTD) method. In this case, based for example on detailed MRT-pictures of the body, analyses have been done with pixel volumes down to 1 mm^3 .

In general the dosimetric calculations are based on the intensity of the field which is characterized by the *plane power density* (W m^{-2}) emitted by an external source. In the case of far-field exposure, i.e., at a distance from the antenna more than twice the wavelength, plane wave conditions can be expected. Furthermore, in dosimetric considerations resonance phenomena of the whole body or of parts of it (e.g., the head) must be considered. The human body, if not grounded, has a resonant absorption frequency close to 70 MHz. For small laboratory animals, such as rats or mice, the resonance frequency is correspondingly higher.

As already mentioned in the introduction to Sect. 4, according to the Grotthus–Draper principle, not the energy penetrating the organism, but only that part which is actually absorbed into the system can be effective. Therefore, a *specific absorption rate* (SAR) in W kg^{-1} is defined to characterize the dose of high frequency exposition (see Eq. 4.23). In contrast to the plane power density (in W m^{-2}), which just depends on technical parameters of the field-emitting device, the SAR is the energy, which per unit of time is absorbed in a unit of mass or volume of the body. Knowing the SAR and the heat capacity of the tissue, thermal effects of absorbed electromagnetic fields can be calculated. In this way SAR is also directly measurable by sensitive thermistors. Because of the inhomogeneity of the tissue, the SAR values can differ at various locations. For this reason SAR values are evaluated for an average mass of 10 g, or in some cases even for 1 g. In the case of time varying radiation, time-averaged values are used.

At frequencies greater than about 10 GHz, because of the limited depth of field penetration, the SAR is proofed not as a good measure for assessing absorbed energy. This is the reason where as at microwave frequencies up to UV-light the incident power density (W m^{-2}) is used as a more appropriate dosimetric quantity.

4.7.2 Biophysical Aspects of High-Frequency Field Interaction

In the previous section (Sect. 4.6.2) we explained that in the frequency range up to about 100 kHz excitations of nerves and muscles, but not thermal effects, are the dominating mechanisms of interaction. At higher frequencies, however, this situation reverses. In the frequency region $\nu > 10^5$ Hz, diathermal heating becomes dominant. *Diathermy* means an inhomogeneous heating of the body corresponding to the inhomogeneity of field absorption, i.e., the inhomogeneous distribution of the SAR-value, as discussed in Sect. 4.7. This word, combined from the Greek words *διὰ* (through) and *ἔρμος* (heat), denotes heat generation inside the body. This inherent difference of diathermy in contrast to conventional heating by thermoconduction from outside makes it useful for various kinds of therapy.

In the discussion of safety aspects of radiofrequency irradiation, it was hypothesized that beside diathermal heating, additionally, there may exist nonthermal effects. In fact the terms *thermal* and *nonthermal* (or *athermal*) are used in different ways. From the empirical point of view, effects are usually called “nonthermal” in situations where the irradiation intensity is so low that changes in temperature of the exposed body are not predictable or not measurable, or if some observed effects do not correspond to such, which appear to the same degree after conventional heating.

Contrary to this, the biophysical definition of this term is based on the types of mechanisms of field interaction. A mechanism is considered as nonthermal if the interaction of the electrical (or magnetic) vector of the electromagnetic field with charges or dipoles of molecules in the living system directly leads to specific

effects, other than heating, or if the system changes its properties in a way that cannot be achieved simply by a temperature increase.

According to this biophysical definition a number of nonthermal effects are well known. This includes, for example, dielectrophoreses or the electrorotation of cells as described in detail in Sect. 3.5.5 (see Fig. 3.47). In these cases the electric field induces dipoles which directly lead to cell movements or deformations. Actually, these effects require field intensities, which cause considerable heating. A field of about 10^4 V m^{-1} , which for example is necessary to induce dielectrophoretic cell movement in a medium with a conductivity of about 0.1 S m^{-1} according to Eq. 4.23 produces a SAR of 10^7 W m^{-3} , or approximately 10^4 W kg^{-1} . This of course leads to an intense heating of the system. As already mentioned in Sect. 3.5.5, the application of these dielectric techniques in biotechnology requires special precautions of thermal protection.

In contrast to these effects at extremely large field strength, nonthermal effects at low field intensities, i.e., without a measurable temperature increase, are biophysically not imaginable, and experimentally not convincingly demonstrated in this frequency region. Occasionally described results of this type are either nonreproducible, or after all a result of local heating.

For this reason the recommendation of exposure limits is based on the circumstance that no established biological and health effects could occur at a rise in the body temperature $>1^\circ\text{C}$. This temperature limit resembles an amount, which corresponds to everyday heating of individuals under moderate environmental conditions. It will arrive at a whole body SAR of about 4 W kg^{-1} . This corresponds to the magnitude of the basic metabolic rate, whereas not the mean metabolic rate of the whole body must be considered to evaluate health risk, but rather that of the most active organs like muscles or brain which may be higher by one order of magnitude. Using an additional safety factor, for whole body irradiation with frequencies between 100 kHz and 10 GHz an average whole body SAR limit of 0.08 W kg^{-1} is recommended. Note, that this is 5 orders of magnitude lower than the above-mentioned SAR in biotechnological applications, allowing cells to live and develop even under permanent exposition for days without observable damage!

The recommendation of the exposure limit is therefore based exclusively on thermal effects of HF radiation. This is justified because “nonthermal” effects in the empirical definition, i.e., without an increase in the temperature, have not been established, as already mentioned. The comparison of effects of diathermal heating with heating to the same temperature in a conventional way, appears to be an unrealistic approach because the resulting temperature gradients are quite different for both cases.

In fact, all of these considerations of the system of thermoregulation of the living body must be taken into account. This has already been described in Sect. 4.1, where in the bioheat equation (Eq. 4.4) the SAR as a generalized heat input from outside is included. Considering the system as illustrated in Fig. 4.2, it cannot be excluded that thermoreceptors can be activated as the result of low-intensity field interaction without a measurable increase in the body temperature, which could

lead to local effects like modifications of blood circulation or EEG. This, however, is comparable with other everyday effects without noxious implications.

The absorption of HF fields in biological systems is basically determined by its water content. In contrast to free water, the dipoles of which are oscillating in the frequency of 18.7 GHz (Sect. 3.5.3), the resonance frequency of bound water can be shifted down, up to tenths of MHz. The broad dispersion of dielectric properties of the biological material (Fig. 3.41) makes it impossible, to find sharp absorption maxima or resonance peaks, and therefore no frequency windows of HF-field interaction are to be expected.

At strong high frequency pulses a *microwave auditory effect*, or *RF-hearing* occurs which is caused by abrupt heating of tissue water in the head. The sudden expansion which is generated in this way, launches acoustic waves which can elicit low intensity sound like a buzz, click, hiss, or knock. RF-hearing has been reported at frequencies ranging from 2.4 MHz to 10 GHz. This effect depends on the energy of the single pulse, and not on the average power density.

These considerations lead to the problem of *microdosimetry*, and the question whether *microthermal* heating is possible as the result of field exposure. In fact, the heterogeneity of dielectric parameters is not only established in anatomical dimensions, as demonstrated in Fig. 3.41 (Sect. 3.5.3) but moreover down to the cellular, subcellular, and even molecular magnitudes. What could the minimal size of a hotspot, i.e., a region with particularly increased temperature caused by increasing absorption of high frequency energy, be?

In the simplest approach, this question can be checked on the basis of heat dynamics of a sphere in an environment with different dielectric properties. Assume that a spherical object of radius r is subject to heating at a given SAR, and is surrounded by unheated material. The maximum temperature increase (ΔT), and the thermal time constant [τ (s)] can be found as a solution of the heat equation (Eq. 4.4 in Sect. 4.1), ignoring the blood flow term, and the heat production by metabolism:

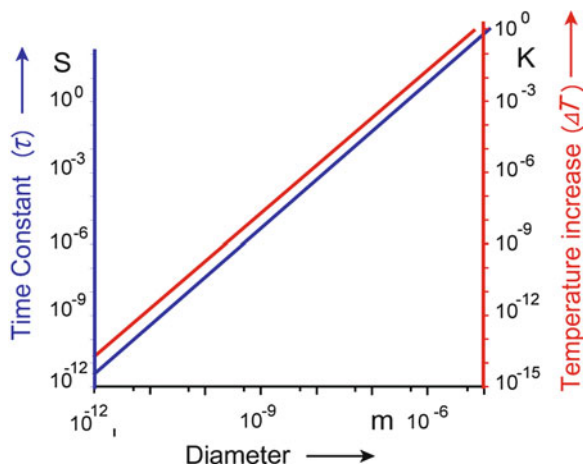
$$\Delta T = \frac{\text{SAR}}{C} \tau \quad (4.25)$$

$$\tau = \frac{\rho C r^2}{\lambda} \quad (4.26)$$

where C is the specific heat capacity, ρ the density, and λ the thermal conductivity.

These equations show that the time constant (τ), and therefore also the steady-state temperature increase (ΔT), are proportional to the square of the dimension of the sphere (r^2). As shown in Fig. 4.31, in fact the temperature fluctuations produced by selective heating of very small structures, as well as the time scale in this dimension are negligible. According to these calculations, points of local temperature increase, i.e., so-called *hot spots*, can occur only in regions of dielectric inhomogeneity with a size at least in millimeter dimensions.

Fig. 4.31 Time constant (thermal response time) and maximum steady-state temperature increase of a sphere as a function of its diameter exposed at a SAR of 10 W kg^{-1} in an unheated medium. For the thermal properties of the spheres the parameters of brain tissue were used (Data from Foster and Glaser 2007)



It should be mentioned, however, that these equations are hardly applicable to processes in molecular dimensions where the phenomenological concept of temperature changes into that of statistical thermodynamics. Conversely, as explained in Sect. 2.1.6 (see Fig. 2.8), the frequencies of vibrational and rotational movements of molecules are above 10^{10} Hz, i.e., essentially in the Terra-Hertz region.

To analyze possible biophysical mechanisms of high-frequency effects of weak fields on biological systems, a number of hypotheses had been proposed. Some of them are based on the idea that high-frequency electromagnetic fields could exhibit classical resonance phenomena, and in this way might absorb energy in excess. The particular analyses, however, suggest that the vibratory motion by biological fluids is severely restricted by the damping properties of water. The absence of a reliable biophysical theory for possible low-intensity, nonthermal effects of high-frequency electromagnetic fields in fact corresponds to the unavailing experimental results. Therefore only thermal, or possibly microthermal effects are to be proposed below electrorotation or dielectrophoreses as nonthermal effects at strong field intensities.

Further Reading

General effects: Barnes and Greenbaum 2006; microthermal effects: Foster and Glaser 2007; RF-hearing: Elder and Chou 2003; mechanisms and resonance phenomena: Adair 2002; guidelines and limits: ICNIRP 1998.

4.8 Visible and Nonvisible Optical Radiation

The spectrum of technically used frequencies of electromagnetic radiation, as discussed in the previous sections, ended in Fig. 4.26 with the “Extremely High Frequency” (EHF) region corresponding to a wavelength of 1 mm. In continuation, the “Terra Hertz Frequency” region follows, which reaches into the infrared region.

Table 4.5 Characteristic parameters of optical radiation

		Wavelength (λ)	Frequency (ν) in Hz	Quantum energy (E) in eV
Terahertz-radiation	<1 mm	$3 \cdot 10^{11}$ – $3 \cdot 10^{12}$	$<1.24 \cdot 10^3$	
Infrared	IRC	3,000 nm–1 mm	$3 \cdot 10^{11}$ – $1 \cdot 10^{14}$	$1.24 \cdot 10^3$ –0.41
	IRB	1,400–3,000 nm	$1 \cdot 10^{14}$ – $2.1 \cdot 10^{14}$	0.41–0.89
	IRA	700–1,400 nm	$2.1 \cdot 10^{14}$ – $4.3 \cdot 10^{14}$	0.89–1.77
Visible light		400–700 nm	$4.3 \cdot 10^{14}$ – $7.5 \cdot 10^{14}$	1.77–3.09
Ultraviolet	UVA	315–400 nm	$7.5 \cdot 10^{14}$ – $9.5 \cdot 10^{14}$	3.09–3.94
	UVB	280–315 nm	$9.5 \cdot 10^{14}$ – $1.1 \cdot 10^{15}$	3.94–4.42
	UVC	100–280 nm	$1.1 \cdot 10^{15}$ – $3.0 \cdot 10^{15}$	4.42–12.4
Ionizing radiation		<100 nm	$>3.0 \cdot 10^{15}$	>12.4

Here the nomenclature of high frequency engineers on the one hand, and the specialists in optical spectroscopy on the other hand overlap. The THz region from 0.3 to 3 THz, corresponding to a wavelength (λ) from 0.1 to 1 mm, overlaps with the far infrared, the so-called IRC, which is defined in the large region between 3 μ m and 1 mm (see Table 4.5). Therefore, the THz range spans the transition from radio-electronics to photonics.

In contrast to Fig. 4.26, in Fig. 4.32, the quantum energy of the radiation is additionally included (E in eV) as a third coordinate. The relation between E and the frequency (ν) is given by Planck's constant $h = 6.626 \cdot 10^{-34}$ J s = $4.136 \cdot 10^{-15}$ eV s, corresponding to the relation $E = h\nu$.

In the lower frequency region the quantum energy of radiation is far too low to be considered as a possible reason for interactions with matter. But if it rises up to the levels of molecular vibrations, this parameter becomes of increasing importance. At first it exceeds the energy of thermal noise (kT). Finally, it achieves the energy of covalent bonds, and therefore the energy of ionization. Therefore, in Fig. 4.32 two remarkable points in the scale are included: First the value of 0.026 eV, which corresponds to the energy of thermal noise at 300 K ($E = kT = 8.617 \cdot 10^{-5} \cdot 300$ eV = 0.026 eV), and then the energy of ionization, i.e., the quantum energy which is sufficient to break a covalent bond of water, which amounts to around 12 eV. This second point in the frequency spectrum is considered as the border line to ionizing radiation (see Sect. 4.9).

It is important to note that these two points indicate the general range of energetic transformations in biological processes. From the point of view of thermodynamics, effective mechanisms of energy transformation are possible only above the level of thermal noise. Conversely, the quantum energy must not be as large, as it could destroy the machinery itself. Therefore, the upper limit of biological existence is the quantum energy of ionization, where the quantum energy of the radiation starts to break the covalent bonds of proteins and other biologically important molecules. In fact, only the tiny frequency region of “visible light” between 400 and 700 nm, i.e., 1.77–3.09 eV is suitable to be used by biological systems to extract energy from sunlight, and to recover optical information.

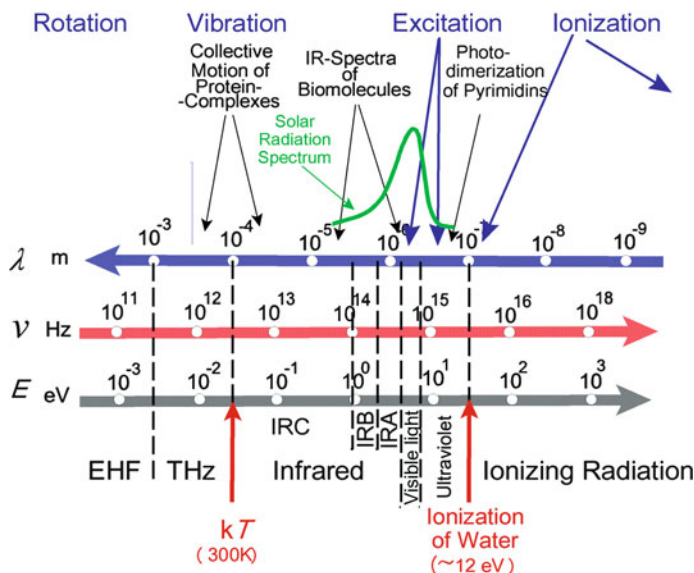


Fig. 4.32 The spectrum of optical frequencies and ionizing radiation (continuation of technical-used frequencies shown in Fig. 4.26)

The scientific literature for this frequency spectrum is largely dominated by spectroscopic investigations. This aspect is quite important for research on the structure and functions of molecular systems, and presented in a number of excellent monographs. Corresponding to the aim of this textbook, the following explanations, however, will concentrate on the mechanisms of molecular interactions, and on aspects of biological influences of this kind of radiation.

It should be noted here that in optical spectroscopy, as a peculiarity, the parameter *wave number* is used to characterize absorption spectra. This parameter is reciprocal to the wavelength. The wavelength (λ) of $3,000 \text{ nm} = 3 \cdot 10^{-4} \text{ cm}$, for example, corresponds to the wave number of: $1/3 \cdot 10^{-4} = 3,330 \text{ cm}^{-1}$. Unfortunately, this parameter is sometimes called “frequency” by spectroscopists, which of course has the correct dimension: $1/\text{s} = \text{Hz}$. Furthermore, the abscissae of absorption spectra are sometimes plotted in a decreasing, and sometimes in an increasing version. In the following text we will only use the three parameters λ (m), ν (Hz), and E (eV) as depicted in Fig. 4.32, plotting them always in the same orientation.

4.8.1 THz and Infrared: The Vibration-Inducing Frequencies

The frequency spectrum which will be considered in this chapter can be characterized in general as thermal radiation, emitted by the sun, by heated materials, and also by the human body. As shown in Figs. 4.32 and 4.37 the thermal

radiation which arrives at the earth's surface from the sun occurs mainly at frequencies $>10^{14}$ Hz, i.e., in the frequency region of IRA, and to a lesser extent at IRB. This also corresponds to the warm sensation of animals, which, however, is considered a temperature sensation according to the Arrhenius law (Sect. 2.1.5), rather than a perception of photons of the infrared radiation. Even the infrared orientation of some snakes, which helps them in the hunting of warm-blooded animals, are not based on particular quantum processes but rather on a temperature-sensitive membrane inside a particular pinhole camera (see Sect. 4.1).

Only in the last decennium has the technical possibilities been developed to produce and to indicate radiation in the THz-frequency range. The THz technologies now receive increasing attention, and various devices use this wavelength. Therefore this frequency band becomes increasingly important in diverse applications. It opens the way to extend infrared spectroscopy into these parts of the spectrum. Because the THz-radiation is strongly absorbed by water molecules, it penetrates just several millimeters of fatty tissue with low water content. This also complicates the THz-spectroscopy of biomolecules in their natural water environment. Conversely the dependence of THz absorption on the water content of the tissue may become important for some diagnostic methods, for example to detect some kinds of epithelial cancer. Some frequencies of terahertz radiation can be used for 3D imaging of teeth and may be more accurate and safer than conventional X-ray imaging in dentistry.

The frequency dependence of absorption of infrared radiation by water and CO_2 is best illustrated by the absorption spectrum of the normal atmosphere, not loaded with fog or rain (Fig. 4.33). It shows a maximum in the frequency region between 10^{12} and 10^{13} Hz. Therefore, it is somewhat unsuitable for use in technical telecommunications, at least under outdoor conditions.

The molecular mechanism of direct interaction of THz and infrared radiation with biological systems can finally be deduced from the data of absorption spectroscopy and by inelastic scattering of photons (*Raman spectroscopy*). The absorption of photons leads to vibrational movements of the molecules and their components. Depending on the structure of the molecule, a more or less large number of vibrational modes are possible. The vibrational modes of proteins are mostly influenced by their secondary structure, and furthermore by their kinds of hydration.

As already discussed in Sect. 2.1.6 (see also Fig. 2.8) the vibration frequency of a C–C bond is in the infrared frequency range of 10^{14} Hz which corresponds to a wavelength of 3 μm . The modes of rotation of these bonds, in contrast, are much lower, and correspond to a frequency below 1 THz. As shown in Fig. 4.32, this is below the energy of thermal movement (kT). The vibration modes, however, may also shift to lower frequencies if collective molecular motions or those of strongly hydrated macromolecules are considered. Collective motions of proteins can involve subdomains of hundreds of atoms with corresponding frequencies up to the THz regime. In these complexes the vibrational energy may flow in a concerted manner, to produce conformational changes.

Recently methods of THz time-domain spectroscopy have been developed. In this case short THz-pulses are applied to the sample material. This allows one to

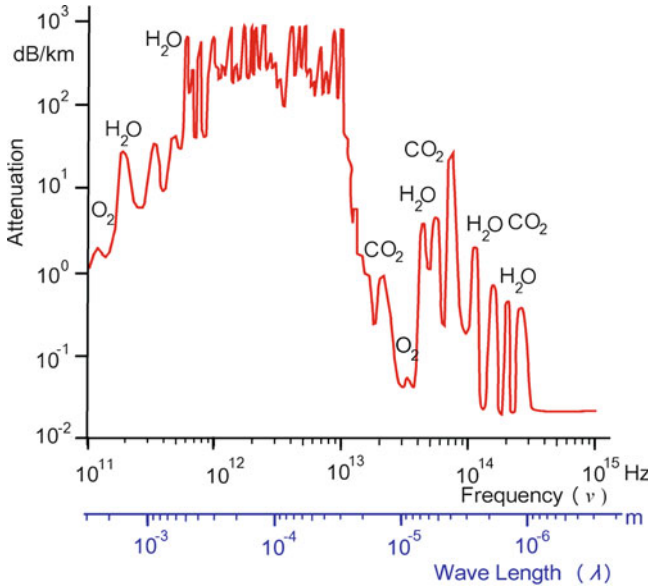


Fig. 4.33 Light absorption in the earth's atmosphere (Data from Sizov 2010)

register both the amplitude and the phase shifts, which contain far more information than a conventional image formed with a single-frequency source.

Exposure of human skin to infrared radiation resulting from solar irradiation, or from IR-emitting devices is used in medical treatments, as a preventative measure, and more recently in the “wellness” sector. Terahertz radiation is nonionizing, and thus is not expected to damage tissues and DNA, unlike X-rays. Single overexposures to IR radiation, however, can lead to acute damage in the form of skin burns, or collapse of the circulatory system. In cases of chronic or frequent repeated exposure, squamous cell carcinoma can result, especially in combination with other sources, like UV radiation. To date the recommendations for protecting humans from the risks of skin exposure to IR are only defined in terms of acute effects.

Further Reading

IR-spectroscopy: Siebert and Hildebrandt 2008; THz-spectroscopy: Plusquellic et al. 2007; Balu et al. 2008; safety aspects: Piazena and Kelleher 2010.

4.8.2 Visible Light: Processes of Excitation and Energy Transfer

The wavelength of visible light lies between 400 and 700 nm. This corresponds to quantum energies between 3.09 and 1.77 eV (see Fig. 4.32 and Table 4.5). This

quantum energy therefore, is far below that of ionizing radiation, a value, which is arrived at only in short-wave ultraviolet light (UVC). Conversely it is larger than that of thermal noise (kT), which at temperatures of biological importance amounts to only about 0.026 eV. As already mentioned in the preface to this Sect. (4.8) the organism uses exactly, this gap between the quantum energy of thermal noise, and the quantum energies causing ionization for photosynthesis and communication with the environment. The photons of light are strong enough for electron excitation, i.e., to lift electrons into a higher energetic state. This is the basic step for effective photochemical processes of energy conversion and photoreception, whilst not endangering the stability of biomolecules.

To understand the basic biophysical processes of photon absorption, let us first consider briefly the process of molecular excitation as illustrated in the so-called *Jablonski diagram* (Fig. 4.34). In a first step, the absorption of a photon leads to the raising of an electron to an orbit of higher quantum number. This can occur in the framework of the following two series of energetic states, which are qualitatively different: S_0, S_1, S_2, \dots , and T_1, T_2, T_3, \dots . Between these steps with increasing quantum numbers additionally, small energetic steps of thermal excitations are positioned. In the case of *singlet states* (S), the electrons of a pair have antiparallel oriented spins. The spin quantum numbers, therefore, have different signs. In the case of *triplet states* (T), the spins of the electrons of the pair are oriented in parallel, thus their spin quantum numbers are identical. The occurrence of electrons in which all quantum numbers are equal is ruled out by *Pauli's exclusion principle* stating that it is impossible for two electrons with identical quantum numbers to occur in the same atom. Thus, if the triplet state represents an electron pair with identical spin quantum numbers, these two electrons must differ with regard to other energy parameters although their orbitals can be energetically very similar. A triplet is a so-called *degenerated state*.

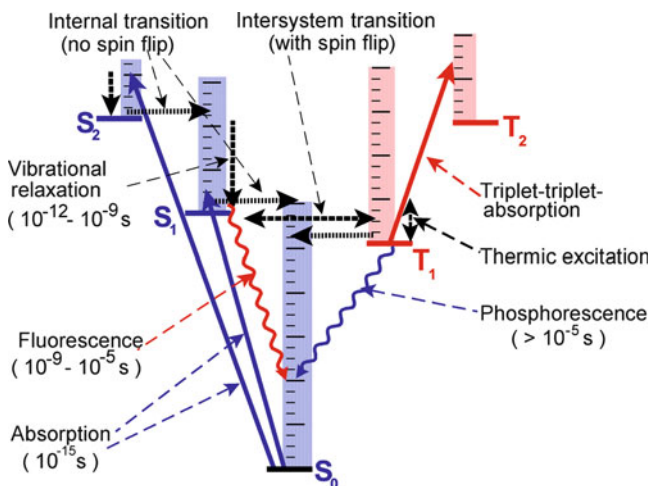


Fig. 4.34 Jablonski diagram of molecular excitations

These two states of excitation differ substantially in their life span. Figure 4.34 shows that the transition $S_1 \rightarrow S_0$ occurs under emission of *fluorescent light* within 10^{-9} – 10^{-5} s, whereas the transition $T_1 \rightarrow S_0$, recordable as *phosphorescence*, will occur at a much slower rate. Consequently, triplet states are characterized by an increased stability when compared with excited singlet states.

In photobiological systems of energy harvesting, as well as in photoreceptor processes the light energy absorbed by the chromophores must be transferred, in order to produce an excitation of the corresponding effector molecules. This transfer from a donor to an acceptor molecule occurs either by charge carriers, i. e., in an electron transfer as a redox-process, by way of fluorescence, or by Förster resonance energy transfer (abbreviated FRET transfer).

The mechanism of this *resonance transfer* can be envisioned as some sort of coupling between oscillating dipoles. It is a process in which a $S_1 \rightarrow S_0$ transition in the donor molecule induces an $S_0 \rightarrow S_1$ excitation in the acceptor. The excited electron of the donor molecule undergoes oscillations and returns to its basic state thus inducing excitation of an electron in the acceptor molecule. This process requires an overlapping of the fluorescent bands of the donor with the absorption band of the acceptor, i.e., the resonance of both oscillators. The smaller the difference of the characteristic frequencies between donator and acceptor, the faster the transfer will be. These so-called strong dipole-dipole couplings are possible to distances of up to 5 nm. This distance is in fact much smaller than the wavelength of emitted light. Therefore a real photon is undetectable, and this mechanism is classified as radiation-less. The term *Förster resonance energy transfer* (FRET) therefore seems to be more appropriate than *fluorescence resonance transfer*.

In general, an *energy transfer by radiation* is also possible. In this case the energy transfer occurs actually by fluorescent radiation emitted by one, and absorbed by the neighboring molecule. Such mechanisms are capable of transferring energy over distances which are large when compared with the other processes described in this context. However, the efficiency of this process is quite low. In fact, such mechanisms do not play a significant role in biological processes.

In contrast to FRET, which usually occurs as a radiation-less singlet-singlet transfer, the *Dexter electron transfer* is a mechanism which allows the energy transfer from triplet states. This is a particular case of electron transfer, in which an excited electron transfers from one molecule (the donor) to a second (the acceptor), maintaining its spin. Typically it may occur at distances below 10 nm.

In the most common metabolic reactions the energy transfer occurs by charge carriers as a classical example of a redox reaction. It consists basically of the transfer of one or two electrons from the donor, to the acceptor molecule. In this way, the donor becomes oxidized, and the acceptor reduced.

For these processes of electron transfer, donor and acceptor molecules must be in exactly defined positions to each other, and at a minimum distance, so that overlapping of respective electron orbitals can occur. In the first place, donor and acceptor will form a complex of highly specific steric configuration, a so-called *charge transfer complex*. This process of complex formation occasionally requires steric transformations of both molecules. It takes place at lower rates when

compared with the energy transfer by induction as discussed earlier. Hence, the charge-transfer complex is an activated transition state which enables redox processes to take place between highly specific reaction partners in the enzyme systems of the cellular metabolism. Because of the oscillating nature of electron transfer, this coupling of two molecules is strengthened by additional electrostatic forces sometimes defined as *charge-transfer forces*.

Further Reading

Montali et al. 2006.

4.8.3 Visible Light: Photobiological Processes

Electromagnetic radiation in the wavelength between 400 and 700 nm is called “visible light” because it is used in cells and organisms at all levels of evolution for information transfer, i.e., as photoreceptors as well as for bioluminescence, and furthermore in the kingdom of eutrophic organisms even as a source of energy.

About 0.05% of the total 10^{22} kJ energy which reaches the earth every year from the sun is assimilated by photosynthesis. This is the general energetic pool for all living processes of the earth. The efficiency of primary reactions of photosynthesis is very high, compared with our recent technical equipment. During subsequent metabolic processes of energy transfer, however, an additional loss of energy occurs. The total efficiency of the process of photosynthesis, in fact, is assumed to vary from 0.1% to 8%, depending on various environmental conditions.

The efficiency of photosynthesis is established by the system of energy transfer from light-harvesting antenna complexes to the entire photosynthetic reaction center. In plants, for example, one reaction center corresponds to nearly 300 antenna molecules. The antenna system, in fact, enlarges the diameter of optical effectiveness of the photoactive pigments by nearly two orders of magnitude. Depending on the species of plant, chlorophyll and various pigments (e.g., carotenoids such as xanthophylls) are part of these antennae (see Fig. 4.35). Size, composition, structure, and absorption spectra of the antennae are quite different for bacteria, algae, and various higher plants. This is understandable, considering the large differences in intensities and in spectral characteristics of the light leading to photosynthesis, in organisms ranging from submarine algae, up to tropical plants.

In eukaryotic plants the process of photosynthesis occurs in the chloroplasts, especially in the thylakoids located there. *Thylakoids* are flat vesicles with a diameter of about 500 nm, which form an ordered complex of a large number of thylakoids, so-called *grana*. One chloroplast contains about 10^3 thylakoids. Every thylakoid contains about 10^6 pigment molecules.

In general, photosynthesis can be considered as a reaction during which water is split, driven by the energy of photons, producing O_2 and transferring hydrogen to the redox system NADPH/NADP⁺, the nicotinamid-adenin-dinucleotid phosphate. Simultaneously, a proton gradient across the thylakoid membrane is generated.

Fig. 4.35 Estimated absorption spectra of chlorophyll a, chlorophyll b and carotenoids in chloroplasts, and action spectrum of photosynthesis (oxygen evolution/incident photon) (Data from Whitmarsh and Govindjee 1999)

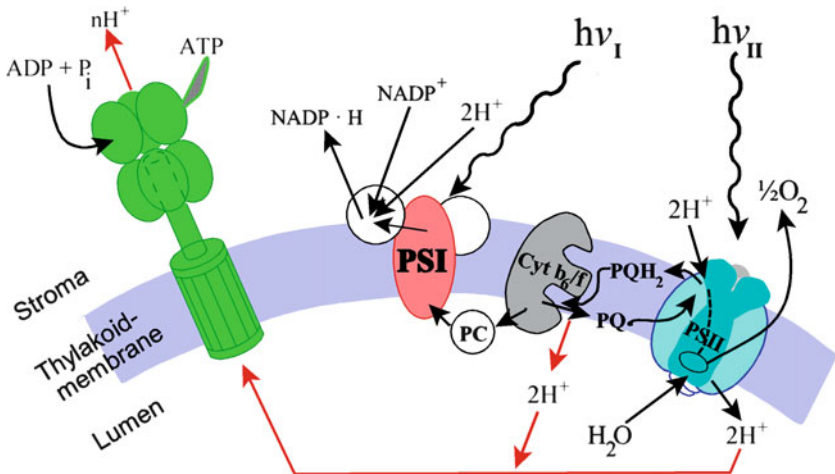
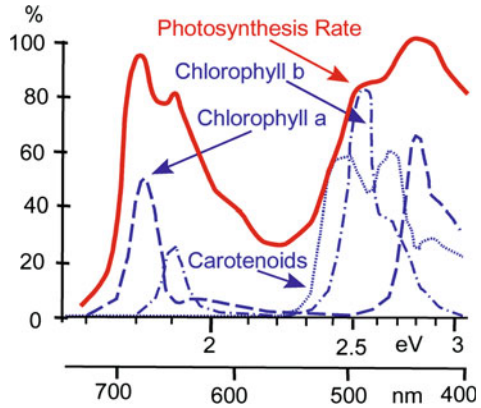


Fig. 4.36 Structural and functional organization of primary processes of photosynthesis in the thylakoid membrane. For an explanation see the text (From Renger 1994, redrawn)

Within a separate process, this leads to the synthesis of ATP. Subsequently, an ATP-consuming synthesis of carbohydrates occurs. This process is usually called the *dark reaction* of photosynthesis occurring in the stroma of chloroplasts.

The energy absorbed by the antenna molecules is transmitted by processes of nonradiant energy transfer to two reaction centers: photosystem I and photosystem II. The antenna molecules of these two photosystems are interconnected to each other. If photosystem II, for example, is overloaded, the absorbed energy will be transferred to photosystem I by a so-called *spillover process*. There are further mechanisms to protect the photosystem from over exposition.

In Fig. 4.36 the primary process of photosynthesis is depicted schematically. The complexes: photosystem I (PSI), photosystem II (PSII), as well as the cytochrome b/f complex (Cyt b/f) are seen as components of the thylakoid membrane.

The light-induced splitting of water occurs in the so-called water oxidizing multi-enzyme complex, structurally connected to photosystem II, where the molecular oxygen becomes free, but the hydrogen, in contrast, is bound by plastoquinon (PQ), forming plastohydroquinon (PQH₂). The cytochrome b/f complex mediates the electron transport between PSII and PSI, reducing two molecules of plastocyanin (PC), using the hydrogen of PQH₂. This light-independent process at the same time extrudes two protons into the internal volume of the thylakoid. In a second photochemical process, photosystem I (PSI), using the redox potential of plastocyanin, transfers one proton to NADP⁺, producing in this way NADPH, one of the energy-rich products of photosynthesis.

In addition to photosynthesis, as the primary process of energy conversion, light furthermore is an essential source of information. There are a number of photobiological processes, which do not end in the gain of chemical or electrochemical energy, but in signal cascades, triggering various kinds of biological response. These photoreceptors can vary from diffusely distributed photoreceptive cells, to complex organs in discrete locations.

The highest level of organization is arrived at in various organs of visual perception. In these the photoreceptor molecules, for example the visual purple rhodopsin, with a highly organized arrangement, are able to localize the source of light, or even reproduce the environment in detailed pictures. Other light-induced processes are *phototaxis* (the movement of an organism by orientation towards light), *phototropism* (orientation of parts of an organism, for example leaves of plants, to optimize received illumination), *photomorphogenesis* (light-controlled differentiation). Irrespective of the differences in location and function, photoreceptors show intriguing similarities in their basic structure and function.

The precondition of all photobiological processes is an effective absorption of corresponding photons. As the polypeptide backbone of the amino acid side chains do not absorb photons in the visible light range, the photoreceptor proteins contain one or more nonprotein components, i.e., chromophores, bound to the protein covalently or noncovalently. These chromophores are color tuned to their environment.

In general the photoreceptor molecules can be classified in the following groups: light-oxygen-voltage (LOV) sensors, xanthopsins, phytochromes, blue-light-using flavin adenine dinucleotides (BLUF), cryptochromes, and rhodopsins. Except for the rhodopsins, which for example are localized in the membrane of photoreceptors, all others are water-soluble, and have quite different cellular locations. The initial changes in the excited state of the chromophore must be very fast. Mostly, beginning with an $S_0 \rightarrow S_1$ excitation, an isomerization process occurs about one or more double bonds.

