Biomedical Applications of Proteomics

Edited by
Jean-Charles Sanchez
Garry L. Corthals
Denis F. Hochstrasser

Biomedical Application of Proteomics
Edited by J.-C. Sanchez, G. Corthals, D. F. Hochstrasser

ISBN: 3-527-30807-5

Related Titles

Reiner Westermeier, Tom Naven

Proteomics in Practice
A Laboratory Manual of Proteome Analysis

2002, 318 pages
ISBN 3-527-30354-5

Jennifer Van Eyk, Michael J. Dunn (eds.)

Proteomic and Genomic Analysis of Cardiovascular Disease

2003, 396 pages
ISBN 3-527-30596-3

Stefan Lorkowski, Paul Cullen (eds.)

Analysing Gene Expression
A Handbook of Methods. Possibilities and Pittfalls

2003, 954 pages
ISBN 3-527-30488-6

Dev Kambhampati (ed.)

Protein Microarray Technology

2003, 256 pages
ISBN 3-527-30597-1

Biomedical Applications of Proteomics

Edited by
Jean-Charles Sanchez
Garry L. Corthals
Denis F. Hochstrasser

WILEY-VCH Verlag GmbH & Co. KGaA

Edited by

Dr. Jean-Charles Sanchez
Dr. Garry L. Corthals
Prof. Dr. Denis F. Hochstrasser

Biomedical Proteomics Research Group
Laboratoire Central de Chimie Clinique
Hopitaux Universitaires de Genève
24, rue Micheli-du-Crest
1211 Genève 14
Switzerland

Library of Congress Card No.: applied for
A catalogue record for this book is available from the British Library.

Bibliographic information published by Die Deutsche Bibliothek
Die Deutsche Bibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data is available in the Internet at <http://dnb.ddb.de>.

Printed in the Federal Republic of Germany
Printed on acid-free paper

Cover design SCHULZ Grafik-Design, Fußgönheim
Composition K+V Fotosatz GmbH, Beerfelden
Printing betz-druck GmbH, Darmstadt
Bookbinding Litges & Dopf Buchbinderei GmbH, Heppenheim

ISBN 3-527-30807-5

To our families at home...

Anne-Catherine, Christiane and Hélène
Jule, Kevin, Lucile, Michael, Rory, Sandrine and Virginie

...and to our BPRG family.

Biomedical Application of Proteomics
Edited by J.-C. Sanchez, G. Corthals, D. F. Hochstrasser

ISBN: 3-527-30807-5

Contents

Biomedical Application of Proteomics
Edited by J.-C. Sanchez, G. Corthals, D. F. Hochstrasser

ISBN: 3-527-30807-5

List of Contributors

Prof. Dr. Ayodele A. Alaiya
Unit of Cancer Proteomics
Department of Oncology
and Pathology
Cancer Center Karolinska
Karolinska Hospital
171 76 Stockholm
Sweden

Dr. Juan Pablo Albar
Laboratorio de Proteómica
Centro Nacional de Biotecnología
CSIC
Cantoblanco
28049 Madrid
Spain

Dr. Laure Allard
Geneva Proteomics Centre, LCCC
Hopital Universitaire de Genève
24, rue Micheli-du-Crest
1211 Genève 14
Switzerland

Prof. Dr. Amos Bairoch
Swiss Institute of Bioinformatics
(ISB/SIB)
University Medical Center
1, rue Michel Servet
1211 Genève 4
Switzerland

Dr. Rosamonde E. Banks
Cancer Research UK Clinical Centre
St. James's University Hospital
Beckett Street
Leeds LS9 7TF
United Kingdom

Dr. Antonio Bernad
Departamento de Inmunología
y Oncología
Centro Nacional de Biotecnología
CSIC
Cantoblanco
28049 Madrid
Spain

Dr. Luca Bini
Functional Proteomic Group
Department of Molecular Biology
University of Siena
Via Fiorentina 1
53100 Siena
Italy

MD Pierre R. Burkhard
Laboratory of CSF 2-DE
Department of Neurology
University Hospital – HUG
24, rue Micheli-du-Crest
1211 Genève 14
Switzerland

Biomedical Application of Proteomics
Edited by J.-C. Sanchez, G. Corthals, D. F. Hochstrasser

ISBN: 3-527-30807-5

Dr. Michel Caron
Laboratoire de Biochimie des Protéines et Protéomique (LBPP)
EA 3408 Immuno-Pathologie et Immuno-Intervention
Université Paris 13
UFR SMBH Léonard de Vinci
93017 Bobigny Cedex
France

Dr. Odile Carrette
Biomedical Proteomics Research Group, LCCC
Genève University Hospital
24, rue Micheli-du-Crest
1211 Genève 14
Switzerland

Dr. Garry L. Corthals
Mass Spectrometry Laboratory
Biomedical Proteomics Research Group (BPRG)
24, rue Micheli-du-Crest
Genève University Hospital
LCCC 5-715
1211 Genève 14
Switzerland

Dr. Yohann Couté
Centre de Génétique Moléculaire et Cellulaire
UMR 5534 CNRS/UCBL1
Batiment Grégor Mendel
43, Boulevard du 11 Novembre 1918
69622 Villeurbanne Cedex
France

David Crettaz
Fondation Centre de Transfusion Sanguine
Croix-Rouge Suisse Lausanne
27, rue du Bugnon
1005 Lausanne
Switzerland

Dr. Pia Davidsson
Experimental Medicine/Molecular Sciences
AstraZeneca Molndal R & D
43083 Molndal
Sweden

Dr. Jean-Jacques Diaz
Centre de Génétique Moléculaire et Cellulaire
CNRS UMR 5534
Université Claude Bernard
43, Boulevard du 11 Novembre 1918
69622 Villeurbanne Cedex
France

Dr. Donna S. Dorow
Center for Molecular Imaging & Trescowthick Research Laboratories
Peter MacCallum Cancer Institute
Melbourne 8006, Victoria
Australia

Michel A. Duchosal
Service d'Hématologie
Centre Hospitalier Universitaire Vaudois
1011 Lausanne
Switzerland

Dr. Ivano Eberini
Gruppo di Studio per la Proteomica e la Struttura delle Proteine
e Centro di Eccellenza sulle Malattie del Sistema Nervoso Centrale e Periferico
Dipartimento di Scienze Farmacologiche
Università degli Studi di Milano
Via G. Balzaretti 9
20133 Milano
Italy

Dr. Giuliano Elia
Department of Chemistry
and Applied Biosciences
Inst. of Pharmaceutical Sciences
Swiss Federal Inst. of Technology
Irchel, Y17 M03
Winterthurerstrasse 190
8057 Zurich
Switzerland

Dr. Maria Livia Famiglietti
Swiss Institute of Bioinformatics
(ISB/SIB)
University Medical Center
1, rue Michel Servet
1211 Genève 4
Switzerland

Dr. Terry B. Farmer
Encysive Pharmaceutical Inc.
7000 Fannin No. 1920
Houston, TX 77030
USA

Elisabeth Gasteiger
Swiss Institute of Bioinformatics
(ISB/SIB)
University Medical Center
1, rue Michel Servet
1211 Genève 4
Switzerland

Dr. Elisabetta Gianazza
Gruppo di Studio per la Proteomica
e la Struttura delle Proteine
e Centro di Eccellenza sulle Malattie
del Sistema Nervoso Centrale
e Periferico
Dipartimento di Scienze
Farmacologiche
Via G. Balzaretti 9
20133 Milano
Italy

Dr. Stéphane Giraud
Centre de Génétique Moléculaire
et Cellulaire
UMR 5534 CNRS/UCBL1
Batiment Grégor Mendel
43, Boulevard du 11 Novembre 1918
69622 Villeurbanne Cedex
France

Igor Golovliov
Department of Clinical Microbiology
Clinical Bacteriology
Umeå University
Univesitetsområdet
90197 Umeå
Sweden

Dr. Anna Greco
Centre de Génétique Moléculaire
et Cellulaire
UMR 5534 CNRS/UCBL1
Batiment Grégor Mendel
43, Boulevard du 11 Novembre 1918
69622 Villeurbanne Cedex
France

Dr. Michael G. Harrington
Molecular Neurology Program
Huntington Medical Research
Institutes
99 North El Molino Avenue
Pasadena, CA 91101
USA

Jana Havlasová
Proteome Center for the Study
of Intracellular Parasitism of Bacteria
Purkyne Military Medical Academy
500 01 Hradec Králové
Trebesská 1575
Czech Republic

Lenka Hernychová
Proteome Center for the Study
of Intracellular Parasitism of Bacteria
Purkyne Military Medical Academy
500 01 Hradec Králové
Trebesská 1575
Czech Republic

Prof. Dr. Denis F. Hochstrasser
Laboratoire Central
de Chimie Clinique
Hopital Cantonal Universitaire
de Genève
24, rue Micheli-du-Crest
1211 Genève 14
Switzerland

Martin Hubálek
Proteome Center for the Study
of Intracellular Parasitism of Bacteria
Purkyne Military Medical Academy
500 01 Hradec Králové
Trebesská 1575
Czech Republic

Dr. Raymonde Joubert-Caron
Laboratoire de Biochimie des Protéines
et Protéomique (LBPP)
EA 3408 Immuno-Pathologie
et Immuno-Intervention
Université Paris 13
UFR SMBH Léonard de Vinci
74, rue Marcel Cachin
93017 Bobigny Cedex
France

Dr. Hermann Lage
Institut für Pathologie
Universitätsklinikum Charité
Campus Charité-Mitte
Schumannstrasse 20–21
10117 Berlin
Germany

Juraj Lenco
Proteome Center for the Study
of Intracellular Parasitism of Bacteria
Purkyne Military Medical Academy
500 01 Hradec Králové
Trebesská 1575
Czech Republic

Dr. Sabrina Liberatori
Functional Proteomic Group
Department of Molecular Biology
University of Siena
Via Fiorentina 1
53100 Siena
Italy

Dr. Juan Antonio López
Unidad de Proteómica
Centro Nacional de Investigaciones
Cardiovasculares Carlos III
Sinesio Delgado, 4
28049 Madrid
Spain

Dr. Didier Lutomski
Laboratoire de Biochimie des Protéines
et Protéomique (LBPP)
EA 3408 Immuno-Pathologie
et Immuno-Intervention
Université Paris 13
UFR SMBH Léonard de Vinci
93017 Bobigny Cedex
France

Ales Macela
Proteome Center for the Study
of Intracellular Parasitism of Bacteria
Purkyne Military Medical Academy
500 01 Hradec Králové
Trebesská 1575
Czech Republic

Dr. Fulvio Magni
University Milano-Bicocca, DIMESAB
Via Cadore 48
20052 Monza (MI)
Italy

Prof. Dr. Paolo Mocarelli
University Department of Laboratory Medicine
Hospital of Desio
Via Mazzini 1
20033 Desio (MI)
Italy

Prof. Dr. Dario Neri
Department of Chemistry
and Applied Biosciences
Institute of Pharmaceutical Sciences
Swiss Federal Institute of Technology
Irchel, Y36 M14
Winterthurerstrasse 190
8057 Zurich
Switzerland

Prof. Dr. Vitaliano Pallini
Functional Proteomic Group
Department of Molecular Biology
University of Siena
Via Fiorentina 1
53100 Siena
Italy

Dr. Julia Poland
Institut für Medizinische
und Chemische Labordiagnostik
LKH-Klagenfurt
St. Veiter Strasse 47A
9020 Klagenfurt
Austria

Dr. Manfredo Quadroni
Institut de Biochimie
Faculté de Médecine de l'Université
de Lausanne
1066 Epalinges
Switzerland

Prof. MD MBA Marc A. Reymond
Department of Surgery
Otto-von-Guericke University
Leipziger Strasse 44
39120 Magdeburg
Germany

Christoph Roesli
Department of Chemistry
and Applied Biosciences
Institute of Pharmaceutical Sciences
Swiss Federal Institute of Technology
Irchel, Y17 M86
Winterthurerstrasse 190
8057 Zurich
Switzerland

Jascha-Nikolai Rybak
Department of Chemistry
and Applied Biosciences
Institute of Pharmaceutical Sciences
Swiss Federal Institute of Technology
Irchel, Y17 M86
Winterthurerstrasse 190
8057 Zurich
Switzerland

Dr. Jean-Charles Sanchez
Biomedical Proteomics Research
Group
Central Clinical Chemistry Laboratory
(5-715)
Genève University Hospital
24, rue Micheli-du-Crest
1211 Genève 14
Switzerland

Dr. Cecilia Sarto
University Department
of Laboratory Medicine
University Milano-Bicocca
Hospital of Desio
Via Mazzini 1
20033 Desio (MI)
Italy

Prof. Dr. Dirk Schadendorf
Klinische Kooperationseinheit
für Dermatoonkologie (DKFZ)
Universitäts-Hautklinik Mannheim
Theodor-Kutzer-Ufer 1
68135 Mannheim
Germany

Simone Scheurer
Department of Chemistry
and Applied Biosciences
Institute of Pharmaceutical Sciences
Swiss Federal Institute of Technology
Irchel, Y17 M86
Winterthurerstrasse 190
8057 Zurich
Switzerland

Gerhard Schmid
Biomedical Proteomics Research
Group
Central Clinical Chemistry Laboratory
Genève University Hospital
24, rue Micheli-du-Crest
1211 Genève 14
Switzerland

Prof. Dr. Philippe Schneider
Fondation Centre de Transfusion
Sanguine
Croix-Rouge Suisse Lausanne
27, rue du Bugnon
1005 Lausanne
Switzerland

Prof. Dr. Peter J. Selby
Cancer Research UK Clinical Centre
St James's University Hospital
Beckett Street
Leeds LS9 7TF
United Kingdom

Prof. Dr. Richard J. Simpson
Joint Proteomic Structure Laboratory
Ludwig Institute for Cancer Research
Royal Melbourne Hospital
Post Office Box 2008
3050 Melbourne, Victoria
Australia

Prof. Dr. Pranav Sinha
Institut für medizinische
und chemische Labordiagnostik
LKH-Klagenfurt
St. Veiter Strasse 47
9020 Klagenfurt
Austria

Anders Sjöstedt
Department of Clinical Microbiology
Clinical Bacteriology
Umeå University
Univesitetsområdet
90197 Umeå
Sweden

Dr. Markus Stoeckli
Novartis Institute for BioMedical
Research
Analytical and Imaging Sciences
WSJ-503.1101
Lichtstrasse 35
4002 Basel
Switzerland

Dr. Jirí Stulík
Proteome Center for the Study
of Intracellular Parasitism of Bacteria
Purkyne Military Medical Academy
500 01 Hradec Králové
Trebesská 1575
Czech Republic

Amalio Telenti
Institut de Microbiologie
de l'Université de Lausanne
Centre Hospitalier Universitaire
Vaudois
1011 Lausanne
Switzerland

Prof. Dr. Jean-Daniel Tissot
Fondation Centre de Transfusion
Sanguine
Croix-Rouge Suisse Lausanne
27, rue du Bugnon
1005 Lausanne
Switzerland

BS. Cristina Valsecchi
University Department of Laboratory
Medicine
University Milano-Bicocca
Hospital of Desio
Via Mazzini 1
20033 Desio (MI)
Italy

Dr. Françoise Vuadens
Fondation Centre de Transfusion
Sanguine
Croix-Rouge Suisse Lausanne
27, rue du Bugnon
1005 Lausanne
Switzerland

Dr. Yum Lina Yip
Swiss Institute of Bioinformatics
(ISB/SIB)
University Medical Center
1, rue Michel Servet
1211 Genève 4
Switzerland

Abbreviations

2-D LPE	two-dimensional liquid-phase electrophoresis
2-D PAGE	two-dimensional polyacrylamide gel electrophoresis
2-D	two-dimensional
2-DE	two-dimensional electrophoresis
β-3AR	beta-3-adrenergic receptor
µLC-ESI-MS/MS	microcapillary liquid chromatography-electrospray ionization tandem mass-spectrometry
B-CLL	chronic lymphocytic leukemia
bFGF	basic fibroblast growth factor
CAF-MALDI	chemically assisted fragmentation MALDI
CEA	carcinoembryonic antigen
CGAP	Cancer Genome Anatomy Project
CIMP	cytosine phosphoguanosine (CpG) island methylated phenotype
CIN	chromosomal instability
CRP	C-reactive protein
CT	computed tomography
CTL	cytotoxic T lymphocyte
DIGE	**di**fferential **g**el **e**lectrophoresis
DPD	Differential Peptide Display™
EC	endothelial cell
ECM	extracellular matrix
ESI	electrospray ionization
EST	expressed sequence tag
FABP	fatty acid-binding proteins
FABP2	fatty acid-binding protein
FACS	fluorescence-activated cell sorters
FAP	familial adenomatous polyposis
FDPs	fibrin-fibrinogen degradation products
FFE	free flow electrophoresis
FOBT	fecal occult blood test
FTD	frontotemporal dementia

Biomedical Application of Proteomics
Edited by J.-C. Sanchez, G. Corthals, D. F. Hochstrasser

ISBN: 3-527-30807-5

FT-ICR	Fourier transform ion cyclotron
FTICR-MS	Fourier transform ion cyclotron resonance mass spectrometry
FT-MS	Fourier transform-MS
GLUT4	glucose transporter 4
HAP-HPLC	hydroxyapatite HPLC
HIF	hypoxia-inducible transcription factor
hK6	human kallikrein 6
HMG-CoA	3-hydroxy-3-methylglutaryl-coenzyme A
HNPCC	hereditary non-polyposis colon cancer
HRE	hypoxia-response element
IAPP	islet-amyloid polypeptide
ICAT	isotope-coded affinity tag
IEF	isoelectric focusing
IL-8	interleukin 8
IMT	intima-media thickness
IPG	immobilized pH gradients
LCM	laser-capture microdissection
LDS	lithium dodecyl sulfate
mAb	monoclonal antibodies
MALDI-TOF-MS	matrix-assisted laser desorption ionization/time-of-flight mass spectrometry
MBP	myelin basic protein
MDR	multidrug resistance
MIN	microsatellite instability
MMPs	matrix metalloproteinases
MMR	mismatch repair
NephGE	non-equilibrium pH gradient electrophoresis
NSE	neuron-specific enolase
PC-2	prohormone convertase 2
PCA	principal components analysis
PD-ECGF	platelet-derived endothelial cell growth factor
PDGF	platelet-derived growth factor
PGDS	prostaglandin D synthase
PGK	phosphoglycerate kinase
Pgp	P-glycoprotein
PLS	partial least squares analysis
PPARγ	peroxisome proliferator-activated receptor gamma
PSD	post source decay
PSMA	prostate-specific membrane antigen
Q-IT	quadrupole ion trap
Q-TOF	quadrupole-TOF
RCC	renal cell carcinoma
RP-HPLC	reverse phase high-performance liquid chromatography
rtPA	recombinant tissue plasminogen activator
RT-PCR	reverse transcription-polymerase chain reaction

RXR	retinoid X receptor
SAGE	serial analysis of gene expression
SAX	strong anonic exchange
SDS-PAGE	sodium dodecyl sulfate-polyacrylamide gel electrophoresis
SELDI	surface-enhanced laser desorption ionization
SHRSP	spontaneously hypertensive stroke-prone rat
T1DM	type 1 diabetes mellitus
TAU	transabdominal ultrasound
TCC	transitional cell carcinoma
TGFα	transforming growth factor α
TIAs	transient ischemic attacks
TIMP	tissue inhibitor of metalloproteinases
Tm	thrombomodulin
TNFα	tumor necrosis factor α
TP	thymidine phosphorylase
TRAP	thrombin receptor-activating peptide
TVS	transvaginal sonography
VEGF	vascular endothelial growth factor
VRCs	viral replication compartments

Introduction

Denis F. Hochstrasser, Garry L. Corthals, and Jean-Charles Sanchez

The Research and Discovery Circle

In medicine, as in any other life sciences discipline, research should start by raising a question (hypothesis) to solve a problem. Then the scientist should select the most appropriated method(s) to answer the question and solve the dilemma. The obtained intermediate and final results may answer the question or may raise additional or new questions. It closes the loop of research and discovery.

Quite often, the development of new methods offers the opportunity to raise new questions and sometimes, although not well accepted, the application of new methods without initial and predefined questions offers new data that can be used to raise many valid interrogations. It is like entering the research circle at a different entry point.

For example, sequencing genomes has currently provides numerous valuable data and can be considered to be the application of new methods without predetermined questions – non hypothesis driven research. In fact, many "omic" experiments, such as genomic, transcriptomic, proteomic, and metabolomic studies, can be classified into this category. Even without well-defined medically driven hypotheses, the spectacular and exponential development of the "omic science" has been an impressive driving force for large-scale method development. Currently, scientists have the range of many powerful tools available to answer very well-defined, pertinent, and difficult hypothesis driven research.

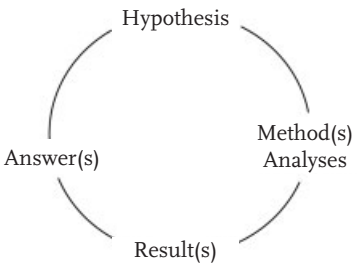


Figure 1.1 The research circle.

Biomedical Application of Proteomics
Edited by J.-C. Sanchez, G. L. Corthals, D. F. Hochstrasser

ISBN: 3-527-30807-5

The Five Dimensions of Diseases

There are five essential elements in medical practice: the diagnosis, the prognosis, the therapy, the prevention and the prediction of diseases. In Greek "*diagnosis*" means "*capable of recognition*" and "*prognosis*" means "*to know in advance*". Ideally, one should predict diseases at an early stage when preventive actions are still possible. Otherwise, early diagnosis should be established to determine the appropriate therapy. The prognosis is good or better when a cure for the disease still exists.

Diseases can be sorted into five categories: genetic, infectious, environmental, due to "storage handling", and apoptosis, time (aging). They often combine several factors from more than one category.

For example, patients with a genetic predisposition and living in a low socioeconomic environment may be predisposed to chronic tuberculosis and to develop amyloidosis due to β_2-microglobulin deposition over time. Patients with the *apoE* genotype 4/4 have a much greater likelihood of developing Alzheimer's disease, where protein accumulation in the brain may cause numerous neuronal death. The genetic predisposition of certain African populations favors Epstein-Barr virus infection and the development of Burkitt's lymphoma.

Genetic diseases or genetic predispositions are best diagnosed by a genomic approach. Environmental diseases could be detected by genomic, proteomic, or metabolomic studies depending at which level the offending agent acts. DNA methylation, protein desialylation, or glycation require genomic, proteomic, and metabolomic approaches, respectively.

Storage mishandling and apoptosis deregulation in degenerative diseases and cancer can be analyzed by proteomic and genomic approaches.

Most infectious agents are detected early by polymerase chain reaction (PCR) or reverse transcription (RT-)PCR but the host response to a new or previous infection is better defined by serology testing (proteins).

So when and where can biomedical applications of proteomics solve questions related to the five elements of medical practice and the five dimensions of diseases? This book offers to the reader potential answers to these crucial questions

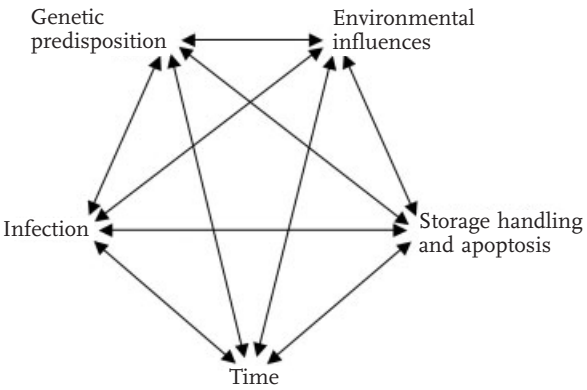


Figure 1.2 The five "origins" of diseases.

and summarizes the strategies that have been developed relatively recently. It is not a comprehensive textbook; like an impressionist painting, it touches several areas of biomedical applications to highlight "the big picture".

Book Content

As discussed above, the first chapter (Part I) on aspects of biomedical research and written by Marc Reymond highlights the need for new diagnostic tools and for better prognosis evaluation. It underscores the need for human samples in proteomic research and the practical difficulties involved in collecting them and also addresses related ethical issues.

In Part II, Dario Neri and his team describe their elegant work on targeting vascular proteins by antibodies in the development of an anti-angiogenic treatment for cancer. They highlight the advantage and pitfalls of several proteomic approaches. Elisabeth Gianazza and Ivano Eberini highlight some aspects of vascular diseases and atherosclerosis and then proteomic finding in those conditions. Laure Allard and co-workers summarize the physiopathology of stroke and discuss some aspects of its current diagnosis and treatment. The authors then describe several proteomic approaches used to unravel new biomarkers of brain diseases in spinal fluid and serum.

In Part III, Rosamonde Banks describes biomedical applications of proteomics in renal cancer, Cecilia Sarto and co-workers summarize available data on heat shock protein 27 in cancer. Richard Simpson and Donna Dorow present and discuss proteomic and genomic approaches for biomarker discovery in colorectal cancer. Ayodela Alaiya underscore the difficulty of finding early diagnostic markers in cancer and especially in ovarian tumor. Juan López and his colleagues describe their phenotypic characterization of mesenchymal stem cells. Raymonde Joubert-Caron and co-workers publish the method that they have used to study lymphoblastoid and lymphoma cells and the results that they have obtained after treatment of those cells by a chemotherapeutic agent.

In Part IV on pharmacology and toxicology, Julia Poland, Dirk Schadendorf, Hermann Lage, and Pranav Sinha study the molecular factors that help in understanding the mechanism of chemoresistance in cancer biology. They summarize their results with several cell lines and discuss potential mechanisms of resistance.

In a different field of pharmacology, the treatment of diabetes, Gerhard Schmid and colleagues describe the different types of diabetes and their pathophysiology. They reference several genomic and proteomic approaches to get a better insight into the molecular cascade involved in this family of diseases.

In Part V on infectious diseases, Luca Bini and his co-workers describe acute-phase proteins and antibody profiles as diagnostic indicators in human plasma. New insights into HIV lymphocyte infection are discussed by Françoise Vuadens and her colleagues and the modifications of host cell induced by herpes simplex virus type 1 are described by Anna Greco and her colleagues. Jirí Stulík and his

colleagues summarize their proteomic work in tularemia, a rare disease caused by an intracellular pathogen.

In Part VI, devoted to the central nervous system, Pierre Burkhard and his colleagues list the biomedical applications of proteomics in the field of neuroscience with numerous references to the literature. Pia Davidsson and Mike Harrington describe the analysis of cerebrospinal fluid by electrophoretic and chromatographic methods. They nicely demonstrate the discovery of new brain disease markers using a proteomic approach. Odile Carrette and her colleagues describe their work in discovering new biomarkers for the diagnosis of Alzheimer's disease.