The mechanisms of photoreception require optimization in the same way as the energy-harvesting processes of photosynthesis. For this the initial process of photon absorption, as well as the transfer of the absorbed photon energy to the protein must be efficient and specific. In this way it is necessary to minimize the energy dissipation into nonproductive processes of vibrational and thermal movement, as well as a fluorescence de-excitation. For this in the receptor proteins various functional domains are localized directly, for example, some for binding the

chromophore, others to promote association with another protein or membrane, or output domains to exhibit light-dependent catalytic activity.

Beside reception, in some organisms even systems for emission of photons occur. *Bioluminescence* is distributed across a broad range of the major groups of organisms from bacteria and fungi to various animals. In insects and some other organisms it is realized by bacterial symbionts, mostly, and particularly in sea water inhabitants, it is produced by the organisms themselves. Bioluminescence is a kind of chemiluminescence typically produced by the oxidation of light-emitting molecules of the luciferin classes in conjunction with luciferase, a corresponding catalyzing enzyme. During the course of this reaction, an electronically excited intermediate is generated that emits a photon of light upon relaxation to the ground state. Because the ability to make light has evolved many times during evolution, a large number of modifications of these molecules occur. The wavelength of the emitted light depends on the chemical structure of the molecule emitter, along with influences of the conditions of the internal microenvironment. The emission spectrum of a particular luciferin-luciferase system can therefore vary.

This property of living organisms has recently been widely used in the so-called *bioluminescence imaging techniques* (BLI). In this case the light-emitting properties of luciferase enzymes are applied to monitor cells and biomolecular processes in living subjects. Luciferases and other proteins which are required for biosynthesis in luminescent bacteria for example, are encoded within a single *lux*-operon. In this way these light-emitting molecules can be introduced into a variety of nonluminescent hosts to track cells, to visualize gene expression, and to analyze collections of biomolecules.

Not only bioluminescence but also the fluorescence of particular marker molecules is used today as a powerful method for microscopic and submicroscopic research. In many cases this method even replaces the radiotracer method. The availability of many kinds of fluorescent tracers together with the technical development of confocal laser microscopes establishes its success.

Especially the discovery of the *green fluorescent protein* (GFP) and related proteins boosted this application. This protein was purified from the jellyfish *Aequorea victoria* by Osamu Shimomura in the 1970s (Nobel Prize 2008). It is supposed that this fluorescent protein (GFP) is colocalized with a bioluminescent counterpart and therefore is important for communication between these animals. The prevalent use of this protein and its modifications in research is because of the possibility to introduce the corresponding genes into any cellular DNA. In this case the cells themselves produce this fluorescent protein. Furthermore, the GFP gene can be fused to genes of particular proteins; the expressed molecules become fluorescent and their position in the cell can be localized.

Further Reading

Photosynthesis: Singhal et al. 1999; Orr and Govindjee 2010; Santabarbara et al. 2010; photoreception: Hegemann 2008; Möglich et al. 2010; energy transfer: Ritz and Damjanovi 2002; bioluminescence: Haddock et al. 2010; bioluminescence imaging: Prescher and Contag 2010; GFP: Nienhaus 2008.

4.8.4 Ultraviolet: The Transition to Ionizing Radiation

As depicted in Fig. 4.32, at the maximum of ultraviolet frequencies, the quantum energy of photons of UV-light forms the border to ionizing radiation. Formally, the UV spectrum is divided into three regions: UVA (315–400 nm), UVB (280–315 nm), and UVC (100–280 nm). UV-radiation in sunlight that reaches the earth's surface is mostly UVA, and partly UVB (Fig. 4.37). The short-wave part of UVB ($\lambda < 290$ nm) is completely screened out by the ozone in the atmosphere, and by the stratospheric ozone layer. The interaction of short-wavelength UVB, as well as UVC with biological systems is therefore only of interest in relation to some technical sources.

All three spectral regions of UV-radiation in different ways affect human health. On one hand UVB exposure induces the production of vitamin D in the skin. On the other hand, UVA, UVB, and UVC can all damage collagen fibers, and therefore induce photo-aging of the skin. Furthermore, it causes skin cancer, and many other detrimental health effects like sunburn, ocular damage, or immune suppression. Because of the absorption of UV from various chromophores in the skin, it penetrates only a few cell layers. Nevertheless, UVA reaches the hypodermis and therefore affects all layers of the skin. In contrast, UVB acts mainly in the epidermis including cells of the basal layer, whereas UVC is mostly absorbed by the *stratum corneum*.

Considering just the normal environmental UV-radiation spectrum, UVB is primarily responsible for most biological effects. In fact, it appears to be three to four orders of magnitude more effective than UVA. However, considering the sun's emission spectrum as depicted in Fig. 4.37, and weighting the spectral intensity, this effectiveness reduces to only the tenfold. UVA may negate or enhance the effects of UVB and vice versa. UVA is also known to affect different biological endpoints which are influenced by UVB. For example, UVA contributes towards photo-aging but mainly causes skin sagging rather than wrinkling, which is largely caused by UVB.

The genotoxic effect of UV-radiation is based on its direct interaction with the DNA molecule as well as indirectly by other photochemical reactions (Fig. 4.37). In contrast to sugar and phosphate residues, the absorption of UV by the DNA molecule is due to the pyrimidine bases. The *direct mechanism* of UV-interaction consists chiefly of the formation of 6,4-pyrimidine-pyrimidone and cyclobutane pyrimidine dimers (Fig. 4.38).

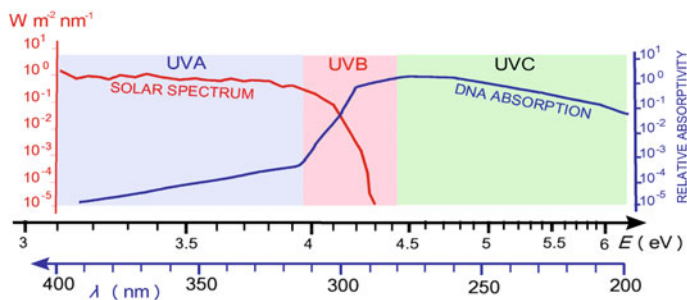


Fig. 4.37 Solar spectral irradiance measured at sea level on a clear day in July (Data from Godar 2005) and absorption spectrum of DNA (Data from Sutherland and Griffin 1981)

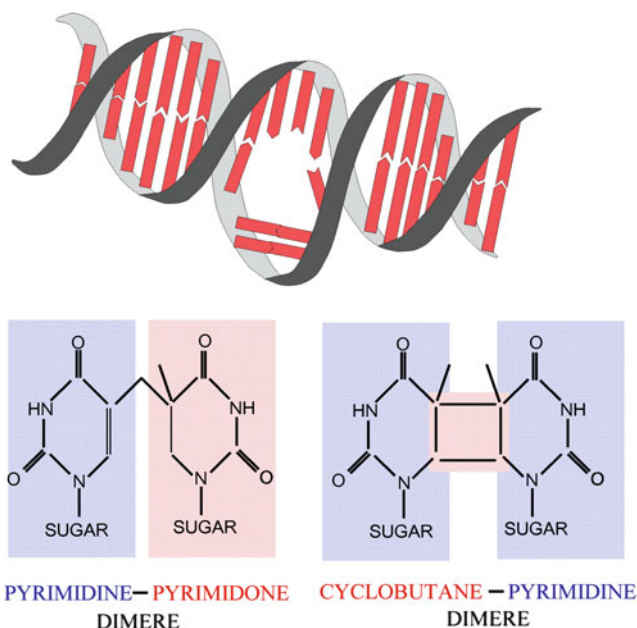


Fig. 4.38 Formation of a thymine-dimer as an example of a 6,4-pyrimidine-pyrimidone, and a cyclobutane pyrimidine dimer, and the consequences to the structure of a DNA double helix

In this case two adjacent pyrimidines of the same strand connect across 5–6 double bonds, and are in this way bound together, or form a cyclobutane ring structure. The yields of the various photoproducts depend on the wavelengths of the radiation. Mostly, thymine dimers appear, but even other combinations are found like cysteine-thymine, cysteine-cysteine, as well as thymine-uracil and cysteine-uracil. These dimers are extremely stable and they can therefore easily be indicated analytically.

The formation of these dimers requires an extensive rotation of the neighboring pyrimidines from their usual position. They were formed more easily in melted and in curved DNA. Conversely, increased DNA rigidity may interfere with this kind of photodimerization, probably by preventing the optimal alignment of the double bonds involved in dimer formation.

These dipyrimidine photoproducts alter the structure of DNA, and consequently inhibit polymerases. In this way they disturb normal replication. Even if they are normally excised by the system of nucleotide repair enzymes, the amount of unrepaired lesions is not negligible, and results in mutations. These are called “*UV-signature mutations*,” which are almost only the result of UV-radiation. They falling into the mutation type of *transition*, i.e., a change from one pyrimidine (cytosine or thymine) or purine base (guanine or adenine) to the other, for example a C:G pair to T:A.

The *indirect mechanism* of UV-interaction is generally caused by the generation of various reactive oxygen species (ROS). These include the superoxide anion radical (O_2^-), hydroxyl radical (OH) and many others. To some extent ROS are natural byproducts of the normal metabolism of oxygen, and have important roles

in cell signaling. Their concentration is normally controlled by a system of redox enzymes, such as superoxide dismutases, catalases, lactoperoxidases, glutathione peroxidases, etc. However, some environmental influences like UV, ionizing radiation (see Sect. 4.9.2), or ultrasound (Sect. 4.3.5) can increase these levels dramatically with detrimental consequences.

It should be emphasized, however, that the mechanisms of generation of these radicals are quite different in these three causes. In contrast to the sonolysis of water by ultrasound, where the ROS production is the result of a strong local increase of temperature, the UVB-induced formation of ROS is mediated by various photosensitizers, i.e., components of the cells absorbing the corresponding wavelength in the UVB range. The ground state photosensitizer absorbs UVB, and is excited to the singlet state which is very short-lived (<1 ns). By intersystem crossing (see Fig. 4.34) the triplet state occurs with much longer lifetime (~ 1 μ s). This triplet photosensitizer transfers its triplet energy to molecular oxygen to form singlet oxygen while the photosensitizer returns to the ground state.

Because UVA in contrast to UVB is only weakly absorbed by DNA itself, the damage induction in this case occurs via this indirect mechanism. These ROS induce DNA damage for example by the formation of the mutagenic modified base, 8-oxoguanine. This can cause G to T transversion mutations by mispairing of template 8oxoG with adenine, or alternatively T to G transversions via incorporation of 8oxodGTP opposite adenine during replication.

Irradiated cells may emit signals which induce DNA and other cellular damage in other cells which are not directly hit by the radiation. This is the so-called *bystander effect* which was first determined as a result of ionizing radiation. The agents of these bystander signals can be some kinds of ROS, as well as various stress response proteins, or byproducts of lipid peroxidation. They can be transferred via gap junctions between adjacent cells, or diffuse over larger distances in the tissue which can lead to effects even deeper in the tissue than the radiation could penetrate.

The intensity of UV-radiation is usually measured as a power density in W m^{-2} . In the distribution spectrum in Fig. 4.37, the dose is applied to a spectral slot, therefore defined as $\text{W m}^{-2} \text{ nm}^{-1}$. According to the so-called *Bunsen–Roscoe reciprocity rule* (Bunsen and Roscoe 1859), taking into account the possibility of accumulation of photolytic products, a photochemical reaction is directly proportional to the total energy dose, and is independent of the particular dose distribution. Therefore, the term $\text{W s m}^{-2} = \text{J m}^{-2}$ is applied. Considering the subsequent biological effects, however, the processes of repair in biological systems must be taken into account. Consequently, the biological effect depends not simply on the total dose, applied in a period, but additionally on the dose rate.

Because of differences in interaction mechanisms of various frequency windows of UV-radiation, and sensibility according to various biological reactions (e.g., erythema, photocarcinogenesis, tanning, melanogenesis) sometimes an *effective dose* is used, by multiplying the measured intensity with a corresponding factor.

Further Reading

Kiefer 1990; Nishigori 2006; Ridley et al. 2009; Svobodova and Vostalova 2010.

4.9 Ionizing Radiation

Studies on effects of ionizing radiation on living organisms have become necessary following the introduction of X-rays in medical therapy and diagnostics at the beginning of the twentieth century, and have acquired a special relevance in the so-called nuclear age. In this book only a few important aspects of radiation biophysics can be focused on.

In the spectrum of electromagnetic waves (Fig. 4.32), ionizing radiation begins at a quantum energy of the order of 10 eV. Usually the energy of ionization of the water molecule, which amounts to 12.46 eV, is taken as a borderline. This corresponds to a range of wavelengths in the upper limit of UVC (see Sect. 4.8).

The term “ionizing radiation” includes the photons of the corresponding part of the electromagnetic spectrum as well as various kinds of corpuscular radiation. These are accelerated elementary particles as well as ions of various mass numbers. Depending on the source of the photon radiation, it is usually called γ -radiation if it is the result of an atomic decay process. In 1895 the German physicist Wilhelm Röntgen detected a mysterious radiation, emitted by a cathode tube, which penetrates various materials. He termed it *X-rays*, using the mathematical designation for something unknown. This term is used even now for photon radiations emitted by various technical devices. In general, the terms γ -radiation and X-rays are synonyms for the same kind of photons, characterized solely by their wavelength resp. their photon energy.

4.9.1 Nature, Properties, and Dosimetry of Radiation

Ionizing radiation in our environment is emitted by extraterrestrial sources, by various technical devices, and by radionuclides. Even if the quality and quantity of human exposition to ionizing radiation has been altered by technical development, a natural exposition of all biological organisms has always existed and has been a part of environmental conditions during evolution.

In general ionizing radiation occurs as photons like X-rays or γ -radiations, as a part of electromagnetic waves (see Fig. 4.32), or as corpuscular radiation. Let us now characterize some properties of corpuscular radiations.

α -rays are the product of atomic decay of several naturally occurring isotopes of radium, uranium, thorium, etc. These are fast-moving helium nuclei with a mass number of 4 and an atomic number of 2 (${}^4_2\text{He}$). The helium nucleus consists of two neutrons and two protons and thus carries two positive charges. Because of this strong charge, there is a correspondingly strong interaction between α -rays and the elementary particles of matter. The energy of α -rays emitted by radionuclides is generally quite high, namely of the order of several millions of electron volts (MeV). As the α -rays pass through matter, this energy is dissipated as a result of ionizing processes and the particles are slowed down. For example, the energy which is required to generate one pair of ions in air is 34.7 eV. It can easily be

calculated that a radiation energy of several 10^6 eV is sufficient to induce about 10^5 ionizations. Consequently, α -particles will leave behind a straight-line track of limited length in irradiated matter which consists of the ionization products. This can be visualized by special methods.

β -rays consist of fast-moving electrons. These particles carry a single negative charge and an extremely small mass. Although it is possible to generate β -radiation of very high energy by means of particle accelerators, the energy of the β -radiation from radioactive nuclides on average is lower than that of α -radiation. In contrast to α -radiation, the energies of electrons of β -emitting radionuclides are not at all equal to each other but are spread across a specific range. Therefore, the terms *mean energy* and *maximum energy* are used when referring to the β -emission of radionuclides. The distance to which a β -particle penetrates matter will, naturally correspond with its energy. The ionization tracks of β -particles in contrast to those of α -particles, do not follow a straight line but their path becomes increasingly curved towards the end. In addition, the density of ionization increases as the energy of the particle decreases so that ionization is more densely packed with low energy radiation than with high energy radiation and, for the same reason, this effect can also be seen towards the ends of the tracks.

Neutron radiation as the result of nuclear fission, occupies a rather special position in the classification of corpuscular radiation. Its particles are electrically neutral and possess a considerable mass when compared with β -particles. For this reason they are able to penetrate an atom and reach the nucleus where they can cause nuclear transformations. In this way the irradiated substance itself becomes radioactive. Neutron emission does not occur when a radionuclide decays spontaneously but takes place during nuclear fission or in the wake of other, externally induced reactions.

Recently various kinds of nuclides have been used in cancer therapy. In this case ions like protons or carbon ions have been used, accelerated by synchrotron- or cyclotron-based facilities up to kinetic energies of several hundreds of MeV.

The process of energy dissipation that occurs during the absorption of electromagnetic radiation by matter follows the same pattern as that during the absorption of corpuscular radiation in that it is not continuous but takes place in steps, i.e., in quantum leaps. The size of energy quanta absorbed in this way depends on the type of interaction. In general, three types of absorption processes may be distinguished according to the relation between the required ionization energy and the available quantum energy:

- With low quantum energies, the *photo effect* is produced. In this, a γ -quantum (i.e., a *photon*) is absorbed and this causes a displacement of an orbiting electron from the shell of an atom. The excess energy, above that which is required for the ionization process, serves to accelerate this so-called *photoelectron*.
- The *Compton effect* occurs at a quantum energy of about 10^5 eV. In this case not only a *Compton electron* is ejected from the atom but this is accompanied by the scattering of secondary γ -radiation. This latter radiation has a quantum energy lower than that which was originally absorbed.

- If the quantum energy of the γ -radiation is above 1.02 MeV then *electron-pair formation* can occur. The quantum energy disappears producing a negative and a positive electron. Such formation of an ion pair can only take place close to the nucleus of an atom, i.e., it can only occur in an absorbing material and not spontaneously in a vacuum.

As ionizing energy moves through the material, a number of secondary particles can be created, at smaller or greater distances. The direct energy loss in the radiation as it moves through a material is called the *linear energy transfer* (LET). It is defined as the energy loss of the radiation per unit length.

To evaluate the dose of ionizing radiation, however, not only the LET, but the total energy is considered which is absorbed in the material, including also the effects of the secondarily produced particles and photons. For this the unit *gray* (abbreviated to Gy) is defined as the energy absorbed per unit mass:

$$1 \text{ Gy} = 1 \text{ J kg}^{-1}$$

In earlier papers the units *roentgen* (R) and *rad* (rd) were used. One roentgen was defined as the amount of radiation which is required to liberate positive and negative charges of one electrostatic unit in 1 cm^3 of dry air at standard temperature and pressure. Therefore, in contrast to the unit gray, it is not a physical equivalent for the dose of absorbed radiation, but rather a representative of the radiation effect. The unit rad is based on the CGS-system of units and is defined as equal to 100 erg g^{-1} . It can be converted as follows:

$$1 \text{ rd} = 100 \text{ erg g}^{-1} = 10^{-2} \text{ Gy}$$

To convert roentgen into gray, the properties of the particular absorbing matter must be considered. For the case of water and tissue it is:

$$1 \text{ R} = 0.93 - 0.98 \text{ rd}$$

In addition, as a measure of the *dose rate* the unit $\text{Gy s}^{-1} = \text{W kg}^{-1}$ is used. This corresponds to the SAR-parameter as explained in Sect. 4.7.1.

Different kinds of radiation have different effects in biological interactions. So, for example the *relative biological effectiveness* (RBE) of α -rays amounts to 20. In the case of neutron radiation, the RBE obviously depends on the actual dose but not on the energy distribution. An average of the maximum RBE of neutrons in relation to γ -rays of ^{60}Co amounts to 86 at the low-dose limit. For carbon ions the RBE is also variable and depends on the actual energy loss in the tissue.

This parameter is the *dose equivalent* with the unit *sievert* (Sv). It is named after the Swedish medical physicist Rolf Sievert. Considering the stated biological effectiveness values gives the following relations: for X-rays: $1 \text{ Gy} = 1 \text{ Sv}$, for α -rays, however: $1 \text{ Gy} = 20 \text{ Sv}$.

Considering the penetration of ionizing radiation into the body, some peculiarities need to be taken into account. For photon radiation, like X-rays, an

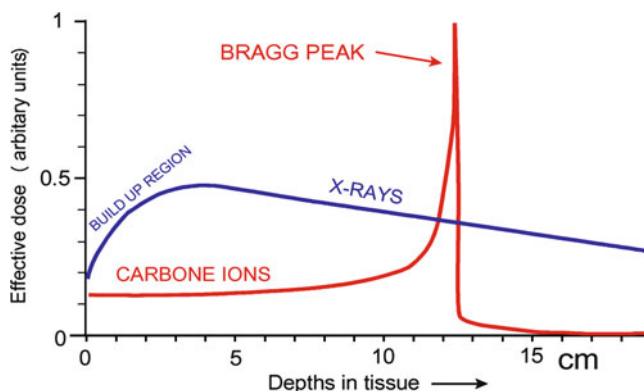


Fig. 4.39 Depth profile of the radiation dose in tissue for photons of X-rays (about 20 MeV) compared with carbon ions of about 300 MeV per nucleon (Modified after Durante and Loeffler 2010)

exponential decay of the intensity occurs, depending on the photon energy, and the density of the medium. Considering, however, the depth dose function it is found that the maximum dose is found at a certain distance from the surface (Fig. 4.39). This effect reflects the difference between the LET and the real doses which takes into account the fluence of the secondary particles. This so-called *build-up phenomenon* is caused by the fact that immediately at the surface access to a number of secondary products of interaction is limited. In the case of high photon energy radiation, the maximum of the curve is shifted several centimeters inside the body, an effect which is used in high voltage therapy.

In the case of beams of accelerated heavy ions the situation is even more complicated. The LET, and the release of secondary products, in this case is inversely proportional to the kinetic energy. The function which indicates the stopping power of the beam, i.e., the amount of ionization, is called the *Bragg curve*. As indicated in Fig. 4.39 this function shows a typical peak deeper in the body. In this way accelerated ions like protons or carbon are a useful tool in radiotherapy to affect particular tumor regions deeper in the body.

Further Reading

Karger et al. 2010; Kiefer 1990.

4.9.2 Primary Processes of Radiation Chemistry

The field of *radiation chemistry* represents the chemical consequences of the physical interaction of ionizing radiation, and it can be distinguished from *radiochemistry*, which is the study of particular chemical properties of radionuclides. As mentioned earlier, at quantum energies of photons, or correspondingly the kinetic

energy of particles for radiation greater than 12 eV, the interaction of the radiation leads to ionization of atoms and molecules. In contrast to the photons of visible light which reversibly raise electrons into a higher excited state (see Fig. 4.34), as explained in Sect. 4.8.2, in the case of ionizing radiation the electrons are fully ejected from the atomic orbitals. The results are on the one hand positively charged radicals, i.e., ions with an unpaired electron, and on the other hand free electrons. Both are short-living unstable products. As a consequence of this alteration, and as a result of the interaction of these primary products of radiolysis, even covalent chemical bonds could be broken.

The entire biological effect starts with interference in a biological cascade of information transfer, i.e., with a DNA molecule or an important enzyme. This can either be the result of a *direct* interaction with the radiation quantum, or the consequence of an interaction with another product of radiolysis (*indirect effect*).

Because water is by far the most common molecule in biological systems, its radiolysis products are of particular interest in the generation of indirect effects. In Fig. 4.40 some steps of radiolysis of water and recombination reactions are illustrated schematically. In a first step within a time span of 10^{-18} – 10^{-16} s an electron is pulled out of the molecular binding, leaving a positively charged water radical ($\text{H}_2\text{O}^{+\bullet}$), and a free electron.

The released electron (e_{aq}) as a charge in an aqueous environment, immediately becomes hydrated, similar to ions in aqueous solutions (Sect. 2.2.2). As an example of this H_2O^- is depicted schematically in Fig. 4.40. Because of this shell of bound water the life span of this free electron, depending on the pH of the solution, can achieve 600 s. It can interact with various other molecules. Reactions with other water molecules in the hydration shell are possible in the following way:

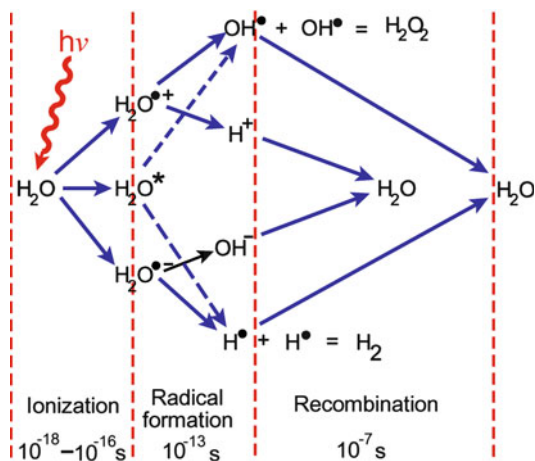
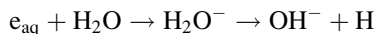
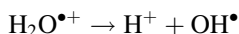
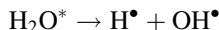


Fig. 4.40 Schematic illustration of the most important reactions of water radiolysis (* – excited states, • – radicals)

The other primary product of radiolysis, the unstable charged radical $\text{H}_2\text{O}^{\bullet+}$, breaks up as follows:



Furthermore, as a primary result of the interaction, an electronically excited water molecule H_2O^* may occur which in a far slower reaction (10^{-13} s) produces two radicals:



The probability of the appearance of these products can be expressed by the so-called *G-value*. It is defined as the number of molecules which become altered per 100 eV absorbed energy. At a pH of the solution in the range between 3 and 10, the following G-values are obtained as a result of ^{60}Co - γ -radiation:

$$G(e_{\text{aq}}) = 2.65; \quad G(\text{H}^\bullet) = 0.55; \quad G(\text{OH}^\bullet) = 2.7$$

This unit corresponds to:

$$1 \text{ molecule}/100 \text{ eV} = 1.036 \times 10^{-7} \text{ mol J}^{-1}$$

As shown in Fig. 4.40 various combinations of radicals may occur, produced by proximate reactions of radiolysis, for example:



Furthermore, various recombinations of these radicals with oxygen are possible, solved in the aqueous phase, which eventually leads to an additional production of H_2O_2 . Furthermore, the oxygen is able to facilitate the transfer of radical sites between molecules or molecular components. As a consequence, oxygen-rich tissue suffers more from irradiation than that with low oxygen content (*oxygen effect*).

It should be noted that these kinds of water radiolysis represent the third way to generate reactive radicals. First, the thermolysis of water was mentioned as the result of local heating through ultrasound cavitation (Sect. 4.3.5). Secondly such radicals were explained as a result of UV-radiation, mediated by photosensitizers (Sect. 4.8.4). These reactive species are together frequently summed up in the all-encompassing designation ROS (*reactive oxygen species*), all being highly reactive due to their content of unpaired valence electrons. This phrase, however, seems to be too imprecise because a greater diversity exists in reactive radicals involved in biological reactions, i.e., those that contain also nitrogen, sulfur, halogens, and carbon.

In fact, free radicals, especially the superoxide radical, are byproducts of the normal cellular metabolism. They are produced inside the mitochondria in the process of oxidative phosphorylation, by membrane-bound NADPH oxidase, or by the nitric oxide synthases. A number of enzymes such as superoxide dismutases, catalases, lactoperoxidases, glutathione peroxidases, or peroxiredoxins are able to regulate this concentration and defend the cells against damage. In fact, this is a complicated system of radical metabolism which by the way is also the subject of cancer research, and many other biomedical topics.

Furthermore, there are a number of intercellular signaling pathways which are based on free radical processes, including ROS and similar species. This explains observations of the so-called *bystander effect*. In this case cells which have not been irradiated, nevertheless show radiation effects, caused by information transfer from irradiated neighboring cells.

In contrast to radicals occurring in normal metabolism, the particular effect of radical species produced by ionizing radiation is caused by the circumstance that they occur directly in the vicinity of biologically important molecules. This especially concerns DNA as a highly charged polyanion, strongly hydrated by water molecules. As a signature of DNA alterations through ionizing radiation, *tandem lesions*, or clusters of DNA lesions occur as the result of various radical-transfer reactions. This means that two or more defects are formed in close proximity on the same DNA strand or even double strand breaks. Even a single OH^\bullet radical is able to induce complex tandem lesions. Because this kind of DNA lesion is untypical in the endogenous processes the efficiency of the cellular repair mechanism is lower.

The hydrated electron reacts primarily with those parts of an organic molecule that have high affinity for electrons such as SH^- , CO^- , NO_2^- , or NO -groups. Either it simply becomes attached to the molecule, thus imparting a negative charge to it, or a dissociative electron capture takes place which, similar to the instances described above, leads again to the formation of a free radical.

In general, two cell defense systems must be considered to cope with free radical DNA damage. They work on very different time scales. On one hand there is a very fast *chemical repair* which occurs at the stage of DNA free radicals. This means that other molecules, so-called *scavengers*, like various thiols containing SH -groups transfer the electron by oxidation. On the other hand an enzymatic repair is possible which is slow, and active if the damage is fully settled. In this respect the double strand helix can be considered as a construction of higher stability against irradiation in relation to the single-stranded DNA helix. This is supported by experiments showing that if the hydrogen bridges between the two molecular filaments are broken by placing the DNA in a solution of urea, then the fragmentation of the DNA by irradiation becomes easier.

Further Reading

Kiefer 1990; Mozumder 1999; O'Neill and Wardman 2009; Sonntag 2006.

4.9.3 Radiobiological Reactions

In this section we leave the narrow region of biophysics and come to the biological effects that are a consequence of the primary processes of radiation chemistry as depicted in the previous section. This leads to the assumption that basically the radiation influences are *stochastic effects*, namely the generation of accidental molecular alterations which are randomly distributed. This in fact is the fundamental problem of low dose irradiation, which may generate genetic defects and cancer. At higher doses of ionizing radiation, leading to acute radiation syndromes, the multitude of these stochastic effects merge to phenomenological, i.e., *nonstochastic* reactions.

Anyway, quantum processes at the molecular level, rather than the total amount of absorbed radiation energy are the reason for the sensitivity of biological systems to ionizing radiation. This can be illustrated easily by considering the real amount of energy which is absorbed in the human body at the lethal dose (LD_{50}) of ionizing radiation which amounts to about 5 Gy, or correspondingly 5 J kg^{-1} (see Fig. 4.42). Considering the specific heat capacity of biological tissue of $4 \text{ J g}^{-1} \text{ K}^{-1}$ this energy absorption therefore leads to a temperature increase of about 0.001 K, an amount that is completely negligible.

Let us first consider the stochastic effects of radiation as the primary biological reactions. As pointed out already in the previous section, the most important effect consists of alterations of the DNA, especially in the generation of double strand breaks. With sequencing of the human genome, the general technological advances in molecular biology, and finally the development of methods of human cell culture, the screening of individual cells for gene defects after irradiation became possible. This allowed detailed investigations of cellular responses even at the low dose range.

It must be emphasized that ionizing radiation induces not only isolated DNA lesions, but also clusters of lesions formed within a few tens of base pairs by a single radiation track. Even a single hydroxyl radical (OH^{\bullet}) may induce such complex tandem lesions. This in fact seems to be the most biologically relevant DNA damage induced by radiation whereas in contrast to this such clustered lesions do not occur endogenously in significant numbers.

As a measure of the radiation effect on DNA, various types of radiation-induced chromosome aberrations are commonly used. These studies indicate that the development of these aberrations in response to ionizing radiation, and the dose–response kinetics appear as a combined effect of radiolysis, and errors in the repair pathways. There are different mechanisms of repair which in fact appear in all phases of the cell cycle. At least for higher radiation doses, nearly all chromosome aberration types show a linear dose effect dependence. At the region of low dose interaction this linearity, however, is a matter of discussion. It may be deflected in both directions mainly by the so-called bystander effect.

As already mentioned in the previous section, the *bystander effect* was established in experiments indicating that radiation effects occur in unirradiated cells in the vicinity of others, hit by radiation. Obviously there exists a direct cell–cell

communication via gap junctions, as well as a mutual influence by the release of various factors from irradiated cells into the medium. This includes reactive oxygen species (ROS), as well as other signaling molecules such as for example various reactive nitrogen compounds. Some of these communications may be effective even over larger distances because several of these species are active for a longer time period. In some cases even removing the medium from irradiated cells, and transferring it to nonirradiated cells induces a corresponding bystander response.

These interactions can be either damaging or protective. Consequently the dose–response curves at low dose may deflect up or down from its linear behavior. In some cases even adaptive responses have been observed, where a small dose of priming radiation reduces the effect of a larger dose, typically when given several hours later.

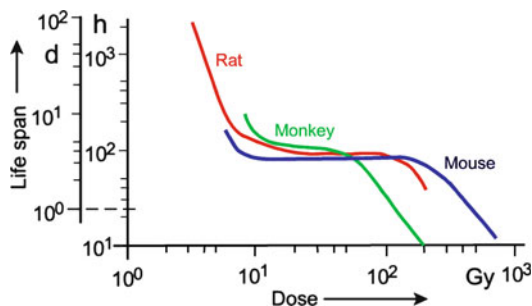
To evaluate the stochastic, i.e., the genetic or carcinogenic effects caused by ionizing radiation, the dose of irradiation must be integrated over the whole life span of an individual, or even over the whole population. The reason for this is not a biophysical, but a genetic mechanism of accumulation. Especially recessive mutations, i.e., such which become effective only in the case where two of them occasionally meet by arbitrary combinations, become dangerous if they become accumulated in the population. The same holds for radiation-induced cancer as a result of somatic mutations.

Radiation effects at the cellular level and even in organs must be characterized as *nonstochastic*, because of the multitude of regulatory processes which respond to ionizing radiation. Already comparatively low doses of radiation reversibly stop the process of cell division. Immediately after irradiation by a sublethal dose the number of dividing cells is reduced. Later these cells undergo mitosis together with the undamaged cells, leading sometimes to a small overshoot of the number of proliferating cells.

In this respect in addition to the influence on the genetic or epigenetic system of the cell, even radiation-induced structural and functional alterations of proteins must be taken into account, especially the behavior of enzymes. The sensitivity of enzymes to radiation shows great variations. Among the most sensitive enzymes are ATPases and catalases. This depends to a large amount on their composition. We already mentioned that proteins with a high percentage of sulfur-containing amino acids are particularly sensitive to radiation. As a consequence of an alteration of protein structures, a number of transport properties of membranes are impaired. These reactions occur only at relatively high doses of radiation, but may have a profound influence on physiological processes, especially on the function of the central nervous system.

All these considerations suggest that juvenile organisms, particularly those in the embryonic phase, are particularly sensitive to irradiation. Damage to the genetic and mitotic apparatus in this period of life can lead to severe malformations or even death of the individual. There are some phases during the development of an organism that are particularly critical in relation to the radiation damage. This, for example, concerns the early stages of cleavage of the ovum up to the morula formation. Radiation damage in these states, however, does not necessarily lead to malformations because the injured cells either die, or are successfully repaired.

Fig. 4.41 Average life span of various mammals as a function of the dose of a single whole-body irradiation with X-rays (After Grodzenskij 1966 redrawn)



Irradiation during the stage when the organs are being laid down is substantially more critical. Exposure during these stages frequently causes malformations. In this context, the development of the central nervous system is particularly sensitive, in contrast to the completely developed brain which is one of the most resistant organs in relation to ionizing radiation.

The complexity of radiobiological reactions helps one to understand the course of acute radiation sickness. Figure 4.41 shows the average life span of some animals as a function of an applied radiation dose. All of the curves similarly show a decline at doses below 10 Gy, then there is a plateau followed by a second decrease at about 80–100 Gy. The shape of these curves reflects two different types of radiation sickness. At low doses the radiation sickness is mainly caused by pathologic alterations in blood-forming organs, which is reflected in dramatic changes in the blood picture. A considerable deficiency of lymphocytes occurs which reduces the resistance of the body to bacterial infections. This can eventually lead to a situation where even the intestinal bacteria become pathogenic. This type of radiation sickness lasts from several days up to weeks and can eventually lead to the death of the individual. The pathological course of this type of radiation sickness is already maximal at a dose of about 10 Gy. A further increase in radiation intensity does not accelerate this process. But if the animals are irradiated with doses over 80 Gy, the central nervous system becomes damaged by a direct influence on the excitation properties of neuronal membranes. In this case the animals die much faster.

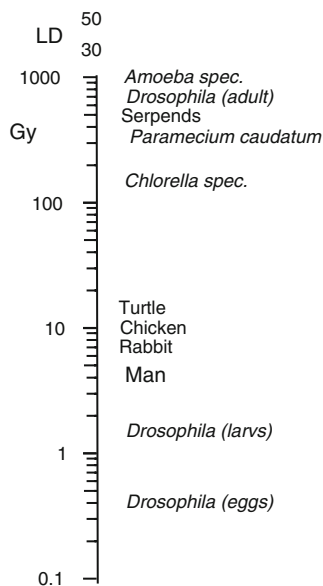
Because of the complexity of radio-biological reactions it is difficult to find benchmark figures as a measure of the resistance of different organisms to irradiation. Similar to the characterizations of poisonous drugs, the parameter LD_{50}^{30} (*mean lethal dose*) is used, i.e., the dose of radiation that causes 50% of the irradiated animals to die within 30 days.

Figure 4.42 indicates that the sensitivity to radiation varies over several orders of magnitude for different species. A comparison of the sensitivity at different ages and stages in the development of a given species clearly shows the embryonic organism to be more sensitive than the adult one.

Further Reading

Heintz et al. 2012, Kiefer 1990; Kogel and Joiner 2009.

Fig. 4.42 Examples for the mean lethal dose LD_{50}^{30} of various organisms after single irradiation with X-rays (Data from Graevskij and Šapiro 1957)



4.9.4 Some Aspects of Radiation Protection

Radiation protection must consider both irradiation from sources outside of the body, as well as from incorporated radionuclides within it. In fact, it is more complicated to evaluate the dose of irradiation caused by accumulated radionuclides, than those from external sources. This concerns on the one hand the diversity of biological effectiveness of the emitted α -, β -, or γ -rays, and on the other hand the pattern of distribution of the corresponding radionuclides in the body, its organs, cells, and molecules.

The investigation of the distribution of natural and man-made radionuclides in the biosphere falls within the field of *radioecology*. This includes their accumulation in various products of the food chain, and finally the degree of resorption, accumulation, distribution, and excretion by man. Empirically obtained data of their distributions, and parameters of their biological half-life times are determined and concentration limits of these nuclides in air, water, and various foods are suggested for legal regulations.

The most important natural radionuclide in the body is the radioisotope ^{40}K , a β -, and γ -emitting nuclide with a half-life time of $1.25 \cdot 10^9$ years. There are 1.2 atoms of ^{40}K for every 10^4 of nonradioactive atoms of potassium. Considering the total potassium content in the human body, this results in about $4.4 \cdot 10^3$ disintegrations per second, i.e., 4.4 kBq.

Another natural radionuclide is the radiocarbon ^{14}C which is continuously produced in the upper layers of the stratosphere by the impact of space radiation. It emits weak β -radiation, and its specific radioactivity, i.e., the relation to the amount of

nonradioactive carbon is lower than that of ^{40}K . Conversely, in contrast to potassium, it is part of all organic molecules, including DNA. The UNSCEAR (United Nations Scientific Committee on the Effects of Atomic Radiation) calculated the annual dose of radiation from ^{40}K as an inherent element of the human body as 0.17 mSv, as a part of the 0.29 mSv of radiation from total ingested radionuclides (Fig. 4.43).

As shown in Fig. 4.43, the largest radiation dose of about 2.4 mSv a^{-1} comes from various natural sources. This concerns the ingestion and inhalation of products of natural radioactive decay, especially the inhalation of the radioactive noble gas radon (Rn) as a decay product of uranium, the radioactive decay products of which eventually, are accumulated in the lung. The data of Fig. 4.43, in fact are global averages, and therefore vary more or less strongly in different regions. The mean value of 1.26 mSv a^{-1} from radon inhalation for example indicates a variation between 0.2 and 10 depending on geological conditions. The value for medical diagnosis also varies between 0.03 and 2 mSv a^{-1} , corresponding to differences in health care. Irradiations from various therapeutic treatments are greater for the individuals concerned, and not included in these data.

This figure indicates the predominance of natural, in relation to that of technical sources of radiation, where the irradiation caused by medical treatments dominates by far. The doses from sources of atomic energy devices are negligible in this relation.

The protection of man against ionizing radiation is based on recommendations of the International Commission of Radiological Protection (ICRP), a nongovernmental organization founded in 1928. The philosophy of these recommendations is the exclusion of any nonstochastic radiation effects, and the minimization of inevitable stochastic effects. These stochastic effects really are the critical point of radiation

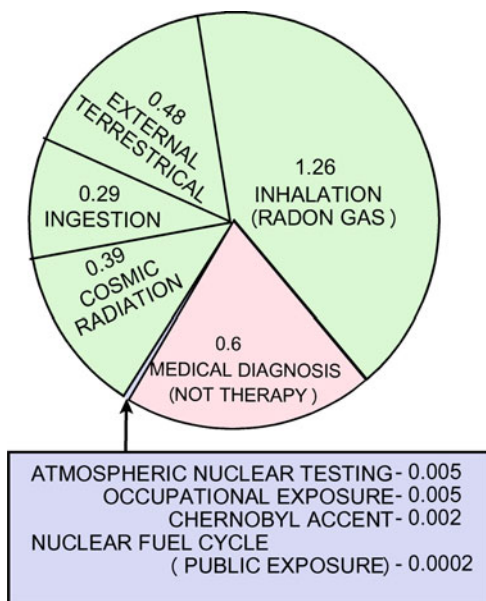


Fig. 4.43 Estimated global annual average dose of humans from natural (*green*), and artificial (*red* and *blue*) sources in mSv a^{-1} (Data from UNSCEAR-Report-2008)

damage. It includes possible genetic aberrations or induction of cancer. In general, one can assume that any increase of irradiation exceeding that of the natural dose of radiation will lead to an increase of the mutation rate. There is no threshold, below which no influence of the mutation rate can be expected. Therefore, all exposure should be kept “as low as reasonably achievable,” a philosophy which is commonly known as the *ALARA-principle*.

Because of the possibility that stochastic effects will be integrated over the whole life span, the limits of maximal exposure are formulated over longer periods of time. Persons with occupational exposure and patients receiving radio-therapeutic treatment are under special supervision. For occupational exposure a maximum yearly irradiation dose of 20 mSv is laid down. For the general population 1 mSv per year is considered to be acceptable. As seen in Fig. 4.43 this is below the amount of natural irradiation which amounts to a mean value of 2.4 mSv a^{-1} , and varies worldwide between 1 and 13 mSv a^{-1} , thus the dose limit applies only to any additional exposure.

Further Reading

UNSCEAR-Reports, recommendations of the ICRP.

4.9.5 Mathematical Models of Primary Radiobiological Effects

Since the first investigations of biological effects of ionizing radiation as early as the 1920s, efforts were made to formulate theoretical approaches, helping to understand the observed phenomena. At this time quantum physics was being developed, which opened up new ways of understanding the interaction between ionizing radiation and biological systems. The starting point was the realization that the total amount of absorbed energy was by far too low to understand the resulting biological effects (see Sect. 4.9.3). Even before anything was known about molecular genetics, biological repair mechanisms, or details of primary molecular reactions the idea was born that particular target effects should occur.

J. A. Crowther, a pioneer of the target theory suggested the picture of “...firing at a swarm of midges with a machine gun,” and he came to the conclusion: “It seems possible, today at least, that the quantum theory must be taken into account in biology as well as in physics, and that a single cell may have a much more direct and painful appreciation of the existence of quanta than is possible to our grosser senses” (Crowther 1926).

This *target theory* can be considered as the first attempt to explain biological phenomena on the basis of stochastic reactions in parallel with the newly developed quantum theory of physics. This eventually proved to become a powerful stimulus to the development of molecular genetics. Even if the knowledge of primary radiation effects, including cellular repair mechanisms indicates a much more complex picture of radiobiological effects, the ideas of these early attempts should not only be seen from a historical perspective but they can still be applied today to solve problems of microdosimetry, and some quantum biological processes, like photosynthesis or scotopic vision, and even for the determination of molecular weights of enzymes.

The starting point of these approaches was the realization that radiation effects are characterized by the following particularities:

- The effects not only depend on the total amount of the absorbed energy, but additionally on the quality of the applied radiation.
- When comparing biological effects obtained by irradiation with rays of the same quality, a dose dependence was found, but not a proportionality between effects and irradiation dose.

These observations have led to the following conclusion: Apparently, there are molecules in living organisms, which must be considered as “neuralgic” points in the system, the destruction of which would produce a sequence of reactions, finally leading to the death of the individual. This sort of primary event which is caused by a reaction with a single energy quantum of radiation was called a “hit” and the corresponding location where the process occurs, a “target.” This leads to the name *target theory*, first applied to suspensions of single cells.

Let us assume that a population consists of n identical individuals, each of them organized physiologically in such a way that a single “hit” would set off a sequence of reactions, leading to death (*single hit process*). According to the laws of probability, with increasing time of irradiation, an increasing amount of individuals would become killed by this *single hit* process. The rate at which the number of living individuals decreases in this way can be expressed by the differential quotient $-dn/dt$. It should be proportional to the number of still-living individuals (n), the time of irradiation (t), and an irradiation constant (σ) which depends on quantity and quality of irradiation:

$$\frac{dn}{dt} = -\sigma n \quad (4.27)$$

Using the initial conditions, where $t = 0$ and $n = n_0$ results in:

$$n = n_0 e^{-\sigma t} \quad (4.28)$$

The reduction of the number of living cells therefore follows a simple exponential function.

In the same way, the number of killed cells (n') can be calculated. Because of the relation: $n' + n = n_0$, we obtain:

$$n' = n_0(1 - e^{-\sigma t}) \quad (4.29)$$

or:

$$\frac{n'}{n_0} = 1 - e^{-\sigma t} \quad (4.30)$$

In Fig. 4.44 the curve for $m = 1$ represents this function. At $t = 0$, $n' = 0$ and therefore: $n'/n_0 = 0$. With increasing irradiation time and doses (σt) the value $n'/n_0 = 1$ will be approached.

In a similar way another model can be calculated. In this case it is assumed that the sequence of processes leading to the death of the individual is only triggered off if the organism has received a minimum number (m) of hits. Blau and Altenburger have calculated the following equation describing this *multi-hit process*:

$$n' = n_0 \left\{ 1 - e^{-\sigma t} \left[1 + \sigma t + \frac{(\sigma t)^2}{2!} + \frac{(\sigma t)^3}{3!} + \dots + \frac{(\sigma t)^{m-1}}{(m-1)!} \right] \right\} \quad (4.31)$$

Some of the curves corresponding to this equation are shown in Fig. 4.44. These functions help to explain experimentally obtained dose-dependent effects.

This theory was extended by J. A. Crowther who introduced a *target area* into this concept. This means he transformed this formal target into a concrete area where the “hits” occur. For this he used the results of experiments with radiation of different wavelengths. As already stated, radiation leaves an ionization trail as it passes through an object. The density of ionizations along this path depends on the character of the radiation, i.e., of its quantum energy. Suppose a “hit” means an ionization of a quite specific molecule or a molecular configuration, then a certain volume must be sensitive to the radiation, which is called the *target area*.

Figure 4.45 schematically illustrates two experiments where a biological object with a given target area is irradiated by two kinds of ionizing radiation. In case A a radiation is used inducing a lower ionization density than in case B. Let us suppose that the biological reaction in this case is triggered by a single hit process, then in

Fig. 4.44 Theoretical curves for one- and multi-hit processes (m corresponds to the number of stars in the “cell”) (After Dessauer 1964)

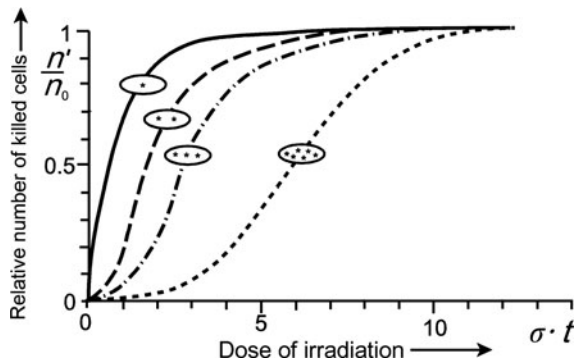
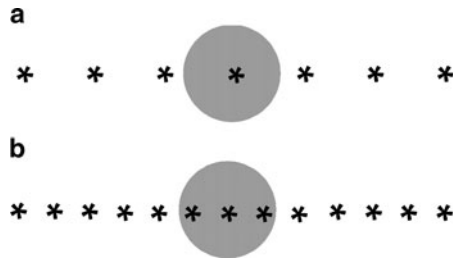


Fig. 4.45 Schematic illustration of the irradiation of a single target area (circular area) with a radiation of lower (a), and higher (b) ionization density. The points mark the locations of ionizations



case B the energy will be squandered because the diameter of the target area is larger than the mutual distance between two points of ionization. This theoretical prediction has been verified by experiments, relating particular dose-effect curves obtained in experiments where different types of radiation were applied.

In a similar way the target theory has been modified step-by-step in adaptation to experimental results. Meanwhile, a considerable number of additional factors have been taken into account influencing the course of radiobiological reactions. For example, it is possible to calculate a multi-hit process, the targets of which have different sensibility. Furthermore, it is possible to take into consideration also some processes of repair. For example, one can assume that in the course of a multi-hit process the efficiency of subsequent “hits” depends on the time period in between. Processes of energy transfer, including for example indirect effects, can be included. Of course any extension of the model leads to the introduction of additional parameters. This finally leads to ambiguous interpretation of experimental results. This seems to be the limit of the extension of these types of calculations.

Nevertheless, the target theory has achieved a certain importance in the analysis of various kinds of radiobiological effects. So, for example, mortality curves for bacteria have been characterized as multi-hit processes. Induction of mutations as a result of irradiation has also been analyzed by this formalism. Some mutations, even those of chromosomes, could be characterized as single-hit processes.

Further Reading

Dessauer 1964; Hug and Kellerer 1966; Kiefer 1990; Zimmer 1961.

Chapter 5

The Kinetics of Biological Systems

This section provides an introduction to the basic ideas of systems theory as applied to living organisms which is an important part of theoretical biophysics. Originally, thermodynamics, as a universal theory of energetic basis of all movements in nature, and kinetics, as a theory of time courses of predicted processes, were considered as two separate theoretical approaches. Thermodynamics answers the questions: what are the reasons for, the driving force of, and the direction of a movement, and what, finally, is the equilibrium situation which will be arrived at if the energetic potentials are equilibrated? Kinetics, on the other hand, just studies the course of a given reaction, its time constants and rates, possible stationary states, stabilities, and oscillations.

Nonequilibrium thermodynamics, evaluating the course of irreversible processes, introduced time as a new variable into the framework of thermodynamics. This was the point when kinetics and thermodynamics became more and more linked. *Synergetics* was the new name proposed to characterize this connection (see Haken 2010).

In recent decades, systems theory experienced an enormous upturn. This was caused by the development of powerful computer systems which are tightly connected to new analytical techniques delivering an enormously increased amount of data and structural insights at all levels of biological organization. This development not only requires methods of storage and managing of this data avalanche but also demands the finding of functional connections, leading to a deeper understanding of complex biological functions. This kind of analysis offers the possibility to analyze existing hypotheses, to check them quantitatively, and to subsequently formulate new concepts for experimental research.

In the framework of this book, and according to its intension, we will impart just a basic understanding of some approaches and ideas of systems theory. In the first part, we will introduce some general approaches to describe the complex network of biological interactions, followed by short introductions to particular applications.

A number of profound textbooks and monographs exist that allow us to go deeper into these theories.

5.1 Some General Aspects of Systems Theory

The basic idea of systems analysis is *isomorphism*. This means that quite different processes in nature, technology, or even sociology, can be formally described by similar mathematical approaches. They show similarities in their properties despite profound differences in the particular details of their elements, or their mechanisms of interaction. For biological systems, this means that there are general principles in the understanding of metabolic and epigenetic, as well as neuronal, networks, and even in systems of ecological interactions.

The general problem in calculating the kinetics of biological systems is their enormous complexity, embedded in a hierarchical structure. For example, the metabolic network of the cell, as complex as it is, is also governed by tissue and organ regulation; this again is controlled by the neuro-humoral system of the organism, and so on. In practice, to answer a particular question, limitations are always necessary. This means defining the system level and limiting the conditions as much as possible.

We will follow here the definition of the term “system” as it has already been introduced in Sect. 2.1.3, and use the depiction of the system properties as explained in Sect. 3.1.1. The definitions of stationary states, and the conditions of their stability (Fig. 3.6), which have been discussed from the thermodynamic point of view in Sect. 3.1.4, are also important here.

Some Textbooks in Biological Systems Theory Alberghina and Westerhoff 2005; Alon 2006; Haken 2010; Heinrich and Schuster 1996; Klipp et al. 2009; Koch et al. 2011.

5.1.1 Basic Equations of Kinetic Processes

Usually, a system is characterized by three types of quantities, which are conventionally named constants, parameters, and variables. *Constants* are quantities which are valid for each natural system, like the Faraday constant (F), the molar gas constant (R), or Avogadro’s number (A). In contrast, *parameters* are the constants just for the particular case. These, for example, can be the rate constants of the observed reactions (k_n), equilibrium constants (K_n), or the Michaelis constants of enzymatic processes. Unfortunately, these quantities also contain the term “constant” in their name, which indicates that this terminology is not obligatory. *Variables* are time-, and probably even space-dependent, quantities which describe the actual state of the system. In thermodynamics (Sect. 3.1), there is not a clear differentiation between the terms “parameter” and “variable” because this can change, for example when considering the same system under isochoric ($V = \text{const}$), isothermal ($T = \text{const}$), or isobaric ($p = \text{const}$) conditions.

In addition, the term *compartment* must be introduced. In some papers, this expression is simply used for a space which is occupied by particular reactants.

For formal kinetic analysis, however, this term must be generalized. Thus, the usual kinetic equations can formally represent the flux rate of a transport process, in the same way as the rate of a chemical reaction, or even some kinds of proliferation rates of cells or organisms. Therefore, a term is needed which can be formally applied to a phase with spatial borders, as well as to a component of a chemical reaction, or even to a population of cells or organisms.

These requirements are fulfilled by the following definition: *A compartment is an element of a system that, with respect to the matter under investigation, has limits which are defined by a specific process, and that contains no gradients within itself.* The term *matter* is used in a generalized form. It can represent a substance in a reaction system, a cell in a proliferation system, an organism in an ecosystem, etc. *Specific process* means the type of the considered movement. It can be an unidirectional flux, as well as the biochemical reaction rate, the proliferation rate of cells, and the birth or death rate of a species. The absence of gradients inside the compartment means that it is to be considered as homogeneous, or *well mixed*. Otherwise, the compartment must be subdivided into smaller compartments, and a multi-compartment system appears. A compartment model is always defined for a special question and is only valid *with respect to the matter under investigation*. In general, it corresponds to the nodes in a complex graph.

The kinetics of a system, i.e. its time-dependent state can be described by the time derivatives of its variables. In this way, differential equations are obtained. If these equations contain only time derivatives, they are called *ordinary differential equations*. If they additionally include derivatives of the variables by space, i.e. by x-, y-, and z-directions, then *partial differential equations* occur.

The simplest reaction is a decay of a substance in a reaction of the first order, the reaction rate of which is proportional to the concentration. The minus sign indicates that the concentration decreases with time.

$$\frac{dc}{dt} = -kc \quad (5.1)$$

transforming this equation into:

$$\frac{dc}{c} = -kdt \quad (5.2)$$

and integrating both sides:

$$\int_0^t \frac{1}{c} dc = \int_0^t -kdt \quad (5.3)$$

it gives:

$$\ln c = -kt + C \quad (5.4)$$

where C is an integration constant. This equation can be transformed into:

$$c = Ce^{-kt} \quad (5.5)$$

It is possible to transform this *general* solution of Eq. 5.1 into a *particular* one, by defining an initial condition, as, for example: at $t = 0$, $C = c_0$:

$$c = c_0 e^{-kt} \quad (5.6)$$

This equation resembles the kinetics of the radioactive decay, or a simple efflux from a compartment. The time at which the concentration is diminished to 50%, i.e. where $c/c_0 = 0.5$, is the *half-life time* ($t_{1/2}$):

$$t_{1/2} = -\frac{\ln 0.5}{k} = \frac{0.693}{k} \quad (5.7)$$

To calculate more complex reactions, the definition of the reaction rate, or velocity (v), is useful. For this simple case, it equals just: $v = kc$, according to Eq. 5.1. For reactions with more components, the degree of advancement ($d\xi$) is a useful parameter (see Sect. 3.1.6, Eq. 3.73) which is a measure of the progress of the reaction ($d\xi = 1/v_i \cdot dn_i$) step by step. Containing the stoichiometric coefficient (v_i) of the particular reactants, it assigns a step of the whole reaction. This allows the defining of the reaction rate of complex reactions as:

$$v = \frac{d\xi}{dt} \quad (5.8)$$

In kinetic equations, the stoichiometric coefficients become signed, depending on the direction considered in the model. In the case of a reaction:



the differential coefficients of the three components can be formulated as:

$$\frac{dc_A}{dt} = -v ; \quad \frac{dc_B}{dt} = -2v ; \quad \frac{dc_C}{dt} = +3v \quad (5.9)$$

In this case, the reaction rate (v) is defined for the reaction, proceeding from left to right. Therefore, the components A and B are diminished if the reaction proceeds in this direction, whereas the component C will be enriched. In biochemical networks with a number of interconnected reactions, the set of stoichiometric coefficients can be listed in matrices, which allow generalized considerations.

The generalized kinetic equation of such a set of reactions is:

$$\frac{dc_i}{dt} = \sum_{j=1}^r v_{ij} v_j \quad \text{for } i = 1, \dots, m \quad (5.10)$$

This indicates that metabolic networks correspond to systems of many differential equations. Their combination results in differential equations of higher orders, i. e. such as with the higher order of derivatives. Furthermore, reactions between several components, i.e. reactions of a higher order in the terminology of chemistry (unfortunately not identical with the mathematical terminology of the order of differential equations!) mostly lead to nonlinear differential equations, containing the product of concentrations.

To understand some basic properties of the steady state, let us consider a simple system, like that shown in Fig. 3.4b (Sect. 3.1.4), where water with a constant flux J is pumped into a vessel, flowing back depending on its filling level. Let the volume of the vessels correspond to the concentration of a component in a chemical reaction. We will formally characterize it by the symbol c :

$$\frac{dc}{dt} = J - kc \quad (5.11)$$

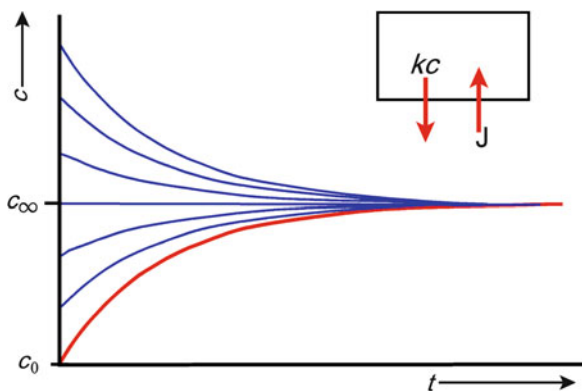
where J indicates the constant influx, and $-kc$ the volume-dependent outflux of the system (see Fig. 5.1). The integration of this equation leads to:

$$c = \frac{J}{k}(1 - e^{-kt}) \quad (5.12)$$

This is already a particular solution of the Eq. 5.11 for the initial condition: at $t = 0$, $c = 0$, i.e. the case where the vessel at the beginning of the experiment was empty. This is represented by the red line in Fig. 5.1. A general solution just leads to an array of curves, some of which are depicted in Fig. 5.1 in blue.

Figure 5.1 indicates that, independent of the initial condition, all curves finally end at the same point (c_∞), thus illustrating a general property of all linear systems, namely the occurrence of only one single stationary state, which depends not on the initial conditions but only on the system parameters. This property is called

Fig. 5.1 An example demonstrating the property of equifinality. It indicates the kinetics of a single compartment system with a constant influx (J) and a concentration-dependent efflux (kc). Some curves show general solutions of the differential equation (Eq. 5.11) (blue lines). The red line corresponds to the particular solution (Eq. 5.12), starting with the concentration $c = 0$



equifinality, which means: all paths, independently of the starting point, lead to the same end.

Extrapolating Eq. 5.12 to $t \rightarrow \infty$, this steady state of equifinality can be obtained as:

$$c_{\infty} = \frac{J}{k} \quad \text{for } t \rightarrow \infty \quad (5.13)$$

The same result can be obtained using the original differential equation (Eq. 5.11) and considering the steady state, which by definition means: $dc/dt = 0$. In this way, even larger sets of differential equations can be transformed into a system of algebraic equations. This makes it possible to calculate the variables $c_{n\infty}$ for the case of the stationary state without further conditions. Therefore, these equilibrium variables only depend on the rate constants (k_{nj}). If all equations are linear, typically only one single solution for each of these values will be obtained.

The property of equifinality occurs only in linear systems. It is easy to understand that solutions of nonlinear algebraic equations obviously provide more than one solution. Some of them may be real, others could just represent mathematical solutions without physical meaning, as, for example, negative concentrations. This multitude of solutions in the case of nonlinear systems corresponds to a system with several stationary states. We will come back to this problem in the next section.

5.1.2 General Features of System Behavior

In general, the behavior of a system can already be predicted by considering the order and the degree of the corresponding differential equations. The *order* of a differential equation corresponds to the order of its highest derivative. The *degree* of an equation is determined by the highest power to which one of its variables, or their derivatives, is raised, provided that the equation has been arranged in such a way that only positive powers of the dependent variables, or their derivatives, occur. The product of two variables has to be considered as a kind of second power. The thermodynamic characteristics of nonlinear processes have already been discussed in Sect. 3.1.4. We will complete these considerations by some kinetic circumstances.

To illustrate this, let us consider a simple mechanical system. Suppose there is a spring hanging from a rigid support (Fig. 5.2). At the lower end of the spring is a pointer which marks the length (l) of the spring. A certain force (F) holds the spring under tension and is in constant equilibrium with the pulling force of the spring according to Young's law ($k_1 l$) (see: Sect. 3.6.3, Eq. 3.226). If the force is suddenly changed (blue line in Fig. 5.2), the spring elongates (red line). Because of a brake which is included in the construction, a fictional force ($F_{\dot{l}}$) occurs, which is taken to be proportional to the rate of displacement (dl/dt) by the factor k_2 . In this case, the action of the force (F) is compensated by the fictional force ($F_{\dot{l}}$) as well as by the

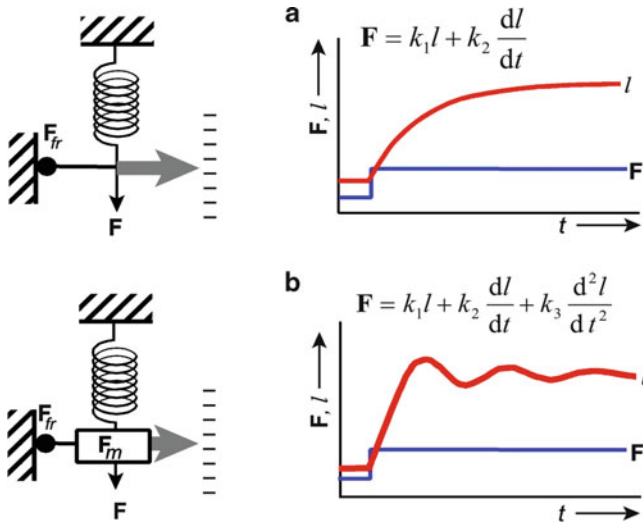


Fig. 5.2 Mechanical constructions illustrating the kinetic behavior of systems of first (a), and second (b) order as response (red lines) to a rectangular function of the extending force F (blue lines)

extension force of the spring. Thus, this system is described by a linear differential equation of the first order (Fig. 5.2a). Its solution and integration leads to a function $l(t)$ as a simple exponential expression. It shows a simple continuous approach to the equilibrium, corresponding to the empirically expected behavior.

Completing this construction by an additional mass (m), which imparts a certain amount of inertia to the system, a further summand must be included in the function (Fig. 5.2b). An inertial force (F_m) occurs, which is proportional (k_3) to the acceleration, i.e. to the second time derivative of the length (l), according to Newton's law. In this way, a differential equation of the second order appears. It is easy to imagine that this second order system will show oscillations of the function $l(t)$ as it settles down, similar to pairs of scales. These oscillations can be made to slow down at different rates by the use of various damping devices, i.e. changing the relationship between the parameters k_2 and k_3 .

A further complication of the behavior of this system would occur at increasing force (F). In this case, the friction becomes a nonlinear function of the elongation velocity (dl/dt) (see also Sect. 3.1.3, Fig. 3.3).

Non-linear systems can exhibit some further kinds of behavior. Figure 5.3 shows in principle how a given variable c in a system can change when it is displaced by an amount Δc from its stationary value. When the system is stable, it returns to its original state. This can occur without oscillations (black line) at a high degree of damping, i.e. a low time constant of reaction. But increasing the time constant above an aperiodic limiting condition causes the oscillations to die away asymptotically (red line). With periodic limiting conditions, non-damped oscillations are

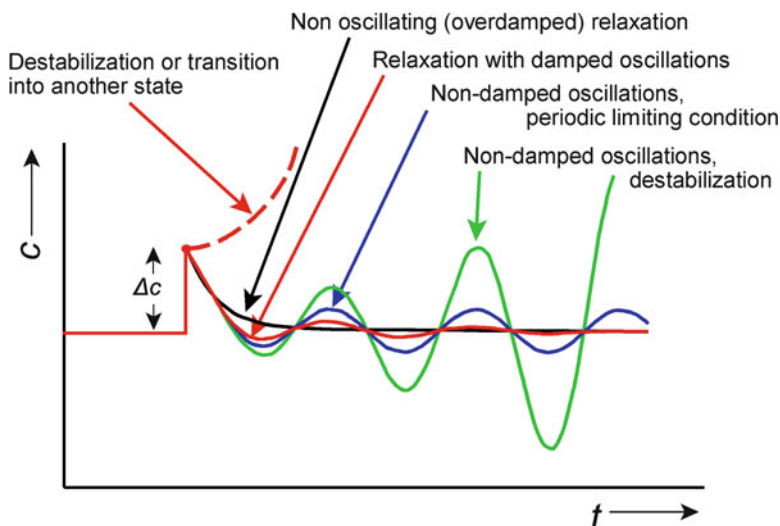


Fig. 5.3 Possible kinds of system behavior, illustrated as a time course of the variable $c(t)$ of a system after a particular displacement Δc

obtained (blue line). Disturbances of the system may also lead to instabilities, possibly leading to another stationary state (dashed red line, and green line).

In the case of more time-dependent variables, two methods are possible to illustrate the system graphically. If, for example, a system contains two variables, c_1 and c_2 , it is possible to characterize it by plotting both curves, $c_1(t)$, and $c_2(t)$, separately (Fig. 5.4, left part). At any moment (t), a value of the variable c_1 corresponds to a particular value of c_2 . The state of the system, therefore, can also be represented by a graph, where c_2 is plotted against c_1 . At the time t_B , for example, the state of the system corresponds to the point B on the righthand graph. Analyzing the functions $c_1(t)$ and $c_2(t)$ on the left plot at the moments $t_A, t_B, t_C, t_D, t_E \dots$, it is possible to construct a *trajectory* following the points A, B, C, D, E... in the graph on the righthand side of Fig. 5.4. This represents a particular change in the system. When there are only two variables, the trajectory is just a line in the plane. The arrow of the trajectory (green line) indicates its time course, showing where the system proceeds.

In systems with n variables, an n -dimensional space is required to depict the behavior of the whole system. In fact, it is a vector space, which means that at each point in it, a signpost is located, indicating the direction of the further direction of system development.

Let us firstly characterize the general kind of system behavior in the case of two variables (n_1, n_2), i.e. in a two-dimensional plane (Fig. 5.5). Later, we will demonstrate this for two particular cases: for a biochemical reaction model (Sect. 5.2.1, Fig. 5.12), and for the depiction of population dynamics (Sect. 5.2.3, Fig. 5.16). Here, we will just discuss some principles of system behavior in a more general way.

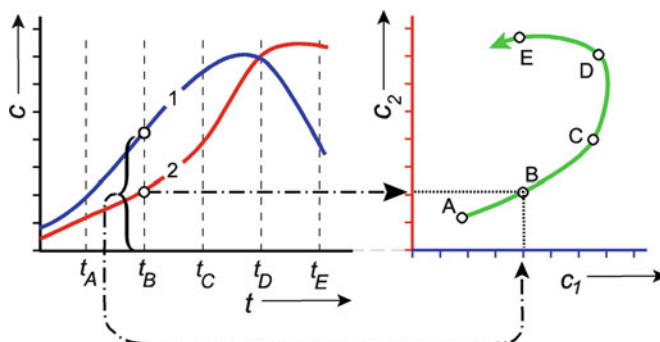
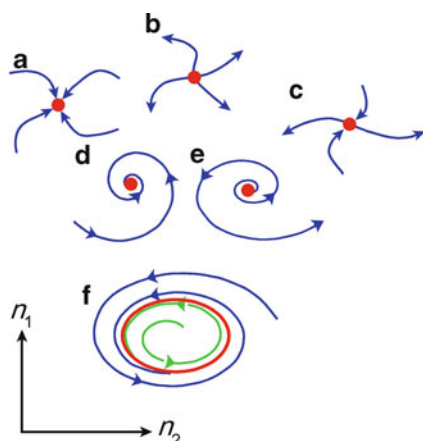


Fig. 5.4 Plot of the time dependent functions of two variables $c_1(t)$ (blue), and $c_2(t)$ (red) on the left, and the depiction of the trajectory (green) of these variables $c_2(c_1)$ on the right (the dash-dotted arrow lines illustrate the method of construction of the trajectory)

Fig. 5.5 Trajectories in a two-dimensional space to characterize various states of the system: (a) stable node, (b) unstable node, (c) saddle point, (d) stable focus, (e) unstable focus, (f) stable limit cycle



The curves in Fig. 5.5 were obtained in the same way as the trajectory in the righthand part of Fig. 5.4. They are the kinetic completion of the energetic schemes in Figs. 3.5 and 3.6, and represent numerical solutions of differential equations, characterizing the kinetic behavior of complete systems.

At first, these presentations show whether the systems indicate a kind of *stability* as an answer to a small perturbation. The curves A, D, and F show reactions of stable systems. The arrows demonstrate a return of the systems to their origin after the perturbation. In contrast to this, the curves B, C, E show that even a small deviation from the starting point produces a further shift, thus a destabilization of the system (see also Fig. 3.5 in Sect. 3.1.4). The curve F represents a system with stable permanent oscillations. In this case, the stability arrives at a *limit circle* which is a sign of undamped oscillations. Metastable states (see Fig. 3.5), which are not plotted in Fig. 5.5, would be characterized by an arrow, leading from one point to another.

In addition, these figures demonstrate different ways of arriving at the stable state. Case A indicates an aperiodic approximation, as shown in Fig. 5.3. The approach by damped oscillations in Fig. 5.3 is represented in case D of Fig. 5.5, and the undamped oscillation corresponds to case E in Fig. 5.5. The distortion of the circle into an ellipse (case F) simply depends on the chosen coordinates. In the case of oscillations superimposed by several frequencies, this ellipse will be further deformed.

A further important property of system behavior can be demonstrated more easily in a three-dimensional system (Fig. 5.6). Because of causal connections between these three parameters, the only possible properties of the system form a *control surface*. This can be best illustrated by the physicochemical properties of water. Take z to be the viscosity or the structure of water, which depends on pressure (x) and on temperature (y). When the temperature falls (trajectory A in Fig. 5.6), the system will reach a *bifurcation point* at 0°C at normal pressure (point P in the green projection). It can either freeze, i.e. move in the lower surface to the point C, or remain as a super-cooled liquid in the upper surface and move to B. The supercooled liquid, however, will spontaneously freeze (trajectory B–C) if there is any disturbance of the system (alteration of pressure, vibration, etc.).