Finally, in the last section, somewhat more technical, Markus Stoeckli and Terry Farmer demonstrate the tremendous potential of MALDI-MS imaging in biomedical research. Yum Lina Yip and her colleagues describe proteome databases and their related tools.

Part I
Aspects in Biomedical Research

Biomedical Application of Proteomics
Edited by J.-C. Sanchez, G. L. Corthals, D. F. Hochstrasser

ISBN: 3-527-30807-5

1
Proteomics in Biomedicine – A Tool, a Science, or an Art?

Marc A. Reymond

1.1
Introduction

A vast range of molecular biology, cellular biology, and advanced validation tools are available in modern medical research today, and some of these tools, capable of separating proteins on gel then identifying them using lasers, have given rise – over the last 25 years – to a novel science called proteomics. Proteomics tools are broadly applied in human disease. They allow the classification of diseases on a molecular basis, giving deep insights into the pathophysiology and prognosis of disorders, and finally providing a systematic search for diagnostic and therapeutic targets. Many pharmaceutical companies have been working for years on the application of proteomics in disease, some of them having invested hundreds of millions of dollars in the necessary infrastructures. Proteomics research consortia are also funded by the state in various countries, in particular in the USA, Canada, France, and Germany.

The first applications, in particular in the diagnostic market, are emerging: these may well change the fate of thousands of patients, for example by allowing cancer diagnosis in early stages, when these cancers can still be cured. Approaches combining the analysis of complex serum protein patterns with advanced bioinformatics have been shown to detect ovarian, prostate, breast, pancreatic, and bladder cancer with outstanding accuracy, and are paving the way for population screening. It appears that, over the last 10 years, proteomics has matured from a collection of tools to a novel science.

What is coming next? Many physicians taking part in the modern saga of proteomics have the strong feeling that these powerful validation tools and the resulting scientific knowledge will change their everyday practice. The proteome of a cell is constantly changing: proteins fold and unfold, shutting on and off their function within a complex society of networked, interacting molecules. What is normal? What is disease? What is aging? Is aging normal? Proteomics tools will give a novel definition to health and disease. "*Tomber malade, vieille notion qui ne tient plus devant les données de la science actuelle*" says Jules Romains' Doctor Knock.

When compared with progress made in genomics, the role of proteomics in biomedicine has so far been limited, but the importance of proteomics studies is ex-

Biomedical Application of Proteomics
Edited by J.-C. Sanchez, G. L. Corthals, D. F. Hochstrasser

ISBN: 3-527-30807-5

pected to increase massively over the next few years. Conventional technologies such as two-dimensional polyacrylamide gel electrophoresis (2-D PAGE) and novel tools such as protein chips allow disease or risk patterns to be recognized and linked with patient-specific outcomes. Proteomics tools, much better than genetic techniques, will allow the investigation of the evolution of disease and the influence of treatment over time. Thus, it is reasonable to expect that, over the next 20 years, proteomics will make a major contribution to the ongoing shift of medicine from a generalized, impersonal scientific knowledge to a singular, individual, personalized practice. As a consequence, proteomics might well contribute to change the art of medicine.

1.2
Diagnosis and Prognosis: an Oxymoron

Two core disciplines of modern medicine are diagnosis and prognosis, which are often considered together. For example, in almost every family an unfortunate destiny illustrates that cancer diagnosis is associated with dismal prognosis. In medical oncology, both disciplines are also closely associated. The best proof is given by the International Union against Cancer (UICC) classification itself, which forecasts a particular patient's prognosis by defining the extent of disease at the time of diagnosis. Modern developments in proteomics tend to perpetuate this traditional association between diagnosis and prognosis, since molecular markers of disease can often be used for both purposes, at least when the question is discussed at the bench. However, when considered from the bedside, associating diagnosis with prognosis is an oxymoron: diagnosis is a generalization, the result of a classification that is independent of the individual case, while the prognosis describes the probable course of the illness in a particular patient. Another distinction has to be made in terms of time. Diagnosis freezes the disease process, providing the physician (and the patient) with a snapshot picture. In contrast, prognosis expands a momentary state into a defined time sequence, a kind of motion picture.

In molecular medicine, a diagnosis can be made by identifying common traits between sick people. For example, common proteomics patterns can be determined in cancer patients: this approach is usually called "expression proteomics". In contrast, to forecast prognosis, only those qualitative and quantitative differences in protein expression that ultimately result in dysfunction in cellular behavior and thus in clinical phenotype over time are selected (so-called functional proteomics).

1.3 The Dimensions of Prognosis

As Hippocrates recognized more than 2000 years ago, forecasting plays a central role in decision making in medicine, and prognosis concerns not only the end of disease, but also its progression and its duration:

> "By foreseeing and foretelling, in the presence of the sick, the present, the past, and the future …, he (the physician) will be the more readily believed to be acquainted with the circumstances of the sick; so that men will have confidence to in trust themselves to such a physician. And he will manage the cure best who has foreseen what is to happen from the present state of matters." (*The Book of Prognosis*)

The patient's history helps to predict prognosis. Different endpoints of interest can be evaluated, which may be multiple (for example, as Hippocrates had already recognized, survival or therapy response). Prognosis also changes with the therapeutic choices made. A prognosis can vary over time. Thus, in contrast to diagnosis, prognosis is multidimensional.

In modern, scientific medical thinking, prognosis depends on a collection of variables called prognostic factors. In cancer, for example, prognostic factors include macroscopic and microscopic characteristics of diseased organs and tissues. However, prognostic factors concern not only the diseased organ, but also the patient and the treatment administered. In solid cancers, the most decisive prognostic factor is the result of the surgical resection – microscopically curative or not – information that cannot and never will be provided by a molecular pattern. Thus, and this is a significant boundary for biomedical application of proteomics, the analysis of diseased tissues or organs alone, for example with proteomics tools, will only provide part of the prognostic information.

As a corollary of the above, molecular data have to be linked to personalized clinical information in prognostic and therapeutic biomedical research, including validation tasks. To be able to achieve accurate forecasting or validation (e.g. therapy response studies), prospective data acquisition is necessary over many months or even years.

In clinical practice, the number of clinical and pathological variables considered in a particular case is usually between 10 and 100. Of course, this conventional prognostic information gives the clinical framework for proteomics studies, and not the opposite. A second consequence will be detailed below, namely that the need for personalized information in proteomics studies generates a series of ethical, information, privacy, property, and consent problems.

1.4
Protein Technologies, Diagnosis, and Prognosis

So far, most current diagnostic tests are based on the detection and quantification of single proteins in body fluids. The direct availability of these proteins in body fluids, in particular in the serum, is an important feature in clinical practice, because repeated sampling is possible with minimal need for invasive tests. For example, tests such as carcinoembryonary antigen (CEA) in colorectal cancer, prostate-specific antigen (PSA) in prostate cancer, alpha-fetoprotein (AFP) in hepatocellular carcinoma, etc. are currently used in clinical practice. Historically, these tests – most of them based on ELISA technology – were developed only on empirical grounds based on observation and correlation of measured levels of single proteins with the diagnosis or recurrence of disease. A common characteristic is their relatively low predictive value, so that they cannot be used alone for diagnostic purposes, and they have to be complemented with other procedures, in cancer usually a biopsy.

In the meantime, as introduced above, novel analytical tools have turned attention to possible improvements of these classical tests in order to improve their specificity and thus their potential for diagnosis, in particular for population screening. Recently, genomic, polymerase chain reaction (PCR)-based diagnostic tests have been introduced into the market. In proteomics research, patterns of protein expression have been shown to yield more biologically relevant and clinically useful information than assays of single proteins. For example, protein chips coupled with SELDI/TOF-MS (surface-enhanced laser desorption ionization/time-of-flight mass spectrometry), when coupled to a pattern-matching algorithm, have led to the identification of biomarker patterns in pancreatic juice, urine, or serum, which correctly classified cancer and non-cancer with high sensitivity and specificity in patient populations suffering from several cancers.

1.5
Individual Protein Patterns in Clinical Practice

A prerequisite for such pattern-based protein studies is access to well-characterized normal and pathological samples obtained from human patients at different stages of disease or from normal volunteers for comparison purposes, as well as access to clinical follow-up data, including treatment and outcome data.

In practice, such access is limited by a number of boundaries. The access to normal tissues is especially difficult in the case of volunteers, who should, of course, not be put into any danger by sampling procedures. Another point is to decide if physicians have the right to use organs, tissues, and/or body fluids obtained from their patients for their own research, or if these patients remain the owners of the parts separated from their body. It is obvious that major economic interests depend on the answer to this question. Marketing of body fluids, for example blood or plasmapheresis products, or of organs such as maternal placenta,

are an example of such an economic interest. The indirect profits to be generated from target validation using human tissue is another example.

The researcher might expect that such questions are to be answered at legal or regulatory levels. The reality is that, so far, the ethical, legal, and regulatory framework for using human tissue in biomedical research is still vague, and varies between different countries. Indeed, important initiatives have been undertaken in this respect by several European countries – in particular France and the UK – as well as in the United States, that might result in precise guidelines, rules or even laws on this matter. Both conditions – missing rules and future unpredictable regulations affecting long-term projects – make industrial and academic researchers currently feel insecure when using samples from human origin. Are they allowed to do so? Can they transfer these samples to an academic or industrial research partner? Should traceability of the patient be guaranteed, and if so, how can patient's anonymity be warranted?

1.6 New Research Tools, Old Problems

The rapid development of biotechnologies over the last 25 years has generated a number of bioethical questions, and although these questions might – at first sight – appear new, a careful analysis shows that this is not the case. The human corpse has held great symbolic significance in various civilizations. In ancient Egypt, performing an incision into the body in order to remove the organs for embalming was considered to be an injury to the integrity of the dead person; the other embalmers present would throw stones at the man who made the cut. Since they weren't really trying to hurt him, this has to be considered as a symbolic part of the ceremony. The ancient Hebrews, in a practice that continues in orthodox Judaism, expressed concerns about the desecration of the body by "mutilation" and have shaped a long tradition of resistance to the dissection of Jews for teaching purposes. Thus, the symbolic meaning attached to human tissue is a permanent feature in history, at least since the Antiquity. In modern time, the problem has become more complex because of the vanishing frontiers in the definition and limits of "human tissues", and because some experimental procedures now allow some human materials to be immortalized.

1.7 What is Human Material?

Human samples include cadavers, organs, tissues, cells, body fluids, hair, nails, and body waste products. A cadaver is defined as a dead human body. Organs mean any human organ (heart, liver, kidney, lung, pancreas, small bowel, etc.). Human tissues include normal or pathologic human tissues obtained from biopsies or surgical specimens, including corneas. Human cells include cells of so-

matic origin and gonadic cells. Somatic cells are cells from any body organ, including stem cells from the umbilical cord or bone marrow, but not sperm, ovula, and embryonic stem cells. Gonadic cells include sperm and ovula. Embryonic (stem) cells are cells obtained from a human embryo. Body fluids include blood, products derived from blood and cerebrospinal fluid. Hair, nails, and body waste products (urine, stool) are components of the human body that are traditionally excluded from the categories above. It is important to note that a given sample can eventually change category (e.g. products derived from an embryonic stem cell might be used in tissue engineering or in organ transplantation).

1.8
Using Human Tissue in Biomedical Research – Potential Pitfalls

When, or preferably before, using human tissue or body fluids in biomedicine, the researcher should be aware of the potential pitfalls. Privacy for the living and for the dead and profit for biomedical researchers are two issues of importance for the patient, who is the primary supplier of human samples, and for their relatives. Recent scandals have highlighted the growing need to enforce patient's consent, and to safeguard the identity and civil rights of donors and relatives. They have also underscored the potential risks for a researcher or for a physician in performing research or transferring human tissue to a third party without having obtained a formal authorization from the patient or their relatives.

UNESCO has contributed to the elaboration of the ethical framework surrounding biomedical research by formulating the principles of the Universal Declaration on the Human Genome and Human Rights, adopted in 1998. Article 5a defines the rights of those who undergo "research, treatment or diagnosis" on their own genome. Article 5c affirms respect for the right of each person to decide whether or not to be informed of the results of a genetic examination. In Europe, the only binding instrument on bioethics is the Convention for the Protection of Human Rights and the Dignity of Human Being with regard to the Application of Biology and Medicine – the so-called Convention on Human Rights and Biomedicine – adopted in 1997 by the Council of Europe. Article 21 of the Convention declares that the human body and its parts shall not, as such, give rise to financial gain. Article 22 states that, "when in the course of an intervention any part of a human body is removed, it may be stored and used for a purpose other than that for which it was removed, only if this is done in conformity with appropriate information and consent procedures".

1.9
Informed Consent

In practice, to facilitate the collection, storage, and appropriate use of human biological materials, but also to protect biomedical researchers against undue claims,

informed consent forms should be submitted to potential subjects, providing them with a sufficient number of options to help them understand clearly the nature of the decision they are about to make. Such options might include, for example, refusing use of their biological materials in research; permitting only unidentified or unlinked use of their biological materials in research; permitting coded or identified use of their biological materials for one particular study only, with no further contact permitted to ask for permission to do further studies; permitting coded or identified use of their biological materials for one particular study only, with further contact permitted to ask for permission to do further studies; permitting coded or identified use of their biological materials for any study relating to the condition for which the sample was originally collected, with further contact allowed to seek permission for other types of studies; or permitting coded use of their biological material for any kind of future study.

As a general rule, the researcher is interested in keeping the definition of the field of research as broad as possible: when no informed consent is available, the risk of doing anonymized, ethically widely accepted biomedical research is low for the researcher in most industrial countries. In the presence of informed consent, use of samples beyond purpose might be a violation of the patient's rights!

Beside informed consent, researchers should address several topics when designing studies using human tissues. Reward for the patient should be kept minimal. Sample transfer should be organized along non-profit lines, a condition in Europe. Sampling procedures should be at no risk for human volunteers, and at low risk for patients. Biosafety aspects should be addressed, in particular when international collaborations are intended (in particular in or with the US, where these aspects are currently playing an extraordinary role), or when collaboration between academia and industry is sought. Regulations exist in some countries (e.g. in France) for the importation and exportation of human tissues. The question of data acquisition and storage should be addressed in accordance with national data protection regulations, in particular when using computerized databases.

If follow-up information is to be taken, the authorization of such information request should be obtained from the beginning. The right for information or for no information should be also addressed: what happens if significant prognostic information is gained during the course of research, or if a novel therapy is discovered that could help a particular patient? Should a patient be informed about potential health risks linked to a particular protein expression profile? What are the actions – if any – when a pathological pattern is recognized in a person that was enrolled as a normal volunteer? As is the case with other topics in this chapter, the information problem in health and disease is not new. Molière noted it as he wrote, about 400 years ago, this famous sentence: "*Les gens bien portants sont des malades qui s'ignorent*". The question is how to decide if, when, and how the physician has to inform the patient about potential, beginning, or threatening disease. The outstanding predictive value of gene expression patterns – when they are matched with large prospective databases – gives a novel dimension to the question.

The issues of patenting and commercialization should be also addressed, for instance by informing the patient that he or she will have no commercial rights on potential research results, or by choosing another strategy. If the samples are intended to be transferred to another institution, research laboratory or private company, this should be told to the patient. Freedom should be given to the patient to request the destruction of the sample and of the related data at any time point during the course of the study. Finally, to optimize the security level both for the donor and for the researcher, traceability of the donor should be ensured.

1.10
Specificity of Proteomics Studies

As we stated at the beginning of this chapter, proteomics studies have so far played only a minor role in clinical medicine when compared with genetic studies, but this role is expected to increase over the next years. Proteomics studies, more than genetic investigations, allow the investigation of disease over time, and often require repeated analyses over the course of disease. Clearly, anonymization of probes is not directly compatible with such follow-up studies, although the problem can be circumvented by contracting a third party for anonymization and follow-up tasks, so that the biological information never comes into contact with the patient's identity.

Genetic studies barely allow the analysis of body fluids, at least when compared with proteomics studies. Ethics committee usually deliver authorization for sampling a few milliliters of blood or serum in patients or in normal volunteers more easily than for performing biopsies. In this respect, proteomics researchers might benefit from favorable framework conditions. However, the problem of serum analysis in biomedical proteomics research is complicated by the large quantities (up to several liters) of serum that are necessary to identify certain peptides present in a very low concentration, so that the serum of hundreds of patients has to be pooled.

Finally, another important specificity of proteomics studies is that they do not amplify genetic information, so that the data protection and privacy issues are less important than when performing, for example, genome-wide cDNA expression studies. However, researchers should not overestimate this difference since even a single but significant piece of information gained from the protein pattern might imply significant privacy issues.

1.11
Conclusion and Summary

Over the last two decades, medical research has begun to make extensive use of products of human origin in diagnostics, therapeutics, and most recently, in predictive medicine. It is now expected that modern medicine will shift from a gener-

alized, impersonal scientific knowledge to a singular, individual, personalized practice. Since proteomics analyze the functional molecules of a normal or diseased cell, and because proteomics take into account dynamic aspects over time, it is expected that this novel science will make a major contribution to the development of medical art.

By analyzing tissues of human origin, linking disease and/or risk-specific molecular patterns with patient identity, the researcher is endorsing potential risks, linked to protection of privacy, violation of existing or future legislation, unauthorized or unethical use of samples, financial claims on research results, non-communication of prognostic information of potential risks, etc.

The ethical questions raised by molecular research, in particular the balancing of the interests of greater control – defended by patients' organizations – versus wider access to human biological samples and related data – supported by industry and insurance companies – will be the subject of renewed controversy. This will challenge the ability of the researchers' community to deliver innovation and the capacity of governments to deal ethically with the protection of the rights of patients.

In everyday practice, in the absence of clearly established national and international frameworks these many issues can accurately be addressed by ensuring transparent informed patient consent.

1.12 Further Reading

1 Steinert, R., Von Hoegen, P., Fels, L.M., Gunther, K., Lippert, H., Reymond, M.A. *Am. J. Pharmacogenomics* **2003**, *3(2)*, 107–115.

2 Steinert, R., Buschmann, T., van der Linden, M., Fels, L.M., Lippert, H., Reymond, M.A. *Technol. Cancer Res. Treat.* **2002**, *1(4)*, 297–304. Review.

3 Reymond, M.A., Steinert, R., Escourrou, J., Fourtanier, G. *Dig. Dis.* **2002**, *20(3/4)*, 257–265.

4 Reymond, M.A., Steinert, R., Eder, F., Lippert, H. *Proteomics* **2003** (in press).

Part II
Blood Vessels

Biomedical Application of Proteomics
Edited by J.-C. Sanchez, G. L. Corthals, D. F. Hochstrasser

ISBN: 3-527-30807-5

2
Antibody-based Vascular Targeting: Proteomic Techniques for the Identification and Quantification of Membrane Proteins on Endothelial Cells

Simone Scheurer, Jascha-Nikolai Rybak, Christoph Roesli, Giuliano Elia, and Dario Neri

2.1 Introduction

Cancer chemotherapy relies on the expectation that anticancer drugs will preferentially kill rapidly dividing tumor cells, rather than normal cells. Since a large proportion of the tumor cells have to be killed in order to obtain and maintain a complete remission, large doses of drugs are typically used, with significant toxicity towards proliferating non-malignant cells. Indeed, the majority of pharmacological approaches for the treatment of solid tumors suffer from poor selectivity, thus limiting dose escalation (i.e. the doses of drug which are required to kill tumor cells cause unacceptable toxicities to normal tissues).

The development of more selective anticancer drugs, with better discrimination between tumor cells and normal cells, is possibly the most important goal of modern anticancer research.

One avenue towards the development of more selective, better anticancer drugs consists in the targeted delivery of bioactive molecules (drugs, cytokines, procoagulant factors, photosensitizers, radionuclides, etc.) to the tumor environment by means of binding molecules (e.g. human antibodies) specific for tumor-associated markers.

Even though the concept of a selective delivery of therapeutics to the tumor environment was first envisioned by Paul Ehrlich at the end of the nineteenth century, several technologies had to be developed before this therapeutic strategy could become a reality. Indeed, in spite of the fact that 13 monoclonal antibodies have already been approved by the US Food and Drug Administration for therapeutic applications, the following considerations outline why today it is more urgent than ever to develop targeted anticancer therapies for the treatment of disseminated solid tumors:

- Most chemotherapeutic agents do not preferentially accumulate at the tumor site. Indeed, the dose of drug that reaches the tumor (normalized per gram of tissue) may be as little as 5–10% of the dose that accumulates in normal organs! [1]. The high interstitial pressure and the irregular vasculature of the tumor account, in part, for the difficult uptake of drugs by tumor cells. On top of that, the activity of multidrug resistance proteins may further decrease drug uptake.

Biomedical Application of Proteomics
Edited by J.-C. Sanchez, G. L. Corthals, D. F. Hochstrasser

ISBN: 3-527-30807-5

- Ligand-based tumor-targeting approaches (e.g. those based on high-affinity monoclonal antibody fragments) allow excellent ligand localization in the tumor environment, with tumor : organ ratios of >10 : 1 just a few hours after intravenous injection [2, 3].

Due to their accessibility, tumor-associated markers located around new blood vessels within the tumor mass (but absent in normal tissues) represent ideal targets for the development of better, more selective therapeutic strategies against cancer and other angiogenesis-related diseases.

This review focuses on current proteomic technologies, which can be used for the discovery of novel markers of angiogenesis, outlining their potential and the challenges which are still to be solved. The review does not describe how vascular targeting agents can be converted into imaging agents and therapeutic products; these issues have been described in recent reviews of our group [4, 5].

2.2 Vascular Targeting

2.2.1 Angiogenesis

Angiogenesis is a process through which new blood vessels develop from pre-existing vessels. Blood vessels are essentially composed of interconnecting, tube-forming endothelial cells (ECs), which are surrounded by an extracellular matrix (ECM) and associated cells such as smooth muscle cells and pericytes. Angiogenesis is essential for the development of the vascular network during embryogenesis. Under physiologic conditions, once the vascular network is established in the adult, ECs remain quiescent apart from certain locally and transiently physiological processes such as wound repair, the female menstrual cycle, hair growth and inflammation [6].

Hypoxia is an important stimulus for angiogenesis in physiological and pathological conditions [7], it triggers vessel growth by signaling through the hypoxia-inducible transcription factor (HIF). HIF is an $\alpha\beta$-heterodimer. HIF-β subunits are constitutive nuclear proteins, whereas HIF-α subunits are inducible by hypoxia [8]. Among three HIF-α isoforms, HIF-1α and HIF-2α appear closely related and are each able to interact with hypoxia-response elements (HREs) to induce transcriptional activity [9, 10]. The transcription of a broad range of genes including angiogenic growth factors like vascular endothelial growth factor (VEGF), but also erythropoietin, glucose transporters, and glycolytic enzymes is initiated by this hypoxic pathway [11].

Angiogenesis initiates with vasodilatation, a process involving nitric oxide. Vascular permeability increases in response to VEGF, thereby allowing extravasation of plasma proteins that lay down a provisional scaffold for migrating ECs. As a next step, ECs need to loosen interendothelial cell contacts in order to emigrate

from their resident site. ECM molecules are degraded by proteinases of the plasminogen activator, matrix metalloproteinase, chymase and heparanase family, which also activates or liberates growth factors (basic fibroblast growth factor, VEGF, and insulin-like growth factor-I) sequestered within the ECM [12]. Proteinases also alter the composition of the ECM, they expose new cryptic epitopes in ECM proteins (such as in collagen IV) or change their structure (fibrillar versus monomer collagen), which induces EC migration [13]. Once sufficient ECM degradation has taken place, proliferating ECs migrate and assemble as solid cords that subsequently form a lumen. The establishment of a functional vascular network further requires that nascent vessels mature into durable vessels. The association of pericytes and smooth muscle cells with newly formed vessels regulates EC proliferation, survival, migration, differentiation, vascular branching, blood flow, and vascular permeability [14].

Angiogenesis is a dynamic process tightly controlled by a large number of pro- and anti-angiogenic factors released by the surrounding tissues and the ECs. Under physiologic conditions, angiogenesis is highly regulated and may be induced by specific endogenous angiogenic molecules such as VEGF, transforming growth factor (TGF), angiogenin, tumor necrosis factor (TNF), basic fibroblast growth factor (bFGF), interleukin 8 (IL-8), or platelet-derived growth factor (PDGF) [7, 15]. Angiogenesis is also thought to be suppressed by a variety of inhibitory molecules such as interferon-*α*, angiostatin, endostatin, platelet factor 4, or thrombospondin-1 [16, 17]. When the tight regulation over stimulatory and inhibitory molecules of angiogenesis is upset, the control over normally quiescent capillary vasculature is lost and the growth of new blood vessels can sustain the progression of a disease.

2.2.2
Angiogenesis-related Disorders

Many diseases are driven by persistent unregulated angiogenesis [18]. In arthritis, new capillary blood vessels invade the joint and destroy cartilage. In diabetes, new capillaries in the retina invade the vitreous, bleed, and cause blindness. Ocular neovascularization is involved in many major eye diseases which may cause loss of vision [19]. In general, vasoproliferation can extend into nearly every mature ocular tissue and affect the cornea, iris, retina, and optic disk. The newly formed vessels are structurally weak and lack integrity, which results in hemorrhage, exudates, and accompanying fibrosis.

Much of our current understanding of angiogenesis stems from studies on tumoral angiogenesis. Many tumors in humans can persist *in situ* without neovascularization for months or even years, surviving as asymptomatic lesions which are rarely larger than 2 mm^3 [20–22]. In these "prevascular" tumors, the high rate of tumor proliferation is balanced by a high rate of tumor cell death, probably caused by the low level of blood perfusion. Only when a group of tumor cells switches to the angiogenic phenotype, by altering its balance between inhibitors and stimulators of angiogenesis, the tumor mass expands and overtakes the rate

of internal apoptosis. It is at this stage that the majority of tumors become clinically detectable and capable of invading the surrounding tissues and metastasizing [6, 22, 23]. This so-called angiogenic switch [23] serves the development of malignant tumors at multiple stages and is triggered by changes in the local microenvironment of a tumor.