In 1969, the French mathematician Rene Thom developed a comprehensive topological theory of dynamic systems which primarily enables the occurrence of multi-stationary states to be systematized. He called a sudden change from one state to another a *catastrophe*. The case of Fig. 5.6 was classified as a *cusp catastrophe*. This system is characterized by the conditions that, in most cases, there is only one point in the z -coordinates that corresponds to a point in the x,y -plane. At the folded region of the control surface, however, three possible values of z occur. At some points in the x,y -coordinates, the system can therefore obtain three different states.

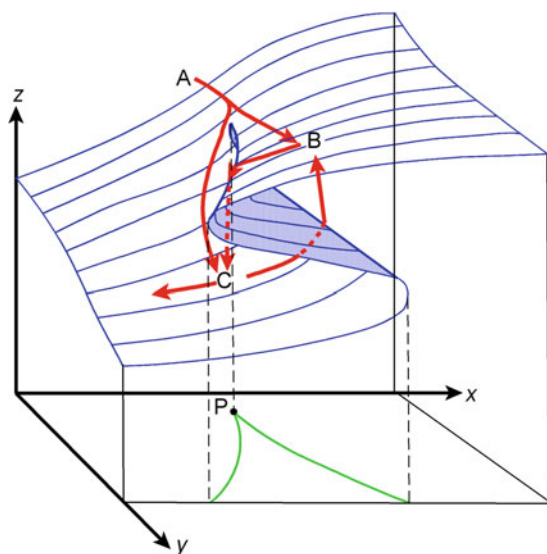


Fig. 5.6 A control surface (blue area) of a system with three variables: x , y , z , as an example of a cusp catastrophe. (A) Trajectory with bifurcation point (P); (B, C) trajectories with a phase jump ("catastrophe") near the borderline of stability; shaded area: region of unstable states

The points on the middle plane (the shaded area in Fig. 5.6), however, indicate unstable states. This can be compared with the thermodynamic picture of a metastable state (Fig. 3.5), i.e. an unstable state between the two stable stationary states.

With respect to the isomorphism of the systems of different natures, these kinds of cusp catastrophe can frequently be observed in various concrete examples. In some kinds, not only the catastrophe B, C (Fig. 5.6) is possible but also the opposite movement (not observed in the case of water).

In Sect. 5.2.1. we will characterize the behavior of a biochemical reaction chain with such a function (Fig. 5.12b). Furthermore, the electrical potential of a nerve cell was explained in this way. Even examples in the behavioral biology can be found, as, for example, the reaction of a dog, which as a result of irritation can either run away (i.e. the upper surface), or attack (lower surface). The running away in panic can suddenly change to an attack or vice versa.

In recent decades, a type of system behavior was investigated which is called *chaotic*. This means an obviously irregular and non-predictable behavior, which nevertheless is the result of the interaction of deterministic processes. Nonlinear systems with several stationary states can jump from one state to another. Sometimes, oscillations occur, but these oscillations may degenerate into chaotic movements. The turbulent flow, which we discussed in Sect. 3.7, may be taken as an example of a chaotic process. In medicine, the disturbance of cardiac cycle is analyzed by approaches of the chaos theory.

Further Reading

Heinrich and Schuster 1996; Thom 1969; Zeeman 1976.

5.1.3 *The Graph-Theory as an Topological Approach to Describe Complex Network Systems*

Biological systems are characterized by an enormous number of interactions on each level of organization. A large number of subcellular and neuronal machines are known, such as metabolic circuits and signaling networks containing hundreds to thousands of different types of components with particular kinds of interactions. The description of all of these processes by particular differential equations leads to a system which is not only impossible to solve analytically but which in addition contains an enormous number of unknown constants, so that even numerical solutions will become quite ambiguous.

To overcome this problem, some generalized approaches of topological mathematics are used, in the end based on *graph theory* which had already been introduced in 1736 by the German mathematician Leonhard Euler. He applied this method at that time to an actual problem of his students. They tried to find a walk that would cross each of the seven bridges of the river Pregel in Königsberg exactly once and return to the starting point. Graph theory first looked like a futile bauble, but nowadays it appears as an important instrument to investigate complex networks of interaction.

A *graph* is a mathematical structure consisting of a number of discrete nodes, linked by a network of connections. It differentiates between *arcs* (or *interactions*) and *edges* (or *links*) (see Fig. 5.7). Arcs, graphically represented by arrows, indicate ordered pairs of nodes. In biochemical networks, they demonstrate, for example, direct interactions between proteins, transcription factors, gene regulation effects, or enzyme substrate interactions. Some of these interactions can occur in a reversible way, in the terminology of chemistry (see Sect. 3.1.4), while others are unidirectional. Edges are represented graphically by single lines. They stand for unordered pairs of nodes, for links indicating a certain influence, but without a clear direction.

A graph, the nodes of which are connected by arcs, is called a *directed graph* (or *digraph*). Cell signaling pathways are commonly represented by mixed graphs. In this case, activation or inhibition processes are represented by arcs, whereas physical protein–protein interactions without a clear-cut directionality are classified as edges. In some cases, processes of self-inhibition or even self-activation are possible. Simple graphs like that in Fig. 5.7, without intersecting edges, allow a planar depiction. In the case of complex networks, nonplanar representations are necessary.

The number of interaction arrows of a particular node is defined as its *degree* (or *size*). This is the most fundamental measure of a network. In the case of reversible reactions, it can be separated into *in-degrees* and *out-degrees*. For a directed graph with m edges and n nodes, an average degree is defined by the number of m/n . Furthermore, a *connection density* of a graph can be calculated, which is the proportion of the actual number of edges in relation to the number of possible ones. This is the simplest estimator of the physical cost, namely of the required energy or other resources.

The degrees of all the nodes show a function of *degree distribution*. In random graphs, each pair of nodes has an equal probability of connection. Therefore, all connections are distributed according to a symmetrical Gaussian function. Most graphs describing real-world networks, however, significantly deviate from the

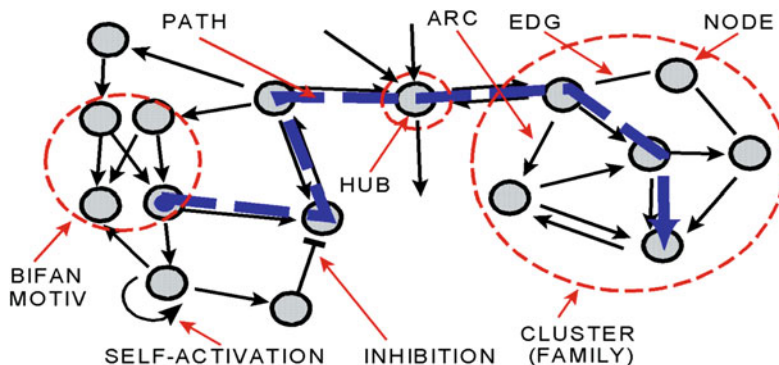


Fig. 5.7 Illustrations of some elements and properties of a graph

simple random-graph model. The analysis of them in some cases shows that the probability function of the number of nodes with a particular degree follows a power function with a negative exponent which often ranges between -2 and -3 . This very gradual decay of the degree distribution implies that the network lacks a characteristic scale. These kinds of graphs are therefore called *scale-free*.

In contrast to this, in biological systems, most graphs occur with high levels of local clustering among nodes even forming particular modules as *families* (or *cliques*). This type of network is called a *small-world* graph. Evidence for small-world properties has been found in most genetic, metabolic, and neural networks. These networks therefore deviate from randomness and reflect a kind of specific functionality. Some families representing small circuits in the complex graph show structural similarities, and collectively are termed *network motifs*. In fact, these are subgraphs that are over-represented in the network. Small motives, for example, are particular feedback loops. In gene regulatory networks, *bifan motifs* occur, which are comprised of two upstream regulators both regulating the same two downstream effectors (see Fig. 5.7).

The degree of clustering of a network can be quantified by the *clustering coefficient*, as a proportion of the number of connections that exist between the nearest neighbors of a node in relation to the maximum number of possible connections. The clustering is associated with high local efficiency of information transfer and robustness.

A further parameter in functional graphs is the *path length*. This is defined as the minimum number of edges that must be traversed to go from one node to another. The efficiency of a function is inversely related to its path length. In some cases, it is also possible to estimate the topological distances between nodes of a graph by the functional characteristics. Central nodes, i.e. those with a high number of the shortest paths crossing them, are called *hubs*. Many complex networks can therefore be divided in a number of modules which frequently indicate a hierarchical structure. Usually, each module contains several densely interconnected nodes, whereas there are relatively few connections between nodes of different modules.

A particular limitation of graph theory applications in analyzing biological networks is their dynamics. The nodes and links of biological networks change with time. The abstraction of developing and self-modifying biological networks by graphs with permanent structures sometimes masks temporal aspects of information flow. Modeling the dynamics of these networks to develop more quantitative hypotheses provides a closer insight into the system's behavior.

In summary, the graph analysis of biological networks has been useful to obtain an overview of the organization, and of different types of reactions, and it suggests a number of topological patterns of behavior. Furthermore, some frequently obtained clusters in metabolic systems have already been analyzed in detail, and can be introduced as a more or less clarified module in the network. These approaches allow conclusions on the dynamical behavior of a system, including its stability, noise filtering, modularity, redundancy, and robustness to failure, as well as variations of kinetic rates and concentrations.

Further Reading

Bullmore and Sporns 2009; Lima-Mendez and Van Helden 2009; Ma'ayan 2009.

5.1.4 Systems with Feedback Regulation

In 1948, Norbert Wiener published his book: “Cybernetics – or Control and Communication in the Animal and the Machine.” In it, he recommended applying experiences with engineering control systems to understand various processes of physiological and neuronal regulations. Meanwhile, systems theory uses various other possibilities for mathematical modeling of biological processes, but for some considerations, especially in various feedback systems, these cybernetic approaches nevertheless seem to be a useful addition. Generalized feedback systems are mostly applied to understand neuromuscular systems and sensory processes in neurophysiology. Especially in the borderline between physiology and techniques, namely in various branches of medical techniques, these systems engineering tools are common. In some cases, these engineering control system approaches are applied even in processes of cellular regulation.

The basic elements of a feedback control system, and their terminology in the language of engineers, are depicted in Fig. 5.8. In its simplest form, it consists of two subsystems, the *controlled system* and the *controller*. Both of them convert one or more input signals into output signals corresponding to their *input–output law*, which in the simplest case is governed by a system of linear differential equations.

The controller may influence the controlled system in the sense of stimulation, or inhibition, producing in this way a positive, or a negative, feedback. The feedback can work by just triggering, increasing, or inhibiting a process, or it may be part of a sensitive control system. In this case, it adjusts a *controlled variable* $y(t)$ to a given standard which is specified by a *command variable* $z(t)$, in spite of a supplementary effect of a *disturbance variable* $x(t)$. If the command variable remains constant, then the network is called a *retaining control circuit*. However, if it is a function of time, which means that the system has an additional control over it, this is a *follow-up control circuit*. The command variable, the controlled variable, and the disturbance variable may differ from each other in their physical character, both in quality and quantity. The regulation of blood pressure (a controlled variable), for example, on the one hand can be controlled through a sequence of nerve impulses acting as the manipulating variable, while on the other hand, it could also be regulated and controlled by the concentration of a hormone in the blood. Hence, the control signals can be classified as either discrete (sequence of nerve impulses) or analog (hormone concentration).

The difference between the command variable and the controlled variable is called *control deviation* $[\Delta y(t)]$. Depending on the type of the controller, the *manipulated variable* $z(t)$ can be calculated in two ways, corresponding to two types of controllers:

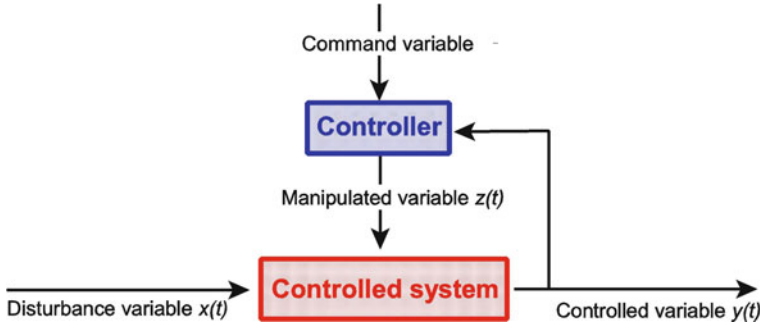


Fig. 5.8 Terminology of the elements of an automatic control system

- In the case of proportional controller (*P-controller*):

$$z(t) = k_p \Delta y(t) \quad (5.14)$$

- In the case of integral controller (*I-controller*):

$$z(t) = k_i \int \Delta y(t) dt \quad (5.15)$$

The P-controller, therefore modified the variable $[z(t)]$ proportional to the control deviation $[\Delta y(t)]$. This type of controller is not able to completely compensate an induced disturbance of the system. As illustrated in Fig. 5.9, there always remains a proportional deviation (*P*), because, if $\Delta y = 0$, then according to Eq. 5.14, $z = 0$.

This problem does not arise with the I-controller. In this case, the modified variable is proportional to the time integral of the control deviation. If the system is brought back to its starting positions, then the parameter z in Eq. 5.15 becomes constant, which means that an optimal compensation for the disturbance variable is achieved.

In biological systems, P- and I-controllers usually exist in combination with each other or are combined with a differential controller (*D-controller*).

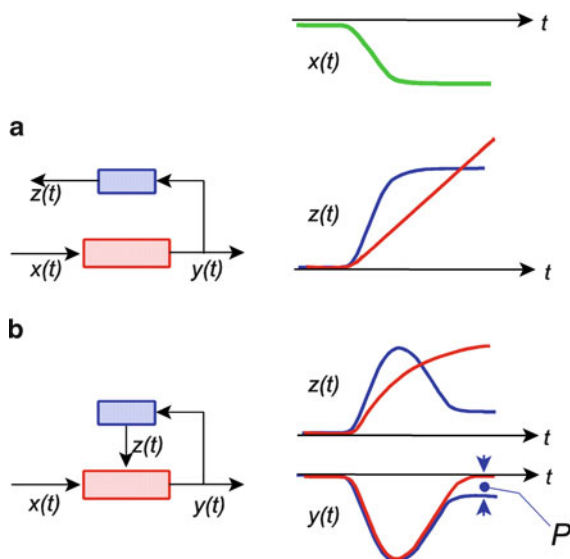
For a D-controller, the following equation holds:

$$z(t) = k_D \frac{d[\Delta y(t)]}{dt} \quad (5.16)$$

This type of controller does not respond to the control deviation but to the rate at which it is changing.

Neither the controller nor the controlled system respond without inertia, i.e. with a certain time delay. In Eqs. 5.14, 5.15, and 5.16, terms have to be included that take into account the time dependence of the adjustments. This will result in differential

Fig. 5.9 Characteristic responses of an open (a), and a closed (b) control loop with P-controller (blue lines), and I-controller (red lines) on a disturbance variable $x(t)$ (green line). P proportional deviation



equations of the n th order. The system therefore behaves like those demonstrated as mechanical examples in Fig. 5.2. This means that the curves representing the functions $y(t)$ and $z(t)$ in Fig. 5.9 are in fact more complex than shown. Curves like those illustrated in Fig. 5.3 will more likely result. At least there are no multi-stationary states to be expected with linear systems. The amplification depends on the factors k_p , k_i and k_D in Eqs. 5.14, 5.15, and 5.16, whereas the attenuation correlates with the time constants of the rates at which the systems make necessary adjustments.

The sooner and stronger a system reacts to a disturbance, the faster it will achieve the required correction. However, because control circuits are systems of higher order, the compensation can overshoot. If a system is deflected by a step-wise function, then the following types of system behavior can occur as the amplification is increased (see also Fig. 5.3):

- At a low degree of amplification, up to the aperiodic limiting condition, the control variable approaches the target correction value asymptotically without an overshoot.
- As the amplification increases, exceeding the aperiodic limiting condition, the system starts to oscillate. It reaches the target correction value faster, but overshoots it and is eventually adjusted by damped oscillations. As a result, it may take longer to achieve a stable value than in the first case.
- A further increase in the amplification leads to the periodic limiting condition whereby undamped oscillations around the correct parameter occur.
- If the periodic limiting condition is exceeded because of a further increase in the amplification, then the system becomes unstable. The oscillations get bigger and bigger and cause the system to become increasingly more deflected.

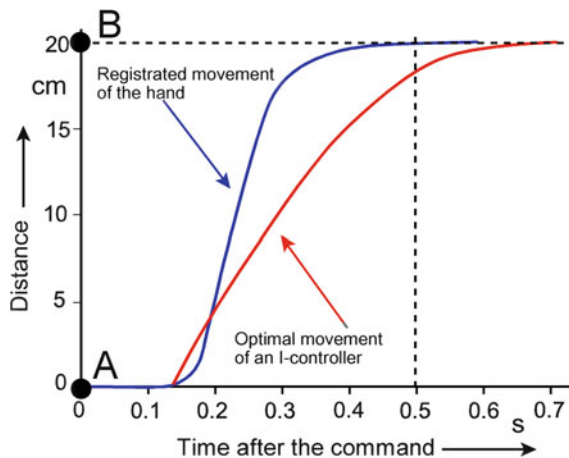
Such instabilities can lead to the destruction which, however, in biological systems is usually limited by the available energy.

These properties illustrate the factors involved in optimizing a control circuit. The fastest possible rate of attaining the target correction value, of course, can to be considered as an optimal condition. This, in fact, cannot be done by simply increasing the amplification. It is best achieved by a combination of different control systems.

Such a kind of complex regulation in the first period of biocybernetic research had already been demonstrated for the case of target-oriented movement of human hand: A person is asked at a given command to immediately move a pointer from position A to position B, whereas the resulting movement is recorded as a function of time (blue line in Fig. 5.10). After a lag period of about 145 ms, the hand starts to move, and reaches the target 500 ms after the signal, without oscillation. This behavior cannot be achieved by a simple I-controller. The aperiodic limiting condition, i.e. the fastest possible reaction of this controller without oscillations, is also shown in the diagram (red line). In fact, the target-oriented movement of the hand is controlled by two processes. First, the distance the hand has to travel is estimated through the eyes, and a rapid movement with a built-in limit is programmed. This takes up the observed lag period of 145 ms. Only during the last phase of the target-oriented movement, do the eyes correct the hand, a process that can be regarded as being performed by an I-controller with a time constant of 47 ms. This model is not only mathematically consistent with the observations but can also be verified by temporarily preventing the subject from seeing what they are doing during the experiment.

Cybernetic approaches to understand physiological reactions, as demonstrated in Fig. 5.10, were chiefly used in the decades after the publication of Norbert Wiener's book. Nowadays, these engineering approaches are directed away from basic biological mechanisms and toward prosthetic applications, and the optimal use of

Fig. 5.10 The course of moving the hand (blue line) after an order, given at time $t = 0$, to move to a point B with a distance of 20 cm from the starting-point A. The red line indicates an optimal reaction of an I-controller with a lag period of 145 ms and an amplification up to the aperiodic limit (Data from Hassenstein 1966)



neural information for their control. Feedback systems in metabolic and neuronal processes are mostly treated by direct mathematical modeling.

Further Reading

Flanders 2011; Rudall 2004.

5.2 Model Approaches to Some Complex Biological Processes

After introducing general approaches of systems analysis, and some general properties of complex systems, this section will give insights into the theoretical treatment of some particular processes. The enormous increase in research in this field, however, only allows to outline these considerations, simply showing their basic ideas and models.

5.2.1 Models of Biochemical Reactions

The first step in investigating biochemical reactions is the construction of the net of interactions as described in Sect. 5.1.3. To concretize these calculations, the particular edges and arcs must be represented by equations. In some cases, the conventional equations of chemical reactions of the first or second order are applicable. Usually, however, biochemical reactions are catalyzed by enzymes. In its simplest form, an enzymatic reaction may be described as follows:



A substrate S forms a complex ES with the enzyme E . This complex then breaks down to give the product P and release the enzyme E . The energetic advantage of these reactions has already been discussed in Sect. 2.1.5 and illustrated in Fig. 2.5. Here, we will concentrate on the kinetic consequences.

Usually, the rate-limiting step of this sequence of reactions is the break-down of the enzyme–substrate complex. Therefore, the rate of this reaction can be described in the following way:

$$J = k c_{ES} \quad (5.17)$$

where k is the rate constant of this break-down reaction. Further assuming that the first step of this reaction, i.e. the formation of the enzyme–substrate complex is always in thermodynamic equilibrium, then by the mass action law the equilibrium constant (K_M) can be evaluated:

$$K_M = \frac{c_S c_E}{c_{ES}} \quad (5.18)$$

The total concentration of the enzyme (c_{E0}) must remain constant.

$$c_{E0} = c_E + c_{ES} \quad (5.19)$$

Combining Eqs. 5.18 and 5.19 one obtains:

$$c_{ES} = \frac{c_S c_{E0}}{K_M + c_S} \quad (5.20)$$

Introducing this into Eq. 5.17, one gets the *Michaelis–Menten equation* which describes the reaction rate (J) of simple enzymatic processes as a function of the substrate concentration (c_S):

$$J = \frac{k c_S c_{E0}}{K_M + c_S} \quad (5.21)$$

K_M , as the dissociation constant of the enzyme–substrate complex, is known as the *Michaelis constant*. It has the units of concentration, and corresponds to the concentration of the substrate when the reaction rate is at half of its maximum value (see Fig. 5.11). It usually lies somewhere between 10^{-2} and 10^{-5} M.

As shown in Fig. 5.11, there are two particular regions in this function where some simplifications are allowed. At high substrate concentrations ($c_S \gg K_M$), a maximum reaction rate J_{\max} is attained. For this case, Eq. 5.21 leads to:

$$J_{\max} = k c_{E0} \quad (5.22)$$

In this case, the flux is independent of substrate concentration (c_S). J_{\max} can be introduced as a constant parameter into the general system of equations.

In contrast, at very low substrate concentrations ($c_S \ll K_M$), Eq. 5.21 becomes:

$$J \approx \frac{k c_S c_{E0}}{K_M} = \frac{J_{\max}}{K_M} c_S \quad \text{for: } c_S \ll K_M \quad (5.23)$$

In this case (dash-dotted line in Fig. 5.11), a linear flux equation similar to Eq. 5.17 can be used. In many particular situations, therefore, simple approaches

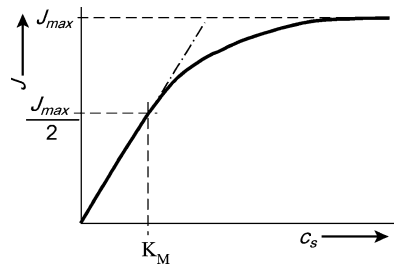


Fig. 5.11 The reaction rate (J) as a function of the substrate concentration (c_S) of an enzymatic reaction with constant enzyme concentration

are possible even in cases of enzymatic reactions. In all other cases, the Michaelis–Menten equation must be directly introduced into the system of flux equations. Obviously, this leads to nonlinear equations.

In fact, enzymatic reactions in metabolic systems are frequently regulated in a more complex way. A number of effectors may bind to the enzyme and interact as inhibitors or activators. Even the product of the enzymatic reaction itself can interact as a positive (product activation) or negative (product inhibition) feedback system. In a similar way, the substrate at higher concentration may occupy a second binding place of the enzyme, forming a ESS-complex that cannot form the product P. As the consequence, the Michaelis–Menten function (Fig. 5.11) with a final reaction rate J_∞ will be modified to a curve like the red line in Fig. 5.12a, which finally arrives at a lower level.

Some consequences of complex enzymatic systems can be demonstrated by a small reaction chain with nonlinear behavior and some stationary states (Fig. 5.12). A product (P) is formed by an enzymatic reaction from a substrate (S). The substrate (S) itself is produced by a first order reaction from a substance A. It can, however, break down with the same kind of kinetics to Z. The time dependence of the concentration of the substrate (c_S) can be formulated as follows:

$$\frac{dc_S}{dt} = J_{AS} - J_{SZ} - J_{SP} \quad (5.24)$$

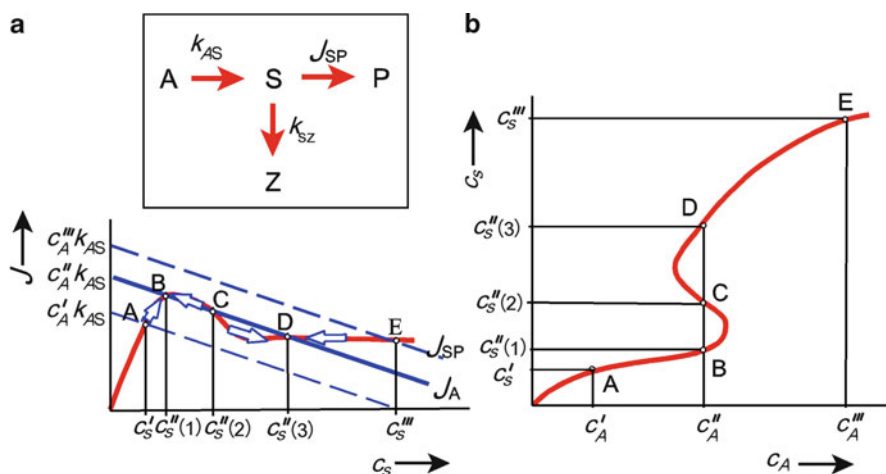


Fig. 5.12 A kinetic model of a simple biochemical reaction chain (insert). (a) The reaction rate (J_{SP}) (red line) as a function of the substrate concentration (c_S) for the enzymatic reaction $S \Rightarrow P$ with substrate inhibition, and the difference of fluxes: $J_A = J_{AS} - J_{SZ}$ (blue lines) at various concentrations of the substance A (c_A' , c_A'' , c_A'''). The intersection points (A, B, C, D, E) of the blue curves with the red indicate stationary states. The arrows are directed towards stable states. (b) The trajectory of the stationary states of the left picture as a function $c_S(c_A)$

Let the flux J_{SP} , which denotes the rate of the enzymatic destruction, be dependent on c_S in a kinetics with substrate inhibition as explained before. The function $J_{SP}(c_S)$ is illustrated in Fig. 5.12a. The fluxes J_{AS} and J_{SP} shall be determined by linear approaches according to Eq. 5.1. In this case, Eq. 5.24 can be transformed in the following way:

$$\frac{dc_S}{dt} = k_{ASC_A} - k_{SZC_S} - J_{SP} \quad (5.25)$$

Introducing formally a flux:

$$J_A = J_{AS} - J_{SZ} = k_{ASC_A} - k_{SZC_S} \quad (5.26)$$

then Eq. 5.25 transforms into:

$$\frac{dc_S}{dt} = J_A - J_{SP} \quad (5.27)$$

For the case of a stationary state, the following condition must apply:

$$\frac{dc_S}{dt} = 0 \quad \text{and therefore} \quad J_A = J_{SP} \quad (5.28)$$

The function $J_A(c_S)$, as defined by Eq. 5.26, is a straight line with a slope of $-k_{SZ}$, intersecting the ordinate at a point k_{ASC_A} (Fig. 5.12a). The stationary states of the system according to Eq. 5.28 are marked by the points, where the blue lines of the functions J_A are cutting the red curve of the function J_{SP} .

Let us first consider the case $c_A = c_A''$. This straight line intersects with the function J_{SP} at points B, C and D. This means that there are three stationary states for one single value of c_A . However, these points are not equivalent to each other. Only states B and D are stable, C being unstable. This can easily be deduced from the curve.

First, consider the point B: a small increase in the substrate concentration will lead to the situation where function $J_{SP} > J_A$. Reference to the balance equation (Eq. 5.27) will show that this will bring about a reduction in c_S . The system, therefore, is moving towards its initial state. If c_S'' is reduced slightly, then $J_{SP} < J_A$, and this results in an increase in c_S . Small deviations from the stationary value B will thus be corrected. This behavior is indicated by the arrows on the function drawn in Fig. 5.12a. Around the point D, the situation is the same.

Near the point C, the system behaves differently. Small reductions in c_S'' cause J_{SP} to become larger than J_A . The substrate concentration will continue to fall, until point B is reached. An increase in the value of $c_S''(2)$ by a small value will cause the system to slide in the direction of state D which is stable like the state B.

The stationary states in this system can be summarized by a trajectory. In Fig. 5.12b, the function of the substrate concentration (c_S) versus the concentration

of the substance A (c_A) is shown. First, the three stationary values for c_A'' shown in Fig. 5.12a are plotted. A change in c_A causes a parallel shift of the curve J_A in Fig. 5.12a (for example, the dotted lines). The three stationary values occur only within a relatively narrow range of c_A (i.e. $J_A = c_A k_A$). If all the stationary states of c_s for different values of c_A are plotted in the coordinates of Fig. 5.12b, then an S-shaped trajectory is formed. This corresponds to a section through the behavioral surface of the cusp catastrophe, which is shown in Fig. 5.6. This particular example discussed here clearly shows that the points on the shaded area of the surface, corresponding to the point C in Fig. 5.12, are unstable states.

For the curve Fig. 5.12b, the flux J_{SZ} is taken to be constant. Any variation in the rate at which the substrate (S) is used, i.e. a modification of the coefficient k_{SZ} , will change the slope of the straight lines representing the function J_A in Fig. 5.12a (according to Eq. 5.26). It can easily be seen that, if the slope of this curve exceeds a critical value, a situation will develop where, independent of c_A , there will only be a single possible intercept. When compared with the functional surface in Fig. 5.6, this situation corresponds to a section through the bifurcation point P. In this instance, k_{SZ} (or J_{SZ}) would have to be plotted on the y-axis of the coordinate.

The analyses of metabolic networks of course lead to systems of many differential equations. The integrations of them are not only complicated, because most of them are nonlinear, but they are furthermore quite often ambiguous. Many possibly occurring stationary states are probably metastable. By small alterations of the parameters of the system, or by changing the initial conditions, the characters of the stationary states may modify in a non-easily predictable way. Even chaotic situations are possible. These properties not only make calculations difficult but they also sometimes lead to a higher degree of system destabilization.

In Sect. 3.1.4, we have already mentioned the time hierarchy of biological systems. In fact, the rate constants of biologically important reactions are distributed over a broad region of magnitudes. In the context of thermodynamic properties of systems, we have already discussed that, as a consequence in a steady state ($\sigma > 0$) of the whole system, fast reacting equilibria ($\sigma = 0$) of subsystems can be established. Considering this time hierarchy from the standpoint of systems analysis, the following conclusion about the behavior of systems can be drawn from mathematical considerations.

Formally, in a system of many differential equations, simplifications occur because differential quotients of fast processes may approach zero. This is the expression of the above-mentioned case where, in a steady state, as a non-equilibrium stationary state, some parts of the system are in equilibrium. Other processes, being very slow, and taking a shorter time scale, can be considered as non-existing, i.e. the corresponding parameters can be taken as constants. These kinds of simplifications not only help in mathematical modeling they automatically reduce the number of stationary states and stabilize the system itself. It seems that the established time hierarchy of biological reactions is the result of biological optimization in the course of evolution.

Further Reading

Heinrich and Schuster 1996; Klipp et al. 2009; Koch et al. 2011; Palmer 1995.

5.2.2 Pharmacokinetic Models

Optimization of modern therapy requires detailed knowledge of the kinetics of drug distribution in the body, as well as the time course of the induced physiological reactions. The first question concerns the *pharmacokinetics*, the second is the subject of *pharmacodynamics*. The latter includes sets of biochemical and physiological reactions, and concerns the problems of the previous sections in this book.

Pharmacokinetics can be considered as a particular kind of compartment analysis. In this case, the compartments, according to the definition given in Sect. 5.1.1, are represented by various organs and tissues of the body. A pharmacokinetic analysis answers the questions: how much, how frequently, in which way, and how long a definite drug must be administered to guarantee an optimal therapeutic efficiency, i.e. the optimal concentration to have the required effect.

The simplest kind of model calculation concerns the intravenously administered drug. In this case, the *bioavailability* is considered as 100%, and the drug is rapidly and homogeneously distributed throughout the whole blood volume. From here, it will be more or less quickly absorbed by other tissues and organs, metabolized, and finally excreted. As the first approximation, this process can be considered as an outflow from the blood volume as a one-compartment system. This can be described by the basic kinetic flux equation (Eq. 5.1, Sect. 5.1.1) and will follow the time function corresponding to Eq. 5.5.

Applied to the conditions of pharmacokinetics, this is:

$$\frac{dm}{dt} = -Cl\,c \quad (5.29)$$

where c is the concentration of the drug in the blood, which equals m/V , i.e. the total amount of the drug administered (m) per volume (V). k is the rate constant, which is displaced by the *clearance constant* Cl as applied by the volume. This *clearance* is defined as the volume of blood that is totally cleared of its content of the drug per unit time. It has the unit m^3s^{-1} , or more commonly: 1/h.

The integration of Eq. 5.29 leads to:

$$c = c_0 e^{-kt} = \frac{m_0}{V} e^{-\frac{Cl}{V}t} \quad (5.30)$$

where m_0 is the total amount of drug applied at time t_0 . As explained in Sect. 5.1.1 (Eq. 5.7), a half-life time of the drug can be obtained by:

$$t_{1/2} = \frac{V \ln 2}{Cl} \quad (5.31)$$

It should be noted that the clearance constant (Cl) is defined as being related to the volume of blood, whereas the volume (V) in this compartment analysis means the total available volume for the distribution of the particular drug. These, in some cases, may differ from each other.

For a particular pharmacokinetic analysis, the distribution of the drug in different parts of the body is of interest. For this, the linear approaches of the compartment analysis have been proved to be useful.

Figure 5.13 demonstrates an example of such a system, which can be regarded as a basic pharmacokinetic model. Schematically, it indicates the case of a single oral dose of a drug which is then taken from the stomach into the blood stream. From here, it is distributed into the tissues of various organs, and at the same time excreted by the kidney.

In this way, the drug distribution can be calculated for various kinds of administrations, considering differences in the time constants of resorption, excretion, and destruction by metabolism. Computer simulations allow one to consider stationary concentrations in particular tissues after repeated administrations of a drug or following a permanent infusion.

Usually, pharmacokinetic studies are based on linear models. In some instances, a concept known as *capacity-limited elimination* is used. This is the excretion of a drug using saturation kinetics. For the direct binding of the drug to a receptor, non-linear approaches should be used, similar to the Michaelis–Menten equation (Eq. 5.17).

Further Reading

Gabrielsson and Weiner (2006), Rowland and Tozer (2011).

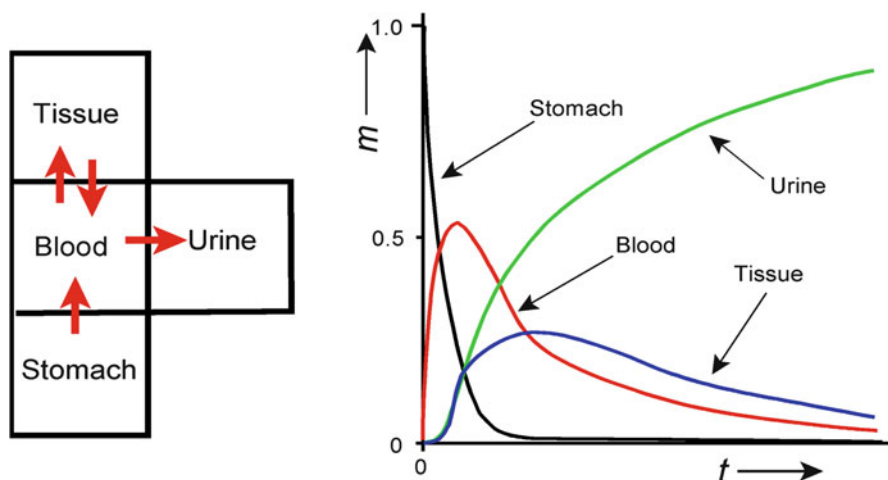


Fig. 5.13 Time course of drug distribution in a basic pharmacokinetic model: At $t = 0$, the drug is taken *per os*, subsequently transferred by the blood stream into the tissue and finally excreted. (The amount of the drug (m) is used as relative unit.) (Data from Knorre 1981)

5.2.3 Models of Propagation and Ecological Interactions

Ecological systems and their interactions are in fact extremely complex. Besides the *abiotic* conditions of the environment, which are mostly time dependent, a large number of *biotic* influences occur, like competition, predation, food webs, migrations, etc. The difficulties of the mathematical treatment of biological systems become evident in ecological models in an extreme way.

Actually, two kinds of models are used to handle this problem. A first approach uses *educational models*, which limit the investigation to one or two essential processes, separated from the complexity of the whole problem. These models are usually restricted to some simple assumptions, but in their favor, they lead to analytically tractable equations. In this way, even reflecting just more or less idealized conditions, this approach is suitable to gain insights into some general properties of the process, and furthermore include the possibility to extend the calculations to more complex models in a next step.

The second kind of approach is to use *practical models* which are based on a larger number of more or less realistic data and assumptions. These models essentially include large numbers of variables and parameters, which to some extent must be evaluated by rough estimations. The corresponding sets of equations are much too complicated for an analytical treatment. Results can be obtained just by numerical simulations. With increasing complexity, these models, however, become more and more ambiguous, and therefore more doubts are indicated whether they in fact describe realistic situations.

Anyway, these models can be formulated mathematically either by deterministic or by stochastic approaches. The *deterministic models* of growth and interactions in biological populations use phenomenological variables, and are finally based on the same kind of mathematical constructions as demonstrated in the previous sections for calculations of chemical reactions or transport processes. The change of the number (n) of individuals, which is described by the differential quotient dn/dt , can be considered, for example, in first approximations as proportional to their number (kn). These kinds of approaches are possible in considering large populations and averaging the whole process. In contrast to this, the *stochastic approaches* start with the consideration of the probability of the event of the birth of a new subject or of the death of an existing one. In this way, the probability $P(n, t)$ can be evaluated, indicating that the number of the individuals in the population increase or decrease.

In fact, there exist some differences in the results of these two approaches, at least in some particular situations. Let us consider firstly a simple deterministic example. Let the time constant (k) of the population be composed of a coefficient of propagation (k_p), and another for the mortality rate (k_m):

$$k = k_p - k_m \quad (k_p, k_m > 0) \quad (5.32)$$

In this way, the proliferation of a population can be described by the following differential equation:

$$\frac{dn}{dt} = k n \quad (5.33)$$

This expression states that both the rate of proliferation as well as the rate of mortality are proportional to the actual number of individuals in the population. Considering that at $t = 0$ the initial number of individuals is n_0 , and $k_p > k_m$, i.e.: $k > 0$, then the integration of Eq. 5.33 leads to the following equation:

$$n = n_0 e^{kt} \quad (5.34)$$

This function shows an exponential growth up to infinity. It was discovered as early as 1798 by the English economist Thomas Robert Malthus. The time course of this function is depicted in Fig. 5.14 (blue line).

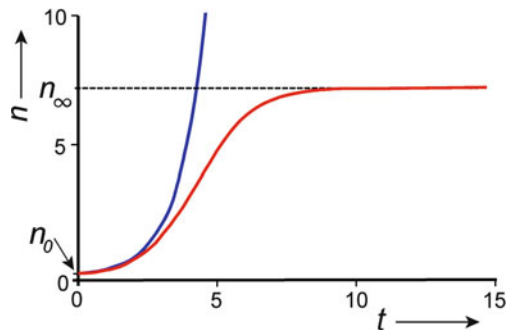
In this case, where k_m dominates, i.e. if the birth rate is lower than the rate of mortality, it becomes $k < 0$, and the number of individuals declines exponentially. From the mathematical point of view, the variable (n) in this way can become extremely small, but never zero. In other words, this approach excludes the realistic possibility of a full extinction of the population. In contrast, in the probabilistic models, this situation is included.

Another difference in these two approaches concerns the real-time dependence of biological propagation. In contrast to the deterministic models which assume a continuous process of birth and death, the stochastic approaches may better consider discrete intervals of time between these processes and even a completely non-overlapping of the generations, as happens, for example, in some insect populations.

In fact, educational models usually consider deterministic models, which we will discuss here, in order to demonstrate some general principles of population growth.

The original assumption of Eq. 5.33 supposes that the external milieu is constant and not affected by the organisms themselves. As a result, an infinitely growing Malthus function occurred (Eq. 5.34). This, however, reflects an idealized condition. Only when the number of subjects in the reservoir, i.e. cells in a suspension, is

Fig. 5.14 The function $n(t)$ according to the Verhulst–Pearl equation (Eq. 5.37) (red line), and the Malthus equation (Eq. 5.34) (blue line). Using arbitrary values in the ordinate and abscissa, the parameters are: $n_0 = 0.1$; $k_1 = 1$; $k_2 = 0.14$. In the case of Malthus curve, $k = k_1$



very small, and the time of consideration is sufficiently short, can this condition apply. In real situations, this in fact marks the initial exponential period of growth.

In most cases, a mutual influence of the individuals on one another is to be expected, as, for example, competition for food. This circumstance can be taken into account by modifying Eq. 5.32. It can be better postulated that k_p and k_m are not constant parameters but themselves functions of the population density. So, for example, one can assume that the rate constant of proliferation (k_p) will decrease and the rate constant of mortality (k_m) will increase, when n increases, because of shortage of food. The degree of physiological variability of these parameters, however, will be limited for genetic reasons. The simplest approach is the following linear equation:

$$k = k_1 - k_2 n \quad \text{for: } k_1, k_2 \geq 0 \quad (5.35)$$

This takes into account an optimal propagation constant, k_1 , determined genetically, in combination with a further parameter, k_2 , leading to a diminution of k_1 with an increase of the population density (n). If $n \rightarrow 0$ then $k_1 \Rightarrow k$. Introducing Eq. 5.26 into Eq. 5.33, one obtains:

$$\frac{dn}{dt} = (k_1 - k_2 n)n = k_1 n - k_2 n^2 \quad (5.36)$$

This approach therefore leads to a first order nonlinear differential equation. The integration of this equation for the same initial conditions as above ($t = 0$; $n = n_0$) gives an equation which in mathematics is known as the *logistic function*. In population kinetics, it is called the *Verhulst–Pearl law*:

$$n = \frac{n_0 k_1 e^{k_1 t}}{k_1 + k_2 n_0 (e^{k_1 t} - 1)} \quad (5.37)$$

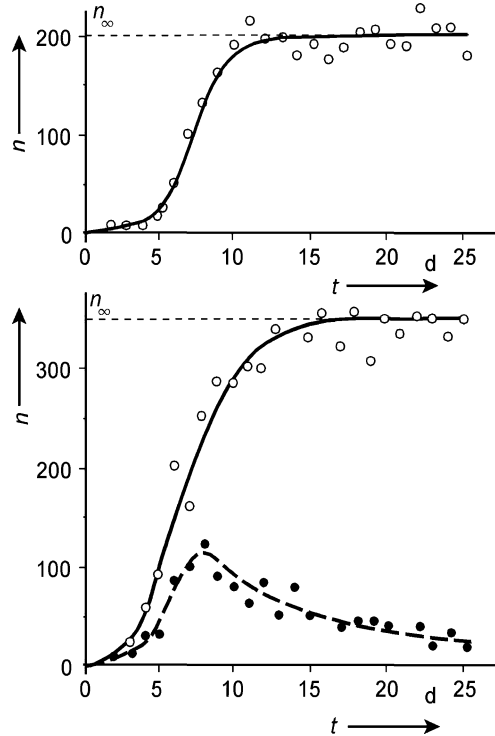
In contrast to Eq. 5.34, in this function, n does not approach infinity at $t = \infty$ but rather tends towards a saturation value (n_∞). This saturation value can easily be determined from Eq. 5.37 by considering that, with increasing t , the exponential terms increase up to a value where neither the subtraction of 1 nor the addition of k_1 become important. This leads to:

$$\lim_{t \rightarrow \infty} n = n_\infty = \frac{k_1}{k_2} \quad (5.38)$$

Furthermore, it is easy to show that, for small values of t , the Eq. 5.37 takes the form of Eq. 5.34. Both functions, for some arbitrarily chosen parameters, are illustrated in Fig. 5.14. An example of experimentally obtained data is shown in Fig. 5.15.

If the system contains individuals of several different species, then the inter-relationship between them must be taken into consideration. This can involve competition for the food supply (Fig. 5.15, below) or a predator–prey relationship

Fig. 5.15 The time course of the growth of a pure population of *Paramecium caudatum* (full line) alone (above), and mixed together with a population of *Paramecium aurelia* (dashed line) (below). The ordinates indicate the number of individuals (n) per 0.5 ml medium. The abscissa gives the time in days (d) (Data from Gause 1935)



(Fig. 5.16). Systems like this had already been studied in the 1930s by Vito Volterra and Alfred Lotka.

The following set of equations describes the simplest case of a predator–prey relation between two species:

$$\left. \begin{aligned} \frac{dn_P}{dt} &= (k_{pP} - k_{mP}n_H)n_P \\ \frac{dn_H}{dt} &= (k_{pH}n_P - k_{mH})n_H \end{aligned} \right\} \quad (5.39)$$

The subscripts p and m have the same meaning as in Eq. 5.32, and the index P in these equations refers to “prey,” whereas H means “hunter.” For the prey organisms, the mortality ($k_{mP}n_H$) will increase as the number of predators, i.e. hunters (n_H) increases. The greater the number of predators, the shorter will be the life expectancy of the prey organisms. Otherwise, an increasing number of prey organisms means better living conditions for the predators. For this reason, the propagation coefficient of the predators (k_{pH}) is multiplied by the amount of prey organisms (n_P) in the second equation.

This set of Eq. 5.39 represents an extremely simplified situation. The interaction of individuals of the same species, for example, which was already reflected in

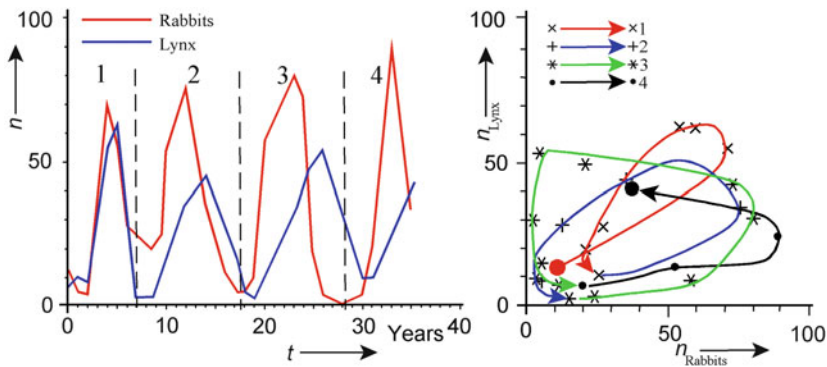


Fig. 5.16 A particular example of population kinetics: the number of pelts obtained from hunted lynx and snow rabbits in Canada after 1900 (numbers in thousands). On the *left*: time course; on the *right*: the same numbers in a parameter plot. The population waves 1. . . 4 of the left part of this figure appear as Volterra cycles in the graph on the right, starting at the *red point* and ending at the *back point* (Data from Haken 2010)

Eq. 5.36, has been ignored here. This example, however, demonstrates that even the simplest ecological models already lead to complicated systems of nonlinear equations. The solution of such systems of equations frequently leads to functions indicating oscillations. This corresponds to experiences where ecological systems oscillate even without additional input of external signals, such as daily or annual oscillations of light and temperature.

These oscillations are called *Volterra cycles*. They can either be plotted versus time, or represented as trajectories. In the case of Eq. 5.39, for example, the functions $n_P(t)$ and $n_H(t)$ can be plotted separately or combined as the trajectory $n_P(n_H)$. Again, a system trajectory would appear like those shown in Figs. 5.4, 5.5, 5.6, and 5.12b.

Figure 5.16 demonstrates a concrete example of such oscillations. It uses data from the hunting lists of the Hudson Bay Company in Canada which were kept from 1845 to 1935. Supposing the amounts of hunted animals were proportional to the number of those really existing, one can roughly use this as a representation of the ecological situation. In Fig. 5.16, the data from 1900 to 1935 were used, indicating the time course with typical oscillations (left) and the same data in a parameter plot showing Volterra cycles (right).

In a similar way to the predator–prey system, a kind of symbiotic inter-relationship of two species can be formulated:

$$\left. \begin{aligned} \frac{dn_A}{dt} &= (k'_{pA} + k''_{pA} n_B - k_{mA}) n_A \\ \frac{dn_B}{dt} &= (k'_{pB} + k''_{pB} n_A - k_{mB}) n_B \end{aligned} \right\} \quad (5.40)$$

This set of equations takes into account the fact that when $k'_{pA} > k'_{mA}$, i.e. $k'_{pB} > k'_{mB}$, then one species of this symbiotic community can survive without the other. The symbiotic cooperation is expressed by the product terms $k''_{pA} n_B$ and $k''_{pB} n_A$.

Mathematical approaches of this kind form the basis for modeling of ecological systems. However, other factors must also be considered such as a large number of trophic levels and complex inter-relationships, time delays caused by processes of developments, annual rhythms, and also ethological and sociological structures.

Further Reading

Ginzburg and Colyvan 2004; Jeffries 1989; Okubo and Levin 2010; Renshaw 1990; Turchin 2003.

5.2.4 Models of Growth and Differentiation

The morphology of an organism, its geometric structure and the corresponding morphogenetic mechanisms, have provided time and again subjects for modeling. In 1917, in his famous book "On Growth and Form," D'Arcy Thompson first formulated the mathematical ideas of morphometry. His theory states that different types of body shape of an animal, or parts of it, if projected onto a system of coordinates, can be systematically transformed into another by means of the mathematically defined distortion, i.e. by a sort of coordinate transformation. D'Arcy Thompson demonstrated this method for the shapes of various species of fishes as well as forms of crab carapaces, mammalian skulls, and even leaves of plants and many other structures. In this way, for example, phylogenetic changes can be expressed mathematically. Such approaches belong more to the realms of biometry. They do not include the mechanisms of differentiation and shape formation. The same applies to *allometric models* of growth which are discussed in Sect. 3.8.

A striking phenomenon of living subjects is their ability of self-organization. What mechanisms finally transform the genetic information of an organism into its final structure? In which way a more or less homogeneous egg, or even a single embryonic stem cell, may produce a well-organized animal? What processes control the specific pattern formation in plants and animals?

Experiments for the purpose of understanding processes of regeneration and development range back to the eighteenth century. The German biologists Hans Spemann and Hilde Mangold were crowned by the Nobel prize in 1935 for their experiments on eggs and embryos of amphibians. They supported the idea of an organizing center, later named *Spemann organizer*, which influences the differentiation of cells around a concentration gradient of a certain activator substance.

In the following, a large number of genetic and epigenetic mechanisms were found, controlling growth and differentiation in various plants and animals by systems of activator- and inhibitor molecules. These substances control the

morphogenetic movements of cells, as well as the activation of the genes that are responsible for differentiation. In this context, reference must be made to the discussions on the possible role of electric fields in embryogenesis, leading to electro-diffusion of growth factors (Sect. 3.5.2).

This progress in molecular biology meets the advantage of the theory of nonlinear processes, and of course was strongly supported by the development of powerful computer techniques. As already explained in Sect. 3.1.4, dissipative structures can be established spontaneously in nonlinear systems as a result of superimposition of two or more processes. The Volterra cycles of population dynamics, as described in Sect. 5.23 (Fig. 5.16), like many other oscillations in biological systems can be considered as dissipative structures in time. Evolutionary patterns, on the other hand, are examples of dissipative structures in a spatial “frozen” state. In this way, they are the consistent result of time-dependent concentration gradients.

A first mathematical formulation of this idea was published by Alan Turing in 1952. He demonstrated that spontaneous pattern formation is possible as the result of interaction of two components with different diffusion rates. This is the origin of the *reaction–diffusion model of morphogenesis*, a mathematical formulation of the physiological idea of Spemanns organizer system.

The basic equation for these calculations is the second Fick’s law (Eq. 3.32, Sect. 3.3.1) which explains the time-dependent formation of a concentration gradient of a substance as the result of diffusion, starting at the point of its production or activation. In the case where the molecules were supported permanently from a source into the target, this system does not have steady states, but just temporarily stable gradients can be formed. To attain a concentration gradient, as a time-independent steady state, the diffusion term must be completed by a certain reaction of molecule depletion:

$$\frac{dc}{dt} = D \frac{\partial^2 c}{\partial x^2} - kc \quad (5.41)$$

This is the simplest approach, because it represents just a one-dimensional system, and furthermore, a linear kinetics of activator depletion where the parameter k represents the time constant of this reaction. In fact, the particular kind of depletion reaction is crucial for the formulation of a realistic model.

Furthermore, a reaction term must be introduced, characterizing the source of the activator. In many models, an autocatalytic kind of reaction is supposed. This, in the simplest way, is a kind of bimolecular reaction which is nonlinear, i.e. determined by the square of concentration. An important aspect was the introduction of an antagonist to the activator, a kind of inhibitor, the formation of which is actually initiated by the activator itself. Finally, it was established that pattern formation is possible only if a locally restricted self-enhancing reaction is coupled with an antagonistic reaction that acts on a longer range.

In the simplest way, two substances, an activator (A) and an inhibitor (I), are considered, which are formed at particular sites and then distributed by diffusion. In this way, complex concentration patterns of both substances are set up, generating a

particular space–time structure. The growth rate, or the mode of differentiation, finally depends on the concentration ratio of these two factors.

Consider again just a simple unidimensional case where the concentration of an activator (c_A) and an inhibitor (c_I) are distributed only in the x -direction. For this case, the following set of equations can be formulated:

$$\left. \begin{aligned} \frac{\partial c_A}{\partial t} &= k_A \frac{c_A^2}{c_I} - k'_A c_A + D_A \frac{\partial^2 c_A}{\partial x^2} \\ \frac{\partial c_I}{\partial t} &= k_A c_A^2 - k'_I c_I + D_I \frac{\partial^2 c_I}{\partial x^2} \end{aligned} \right\} \quad (5.42)$$

This is a system of partial, nonlinear differential equations: *partial*, because the variables c_A and c_I are differentiated both with respect to time and position, and *nonlinear*, because the square of the concentrations appears as the result of an autocatalytic process.

In particular, the following assumptions are made in this model. The rate of formation of the activating factor A is determined by a process that is autocatalytically promoted by c_A , and furthermore suppressed by the concentration of the inhibitor (c_I). This corresponds to the term: $(k_A \frac{c_A^2}{c_I})$. This quadratic term of activator formation overcomes the decay of the activator which is expressed by the simple linear term $(-k'_A c_A)$. The diffusion along the x -axis $(D_A \frac{\partial^2 c_A}{\partial x^2})$ again follows the second Fick's law. The rate of inhibitor production also depends on the square of the activator concentration. It differs from that of the activator (A) only by the circumstance that the autocatalytic activation $(k'_A c_A)$ does not include an additional inhibitor term.

The behavior of such a system depends to a great extent on the choice of the constants. The relation of the diffusion coefficients D_A and D_I is of particular importance. If the diffusion constant for the inhibitor (D_I) is much larger than that for the activator (D_A), then the distribution of the factor I can be regarded as being homogeneous.

This figure shows an extremely simplified case. However, it illustrates the formation of a more or less periodic structure in a one-dimensional system. The question arises: Why there is no self-inhibition of the heterocysts by their own inhibitor production? The answer is the c_A^2 -terms in Eq. 5.42. This nonlinear self-enhancement, together with the linear process of lateral inhibition, makes a uniform distribution unstable and eventually leads to the periodic structure.

The simplified one-dimensional approach is applied to a particular case: the cell differentiation in the blue-green algae *Anabaena*. Under nitrogen-limiting conditions, their filaments at some intervals differentiate vegetative cells into larger non-dividing heterocysts, which are specialized for nitrogen fixation. The distance between two heterocysts in the filament usually is larger than 12–14 cells.

To understand this reaction, it is postulated that the activator of heterocyte-differentiation, which has actually been identified as the transcription factor HetR, which activates itself in an autocatalytic reaction like Eq. 5.42. In the same way, however, it activates the formation of a small peptide PatS as an inhibitor. In contrast to the inhibitor, which diffuses along the filament, the activator does not

leave the cell ($D_A = 0$). The result of this proposal is illustrated schematically in Fig. 5.17.

Another model shows an approach to explain the formation of filamentous branched structures in a two-dimensional system with a similar activator–inhibitor mechanism (Fig. 5.18). In this case, it is assumed that at first the plane contains a homogeneous distribution of a certain trophic factor which allows cell differentiation. If this occurs at a particular point, a trace starts to develop behind a moving local signal. Furthermore, the differentiated cells lower the concentration of the trophic substrate in their vicinity. Only in some cases, where this reaction is not strong enough, do branches of the filament appear.

Figures 5.17 and 5.18 show two simple kinetic models of differentiation processes, just to demonstrate the basic ideas of this reaction–diffusion model. In addition to the initial idea of Turing (1952), an autocatalytic activator–inhibitor mechanism must be included, to attain these results. The molecular basis of such systems, meanwhile, has been found experimentally in many concrete systems. In this way, this model, starting with more or less hypothetical assumptions, had been verified and concretized in many cases.

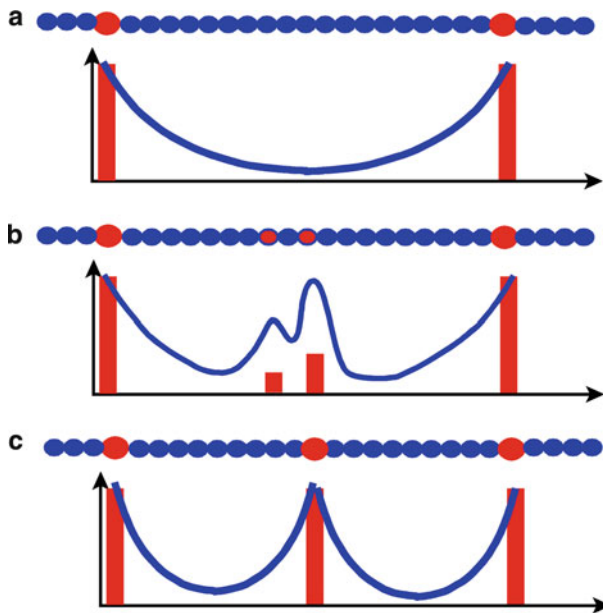


Fig. 5.17 A scheme of heterocyst formation (red spheres) in a filament of normal vegetative cells (blue spheres) of *Anabaena* as an interplay of a non-diffundable activator (red bars) produced in individual cells, and an inhibitor (blue lines), which itself is produced catalytically by the activator (corresponding to an equation similar to Eq. 5.42). If the distance between two inhibitor producing heterocysts is large enough (a), new heterocysts began to differentiate. They are in competition among each other, because of their own inhibitor production (b). Finally, only one of them becomes effective (c) (Modified after Meinhardt 2008)

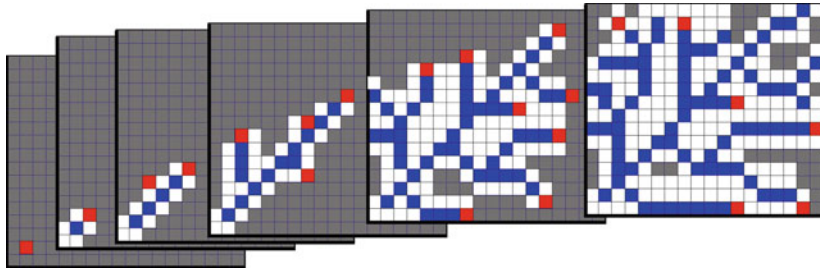


Fig. 5.18 Formation of a branched structure behind moving signals. A homogeneously distributed guiding factor (*gray squares*) stimulates a spontaneous cell differentiation. Starting with a single differentiated cell (*red square, left plane*) an elongation occurs (*blue squares*), whereas the differentiated cells remove the trophic substrate in their neighborhood (*white squares*). In some cases, where the lateral inhibition temporary is not strong enough, branches can be formed (After Meinhardt 2008, redrawn)

The one-dimensional approach of Fig. 5.17 in this way was extended to the two-dimensional system in Fig. 5.18. Finally, it is possible to formulate even three-dimensional systems of differentiation. In many cases, however, the three-dimensional orientation is finally based on two-dimensional events, because it takes place in two-dimensional cell sheets. In organs such as three-dimensional structures, firstly a two-dimensional pattern is generated which subsequently folds or form tubes or other epithelial structures.

Starting from more or less simple models, like tentacle formation of the freshwater polyp hydra, or the regeneration of the flatworm planaria, which are supported by a sufficient experimental basis, similar approaches are meanwhile used to explain differentiation processes on higher organized animals. So, for example, the embryonal orientation of the main body axis and the patterning of limbs in vertebrates, the formation of insect legs, but also many examples of pattern formation in reptiles and mammals, have been investigated. Some of these models are purely hypothetical, while others are based on molecular biological data.

Further Reading

Maini and Othmer 2000; Meinhardt 2008, 2009; Murray 2003; Wartlick et al. 2009.

5.2.5 Models of Origin and Evolution of Life

The interest in the question, how life could have originated and developed, is not new. It is stimulated to a great extent by progress in our knowledge of the early history of the earth, by results in molecular biology, and last but not least by the advances in systems theory in general. In the 1950s, Stanley L. Miller showed experimentally that electric discharges applied to a mixture of CH_4 , NH_3 , H_2 , and water vapor could generate a number of amino acids. These experiments confirmed

earlier speculations but at the same time they provoked a number of new questions: What concrete physicochemical conditions had led to the origin of the first biologically significant molecules? Were these polypeptides first, or instead the nucleic acids? How could the statistically completely improbable amino-acid sequence of functioning proteins occur? What is the origin of the genetic code? The answers to these questions are expected from geophysics, from physical chemistry, and from various theoretical approaches.

The chemical evolution could have happened only after the temperature on the earth surface allowed the existence of liquid water, which was considered to have occurred only about 3.8 billion years ago. On the other hand, the first fossil cells have been assigned an age of 3.5 billion years. The period of chemical evolution, therefore, probably took less than 300 million years. Considering the above-mentioned hen-and-egg problem, it is assumed that probably the polypeptides developed before the nucleic acids. This is based on some arguments: A chemical evolution is possible only with the help of some enzyme-controlled mechanisms, which can be realized even by particular polypeptides. On the other hand, the phosphates, which are essential for the formation of RNA, would have been precipitated by heavier metal ions in the “primordial soup,” and therefore not have been easily available. Furthermore, these polypeptides in contrast to nucleic acids indicate a larger resistance to the extremely high UV-irradiation at that time.

The formation of polypeptides seems to be catalyzed by surface reactions of minerals like kaolinite, montmorillonite, hectorite, or other smectites. This was probably promoted by local heating—drying—wetting cycles in this period. In this context, the hypothesis of the *salt-induced peptide formation (SIPF) reaction* should be mentioned. It considers that, above a concentration of 3 M, the NaCl can be considered as a condensation reagent, helping amino acids to form oligomers through peptide linkages.

So far, the physicochemical background of recent speculations on the origin of the first bio-organic molecules, concerning the first two questions which are mentioned above, have been considered. The next questions are the subject of various theories of probability, as well as information, game, and network theory. This mainly concerns the problem of the generation of a particular amino-acid sequence transforming a polypeptide with stochastic composition into a functional protein.

Consider a particular enzymatic biopolymer, a rather small one with an information content of only 10^3 bits. As a product of stochastic combination, it would result, on average, once in the course of $2^{1,000} \approx 10^{300}$ chances. This means that, out of 10^{300} resulting molecules, only one could be considered as being able to realize this particular function. To illustrate this abstract situation in a more realistic picture, one should realize that the whole hydrosphere of the earth contains only about 10^{32} molecules. Even considering the whole time of 10^{17} s of the existence of the earth and an extremely fast recombination reaction of 10^{-9} s, the probability of stochastic origin of this enzyme would be minute (see also Sects. 2.1.2 and 2.1.3). Albert Einstein once compared this probability with the chance of a complete dictionary resulting from an explosion in a printing office!

A solution of this problem offers the *quasispecies theory*, proposed by Manfred Eigen (1971). This suggests that, already at the beginning of the chemical evolution, the principles of competition are responsible for the development of particular chains of reactions and patterns of polymerizations. The formation of a functional amino-acid sequence of proteins is therefore not a product of chance but of selection, as in the case of further evolution, according to the principles of Charles Darwin.

These effects of *prebiotic molecular selection* can be considered as processes of catalysis and auto-catalysis. Auto-catalysis can be regarded in this context as a forerunner of self-reproduction.

The number of monomers in a given region is limited, and this will lead to a form of competition between the different kinds of polymerization reactions. However, such a system would rapidly reach an equilibrium state if it were not exposed to a continuous flow of energy and entropy. In fact, it will be disturbed permanently by geophysical effects of heating, cooling, drying, etc. From the thermodynamic point of view, the resulting combinations can be considered as a kind of dissipative structures (see Sect. 3.1.4).

Under these conditions, a selection of polymers with high catalytic activity will take place. A network of catalytic relationships between these individual molecules is built up (Fig. 5.19). If a closed chain occurs in such a network, this will form an auto-catalytic cycle. This is a structure with a good chance for survival. Such cycles sustain themselves, and in doing so, they can modify their own structure. Side chains, which are formed through the poly-catalytic activity of several molecules, are removed and distensible intermediate linkages are eliminated. In this way, the cycle becomes smaller. At the same time, it can be shown that

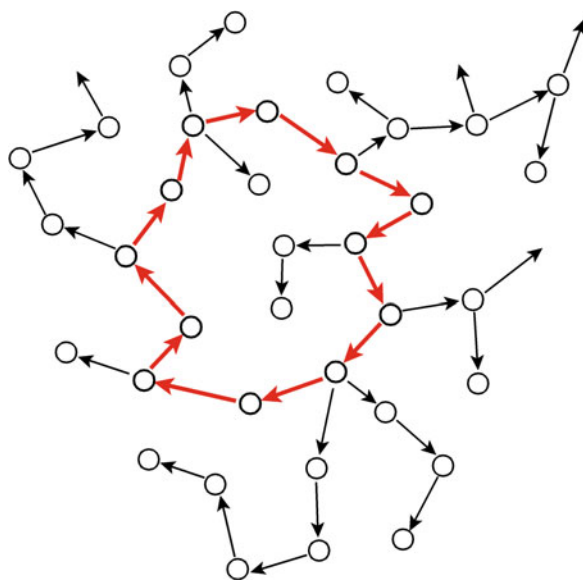


Fig. 5.19 The network of catalytic correlations in a mixture of polymers. The red arrows indicate an auto-catalytic hypercycle

heterogeneous cycles, made up of alternating proteins and nucleic acids, possess a selection advantage. Eigen called such structures *hypercycles*.

This concept of evolution, therefore, indicates a dialectic unity of stochastic and deterministic processes. The stochastic component is derived from the stability of bonds, and from the stochastics of molecular contacts in the course of the reaction. At higher levels of development, this stochastic element is realized by the mutation rate or the dependability of self-reproduction.

This theory is backed up by mathematical modeling and computer simulations. It is possible to formulate equations for systems of chemical reactions as well as for processes of reproduction and selection at the highest levels of evolution. With regard to the equations that were given in the previous sections, the following simplified approach can be demonstrated:

$$\frac{dn_i}{dt} = (k_{pi}q_i - k_{mi} - k'_i)n_i + \sum_{j \neq i} k_{m,ji}n_j \quad (5.43)$$

This is the equation for the time course of the number of individuals (n_i) of a species i , which embodies the following postulates: The individuals propagate at a rate that bears a linear relationship to the existing number of individuals and is determined by a time factor k_{pi} . This coefficient is, however, reduced by the factor q_i being a little less than one. In this way, the fact is taken into account that, by mutations, the identical reproduction of the individuals is disturbed. The coefficient governing the mortality (k_{mi}) is the same as in Eq. 5.32. In addition, another coefficient, k'_i , is introduced, expressing the possibility that some individuals will leave the area, i.e. will diffuse away. The summing term at the end of the equation reflects the possibility of the emergence of an individual of the species i from another of the species j by mutation. In contrast to the ecological equations in Sect. 5.2.3, n_i is thus capable of increasing even under the initial conditions, when $t = 0$ and $n_i = 0$. This sum term is a prerequisite for the neogenesis of a species.

This is, of course, a very primitive approach. It presupposes that propagation is based on linear relationships, an assumption having limited validity as has already been discussed in Sect. 5.2.3. The relationships between species have also not been considered. Nevertheless, evolution because of competition between species can be demonstrated on the basis of this approach if the following mass balance is used:

$$\sum_{i=1}^m n_i = \text{const}, \quad \text{therefore: } \sum_{i=1}^m \frac{dn_i}{dt} = 0. \quad (5.44)$$

The advantage for selection of a species will then result from the relationships of the chosen system constants, one to another. With more complex evolutionary systems, weighting functions will have to be introduced but these must necessarily be of a speculative nature.

Besides the Eigen model, J. F. Crow and M. Kimura in 1970 introduced another quasispecies model of evolution. In contrast to the Eigen model, where the

sequences are subjected to point mutations only during the process of replication (*connected mutation-selection*), in the Crow–Kimura model, mutation and selection are two independent processes (*parallel mutation-selection*). The relationships of these models indicate that the optimum mutation rate in the connected mutation–selection model is lower than that in the parallel mutation–selection model. The Crow–Kimura model therefore gives an adaptive advantage under changing environmental conditions.

A further problem of prebiotic evolution concerns the development of the genetic code which translates the four-letter language of the nucleic bases, A, T, G, and C, into the 20-letter language of the amino acids. In fact, it is highly redundant, with the consequence that all amino acids, except methionine and tryptophan, are encoded by multiple codons. It is suggested that, in a primitive living systems, amino acids and nucleotides coexist and bind non-specifically. This kind of random association of nucleotides nevertheless could catalyze the synthesis of random polypeptides, barely functioning as proto-proteins. The evolution from the *non-coding* to the *coding* state is considered as a result of the interplay of the three conflicting forces: the need for diverse amino acids, for error-tolerance, and for minimal cost of resources. Therefore:

$$(\text{fitness}) = (\text{diversity}) - (\text{error} - \text{load}) - (\text{cost})$$

These relationships can be evaluated by mathematical expressions. The result indicates a certain degree of smoothness. In fact, the view of some codes in bacteria, protozoa and mitochondria indicate that the standard genetic code is not strictly universal. This suggests that the code was open, at least for some time, to evolutionary change.

Interestingly, these sorts of calculations lead to the *map coloring problem*. The question is about the minimal number of colors which are required to color an arbitrary map on a surface such that no two bordering countries have the same color. The result is known as the *coloring*, or *chromatic*, *number*.

Further Reading

Ancliff and Park 2010; Dyson 1999; Eigen 1992; Obermayer and Frey 2010; Rode et al. 2007; Tlusty 2010.

5.2.6 Models of Neuronal Processes

The neuronal system, starting from the excitation mechanism of a single neuron up to reactions of neuronal networks in the brain, is extremely complex and multilayered. Correspondingly, the theoretical approaches attempting to explain and model these processes are quite heterogeneous. A number of newly introduced experimental techniques for *in vivo*, as well as *in vitro* investigations provide a mass of new information requiring to be functionally connected, and included in model

considerations. This concerns methods for simultaneous recording of multiple single units and optical recordings with voltage and ion-sensitive dyes, as well as large-scale measurements of brain structure and activity with positron emission tomography (PET), magnetoencephalogram (MEG), 2-deoxyglucose (2-DG), and magnetic resonance imaging (MRI). In parallel, the enormous development of computer techniques in the last decades has stimulated the efforts of theoretical modeling.

The development of neuronal and brain models is helpful for a number of questions. In general, it makes the brain system with many interacting components more accessible. Furthermore, it is able to simulate and predict some properties of the neuronal system, which leads to new experiments, and makes it possible to understand and concentrate experimentally obtained data. In some cases, experiments that are difficult or even impossible to perform in living tissue can be simulated by the use of a model. The major research objective of computational neuroscience, however, is to discover the algorithms used in the neuronal systems.