When cell masses grow, they can change their microenvironment in a number of ways. They can alter local pH [24] and the concentration of nutrients and metabolites, increase interstitial pressure as a consequence of an expanding cellular volume and increased vascular permeability, and induce hypoxia by increasing local oxygen consumption [25]. The latter is particularly true of cells that are very metabolically active, such as those found in tumors. Furthermore, the oxygen delivery decreases parallel to an increase in diffusion distances between capillaries and the center of the expanding cell mass. This leads to a state of hypoxia in the microenvironment. Even though tumors are able to adapt their metabolism to survive under conditions of reduced oxygen availability by increasing glycolysis to maintain ATP production [26], they depend on adequate oxygen delivery. Mammalian cells are able to sense prolonged decreases in oxygen tension through a conserved hypoxic response pathway, which aims at increasing the local oxygen concentration by several actions. An important mediator in this process is the HIF complex, which increases the transcription of a broad range of genes including angiogenic growth factors like VEGF, but also erythropoietin, glucose transporters and glycolytic enzymes [11]. Furthermore, hypoxia promotes the stabilization of the VEGF mRNA, which is rapidly degraded under normoxia [27, 28]. Those two events contribute to VEGF up-regulation.

VEGF specifically stimulates the growth of ECs, which in turn produce many other non-specific angiogenic stimulators, including bFGF, acid fibroblast growth factor (aFGF), transforming growth factor α (TGFα), transforming growth factor β (TGFβ) and platelet-derived endothelial cell growth factor (PD-ECGF). In addition, tumor cells and endothelial cells excrete proteolytic enzymes, such as matrix metalloproteinases (MMPs) and serine proteases (e.g. tissue plasminogen activator and urokinase plasminogen activator), which break down the ECM. Cell adhesion molecules, such as integrins ($\alpha v\beta 3$ and $\alpha v\beta 5$) are expressed on the surface of ECs and mediate interactions with the ECM. Laminin, tenascin, and type IV collagen are also produced to provide new basement membrane components.

In addition, expression of VEGF receptors, Eph receptors, members of the ephrin family [29–31] and the Tie-1 and Tie-2 receptors [32, 33] are up-regulated. The latter interact with angiopoietins to signal capillary organization. Furthermore, the induction of cytokines and chemokines also recruits monocytes and leukocytes, producing local inflammatory reactions that aid in the process.

Up-regulation of angiogenic factors, however, is not sufficient in itself for a tumor to become angiogenic: negative regulators of vessel growth have to be down-regulated [34].

Thrombospondin was the first protein for which a down-regulation during tumorigenesis was demonstrated [35]. A number of endogenous angiogenesis inhibitors have been identified since. Some of them are cryptic regions or frag-

ments of proteins, which in their intact form do not possess any anti-angiogenic activity. Examples include a fragment of platelet factor 4 [36], or an antithrombin III fragment [37], of which both precursor proteins are members of the clotting/fibrinolytic pathways. Often, however, the intact proteins represent components of the extracellular matrix, as is the case for angiostatin (plasminogen) [38], endostatin (collagen XVIII) [39], and PEX (matrix metalloproteinase 2) [40].

The development of metastases is also dependent on angiogenesis [6]. First, metastatic cells are not shed from a primary tumor until the tumor has become neovascularized. Second, once metastatic cells have colonized a target organ, they will again only grow to a metastasis of clinically detectable size if they can induce neovascularization.

Tumor blood vessels are architecturally different from their normal counterparts – they are irregularly shaped, dilated, tortuous, and can have dead ends [41]. During physiological angiogenesis, new blood vessels rapidly mature and become stable, while tumor blood vessels fail to become quiescent. Consequently, the tumor vasculature develops unique characteristics and becomes distinct from the normal blood supply system. Characteristic molecular species (i.e. proteins) which are more abundant in tumoral blood vessels than in normal tissues may serve as markers for angiogenesis. The identification of such molecular markers has extended the field of anti-angiogenic therapy. The new field of vascular targeting features the use of molecular vehicles that selectively localize in the neovasculature and that allow the targeted delivery of therapeutic agents.

2.2.3 Markers of Angiogenesis

Markers of angiogenesis have to be specific, accessible, and abundant if they are to serve as targets for therapeutic or diagnostic intervention. To date, only few good quality markers of angiogenesis located either on ECs or in the modified ECM are known. Most existing candidate markers are also expressed in some normal tissues, thus limiting their usefulness. Systematic *ex vivo* investigations of tumor endothelial structures using proteomic techniques [42, 43], biopanning of phage display libraries [44–50], or transcriptomic techniques, such as serial analysis of gene expression [51], are revealing new candidate tumor endothelial markers. Their validation, however, requires the generation of specific monoclonal antibodies, a comprehensive immunohistochemical analysis, and quantitative biodistribution studies in animal models of angiogenesis-related diseases. Markers for angiogenesis located on the cell surface of tumor endothelial cells are the integrins $\alpha v\beta 3$ [52, 53], endoglin [54], VEGF and VEGF receptor complex [55], prostate-specific membrane antigen (PSMA) [56], aminopeptidase N [57], CD62E [58], and also phosphatidylserine phospholipids [59]. By using serial analysis of gene expression, novel genes (TEM1, TEM5, and TEM8), encoding tumor endothelial markers have been found [51, 60]. Other markers of angiogenesis are part of the modified ECM.

Probably one of the most selective markers of angiogenesis, the extra domain B of fibronectin (ED-B), is located in the modified ECM within the tumor environ-

ment [48, 49]. The ED-B domain is highly conserved in different species and can be inserted into the fibronectin mRNA by a mechanism of alternative splicing, which only occurs in transformed cells, malignancies, and under strictly controlled physiologic conditions. With some very rare exceptions (uterus, ovaries), ED-B is undetectable in normal tissues, but exhibits a much greater expression in fetal and tumor tissues as well as during wound healing [50]. Its selective accumulation around neovascular structures has been demonstrated in studies on many different tumor types, in particular on invasive ductal carcinoma and brain tumors [50], as well as in ocular angiogenesis [61]. The ED-B domain of fibronectin has been extensively validated as a target for antitumor intervention in animal models by biodistribution analysis *in vivo* [2, 62–66] and in patients with cancer [3, 67]. Another interesting marker of the tumor-associated ECM is the large tenascin C isoform [68].

2.2.4 Ligand-based Vascular Targeting

The markers described above may serve as targets for the molecular imaging and therapy of tumors and other diseases characterized by vascular proliferation. Therapeutic application of vascular targeting includes the targeted delivery of bioactive molecules (drugs, cytokines, procoagulant factors, photosensitizers, radionuclides, etc.) to the tumor environment by means of binding molecules (e.g. human antibodies) specific for tumor-associated markers [69]. The molecular targeting of tumor endothelial cells features several advantages compared with the targeting of tumor cells. Vascular endothelial cells are freely accessible from the blood, while barriers separate tumor cells from the bloodstream. In principle, antivascular therapies could be used for the treatment of different types of solid tumors, since different aggressive cancers often share upregulated markers of angiogenesis. Endothelial cells, unlike cancer cells, have stable genomes and are less prone to mutate [6, 70], therefore, antigen loss or resistance to therapy is less likely to occur when a therapeutic agent is directed against the vasculature. Large numbers of tumor cells depend heavily upon a few vessel-forming ECs for their blood supply [71, 72]. Thus, damaging the tumor vasculature would have a vast amplification or bystander effect by causing an avalanche of tumor cell death.

To achieve selective targeting of the tumoral vasculature, antibodies, small molecular compounds, or other molecules must be directed towards abnormally expressed or overexpressed proteins, carbohydrates, or lipids, so-called markers of angiogenesis. Membrane proteins on tumoral ECs, if present in sufficient abundance and if specific enough, are likely to be ideal targets for diagnostic or therapeutic intervention because of their accessibility from the bloodstream. The following section of this review focuses on the proteomic technologies available to study differentially expressed membrane proteins on ECs. *In vitro* and *in vivo* model systems are presented, which have been applied for the investigation of tumoral and normal vasculature.

2.3 Technologies for the Quantitation of Membrane Proteins in Different Cell Types

Membrane proteins are either situated at the interface between the cell and the surrounding environment or at the interface of subcellular compartments, and perform key functions such as cell-to-cell recognition and transport of ions and solutes, as well as acting as receptors for forwarding the diverse signals that reach the cell or subcellular compartment. Since membrane proteins are so important for a cell's life, they are often suitable targets for pharmaceutical intervention and, indeed, more than two-thirds of the known protein targets for drugs are membrane proteins [73].

The expression of membrane proteins can alter as a consequence of disease, potentially providing targets for the selective delivery of pharmaceuticals.

The investigation and quantitation of membrane proteins remains a challenging task for several reasons. First, membrane proteins are typically low-abundance proteins. The dynamic range in protein abundance (copy numbers per cell or tissue) is believed to cover up to nine orders of magnitude, and thus the analysis of low-abundance proteins is very difficult using any technique [74]. Second, membrane proteins are hydrophobic proteins designed to be soluble in lipid bilayers and consequently are difficult to solubilize in aqueous media [75].

The following sections outline some methods that have been applied for the investigation of membrane proteins.

2.3.1 Gel-based Quantitative Profiling of Membrane Proteins

2.3.1.1 Two-dimensional Polyacrylamide Gel Electrophoresis

Two-dimensional polyacrylamide gel electrophoresis (2-D PAGE) is a powerful method for the analysis of complex protein mixtures extracted from cells, tissues, or other biological samples. This technique, which was first introduced by O'Farrell [76] and Klose [77], sorts proteins according to two independent properties in two discrete steps: the first dimension step, isoelectric focusing (IEF), separates proteins according to their isoelectric points (pI); the second dimension step, sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE), separates proteins according to their molecular weights (MW).

Two-dimensional PAGE is a core technique in proteome analysis, which usually includes sample preparation, 2-D PAGE, post-separation image analysis of the stained gel, and protein characterization by mass spectrometry. Technical innovations within the last 20 years have made 2-D PAGE a widely applied analytical tool. The replacement of classical first-dimension carrier ampholyte pH gradients by well-defined immobilized pH gradients (IPG) resulted in higher resolution, improved interlaboratory reproducibility, higher protein loading capacity, and an extended basic pH limit for 2-D PAGE [78]. Furthermore, microanalytical techniques were developed, which allowed the identification of proteins at the amounts available from 2-D PAGE. These microanalytical techniques were first

Edman sequencing [79] and, more recently, mass spectrometry, which has greatly increased the sensitivity and throughput of the protein identification [80, 81].

Software is available that allows the routine computerized evaluation and semi-quantitation of the highly complex two-dimensional patterns. Data about entire genomes (or substantial fractions thereof) are available for a number of organisms, allowing rapid identification of the gene encoding a protein separated by 2-D PAGE. The World Wide Web provides simple, direct access to spot pattern databases for the comparison of electrophoresis results and to genome sequence databases for assignment of sequence information.

However, despite all these upgrades, there are still limitations of 2-D PAGE, which are linked to the chemical diversity of proteins in a cell or tissue and to their different abundance.

The comparison between the number of actin molecules (about 10^8 molecules per cell) and the number of some cellular receptors (ca. 100–1000 molecules per cell) present in a cell, reveals a dynamic range of up to 1 000 000 between the most-abundant and least-abundant proteins [74]. In contrast, 2-D PAGE can only display differences in protein concentration in the range of 100–10 000. Since membrane proteins are typically of low abundance, they are often underrepresented on 2-D PAGE gels. Furthermore, complex protein mixtures lead to comigrating proteins which compromise quantitative analysis based on the assumption that one protein is present per spot [82]. To overcome this problem, highly expressed proteins can be purified away by sample prefractionation or enrichment strategies. Membrane proteins can be enriched by cellular fractionation [83], but a difficult solubilization during sample preparation for 2-D PAGE often limits the applicability of this technology. Furthermore, membrane proteins tend to precipitate during IEF, when they concentrate at their pI.

IEF sample buffer for 2-D PAGE using an IPG usually contains a chaotropic agent to solubilize and denature proteins, a non-ionic or zwitterionic detergent to solubilize hydrophobic proteins, and a reducing agent, which cleaves disulfide bonds to allow a more complete protein unfolding. Additional components include a carrier ampholyte mixture, which can improve separations and sample solubility, and a tracking dye. The strong anionic detergent SDS, which is known to solubilize almost any protein, interferes with the IEF step in the first dimension of 2-D PAGE and can only be used in a concentration below 0.25% [84, 85]. IEF requires the maintenance of the intrinsic surface charge of proteins. SDS, which binds highly to proteins, produces a strong charge shift, impairing subsequent IEF. Salts that help to solubilize proteins lead to horizontal streaking in the 2-D gels and therefore have to be avoided.

Several approaches have been taken to increase the representation of membrane proteins on 2-D gels. Molloy and colleagues extracted membrane proteins of a whole cell lysate from *Escherichia coli* by differential solubilization [86]. In a three-step sequential solubilization protocol, membrane proteins are separated from other cellular proteins by their insolubility in solutions conventionally used for IEF. Eleven membrane proteins could be identified in the final membrane-rich pellet. Another approach attempts to isolate membrane proteins from *E. coli* by organic solvent ex-

traction prior to 2-D PAGE [87]. The use of an organic solvent extraction has selectively enriched hydrophobic proteins, which could be resolubilized with denaturing conditions to permit 2-D PAGE. However, no highly hydrophobic protein typical of an *E. coli* cytoplasmic membrane was identified with this approach.

Other groups have aimed at optimizing the composition of the sample buffer for a more efficient recovery of membrane proteins in the first dimension. In a recent study, Luche and colleagues compared the efficiency in membrane protein solubilization of several non-ionic, commercially available detergents [88]. The non-ionic detergents dodecyl maltoside and decaethylene glycol mono hexadecyl ether were the most efficient membrane protein solubilizers. Other groups have also developed new detergents to achieve a better representation of membrane proteins on 2-D gels [89, 90].

However, membrane protein solubility problems are still encountered with 2-D PAGE. As a consequence, alternative approaches rely on one-dimensional SDS-PAGE for the separation of membrane proteins, replacing the IEF by another separation technique in the first dimension.

2.3.1.2 **Combination of Chromatography and 1-D SDS-PAGE**

The solubility problem of membrane proteins during the IEF step in 2-D PAGE stimulated the search for alternative separation methods, orthogonal to SDS-PAGE. Complex protein mixtures can be separated according to several parameters, including the retention on a chromatographic column and/or the molecular weight. The separation of individual fractions after chromatography by means of 1-D SDS-PAGE generates two- or multi-dimensional patterns and facilitates the resolution of different proteins.

A wide variety of high-performance liquid chromatography (HPLC) systems have been employed, originally for the purification of membrane proteins [91, 92]. These include size-exclusion HPLC, ion-exchange HPLC, bioaffinity chromatography, reverse phase HPLC (RP-HPLC), and hydroxyapatite HPLC (HAP-HPLC) with SDS. Since those chromatographic methods allow the separation of contaminants from membrane proteins, the question arises whether one can also part different membrane proteins contained in a membrane extract by applying chromatography in the first dimension.

Horigome and colleagues combined ceramic HAP-HPLC with 1-D SDS-PAGE for the investigation of membrane proteins isolated from rat erythrocyte membranes and rat liver microsomes [93]. The HAP-HPLC was performed with a buffer system, which contained 1% of SDS and sodium phosphate up to a concentration of 0.5 M. In another study, membrane proteins from rat liver rough microsomes were efficiently resolved with a protein recovery of more than 90% by HAP-HPLC, using 1% sodium cholate as detergent [94]. Hydroxyapatite chromatography was introduced in 1956 by Tiselius and Hjerten [95]. In HAP-HPLC, biomolecules are separated according to their different interactions with hydroxyapatite, whose molecular formula is $Ca_{10}(PO_4)_6(OH)_2$. Positively charged ammonium

groups (e.g. side chains of lysine residues) are attracted by the phosphate groups on the column and repelled by the calcium ions; the situation is the opposite for carboxylic acids [96–98].

HAP-HPLC allows the use of strong detergents, which is very advantageous when working with membrane proteins. However, the same proteins are often found in more than one fraction, thus hampering the overall resolution of the two-dimensional separation process. Excellent resolution is imperative if gel-based methods are to be used for the comparison of protein abundance in different samples.

In 1988, a group reported the development of the chromatophoresis process, which couples RP-HPLC to SDS-PAGE in a real-time automated system [99, 100]. The real potential of this method still remains to be demonstrated.

Using SDS-PAGE for the one-dimensional separation of membrane proteins and microcapillary liquid chromatography-electrospray ionization tandem mass-spectrometry (μLC-ESI-MS/MS) for the analysis of peptides generated by digesting the protein migrating to a particular zone of the gel, Simpson and co-workers identified 284 proteins, including 92 membrane proteins [101]. Although this method is suitable for cataloguing proteins contained in membrane fractions, it is inherently not quantitative and therefore not suitable for the detection of differences in the membrane protein profile of cells representing different states.

Sharov and colleagues developed a two-dimensional method for the characterization of post-translational modifications of rabbit sarco/endoplasmic reticulum Ca-ATPases (SERCA) using a combination of RP-HPLC and 1-D SDS-PAGE followed by liquid chromatography tandem mass spectrometry (LC-MS/MS) analysis of in-gel tryptic digests [102]. The SERCA, which belong to a group of large hydrophobic proteins, are likely to be underrepresented on 2-D PAGE because of reasons outlined above. Applying this method, SERCA proteins were successfully separated from other proteins present in native sarcoplasmic reticulum vesicles. The degree of artificially induced post-translational modifications of SERCA was investigated by tandem mass spectrometry analysis. The authors claim that their method might be applicable to the proteomic analysis of membrane proteins of whole tissue or selected organelles, especially when post-translational modifications need to be monitored.

In general, the combination of (a) fractionation of subcellular organelles, (b) chromatography in the presence of SDS, and (c) 1-D SDS-PAGE appears to be a robust avenue for the comparison of relative protein abundance in different cells/tissues.

Automation of proteome analysis will be important in order to increase inter-laboratory reproducibility and also to make the process less labor-intensive. Furthermore, automation would increase the overall throughput of sample analysis.

2.3.2 Gel-independent Quantitative Profiling of Membrane Proteins

2.3.2.1 Mass Spectrometry-based Methods

Even when protein bands (or spots) are well resolved in PAGE, the routine use of this technique for the comparison of the relative abundance of proteins in different samples is limited to the detection of big changes (e.g. >3-fold change in relative abundance). Special techniques (such as biosynthetic or post-separation isotope labeling, 2-D difference fluorescence gel electrophoresis) are not always compatible with the requirements for the discovery of novel markers of angiogenesis (e.g. *in vivo* protein biotinylation), but have been used with success for the studies of cells cultured in different conditions [103]. Furthermore, methodologies such as 2-D PAGE are labor-intensive and do not lend themselves easily to the analysis and comparison of several dozens of samples.

Continuous advances in the field of biological mass spectrometry have now made it possible to routinely identify proteins from the corresponding endoproteolytic peptides, at total protein amounts in the femtomole range. The combined use of multidimensional liquid chromatography (LC) and tandem mass spectrometry (MS/MS) has made it possible to identify hundreds of proteins in the same sample containing tryptic peptides [104]. In a large-scale analysis of the yeast proteome, more than 100 membrane proteins were identified [105]. More recently, a modified methodology has been developed for the application of mass spectrometry to the study of membrane proteins [106]. A combination of membrane sheets enrichment at high pH, followed by proteinase K treatment and LC/MS analysis, has allowed the identification of >400 membrane proteins in a rat brain homogenate. All these methodologies, however, do not yield information about relative protein quantity at present, and cannot be used for the comparison of membrane protein abundance in different complex specimens.

In 1999, Aebersold and co-workers introduced the concept of isotope-coded affinity tags ("ICAT") for the stable non-radioactive isotopic labeling of proteins, compatible with protein identification and relative quantitation in different biological specimens [107]. In its original implementation, the ICAT technology consists in the biotinylation of cysteine residues in proteins with reactive derivatives of biotin, carrying a linker arm with hydrogen or deuterium atoms. These "light" and "heavy" biotin derivatives serve a dual purpose. First, they reduce the complexity of tryptic peptides to be analyzed in a gel-free MS experiment (they can be purified on affinity resins; only few tryptic peptides in a protein contain a cysteine residue). Second, the labeling of peptides from two different samples with a light or heavy tag allows the use of LC-MS/MS methodologies for the relative comparison of protein abundance in the two samples. The relative protein abundance is, in fact, reflected in the relative intensity of the MS signals of the corresponding biotinylated peptides, which are separated in the m/z axis by the number of daltons corresponding to the number of atoms which are either hydrogen or deuterium in the biotin derivative tag. A number of modified ICAT implementations have been developed in the last few years. They include the use of ICAT for the relative

quantitation of spots in 2-D gels [108], the use of solid-phase isotope tagging [109], and the use of special chemical procedures for the analysis of post-translational modifications, such as glycosylation and phosphorylation [110].

The ICAT technology has recently been used for the quantitative profiling of differentiation-induced microsomal proteins. The method was used to identify and determine the ratios of abundance of each of 491 proteins contained in the microsomal fractions of naive and *in vitro* differentiated human myeloid leukemia cells [111]. In this study, the authors recognize that a subset of proteins that lack cysteine residues, very low-abundance proteins, and very hydrophobic proteins would not be analyzed with this technique.

The substitution of thiol-reactive ICAT reagent with a similar ICAT reagent, capable of reaction with primary amino groups (e.g. N-terminus and side chains of lysines, as would be desirable for *in vivo* biotinylation reactions) is not straightforward. The higher number of lysine residues in proteins (compared with cysteines) may lead to incomplete chemical coupling, and to a distribution of patterns of (incomplete) lysine labeling, thus complicating the downstream LC-MS/MS analysis.

Stimulated by the work of Schrader, Schulz-Knappe and colleagues [112, 113], which have routinely used the orthogonal combination of chromatography and matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF-MS) for the relative quantitation of peptides in biological fluids (e.g. sera and cerebrospinal fluids), we have investigated the extent to which MALDI-TOF-MS can be used for the relative quantitation of peptides, obtained by proteolytic degradation of proteins. We tested the reproducibility of the correlation between MALDI-TOF-MS signal intensity and sample amount, using a dilution series of a model peptide (amounts: 0.3, 3, and 30 pmol). Two different sample target plates were evalu-

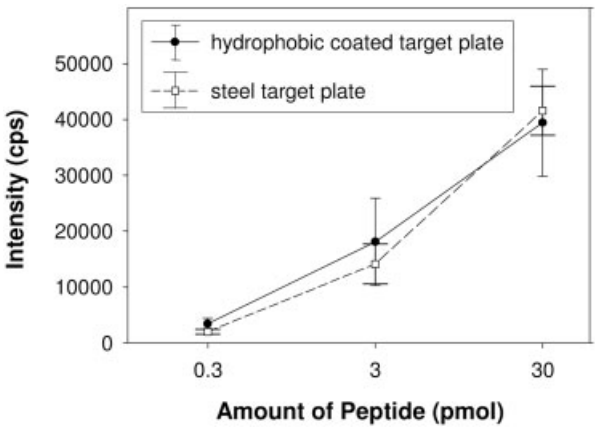


Figure 2.1 Correlation between MS signal intensity relative to the amount of peptide. Samples of 0.3, 3 and 30 pmol of a test peptide were measured with MALDI-TOF-MS. For each peptide amount tested, the spectra from 20 replicates were averaged and the resulting signal intensity (counts per second, cps) and standard deviation were plotted versus the corresponding peptide amount.

ated: a steel target plate and a hydrophobic coated target plate. The test peptide was mixed with a saturated solution of the matrix (alpha-cyano-4-hydroxycinnamic acid). The measurements were performed with the Voyager™ Elite workstation from PerSeptive Biosystems (Framingham, MA, USA) in the reflectron (acceleration 20 kV) and the delayed extraction mode. Figure 2.1 illustrates the results of measuring 20 replicates of each peptide amount: the mean values and the corresponding standard deviations are plotted relative to the amount of peptide. For both target plates a clear correlation between MS intensity and peptide amount can be observed, allowing a discrimination of peptides which differ by at least a factor 4–5 in relative amount.

However, the results can be dramatically improved by including an internal control peptide for the normalization of signal intensity (data not shown). Comforted by this observation, we have used a combination of membrane protein biotinylation, purification on streptavidin resin and LC-MALDI-TOF analysis for the determination of the relative abundance of membrane proteins in human umbilical vein endothelial cells, cultured in normoxic (20% O_2) and hypoxic (2% O_2) conditions [114].

It is almost certain that modern mass spectrometric procedures and instrumentation (such as the use of Fourier transform ion cyclotron (FT-ICR) and MALDI-TOF time-of-flight (MALDI-TOF-TOF) spectrometers [103]), which offer unprecedented resolution and sensitivity, will contribute to the increased use of MS-based methods for gel-free proteomic analysis. However, the study of membrane proteins will also require improved methodologies for the chemical modification and recovery of peptides from these low-abundance, hydrophobic proteins.

2.3.2.2 **Ligand-based Methods**

Traditionally, the first markers of angiogenesis have been discovered either by limited proteolysis of purified protein preparations [48], or by animal immunization with biological samples derived from tumors, followed by an extensive immunohistochemical analysis of the resulting hybridomas [115, 116].

The introduction of recombinant antibody technologies, and in particular of antibody phage technology [117], has greatly facilitated the production of good-quality monoclonal antibodies without immunization. These technologies are particularly efficient when pure antigen preparations are available [118], but antibodies have also been generated from phage libraries against "difficult" antigens [46].

It is difficult to imagine that antibody-based chips may facilitate the study of the relative abundance of membrane proteins in different biological specimens [119]. Nonetheless, if larger public-domain collections of monoclonal antibodies become available in the future, it should be possible to use fluorescence-activated cell sorters (FACS) and/or immunohistochemistry with tissue arrays [120] for the relative quantitation of membrane proteins in different cells/tissues.

Ruoslahti, Pasqualini and co-workers have pioneered the *in vivo* biopanning of peptide phage libraries, in an attempt to identify binding specificities against different vascular addresses in different tissues and/or tumors [121, 122]. Among others, peptides specific to integrins and to CD13 were identified with this procedure. However, the real potential of this technology remains to be demonstrated,

considering that the use of peptides on tissue sections is often less efficient than the use of antibodies (which normally display a higher affinity for the antigen), and in the absence of quantitative biodistribution studies and clinical studies with purified preparations of the vascular-targeting peptides.

2.4 Model Systems for the Identification of Vascular Targets

2.4.1 *In vitro* Model Systems for the Study of Gene Expression in Response to Environmental Changes

In spite of the many differences that can be observed between different types of cancer, some common features are characteristic for aggressive, rapidly growing solid tumors. In addition to features such as the genomic instability of tumor cells and their deranged proliferation and attachment behavior, the tumor environment is also characterized by the presence of new blood vessels and (in spite of them) by the insufficient blood perfusion of the tumor mass, leading to hypoxia, serum starvation, and pH changes.