As the result, the number of neuronal models becomes nearly unmanageably large. In the second half of the 1980s, the term *computational neuroscience* was introduced, some years before the coining of the expression *systems biology*. To some extent, this term replaced the *biocybernetics* which in the 1960s had been used to summarize research on many neuronal models.

Nevertheless, the trend to display neuronal functions by simplified analog models is as old as neurobiology itself. Already in 1907, Louis Lapicque used an electronic trigger circuit to model this process. A capacitor is charged up to a certain threshold of voltage, leading to a reaction of instantaneous discharge. This model was completed in the 1930s by Archibald Hill.

After this, various models have been proposed using approaches of Boolean algebra. In the 1940s, McCulloch and Pitts investigated networks consisting of *formal neurons*. These are elements performing basic logical functions, resembling in some instances nerve cells with synaptic connections. In the late 1950s, John von Neumann introduced his theory of self-organization in cellular automata, suggesting the intimate relationship between computers and the brain. This basic model has been modified in subsequent years. So, for example, the refractory period of nerves was included, which means that, for some milliseconds after an input signal, the nerve is not able to accept a further one. These concepts of neuronal networks mostly support the development of new technical systems of artificial intelligence and intelligent computing.

The central problem of these models, however, consists in the requirement of their synchronous reactions. They require a synchronization of all impulses in the network, i.e. the introduction of a certain sequence of operations. Despite the fact that in some particular brain neurons a certain sequence of firing exists, for example in cortical cell columns, there is no evidence of an overall synchronization in the brain. In contrast, it is necessary to take into account the random times of arrival of the synaptic inputs.

In the 1960s, a number of approaches using these digital elements led to models resembling various kinds of behavior and sensory perception. This includes some

systems of feedback regulation as explained in Sect. 5.1.4. As already mentioned, these models at that time were subsumed under the term *biocybernetics*.

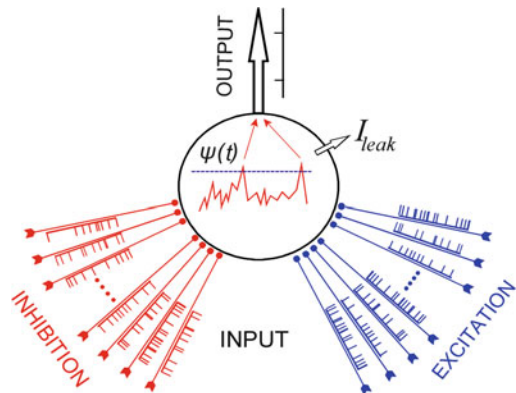
In general, neuronal and brain models according to their approaches can be classified as realistic and as simplified ones. *Realistic brain models* start with the entire process of nerve excitation, for example with the Hodgkin–Huxley model as explained in detail in Sect. 3.4.4. In this case, one tries to include as much of molecular and cellular details as are available. The extremely large number of parameters of these models makes it impossible for them to be analyzed analytically. They can be handled only using numerical simulations. Furthermore, the large number of parameters and their variability make the results of such computations sometimes rather ambiguous. The large amount of detail contained in these models can often obscure rather than illuminate the problem. One way of overcoming this dilemma is the more or less rigorous reduction of the Hodgkin–Huxley equations to some essential properties.

As a consequence, this finally leads to the *simplified brain models*. In exchange for analytical tractability, they do abstract from the complexity of individual neurons and the patterns of connectivity. These simplifying models, of course, are essential, but to some extent they are also dangerously seductive because they can become an end in themselves and lose touch with reality.

Some of these models consider *spiking* neurons, reacting simply as digital elements, in contrast to *firing-rates* models, where the code is based on the frequency of the firing rate. Such frequency coded signals in fact occur in most sense organs. In some cases, however, the reaction times are too short to allow long temporal analysis.

The *integrate-and-fire model* (Fig. 5.20) has become widely accepted as one of the canonical models for the study of neural systems. It can be considered as a combination of the earliest electronic trigger-circuit models with the approaches of Boolean algebra. This model considers the state of the neuron as a function of its membrane potential $[\psi(t)]$. This is controlled by stochastic excitatory or inhibitory

Fig. 5.20 Schematic representation of an integrate-and-fire model. The membrane potential $\psi(t)$ of the neuron is decreased by stochastic inhibitory (red) and increased by excitatory (blue) synaptic inputs. Furthermore, a leak current (I_{leak}) (small black arrow) permanently degrades the membrane potential with a time constant (τ_m). If the membrane potential arrives a critical value (dashed line), output spikes (black) are generated



synaptic inputs. These inputs are modeled either as injected currents (current synapses) or as a change in the membrane conductance (conductance synapses). Furthermore, due to ion leakage, the membrane potential permanently decays with a characteristic time constant (τ_m). When the membrane potential reaches a definite maximal threshold, the neuron generates an output spike.

With respect to the membrane capacity (C_m), this process can be characterized as follows:

$$C_m \frac{d\psi(t)}{dt} = I_{leak}(t) + I_{syn}(t) \quad (5.45)$$

whereas the current (I_{syn}) characterizes the input current from the synapses, and (I_{leak}) stands for the leak current which results from:

$$I_{leak} = -\frac{C_m}{\tau_m} [\psi(t) - \psi_o] \quad (5.46)$$

This includes the time constant (τ_m) which results from product of the capacity and resistance of the membrane as explained in Sect. 3.5.4 (Eq. 3.214). ψ_o indicates the resting potential of the neuron.

The crucial point of the integrate-and-fire model, and the source of many variations of it, is the formulation of the term (I_{syn}). This is an integral expression which includes the weighted amount of stochastic negative, or positive pulses, depending on the inhibitory or excitatory character of the synapses. In the case of *current synapses*, it is the linear sum of weighted current pulses of various strengths, durations, and polarity. In the case of *conductance synapses*, a nonlinear relationship occurs because the currents, produced by individual pulses, depend on the degree of the corresponding conductance variations, and additionally on the actual membrane potential. The arrival time of the synaptic input pulses is generally modeled as a time-homogeneous stochastic process based on Poisson distribution. Under certain assumptions, it is possible to calculate the output spiking rate of the neuron for this case.

To understand processes coded by the firing rate of pulses, the integrate-and-fire model was extended to inhomogeneous and especially to periodic synaptic input. This also includes the mechanism of stochastic resonance which is important to analyze periodic processes, superimposed by stochastic noise (see Sect. 3.1.5). This, for example, allowed the analyses of auditory neural processing, where neural responses show activity that is phase-locked to periodic acoustical signals.

The integrate-and-fire neuron model appears as a very useful tool to understand a number of neurological observations, how neural systems function, and how they process information. On the one hand, it is sufficiently simple to allow mathematical analysis, while on the other hand, it contains experimentally measurable parameters, such as the firing threshold of the neuron, its resting potential, functional spikes, the response of the membrane potential to stochastic synaptic inputs, etc. Of course, there are a number of problems associated with the simplifying

assumptions. Nevertheless, it represents a balance between conceptual simplicity and biological accuracy.

Beside these models, which at least contain basic neurobiological facts, there are some others which in fact only show some formally analog phenomena. A frequently used example is the *spin-glass model*. This considers the cooperative behavior of molecules in amorphous materials, theoretically treated by approaches of statistical thermodynamics.

Spin-glasses are viscous materials containing particles of different magnetic properties. These include ferromagnetic particles, orienting themselves by mutual influences parallel to each other, as well as anti-ferromagnetic particles, orienting anti-parallel (Fig. 5.21). The analogy between spin-glasses and neuronal networks is the direction of the particle and the state of the neuron as well as their mutual interactions. A parallel or anti-parallel orientation of a particle in the spin-glass would correspond to a “silent” or an “excited” cell in a neural network. The ferromagnetic, i.e. anti-ferromagnetic interaction of the particles, simulates the excitation or inhibitory interaction of the nerve cells.

In this respect, spin-glasses may show a quite complicated pattern of self organization. Of particular interest are the situations where two opposite kinds of interaction come together in one point. In this case, a particle will be influenced at the same time by two forces tending to orientate it in opposite ways. This disturbs the stability of the whole system. This situation is called a *frustrated* state. It can be considered as a sort of metastability and could reflect a kind of memory state of a neuronal network. It is a special kind of phase transition which cannot be described like a catastrophe as depicted in Fig. 5.6, but instead shows a more continuous kind of transition. These spin-glass models, therefore, resemble not only the problem of self-organization but also directly the function of neuronal networks.

The development of new analytical techniques, combined with the sharp increase in computational power, have led to formidable progress in neuroscience in recent decades. This, however, should not mask the fact that the brain is a multilayered functional system, and the path from the action potential of a single neuron up to complex neurological processes such as thinking or even consciousness cannot be modeled by a single mathematical approach. It is the merit of the

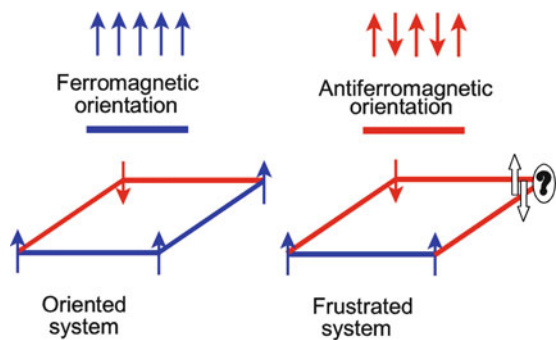


Fig. 5.21 Ferromagnetic and antiferromagnetic interactions in spin-glasses and the occurrence of frustrated systems

research of the last 50 years to realize that there is a principal difference between the data processing in brain versus that in the recent kinds of digital computers.

Further Reading

Banerjee and Chakrabarti 2008; Burkitt 2006; Cessac and Samuelides 2007; Faugeras et al. 2009; Hanrahan 2011; Prettejohn et al. 2011.

References

- Adair RK (2002) Vibrational resonances in biological systems at microwave frequencies. *Biophys J* 82:1147
- Ahlborn BK (2004) *Zoological physics*. Springer, Berlin
- Ainslie MA, McColm JG (1998) A simplified formula for viscous and chemical absorption in sea water. *J Acoust Soc Am* 103:1671
- Alberghina L, Westerhoff HV (eds) (2005) *Systems biology. Definitions and perspectives*. Springer, Berlin
- Alexander RMcN (2003) *Principles of animal locomotion*. Princeton University Press, Princeton
- Alon U (2006) *An introduction to systems biology: design principles of biological circuits*. Chapman and Hall/CRC, Boca Raton, Mathematical & computational biology
- Ancliff M, Park JM (2010) Optimal mutation rates in dynamic environments: the Eigen model. *Phys Rev E* 82:21904
- Andrä W, Nowak HE (2006) *Magnetism in medicine*. Wiley-VCM, Berlin
- Anishchenko V, Astakhov V et al (2006) *Dynamics of chaotic and stochastic systems*, 2nd edn. Springer, Berlin
- Au WWL (1994) *Sonar of dolphins*. Springer, New York
- Amman D (1986) *Ion-selective microelectrodes*. Springer, Berlin
- Arnason BT, Hart LA et al (2002) The properties of geophysical fields and their effects on elephants and other animals. *J Comp Psychol* 116:123
- Azuma A (1992) *The biokinetics of flying and swimming*. Springer, Tokyo
- Balu R, Zhang H et al (2008) Terahertz spectroscopy of bacteriorhodopsin and rhodopsin: similarities and differences. *Biophys J* 94:3217
- Banerjee B, Chakrabarti BK (eds) (2008) *Models of brain and mind. Physical computational and psychological approaches*, vol 168, *Progress in brain research*. Elsevier, Amsterdam
- Barnes FS, Greenbaum B (eds) (2006) *Handbook of biological effects of electromagnetic fields, bioengineering and biophysical aspects of electromagnetic fields*, 3rd edn. CRC Press, Boca Raton
- Barry PH (1998) Derivation of unstirred-layer transport number equations from the Nernst-Planck flux equations. *Biophys J* 74:2903
- Bartlett DH (ed) (2010) *High-pressure bioscience and biotechnology*, *Ann NY Acad Sci* 1189:1
- Bashford CL, Pasternak CA (1986) Plasma membrane potential of some animal cells is generated by ion pumping, not by ion gradients. *Trends Biochem Sci* 11:113
- Bauer J (ed) (1994) *Cell electrophoresis*. CRC Press, Boca Raton
- Bazylinski DA, Frankel RB (2004) Magnetosome formation in prokaryotes. *Nat Rev Microbiol* 2:217
- Bejan A, Lorente S (2008) *Design with constructal theory*. Wiley, Hoboken

- Bejan A, Marden JH (2009) The constructal unification of biological and geophysical design. *Phys Life Rev* 6:85
- von Bekesy G (1960) Experiments in hearing. McGraw Hill, New York
- Bels VL, Gasc JP et al (eds) (2003) Vertebrate biomechanics and evolution, *Experimental biology Reviews*. BIOS Scientific Publishers, Oxford
- Benz R, Conti F (1986) Effects of hydrostatic pressure on lipid bilayer membranes. *Biophys J* 50:91, and 99
- Bernhardt JH (1988) The establishment of frequency dependent limits for electric and magnetic fields and evaluation of indirect effects. *Radiat Environ Biophys* 27:1
- von Bertalanffy L (2001) General system theory: foundations, development, applications. Braziller, New York
- von Bertalanffy L (1950) The theory of open systems in physics and biology. *Science* 111:23
- Bezanilla F (2008) How membrane proteins sense voltage. *Nat Rev Mol Cell Biol* 9:323
- Bingeli R, Weinstein R (1986) Membrane potentials and sodium channels: hypotheses for growth regulation and cancer formation based on changes in sodium channels and gap junctions. *J Theor Biol* 123:377
- Blumenfeld LA, Tikhonov AN (1994) Biophysical thermodynamics of intracellular processes. Springer, Berlin
- Brown BR (2003) Sensing temperature without ion channels. *Nature* 421:495
- Bullmore E, Sporns O (2009) Complex brain networks: graph theoretical analysis of structural and functional systems. *Nat Rev Neurosci* 10:186
- Bullock TH, Hopkins CD et al (eds) (2005) Electoreception. Springer handbook of auditory research. Springer, Berlin
- Bunsen R, Roscoe H (1859) Photochemische Untersuchungen. *Annalen der Physik und Chemie* 108:193
- Burkitt AN (2006) A review of the integrate-and-fire neuron model: I. Homogeneous synaptic input; II. Inhomogeneous synaptic input and network properties. *Biol Cybern* 95(1):97
- Butt HJ, Graf K et al (2006) Physics and chemistry of interfaces, 2nd edn. Wiley-VCH, Weinheim
- Campbell GS, Norman JM (1998) An introduction to environmental biophysics, 2nd edn. Springer, New York
- Cantor CR, Schimmel PR (2002) Biophysical chemistry part I: the conformation of biological macromolecules. W.H. Freeman, New York
- Cantor CR, Schimmel PR (1980) Biophysical chemistry. Part III the behavior of biological macromolecules. W.H. Freeman, New York
- Carter DR, Beaupré GS (2001) Skeletal function and form. *Mechanobiology of skeletal development, aging, and regeneration*. Cambridge University Press, Cambridge
- Cessac B, Samuelides M (2007) From neuron to neural network dynamics. *Eur Phys J* 142:7, Special topics
- Cevc G (1990) Membrane electrostatics. *Biochim Biophys Acta* 1031:311
- Charry JM, Kavet R (1987) Air ions: physical and biological aspects. CRC Press, Boca Raton
- Chein S, Chen PCY et al (2008) An introductory text to bioengineering, vol 4, *Advanced series in biomechanics*. World Scientific Publishers, Singapore
- Chen YW, Ding F et al (2008) Protein folding: then and now. *Arch Biochem Biophys* 469:4
- Clay JR (2005) Axonal excitability revisited. *Prog Biophys Mol Biol* 88:59
- Collins KD, Washabaugh MW (1985) The Hofmeister effect and the behaviour of water at interfaces. *Q Rev Biophys* 18:323
- Conn M (ed) (2008) Molecular biology of protein folding, vol 83, *Progress in molecular biology and translational science*. Academic, Amsterdam/Boston
- Crow JF, Kimura M (1970) An introduction to population genetics theory. Harper and Row, New York
- Crowther JA (1926) X-Rays and living matter. *Nature* 118:86
- Curtis RA, Lue L (2006) A molecular approach to bioseparations: protein-protein and protein-salt interactions. *Chem Eng Sci* 61:907

- Da Silva JKL, Garcia GJM et al (2006) Allometric scaling laws of metabolism. *Phys Life Rev* 3:229
- De Cock Buning TJ (1983) Thresholds of infrared sensitive tectal neurons in *Python reticulatus*, *Boa constrictor* and *Agkistrodon rhodostoma*. *J Comp Physiol* 151:461
- deNoordhout AM (1998) Cortical magnetic stimulation. *Clin Neurophys* 28:9
- Dessauer F (1964) *Quantenbiologie*. Springer, Berlin
- Devaux P, Herrmann A (eds) (2011) *Transmembrane dynamics of lipids*, Wiley series in protein and peptide science. Wiley, Hoboken
- Dieckmann S (1984) *Einwirkung mechanischer Schwingungen (Vibrationen) auf den Menschen*. In: *Arbeitsmedizin aktuell, ein Loseblattwerk für die Praxis*. Gustav Fischer Verlag, Stuttgart 14
- Digel I et al (2008) Molecular processes in biological thermosensation. *J Biophys* 2008:1
- Dillon PF et al (2006) Molecular shielding of electric field complex dissociation. *Biophys J* 901:1432
- Donath E, Voigt A (1986) Electrophoretic mobility of human erythrocytes. Theory and experimental applicability. *Biophys J* 49:493
- Dukhin SS, Zimmermann R et al (2004) Intrinsic charge and Donnan potentials of grafted polyelectrolyte layers determined by surface conductivity data. *J Colloid Interface Sci* 274:309
- Durante M, Loeffler JS (2010) Charged particles in radiation oncology. *Nat Rev Clin Oncol* 7:37
- Dyson F (1999) *Origins of life*, 2nd edn. Cambridge University Press, Cambridge
- Eigen M (1971) Selforganization of matter and the evolution of biological macromolecules. *Naturwiss* 58:465
- Eigen M (1992) *Steps toward life*. Oxford University Press, Oxford
- Elder JA, Chou CK (2003) Auditory response to pulsed radiofrequency energy. *Bioelectromagnetics* 6:162
- Engelman DM (2005) Membranes are more mosaic than fluid. *Nature* 438:578
- Everest FA, Pohlmann KC (2009) *Master handbook of acoustics*, 5th edn. McGraw-Hill, New York
- Evtodienco VY, Antonenko YN et al (1998) Increase of local hydrogen ion gradient near bilayer lipid membrane under the conditions of catalysis of proton transfer across the interface. *FEBS Lett* 425:222
- Eyal E, Bahar I (2008) Toward a molecular understanding of the anisotropic response of proteins to external forces: insights from elastic network models. *Biophys J* 94:3424
- Faller LD (2008) Mechanistic studies of sodium pump. *Arch Biochem Biophys* 476:12
- Faugeras O, Touboul J et al (2009) A constructive mean-field analysis of multi-population neural networks with random synaptic weights and stochastic inputs. *Front Comput Neurosci* 3:1
- Feistel R, Ebeling W (2011) *Physics of self-organization and evolution*. Wiley-VCH, Heidelberg
- Flanders M (2011) What is the biological basis of sensorimotor integration? *Biol Cybern* 104:1
- Fortune ES (2006) The decoding of electrosensory systems. *Curr Opin Neurobiol* 16:474
- Foster KR, Glaser R (2007) Thermal mechanisms of interaction of radiofrequency energy with biological systems with relevance to exposure guidelines. *Health Phys* 92:609
- Fowlkes JB (2008) American institute of ultrasound in medicine. Consensus report on potential bioeffects of diagnostic ultrasound. Executive summary. *J Ultrasound Med* 27:503
- Francis JT, Gluckman BJ et al (2003) Sensitivity of neurons to weak electric fields. *J Neurosci* 23:7255
- Fraser JA, Huang CLH (2007) Quantitative techniques for steady-state calculation and dynamic integrated modeling of membrane potential and intracellular ion concentrations. *Prog Biophys Mol Biol* 94:336
- Frauenfelder H, McMahon B (1998) Dynamics and function of proteins: the search for general concepts. *Proc Natl Acad Sci USA* 95:4795
- Frenkel V (2010) *Therapeutic ultrasound: mechanisms to applications*. Nova Science Publishers, Hauppauge

- Fritz M (1998) Three-dimensional biomechanical model for simulating the response of the human body to vibration stress. *Med Biol Eng Comput* 36:686
- Fry CH, Langley SEM (2005) Ion-selective electrodes for biological systems. Taylor & Francis, e-Library
- Fuhr G, Hagedorn R (1996) Cell electrorotation. In: Lynch PT, Davey MR (eds) *Electrical manipulation of cells*. Chapman & Hall, New York, p 38
- Fuhr G, Zimmermann U et al (1996) Cell motion in time-varying fields: principles and potential. In: Zimmermann U, Neil GA (eds) *Electromanipulation of cells*. CRC Press, Boca Raton, p 259
- Fullerton GD, Kanal KM et al (2006) On the osmotically unresponsive water compartment in cells. *Cell Biol Int* 30:74
- Fullerton GD, Cameron IL (2007) Water compartments in cells. *Methods Enzymol* 428:1
- Fung YC (1984) *Biodynamics*. Springer, New York
- Fung YC (1993) *Biomechanics*, 2nd edn. Springer, New York
- Gabriel C, Gabriel S et al (1996) The dielectric-properties of biological tissues. I. Literature survey. *Phys Med Biol* 41:2231
- Gabrielsson J, Weiner D (2006) *Pharmacokinetics and pharmacodynamics. Data analysis, concepts and applications*, 4th edn. Swedish Pharmaceutical Press, Stockholm
- Gadsby DC (2009) Ion channels versus ion pumps: the principal difference, in principle. *Nat Rev Mol Cell Biol* 10:344
- Garlid KD (2000) The state of water in biological systems. *Int Rev Cytol-A: Surv Cell Biol* 192:281
- Garstang M (2004) Long-distance, low-frequency elephant communication. *J Comp Physiol A Neuroethol Sens Neural Behav Physiol* 190:791
- Gause GF (1935) Experimentelle Untersuchungen über die Konkurrenz zwischen *Paramecium caudatum* und *Paramecium aurelia*. *Archiv Protistenkunde* 84:207
- German B, Wyman J (1937) The titration curves of oxygenated and reduced hemoglobin. *J Biol Chem* 117:533
- George MS, Nahas Z et al (1998) Transcranial magnetic stimulation: a new method for investigating the neuroanatomy of depression. *Adv Biol Psychiatry* 19:94
- Georgieva R, Neu B et al (1998) Low frequency electrorotation of fixed red blood cells. *Biophys J* 74:2114
- Gimsa J, Wachner D (1999) A polarization model overcoming the geometric restrictions of the Laplace solution for spheroidal cells: obtaining new equations for field-induced forces and transmembrane potential. *Biophys J* 77:1316
- Giebisch G, Tosteson D, Ussing HE (eds) (1979) *Membrane Transport in Biology*. Springer, Berlin
- Gimsa J, Wachner D (1998) A unified resistor-capacitor model for impedance, dielectrophoresis, electrorotation, and induced transmembrane potential. *Biophys J* 75:1107
- Ginzburg L, Colyvan M (2004) *Ecological orbits*. Oxford University Press, New York
- Glansdorff P, Prigogine I (1985) *Thermodynamic theory of structure, stability and fluctuations*. Wiley Interscience, London
- Glaser R (2008) *Heilende Magnete – strahlende Handys. Bioelektromagnetismus: Fakten und Legenden*. Wiley-VCH, Weinheim
- Glaser R (1989) *Grundriß der Biomechanik*, 2nd edn. Akademie-Verlag, Berlin
- Glaser R (1996) The electric properties of the membrane and the cell surface. In: Zimmermann U, Neil GA (eds) *Electromanipulation of cells*. CRC Press, Boca Raton
- Glaser R, Brumen M et al (1980) Stationäre Ionenzustände menschlicher Erythrozyten. *Biol Zbl* 99:429
- Glaser R, Donath J (1984) Stationary ionic states in human red blood cells. *Bioelectrochem Bioenerg* 13:71
- Godar DE (2005) UV doses worldwide. *Photochem Photobiol* 81:736
- Goldman DE (1943) Potential, impedance and rectification in membranes. *J Gen Physiol* 27:37

- Gracheva EO, Ingolia NT et al (2010) Molecular basis of infrared detection by snakes. *Nature* 464:1006
- Graevskij EJ, Šapiro NI (1957) *Sovremennye voprosy radiobiologii*. Nauka, Moscow
- Griffin DR (1958) *Listening in the dark*. Yale University Press, New Haven
- Griffith JH, Scheraga HA (2004) Statistical thermodynamics of aqueous solutions. I. Water structure, solutions with non-polar solutes, and hydrophobic interactions. *J Mol Struct (THEOCHEM)* 682:97
- Grodzenskij DE (1966) *Radiobiologija*. Nauka, Moscow
- Grosse C, Schwan HP (1992) Cellular membrane potentials induced by alternating fields. *Biophys J* 63:1632
- Guy Y, Sandberg M et al (2008) Determination of zeta-potential in rat organotypic hippocampal cultures. *Biophys J* 94:4561
- Haddock SHD, Moline MA et al (2010) Bioluminescence in the sea. *Ann Rev Mar Sci* 2:443
- Haggis GH, Michie D et al (1965) *Introduction to molecular biology*. Wiley, New York
- Haken H (2010) Information and self-organization: a macroscopic approach to complex systems, Springer series in synergetics. Springer, Berlin
- Hänggi P, Inchiosa ME et al (2000) Nonlinear stochastic resonance: the saga of anomalous output–input gain. *Phys Rev E* 62:6155
- Hanrahan G (2011) Artificial neural networks in biological and environmental analysis. Analytical chemistry series. CRC Press, Boca Raton
- Hassenstein B (1966) Kybernetik und biologische Forschung. In: Gessner F (ed) *Handbuch der Biologie*, 1. Frankfurt a.M, p 529
- Hegemann P (2008) Algal sensory photoreceptors. *Annu Rev Plant Biol* 59:167
- Heinrich R, Schuster S (1996) Modeling of metabolic systems. Structure, control and optimality. Chapman and Hall, New York
- Heintz PH, Kelsey CA et al (2012) *Primer on radiation biology*. Wiley-Blackwell
- Heinz WF, Hoh JH (1999) Relative surface charge density mapping with the atomic force microscope. *Biophys J* 76:528
- Hemilä S, Nummela S et al (1995) What middle ear parameters tell about impedance matching and high frequency hearing. *Hear Res* 85:31
- Hertel U (1963) *Struktur, Form, Bewegung, Krauskopf*, Mainz
- Hianik T, Passechnik VI (1995) Bilayer lipid membranes. Structure and mechanical properties. Kluwer, Dordrecht
- Hirata A, Fujiwara O (2009) Modeling time variation of blood temperature in a bioheat equation and its application to temperature analysis due to RF exposure. *Phys Med Biol* 54:N189
- Hodgkin AL, Huxley AF (1952) A quantitative description of membrane current and its application to conduction and excitation in nerve. *J Physiol* 117:500
- Hofmeister F (1888) Zur Lehre von der Wirkung der Salze. *Archiv für experimentelle Pathologie und Pharmakologie* 24:247
- Holder DS (ed) (2005) *Electrical impedance tomography: methods, history and applications*, Series in medical physics and biomedical Engineering. Institute of Physics Publishing, Bristol
- Honig BH et al (1986) Electrostatic interactions in membranes and proteins. *Ann Rev Biophys Chem* 151:163
- Hosford WF (2009) *Mechanical behavior of materials*, 2nd edn. Cambridge University Press, Cambridge
- Hudspeth AJ (2008) Making an effort to listen: mechanical amplification in the ear. *Neuron* 59:530
- Hug O, Kellerer AM (1966) *Stochastik der Strahlenwirkung*. Springer, Heidelberg
- Huxley A (2002) From overshoot to voltage clamp. *Trends Neurosci* 25:553
- ICNIRP (2009) Guidelines on limits of exposure to static magnetic fields. *Health Phys* 96:504
- ICNIRP (1998) Guidelines for limiting exposure to time-varying electric, magnetic, and electromagnetic fields (up to 300 GHz). *Health Phys* 74:494
- Imanidis G, Luetolf P (2006) An extended model based on the modified Nernst-Planck equation for describing transdermal iontophoresis of weak electrolytes. *J Pharm Sci* 95:1434

- Israelachvili J (1994) Intermolecular and surface forces, 2nd edn. Academic, London
- Jaffe LF (1979) Control of development by ionic currents. In: Cone RA, Dowling J (eds) Membrane transduction mechanisms. Raven, New York
- Jeffries C (1989) Mathematical modeling in ecology. A workbook for students. Birkhäuser, Boston
- Kalmijn AJ, Gonzalez IF et al (2002) The physical nature of life. *J Physiol Paris* 96:355
- Karger CP, Jakel O et al (2010) Dosimetry for ion beam radiotherapy. *Phys Med Biol* 55:R193
- Katchalsky A, Curran PF (1965) Nonequilibrium thermodynamics in biophysics. Harvard University Press, Cambridge
- Kauffman SA (1993) The origin of order. Oxford University Press, New York
- Keynes RD (1994) The kinetics of voltage-gated ion channels. *Q Rev Biophys* 27:339
- Kiefer J (1990) Biological radiation effects. Springer, Berlin
- Kjelstrup S, Bedeaux D (2008) Non-equilibrium thermodynamics of heterogeneous systems, vol 16, Series on advances in statistical mechanics. World Scientific Publishing, Singapore
- Klipp E, Liebermeister W et al (2009) Systems biology, a textbook. Wiley-VCH, Weinheim
- Kloss E, Courtemanche N et al (2008) Repeat-protein folding: new insights into origins of cooperativity, stability, and topology. *Arch Biochem Biophys* 469:83
- Knorre WA (1981) Pharmakokinetik. Akademie-Verlag, Berlin
- Koch I, Reisig W et al (eds) (2011) Modeling in systems biology. The petri net approach. Springer, London
- Kogel Avd, Joiner M (eds) (2009) Basic clinical radiobiology. Hodder Arnold
- Kolb JF, Kono S et al (2006) Nanosecond pulsed electric field generators for the study of subcellular effects. *Bioelectromagnetics* 27:172
- Kontturi K, Murtomaki L et al (2008) Ionic transport processes: in electrochemistry and membrane science. Oxford University Press, Oxford
- Kunz W, Lo Nostro P et al (2004) The present state of affairs with Hofmeister effects. *Curr Opin Colloid Interface Sci* 9:1
- Kuyucak S, Bastug T (2003) Physics of ion channels. *J Biol Phys* 29:429
- Lalchev Z, Todorov R et al (2008) Thin liquid films as a model to study surfactant layers on the alveolar surface. *Curr Opin Colloid Interface Sci* 13:183
- Lammert PE, Prost J et al (1996) Ion drive for vesicles and cells. *J Theor Biol* 178:387
- Lauffer MA (1975) Entropy-driven processes in biology. Springer, Berlin
- Läuger P (1991) Electrogenic ion pumps. Sinauer Associates, Sunderland
- Le Pichon A, Blanc E et al (eds) (2010) Infrasound monitoring for atmospheric studies. Springer, Dordrecht
- Leff HS, Rex AF (1990) Maxwell's demon. Entropy, information, computing. Adam Hilger, Bristol
- Lesk AM (2002) Introduction to bioinformatics. Oxford University Press, Oxford
- Leitgeb N (1990) Strahlen, Wellen, Felder. Ursachen und Auswirkungen auf Umwelt und Gesundheit. Georg Thieme Verlag, Stuttgart
- Lerche D, Bäuml H (1984) Moderate heat treatment of only red blood cells slows down the rate of RBC-RBC aggregation in plasma. *Biorheology* 21:393
- Leventhall G (2007) What is infrasound? *Prog Biophys Mol Biol* 93:130
- Levitt DG, Mlekoday HJ (1983) Reflection coefficient and permeability of urea and ethylene glycol in the human red cell membrane. *J Gen Physiol* 81: 239
- Leyton L (1975) Fluid behaviour in biological systems. Clarendon, Oxford
- Lilaonitkul W, Guinan JJ Jr (2009) Reflex control of the human inner ear: a half-octave offset in medial efferent feedback that is consistent with an efferent role in the control of masking. *J Neurophysiol* 101:1394
- Lima-Mendez G, Van Helden J (2009) The powerful law of the power law and other myths in network biology. *Mol Biosyst* 5:1482
- Lipowsky R, Klumpp S (2005) 'Life is motion': multiscale motility of molecular motors. *Phys A Stat Mech Its Appl* 352:53

- Loehe JR, Donohue MD (1997) Recent advances in modeling thermodynamic properties of aqueous strong electrolyte systems. *AIChE J* 43:180
- Lommerse PHM, Spaink HP et al (2004) In vivo plasma membrane organization: results of biophysical approaches. *Biochim Biophys Acta* 1664:119
- Luckey M (2008) Structural biology with biochemical and biophysical foundations. Cambridge University Press, Cambridge
- Lynch PT, Davey MR (eds) (1996) Electrical manipulation of cells. Chapman & Hall, New York
- Lynden-Bell RM, Morris SC et al (eds) (2010) Water and life. The unique properties of H₂O. CRC Press, Boca Raton
- Ma'ayan A (2009) Insights into the organization of biochemical regulatory networks using graph theory analyses. *J Biol Chem* 284:5451
- MacGinitie LA (1995) Streaming and piezoelectric potentials in connective tissues. *Adv Chem Ser* 250:125
- MacKinnon R (2003) Potassium channels. Minireview. *FEBS Lett* 555:62
- Mahmud G et al (2009) Directing cell motions on micropatterned ratchets. *Nat Phys* 5:606
- Maini PK, Othmer HG (eds) (2000) Mathematical models for biological pattern formation. Springer, New York
- Makarov VA, Feig M et al (1998) Diffusion of solvent around bimolecular solutes: a molecular dynamics simulation study. *Biophys J* 75:150
- Maniewski R (1991) Magnetic studies on mechanical activity of the heart. *Crit Rev Biomed Engin.* 19:203
- Mansfield NJ (2005) Impedance methods (apparent mass, driving point mechanical impedance and absorbed power) for assessment of the biomechanical response of the seated person to whole-body vibration. *Ind Health* 43:378
- Maret G, Boccarda, N et al (1986) Biophysical Effects of Steady Magnetic Fields. Springer, Berlin
- Martin C, Bhatt K et al (2001) Shaping in plant cells. *Curr Opin Plant Biol* 4:540
- Martínez-Reina J, García-Aznar JM et al (2009) A bone remodelling model including the directional activity of BMUs. *Biomech Model Mechanobiol* 8:111
- Matthew JB (1985) Electrostatic effects in proteins. *Annu Rev Biophys Biophys Chem* 14:387
- McCaig CD, Rajnicek AM et al (2005) Controlling cell behavior electrically: current views and future potential. *Physiol Rev* 85:943
- McDonnell MD, Abbott D (2009) What is stochastic resonance? Definitions, misconceptions, debates, and its relevance to biology. *PLoS Comput Biol* 5:1
- McLauchlan KA, Steiner UE (1991) The spin-correlated radical pair as a reaction intermediate. *Mol Phys* 73:241
- McLaughlin S (1989) The electrostatic properties of membranes. *Annu Rev Biophys Biophys Chem* 18:113
- McLaughlin S, Murray D (2005) Plasma membrane phosphoinositide organization by protein electrostatics. *Nature* 43:8605
- McMahon HT, Gallop JL (2005) Membrane curvature and mechanisms of dynamic cell membrane remodelling. *Nature* 43:8590
- Meinhardt H (2008) Models of biological pattern formation: from elementary steps to the organization of embryonic axes. *Curr Top Dev Biol* 81:1
- Meinhardt H (2009) Models for the generation and interpretation of gradients. *Cold Spring Harb Perspect Biol* 1:a001362
- Mewis J, Wagner NJ (2009) Thixotropy. *Adv Colloid Interface Sci* 147–148:214
- Miyakoshi J (2005) Effects of static magnetic fields at the cellular level. *Prog Biophys Mol Biol* 87:213
- Möglich A, Yang X et al (2010) Structure and function of plant photoreceptors. *Annu Rev Plant Biol* 61:21
- Møller H (1984) Effects of infrasound on man. Aalborg University Press, Aalborg
- Montali M, Credi A et al (2006) Handbook of photochemistry, 3rd edn. CRC Press, Boca Raton

- Moore LR, Zborowski M et al (1998) Lymphocyte fractionation using immunomagnetic colloid and a dipole magnet flow cell sorter. *J Biochem Biophys Methods* 37:11
- Mozumder A (1999) Fundamentals of radiation chemistry. Academic, San Diego
- Muñoz V (2007) Conformational dynamics and ensembles in protein folding. *Annu Rev Biophys Biomol Struct* 36:395
- Murray JD (2003) Mathematical biology, 3rd edn. Springer, New York
- Mycielska ME, Djamgoz MBA (2004) Cellular mechanisms of direct-current electric field effects: galvanotaxis and metastatic disease. *J Cell Sci* 117:1631
- Netter H (1959) Theoretische Biochemie. Springer, Berlin
- Nienhaus GU (2008) The green fluorescent protein: a key tool to study chemical processes in living cells. *Angew Chem Int Ed Engl* 47:8992
- Nigg BN, Herzog W (eds) (2007) Biomechanics of the musculo-skeletal system, 3rd edn. Wiley, Chichester
- Niklas KJ (1992) Plant biomechanics. An engineering approach to plant form and function. The University of Chicago Press, Chicago
- Nishigori C (2006) Cellular aspects of photocarcinogenesis. *Photochem Photobiol Sci* 5:208
- Nuccitelli R, Nuccitelli P et al (2008) Imaging the electric field associated with mouse and human skin wounds. *Wound Repair Regen* 16:432
- Nunez PL, Srinivasan R (2006) Electric fields of the brain: the neurophysics of EEG. Oxford University Press, Oxford
- Obermayer B, Frey E (2010) Error thresholds for self- and cross-specific enzymatic replication. *J Theor Biol* 267:653
- Okubo A, Levin SA (2010) Diffusion and ecological problems: modern perspectives, 2nd edn. Springer, New York
- Ogawa N, Oku H et al (2006) A physical model for galvanotaxis of paramecium cell. *J Theor Biol* 242:314
- Ohta S, Suzuki K et al (2008) Gene transduction by sonoporation. *Dev Growth Differ* 5:517
- Okada Y (2004) Ion channels and transporters involved in cell volume regulation and sensor mechanisms. *Cell Biochem Biophys* 41:233
- Oleinikova A, Smolin N et al (2007) Influence of water clustering on the dynamics of hydration water at the surface of a lysozyme. *Biophys J* 93:2986
- Olsvik Ø, Popovic T et al (1994) Magnetic separation techniques in diagnostic microbiology. *Clin Microbiol Rev* 7:43
- O'Neill P, Wardman P (2009) Radiation chemistry comes before radiation biology. *Int J Radiat Biol* 85:9
- Oomens C, Brekelmans M et al (2009) Biomechanics. Concept and computation, Cambridge texts in biomedical engineering. Cambridge University Press, Cambridge
- Orazem ME, Tribollet B (2008) Electrochemical impedance spectroscopy. Wiley
- Orr L, Govindjee (2010) Photosynthesis online. *Photosynth Res* 105:167
- Overbeek JTG (1956) The Donnan equilibrium. *Prog Biophys Biophys Chem* 6:58
- Owens RGA (2006) New microstructure-based constitutive model for human blood. *J Non-Newtonian Fluid Mech* 140:57
- Özkaya N, Nordin M (1999) Fundamentals of biomechanics. Equilibrium, motion, and deformation, 2nd edn. Springer, New York
- Pain RH (2000) Mechanism of protein folding, 2nd edn. Oxford University Press, Oxford
- Palmer T (1995) Understanding enzymes. Prentice Hall, London
- Parsegian VA (2002) Protein-water interactions. *Int Rev Cytol Surv – Cell Biol* 215:1
- Parsegian VA (2006) Van der Waals forces: a handbook for biologists, chemists, engineers and physicists. Cambridge University Press, Cambridge
- Pauwels F (1980) Biomechanics of the locomotor apparatus. Springer, Berlin
- Pearson K (1892) The grammar of science. Walter Scott, London
- Penzlin H (1991) Lehrbuch der Tierphysiologie, 5th edn. Fischer Verlag, Jena

- Péqueux AJR, Gilles R (1985) High pressure effects on selected biological systems. Springer, Berlin
- Peters RC, Eeuwes LBM et al (2007) On the electroreception threshold of aquatic vertebrates with ampullary or mucous gland electroreceptor organs. *Biol Rev* 82:361
- Pethig R (1979) Dielectric and electronic properties of biological materials. Wiley, Chichester
- Pethig R, Kell DB (1987) The passive electrical properties of biological systems: their significance in physiology, biophysics and biotechnology. *Phys Med Biol* 32:933
- Piazena H, Kelleher DK (2010) Effects of infrared-a irradiation on skin: discrepancies in published data highlight the need for an exact consideration of physical and photobiological laws and appropriate experimental settings. *Photochem Photobiol* 86:687
- Plotkin SS, Onuchic JN (2002) Understanding protein folding with energy landscape theory. Part I: basic concepts. *Q Rev Biophys* 35:111
- Plusquellic DF, Siegrist K et al (2007) Applications of terahertz spectroscopy in biosystems. *Chemphyschem* 8:2412
- Pohl P, Saparov SM et al (1998) The size of the unstirred layer as a function of the solute diffusion coefficient. *Biophys J* 75:1403
- Pollak GD, Casseda JH (1989) The neural basis of echolocation in bats. Springer, New York
- Prat H, Roberge L (1960) Variations de la thermogenèse en fonction du vieillissement chez la souris. *Rev Can Biol* 19:80
- Precht H, Christophersen J et al (1955) *Temperatur und Leben*. Springer, Berlin
- Prescher JA, Contag CH (2010) Guided by the light: visualizing biomolecular processes in living animals with bioluminescence. *Curr Opin Chem Biol* 14:80
- Prettejohn BJ, Berryman MJ et al (2011) Methods for generating complex networks with selected structural properties for simulations: a review and tutorial for neuroscientists. *Front Comput Neurosci* 5:1
- Prigogine I (1967) Introduction to thermodynamics of irreversible processes, 3rd edn. Wiley Interscience, New York
- Purves D, Augustine GJ et al (2008) *Neuroscience*, 4th edn. Sinauer Associates, Sunderland
- Ramamoorthy S, Zha DJ et al (2010) The biophysical origin of traveling-wave dispersion in the cochlea. *Biophys J* 99:1687
- Raschke TM (2006) Water structure and interactions with protein surfaces. *Curr Opin Struct Biol* 16:152
- Rashevsky N (1960) *Mathematical biophysics, physico-mathematical foundations of biology*. Dover, New York
- Raudino A, Mauzerall D (1986) Dielectric properties of the polar head group region of zwitterionic lipid bilayers. *Biophys J* 50:441
- Reilly JP (1998) *Applied bioelectricity. From electrical stimulation to electropathology*. Springer, New York
- Reiter R (1992) *Phenomena in atmospheric and environmental electricity*. Elsevier, Amsterdam
- Renger G (1994) Biologische Wasserspaltung durch Sonnenlicht im Photosyntheseapparat. *Chemie in unserer Zeit* 28:118
- Renshaw E (1990) *Modelling biological populations in space and time*. Cambridge University Press, Cambridge
- Rhumbler L (1898) *Physikalische Analyse von Lebenserscheinungen der Zelle. 1. Bewegung, Nahrungsaufnahme, Defäkation, Vacuolen-Pulsation und Gehäusebau bei lobosen Rhizopoden*. *Arch Entwicklungsmechanik der Organismen* 7:103
- Ridley AJ, Whiteside JR et al (2009) Cellular and sub-cellular responses to UVA in relation to carcinogenesis. *Int J Radiat Biol* 85:177
- Riesz P, Kondo T (1992) Free radical formation induced by ultrasound and its biological implications. *Free Radic Biol Med* 13:247
- Rittweger J (2010) Vibration as an exercise modality. How it may work, and what its potential might be. *Eur J Appl Physiol* 108:877
- Ritz T, Damjanovi A et al (2002) The quantum physics of photosynthesis. *Chemphyschem* 3:243

- Riu PJ, Rosell J et al (eds) (1999) Electrical bioimpedance methods. *Ann N Y Acad Sci* 873:1
- Rode BM, Fitz D et al (2007) The first steps of chemical evolution towards the origin of life. *Chem Biodivers* 4:2674
- Roduit C, Van Der Goot FG et al (2008) Elastic membrane heterogeneity of living cells revealed by stiff nanoscale membrane domains. *Biophys J* 94:1521
- Roederer JG (2008) Introduction to the physics and psychophysics of music, 4th edn. Springer, New York
- Rossing TD, Fletcher NH (2004) Principles of vibration and sound, 2nd edn. Springer, New York
- Roux B, Allen T et al (2004) Theoretical and computational models of biological ion channels. *Q Rev Biophys* 37:15
- Rowbottom M, Susskind C (1984) Electricity and medicine. History of their interaction. San Francisco Press, San Francisco
- Rowland M, Tozer TN (2011) Clinical pharmacokinetics and pharmacodynamics. Concepts and applications, 4th edn. Wolters Kluwer, Philadelphia
- Rozhok A (2008) Orientation and navigation in vertebrates. Springer, Berlin
- Rubinacci A, Covini M et al (2002) Bone as an ion exchange system. Evidence for a link between mechanotransduction and metabolic needs. *Am J Physiol Endocrinol Metab* 282:E851
- Rudall BH (2004) Contemporary systems and cybernetics – new advances in biocybernetics. *Kybernetes* 33:1084
- Rumler L (1898) Physikalische Analyse der Lebenserscheinungen der Zelle. *Arch Entw Mech* 7:103
- Santabarbara S, Galuppini L et al (2010) Bidirectional electron transfer in the reaction centre of photosystem I. *J Integr Plant Biol* 52:735
- Schenck JF (2005) Physical interactions of static magnetic fields with living tissues. *Prog Biophys Mol Biol* 87:185
- Schienze A, Stark R et al (1997) Effects of low-frequency magnetic fields on electrocortical activity in humans: a sferics simulation study. *Int J Neurosci* 90:21–36
- Schmid R (2001) Recent advances in the description of the structure of water, the hydrophobic effect, and the like-dissolves-like rule. *Monatshefte für Chemie* 132:1295
- Schmidt-Nielsen K (1999) Scaling. Why is animal size so important. Cambridge University Press, Cambridge
- Schmid-Schönbein GW, Woo SL-Y et al (1986) Frontiers in biomechanics. Springer, New York
- Schnakenberg J (1981) Thermodynamic network analysis of biological systems, 2nd edn. Springer, Berlin
- Schneck DJ (ed) (1980) Biofluid mechanics. Plenum, New York
- Schrödinger E (1944) What is life? Cambridge University Press, London
- Schroer MA, Paulus M et al (2010) High-pressure SAXS study of folded and unfolded ensembles of proteins. *Biophys J* 99:3430
- Schwan HP (1957) Electrical properties of tissue and cell suspensions. *Adv Biol Med Phys* 5:147
- Shannon CE, Weaver W (1962) The mathematical theory of communication. University of Illinois Press, Urbana
- Shi R, Borgens RB (1995) Three-dimensional gradients of voltage during development of the nervous system as invisible coordinates for the establishment of embryonic pattern. *Dev Dyn* 202:101
- Siebert F, Hildebrandt P (2008) Vibrational spectroscopy and life science. Wiley-VCH, Weinheim
- Simon R, Holderied MW et al (2011) Floral acoustics: conspicuous echoes of a dish-shaped leaf attract bat pollinators. *Science* 333:631
- Singhal GS, Renger G et al (eds) (1999) Concepts in photobiology and photogenesis. Kluwer, Boston
- Singer SJ, Nicolson GL (1972) The fluid mosaic model of the structure of cell membranes. *Science* 175:720
- Sizov F (2010) THz radiation sensors. *Opto-electron Rev* 18:10
- Skalak R, Chien S (eds) (1987) Handbook of bioengineering. McGraw-Hill, New York

- Solov'yov IA, Schulten K (2009) Magnetoreception through cryptochrome may involve superoxide. *Biophys J* 96:4804
- von Sonntag C (2006) Free-radical-induced DNA damage and its repair. A chemical perspective. Springer, Berlin
- Starke-Peterkovic T, Turner N et al (2005) Electric field strength of membrane lipids from vertebrate species membrane lipid composition and Na⁺K⁺ATPase molecular activity. *Am J Physiol* 288:R663
- Stein WD (1990) Channels, carriers, and pumps. An introduction to membrane transport. Academic, San Diego
- Strait BJ, Dewey TG (1996) The Shannon information entropy of protein sequences. *Biophys J* 71:148
- Sukhorukov VL, Mussauer H et al (1998) The effect of electrical deformation forces on the electropemabilization of erythrocyte membranes in low- and high-conductivity media. *J Membr Biol* 163:235
- Sun DD, Guo XE et al (2004) The influence of the fixed negative charges on mechanical and electrical behaviors of articular cartilage under unconfined compression. *J Biomech Eng* 126:6
- Suslick KS, Doktycz SJ (1991) Effects of ultrasound on surfaces and solids. In: Mason TJ (ed) *Advances in sonochemistry*. CTJAI Press, Greenwich
- Sutherland JC, Griffin KP (1981) Absorption spectrum of DNA for wavelengths longer than 320 nm. *Radiat Res* 86:399
- Svobodova A, Vostalova J (2010) Solar radiation induced skin damage: review of protective and preventive options. *Int J Radiat Biol* 86:999
- Syganov A, von Klitzing E (1999) (In)validity of the constant field and constant currents assumptions in theories of ion transport. *Biophys J* 76:768
- Takahashi S, Casaleggio A, et al. (1989) Study of bound water of poly-adenine using high frequency dielectric measurements. *Biophys J* 49:1003
- Takahashi M, Chiba K et al (2007) Free-radical generation from collapsing microbubbles in the absence of a dynamic stimulus. *J Phys Chem B* 111:1343
- Takahashi S (1963) Dielectric dispersion of DNA. *J Mol Biol* 7:455
- Talbot L, Berger SA (1974) Fluid-mechanical aspects of the human circulation. *Am Sci* 62:671–682
- ter Haar G (2007) Therapeutic applications of ultrasound. *Prog Biophys Mol Biol* 93:111
- Thom R (1969) Topological models in biology. *Topology* 8:313
- Thomas JA, Moss CF et al (eds) (2004) *Echolocation in bats and dolphins*. University Chicago Press, Chicago
- Thompson DW (1966) *On growth and form*. University Press, Cambridge
- Tinoco I, Sauer K et al (2002) *Physical chemistry, principles and applications in biological sciences*, 4th edn. Prentice-Hall, New Jersey
- Thlusty TA (2010) Colorful origin for the genetic code: information theory, statistical mechanics and the emergence of molecular codes. *Phys Life Rev* 7:362
- Toyoshima C, Nakasako M et al (2000) Crystal structure of the calcium pump of sarcoplasmic reticulum at 2.6 Å resolution. *Nature* 405:647
- Turchin P (2003) *Complex population dynamics*. Princeton University Press, Princeton
- Turing AM (1952) The chemical basis of morphogenesis. *Philos Trans R Soc Lond B Biol Sci* 237:37
- Ussing HH (1949) The distinction by means of tracers between active transport and diffusion. The transfer of iodide across the isolated frog skin. *Acta Physiol Scand* 19:43
- Varma R, Selman JR (1991) *Characterization of electrodes and electrochemical processes*. Wiley, New York
- VDI-2057 (2002) Human exposure to mechanical vibrations. Whole-body vibration. Düsseldorf
- Veldhuizen EJA, Haagsman HP (2000) Role of pulmonary surfactant components in surface film formation and dynamics. *Biochim Biophys Acta* 1467:255

- Verkman AS, Alpern RJ (1987) Kinetic transport model for cellular regulation of pH and solute concentration in the renal proximal tubule. *Biophys J* 51: 533
- Verwey EJW, Overbeek JTG (1948) *Theory of the stability of lyophobic colloids*. Elsevier, Amsterdam
- Videler J (1993) *Fish swimming*. Chapman and Hall, London
- Vogel S (1994) *Life in moving fluids. The physical biology of flow*, 2nd edn. Princeton University Press, New Jersey
- Voet D, Voet JG (2011) *Biochemistry*, 4th edn. Wiley, New York, International Edition
- Voigt A, Donath E (1989) Cell surface electrostatics and electrokinetics. In: Glaser R, Gingell D (eds) *Biophysics of the cell surface*, vol 5, Springer series of biophysics. Springer, Berlin, p 75
- Waite L, Fine J (2007) *Applied biofluid mechanisms*. McGraw Hill, New York
- Wang ET, Yin YL et al (2003) Physiological electric fields control the G(1)/S phase cell cycle checkpoint to inhibit endothelial cell proliferation. *FASEB J* 17:U26
- Wartlick O, Kicheva A et al (2009) Morphogen gradient formation. *Cold Spring Harb Perspect Biol* 1:a001255
- Wassermann EM (1998) Risk and safety of repetitive transcranial magnetic stimulation: report and suggested guidelines from the international workshop on the safety of repetitive transcranial magnetic stimulation. *Clin Neurophysiol* 108:1
- Watanabe H, Vriens SH et al (2002) Heat-evoked activation of TRPV4 channels in a HEK293 cell expression system and in native mouse aorta endothelial cells. *J Biol Chem* 277:47044
- Waterman TH, Morowitz HJ (eds) (1965) *Theoretical and mathematical biology*. Blaisdell, New York
- Weaver JC, Vaughan TE et al (1998) Theoretical limits on the threshold for the response of long cells to weak extremely low frequency electric fields due to ionic and molecular flux rectification. *Biophys J* 75:2251
- Webb PW, Weihs D (eds) (1983) *Fish – biomechanics*. Praeger, New York
- Wernet P, Nordlund D et al (2004) The structure of the first coordination shell in liquid water. *Science* 304:995
- West GB, Brown JH (2004) Life's universal scaling laws. *Phys Today* 57:36
- Westerhoff HV, van Dam K (1987) *Thermodynamics and control of biological free-energy transduction*. Elsevier Science, New York
- Whitmarsh J, Govindjee The photosynthetic process. See: Singhal et al (1999)
- Wiener N (1961) *Cybernetics: or control and communication in the animal and the machine*. Hermann & Cie/MIT, Paris/Cambridge
- Withers PC (1992) *Comparative animal physiology*. Saunders College Publishing, Fort Worth
- Wiltschko R, Wiltschko W (1995) *Magnetic orientation in animals*. Springer, Berlin
- Wilson BW, Stevens RG et al (eds) (1990) *Extremely low frequency electromagnetic fields: the question of cancer*. Battelle, Columbus
- Wilson CE (1994) *Noise control measurement, analysis, and control of sound and vibration*. Krieger, Melbourne
- Winter R (ed) (2003) *Advances in high pressure bioscience and biotechnology*. Springer, Berlin
- Wolff J (1986) *Das Gesetz der Transformation der Knochen*. (1892) (The law of bone remodeling), (trans: Marquet P, Furlong R). Springer, New York
- Yang F, Cui YY et al (2010) Thermosensitive TRP channel pore turret is part of the temperature activation pathway. *Proc Natl Acad Sci USA* 107:7083
- Yockey HP (2005) *Information theory, evolution and the origin of life*. Cambridge University Press, Cambridge
- Zborowski M, Osters GR et al (2003) Red blood cell magnetophoresis. *Biophys J* 84:2638
- Zeeman EC (1976) Catastrophe theory. *Sci Am* 234:65
- Zhang YJ, Cremer PS (2006) Interactions between macromolecules and ions: the Hofmeister series. *Curr Opin Chem Biol* 10:658
- Zhao F, Koike T et al (2009) Finite element analysis of the middle ear transfer functions and related pathologies. *Med Eng Phys* 31:907

- Zhao M, Song B et al (2006) Electrical signals control wound healing through phosphatidylinositol-3-OH kinase-g and PTEN. *Nature* 442:457
- Zhou S-A, Uesaka M (2009) An analysis of ion channels in time-varying fields by the generalized Poisson-Nernst-Planck theory. *Int J Appl Electromagn Mech* 29:25
- Zielkiewicz J (2005) Structural properties of water. Comparison of the SPC, SPCE, TIP4P, and TIP5P models of water. *J Chem Phys* 123:104501
- Ziemann F, Radler J et al (1994) Local measurements of viscoelastic moduli of entangled actin networks using an oscillating magnetic bead micro-rheometer. *Biophys J* 66:2210
- Zimmer KG (1961) *Studies of quantitative radiation biology*. Oliver and Boyd, Edinburgh
- Zimmermann U (1989) Water relations of plant cells: pressure probe techniques. *Meth Enzymol* 174:338
- Zimmermann U, Rüger S et al (2010) Effects of environmental parameters and irrigation on the turgor pressure of banana plants measured using the non-invasive, online monitoring leaf patch clamp pressure probe. *Plant Biol* 12:424
- Zimmermann U, Schneider H et al (2004) Water ascent in tall trees: does evolution of land plants rely on a highly metastable state? *New Phytol* 162:575
- Zimmermann U, Neil GA (1996) *Electromanipulation of cells*. CRC Press, Boca Raton
- Zimmermann U, Steudle E (1978) Physical aspects of water relation of plant cells. *Adv Bot Res* 6:45
- Zotín AI (1990) *Thermodynamic bases of biological processes: physiological reactions and adaptations*. DeGruyter, Berlin

Subject Index

- α -helix
 - dipole moment, 38
 - structure, 58f
- α -radiation, 317f
- β -dispersion, 196f
- β -pleated sheet of proteins, 58f
- β -radiation, 318
- γ -radiation, 317ff
- π -helix of proteins, 59
- A**
- α -helix
 - dipole moment, 38
 - structure, 58f
- α -radiation, 317f
- Acid-base equilibrium, 64f
- Action potential, 177ff
- Activated state, 21f
- Activation energy, 19ff
- Active transport, 166ff
- Activity
 - chemical, definition, 104
 - coefficient, 104f
 - measurement, 142f
- Adenosine triphosphate. *See* ATP
- Admittance, 193
- Affinity, chemical, 126
- Agglutination, 55
- Air
 - electric conductivity, 287
 - ions, 287
 - kinematic viscosity, 227
 - magnetic susceptibility, 280
 - thermal conductivity, 250
- ALARA-principle, 274, 329
- Allometric relations
 - basic equation, 240
 - Kleiber's approach, 241
 - of metabolism, 242
 - Rubner's approach, 240f
 - of wing area, 237
- Amino acids
 - generation, 366f
 - isoelectric point, 65
 - sensibility to radiation, 312
 - sequences in proteins, 57, 367f
 - transport, 167
- Amoeba, movement, 69
- Ampholytes, 38f, 63ff
- Ampullae of Lorenzini, 289f
- Ampullar organ, 288f
- Anisotropy mechanical, 219
- Anisotrope systems, 5f, 112
- Antifield rotation, 204f
- Antiport, 166f
- Aperiodic
 - limit, 339, 348
 - structure, 15
- Area moment of inertia, 223
- Arrhenius
 - equation, 19ff
 - plot, 20f
- Atmospherics (=Spherics), 294
- Atomic force microscopy (AFM), 79
- ATPases, 166ff, 171f, 184
- Auditory effect of Microwaves, 302
- Avogadro number, 18
- Axon, propagation of impulses, 176f

B

Barotrauma, 253
 Basal metabolic rate (BMR), 241, 297
 Basic restrictions of EMF interactions, 294
 Bat, echolocation, 274ff
 Békésy-theory of hearing, 264ff
 Bending of a rod, 220ff
 Bernard cells, 118
 Bernoulli-equation, 235
 Bifurcation, 342
 Bilayer couple mechanism, 79
 Bilayer of lipids (BLM), 71f
 Bimolecular lipid membrane (BLM), 71f
 Bingham-plastic fluid, 210
 Bioavailability of drugs, 355
 Biochemical reaction kinetics, 350ff
 Biocybernetics, 3, 346, 371f
 Bioelectricity, 1
 Bioheat equation, 251f, 271, 301f
 Bioinformatics, 12
 Bioluminescence, 313
 Bioluminescence imaging techniques (BLI), 313
 Biomechanics, 207f
 Bionics, 236
 Biophysics
 history, 2f
 subject, 1f
 Birds
 area of wings, 237
 echolocation, 276
 flying mechanisms, 237f
 Reynolds numbers, 228, 236
 Bit (binary digit), 10
 Black body, 250
 Blau-Altenburger equation, 331
 BLM. *See* Bimolecular lipid membrane (BLM)
 Blood flow
 critical Reynolds numbers, 233f
 diagnosis by ultrasound, 233
 entrance effect, 230f
 Fahraeus-Lindqvist effect, 230
 microrheology, 230
 Moens-Korteweg equation, 231
 non-Newtonian behavior, 213
 pulsed flow, 231
 turbulences, 233
 velocities, 233
 viscosity, 213, 229
 Blood vessels
 branching, 233f
 diameter, 233
 eddy currents, 293

Reynolds number, 233
 smooth muscles, 233
 viscoelasticity, 232f
 Bohr's radius, 55
 Boltzmann constant, 8, 12
 Boltzmann's equation for entropy, 8ff, 16f
 Boltzmann's law of energy distribution, 16ff, 47f
 Bond energy
 of covalent bonds, 25
 of hydrogen bonds, 26
 Bonds
 by hydrogen bridges, 26
 hydrophobic, 44f
 intermolecular, 50ff
 by Van-der-Waals force, 52ff
 Bone
 healing by diathermy, 273
 healing by electric field application, 291
 mechanical stability, 220f
 passive electrical properties, 196
 piezoelectric properties, 192
 streaming potential, 192
 Borelli, Alfonso, 2, 208
 Boundary layers, 234ff, 250
 Bragg curve of ionization, 320
 Brain models, 372f
 Bronchial dynamics, 69
 Bronchial dynamics concept of transport, 185
 Brownian
 motion, 30ff
 ratchet, 122f
 Buffer properties, 64f
 Build-up phenomenon, 320
 Bunsen-Roscoe reciprocity rule, 316
 Bystander effect
 of ionizing radiation, 323ff
 of UV-radiation, 316

C

Cable theory of propagation of nerve
 impulse, 178
 Caisson disease, 253
 Calcium
 ATPase, 168, 184
 role in processes of cellular control, 56
 signal system, 168f
 Cancer
 diagnosis, 197, 306
 induction by radiation, 314, 324f, 329
 therapy, 280, 318
 Capacitor equation, 174

- Capacity, specific of cell membrane, 74, 89f
- Capillary viscosimeter, 212
- Carotenoids, 310f
- Cartilage
 - Donnan-osmotic properties, 147
 - piezoelectric properties, 192
 - viscoelastic properties, 214
- Catastrophe theory, 342
- Cauchy condition, 98
- Caveolae, 76
- Cavitation, 272ff
- Cell
 - deformation by electric fields, 204
 - deformation by shear stress, 213
 - electrophoresis, 83ff
 - fusion, 203
 - galvanotaxis, 290
 - manipulation by electric fields, 202ff
 - proliferation, correlation with membrane potential, 176
 - swelling, 78
 - viscoelastic properties, 218f
 - volume regulation, 138, 147
- Cell membrane
 - action potential, 176ff
 - active transport, 165ff
 - anchored-protein picket model, 76
 - conductivity, 89, 172ff
 - capacity, specific, 74, 89f, 201
 - caveolae, 76
 - deformation, 76ff
 - dielectric constant, 89, 174
 - diffusion potential, 162ff, 171ff
 - elasticity modulus, 77f, 219
 - electric break down, 202f
 - electric field strength, 37, 92f, 175, 183, 203
 - electric potential, 89ff, 199f
 - electrostatic structure, 87
 - excitability, 295ff
 - fluidity, 74, 214f
 - fluid-mosaic-model, 74
 - function in general, 67
 - induction of electrical potential, 200ff
 - ion transport, 165ff
 - magnetic anisotropy, 283
 - mechanical properties, 21, 76ff
 - Ohm's resistance, 109, 173
 - properties as capacitor, 74, 189ff
 - skeleton, 74ff
 - skeleton fence model, 76
 - structure, 74ff
 - surface charge density, 86, 91, 174
 - surface charges, 87ff
 - surface tension, 44, 68ff, 72
 - transport properties, 165ff
 - viscosity, 77
- Cell wall
 - ion-exchange properties, 87
 - mechanical properties, 138, 144
- Center of gravity of macromolecules, 28
- Chaos theory, 343
- Chaotropic ions, 43
- Chaperons, 61
- Charge
 - density, 48
 - electric, 34f
 - transfer complex, 309
- Chemical activity, definition, 104
- Chemical potential, definition, 104
- Chloride
 - activity, 140f
 - distribution in cells, 140
 - transport, 113f, 140
- Chlorophyll, 310ff
- Chloroplasts, 167, 208, 310f
- Chromosome aberrations, 315, 324
- Clearance constant, 355f
- Coagulation of particles, 54f
- Cochlea, viscoelastic properties, 265ff
- Coefficient
 - of activity, 104f
 - of friction, 109f
 - of mobility, 151
 - of mortality, 357f
 - of proliferation, 358f
 - of work, 100
- Cofield rotation, 205f
- Coherent resonance, 121
- Collagen
 - elastic properties, 218, 270
 - piezoelectric properties, 192
 - structure, 46
 - UV-damage, 314
- Colloid osmotic pressure, 133
- Colomb's law, 34, 51
- Command variable, 346
- Compartment
 - analysis, 161, 356
 - definition, 335
- Compass orientation, 283f
- Compton effect, 318
- Computational neuroscience, 3, 371
- Conductivity
 - electrical, 192f
 - of air, 287

- of biological tissue, 295
 - of cell membrane, 74, 177, 198
- hydraulic, 155
- thermal, 249
 - of air, 250
 - of biological tissue, 250
 - of water, 250
- vectorial, 112
- Conjugated flux-force relation, 110f
- Conservative System, 107f, 113
- Constant field theory (Goldman), 158f, 163f
- Constants, definition, 334
- Continuum Physics, 95, 214
- Control deviation, 346f
- Controlled variable, 348
- Controller, types of, 346f
- Coordinate of work, 100
- Cosmotropic ions, 43
- Cotransport, 165ff
- Couette viscosimeter, 212
- Coulomb's law, 34, 51
- Coupling coefficient, 111
- Covalent bonds
 - energy, 25
 - rotation, vibration frequency, 27
- Cryoscopy, 132
- Crystal radius of ions, 42
- Cusp catastrophe, 342f
- Curie-Prigogine principle, 112
- Curvature of bending, 221
- Cybernetics, 346ff
- Cytochrome, 312
- Cytoskeleton, 74, 76, 79f, 92, 253

- D**
- Dark reaction of photosynthesis, 311
- DC-field, 286ff
- Debye
 - dispersion, 197
 - unit of dipole momentum, 37f
- Debye-Hückel
 - parameter, 49f, 82
 - theory of electrolytes, 47ff
- Decibel (dB), 258
- Deep sea, pressure influence, 44f, 253
- Degenerated state of electrons, 308
- Degree of
 - advancement, 126
 - freedom, 61, 102
- Depot lubrication of joints, 214
- Desoxyribonucleic acid (DNA)
 - content of information, 10f
 - dipole momentum, 38
 - double strand breaks, 324
 - prebiotic evolution, 369f
 - radiation sensibility, 322ff
 - single strand breaks, 323
 - stability, 26
 - structure, 57
 - UV-sensibility, 315
- Deterministic behavior, 5
- Dexter electron transfer 309
- Diamagnetic properties of biological materials, 279, 282f
- Diathermy
 - caused by electromagnetic fields, 2, 300ff
 - caused by ultrasound, 271, 273
- Dielectric constant
 - of cell membrane, 89, 174
 - definition, 34
 - local, 47, 56
 - of various tissues, 192ff
 - of water, 34
- Dielectrophoresis, 204ff
- Differential
 - total, 97f
 - of work, 99f
- Differential controller (D-controller), 347
- Differential equations
 - degree of, 338
 - general solution, 338
 - order of, 338
 - ordinary, 335f
 - partial, 335f
- Differential quotient, partial, 97
- Diffuse double layer, 81
- Diffusion
 - coefficient, 32f, 151f
 - of ions, 157ff, 165ff
 - layer, 152
 - of nonelectrolytes, 150ff
 - potential, 140ff, 162ff
- Dilatant fluid, 210f
- Dipole
 - fluctuating, 38
 - induced, 37f, 52f
 - interactions, 51ff
 - momentum, 37f
- Dispersion forces, 52ff
- Dispersion regions of electromagnetic fields, 196f
- Displacement current, 194
- Dissipation function (Rayleigh), 116f, 153ff, 170f, 216, 230, 241, 245, 252, 297

- Dissipative structures, 16, 96, 116, 118f, 226, 363, 368
- Disulfide bridges, 59
- Divergence (div)-operator, 107
- DLVO-theory, 53ff
- DNA. *See* Desoxyribonucleic acid (DNA)
- Dolphin
 - Reynolds number, 228, 238
 - sonar system, 276f
 - swimming behavior, 236
- Donnan
 - equilibrium, 87, 90, 143ff, 147ff
 - osmotic cytolysis, 149, 169, 208
 - osmotic pressure, 91, 144, 146ff, 168
 - potential, 144f, 148
 - ratio, 145
- Doppler effect
 - in ultrasound diagnosis, 233, 270
 - in ultrasound location, 276
- Dorn effect, 87
- Dose equivalent (Sv), 319
- Dosimetry
 - of ionizing radiation, 319
 - of low frequency EMF, 291ff
 - of radiofrequency EMF, 298ff
 - of UV-radiation, 316
- Double layer, electrical, 80ff
- Double strand breaks of DNA, 323f

- E**
- Ear, human
 - anatomical organization, 261f
 - as frequency filter, 262
- Echolocation
 - in bats, 274ff
 - in birds, 276
 - in dolphins, 277f
 - mechanisms, 274ff
- Ecological models, 357ff
- Eddy currents in the body, 278, 285, 288, 293, 297f
- Effective parameters, 5f, 29, 43, 50, 76f
- Einstein equation
 - for diffusion coefficient, 32f, 214
 - for viscosity of suspensions, 210
- Einstein-Smoluchowsky equation, 32
- Einthoven triangle, 191
- Elasticity module, 77f, 215ff, 232
- Elastin, 218
- Electric
 - charge, 34f
 - potential, definition, 35
- Electrical double layers, 80ff
- Electric break down of cell membranes, 202f
- Electric field
 - calculation, 35f
 - field strength, 35
 - interaction with biological systems, 286ff, 295
 - in membrane, 36f
 - of natural environment, 286f
- Electric organs of fishes, 191f
- Electric potential, definition, 35
- Electro-aerosol therapy (EAT), 287
- Electrocardiogram (EKG), 187, 189f
- Electrochemical
 - energy accumulation in cells, 165f
 - potential, 106
- Electrodes
 - Ag/AgCl/KCl, 141f
 - ion selective, 141
 - suspension effect, 143
 - tip potential, 143
- Electrodiffusion, 165ff
- Electroencephalogram (EEG), 187f
- Electrofusion of cells, 203f
- Electrogenic transport, 165f
- Electroinjection (=permeabilization), 202f
- Electrokinetic
 - phenomena, 80ff
 - potential, 85
- Electrolyte flux, 157ff
- Electromagnetic field
 - application in therapy, 291, 298, 300
 - applications in biotechnology, 202ff, 301
 - dosimetry, 291ff
 - frequency classification, 291f
 - limits (*see* Limits of exposition)
 - microdosimetry, 302, 329
 - of natural origin, 294
 - penetration in the body, 292f
 - screening, 292f
 - spectrum, 292
 - window effects, 302
- Electron spin resonance (ESR), 77, 214
- Electroosmosis, 85
- Electropermeation, 202f
- Electrophoresis, 83ff
- Electrophoretic potential, 85
- Electroplax, 177, 191f
- Electroporation, 202f
- Electroreception, 288f
- Electrorotation, 204f