Our group and others have tried to mimic the tumor environment *in vitro*, using cell cultures (primary endothelial cells, primary cultures of fibroblasts, etc.) and studying the changes in gene expression as a response to environmental changes (pH, serum starvation, hypoxia, etc.). In our experience, the combined use of transcriptomic analysis (using the Affymetrix gene chip system) and proteomic investigations (by 2-D PAGE and MS) is often beneficial, as the two technologies are often complementary in identifying candidate genes whose patterns of expression are regulated by environmental changes [114].

The study of gene expression in endothelial cells isolated from tumors and normal tissues may allow a closer analysis of the patterns of gene expression in tumor vascular structures. The quality of endothelial cell purification is crucial for these types of studies, both in terms of separation from other cell types and in terms of speed, preventing post-separation changes in the abundance of mRNAs and proteins. Most studies performed so far have studied levels of gene expression using transcriptomic methods, such as cDNA subtractive hybridization procedures (e.g. [123]), serial analysis of gene expression (SAGE) [60] and gene chip-based methods. A group at Genentech has crossed gene expression data obtained by comparing colorectal cancer and normal mucosa with a database of genes known to be expressed in endothelial cells, thus identifying putative markers with differential expression in tumor endothelial cells compared to endothelial cells in normal tissues [124–126]. Interestingly, the most attractive candidate resulting from this study (stanniocalcin) is also one of the most strongly up-regulated genes in *in vitro* model experiments when shifting cell culture pH to an acidic value [127].

In principle, proteomic investigations should be possible if sufficient amounts of endothelial cells and the associated ECM components can be recovered from tumors, either by cell purification [128] or by laser capture microdissection [129].

The experimental approaches described in this chapter can, at best, suggest candidate markers of angiogenesis. An experimental confirmation, however, requires the generation of specific monoclonal antibodies and an extensive immunohistochemical analysis of expression patterns in normal and pathological specimens.

2.4.2 *In vivo* Model Systems for the Identification of Vascular Targets

In principle, the most direct way to assess differences in protein abundance between the tumor endothelium and the normal endothelium would consist in the *in vivo* labeling of vascular structures, followed by rapid recovery and comparative proteomic analysis of the proteins in the two samples.

The group of Jan Schnitzer has pioneered the use of colloidal silica for the *in vivo* coating of vascular structures in tumors and in normal organs [130, 131]. This physical modification allows the recovery (by centrifugation and fractionation) of silica-coated structures (luminal cell plasma membranes and caveolae of the endothelium), providing ideal material for proteomic investigations, for example by immunization [132] or by 2-D PAGE.

De La Fuente et al. have described the artificial perfusion of lungs isolated from normal and hyperoxic rats with sulfo-*N*-hydroxysuccinimide ester of biotin LC (sulfo-NHS-LC-biotin) [133]. After SDS-PAGE, the biotinylated proteins were visualized using a chemiluminescence substrate for the streptavidin/horseradish peroxidase conjugate, outlining differences in rats exposed to hyperoxia for 48–60 hours.

Our group is using the terminal perfusion of tumor-bearing mice with sulfo-NHS-LC-biotin solution as an avenue for the *in vivo* covalent modification of amine-containing phospholipids and proteins, which are accessible to the reagent during the perfusion (unpublished). The experimental methodology is depicted in Figure 2.2. After anesthesia, mice are first perfused with saline solution, to remove circulating cells, proteins and other primary amine-containing compounds. A few minutes later, perfusion is continued with an aqueous solution of sulfo-NHS-LC-biotin, followed by a primary amine (e.g. Tris buffer) to quench unreacted ester derivatives of biotin.

This methodology leads to reliable and efficient labeling of accessible structures (mainly vascular structures) *in vivo*, and is ideally suited for proteomic investigations.

The resulting biotinylated proteins (or the corresponding peptides generated by endoproteolytic cleavage) can be purified on streptavidin-coated resins in the presence of SDS. However, the choice of detergent and of purification protocol depends on the experimental strategy chosen for proteomic investigations. At present, it is not clear which approach among (a) 2-D PAGE, (b) 1-D SDS-PAGE with chromatographic pre-fractionation, or (c) gel-free spectrometric comparative analysis of biotinylated peptides is the most suitable method to perform a comprehensive proteomic investigation of the differences in biotinylation in various organs and tumors. In principle, the *in vivo* biotinylation method presents several attrac-

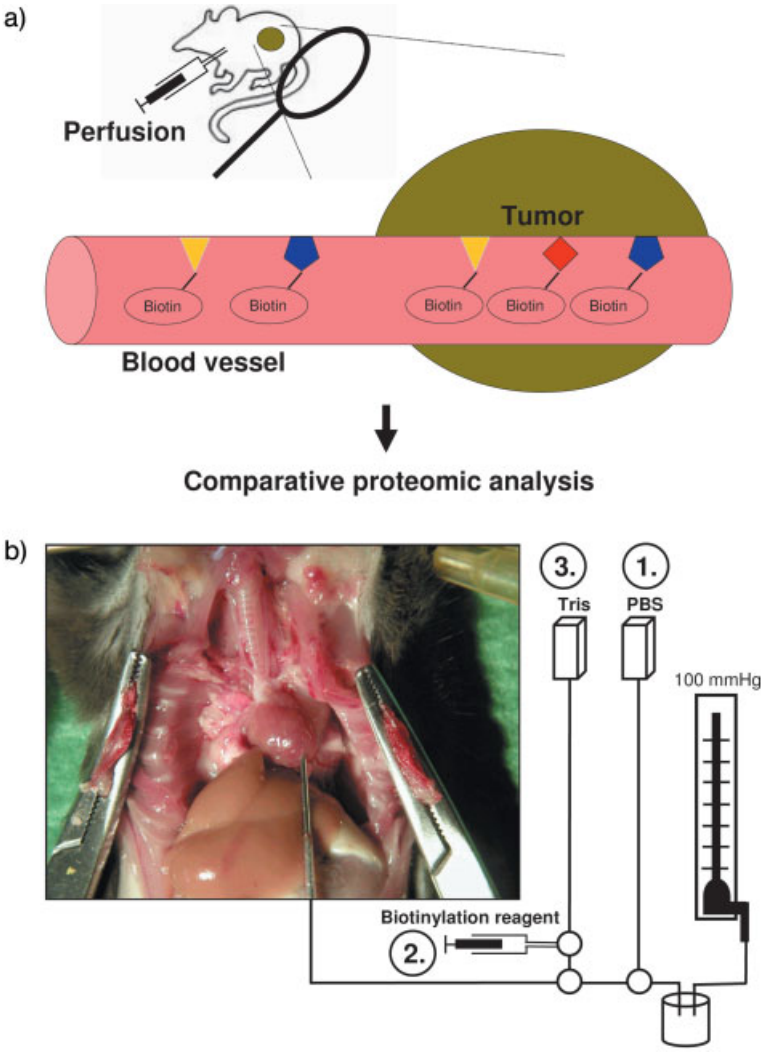


Figure 2.2 Terminal perfusion of tumor-bearing mice with Sulfo-NHS-LC-biotin. (a) The principle of the terminal perfusion and *in vivo* biotinylation. Tumor-bearing rodents are perfused with a sulfo-NHS-LC-biotin solution. Accessible structures (e.g. proteins) carrying a primary amino group are biotinylated followed by the isolation and purification of the biotinylated proteins. A differential proteome analysis between non-tumoral and tumoral samples aims at identifying accessible, tumor-associated markers from the bloodstream. (b) The typical *in vivo* biotinylation protocol. After anesthesia, the mice are first perfused with saline solution to remove circulating cells, proteins and other primary amine containing compounds. The perfusion is continued with an aqueous solution of sulfo-NHS-LC-biotin, followed by a primary amine (e.g. Tris buffer) to quench unreacted ester derivatives of biotin. The perfusion is carried out at a constant pressure of 100 mmHg.

tive features, as it allows a direct investigation of those accessible targets that are likely to be amenable to targeted anticancer imaging and therapeutic strategies. As protein biotinylation lends itself not only to purification strategies (special precautions for elution must be chosen, considering the high-affinity interaction with streptavidin!), but also to biochemical analysis (e.g. by blotting or microscopic analysis with streptavidin-based detection reagents), it is possible to monitor the efficiency of the biotinylation reaction in various organs prior to proteomic analysis.

2.5 Conclusions

The area of ligand-based vascular targeting is developing rapidly, and promises to deliver novel diagnostic and therapeutic agents for cancer and a variety of other diseases, including blinding ocular disorders, chronic inflammatory conditions, atherosclerosis, and amyloidosis. A number of complementary technologies will contribute to developments in this field, including methods for quantitative protein profiling, the rapid generation of specific ligands (e.g. recombinant antibody technology and combinatorial chemistry), and the availability of suitable animal models of pathology.

In vivo biotinylation procedures under terminal perfusion of animal models appears to be one of the most exciting avenues for the unbiased comparative analysis of proteins in accessible vascular structures. The real potential of this approach, however, will crucially rely on advances in methodology for the separation, identification, and quantitation of membrane proteins and ECM proteins.

The results of the ongoing clinical studies with biopharmaceuticals directed against the markers of angiogenesis already validated (such as the ED-B domain of fibronectin, PSMA, and the integrin $\alpha v\beta 3$) will shed light on the potential and pitfalls of ligand-based approaches to molecular vascular targeting, thus stimulating further research for the discovery of better vascular targets.

2.6 Acknowledgements

This work is supported, in part, by the Gebert-Rüf Stiftung, the Swiss National Science Foundation and the Bundesamt für Bildung und Wissenschaft (EC Projects to D. N.).

The authors would like to express their gratitude to the Proteinservice-Lab Hönggerberg of the ETH Zurich and to the Functional Genomics Center Zurich for providing technical support.

Simone Scheurer is a recipient of a fellowship from the Roche Research Foundation.

2.7 References

1 K. Bosslet, R. Straub, M. Blumrich et al. *Cancer Res.* **1998**, *58*, 1195.
2 L. Borsi, E. Balza, M. Bestagno et al. *Int. J. Cancer* **2002**, *102*, 75.
3 M. Santimaria, G. Moscatelli, G.L. Viale et al. *Clin. Cancer Res.* **2003**, *9*, 571.
4 C. Halin, L. Zardi, D. Neri. *News Physiol. Sci.* **2001**, *16*, 191.
5 C. Halin, D. Neri. *Crit. Rev. Ther. Drug Carrier Syst.* **2001**, *18*, 299.
6 J. Folkman. *Nature Med.* **1995**, *1*, 27.
7 P. Carmeliet. *Nature Med.* **2000**, *6*, 389.
8 C.W. Pugh, P.J. Ratcliffe. *Nature Med.* **2003**, *9*, 677.
9 H. Tian, S.L. McKnight, D.W. Russell. *Genes Dev.* **1997**, *11*, 72.
10 M.S. Wiesener, H. Turley, W.E. Allen et al. *Blood* **1998**, *92*, 2260.
11 G.L. Semenza. *Annu. Rev. Cell Dev. Biol.* **1999**, *15*, 551.
12 L.M. Coussens, W.W. Raymond, G. Bergers et al. *Genes Dev.* **1999**, *13*, 1382
13 M. Hangai, N. Kitaya, J. Xu et al. *Am. J. Pathol.* **2002**, *161*, 1429.
14 R.K. Jain. *Nature Med.* **2003**, *9*, 685.
15 J. Folkman, *N. Engl. J. Med.* **1995**, *333*, 1757.
16 F. Bussolino, A. Mantovani, G. Persico. *Trends Biochem. Sci.* **1997**, *22*, 251.
17 R.S. Kerbel. *Carcinogenesis* **2000**, *21*, 505.
18 J. Folkman, Y. Shing. *J. Biol. Chem.* **1992**, *267*, 10931.
19 P. Lee, C.C. Wang, A.P. Adamis. *Surv. Ophthalmol.* **1998**, *43*, 245.
20 J. Folkman. *N. Engl. J. Med.* **1971**, *285*, 1182.
21 J. Folkman. *Ann. Surg.* **1972**, *175*, 409.
22 C.H. Blood, B.R. Zetter. *Biochim. Biophys. Acta* **1990**, *1032*, 89.
23 D. Hanahan. *Nature Med.* **1998**, *4*, 13.
24 P. Vaupel, P. Okunieff, L.J. Neuringer. *Adv. Exp. Med. Biol.* **1989**, *248*, 835.
25 F.J. Giordano, R.S. Johnson. *Curr. Opin. Genet. Dev.* **2001**, *11*, 35.
26 G.L. Semenza. *Trends Mol. Med.* **2002**, *8*, S62.
27 A. Damert, M. Machein, G. Breier et al. *Cancer Res.* **1997**, *57*, 3860.
28 J.A. Dibbens, D.L. Miller, A. Damert, W. Risau, M.A. Vadas, G.J. Goodall. *Mol. Biol. Cell* **1999**, *10*, 907.
29 D.J. Easty, S.P. Hill, M.Y. Hsu et al. *Int. J. Cancer* **1999**, *84*, 494.
30 P.M. Helbling, D.M. Saulnier, A.W. Brandli. *Development* **2000**, *127*, 269.
31 N. Holder, R. Klein. *Development* **1999**, *126*, 2033.
32 M.J. McCarthy, M. Crowther, P.R. Bell, N.P. Brindle. *FEBS Lett.* **1998**, *423*, 334.
33 G. Siemeister, M. Schirner, K. Weindel et al. *Cancer Res.* **1999**, *59*, 3185.
34 J. Folkman. *Semin. Med. Beth Israel Hosp. Boston* **1995**, *333*, 1757.
35 M. Tenan, G. Fulci, M. Albertoni et al. *J. Exp. Med.* **2000**, *191*, 1789.
36 S.K. Gupta, T. Hassel, J.P. Singh. *Proc. Natl Acad. Sci. USA* **1995**, *92*, 7799.
37 M.S. O'Reilly, S. Pirie-Shepherd, W.S. Lane, J. Folkman. *Science* **1999**, *285*, 1926.
38 M.S. O'Reilly, L. Holmgren, Y. Shing et al. *Cell* **1994**, *79*, 315.
39 M.S. O'Reilly, T. Boehm, Y. Shing et al. *Cell* **1997**, *88*, 277.
40 P.C. Brooks, S. Silletti, T.L. von Schalscha, M. Friedlander, D.A. Cheresh. *Cell* **1998**, *92*, 391.
41 J. Denekamp. *Br. J. Cancer* **1982**, *45*, 136.
42 J.E. Schnitzer. *N. Engl. J. Med.* **1998**, *339*, 472.
43 J.E. Schnitzer, P. Oh, B.S. Jacobson, A.M. Dvorak. *Proc. Natl Acad. Sci. USA* **1995**, *92*, 1759.
44 E. Ruoslahti. *Semin. Cancer Biol.* **2000**, *10*, 435.
45 M. Kolonin, R. Pasqualini, W. Arap. *Curr. Opin. Chem. Biol.* **2001**, *5*, 308.
46 H.R. Hoogenboom, J.T. Lutgerink, M.M. Pelsers et al. *Eur. J. Biochem.* **1999**, *260*, 774.
47 M. Rousch, J.T. Lutgerink, J. Coote, A. de Bruine, J.W. Arends, H.R. Hoogenboom. *Br. J. Pharmacol.* **1998**, *125*, 5.
48 L. Zardi, B. Carnemolla, A. Siri et al. *EMBO J.* **1987**, *6*, 2337.

49 B. Carnemolla, E. Balza, A. Siri et al. *J. Cell Biol.* **1989**, *108*, 1139.

50 P. Castellani, G. Viale, A. Dorcaratto et al. *Int. J. Cancer* **1994**, *59*, 612.

51 B. St Croix, C. Rago, V. Velculescu et al. *Science* **2000**, *289*, 1197.

52 D.A. Sipkins, D.A. Cheresh, M.R. Kazemi, L. Nevin, M.D. Bednarski, K.C. Li, *Nature Med.* **1998**, *4*, 623.

53 W. Arap, R. Pasqualini, E. Ruoslahti, *Science* **1998**, *279*, 377.

54 F.J. Burrows, E.J. Derbyshire, P.L. Tazzari et al. *Clin. Cancer Res.* **1995**, *1*, 1623.

55 R.A. Brekken, X. Huang, S.W. King, P.E. Thorpe. *Cancer Res.* **1998**, *58*, 1952.

56 S.S. Chang, V.E. Reuter, W.D. Heston, N.H. Bander, L.S. Grauer, P.B. Gaudin. *Cancer Res.* **1999**, *59*, 3192

57 R. Pasqualini, E. Koivunen, R. Kain et al. *Cancer Res.* **2000**, *60*, 722.

58 R. R. Langley, J. Russell, M. J. Eppihimer et al. *Am. J. Physiol.* **1999**, *277*, H1156.

59 S. Ran, A. Downes, P.E. Thorpe. *Cancer Res.* **2002**, *62*, 6132

60 E.B. Carson-Walter, D.N. Watkins, A. Nanda, B. Vogelstein, K.W. Kinzler, B. St Croix. *Cancer Res.* **2001**, *61*, 6649.

61 M. Birchler, F. Viti, L. Zardi, B. Spiess, D. Neri. *Nature Biotechnol.* **1999**, *17*, 984.

62 L. Tarli, E. Balza, F. Viti et al. *Blood* **1999**, *94*, 192.

63 F. Viti, L. Tarli, L. Giovannoni, L. Zardi, D. Neri. *Cancer Res.* **1999**, *59*, 347.

64 F. Nilsson, H. Kosmehl, L. Zardi, D. Neri. *Cancer Res.* **2001**, *61*, 711.

65 B. Carnemolla, L. Borsi, E. Balza et al. *Blood* **2002**, *99*, 1659.

66 C. Halin, S. Rondini, F. Nilsson et al. *Nature Biotechnol.* **2002**, *20*, 264.

67 P. Castellani, L. Borsi, B. Carnemolla et al. *Am. J. Pathol.* **2002**, *161*, 1695.

68 L. Borsi, G. Allemanni, B. Gaggero, L. Zardi. *Int. J. Cancer* **1996**, *66*, 632.

69 P. Alessi, C. Ebbinghaus, D. Neri. *Biochim. Biophys. Acta* (in press).

70 R.S. Kerbel. *Nature* **1997**, *390*, 335.

71 J. Denekamp. *Cancer Metastasis Rev.* **1990**, *9*, 267.

72 F.J. Burrows, P.E. Thorpe. *Pharmacol. Ther.* **1994**, *64*, 155.

73 T.J. Stevens, I.T. Arkin. *Proteins* **2000**, *39*, 417.

74 T. Rabilloud. *Proteomics* **2002**, *2*, 3.

75 V. Santoni, M. Molloy, T. Rabilloud. *Electrophoresis* **2000**, *21*, 1054.

76 P.H. O'Farrell. *J. Biol. Chem.* **1975**, *250*, 4007.

77 J. Klose. *Humangenetik* **1975**, *26*, 231.

78 A. Gorg, W. Postel, S. Gunther. *Electrophoresis* **1988**, *9*, 531.

79 P. Matsudaira. *J. Biol. Chem.* **1987**, *262*, 10035.

80 J. R. Yates III, S. Speicher, P.R. Griffin, T. Hunkapiller. *Anal. Biochem.* **1993**, *214*, 397.

81 P. James, M. Quadroni, E. Carafoli, G. Gonnet. *Protein Sci.* **1994**, *3*, 1347.

82 S.P. Gygi, G.L. Corthals, Y. Zhang, Y. Rochon, R. Aebersold. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 9390.

83 P. Walter, G. Blobel. *Methods Enzymol.* **1983**, *96*, 84.

84 G.F. Ames, K. Nikaido. *Biochemistry* **1976**, *15*, 616.

85 A. Harder, R. Wildgruber, A. Nawrocki, S.J. Fey, P.M. Larsen, A. Gorg. *Electrophoresis* **1999**, *20*, 826.

86 M.P. Molloy, B.R. Herbert, B.J. Walsh et al. *Electrophoresis* **1998**, *19*, 837.

87 M.P. Molloy, B.R. Herbert, K.L. Williams, A.A. Gooley. *Electrophoresis* **1999**, *20*, 701.

88 S. Luche, V. Santoni, T. Rabilloud. *Proteomics* **2003**, *3*, 249.

89 E. Gianazza, T. Rabilloud, L. Quaglia et al. *Anal. Biochem.* **1987**, *165*, 247.

90 T. Rabilloud, E. Gianazza, N. Catto, P.G. Righetti. *Anal. Biochem.* **1990**, *185*, 94.

91 T.C. Thomas, M.G. McNamee. *Methods Enzymol.* **1990**, *182*, 499.

92 G.W. Welling, R. van der Zee, S. Welling-Wester. *J. Chromatogr.* **1987**, *418*, 223.

93 T. Horigome, T. Hiranuma, H. Sugano. *Eur. J. Biochem.* **1989**, *186*, 63.

94 T. Ichimura, N. Ikuta, Y. Uda, T. Horigome, S. Omata. *Anal. Biochem.* **1995**, *224*, 250.

95 A. Tiselius, S. Hjerten. *Arch. Biochem. Biophys.* **1956**, *65*, 132.

96 M. J. Gorbunoff. *Anal. Biochem.* **1984**, *136*, 425.

97 M. J. Gorbunoff. *Anal. Biochem.* **1984**, *136*, 433.

98 M. J. Gorbunoff, S. N. Timasheff. *Anal. Biochem.* **1984**, *136*, 440.

99 W. G. Burton, K. D. Nugent, T. K. Slattery, B. R. Summers, L. R. Snyder. *J. Chromatogr.* **1988**, *443*, 363.

100 K. D. Nugent, W. G. Burton, T. K. Slattery, B. F. Johnson, L. R. Snyder. *J. Chromatogr.* **1988**, *443*, 381.

101 R. J. Simpson, L. M. Connolly, J. S. Eddes, J. J. Pereira, R. L. Moritz, G. E. Reid. *Electrophoresis* **2000**, *21*, 1707.

102 V. S. Sharov, N. A. Galeva, T. V. Knyushko, D. J. Bigelow, T. D. Williams, C. Schoneich. *Anal. Biochem.* **2002**, *308*, 328.

103 R. Aebersold, M. Mann. *Nature* **2003**, *422*, 198.

104 A. J. Link, J. Eng, D. M. Schieltz et al. *Nature Biotechnol.* **1999**, *17*, 676.

105 M. P. Washburn, D. Wolters, J. R. Yates III. *Nature Biotechnol.* **2001**, *19*, 242.

106 C. C. Wu, M. J. MacCoss, K. E. Howell, J. R. Yates. *Nature Biotechnol.* **2003**, *21*, 532.

107 S. P. Gygi, B. Rist, S. A. Gerber, F. Turecek, M. H. Gelb, R. Aebersold. *Nature Biotechnol.* **1999**, *17*, 994.

108 M. Smolka, H. Zhou, R. Aebersold. *Mol. Cell. Proteomics* **2002**, *1*, 19.

109 H. Zhou, J. A. Ranish, J. D. Watts, R. Aebersold. *Nature Biotechnol.* **2002**, *20*, 512.

110 H. Zhou, J. D. Watts, R. Aebersold. *Nature Biotechnol.* **2001**, *19*, 375.

111 D. K. Han, J. Eng, H. Zhou, R. Aebersold. *Nature Biotechnol.* **2001**, *19*, 946.

112 P. Schulz-Knappe, H. D. Zucht, G. Heine, M. Jurgens, R. Hess, M. Schrader. *Comb. Chem. High Throughput Screen* **2001**, *4*, 207.

113 H. Tammen, R. Hess, S. Uckert et al. *Urology* **2002**, *59*, 784.

114 S. B. Scheurer, J. N. Rybak, C. Roesli, D. Neri, G. Elia. Proteomics 2004, in press.

115 H. Liu, P. Moy, S. Kim et al. *Cancer Res.* **1997**, *57*, 3629.

116 S. S. Chang, D. S. O'Keefe, D. J. Bacich, V. E. Reuter, W. D. Heston, P. B. Gaudin. *Clin. Cancer Res.* **1999**, *5*, 2674.

117 G. Winter, A. D. Griffiths, R. E. Hawkins, H. R. Hoogenboom. *Annu. Rev. Immunol.* **1994**, *12*, 433.

118 F. Viti, F. Nilsson, S. Demartis, A. Huber, D. Neri. *Methods Enzymol.* **2000**, *326*, 480.

119 G. Elia, M. Silacci, S. Scheurer, J. Scheuermann, D. Neri. *Trends Biotechnol.* **2002**, *20*, S19.

120 P. Schraml, J. Kononen, L. Bubendorf et al. *Clin. Cancer Res.* **1999**, *5*, 1966.

121 R. Pasqualini, E. Ruoslahti. *Nature* **1996**, *380*, 364.

122 D. Rajotte, W. Arap, M. Hagedorn, E. Koivunen, R. Pasqualini, E. Ruoslahti. *J. Clin. Invest.* **1998**, *102*, 430.

123 L. Wyder, A. Vitaliti, H. Schneider et al. *Cancer Res.* **2000**, *60*, 4682.

124 J. Kahn, F. Mehraban, G. Ingle et al. *Am. J. Pathol.* **2000**, *156*, 1887.

125 Y. Fujiwara, Y. Sugita, S. Nakamori et al. *Int. J. Oncol.* **2000**, *16*, 799.

126 M. E. Gerritsen, R. Soriano, S. Yang et al. *Physiol. Genomics* **2002**, *10*, 13.

127 M. A. Bumke, D. Neri, G. Elia. *Proteomics* **2003**, *3*, 675.

128 G. Alessandri, R. G. Chirivi, S. Fiorentini et al. *Clin. Exp. Metastasis* **1999**, *17*, 655.

129 R. A. Craven, N. Totty, P. Harnden, P. J. Selby, R. E. Banks. *Am. J. Pathol.* **2002**, *160*, 815.

130 B. S. Jacobson, J. E. Schnitzer, M. McCaffery, G. E. Palade. *Eur. J. Cell. Biol.* **1992**, *58*, 296.

131 M. Czarny, J. Liu, P. Oh, J. E. Schnitzer. *J. Biol. Chem.* **2003**, *278*, 4424.

132 D. P. McIntosh, X. Y. Tan, P. Oh, J. E. Schnitzer. *Proc. Natl Acad. Sci. USA* **2002**, *99*, 1996.

133 E. K. De La Fuente, C. A. Dawson, L. D. Nelin, R. D. Bongard, T. L. McAuliffe, M. P. Merker. *Am. J. Physiol.* **1997**, *272*, L461.

3
Vasculature, Vascular Disease, and Atherosclerosis

Elisabetta Gianazza and Ivano Eberini

3.1
Introduction

The current view on the natural history of atherosclerosis focuses on the "response-to-injury" as the mechanism of disease [1]. Each step in its progression corresponds to a different stage in a chronic process affecting large and medium-sized elastic and muscular arteries. The compensatory responses to injury alter the normal homeostatic properties of the endothelium; if unabated and excessive, the process eventually results in a lesion.

The first step towards atherosclerosis is endothelial dysfunction, triggered by: elevated and modified (oxidized, glycated, aggregated, associated with proteoglycans, incorporated into immunocomplexes) low-density lipoproteins (LDL); free radicals (caused by cigarette smoking, hypertension and diabetes mellitus); genetic alterations (mutations in the genes coding for apolipoprotein B, CII, and E, lipoprotein lipase and the LDL receptor); elevated plasma homocysteine levels; and infections (herpesvirus, *Chlamydia pneumoniae*). The different forms of injury increase the adhesiveness of endothelium versus leukocytes and platelets, as well as its permeability. The injury also induces the conversion of the endothelium from an antithrombotic to a prothrombotic state, that is, producing vasoactive molecules, cytokines, and growth factors. The ensuing inflammatory response stimulates migration and proliferation of smooth muscle cells; at first thickening of the arterial wall is compensated by gradual dilation (remodeling). Then, monocytes/macrophages and lymphocytes migrate from the blood and multiply within the lesion; macrophages accumulate lipids and become foam cells. Activation of these cells leads to the release of hydrolytic enzymes, cytokines, chemokines, and growth factors, which can induce further damage and eventually lead to focal necrosis. Cycles of accumulation of mononuclear cells, migration and proliferation of smooth muscle cells, and formation of fibrous tissue lead to enlargement and restructuring of the lesion, which becomes covered by a fibrous cap overlying a core of lipid and necrotic tissue. When the artery can no longer compensate by dilation, the lesions intrude into the lumen and reduce the flow of blood. Atherosclerotic lesions can thus lead to ischemia of the heart, brain, or extremities, resulting in infarction or gangrene.

Biomedical Application of Proteomics
Edited by J.-C. Sanchez, G. L. Corthals, D. F. Hochstrasser

ISBN: 3-527-30807-5

Table 3.1 Classification of hyperlipidemias.

Electrophoretic phenotype	*Elevated Lp class*	*Elevated lipid class*	*Relative frequency (%)*	*Genetic defect*	*Atherogenicity*
Type I	Presence of chylomicrons	Triglycerides	<1	Mutation in lipoprotein lipase gene	none
Type II a	Elevated β-Lp (LDL)	Cholesterol	10	Mutation in LDL receptor or apolipoprotein B genes	+++
Type II b	Elevated β-Lp and pre β-Lp (LDL and VLDL)	Cholesterol and triglycerides	40	Mutation in LDL receptor or apolipoprotein B genes	+++
Type III	Fusion and elevation of β-Lp and pre β-Lp (β-VLDL)	Triglycerides and cholesterol	<1	Mutation in apolipoprotein E gene	+
Type IV	Elevated pre β-LP (VLDL)	Triglycerides	45	Mutation in lipoprotein lipase gene	+
Type V	Presence of chylomicrons and elevated pre β-Lp (VLDL)	Triglycerides and cholesterol	5	Mutation in apolipoprotein CII or lipoprotein lipase genes	+

Atherosclerosis is a most typical multifactorial and epigenetic disease. The predisposing genetic background is known in some detail [2]. As summarized in Table 3.1, mutations in the genes coding for apolipoprotein B, CII, and E, for lipoprotein lipase and for the LDL receptor are at the root of dyslipoproteinemias [3], whose outcome is atherosclerosis and eventually cardiovascular disease. However, the actual development of the disease is heavily influenced by further genetic factors and by a number of non-genetic factors (many aspects of lifestyle), and its severity varies among affected individuals. The development of atherosclerosis covers decades in one's life. As outlined above, the sequence of events is then long and complex; a thorough elucidation of all steps in the development of atherosclerotic lesions and the recognition of all cellular and enzymatic systems involved is required in order to plan for interventions aimed at blocking the progression of the pathological status.

Accordingly, many investigations deal with the characterization of atherosclerotic lesions, compared with healthy vessel walls and with fatty streaks, assumed to be the antecedents of overt lesions. A common finding is the infiltration of plas-

ma proteins in atherosclerotic plaques. The preferential adsorption of some proteins suggests specific pathological mechanisms; structural alterations of the infiltrated proteins reveal the occurrence of enzymatic or non-enzymatic processing, such as proteolysis or oxidation. In comparison with other fields in cardiovascular research only few such investigations have been carried out in recent years [4].

3.2
Protein Composition of Human Aorta in Atherosclerosis – *Ex vivo* Studies

3.2.1
Cellular Proteins

Stastny et al. [5] began their survey on normal and pathological arteries in the early 1980s, when many of the technical problems related to two-dimensional electrophoresis (2-DE) had been defined and were beginning to be solved, but none of the micromethods allowing identification of the proteins in the 2-DE maps were yet available. Except for immunological detection – with monoclonal antibodies to cellular proteins hardly available – only comigration with purified proteins could be used as an identification criterion. Accordingly, a number of proteins for which significant differences in abundance were detected are referred to in their publication by means of relative charge and/or molecular size.

Purification, partial proteolysis, and immunostaining had been used in the pioneering work by Giometti and Anderson [6] who could characterize muscle and non-muscle cells on the basis of the isoforms of tropomyosin expressed in the various cell types. The rationale for this heterogeneity, which entails the expression in a single cell type of more than one of the several genes coding for tropomyosin, may be related to the different functions they perform [6]. The above classification is relevant to the issue of atherosclerosis in that smooth muscle cells in atheromas shift from a contractile to a proliferative-synthetic phenotype.

Indeed, for atheromas, few data are available on cellular proteins proper. When comparing grossly and histologically normal human aortic intima and human aortic intima with fatty streaks or fibro-fatty lesions, actin, tropomyosin-like proteins, tubulin, glycoprotein G35, and two myosin light chains are present in high amounts in normal specimens and slightly decrease in fatty streaks. Several unidentified polypeptides (P15, P18, P60, P110b; P56, P190) that are not observed or present in only very low amounts in controls are found in significant amounts in fibro-fatty lesions. Statistically significant increases of the albumin/actin ratio are found in fatty streaks, compared with adjacent normal intimas (i.e. from the same subjects). The mean value of the albumin/actin ratio is also increased in the fibro-fatty lesions compared with the fatty streaks or normal intima. No significant differences are observed in the intima protein pattern and quantities of selected intima proteins between paired thoracic and abdominal aortas [7].

While all the above data were obtained on *ex vivo* specimens, atherosclerotic plaques may be obtained by percutaneous atherectomy with the Simpson device. In a

limited number of such samples, 2-DE shows the presence of actin in α-, β- and γ-isoforms, with predominance of the β-isoforms [8].

Some of the pathological features in atherosclerotic plaques exacerbate changes produced by age in normal specimens. After investigation by 1-DE [9], 2-DE demonstrated that plasma-derived proteins are either present solely in the intimal layer of aortas from older subjects (α_1-antichymotrypsin, haptoglobin, immunoglobulin G) or in higher concentrations in specimens from older than from younger subjects (albumin, α_1-antitrypsin) [10]. Conversely, typical intracellular protein components (actin, cytoskeletal proteins, tropomyosin-like proteins) decrease in abundance in the intimal layer of aortas from older subjects [10].

Only one recent investigation is devoted to atherosclerotic plaques. You et al. report strikingly contrasting data on ferritin light chains and their mRNA: a 90% increase versus a 40% reduction, respectively, in coronary arteries from patients with atherosclerotic lesions versus appropriate controls [11]. The quantitative protein data are obtained both through general protein staining and MS identification and through immunostaining. The latter check is critical to take into account all molecular forms of the protein of interest. Indeed, for a given protein, several post-translationally processed forms often exist, which includes the occurrence of degradation products. Although genomics, transcriptomics, and proteomics all describe the same systems at different levels of integration, little correlation has been reported in a number of instances between mRNA and protein levels [12]; significant differences in half-lives and pool sizes between the two types of macromolecules have been suggested as the general explanation for such discrepancies. A correlation between excessive iron storage and increased risk of atherosclerosis has been reported as one of the aspects connecting oxidative stress and degenerative disease [13]. The data by You et al. indeed provide "*in situ* proteomic evidence" for such a correlation but cannot directly settle at which level in a complex sequence of events the observed alteration occurs.

3.2.2 Infiltrating Proteins

Much more is known about how discontinuity in the endothelial layer allows access of plasma proteins to the intimal matrix. Fatty streaks are infiltrated by several plasma proteins (immunoglobulin G, fibrinogen, transferrin, albumin, α_2-HS-glycoprotein, α_1-antitrypsin, α_1-antichymotrypsin, apolipoprotein A-I and A-II) [5]. The extent of transferrin, albumin, α_1-antitrypsin and apolipoprotein A-I infiltration appears to increase with the severity of fatty streaks and to become massive in aortic intima with fibro-fatty lesions [7, 14].

Data from 2-DE extend previous findings on tissue infiltration by plasma proteins using immunological techniques such as radial immunodiffusion. The intima layer in fibrous plaque and its adjacent area presenting intimal thickening contain all test substances – a panel of 14 plasma proteins – at higher levels than fatty streaks; significant differences are found for IgG and IgA; the third and fourth components of the complement, C3 and C4, are found in activated form.

In comparison with the other studied proteins, immunoglobulins and complement components are present in higher intima/serum and in lower media/intima retention ratios. This preferential retention results from an altered permeability and may in turn influence specific protein function [15].

3.2.2.1 Infiltrating Lipoproteins

Lipoproteins are a class of plasma proteins of major relevance for atherosclerosis. Hoff et al. devoted a comprehensive investigation to lipoproteins infiltrating control and pathological aortas, using immunohistological [16, 17] and biochemical procedures; for the latter, differential flotation was the fractionation and characterization protocol of choice.

The $1.006 < d < 1.063$ g/ml density fraction extracted from aorta demonstrates higher electrophoretic mobility (possibly due to oxidation) than plasma LDL. The lipid composition of isolates from normal intima is similar to that of LDL whereas the plaque fractions show a significant decrease in the cholesteryl ester to free cholesterol ratio and in the triglyceride content in comparison to LDL and to fractions from normal intima. The fatty acid pattern of the cholesteryl ester fraction from isolates of both normal and plaque aortic homogenates demonstrates a significant decrease in the linoleate to oleate ratio compared with that of LDL [18]. Lipoprotein(a) (Lp(a)) accumulates preferentially to LDL in plaques, in a form that is not readily available (requiring guanidinium chloride rather than PBS for its extraction). Plaque apo(a) level correlates with plasma apo(a) level [19]. Quantitation of apolipoprotein A-I infiltrating the intimal layer in raised atherosclerotic lesions of human aortas detects similar amounts in normal tissue and in aortic plaques. No apolipoprotein A-I is present in the tunica media. However, the lipid content of the infiltrating lipoproteins does change between normal and pathological specimens: by ultracentrifugation the proportion of apoA-I-containing particles with a buoyant density $d < 1.063$, $1.063 < d < 1.21$ and $d > 1.21$ g/ml is 1, 94 and 5% in normal intima, which is not very different from that in control plasma (1, 89, 10%). In contrast, in atherosclerotic plaques the relative proportions are 19, 31 and 50%. The higher abundance $1.063 < d < 1.21$ fraction in fibrous plaques is organized in particles 6–12 nm in diameter that contain more free cholesterol and phospholipids and less cholesteryl ester than the plasma HDL [20].

3.2.2.2 Effect on Apolipoproteins of Proteases Released in Atherosclerotic Lesions

Proteins infiltrating fatty streaks and fibrous lesions may have a different fate and different physiological effects. Recent investigations have concentrated on two classes of proteases relevant in the vessel walls under pathological conditions and on their effect on infiltrating lipoproteins.

Matrix metalloproteinases (MMPs), a family of calcium- and zinc-dependent endoproteinases with catalytic activity against extracellular matrix components [21], have been implicated in the progression of atherosclerotic plaques. In their

active form they may contribute to vascular remodeling and to plaque disruption [22–25].

After incubation with purified MMP-3, -7, or -12, the ability of HDL_3 to induce the high-affinity component of cholesterol efflux from macrophage foam cells is markedly reduced, whereas preincubation with MMP-1 reduces cholesterol efflux only slightly and MMP-9 is ineffective [26].

All cell types involved in atherosclerotic plaque formation – macrophages, endothelial and smooth muscle cells – have been reported to release MMPs. Following culture for 24 h and incubation of HDL_3 in cell-free conditioned media, proteolytic degradation of apolipoprotein A-I is observed with macrophage-, but not with endothelial- or muscle cell-conditioned supernatants. The identified apolipoprotein A-I fragments have sizes of 26, 22, 14, and 9 kDa. The higher abundance of C- than N-terminus cleaved peptides agrees with literature data for a fully structured α-helix around Tyr18 versus an unstructured region around Gly185 and Gly186. Macrophages are the first cells to migrate to the lesion and the action of their MMPs on HDL is to hamper the mechanism of cholesterol efflux. Conversely, apoA-I entering the lesion behaves as a suicide substrate which spares matrix degradation and helps to stabilize the plaque [27].

Chymase is a cytoplasmic, chymotrypsin-like neutral protease secreted by activated mast cells. Proteolysis with chymase for up to 24 h does not alter the integrity of the α-migrating HDL, whereas a minor peak containing particles of smaller size with pre-β mobility rapidly disappears. Incubation with chymase reduces the ability of HDL_3 to induce high-affinity efflux of cholesterol from macrophage foam cells, without reducing the ability of HDL_3 to activate lecithin : cholesterol acyltransferase (LCAT) [28]. Chymase cleaves apoA-I contained in reconstituted HDL either at the N- (Tyr18 or Phe33) or at the C-terminus (Phe225), generating three major truncated polypeptides that remain bound to the rHDL. When present alone in reconstituted HDL, apoA-II is resistant to degradation. However, when together with apoA-I, apoAII is degraded by chymase [29].

In a pilot study on patients referred to a coronary unit, fragments of apoA-I matching the size of metalloprotease-degraded HDL could be detected in sera from 3 of 7 patients with acute myocardial infarction (I. Eberini et al., unpublished data).

3.3
Protein Composition of Human Aorta in Atherosclerosis – *In vitro* Studies

Smooth muscle cells may be cultured and their proteomes characterized as a function of experimental conditions. For instance, 14 proteins are differentially expressed in smooth muscle cells cultured from newborn versus aged rats, 10 being up-regulated in older animals and four being down-regulated [30].

The effect of growth factors on smooth muscle cells has been taken as a model for abnormal growth in atherosclerosis. Growing vascular smooth muscle cells in the presence of various growth factors – 10% calf serum, platelet-derived growth

factor or angiotensin II – results in similar proteomic changes [31]. Up-regulation involves mediators of protein folding: heat shock protein 60 and 70, protein disulfide isomerase (and its isozyme Q-2), and calreticulin; and a component of the protein synthesis apparatus: elongation factor EF-1β together with vimentin and actin. An isoform of myosin heavy chain is down-regulated.

The composition of the extracellular matrix underlying endothelial cells (deoxycholate-insoluble) changes when a steady laminar shear stress is applied in comparison with culture under static conditions [32]. General protein stains after 2-DE detect four (unidentified) spots increasing in abundance after 3–6 hours of exposure to flow. Immunostain for fibronectin detects a decrease after 12 hours followed by an increase between 24 and 48 hours; for laminin an increase between 24 and 48 hours; and no changes for vitronectin. Immunofluorescence microscopy demonstrates grouping of fibrils into thicker filaments – fibronectin fibrils aligned to the flux, laminin and collagen IV fibrils randomly oriented.

3.4 Lipoproteins and Apolipoproteins as Disease Factors

3.4.1 Typing of Apolipoprotein E Phenotype in Humans

The first typing of apolipoprotein E phenotype was obtained by 2-DE [33]. A specific subclass, at first defined as βIV and later as apoE4, is associated with type III hypertriglyceridemia [3], a disease characterized by xanthomatosis and premature atherosclerosis, and presenting with a specific pattern upon electrophoresis (fusion and increase of β-Lp with pre-β-Lp). A further association of apoE4, with Alzheimer's disease, has been reported [34]. ApoE typing on untreated serum samples has been reported, using focusing on immobilized pH gradients followed by print immunoblotting with an in-house produced polyclonal antiserum [35]. However, since few antisera of adequate quality for such an application are commercially available, apoE typing is still most often performed with either 1-DE (according to isoelectric point) or 2-DE procedures, after purification of the lipoprotein fraction (described in detail in [36]; see also [37] and [38]) or by 2-DE on whole serum [39].

3.4.2 Studies in Transgenic or Knockout Mice

Transgenic animals – most frequently mice but also, in a few cases, rabbits – have been used to assess the functions of the various apolipoproteins, enzymes, and receptor proteins involved in lipoprotein metabolism. Overexpression of human genes was the preferred strategy; indeed, this is not only the most direct technical approach but also, the lipoprotein profile being different in mice and in humans, the extra proteins do not simply add up to an already high background. In a few

cases, such as for apolipoprotein A-I, targeted replacement has been applied. Knocking-out of the apolipoprotein E gene has established a mouse line that has become a standard for investigating spontaneous hypercholesterolemia in the absence of a dietary challenge.

The proteomic approach has been mainly devoted to monitoring the changes induced by transgenes in the overall lipoprotein set-up. For instance, 2-DE under denaturing but either non-reducing or reducing conditions demonstrates a differential secretion from liver of two allelic variants of apolipoprotein A-I. The relevant transgenic mice have been obtained by targeted replacement of mouse apoA-I with human apoA-I or the mutant apoA-I_{Milano} [40]. ApoA-I_{Milano} is a molecular variant of apoA-I characterized by the Arg173 → Cys substitution [41]. The presence of a cysteine residue results in the formation of homodimers and heterodimers with apoA-II. The carriers of this mutation are all heterozygotes that exhibit hypertriglyceridemia with markedly reduced HDL and apoA-I levels; none of them has clinical signs of atherosclerotic disease.

More frequently, the whole lipoproteins have been separated and quantitated under native conditions. Native 2-DE, i.e. the sequence of agarose electrophoresis and of gradient gel electrophoresis followed by general protein stain or better immunostain for various lipoproteins, is able to resolve intact lipoproteins while characterizing them according to charge density, in the first dimension, and to particle volume, in the second dimension. Although setting up such an approach predates the definition of protocols for high-resolution 2-DE under denaturing conditions, we suggest that native 2-DE still fits the current perspectives of proteomics. Indeed, it addresses the analysis of a relevant subproteome and allows directly definition of the protein–protein interactions that are the topic of functional proteomics.

While in control mice all apoA-I migrates in α-particles, mice transgenic for human apolipoprotein A-I have 30% in particles with pre-β mobility and mice transgenic for both human apoA-I and apoA-II have 38% of apoA-I in particles with pre-β mobility [42].

Low-dose expression of a human apolipoprotein E3 transgene in deficient (E–/–) mice restores cholesterol efflux capacity from fibroblasts and macrophages mediated by lipoprotein particles of the same size as resolved from wild-type serum [43].

In knockout mice for lecithin : cholesterol acyltransferase (LCAT) native 2-DE demonstrates the presence of heterogeneous pre-β-migrating HDL as well as of triglyceride-enriched VLDL [44]. A very high increase (10-fold) in the expression of scavenger receptor class B type I (SR-BI) in transgenic mice hemizygous for a human apolipoprotein B transgene results in grossly altered structural changes of HDL; α-migrating particles are almost completely absent [45]. In the same animals no reduction of diet-induced fatty streak lesions is observed in comparison with apoB transgenics whereas transgenic mice expressing lower levels of SR-BI (2-fold increase) show a more than 2-fold decrease in lesions.

Only one recent investigation has followed the comprehensive approach typical of proteomic investigations, analyzing the level of expression of all serum proteins in a transgenic animal. An apolipoprotein E variant with a reduced affinity for the

LDL receptor (apoE3$_{Leiden}$) results in the dominant occurrence of type III hyperlipoproteinemia in human carriers. In transgenic mice, the expression of apoE3$_{Leiden}$ is associated with severe spontaneous arterial disease, hyperlipidemia, and elevated serum levels of haptoglobin, a typical acute-phase reactant in this species [46]. Shehel et al. thus suggest haptoglobin may provide a marker for the onset of atherosclerosis in humans (see next section for a further discussion of inflammatory proteins as disease markers and prognostic indicators).

3.5 Pathogenetic Mechanisms

An increasing number of reports have connected ongoing inflammatory conditions with onset and progression of atherosclerosis and have identified C-reactive protein (CRP) as a strong predictor of myocardial infarction [47, 48] and of ischemic brain disease [48–50]. From immunohistochemical studies, CRP colocalizes with complement C5b-9 in early atherosclerotic lesions; CRP also binds to phosphorylcholine groups that become exposed in enzymatically degraded, non-oxidized LDL particles [51]. CRP may thus be the link between LDL deposition and complement activation in the development of atherosclerotic lesion.

In humans, C-reactive protein is the most sensitive marker of many inflammatory conditions, as its levels increase by one order of magnitude within a few hours of the noxious stimulus [52]. It should be noted, however, that in any given species the expression of many proteins is affected by inflammation, if with a different time-course. Moreover, the set of proteins involved in acute-phase reactions differs depending on the species [53, 54]: among laboratory animals, the prominent positive acute-phase reactant is thiostatin in rats but haptoglobin in mice.

Besides CRP, levels of α_1-antitrypsin, α_1-acid glycoprotein, α_2-macroglobulin, ceruloplasmin, and Lp(a) are significantly associated with the severity of coronary atherosclerosis as determined by the Gensini score and can serve as independent indicators of the progression of coronary atherosclerosis [55]. In a cohort study based on over 6000 men in their mid-forties at the time of the screening examination, and with a follow-up of 16 years, on average, the incidence of cardiac events and mortality increases stepwise with the plasma levels of fibrinogen, α_1-antitrypsin, haptoglobin, and ceruloplasmin (defined as quartiles) [56]. However, complex relationships hold among acute-phase reactants, as is seen by analyzing data about fibrinogen with and without taking into account other test proteins.

The haptoglobin 2–2 type appears to be associated with accumulation of atherosclerotic lesions in essential hypertension [57]. Another indirect connection with atherosclerosis is given by the finding that haptoglobin β-chains are present in higher concentrations in macular extracts from patients with age-related maculopathy than from control subjects [58]. A relationship between age-related maculopathy, the most common cause of blindness in the elderly, and atherosclerosis has been suggested by epidemiological studies [59].

While in the above investigations 2-DE has been seldom used as the analytical tool, two aspects of the pathogenetic mechanisms leading to atherosclerosis are being analyzed with a typical proteomic approach.

2-DE was used as the final micro-preparative step for the purification of a protein from the plasma membrane of human cell lines and primary cultured hepatocytes with affinity for HDL [60]. This was demonstrated to be heat shock protein 60 (hsp60), which provides a potential mechanism to explain the known association between immunity developed against hsp60 and the development of atherosclerosis.

Complex strategies are being implemented for the recognition of surface proteins of *Chlamydia pneumoniae*, a human pathogen probably contributing to the development of atherosclerosis [61]. These include *in silico* prediction from the available genome sequences of peripherally located proteins, heterologous expression and purification of selected proteins, production of mouse immune sera against the recombinant proteins to be used in immunoblotting and fluorescence-activated cell sorter analyses for the identification of surface antigens, and mass spectrometry analysis of 2-DE maps of chlamydial protein extracts to confirm the presence of the fluorescence-activated cell sorter (FACS)-positive antigens in the chlamydial cell.

3.6
End Pathologies: Myocardial and Cerebral Infarction

Serum was investigated by 2-DE after myocardial infarction with the aim of detecting specific disease markers. Serum proteins acting as acute-phase reactants change in their level with specific time-course. A protein (M_r = 27 000; pI = 5.2) tentatively identified as myosin light chain and three spots (M_r = 13 000; pI = 6.2, 6.7, and 7.5, not detected in the 2-DE patterns of healthy myocardium, infarcted myocardium, pectoral muscle, or tongue) appear simultaneously approximately 30 h after infarction, reach maximum intensity after 48 h and progressively decline thereafter [62, 63]. Unfortunately, at the time when this investigation was performed, no identification by N-sequencing or MS was possible on the resolved spots.

The protein pattern of human cerebrospinal fluid under control conditions has been described in detail [64–67]. In contrast, due to obvious problems with sampling from acutely ill patients, no thorough proteomic investigation has yet been carried out on CSF from stroke patients and only a few data are available. One of the common findings is that of variability among subjects and poor to no correlation between biochemical parameters and clinical findings. Measurements of CSF albumin, and serum/CSF albumin in two groups of patients with small and large infarcts fail to show blood/brain barrier permeability to albumin in about half the patients [68]. Significantly increased concentrations of total protein and albumin are found in patients with acute non-embolic and embolic infarctions and bleeding, compared with patients having old infarctions or with controls. Infarction

sizes on computerized tomography scan does not generally correlate with the CSF parameters. High CSF concentrations of total protein and albumin in acute infarctions are related to a poor short-term prognosis [69]. After acute cerebrovascular disease, plasma creatine kinase, lactic dehydrogenase and the α_1-protein fraction show a moderately close relationship to the severity of stroke. In CSF no relation is observed between either enzymatic activities or CSF protein, or serum/CSF albumin ratio, and the severity of disease [70]. Prostaglandin D (PGD) synthase [71], also known as β-trace protein, has been investigated as a possible marker of brain disorders, with controversial results. Melegos et al. conclude that measurements of PGD synthase levels in CSF or plasma has no clinical utility in diagnosing cerebrovascular disease [72].

3.7 Surgical Treatments

Reduced flow through vessels occluded by atheromas may be treated by balloon angiography or surgical bypass. Saphenous veins and internal mammary arteries are most often used as grafts. Bypass surgery using veins depends on the successful adaptation of the vein to high pressure and pulsatile flow of the arterial circulation. Migration of smooth muscle cells to the intima, with alteration to a synthetic, proliferative phenotype, that causes intimal hyperplasia, is an important underlying cause of vein graft failure; similar cellular changes are observed in restenosis following angioplasty of diseased arteries. A reference map of human saphenous vein medial smooth muscle has been published recently [73] (Fig. 3.1). Two-dimensional electrophoresis as well as ultrastructural and immunochemical data on the distal part of left internal mammary artery obtained at the time of surgery from patients undergoing coronary bypass confirm that most smooth muscle cells in these specimens are in a differentiated state [74]; this finding explains their resistance to atherosclerosis and the better outcome of arterial versus vein bypass. Proteomic differences between arteries and veins are obvious in 2-DE maps of animal specimens (Fig. 3.2).