- Electrostatic field
 - of cell membrane, 92f
 - interaction with body hairs, 288
 - of natural environment, 286f
 - receptors, 289
 - Electrostatic unit (eu), 34
 - Embryo
 - morphogenic role of electric fields, 188f
 - sensibility to radiation, 235f
 - Endothermic reaction, 103
 - Energy
 - of activation, 19ff
 - definition, 97
 - distribution according to Boltzmann, 16ff, 47f
 - of formation, 103
 - Gibbs, free, 102ff
 - gradient, 106f
 - Helmholtz, free, 102
 - of hydrogen bonds, 26
 - internal, 99
 - of ionization, 304f
 - kinetic, 17
 - of photons, 304
 - of thermic noise, 19
 - transfer, 307ff
 - Enthalpy
 - of activation, 23f
 - definition, 102
 - Entrance effect, 230f
 - Entropy
 - account of a system, 113
 - of activation, 24
 - connection with information, 9ff
 - driven process, 44f, 60f
 - elasticity (= rubber elasticity), 30, 217
 - flux, 113
 - phenomenological definition, 100
 - production, 112ff
 - in statistical thermodynamics, 7ff
 - Environmental biophysics, 1, 245ff
 - Enzymatic reaction, 22, 350ff
 - Enzyme
 - activation energy, 21f, 24, 124
 - membrane bound, 6, 83, 93
 - reaction kinetics, 350ff
 - sensitivity to radiation, 321, 323
 - structure forming, 61
 - substrate complex, 115
 - Equifinality, 337f
 - Equilibrium
 - constant, 23, 64
 - global, 116
 - locale, 116
 - structures, 16
 - thermodynamic, 16f, 115ff
 - thermodynamic, definition, 113
 - Equilibrium constant, 22f, 64, 125, 334, 350
 - Equipotential lines, 35
 - Equivalent conductivities of ions, 42
 - Erythrocytes
 - aggregation, 213
 - deformability, 78, 213
 - ionic conditions, 113f
 - life span, 119
 - magnetic susceptibility, 279ff
 - membrane potential, 148
 - osmotic properties, 136, 157
 - pressure resistance, 138
 - reflection coefficient, 137, 157
 - surface charge, 85f, 90
 - swelling, 149
 - ESR. *See* Electron spin resonance (ESR)
 - Evaporation, temperature regulation, 251
 - Evolution
 - Crow-Kimura model, 369f
 - Eigen model, 368f
 - prebiotic, 16, 368
 - Exchange flux, 155
 - Exothermic reaction, 103
 - Extensive parameters, 97
 - Extracellular electric fields, 187ff
- F**
- Fahraeus-Lindqvist effect, 230
 - Faraday constant, 34, 49
 - Feedback systems, 346ff
 - Ferromagnetic implants, 285
 - Ferromagnetic properties, 279f
 - Fick's laws of diffusion
 - first, 33, 151f
 - second, 151f, 250, 363
 - Field lines. *See* Equipotential lines
 - Filtration coefficient, 155
 - Finite-difference time-domain method (FDTD), 299
 - Fish
 - electric organs, 191f
 - electric sensibility, 288f
 - Reynolds numbers, 228
 - swimming behavior, 234ff
 - Flickering clusters of water, 41
 - Flight mechanisms, 234ff
 - Flippase, 74
 - Fluctuating dipoles, 38
 - Fluctuations. *See* Thermic noise

Fluidity, 209
 Fluid-mosaic structure of cell membrane, 74f
 Fluorescence
 life span, 308f
 markers, 77, 92, 313
 resonance energy transfer, 309
 role in energy transfer, 308
 Flux
 conservative, 107
 definition, 107
 electrogen, 166ff
 formal approaches, 107ff
 matrix, 110ff
 of non charged molecules, 150ff
 rheogen, 166
 scalar, 111f
 unidirectional, 160ff
 Flying mechanisms, 237ff
 Force, generalized, 106ff
 Formal neuron models, 371ff
 Förster resonance energy transfer (FRET), 309
 Fourier's law of thermal conductivity, 249
 Free radicals. *See* Radicals
 Frequency classification of electromagnetic fields, 291f
 Friction, 109
 "Frozen" dissipative structures, 118

G

Galvanotaxis, 189, 290
 Galvanotropism, 290f
 Gap junctions, electrical conductivity, 202
 Gas constant, molar (R), 18
 Gating current, 183
 Gauss
 plane of imaginary numbers, 194f
 statistics, 28
 Genetic effects or radiation, 324ff
 Geoelectric field, 286f
 Geomagnetic field, 278, 280ff, 288
 Gerdien tubes, 287
 Giant axon, 176
 Gibbs-Duhem equation, 102f
 Gibbs free energy, 23, 61, 97f, 102f, 124
 Gibbs fundamental equation, 99ff, 106, 124ff, 252, 282
 Glansdorff-Prigogine principle, 116ff
 Glucose
 reflection coefficient, 137
 transport, 166f, 169ff
 Glycine, buffer capacity, 64f
 Glycocalyx (=surface coat), 75, 91f

Glycoproteins, 75, 90f, 147, 214, 289f
 Goldman equation of ionic flux, 158f, 163
 Goldman-Hodgkin-Katz equation,
 164, 172, 181
 Gouy-Chapman double layer, 80ff, 85
 Gradient (grad), 106
 Graph theory, 343ff
 Gray (Gy), 319
 Gray's paradox, 236
 Green fluorescent protein (GFP), 313
 Grotthus-Draper principle, 245, 300
 Growth, models of, 362ff
 G-value (eV-gain), 322

H

Hagen-Poiseuille law, 229
 Hair cells in inner ear, 267ff
 Half life time, 336
 Hamaker constant, 52
 Hearing
 angular resolution, 261f
 frequency dependence, 259f
 mechanistic theories, 264ff
 Heat
 capacity of proteins, 62f
 conductivity, 250
 differential, 99
 flux, 249
 production of a chemical reaction, 103
 Helmholtz
 double layer, 81
 free energy, 102
 potential, 81ff, 85
 theory of hearing, 264
 Hematocrite, 210
 Hemoglobin
 buffer capacity, 65f
 concentration in erythrocytes, 148
 isoelectric point, 66
 osmotic coefficient, 135
 Hemorheology, 228ff
 Henderson-Hasselbalch equation, 64f
 Hierarchy
 of structural levels, 15
 of time constants, 119f, 127, 147, 354
 High-frequency EMF, 298ff
 High-pressure biotechnology, 253
 "Hit" process of radiation, 329ff
 Hodgkin-Huxley
 theory of nerve excitation, 121ff, 179ff
 Hofmeister series, 55f
 Hooke's law, 215, 222, 232

Hot spots in diathermy, 271, 302f
 Hovering flight, 237f
 Hydration, first and second order, 45f
 Hydraulic conductivity, 155
 Hydrodynamic radius. *See* Stokes' radius
 Hydrogen bonds, 26, 33, 39ff, 45f, 50, 58f, 75
 Hydrophobic
 interactions, 44f, 51, 55, 59f, 68, 70ff,
 74f, 253
 regions, 71f, 74, 89, 93, 183f
 Hypercycle (Eigen), 369
 Hypertonic solution, 135
 Hypotonic solution, 135f

I

Ice structure of water, 40f
 Immunomagnetic cell separation, 280
 Impact pressure, 235
 Impedance
 acoustic, 258
 electric, 193ff
 mechanic, 217
 Impedance spectroscopy, 197
 Impedance transformer (mechanical), 262f
 Impulse compression method in sonar
 systems, 276
 Indifferent state, 114f
 Induction, electric, 279, 284f, 293, 296f
 Information
 of biomacromolecules, 10f, 56f, 367
 definition (Shannon), 9f
 entropy interconnection, 12
 input by environment, 245, 254, 276,
 304, 310
 semantic, 11, 14
 syntactic, 11, 14
 transfer in thermoreception, 249
 transfer, molecular, 321ff, 345, 362
 transfer, neuronal, 178, 373
 units, 10
 Infrared
 absorption by black body, 250f
 absorption in the atmosphere, 307
 definition of the regions A,B,C, 304
 reception in snakes, 247, 306
 resonance with molecular movement, 26
 spectroscopy, 306
 therapeutic use, 397
 Infrasound, 254, 259, 268f
 Insects
 echolocation, 274
 flight mechanism, 237f
 luminescence, 313

 magnetic orientation, 284
 resistance to radiation, 327
 Reynolds number, 228
 sensitivity to vibrations, 257
 thermosensitive, 247
 wing loading, 237, 242
 Integral controller (I-controller), 348f
 Integrate-and-fire model of neurons, 372f
 Intensive variables, 97
 Intercellular clefts
 Donnan-osmotic properties, 91, 187
 electric field, 91
 electroosmotic flow, 88
 ion transport, 187f
 Interfacial tension, 67f, 69
 Intermolecular interactions, 50ff
 Internal energy (U), 99f
 Intersystem crossing, 316
 Intrinsic
 pK, 66
 viscosity, 210
 Ione
 channels, 182ff
 cloud, 48f
 crystal radius, 42
 equivalent conductivity, 42
 hydration, 39ff, 42
 locale milieu, 169
 pumps, 165ff, 184
 strength, 49f
 transport, 165ff
 Ionization
 density, 331
 energy, 304, 317ff
 Ionizing radiation
 annual average dose, 328
 definition, 317
 dosimetry, 319
 kinds of, 318f
 lethal doses (LD₅₀³⁰), 326f
 from medical diagnostics, 328
 natural sources, 327f
 primary processes, 320ff
 protection, 327ff
 repair processes, 323
 stochastic effects, 324f, 329ff
 Iontophoresis, 88
 Isoelectric focusing, 66f
 Isoelectric point, 38, 65, 90, 148
 Isomerases, 61
 Isomorphism, 334
 Isotonic solution, 135f, 148f, 157
 Isotropy, 3, 5f, 67, 112

J

Jablonski diagram, 308f

K

KcsA channel, 182f
 Kelvin-Voigt-element, 216
 Kidney cell, ion transport, 167
 Kinetosis, 257
 Kirchhoff's law, 197

L

Laminar flow, 118, 208ff, 225ff
 Laminar profile, 225ff
 Langmuir trough, 71f
 Laplace equation of pressure, 68, 232
 Lateral movement of membrane compounds, 77ff
 LD₅₀ unit, 326f
 Le Chatelier principle, 44, 72, 252f
 Lethal dose of ionizing radiation. *See* LD₅₀³⁰
 Levinthal paradox, 59f
 Life, origin, theories, 366ff
 Light
 absorption in the atmosphere, 307
 absorption by chromophores, 309, 312
 dependent catalysis, 313
 emitting molecules, 313
 energy content of solar radiation, 310
 influence on magnetoreception, 384
 processes of excitation, 307ff
 spectral distribution, 304f
 Limit circle of system behavior, 341
 Limits for exposition by
 high-frequency EM-fields, 300ff
 infrasound, 269f
 ionizing radiation, 327f
 low-frequency EM-fields, 296
 noise, 260f
 static magnetic fields, 285
 ultrasound, 274
 Linear energy transfer (LET), 319
 Linear force-velocity approach, 96
 Lipid rafts, 76
 Lipids
 behavior at boundaries, 70ff
 bilayers (BLM), 71f
 dielectric constant, 34
 lateral diffusion in membrane, 74f
 phase transitions, 80, 175
 protein interaction, 75
 shape factor, 79f

structure, 70f

transversal distribution, flippase, 74

Liposomes, 72

Liquid junction potential. *See* Diffusion potential

Lithotripsy treatment, 273

Liver

 magnetic susceptibility, 279

 passive electric properties, 196

Logistic function, 359

London dispersion forces, 52ff

Long-range interaction forces, 53ff

Lorenzini ampullae, 289f

Loschmidt number (= Avogadro number), 18

Low frequency electromagnetic fields, 291ff

Luciferase, 313

Lung

 air flow, 231

 diagnostics by ferromagnetic particles, 285

 diagnostics with impedance spectroscopy, 197

 magnetic susceptibility, 279

 radon accumulation, 328

 resonance with infrasound, 269

 surfactants, 68f

Lyso-phospholipids, 80

M

Macromolecule

 center of gravity, 28

 entropy elasticity, 30, 218

 interactions, 50ff

 life span, 25f

 movement, 26ff

 random flight model, 27f

 size, 28f

 Stokes radius, 29

 structure levels, 15

Magnetic

 anisotropy, 383

 field strength, 287f

 flux density, 287f

 induction, 278

 permeability, 278f

 susceptibility, 279

Magnetic field

 diagnostic methods, 285

 on earth surface, 281f

 influence on biological objects, 281ff

 orientation of organisms, 281ff

Magnetites, 280

- Magnetocardiogram, 285
- Magnetoencephalogram, 285
- Magnetokinetics, 282
- Magnetophoresis, 383
- Magnetophosphenes, 285
- Magnetoseparation, 280
- Magnetosomes, 280
- Magnetostatic bacteria, 281f
- Magnetotaxis, 281
- Malthus law of exponential growth, 358
- Map coloring problem, 370
- Maximal possible dose. *See* Limits of exposition
- Maxwell
 - demon, 12f, 32, 113, 123
 - element of viscoelasticity, 216f
 - equations of electromagnetic fields, 186, 291
 - equation for velocity distribution, 18f
- Maxwell-Wagner dispersion, 196f
- Maxwell-Weichert model of viscoelasticity, 216
- McCulloch-Pitts-neuron model, 371f
- Medical physics, 1, 208, 245
- Meissner corpuscles, 257
- Membrane. *See also* Cell membrane
 - bending, 78
 - biological structure, 74ff
 - as boundary, 67
 - capacity, 74, 89, 174
 - caveolae, 76
 - elasticity, 77f
 - electric field strength, 36f, 175
 - field intensity, 37, 92f, 175, 183
 - fluidity, 74, 77
 - formation, 70ff
 - functions, 67
 - mechanical properties, 76ff
 - models, 74ff
 - phase transitions, 93, 175
 - potential, 89f, 140f, 165ff, 199f, 295
 - resistance, electrical, 89
 - semipermeable, 127f
 - skeleton, 74ff
 - transport, molecular aspects, 181ff
 - viscosity, 77
- Metabolic
 - rate, 116, 241, 297, 301, 309
 - network, 337, 343ff, 354
- Metastability of systems, 115, 117f, 341ff, 354, 374
- Michaelis-Menten equation, 334, 351f, 356
- Microelectrodes, 142f
- Microphonic potential of ear, 264
- Microphysical behavior, 5f, 77f
- Microtubules, pressure dependence, 45, 253
- Middle ear, mechanics, 263f
- Mitochondria
 - ATP-synthesis, 167
 - electromanipulation, 203
 - free radicals, 323
 - osmotic behavior, 136
- Mitosis
 - change of membrane potential, 175
 - influence of radiation, 325
 - time constants, 119
- Mobil telephone, electromagnetic fields, 299
- Models
 - of biochemical reaction networks, 350ff
 - of compartment systems, 344ff
 - of drug distribution in the body, 355f
 - of ecological systems, 357ff
 - of evolution, 393f
 - of growth, 362f
 - of morphogenesis, 363f
 - of neuronal excitation, 176ff
 - of neuronal networks, 370ff
 - of radiobiological effects, 329ff
- Moens-Korteweg equation, 231
- Molar gas constant (R), 18
- Molecular
 - dynamics concept (MD), 184f
 - movement, thermal, 26ff
 - radius, effective, 28f
 - volume, partial, 98f
- Mole fraction, 104f
- Molten-globe structure of proteins, 60f
- Monolayer of lipids, 71f
- Morphogenesis
 - models, 362f
 - role of internal electric fields, 188f
- Motor, molecular, 123f
- Muscle
 - action potential, 177
 - excitability by AC current, 296f
 - extracellular potentials, 191
 - heat conduction, 250
 - magnetic susceptibility, 280
 - maximal power, 218
 - membrane potential, 175
 - passive electrical properties, 196
 - penetration of ultrasound, 270
 - thermal conductivity, 250
- Mutation
 - induced by γ -radiation, 324f, 332
 - induced by UV, 315f
 - role in evolution, 26
 - target theoretical approach, 369f
- Myoglobin, 63

N

Nabla-Operator, 35, 48, 107
 NADPH. *See* Nicotinamid-adenin-dinucleotid phosphate (NADPH)
 Na-K-ATPase, 166, 172, 184
 Negentropy, 12
 Nernst equation, 138ff, 173
 Nernst-Planck equation, 158ff, 185f
 Nerve
 action potential, 176ff
 conductivity electrical, 181
 external electric field, 190
 Hodgkin-Huxley theory, 179ff
 impulse propagation, 178ff
 refractory period, 178
 regeneration, 290f
 sensitivity to radiation, 326
 stimulation by LF-EMF, 296f
 stochastic resonance effects, 121f
 unmyelinated, 178
 velocity of pulse propagation, 178
 Network theory, 343ff
 Neuraminic acid (=sialic acid), 75, 90f
 Neuronal networks, 370ff
 Neuston, 68
 Neutron radiation, 318
 Newtonian fluid, 210f
 Nicotinamid-adenin-dinucleotid phosphate (NADPH), 310f
 Nodes of Ranvier, 178
 Noise acoustic, 259ff, 264, 269f
 Noise, thermal
 effect, 22, 26ff, 33, 41, 47f, 62, 120ff, 248f, 304, 308
 enhanced processes, 120ff
 Non-Arrhenius behavior, 248
 Nonlinear
 differential equations, 148, 180f, 337ff, 352, 359ff
 processes in general, 108ff, 120ff, 268, 343, 363f, 373
 streaming behavior, 110, 226
 thermodynamic approaches, 96, 108f, 116ff
 Non-Newtonian viscosity, 210f, 229f
 Nonthermal effect, definition, 300f
 Nuclear Magnetic Resonance (NMR), 39, 42, 46, 66

O

Ohm's law, 109
 Ontogenesis, entropy production, 117

Onsager coefficients, 111
 Onsagers law on reciprocal relations, 111
 Order
 of biological systems, 6ff
 of chemical reactions, 335, 350, 352
 of differential equations, 48, 152, 338f, 337f, 348, 359
 and entropy, 7ff
 of systems behavior, 339f
 Organ of Corti, 261ff
 Origin of life, 366ff
 Oscillations
 of controllers, 348ff
 damped, 339f, 348
 of ecological systems, 361ff
 as a kind of dissipative structure, 14, 118
 mechanical, 254ff, 339, 339ff
 molecular, 26f, 52
 Osmolality (osmol/kg), 132
 Osmolarity (osmol/l), 132
 Osmoregulation, 138
 Osmotic pressure, 127ff
 coefficient, 132f
 definition, 127
 measurement, 132f
 Ostwald viscosimeter, 212
 Otoacoustic emission, 268
 Oxygen effect of radiation damage, 322

P

Paddle model of voltage sensitivity, 184
 Paramagnetic properties, 279f
 Parameter, definition, 334
 Parameter plot of a function, 361
 Parameters
 effective, 5f, 29, 43, 50, 76f
 extensive, 7, 97
 intensive, 7, 97
 phenomenological, 95
 Partial differentiation, 97f
 Partial mole volume, 98f, 130f
 Pascalization, 253
 Passive electrical properties, 192ff
 Pauli's exclusion principle, 308
 PEMF. *See* Pulsed electromagnetic fields (PEMF)
 Peptide bond, 57
 Permeability coefficient, 34
 Permittivity, 195
 Perpetuum mobile, 12, 123
 Pfaffian differential equation, 98
 Pfeffer's osmometer, 127, 132, 156

- pH
 electrodes, 141f
 local at charged surfaces, 84
 Phagocytosis, 69
 Pharmacodynamics, 355
 Pharmacokinetics, 355ff
 Phase
 boundaries, 70ff
 transitions, 80, 93, 175
 Phenomenological
 flux equations, 110f
 parameters, 95
 Phon, 259
 Phosphatidyl
 choline, 70f
 ethanol amine, 71, 80
 serine, 70, 90f
 Phospholipids. *See* Lipids
 Phosphorescence, 309
 Photo
 effect of ionizing radiation, 318
 electrons, 318
 Photomorphogenesis, 311
 Photosynthesis
 efficiency, 310
 mechanism, 311f
 spectral dependence, 311
 Photosystems I and II, 311
 Phototaxis, 312
 Phototropism, 312
 Piezoelectric properties of biological
 materials, 192
 Piezophile organisms, 253
 Piezophysiology, 253
 pK
 intrinsic, 66
 value, 64ff
 Planck's constant, 8, 304
 Plane power density, 299f
 Poise (P), 209
 Poisson-Boltzmann equation, 48f, 50,
 81f, 91, 185
 Poisson equation, 48
 Poisson-Nernst-Planck-theory (PNP-theory),
 160, 184f, 186
 Polar moment of inertia, 223f
 Polarization of
 bonds, 37
 molecules, 39f, 56
 Polyethylene glycole, 134f
 Population dynamics, 340, 363
 Potential
 chemical, 104
 electric, definition, 35
 electrochemical, 106
 Power lines, electromagnetic field, 292, 294
 Prebiotic evolution, 16, 368ff
 Pressure
 application in biotechnology, 153
 influence on cells, 252
 physiological influences, 252
 Pressure, influence
 on biological systems, 252ff
 on supramolecular structures, 44f, 72
 Prigogine principle of minimal entropy
 production, 116f
 Principles of Thermodynamics
 first, 17, 99
 second, 7, 12, 15, 17f, 44, 99, 112, 116
 third, 103
 Probability
 mathematical, 8ff
 thermodynamic, 7ff, 17f
 Process
 entropy driven, 44f, 60f
 irreversible, 113
 non-linear, 108ff
 Profile drag, 234f
 Proportional controller (P-controller), 347
 Proportional deviation, 347
 Proteins
 β -pleated sheet, 58
 behavior near boundaries, 70ff
 charges, pH-dependence, 65f
 dipole momentum, 38
 electric response signals, 38
 fluctuations, 62f
 folding, 57
 folding rules, 59ff
 heat capacity, 62f
 hydration kinetics, 46f
 information contents, 10f
 isoelectric point, 65f
 in membranes, 74ff
 prebiotic evolution, 368f
 self assembly, 56ff
 stability, 61f
 stochastic coil, 62
 structure, 56ff
 structure information, 13
 Proteoglycans, 214
 Proton
 acceptor, donor, 46, 64
 gradient, 310ff
 transport, 166f, 170
 Pseudoplastic fluid, 210f

Pulsed electromagnetic fields (PEMF),
297f

Purkinje-fiber, action potential, 177

Q

Q_{10} -parameter, 20f, 247f

Quantum mechanical approaches, 50, 52, 95

Quasi

elastic neutron scattering, 47

equilibrium, 119f, 127, 147

reversible process, 100

species theory (Eigen), 368f

stationary state, 119

R

rad (rd), 319

Radiation

chemistry, 320

corpuscular, 317ff

disease, 324ff

effects, direct and indirect, 321f

lethal doses (LD_{50}^{30}), 326f

mathematical model of interaction, 329ff

oxygen effect, 322

protection, 327ff

repair mechanisms, 316, 323

sensibility of various organisms, 326f

stochastic effects, 324, 329ff

threshold dose, 329

Radical pair reaction, 282ff

Radicals, free. *See also* ROS

induced by radiation, 320ff

induced by ultrasound, 272

in metabolism, 323

Radiochemistry, 320

Radioecology, 327

Radiofrequency EMF, 298ff

Radiolysis of water, 320ff

Radius

of bending, 221

of gyration of a macromolecule, 28f

Ramachandran bridges, 46

Ramachandran diagram, 57

Raman spectroscopy, 306

Random flight chain (= random coil), 27f

Ranvier nodes, 178

Rate constant, 20, 23

Rayleigh's dissipation function, 116f,
153ff, 170f, 216, 230, 241, 245,
252, 297

Raynaud's phenomenon (vibration), 257

RC-circuit, 193, 195ff

Reaction

bimolecular, 363

coordinates, 21f

degree of advancement, 126, 336

enthalpy, free, 23

isotherm (Van't Hoff), 125

kinetics, biochemical, 350ff

order of, 335, 350, 352

rate constant, 350

theory of absolute reaction rate, 19ff

Reaction-diffusion model of morphogenesis,
363ff

Reaction coordinate, 22

Reactive oxygen species (ROS),

315f, 322f, 325

Reciprocity relation by Onsager, 111

Reference

electrodes, 141ff

intensities of sound, 294f

levels of electromagnetic field interactions,
294ff

Reflection coefficient (Staverman), 137, 156f

Relative biological effectiveness of ionizing
radiation (RBE), 319

Resilience, 216

Resistance

electrical, 192ff, 201

phenomenological coefficient, 109

Resonance frequencies

of bound water, 302

of cavitation, 272f

electrical for whole body, 299, 303

of human body for vibrations, 256f

for infrasound, 269

of outer and middle ear, 262ff

Resonance transfer of energy, 309

Reversibility, in physics, chemistry and
biology, 112f

Reynolds number

of animal movement, 228, 235f

of blood circulation, 233

critical, 227f

definition, 227

Rheogenic ion transport, 165f

Rheopectic viscosity behavior, 211

Ribonucleic acid. *See* Desoxyribonucleic acid

Riboswitch, 247f

Roentgen (R), 319

ROS. *See* Reactive oxygen species (ROS)

Rotation of molecules, 26f

Rubber elasticity (=entropy elasticity),
78, 217f

S

Safety standards. *See* Limits for exposure

Saltatory nerve conduction, 178

Salt-out effect, 55

SAR. *See* Specific absorption rate (SAR)

Scavenger of radicals, 323

Schrödinger, 3, 12

Schrödinger equation, 185f

Schulze-Hardy rule, 56

Schumann resonance, 294

Schwann cells, 178

Sedimentation potential, 85f

Self assembly, 16, 70ff

Self diffusion, 42, 46, 114

Self organization, 16, 70ff, 362, 371

Semipermeable membrane, 127f

Shannon

entropy, 12

equation, 10

Shear

rate (=velocity gradient), 208f

stress, 211, 213f, 230

Short-range interactions of particles, 53ff

Sialic acid (=neuraminic acid), 75, 90f

Sievert (Sv), 319

Signal-to-noise ratio (SNR), 120ff

Silver-silver chloride electrode, 141f

Similarity analysis, 227, 231, 240ff

Singlet state of electrons, 308f

Skin

depth of EMF penetration, 298f

electrical potential differences, 189

electrical resistance, 289

iontophoresis of drugs, 88

thermal conductivity, 250

UV-sensibility, 314

vitamin D production, 314

Skin effect of HF EMF, 298f

Skin friction, 234

Small world graph, 345

Smoluchowski equation, 32, 84

Sonar system, 271, 274ff

Sonophoresis, 273

Sonoporation, 273

Sound

audible, 258ff

auditory threshold, 259f

intensity, 258

limit of maximal exposure, 260

pain threshold, 259f

pressure, 258

Specific absorption rate (SAR),

251f, 297, 300ff

Spectrin, 76

Spemann organizer of differentiation, 362

Spherics (= Atmospheric), 294

Spillover process in photosynthesis, 311

Spin-glass models of brain, 374

SQUID. *See* Superconductive quantum interference device (SQUID)

Stability criteria

of macromolecules, 25f

of a system, 114, 341f

Stabinger viscosimeter, 212

State

degenerated, 308

frustrated, 374

indifferent, 114f

metastable, 114f

nonequilibrium, 113, 116

quasistationary, 119

stationary, 113ff

variables, 97

Statistical thermodynamics, 7, 17, 19, 21, 25, 27, 33, 95, 186, 303

Stationary

motion, 108

state, 113f

Staverman's reflection coefficient, 137, 156f

Steady state

definition, 113f

equifinality, 337f

stability conditions, 114f

Steel elasticity, 78f, 218

Stefan-Boltzmann law of heat radiation, 251

Stern's double layer, 81ff

Stochastic

composition of amino acids in proteins, 367ff

distribution, 6ff, 11

effects of radiation, 324f, 329f

fluctuations, 62

models for ecological systems, 357

movement, 30ff

nerve excitations, 372f

properties of biological systems, 5f

resonance, 120ff, 290

transition to turbulent flow, 227

Stokes'

equation, 29

radius, 28f, 42f

Stop-flow technique, 157

Streaming potential, 85f, 187

Stress-strain diagram, 215

Structure

aperiodic, 15

biological, 6, 14ff

dissipative, 16, 96, 118, 226, 363, 368

electric, of organism, 185ff

information, 13, 15

- levels of macromolecules, 15
 - mathematical definition, 14
 - periodic, 15
 - in time, 14ff
 - of water, 39ff
- Structure braking ions, 42f
- Sunspot activity, 283
- Superconductive quantum interference device (SQUID), 120, 285
- Supraparamagnetic properties, 280
- Surface
 - charge, 81ff, 85f
 - charge of cell membrane, 88ff
 - coat of cell membrane, 75
 - energy, specific, 68f
 - friction, 227, 234
 - potential, 81ff
 - tension, 67ff
- Surfactants, pulmonary, 69f
- Susceptibility, magnetic, 279ff
- Suspension effect of electrodes, 143
- Swimming mechanisms, 234ff
- Symbioses, models, 361
- Symport, 165f
- Synapses, types, 373
- Synergetics, 333
- Synovial fluid, 214
- System
 - analysis, 338ff, 343ff
 - anisotropic, 5f, 112
 - biology, 371
 - closed, 97
 - chaotic, 343
 - conservative, 108
 - definition, 14f, 96ff
 - of electric units, 34
 - isolated, 97
 - isotropic, 112
 - linear, 108f
 - metastable, 115f, 118, 341ff, 354
 - nonlinear, 3, 96, 108ff, 116ff, 120f, 181, 225, 268, 338f, 343, 352, 356, 359ff, 363f, 373
 - open, 96f, 113
 - stability, 114f
 - stationary, 113f
- Système International d' Unités (SI), 34
- T**
- Target
 - area of radiation effect, 331
 - theory of radiation, 329ff
- Temperature
 - anomalies of water, 39
 - dependence of biological processes, 20f
 - increase in HF-fields, 300ff
 - regulation, 247f
 - sensation, 247ff
- Tendons
 - elasticity, 217f
 - energy storage, 218
- TENS. *See* Transcutaneous electric nerve stimulation (TENS)
- Tension, isotonic, isomeric, 217
- Terra-Hertz radiation, 250, 303f, 305ff
- Theory of
 - absolute reaction rates, 21f
 - biological amplification, 246
 - graphs, 343ff
 - similarity, 227, 231, 240ff
 - transition states, 22
- Thermal
 - conductivity, 249f
 - convection, 250
 - effect of electromagnetic fields, 300ff
 - energy, 22
 - movement, 26ff
 - noise, 22, 26ff, 33, 41, 47f, 62, 304
 - radiation, 250f
- Thermodiffusion, 249
- Thermoreceptors, 247ff
- Thermotaxis, 247
- Thixotropy, 211, 214
- Threshold values. *See* Limits
- Thylakoid, 310
- Time hierarchy in biological systems, 119, 354
- Tip potential of microelectrodes, 143
- Titration of ampholytes, 64f
- Torsion, 224
- Trajectory of system behaviour, 340ff
- Transcutaneous nerve stimulation (TENS), 297
- Translation of molecules, 26ff
- Transport
 - active, 165ff
 - electrogenic, 165ff
 - frequency, of ions in membrane, 183
 - molecular aspects, 181ff
 - proteins, 181ff
 - rheogenic, 165f
 - voltage dependent, 179, 184, 247
- Traveling wave dielectrophoresis, 207
- Tribology (biotribology), 214
- Triplet state of electrons, 308
- TRPV-proteins, 21, 247f
- Tubercous organ, 289
- Tumor
 - growth-dissipation function, 177
 - radiotherapy, 320

Tumor cells, membrane potential, 175f
 Turbulent flow, 110, 118, 225ff, 233ff
 Turgor pressure, 138
 Turing model of morphogenesis, 363ff

U

Ubbelohde viscosimeter, 211
 Ultrafiltration coefficient, 156
 Ultrasound
 cavitation effect, 272ff
 definition, 269
 diagnostics, 270f
 diathermy effect, 271f
 echolocation, 274ff
 frequency spectrum, 271
 intensity limits, 274
 lithotripsy, 273
 microthermal heating rate, 273
 penetration in tissue, 270
 radical formation, 272f
 risks in medical diagnosis, 274
 use for vesicle preparation, 72
 Ultraviolet light (UV)
 absorption in the body, 314
 effective dose, 316
 frequency regions, 305, 314
 indirect mechanism, 315
 interaction with biological systems, 314ff
 signature mutations, 315
 solar spectrum, 314
 Unidirectional flux, 160f
 Unstirred layer, 152
 Urea
 dipole moment, 38
 interaction with hydrogen bridges, 323
 reflection coefficient, 136f
 Ussing's equation, 160f

V

val, 34
 Valinomycin, 173f
 Van der Waals
 interactions, 52ff
 radius, 57
 Van't Hoff's
 equation for osmotic pressure, 131, 133
 reaction isotherm, 125
 rule for temperature dependence, 20
 Vapor pressure osmometer, 132
 Variable, definition, 334
 Vater-Pacini corpuscles, 257

Vector space, 340
 Velocity
 distribution of particles, 18
 gradient, 208f
 profile of flow, 209
 Verhulst-Pearl law, 358f
 Vibration
 electrode, 188f
 of molecules, 26f
 Vibration, mechanical
 influence on humans, 256
 intensity, 255
 perceiving by humans and animals, 257
 relative frequency sensibility, 255
 Viscoelasticity, 215ff
 Viscosimeter, 211f
 Viscosity
 definition, 209
 kinematic, 209
 methods of measurement, 211f
 reduced, 210
 relative, 210
 specific, 210
 of suspensions, 219f
 temperature dependence, 40f, 209
 Voigt element of viscoelasticity, 216f
 Voltage clamp, 177, 181
 Voltage dependent channels, 179, 184, 247
 Volterra cycle, 361
 Volume
 compressibility of proteins, 62
 effective of molecules, 65, 80
 regulation of cells, 135ff, 147, 150
 Von Neumann neuronal model, 371

W

Water
 absorption of ultrasound, 276f
 acoustic impedance, 258
 anomalies, 39
 bridges, 46
 cavitation, 272ff
 density, 39f, 258
 dielectric constant (bulk), 34
 dielectric constant (near surfaces), 47
 dipole moment, 38ff
 dispersion of dipoles, 197
 entropy changes, 44ff
 evaporation in the process of heat
 regulation, 251
 hydration kinetics, 46f
 hydrogen bonds, 39ff

- kinematic viscosity, 227
- lifetime of clusters, 41
- magnetic susceptibility, 279
- molecular orbital, 39
- radical formation, 272, 320ff
- radiolysis, 321f
- structure, 39ff
- surface tension, 68
- temperature anomalies, 39
- thermal conductivity, 250
- velocity of sound propagation, 258
- viscosity, 40
- Wave number, 305
- Weak bonds, 33
- Weather, sensitiveness, 294
- Weber-Fechner's law, 258f
- Weighting scales of sound, 260
- Wilson-Sommerfeld relation, 23
- Windows effect of radiation, 302, 316
- Wolff's law, 219, 221
- Work
 - coefficient, 100
 - coordinates, 100
 - differential, 99ff
- Wound
 - healing, electric field effects, 291
 - potential, 189f
- X**
 - Xenopus (clawed frog), 183, 188, 283
 - X-ray crystallography, 181
 - X-rays, 317ff
- Y**
 - Young's modulus, 77f, 215ff
- Z**
 - Zeta-potential (ζ -potential), 85f, 88, 273
 - Zwitterion, 65