Accelerated coronary atherosclerosis is a major risk after heart transplants that limits long-term survival. It is associated with dyslipoproteinemia even in subjects with no such problems before surgery. By native 2-DE, a predominance of HDL_2-like patterns is observed in transplant recipients [75].

3.8 Pharmacological and Dietary Treatments

Hypercholesterolemia is one of the major risk factors for atherosclerosis. The effect of various dietary and pharmacological treatments on liver proteins was one of the first biochemical investigations carried out using the comprehensive approach allowed by 2-DE [76]. However only recently could the initial data be re-

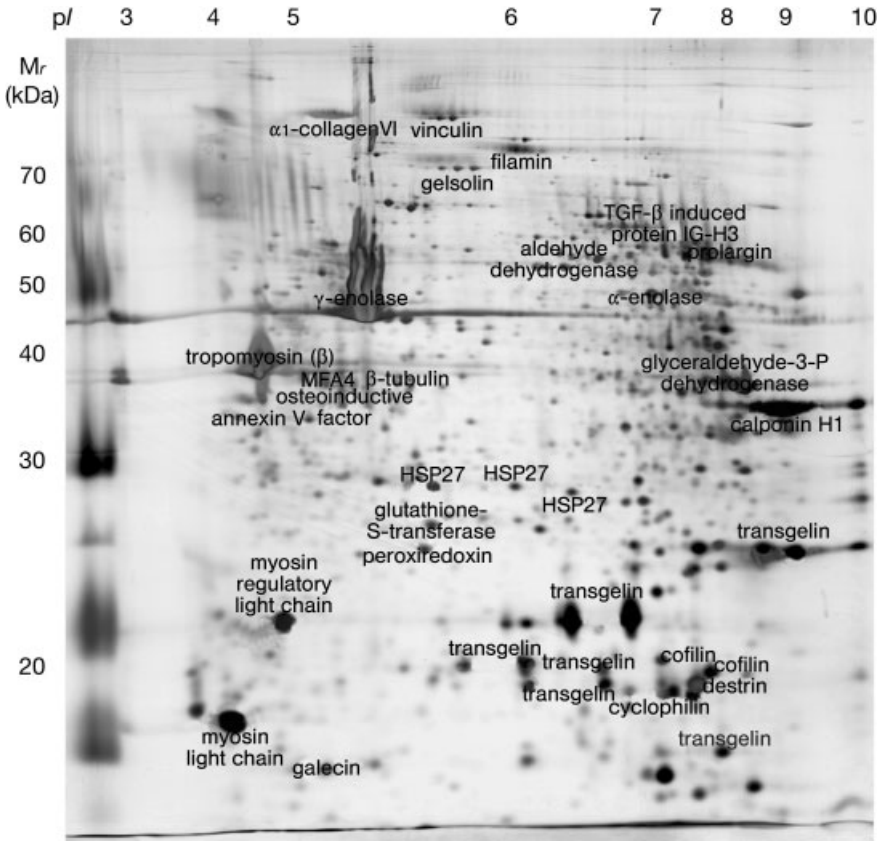


Figure 3.1 Two-dimensional electrophoresis map of human saphenous vein medial smooth muscle proteins. Proteins (400 g loading) were separated by IEF using 180 mm, immobilized, non-linear pH gradient strips (IPG Dry-Strips) of pH 3–10, followed by 12% T gradient SDS-PAGE gels. Proteins were visualized by silver staining. Proteins were identified by MALDI-MS and nanospray-MS. Major spots are marked by their names; for identification of all 150 spots analyzed by the authors, see Table 1 in [74]. Modified from [74] with permission of the publisher.

assessed with the identification of the enzymes up- or down-regulated under the given experimental conditions [77, 78]. Treatment with lovastatin and fluvastatin, cholesterol-lowering drugs that inhibit 3-hydroxy-3-methylglutaryl-coenzyme A (HMG-CoA) reductase, affects the levels of 58 liver proteins in rats. Major effects are evident in the cholesterol biosynthesis pathway, including the induction of enzymes upstream and downstream of HMG-CoA reductase. Treatment also affects key enzymes of carbohydrate metabolism and cellular stress proteins involved in cytoskeletal structure, calcium homeostasis, and protease activity. The latter set of protein alterations indicates that high-dose treatment may result in hepatotoxicity. From these findings Steiner et al. suggest that HMG-CoA synthase and isopente-

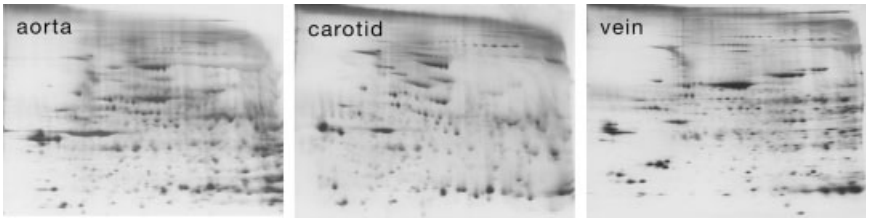


Figure 3.2 Two-dimensional electrophoresis of extracts from (*top to bottom*) thoracic aorta and carotid artery, and cava veins, from Sprague-Dawley rats. Tissues were homogenized in 2% SDS/2% 2-mercaptoethanol; aliquots were acetone precipitated. The pellets were redissolved in 8 M urea, 2% 2-mercaptoethanol, 2% NP40, and 2% CA in the pH range 8–10.5. IEF was run on non-linear 4–10 IPG, SDS-PAGE on 4–16% T PAA gradients. Protein pattern was silver stained.

nyl-diphosphate Δ-isomerase may be explored as alternative drug targets and that the induction levels of these enzymes may serve as a measure of potency of individual statin drugs [78].

The effect of diets on the occurrence and severity of atherosclerosis and its sequels has been extensively investigated. Epidemiological data have shown a correlation between moderate alcohol consumption and prevention of cardiovascular disease [79, 80]. However, wine appears to rank better than beer, and liquors have little if any positive effect. Indeed, in wine a few components with antioxidant properties, such as resveratrol, have been identified as the possible effectors of the protective action [81]. Data by Gorinstein et al. [82] demonstrate that after 30 days of consumption of 20 g alcohol per day, provided by beer, fibrinogen levels appear to be reduced, as assessed by quantitation after 2-DE. Circular dichroism and Fourier transform infrared spectroscopy demonstrate reduced protein stability, including a decrease in α-helix and increase of β-sheet content.

3.9
Animal Models of Atherosclerosis and its Complications

In an evaluation of the effects of myocardial ischemia and reperfusion in rabbits, complement inhibition with the synthetic serine protease inhibitor nafamstat mesilate (FUT-175) significantly reduces myocardial necrosis and leukocyte accumulation as well as preserves the expression of superoxide dismutase and αB-crystallin in comparison with vehicle-treated and sham-treated animals [83]. The spontaneously hypertensive stroke-prone rat (SHRSP) provides an inbred animal model for a complex form of cerebrovascular pathology resembling, in many aspects, the human disease [84]. In SHRSPs subjected to salt loading (to hasten the progression of the disease) an atypical inflammatory condition and widespread alterations of vascular permeability develop prior to the appearance of anomalous features in the brain detected by magnetic resonance imaging [85]. Markers of an inflammatory response, including very high levels of thiostatin, are detected in the serum

of SHRSPs at least 4 weeks before stroke develops. Several proteins are excreted in urine after some weeks of salt loading, and in advance of stroke (transferrin, hemopexin, albumin, α_2-HS-glycoprotein, kallikrein-binding protein, α_1-antitrypsin, Gc-globulin, transthyretin). Their urinary concentrations correlate with time before stroke occurs [86]. Blood/brain barrier leakage, as evident from plasma protein extravasations towards brain interstitial fluid and CSF, occurs in synchrony with the appearance of MRI signs of brain abnormalities. High levels of plasma proteins are detected in the CSF (with molecular masses up to >130 kDa) and in tissue homogenates of the infarcted brain areas. In contrast, in rats still without brain abnormalities only an increase of total protein concentration is measured, with no signs of plasma protein inflow. Gadolinium diffusion shows impairment of blood/brain barrier localized in a single or a few small foci. Tissue lesions are initially localized at these foci then spread throughout the brain in the form of fibrinoid necrosis (L. Sironi et al., unpublished).

3.10 Conclusions

The atherosclerotic plaque has a complex composition. In this respect, its analysis offers a case similar to that of cancer specimens: microdissection and evaluation of the various cellular types plus the vast array of extracellular components would help to discriminate in an objective way the different stages in the progression of the lesion. Even undissected specimens, however, may provide some more interesting observations: up-to-date analytical procedures still have to be applied to in-depth investigations in the field.

3.11 Acknowledgements

The author's current research is funded by grants from MIUR, from Università degli Studi di Milano, from INBIFO and from Fondazione CARIPLO.

Emma Mc Gregor and Robin Wait kindly provided original image files for Figure 3.1. Rossana Ballerio helped with rat sample preparation for 2-DE in Figure 3.2. The authors are indebted to Professor Cesare Sirtori for critically reading the manuscript.

In memory of Professor Giovanni Galli, a colleague who is no longer with us.

3.12 References

1 R. Ross. *Ann. Rev. Physiol.* **1995**, *57*, 791–804.
2 N. E. Miller. *J. Clin. Pathol.* **1979**, *32*, 639–650.
3 D. S. Fredrickson, R. I. Levy, R. S. Lees. *N. Engl. J. Med.* **1967**, *276*, 34–42, 94–103, 148–156, 215–225 and 273–281.
4 D. K. Arrell, I. Neverova, J. E. van Eyk. *Circ. Res.* **2001**, *88*, 763–773.
5 J. Stastny, E. Fosslien, A. L. Robertson, Jr. *Atherosclerosis* **1986**, *60*, 131–139.
6 C. S. Giometti, N. L. Anderson. *J. Biol. Chem.* **1984**, *259*, 14113–14120.
7 J. Stastny, E. Fosslien. *Exp. Mol. Pathol.* **1992**, *57*, 205–214.
8 P. C. Dartsch, G. Bauriedel, I. Schinko, H. D. Weiss, B. Hofling, E. Betz. *Atherosclerosis* **1989**, *80*, 149–157.
9 J. Song, J. Stastny, E. Fosslien, A. L. Robertson, Jr. *Exp. Mol. Pathol.* **1985**, *43*, 233–241.
10 J. Song, J. Stastny, E. Fosslien, A. L. Robertson, Jr. *Exp. Mol. Pathol.* **1885**, *43*, 297–304.
11 S. A. You, S. R. Archacki, G. Angheloiu et al. *Physiol. Genomics* **2003**, *18*, 25–30.
12 L. Anderson, J. Seilhamer. *Electrophoresis* **1997**, *18*, 533–537.
13 J. Balla, H. S. Jacob, G. Balla, K. Nath, J. W. Eaton, G. M. Vercellotti. *Proc. Natl Acad. Sci. USA* **1993**, *90*, 9285–9289.
14 J. Stastny, A. L. Robertson, Jr., E. Fosslien. *Exp. Mol. Pathol.* **1986**, *45*.
15 R. Vlaicu, H. G. Rus, F. Niculescu, A. Cristea. *Atherosclerosis* **1985**, *55*, 35–50.
16 H. F. Hoff, J. W. Gaubatz. *Exp. Mol. Pathol.* **1977**, *26*, 214–227.
17 H. F. Hoff, C. L. Heideman, J. W. Gaubatz. *Stroke* **1977**, *8*, 366–370.
18 H. F. Hoff, W. A. Bradley, C. L. Heideman, J. W. Gaubatz, M. D. Karagas, A. M. Gotto, Jr. *Biochim. Biophys. Acta* **1979**, *573*, 361–374.
19 J. M. Pepin, J. A. O'Neil, H. F. Hoff. *J. Lipid Res.* **1991**, *32*, 317–327.
20 C. L. Heideman, H. F. Hoff. *Biochim. Biophys. Acta* **1982**, *711*, 431–444.
21 J. F. Woesser, H. Nagase. *Matrix Metalloproteinases and TIMPs*. Oxford University Press, Oxford, **2000**.
22 C. M. Dollery, J. R. McEwan, A. M. Henney. *Circ. Res.* **1995**, *77*, 863–868.
23 D. C. Celentano, W. H. Frishman. *J. Clin. Pharmacol.* **1997**, *37*, 991–1000.
24 P. Libby, U. Schoenbeck, F. Mach, A. P. Selwyn, P. Ganz. *Am. J. Med.* **1998**, *104*, 14S–18S.
25 G. Pasterkamp, A. H. Schoneveld, D. J. Hijnen et al. *Atherosclerosis* **2000**, *150*, 245–253.
26 L. Lindstedt, J. Saarinen, N. Kalkkinen, H. Welgus, P. T. Kovanen, *J. Biol. Chem.* **1999**, *274*, 22627–22634.
27 I. Eberini, L. Calabresi, R. Wait et al. *Biochem. J.* **2002**, *366*, 245–253.
28 M. Lee, P. Uboldi, D. Giudice, A. L. Catapano, P. T. Kovanen. *J. Lipid Res.* **2000**, *41*, 975–984.
29 M. Lee, P. T. Kovanen, G. Tedeschi, E. Oungre, G. Franceschini, L. Calabresi. *J. Lipid Res.* **2003**, *44*, 539–546.
30 O. Cremona, M. Muda, R. D. Appel et al. *Exp. Cell Res.* **1995**, *217*, 280–287.
31 W. F. Patton, H. Erdjument-Bromage, A. R. Marks, P. Tempst, M. B. Taubman. *J. Biol. Chem.* **1995**, *270*, 21404–21410.
32 O. Thoumine, R. M. Nerem, P. R. Girard. *Lab. Invest.* **1995**, *73*, 656–576.
33 V. I. Zannis, P. W. Just, J. L. Breslow. *Am. J. Human, Genet.* **1981**, *33*, 11–24.
34 E. H. Corder, A. M. Saunders, W. J. Strittmatter et al. *Science* **1993**, *261*, 921–923.
35 M. Kohlmeier, H. J. Drossel, P. Sinha, E. Kottgen. *Electrophoresis* **1992**, *13*, 258–261.
36 U. de Faire, L. Friberg, T. Lundman. *Prev. Med.* **1975**, *4*, 509–517.
37 G. Utermann, A. Steinmetz, W. Weber. *Hum. Genet.* **1982**, *60*, 344–351.
38 J. M. Ordovas, L. Litwak-Klein, P. W. F. Wilson, M. M. Schaefer, E. J. Schaefer. *J. Lipid Res.* **1987**, *28*, 371–380.
39 D. F. Hochstrasser, J.-D. Tissot. In *Advances in Electrophoresis* (Chrambach, A., Dunn, M. J., and Radola, B. J., Eds). VCH, Weinheim, **1993**, pp 270–275.
40 C. Parolini, G. Chiesa, Y. Zhu et al. *J. Biol. Chem.* **2003**, *278*, 4740–4746.

41 K. H. Weisgraber, S. C. Rall, T. P. Bersot, R. W. Mahley, G. Franceschini, C. R. Sirtori. *J. Biol. Chem.* **1983**, *258*, 2508–2513.

42 G. Castro, L. P. Nihoul, C. Dengremont et al. *Biochemistry* **1997**, *36*, 2243–2249.

43 Y. Zhu, S. Bellosta, C. Langer et al. *Proc. Natl Acad. Sci. USA* **1998**, *95*, 7585–7590.

44 N. Sakai, B. L. Vaisman, C. A. Koch et al. J. Biol. Chem. **1997**, *272*, 7506–7510.

45 Y. Ueda, E. Gong, L. Royer, P. N. Cooper, O. L. Francone, E. M. Rubin. *J. Biol. Chem.* **2000**, *275*, 20368–20373.

46 J. M. Skehel, K. Schneider, N. Murphy et al. *Electrophoresis* **2000**, *21*, 2540–2545.

47 K. O. Pietila, A. P. Harmoinen, J. Jokiniitty, A. I. Pasternack. *Eur. Heart J.* **1997**, *17*, 1345–1349.

48 P. M. Ridker, C. H. Hennekens, J. E. Buring, N. Rifai. *N. Engl. J. Med.* **2000**, *342*, 836–843.

49 P. M. Ridker, M. Cushman, M. J. Stampfer, R. P. Tracy, C. H. Hennekens. *Circulation* **1998**, *97*, 425–428.

50 K. W. Muir, C. J. Weir, W. Alwan, I. B. Squire, K. R. Lees. *Stroke* **1999**, *30*, 981–985.

51 S. Bhakdi, M. Torzewski, M. Klouche, M. Hemmes. *Arterioscler. Thromb. Vasc. Biol.* **1999**, *19*, 2348–2354.

52 M. B. Pepys, M. L. Baltz. *Adv. Immunol.* **1983**, *34*, 141–212.

53 G. Schreiber, G. Howlett, M. Nagashima et al. *J. Biol. Chem.* **1982**, *257*, 10271–10277.

54 I. Kushner. In *Marker Proteins in Inflammation* (Arnaud, P., Benvenu, J., Laurent, P., Eds). W. deGruyter, Berlin, **1984**, pp 3–14.

55 T. Mori, J. Sasaki, H. Kawaguchi et al. Am. Heart J. **1995**, *129*, 234–238.

56 P. Lind, B. Hedblad, L. Stavenow, L. Janzon, K. F. Eriksson, F. Lindgarde. *Arterioscl. Thromb. Vasc. Biol.* **2001**, *21*, 452–458.

57 J. Delanghe, B. Cambier, M. Langlois et al. *Atherosclerosis* **1997**, *132*, 215–219.

58 M. Kliffen, P. T. de Jong, T. M. Luider. *Lab. Invest.* **1995**, *73*, 267–272.

59 R. Klein, L. Clegg, L. S. Cooper et al. Arch. Ophthalmol. **1999**, *117*, 1203–1210.

60 A. V. Bocharov, T. G. Vishnyakova, I. N. Baranova et al. *Biochem. Biophys. Res. Commun.* **2000**, *277*, 228–235.

61 S. Montigiani, F. Falugi, M. Scarselli et al. *Infect. Immun.* **2002**, *70*, 368–379.

62 T. Marshall, J. Williams, K. M. Williams. *Electrophoresis* **1988**, *9*, 672–675.

63 T. Marshall, J. Williams, K. Williams. *Electrophoresis* **1989**, *10*, 584–588.

64 M. Yun, W. Wu, L. Hood, M. Harrington. *Electrophoresis* **1992**, *13*, 1002–1013.

65 J. C. Sanchez, R. D. Appel, O. Golaz et al. *Electrophoresis* **1995**, *16*, 1131–1151.

66 J. Raymackers, A. Daniels, V. De Brabandere et al. *Electrophoresis* **2000**, *21*, 2266–2283.

67 A. Sickmann, W. Dormeyer, S. Wortelkamp, D. Woitalla, W. Kuhn, H. E. Meyer. *Electrophoresis* **2000**, *21*, 2721–2728.

68 S. Al-Kassab, T. Skyhoj Olsen, E. B. Skriver. *Acta Neurol. Scand.* **1981**, *64*, 438–445.

69 R. Palm, T. Strand, G. Hallmans. *Acta Neurol. Scand.* **1986**, *74*, 308–313.

70 K. J. Lamers, H. C. Schoonderwaldt, M. V. Borkent, A. G. Theeuwes, W. H. Doesburg, R. A. Wevers. *Clin. Neurol. Neurosurg.* **1987**, *89*, 23–29.

71 M. C. Peitsch, M. S. Boguski. *Trends Biochem. Sci.* **1991**, *16*, 363–363.

72 D. N. Melegos, M. S. Freedman, E. P. Diamandis. *Prostaglandins* **1997**, *54*, 463–474.

73 E. McGregor, L. Kempster, R. Wait et al. *Proteomics* **2001**, *1*, 1405–1414.

74 S. T. Nikkari, T. Sisto, T. Nikkari. *Atherosclerosis* **1989**, *79*, 129–138.

75 V. Atger, M. Cambillau, R. Guillemain et al. *Atherosclerosis* **1990**, *81*, 103–110.

76 N. L. Anderson, R. Esquer-Blanco, J.-P. Hofmann, N. G. Anderson. *Electrophoresis* **1991**, *12*, 907–930.

77 S. Steiner, C. L. Gatlin, J. J. Lennon et al. *Electrophoresis* **2000**, *21*, 2129–2137.

78 S. Steiner, C. L. Gatlin, J. J. Lennon et al. *Toxicol. Lett.* **2001**, *120*, 369–377.

79 T. A. Pearson. *Circulation* **1996**, *94*, 3023–3025.

80 E. B. Rimm, A. Klatsky, D. Grobbee, M. J. Stampfer. *BMJ* **1996**, *312*, 731–736.

81 J. Constant. *Coron. Artery Dis.* **1997**, *8*, 645–649.

82 S. Gorinstein, A. Caspi, I. Goshev et al. *J. Agric. Food Chem.* **2003**, *51*, 822–827.

83 H. Schwertz, T. Längin, H. Platsch et al. Proteomics **2002**, *2*, 988–995.

84 K. Okamoto, Y. Yamori, A. Nagaoka. *Circ. Res.* **1974**, *33/34*, I143–I153.

85 U. Guerrini, L. Sironi, E. Tremoli et al. *Stroke* **2002**, *33*, 825–830.

86 L. Sironi, E. Tremoli, I. Miller et al. *Stroke* **2001**, *32*, 753–760.

4
Discovery of New Diagnostic Markers of Stroke

Laure Allard, Denis F. Hochstrasser, and Jean-Charles Sanchez

4.1
Introduction

A vascular cerebral accident, also called stroke or brain attack, is an interruption of the blood supply by blockage or rupture of a blood vessel to any part of the brain, resulting in damaged brain tissue. It has a devastating impact on public health and remains a leading cause of death and disability in industrialized countries. An early diagnosis of the cerebral accident associated with an appropriate treatment would reduce the risk of death and enhance the chances of recovery. Moreover, two different types of stroke exist – ischemic and hemorrhagic – with an occurrence of over 80% for the ischemic type. When the diagnosis of stroke is established, the physician needs to know the nature (ischemic or hemorrhagic), the extent, and the location of the accident in order to orientate and prescribe the most suitable treatment. As no specific and unique symptom as well as no early diagnostic marker or universal treatment is currently available, it is of major interest to develop new approaches in the research and discovery arena of new early diagnostic and prognostic markers of stroke. This review covers an overview of the pathogenesis of the stroke and the current diagnosis of stroke. Finally it also reviews the literature dedicated to the proteomic-based approaches for the discovery of stroke diagnostic markers.

4.2
Stroke Features

4.2.1
Brain Anatomy

With a mean weight of 1500 g and containing around 10 billion neurons, the adult brain represents about 2% of the total adult weight and requires 20% of the total energy. It consumes continuously 150 g of glucose and 72 L of oxygen every 24 hours. Interruption of this supply for just a few minutes can lead to dramatic brain damage. The manifestation and consequences of stroke depend on the location and extent of the lesions. It is thus of a great interest to understand the orga-

Biomedical Application of Proteomics
Edited by J.-C. Sanchez, G. L. Corthals, D. F. Hochstrasser

ISBN: 3-527-30807-5

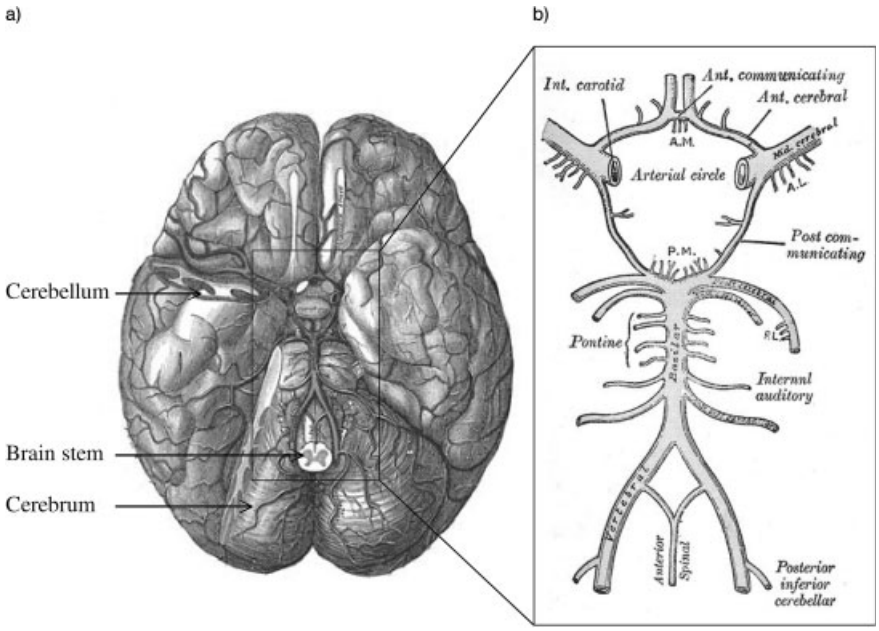


Figure 4.1 (a) The arteries of the base of the brain. The temporal pole of the cerebrum and a portion of the cerebellar hemisphere have been removed on the right side (http://www.yahooligans.com/reference/gray/146.html#i516). (b) Diagram of the arterial circu-lation at the base of the brain. A.L., antero-lateral; A.M., antero-medial; P.L., postero-lateral; P.M., posteromedial ganglionic branches (http://www.yahooligans.com/reference/gray/147.html). Reproduced with permission from [58].

nization of the brain. The three main components of the brain – the cerebrum, the cerebellum and the brainstem – are shown in Figure 4.1 a.

The cerebrum, the most important part of the brain, is divided into a right and a left hemisphere. The left hemisphere controls the majority of functions on the right side of the body, while the right hemisphere controls most of the functions on the left side of the body. The crossing of nerves takes place in the brainstem. Thus, injury to the left cerebral hemisphere produces sensory and motor deficits on the right side and vice versa. The cerebrum is responsible for functions such as initiation and coordination of movement, temperature, touch, vision, hearing, judgment, reasoning, problem solving, emotions, and learning.

The cerebellum is located at the back of the head. Its function is to coordinate voluntary muscle movements and to maintain posture, balance, and equilibrium.

The brainstem is responsible for movement of the eyes and mouth, relaying sensory messages, hunger, respirations, consciousness, cardiac function, body temperature, involuntary muscles movements, sneezing, coughing, vomiting, and swallowing.

Thus, according to the location of the lesion, the patient will display different disabilities.

4.2.2
Cerebrovascular Blood Circulation

Each hemisphere is supplied by one internal carotid artery, which originates from the common carotid artery, and by two vertebral arteries. The internal carotid artery and the vertebral arteries anastomose at the base of the brain to form the circle of Willis (Fig. 4.1 b). After the internal carotid arteries enter the skull from each side, they each trifurcate into the anterior cerebral artery, middle cerebral artery, and posterior communicating artery. The vertebral arteries, entering the skull posteriorly, join to form the basilic artery. The basilic artery will soon bifurcate again into the right and left posterior cerebral arteries. The posterior communicating arteries join the posterior cerebral arteries, thereby completing the circle of Willis (see websites http://science.nhmccd.edu/biol/cardio/willis.htm and http://pathology.mc.duke.edu/neuropath/nawr/blood-supply.html). Blood is then drained from the brain through a network of sinuses that drain into the right and left internal jugular veins.

The location of infarcts may be established after subdivision of the posterior circulation into proximal, middle, and distal intracranial posterior territories [1]. The proximal circulation territory correspond to regions supplying mainly the cerebellum *via* intracranial vertebral arteries, the middle circulation territory comprises the brain supplied by the basilar artery, and the distal intracranial posterior circulation supplies the midbrain and the thalamus *via* mainly posterior cerebral arteries and derived penetrating branches.

4.2.3
Aetiology and Pathology of Stroke

A stroke is an interruption of the blood supply to any part of the brain, resulting in damaged brain tissue due to hypoperfusion and hypoxia. Cell death is secondary to a failure of energy production. This definition implies (a) a parenchymal lesion, which can be ischemic or hemorrhagic, and (b) a vascular lesion responsible for the accident. Whatever the type of stroke, the location of the lesion determines the type of symptoms and the patient's overall outcome.

Many factors may influence the occurrence of a stroke. The main risks factors are hypertension, transient ischemic attacks (TIAs), hyperlipidemia, diabetes mellitus, tobacco, excessive alcohol and drugs, atherosclerosis, heart disease, disorders in hemostasis, genetic factors, physical inactivity, and obesity [2].

4.2.4
Pathophysiology – Different Types of Stroke

Two main different types of stroke exist according to the etiology of the lesion: ischemic or hemorrhagic [3]. Ischemic strokes represent more than 80% of total strokes and are divided into thrombotic and embolic strokes according to the location of the blood clot. In an atherothrombotic stroke, blood flow is impaired by a

blood clot (also called thrombus) in one or more arteries supplying blood in the brain. An embolic stroke occurs when a blood clot is formed somewhere in the body (usually the heart) and travels through the bloodstream to the brain. Once in the brain, the blood clot (also called embolus) eventually blocks a small vessel, causing the ischemic stroke.

TIAs are "small strokes" that produce stroke-like symptoms during less than 24 hours but no lasting damages (10% of total cerebrovascular accidents). They are extremely important predictors of strokes.

An intracerebral hemorrhage is the result of the rupture of a vessel within the brain parenchyma. These represent 10–20% of total strokes. The primary causes of these ruptures are hypertension and amyloid angiopathy. Secondary precipitating factors are aneurysms, arteriovenous malformations, neoplasms, trauma, anticoagulation, use of thrombolytics, or hemorrhagic conversion of ischemic stroke. Among the hemorrhagic strokes we can distinguish (a) parenchymal hemorrhages (or ICH for intracerebral hemorrhage), which represent 10–15% of hemorrhages and are mainly caused by hypertensive arterial disease, alcoholism, or amyloid angiopathy, and (b) subarachnoid hemorrhages (SAHs), which represent around 5% of hemorrhages and occur when a blood vessel on the brain's surface ruptures and bleeds into the space between the brain and the skull (but not into the brain itself). SAHs are mainly caused by arteriovenous malformation or aneurysm (80% of cases) and, like TIAs, are not considered to be strokes. In contract, a cerebral hemorrhage occurs when a defective artery bursts in the brain.

The vascular lesion may be of two types: large vessel disease or small vessel disease (or lacunar infarction) [4]. Arterial lesions are most frequent and are due to malformations (aneurysm, angioma), alteration of the arterial surface (arterial hypertension, atheroma) or amyloid angiopathy (in the elderly). Small vessel lesions result from occlusion of arteries or arterioles in the brain.

4.2.5 Epidemiology

Stroke is the third highest cause of death in industrialized countries, after ischemic heart disease and cancer, and it is the commonest cause of adult disability [5]. About 700 000 Americans each year have a new or recurrent stroke. On average, a stroke occurs every 45 seconds. Stroke kills nearly 150 000 people a year in the USA, which corresponds to about 1 of every 14 deaths. Stroke is more common in men than women in most age groups. However, of every five deaths from stroke, two are men and three are women. This is because the average life expectancy for women is greater than for men, and the highest rate for stroke is in the oldest age group [6, 7]. Black populations in the US have almost twice the risk of first-ever stroke compared with white populations. Almost the same results have been observed in western Europe [8] (see website of the American Heart Association http://www.americanheart.org/presenter.jhtml?identifier=1928).

4.2.6
Treatment

As the majority of strokes are due to obstruction of an artery in the brain by a blood clot, most treatments use thrombolytic drugs, such as urokinase and streptokinase, for the first-generation agents and recombinant pro-urokinase and recombinant tissue plasminogen activator (rtPA) for the second-generation products. rtPA is the first drug shown to be effective for acute ischemic stroke, but only when it is administrated within the first 3 hours of symptoms [9]. It is also the first FDA-approved acute treatment for ischemic stroke. A third generation of thrombolytic agents is now available, among them tenecteplase, reteplase, and lanoteplase [10, 11]. Nevertheless, such therapies increase death within the first 7–10 days, mainly due to intracerebral hemorrhage (for review of 17 trials, see [12]). These risks are offset by a reduction in disability in survivors, so that there is, overall, a significant net reduction in the proportion of patients dead or dependent in activities of daily living. Antiplatelets such as aspirin have been found to be of significant benefit during the acute phase of stroke [13, 14], unlike the heparin anticoagulant [14].

Representing about 20% of all strokes, intracerebral hemorrhagic stroke is more likely to result in death and major disability, but few investigations have been published compared with those on ischemic stroke or SAH [15]. Some randomized trials for medical therapies have been performed, using dexamethasone [16], intravenous glycerol [17], or steroid *versus* placebo. None of them showed significant benefit and the treatment of intracerebral hemorrhagic stroke seems to remain empirical. The benefit/risk of surgery treatment of ICH still displays insufficient evidence [18, 19].

4.3
Current Diagnosis of Stroke

The diagnosis of stroke is sometimes difficult to establish due to the lack of an early specific marker. Nevertheless, a rapid assessment and an early intervention may improve the prognosis. Many investigations tend to organize stroke management [20] and obviate, via informatics algorithms, the diagnosis [21–23]. The diagnosis takes into account several analyses, including physician evaluation associated with stroke scores such as the National Institutes of Health (NIH) Stroke Scale [24] or Glasgow scale, CT scanner, and/or magnetic resonance imaging (MRI). According to the results of these medical assessments, the final diagnosis will exclude any other brain disorders such as meningitis, cancer, migraine, viral infection, or metabolic disorders.

4.3.1 Physician's Evaluation

The first step for a contemporary diagnosis of stroke is the physician's evaluation. Indeed, many symptoms may orientate the physician to establish the diagnosis. Alteration in consciousness [25], headache, aphasia, facial weakness, or asymmetry are common manifestations of the vascular cerebral accident. Loss of coordination, weakness, paralysis, or sensory loss of one or more limbs is usually observed, as well as ataxia, visual loss, and other symptoms such as intensive vertigo, double vision, unilateral hearing loss, nausea, vomiting, photophobia, or phonophobia. Stroke scores are determined on the patient's arrival to evaluate the extent and the location of the lesions.

4.3.2 Imaging

Imaging has seen major advances in the field of cerebrovascular diseases [26–28]. Computed tomography (CT) scanning uses X-ray to generate an image of the brain. It is mainly used to identify hyperdense images in white or deep gray matter early after hemorrhagic stroke. Such a picture is the major exclusion criterion for antithrombolytic treatments. In contrast, CT scan parenchymal hypodensity is representative of an ischemic stroke, but the sensitivity remains low in the 24 hours after onset of symptoms. MRI has a very good sensitivity in diagnosing an ischemic stroke. MRI uses a magnetic field to detect any modification in water content and it is the best technique to visualize edema. Nevertheless, MRI generally fails in detecting an hemorrhagic stroke in the first 3 days. Two new imaging techniques, MR diffusion-weighted and MR perfusion-weighted imaging (DWI and PWI), permit accurate, reliable diagnosis and characterization of ischemic stroke within the critical first 6-hour time period needed to initiate thrombolytic therapy [29]. DWI and PWI allow rapid assessment of the underlying pathophysiology in acute ischemic stroke. Positron emission tomography (PET) is the 'gold standard' technique for the investigation of cerebral disorders. It allows physicians to measure the body's abnormal molecular cell activity, but it cannot be routinely used because of its cost.

4.3.3 Lumbar Puncture

A lumbar puncture is often performed when an SAH is suspected. In that case, the cerebrospinal fluid (CSF) displays xanthochromy instead of being clear.

4.3.4 Biochemical Markers of Stroke

If the diagnosis of stroke still relies on clinical evaluation reinforced by neuroimaging, some early ischemic changes may be absent or subtle on CT scans. Despite their lack of sensitivity and/or specificity, some biochemical markers are currently used as indicators of brain damage [30, 31]. Among them, neuron-specific enolase (NSE) [32], S-100β protein [33], myelin basic protein (MBP) [34], creatine phosphokinase isoenzyme BB (CK-BB) [35] for the commonest, but also tau [36] thrombomodulin (Tm), adenylkinase, lactate, lactate deshydrogenase, aspartate aminotransferase, glutathione, vasointestinal neuropeptide, and 7B2 specific neuropeptide [31]. In 1997, Fassbender et al. [37] quantified S-100 and NSE proteins in peripheral blood at the acute and subacute phase of ischemic stroke. Blood was collected from 24 patients at hours 4, 8, 10, and 72 after onset of symptoms. S-100 was detected in 17 out 24 stroke patients and was correlated with extent of the infarction. It was not detected in the matched controls. In contrast, release of NSE in the blood displayed no significant association to the outcome or extent of brain damage. Hill et al. [30] demonstrated on 28 patients of mean age 65 years that at admission, elevated levels of NSE were found in 89% of patients, Tm in 43%, MBP in 39% and S-100β in 32%.

Recently, Cao et al. [38] investigated whether elevated C-reactive protein (CRP) is a risk factor for ischemic stroke, independent or not of carotid intima-media thickness (IMT). Indeed, elevated CRP and increased carotid arteria IMT are both associated with the occurrence of stroke. The statistical investigation relied on 5417 participants recruited for the Cardiovascular Health Study. Authors have demonstrated that elevated CRP was an independent risk factor for future ischemic stroke and that CRP and IMT were closely associated; the higher the CRP, the greater the carotid atherosclerosis as measured by carotid IMT.

4.4 Proteomic-based Approach for the Discovery of Early Diagnostic Stroke Markers

New imaging technologies display considerable advantage in identifying and localizing stroke lesions with good sensitivity. However, the cost and the time required to perform imaging, as well as the difficulty in analyzing the results, leaves plenty of scope for the development of alternative diagnostic tests. There is little in the literature on the specific subject of a proteomic approach towards the discovery of early diagnostic markers of stroke. Although no good plasma or CSF markers are currently available for routine diagnosis of stroke, some proteins have proved of interest for complementary and earlier diagnosis of stroke as detailed below. Their lack of specificity and/or sensitivity does not allow any of them to be used as a single specific marker and the association of these potential markers to clinical outcome and extent of brain damage remains to be proved. The main objective of basic research is to discover new diagnostic, monitoring, and prognosis markers

of stroke, ideally displaying the following features: (1) it should be brain specific, (2) it should be a plasma marker, plasma samples being easier to obtain than CSF, (3) its expression and detection in plasma samples should be as early as possible after symptoms onset, (4) its expression level should reflect the extension of the lesions, (5) it should have a prognosis value, and (6) it should additionally be able to differentiate between transient and ischemic stroke as well as between ischemic and hemorrhagic.

All these aims should be solved using proteomic approaches. Indeed, as proteomics is the investigation of proteins expressed by a genome in a given tissue or cell at a given time, differences in specific protein levels should reflect as early as possible the nature, extent, and location of a specific disorder. Hence, comparative proteomics is currently the major strategy used to investigate diseases [39, 40].

The main proteomic techniques commonly used to answer biomedical questions are one- and two-dimension gel electrophoresis (or SDS-PAGE for sodium dodecyl sulfate-polyacrylamide gel electrophoresis) as well as surface-enhanced laser desorption ionization (SELDI) followed by matrix-assisted laser desorption ionization/time-of-flight mass spectrometry (MALDI-TOF-MS). This procedure represents a powerful tool in the discovery of diagnostic markers for brain-associated disorders [41].

4.4.1 One-dimensional Gel Electrophoresis

One-dimensional gel electrophoresis (1-DE) allows separation of proteins according to their molecular weights. It is widely used to compare protein patterns but also protein modifications. Lu et al. [42] have investigated the effect of focal ischemia of the brain on the content of specific proteins in the damaged tissue following unilateral occlusion of the middle cerebral artery in rats. They compared ischemic brain proteins pattern with controls by SDS-PAGE. Two proteins, at 66 and 80 kDa, appeared significantly increased in ischemic tissues whereas a 260 kDa component was markedly reduced in ischemic tissues. Western blot using antitransferrin antibody identified the 80 kDa component as transferrin and the 66 kDa protein displayed features similar to those of albumin. The authors did not give any complementary information concerning these identifications. Validation of this study would be welcome to reinforce their results.

The etiology of stroke appears to be a multifactorial disorder and one of the factors is genetic. Lounes et al. [43] reported the case of a man diagnosed designated as having "fibrinogen Alès" with a history of two thrombotic strokes before the age of 30. His plasma thrombin clotting time was dramatically prolonged and no clotting time was observed with reptilase (an enzyme derived from a venom, which clots fibrinogen). SDS-PAGE of purified fibrinogen Alès indicated a higher mobility for the pathologic γ-chain than that of the normal γ-chain, indicating that the structural defect was located in the γ-chain. Using molecular biology, the authors demonstrated that it was homozygous for a single base substitution in the codon γ-Asp330, converting this residue to a valine. The difference in appar-

ent molecular weight was unexpected, but has been ascribed to an alteration in hydrophobicity or a local conformation that then affects the mobility of the protein/SDS complex. In this case, electrophoresis has helped to establish a diagnosis of a congenital homozygous dysfibrinogenemia (γ-Asp330 → Val), characterized by a defective fibrin polymerization site "a".

Atherosclerosis is an arterial disease in which raised areas of degeneration and cholesterol deposit forms the inner surfaces of arteries. This fatty material thickens, hardens and may eventually rupture or impede the arteries. It is one of the main causes leading to vascular diseases such as myocardial infarct and stroke. Identification of atherosclerosis-specific molecules in plasma samples could be of great interest in the prediction and diagnosis of stroke. In a recent paper, Matuszek et al. [44] hypothesized that proteins eluted from the arterial wall may discriminate between atherosclerotic and non-atherosclerotic coronary arteries. They analyzed perfusates of 30 patients with and 7 patients without atherosclerosis during cardiac bypass surgery. They consistently observed a 40 kDa band by one-dimensional SDS-PAGE. This 40 kDa band was characterized by Edman degradation after electrophoretic transfer and N-terminal sequencing. It revealed a 100% homology with the haptoglobin (Hpt) β-chain. Western blot performed on all the fractions confirmed that Hpt was present in coronary perfusate fractions. To account for its normal presence in blood, Hpt was quantified relative to red cell hemoglobin and its concentration was significantly greater in eluent from coronary arteries with than without extensive atherosclerotic ($P=0.0027$ after correction for blood contamination). Nephelometric analysis performed on the plasma of the same patients before operation validated this result. It was possible to distinguish patients with and without atherosclerosis ($P=0.002$). This study validates haptoglobin as a good indicator of arterial pathology and appears to discriminate between minor and extensive coronary atherosclerosis.

In our laboratory, haptoglobin levels were also evaluated in plasma samples by nephelometry. No significant difference between a pool of stroke plasma samples (37 different patients) and a pool of control plasma samples (20 different patients) was observed (P. Lescuyer, per. commun.).

Platelet hyperactivity has been found to be a hallmark of both thrombotic and hemorrhagic stroke. Srivastava and Dash [45] have isolated platelets from the blood of thrombotic stroke patients by differential centrifugation. Platelet activation/aggregation is associated with phosphorylation on tyrosine residues. Washed platelets were run on one-dimensional gel electrophoresis and transferred onto polyvinylidene difluoride (PVDF) membranes. Then the authors analyzed the profile of tyrosine-phosphorylated proteins in the platelets from cerebrovascular ischemia using antiphosphotyrosine antibodies. They showed that proteins from non-stimulated platelets (with any aggregation inducer such as thrombin receptor-activating peptide (TRAP) or calcium ionophore) with the relative mobilities of 125, 115, 80, 35, and 30 kDa were found to be significantly more tyrosine phosphorylated in 85% of the 20 stroke patients studied than their healthy counterpart. Besides that, they found extra tyrosine-phosphorylated bands at 131, 100, 47, and 38 kDa regions in the platelets from stroke, which were absent in the resting platelets from the control patients.

In conclusion, hyperactivity of the platelets in thrombotic stroke was demonstrated to be associated with increased membrane viscosity, calpain activation, and higher phosphotyrosine content of the platelet proteins subject to abnormal signal transduction in these cells. All these could contribute to the etiopathogenesis of stroke. Monitoring the profile of tyrosine-phosphorylated proteins may have prognosis potential for the detection of cerebrovascular insufficiencies.

4.4.2 Two-dimensional Gel Electrophoresis

Two-dimensional gel electrophoresis (2-DE) has been widely used to describe protein patterns of healthy *versus* pathological fluids or tissues [46]. This technology allows proteins to be separated, in a first step according to their net charge, and in a second step according to their molecular weight. Krapfenbauer et al. [47] have recently published a rat brain protein expression map including cytosolic and enriched mitochondrial and microsomal fractions. They produced enriched cellular subfractions of brain homogenates by ultracentrifugation and pre-electrophoretic chromatographical separation (ion-exchange chromatography) to display underexpressed proteins. Brain homogenates were first analyzed by 2-DE and protein spots were excised and identified by MALDI-TOF-MS. Pre-electrophoretic fractionation of the cytosolic fractions allowed identification of 437 proteins specific to the subcellular compartment. This work validates the fractionation of brain tissue, while the mapping provides a powerful tool for neuroscientists.

Enrichment of low-abundance brain proteins by preparative electrophoresis was also described by Fountoulakis and Juranville [48]. Fifty milligrams of rat brain cytosolic proteins were fractionated over a cylindrical 11% acrylamide gel electrophoresis (PrepCell system from Bio-Rad) in lithium dodecyl sulfate (LDS) (which is easier to remove from samples than SDS and does not interfere with 2-DE). Preparative electrophoresis separates proteins on the basis of their size. Some selected fractions, among the 80 10-mL collected fractions, were concentrated by ultrafiltration and analyzed onto 2-D gel electrophoresis. About 4000 spots were excised from 13 selected 2-D gels, leading to the identification by MALDI-TOF-MS of about 1000 proteins, which were the products of 207 different genes. This fractionation technology provides an interesting tool to overcome the problem of low-abundance proteins.

As described above, atherosclerosis is a leading cause of vascular accident and several groups have developed 2-DE comparative proteomic experiments in an attempt to find markers of vessel damage due to atherosclerosis plaques. Duran et al. [49] investigated secreted proteins, i.e. supernatant of cultured artery segments. They studied three types of segments: normal artery segment, non-complicated plaques, and complicated plaques with thrombus. Forty-two spots were obtained for normal artery segments, 154 for segments bearing a plaque, and 202 for artery segments with ruptured plaque and thrombus. Spots of interest were excised from the gel, digested and analyzed by MALDI-TOF-MS. Proteins secreted by human carotid artery segments containing a non-complicated plaque are involved in

cholesterol transport (apolipoprotein B-100 and apolipoprotein A-1) or participate in the elimination of toxic radicals (superoxide dismutase and peroxiredoxin). Their presence in the supernatant from artery segments containing non-complicated plaques is easily understood, unlike some others such as β-galactosidase or ubiquitin. These results clearly demonstrated that the number of secreted proteins is directly proportional to the severity of the lesion, suggesting the production of specific proteins relating to the complexity of the atherosclerotic lesion.

In another study aimed at finding biomarkers of coronary atherosclerosis, You et al. [50] analyzed proteins extracted from 10 diseased and 7 normal coronary arteries, through protein delipidation and precipitation. Two-dimensional silver-stained gels containing proteins of the diseased tissues were compared with those of normal tissues by direct visualization. A protein spot of interest with a consistently higher expression level in the diseased arteries was cut out from the gel and analyzed by MS. Ferritin light chain was identified and its expression level was determined using statistical imaging software. It appeared that ferritin light chain expression was 1.9 higher in the diseased arteries. This result was confirmed by western blot as well as by reverse transcription-polymerase chain reaction (RT-PCR) in diseased and normal coronary arteries. As ferritin light chain protein mediates storage of iron in cells, this study reinforces the theory of an association between excessive iron storage and risk of coronary artery disease.

A proteomic approach was reported by Sironi et al. [51] to describe acute-phase proteins before ischemia in stroke-prone rats in serum and urine. They used spontaneously hypertensive stroke-prone rats (SHRSP) subjected to salt loading. These inbred animal models display a complex form of cerebrovascular pathology resembling, in many aspects, human stroke disease. MRI and histology monitored brain alterations and 1-D gels were performed to assess the relative concentrations of the major components. Proteinuria (determined using Bradford) appeared to be strongly increased immediately before stroke and declined thereafter. Two-dimensional gel electrophoresis of the urine indicated the presence of transferrin, hemopexin, albumin, α_2-HS-glycoprotein, kallikrein-binding protein, α_1-antitrypsin, Gc-globulin, and transthyretin. Markers of an inflammatory response, including very high levels of thiostatin, were detected in the serum of SHRSPs at least 4 weeks before a stroke occurred. Salt loading in SHRSPs induces an atypical inflammatory condition as well as alteration of the vascular permeability developed before the appearance of brain damage detected by MRI. The urinary concentration of each of the excreted serum proteins correlated positively with time before stroke occurred. This cartography of proteinuria could be a good tool for the prevention and/or treatment of brain disease. Proteins maps of serum, urine, and CSF from three different rat strains were also published [52].

Zimmermann-Ivol et al. [53] have used 2-DE to model major brain damage by comparing the CSF of deceased patients with non-pathological CSF. Two protein spots came out overexpressed in the CSF map of deceased patients as compared to healthy subjects (Fig. 4.2). These spots were identified as epidermal and adipocyte fatty acid-binding proteins (FABPs) by MALDI-TOF-MS and tandem MS sequencing. It established FABPs as potential markers of brain damage to be

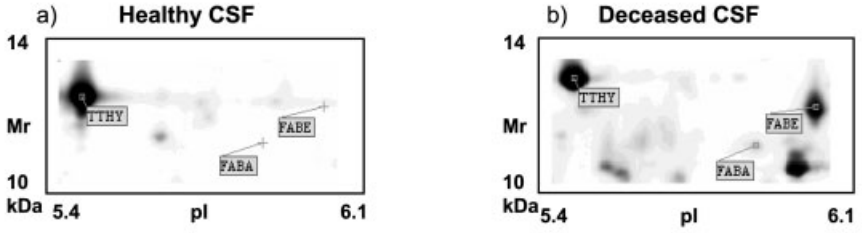


Figure 4.2 Enlargement of 2-DE maps of CSF from (a) healthy and (b) deceased subjects. Forty-five micrograms of protein were loaded on an immobilized pH gradient (IPG) gel (pH 3.5–10 NL, 18 cm). The second dimension was a vertical gradient slab gel (9–16% T), stained with ammoniacal silver [53].

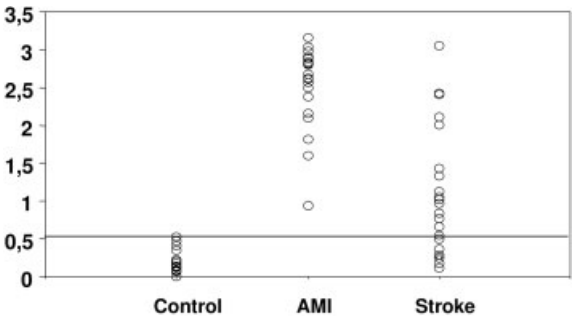


Figure 4.3 Optical densitometry measurement of H-FABP expression in the control, the acute myocardial infarction (AMI) and the stroke groups. The cut-off value was established at 0.531 of OD [53].

further assessed for the diagnosis of stroke. ELISA was performed to validate heart FABP (H-FABP) as a potential plasma marker of stroke on samples from 22 stroke patients and 22 controls (Fig. 4.3). The results were 68% sensitivity (15 out of 22 stroke patients above cut-off), 100% specificity, 100% positive predictive value, and 75.9% negative predictive value. Among the 7 false negative results, 2 had a rapid and complete recovery of their neurological deficit within 24 hours, consistent with a TIA, 2 had a lacunar stroke on MRI imaging and one was located in the brainstem. The performance of the H-FABP was compared with that of two other potential markers of stroke, NSE and S-100B proteins. Sensitivity and specificity were found to be much lower with these two proteins (55% sensitivity and 36.5% specificity for NSE, 15% sensitivity and 95% specificity for S-100B). According to this study, H-FABP appears to be the first valid plasma biomarker for early diagnosis of stroke.

4.4.3 SELDI-TOF

SELDI-TOF (surface-enhanced laser desorption ionization/time-of-flight from Ciphergen, Fremont, CA, USA) is a proteomic technology that allows proteins to be captured onto different ProteinChip® arrays (ion exchange, normal phase, immobilized metal affinity chromatography (IMAC)) and subsequently analyzed by laser desorption/ionization TOF-MS [54]. Compared with electrophoresis, SELDI-TOF allows separation of low molecular weight proteins and peptides, which could be of interest in the field of diagnosis. Alzheimer's disease and cancer are the most documented areas using SELDI-TOF technology.

Mannes et al. [55] previously identified cystatin C as a protein whose expression was induced by peripheral inflammation. Cystatin C is a small (13 kDa), secreted cysteine protease inhibitor, which may be involved in controlling proteolytic modifications of the extracellular matrix during periods of synaptic activity or neuronal stress. CSF samples of 2 μL from subjects (five female patients at term in active labour) and controls (five pregnant patients at term who were scheduled for a caesarean) were spotted on reverse phase and normal phase SELDI chips. Three peaks (around 12 kDa, at 12.8 and 13.7 kDa) appeared consistently elevated in the labour pain patients. The 12.8 kDa matches the cystatin C peak. When the area under the curve of the 12.8 kDa peaks was quantified, an increase of 394% for the labour pain patients compared with controls was observed, consistent with immunoassay results.

This result was confirmed by antibody capture followed by SELDI-TOF. Protein G was covalently attached to the surface of a pre-activated PS2 chip. Anti-cystatin antibody preparation was incubated with the derivatized protein G chip, then in-

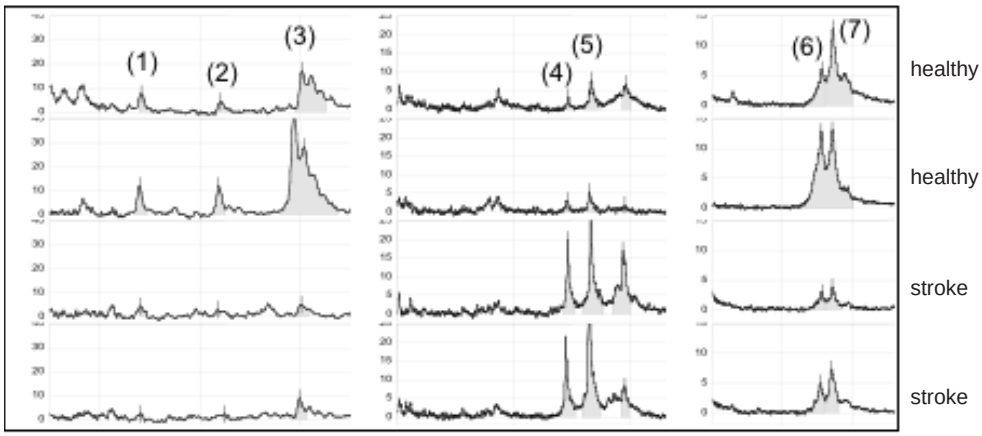


Figure 4.4 SELDI spectra of two of the 21 healthy controls and two of the 21 stroke plasma samples analyzed. Numbers indicated peaks differentially expressed between controls and stroke patients ($P < 0.05$).

cubated with CSF to bind cystatin C and finally analyzed by SELDI-TOF. This experiment confirmed the presence of cystatin C with a higher level in the CSF of pain patients. This technology has been widely used in our laboratory to discover new CSF and plasma diagnostic markers of neurodegenerative disorders such as Alzheimer's disease [56], Creutzfeldt-Jakob disease as well as stroke.

A pilot study was performed with 21 healthy and 21 stroke plasma samples. Crude plasma samples were directly applied on a strong anonic-exchange (SAX) ProteinChip® and analyzed with a SELDI-TOF (PBS2 from Cyphergen). Using the Mann-Whitney statistical tool (Fig. 4.4), seven peaks were found to be significantly ($P < 0.05$) differentially expressed. Further investigations are needed to identify and validate these new polypeptides as potential markers of stroke. A similar study is underway to differentiate ischemic and hemorrhagic strokes.

4.5 Conclusions

Many investigations have confirmed the release of brain proteins in the blood after stroke [30], which reinforces the idea of the existence of one or more specific stroke biomarkers. However, none of the plasma biomarkers currently described in the literature are used routinely by physicians as they display poor sensitivity and specificity. H-FABP, reported by Zimmermann-Ivol [53], is the first one with 100% specificity and is of major interest in avoiding unnecessary treatment of "healthy" patients. Nevertheless, because the etiology of stroke is multisystemic, involving mechanisms of thrombotic and neurotoxic coupling, diagnosis is also probably multifactorial. To overcome this problem Dambinova et al. [57] have proposed the use of a panel of biochemical markers. They assessed three biomarkers that are independently associated with neurotoxicity and that can be measured in blood. Glutamate and homocysteine correlated with large and middle artery dysfunction, and the presence of *N*-methyl-D-aspartate (NMDA) receptor autoantibodies (aAb) was a criterion of microvascular damage independently associated with neurotoxicity and thrombosis in patients with TIA/stroke. They demonstrated that simultaneous assessment of these three biomarkers allowed neurotoxicity and thrombosis to be correlated with severity of cerebral ischemia. This result confirms the finding that such complex neurodegenerative disorders should be diagnosed using several biomarkers simultaneously.

Using SELDI-TOF technology, Carrette et al. [56] determined five differentially expressed peaks for CSF samples from nine Alzheimer's patients and ten controls. The combination of the five polypeptides for the diagnosis of Alzheimer's disease allowed 6 out of the 9 Alzheimer's patients and all 10 controls to be classified, giving 100% specificity and 66% sensitivity.

Identification of differentially expressed proteins associated with cerebral vascular accident has several long-term objectives. First, it should allow clinicians to apply preventive strategies to patients with a high risk for stroke. Second, the discovery of proteins directly or indirectly involved in the disease conditions should lead

to a better understanding of the pathogenic processes. Third, the establishment of appropriate protein markers should permit the precise classification of stroke patients in subgroups for further, more effective targeted therapies. Finally, it should allow the evaluation of new potential therapies. In conclusion, proteomic approaches should provide in the near future new clues for the elucidation of the pathogenesis of the various defects involved in stroke and provide major insights into new more effective diagnostic and prognostic markers.

4.6 Acknowledgements

The authors are grateful to Pierre Burkhard and Denis Hochstrasser for the precious scientific discussions in neurosciences. We would like to thank the Biomedical Proteomic Group, Catherine Zimmermann, and Pierre Lescuyer. Finally, the authors are very grateful to Proteome Sciences plc both for their financial support and for their patience in the stroke project.

4.7 References

1 L. Caplan. *Stroke* **2000**, *31*, 2011–2023.
2 T.A. Pearson, S.N. Blair, S.R. Daniels et al. *Circulation* **2002**, *106*, 388–391.
3 A. Rudd, C. Wolfe. *Hosp. Pharmac.* **2002**, *9*, 32–36.
4 J. Bogousslavsky, M.G. Bousser, J.L. Mas. *Accidents vasculaires cérébraux*. Doin, Rueil-Malmaison, **1993**, 683.
5 A.M. Minino, E. Arias, K.D. Kochanek, S.L. Murphy, B.L. Smith. *Natl Vital Stat. Rep.* **2002**, *50*, 1–119.
6 J. Roquer, A.R. Campello, M. Gomis. *Stroke* **2003**, *34*, 1581–1585.
7 M.G. Bousser. *Circulation* **1999**, *99*, 463–467.
8 C. Sarti, D. Rastenyte, Z. Cepaitis, J. Tuomilehto. *Stroke* **2000**, *31*, 1588–1601.
9 *N. Engl J. Med.* **1995**, *333*, 1581–1587.
10 A.R. Xavier, A.M. Siddiqui, J.F. Kirmani, R.A. Hanel, A.M. Yahia, A.I. Qureshi. *CNS Drugs* **2003**, *17*, 213–224.
11 A.I. Qureshi, R.U. Pande, S.H. Kim, R.A. Hanel, J.F. Kirmani, A.M. Yahia. *Curr. Opin. Invest. Drugs* **2002**, *3*, 1729–1732.
12 J.M. Wardlaw, G. del Zoppo, T. Yamaguchi. *Cochrane Database Syst. Rev.* **2000**, CD000213.
13 P. Sandercock, G. Gubitz, P. Foley, C. Counsell, *Cochrane Database Syst. Rev.* **2003**, CD000029.
14 *Lancet* **1997**, *349*, 1569–1581.
15 J.P. Broderick, H.P. Adams, Jr., W. Barsan et al. *Stroke* **1999**, *30*, 905–915.
16 N. Poungvarin, W. Bhoopat, A. Viriyavejakul et al. *N. Engl. J. Med.* **1987**, *316*, 1229–1233.
17 Y.L. Yu, C.R. Kumana, I.J. Lauder et al. *Stroke* **1992**, *23*, 967–971.
18 G.J. Hankey, C. Hon. *Stroke* **1997**, *28*, 2126–2132.
19 H.M. Fernandes, B. Gregson, S. Siddique, A.D. Mendelow. *Stroke* **2000**, *31*, 2511–2516.
20 O. Benavente, R.G. Hart. *Am. Fam. Physician* **1999**, *59*, 2828–2834.
21 M. Camerlingo, L. Casto, B. Censori et al. *Neurol. Sci.* **2001**, *22*, 357–361.
22 D.L. Tirschwell, W.T. Longstreth, Jr. *Stroke* **2002**, *33*, 2465–2470.
23 A. Al-Wabil, M.A. Smith, L.A. Moye, W.S. Burgin, L.B. Morgenstern. *J. Clin. Epidemiol.* **2003**, *56*, 351–357.
24 D.L. Tirschwell, W.T. Longstreth, Jr., K.J. Becker et al. *Stroke* **2002**, *33*, 2801–2806.

25 B. Cucchiara, S.E. Kasner, D.A. Wolk et al. *J. Neurol. Neurosurg. Psychiatry* **2003**, *74*, 889–892.
26 J.V. Guadagno, C. Calautti, J.C. Baron. *Br. Med. Bull.* **2003**, *65*, 145–157.
27 J.J. Smith, A.G. Sorensen, J.H. Thrall. *Radiology* **2003**, *227*, 633–638.
28 S. Warach. *Stroke* **2003**, *34*, 345–347.
29 P.E. Ricci. *Colorado Neurol. Inst. Rev.* **2000**, *11*, http://www.thecni.org/reviews/11–2-p13-ricci.htm
30 M.D. Hill, G. Jackowski, N. Bayer, M. Lawrence, R. Jaeschke. *CMAJ* **2000**, *162*, 1139–1140.
31 P. Johnsson. *J. Cardiothorac. Vasc. Anesth.* **1996**, *10*, 120–126.
32 R.T. Cunningham, I.S. Young, J. Winder et al. *Eur. J. Clin. Invest.* **1991**, *21*, 497–500.
33 M. Takahashi, A. Chamczuk, Y. Hong, G. Jackowski. *Clin. Chem.* **1999**, *45*, 1307–1311.
34 J. Matias-Guiu, J. Martinez-Vazquez, A. Ruibal, R. Colomer, M. Boada, A. Codina. *Acta Neurol. Scand.* **1986**, *73*, 461–465.
35 G. Capocchi, C. Tassi, S. Ricci, M. Zampolini, R. Fausti, A. Rossi. *Ital. J. Neurol. Sci.* **1987**, *8*, 567–570.
36 A. Bitsch, C. Horn, Y. Kemmling et al. *Eur. Neurol.* **2002**, *47*, 45–51.
37 K. Fassbender, R. Schmidt, A. Schreiner et al. *J. Neurol. Sci.* **1997**, *148*, 101–105.
38 J.J. Cao, C. Thach, T.A. Manolio et al. *Circulation* **2003**, *108*, 166–170.
39 S. Hanash. *Nature* **2003**, *422*, 226–232.
40 R.E. Banks, M.J. Dunn, D.F. Hochstrasser, J.C. Sanchez, W. Blackstock, D.J. Pappin, P.J. Selby. *Lancet* **2000**, *356*, 1749–1756.
41 M. Fountoulakis. *Amino Acids* **2001**, *21*, 363–381.
42 A. Lu, H. Yu, K. Chen, S.S. Koide, X. Li. *Life Sci.* **1999**, *65*, 493–500.
43 K.C. Lounes, C. Soria, S.S. Mirshahi et al. *Blood* **2000**, *96*, 3473–3479.
44 M.A. Matuszek, L.P. Aristoteli, P.G. Bannon, et al. *Atherosclerosis* **2003**, *168*, 389–396.
45 K. Srivastava, D. Dash. *Mol. Cell. Biochem.* **2001**, *224*, 143–149.
46 T. Marshall, K.M. Williams. *Electrophoresis* **1991**, *12*, 461–471.
47 K. Krapfenbauer, M. Fountoulakis, G. Lubec. *Electrophoresis* **2003**, *24*, 1847–1870.
48 M. Fountoulakis, J.F. Juranville. *Anal. Biochem.* **2003**, *313*, 267–282.
49 M.C. Duran, S. Mas, J.L. Martin-Ventura et al. Proteomics **2003**, *3*, 973–978.
50 S.A. You, S.R. Archacki, G. Angheloiu et al. *Physiol. Genomics* **2003**, *13*, 25–30.
51 L. Sironi, E. Tremoli, I. Miller et al. *Stroke* **2001**, *32*, 753–760.
52 R. Wait, E. Gianazza, I. Eberini, L. Sironi, M.J. Dunn, M. Gemeiner, I. Miller. *Electrophoresis* **2001**, *22*, 3043–3052.
53 C.G. Zimmermann-Ivol, P. Burkhard, J. Le Floch-Rohr et al. *Mol. Cell. Proteomics* **2003**, Oct 26 (Epub ahead of print).
54 H.J. Issaq, T.P. Conrads, D.A. Prieto, R. Tirumalai, T.D. Veenstra. *Anal. Chem.* **2003**, *75*, 148A–155A.
55 A.J. Mannes, B.M. Martin, H.-Y.T. Yang et al. *Pain* **2003**, *102*, 251–256.
56 O. Carrette, I. Demalte, A. Scherl et al. *Proteomics* **2003**, *3* (8), 1486–1494.
57 S.A. Dambinova, G.A. Khounteev, A.A. Skoromets. *Stroke* **2002**, *33*, 1181–1182.
58 H. Gray. *Anatomy of the Human Body.* Lea & Febiger, Philadelphia, **1918**, 1396.

Part III
Cancer

Biomedical Application of Proteomics
Edited by J.-C. Sanchez, G. L. Corthals, D. F. Hochstrasser

ISBN: 3-527-30807-5

5
Unravelling Biological Pathways and the Identification of Clinical Markers and Targets in Renal Cancer

Rosamonde E. Banks and Peter J. Selby

5.1
Renal Cancer – The Clinical Perspective

Renal cancer represents a relatively uncommon but important challenge in oncology. Globally it accounts for approximately 3% of all cancers and almost 100 000 deaths per annum. More common in North America and Northern Europe, it is rare in India, China, and Japan. In the United Kingdom it is the 10th commonest cancer in men (3600 new cases per year) and the 14th in women (2200 new cases per year). However the incidence is increasing, with a 126% increase in incidence rate since the 1950s in the USA [1] and a reported 22% increase in incidence rate in the last 10 years in females in the UK for example, the largest rise in incidence rate of any cancer [2]. The vast majority of renal cancers are of the epithelial type involving the parenchyma and are termed renal cell carcinoma (RCC).

5.1.1
Epidemiology

Environmental factors have been shown to be important in the development of RCC, with the most important being tobacco smoking with a relative risk of 1.5 [3, 4]. However, obesity [3, 5–7], reproductive hormones [8], drugs such as phenacetin, diuretics, and beta-blockers (although it is unclear whether antihypertensives themselves or the underlying hypertension are the actual risk factor) [9, 10] and exposure to some industrial chemicals such as trichloroethene, cadmium, asbestos, and petrochemicals [11–13] have also been implicated. Information about the influence of diet is far from clear-cut, with possible effects of vegetable and fruit intake, some vitamins and high intake of meat and milk [14–18]. Patients with end-stage renal disease with long-term hemodialysis have a higher risk of renal cancer with a nine times increased incidence [19].

Familial cases of clear cell RCC have been recognized in the autosomal dominant Von Hippel-Lindau (VHL) syndrome, but Mendelian inheritance only accounts for 1–3% of renal cancer cases [20, 21]. Mutations in fumarate hydratase [22] have also been described in association with papillary renal cancer and hereditary papillary RCC has been associated with mutations in the *MET* gene in most families [23,

Biomedical Application of Proteomics
Edited by J.-C. Sanchez, G. L. Corthals, D. F. Hochstrasser

ISBN: 3-527-30807-5

24]. The importance of genes on 3p and in particular the *VHL* gene on 3p25–26 is now well recognized in sporadic clear cell RCC with loss of 3p in 97% of cases and methylation and mutation of the remaining allele in up to 80% [25–32]. In a population-based familial aggregation analysis of 1078 RCC cases using an extensive genealogical database for the population of Iceland, risk of RCC was significantly higher for members of the extended family of an affected individual as well as the nuclear family, indicating that germline mutations are probably significantly involved in what has been previously regarded as sporadic renal cell carcinoma [33].

5.1.2
Current Clinical Approaches and Clinical Challenges

If diagnosed early, renal cancer is readily curable by radical surgery but late diagnosis is associated with large and clinically symptomatic masses arising from the kidney and a high risk of metastatic disease. Diagnosis is usually initially made by an imaging test of the abdomen, most frequently an ultrasound scan. Over half of patients commonly present with either locally advanced or distant metastatic disease. The most common clinical symptoms are hematuria and flank and back pain and there are a wide range of less specific symptoms including fatigue and anemia. The most common metastatic sites are to the lungs, bone, and liver, but rarely renal cancer can metastasize to almost any site [34]. Increasingly, the disease is made as an incidental finding on imaging tests carried out because of non-specific symptoms or other indications, with such patients tending to have early-stage small-volume tumors and a good prognosis.

Pathological classification of renal tumors plays an important part in their management. Although there has been controversy over the years, in 1997 an international convention defined a new classification system which is used in most major laboratories and which has been associated with characteristic cytogenetic findings [35–38] as shown in Table 5.1. The majority of tumors are of the clear cell type. The outcome for patients with renal cancer is very variable, with prognostic indicators being tumor size, the stage of evolution of the disease, the nuclear grade, and the presence/absence of vascular invasion [39]. The clinician therefore approaches each patient to identify on these grounds the likely outcome and therefore select the most appropriate therapy. Clinical staging is now carried out according to the modified UICC/AJCC TNM staging system [40] which is based on tumor size and extent of local invasion, nodal involvement, and presence or absence of distant metastases. These important tumor characteristics allow the patients to be placed into stage groupings which provide prognostic information. For example, patients with stage I disease which is a tumor confined to the kidney with no nodal involvement or metastatic spread have 5-year survival rates of 91–100%, compared with only 0–25% for patients with stage IV disease, which represents tumors that have invaded beyond Gerota's fascia, involve more than one regional node, or have metastasized to distant sites [40–43].

In the absence of evidence of local or distant spread, the affected kidney is resected by a radical nephrectomy, or, in the case of patients with small tumors, par-

Table 5.1 Histological classification (Heidelberg classification, [36]) of renal epithelial neoplasms, both malignant (RCC) and benign (RCA) with an indication of some of the major characteristic cytogenetic changes and incidence of tumor type [38, 206].

Tumor type	*Main cytogenetic characteristics*	*Relative frequency*
Malignant neoplasms		
Conventional (clear-cell) RCC	–3p (97%), +5q (50%), –14q, –8p, –9p, –6q	75–80%
Papillary RCC	–Y (80%), +17, +7 (85%), +16q (55%), +3q, +8p, +12q, +20q, t(X;1)(p11.2;q21)	10–15%
Chromophobe RCC	–1p, –Y, –2, –6, –10, –13, –17, –21	4–5%
Collecting duct RCC		<1%
Unclassified RCC		1–3%
Benign neoplasms		
Papillary RCA	–Y, +17, +7	
Oncocytic RCA	–Y, –1p, –14q	3–5%
Metanephric RCA	–Y, +17, +7	<1%

RCC, renal cell carcinoma; RCA, renal cell adenoma.

ticularly those with impaired renal function, by a partial nephrectomy. Even after full resection the risk of relapse remains, especially for patients who had large tumors, involvement of lymph nodes, or whose tumors had microscopic invasion into blood vessels necessitating more extensive surgery. In the presence of metastatic disease, the primary may still be resected, but therapy will depend upon systemic treatments. As renal cancer is resistant to conventional chemotherapy, immunological therapies are mostly used, such as interleukin 2 (IL-2) and interferon (IFN), which can produce remissions and prolong survival but are rarely curative.

For the clinician, therefore, the challenges presented by renal cancer are:

- To identify high-risk populations who can be screened using new approaches.
- To establish suitable means of making early diagnoses whilst the disease is manageable and curable by simple measures.
- To develop better systemic treatments, which may be immunological or chemotherapeutic, that depend upon the characterization of appropriate targets within the tumor which are not present or are present in low quantities in normal tissues.
- To develop appropriate monitoring tests to follow up patients after surgery or to monitor the effect of treatment.

5.1.3 Immunotherapy for Renal Cancer

With the resistance to radiotherapy and chemotherapy, the mainstay of clinical management of metastatic disease is with immunotherapy, with IL-2 and IFN being the established agents. The rationale for the use of immunotherapy arose particularly from the recognition that spontaneous regression of metastatic disease is a real, if rare, phenomenon occurring in up to 6% of patients [44–46]. Histological observations show that renal carcinomas commonly demonstrate infiltration by lymphocytes and macrophages.

5.1.3.1 Systemic Cytokine Therapy

Both IFNα and recombinant IL-2 have been shown to produce regression of the disease in a substantial minority of patients with renal cancer. These treatments can be associated with significant side effects. For single agent IFNα, the most recent large study has been reported by the Medical Research Council in the UK [47] with response rates in the order of 15% and prolongation of survival of only a few months. Interleukin 2 was first evaluated in the 1980s with a wide range of schedules. Validated response rates are in the order of 20–30% but there is a definite complete response rate where disease is no longer visible. A proportion of these patients remain in complete remission for prolonged periods of time and this may represent a small but important cured subpopulation. These treatments seem to be most effective in fit patients with a relatively small burden of disease [48]. Combinations of IL-2 and IFN sometimes given with chemotherapy appear to increase response rates [49] but there is no definitive evidence yet that this combination will produce prolonged survival or increased cure rates in patients with metastatic renal cancer.

5.1.3.2 Cellular-based Therapies

Treatment of circulating lymphoid cells with IL-2 (LAK cells) and reinfusion has been associated with some evidence of regression [50] but probably is not superior to treatment with IL-2 alone [51]. Tumor-infiltrating leukocytes can also be expanded *in vitro* and may result in a significant response rate in patients [52, 53]. Neither of these approaches have yet found a standard place in therapy and probably add little to conventional cytokine therapy [54]. More recent developments include allogeneic hematopoietic stem cell transplantation which may harness a graft-versus-tumor effect and may be responsible for regression of renal cancer in a small number of patients [55]. This is, however, a complex and dangerous therapy and adoptive immunotherapy using *in vitro* expansion of immunoreactive cells is probably applicable in a larger proportion of patients.

More recently, antigen-presenting cells, usually dendritic cells (DCs), have been harnessed clinically for the development of renal cancer vaccines and DCs treated

with whole tumor lysates can produce immunological changes *in vitro* [56]. Early clinical studies [57–59] in which renal cancer tissues taken from patients were lysed, irradiated, and used to pulse DCs isolated from peripheral blood, which were then used as a tumor vaccine for the patients resulted in some clinical responses, although these were infrequent and claims for stabilization of the disease were more common.

5.1.4 Existing Markers or Therapeutic Targets Undergoing Clinical Evaluation

Immunologically based therapies are promising but with their associated toxicity and effectiveness still being limited to a proportion of patients, considerable effort is being put into developing "cancer vaccines" involving autologous tumor cells transfected with cytokine or other immune-modulating genes, antigen-presenting cells such as DCs pulsed with tumor lysate or antigens or transfected with the antigen genes, injection of viral vectors or naked DNA encoding tumor antigens, or antibody-mediated therapies. The need for identification of tumor antigens (which may also be of use as markers) is therefore paramount.

5.1.4.1 **G250**

The G250 antibody was raised following immunization of mice with lysates of renal carcinoma cells and reacts with the majority of primary and metastatic RCC but not with normal tissues with the exception of bile duct cells and gastric mucosal cells [60–63]. The antigen recognized by the anti-G250 antibody is now known to be identical to the MN tumor-associated gene product first described in HeLa cells and subsequently identified as carbonic anhydrase IX [62, 64, 65], which is up-regulated at least in part in RCC by loss of VHL [66]. Reverse transcription-polymerase chain reaction (RT-PCR) has been employed to detect G250-expressing cells in the peripheral circulating blood of renal cell patients [67] and radiolabeled antibody has been used as a scanning and targeting reagent in several cancers [61, 68, 69] with preliminary experiments indicating some clinical activity against renal cancer [69]. More recently a dendritic cell vaccine transduced with a GM-CSF/CA-9 fusion gene showed promising results in mice, inducing DC maturation and CA-9-specific cytotoxic T lymphocyte (CTL) activity [70].

5.1.4.2 **RAGE-1**

The RAGE antigen (renal antigen) was identified as being a shared antigen recognized by a CTL clone raised against RCC cell lines. Although expressed in 37% of RCC lines and expression in normal tissues restricted to retina, it is only rarely (2–20%) expressed in RCC tumors [71, 72] so its role as a diagnostic marker or target for therapy may be restricted.

5.1.4.3 **Her-2/neu**

Her-2/neu or ErbB2, which is a member of the epidermal growth factor receptor (EGFR) family is expressed in a range of normal tissues but is overexpressed in several cancers, most notably ovarian and breast [73], in which it is being used as a therapeutic target either via antibody-mediated therapy or as the basis for vaccines. CTLs against Her-2/neu peptides have been raised and found to be capable of killing RCC cell lines [74], indicating processing and presentation of peptides. However, significant Her-2/neu expression is found in normal kidney and there is currently no agreement about whether Her-2/neu is expressed widely in RCC tissue [75–78] so the use of this as a therapy in RCC is not yet proven.

5.1.4.4 **Other Proteins**

A range of other proteins are being investigated as candidate markers that are associated with, but rarely unique to, renal carcinomas. These include the adhesion molecule CD44 [79], the oncoprotein *bcl*-2 [80], the telomerase reverse transcriptase enzyme [81], and vimentin [82–84]. A number of angiogenic factors are expressed in a range of cancers including RCC such as platelet-derived endothelial growth factor (PD-ECGF)/thymidine phosphorylase [85–89] and vascular endothelial growth factor (VEGF) [90–100] and which are being pursued as potential aids to prognosis or therapeutic targets.

5.2
Proteomic Studies

Although genetic/cytogenetic markers are widely used in the diagnosis and follow-up of leukemias and lymphomas, the majority of tumor markers in routine use for epithelial cancers are based on the measurement of specific proteins, either immunohistochemically or as a circulating form. Cytogenetic abnormalities contribute to the classification of renal cancer but this is not performed routinely, and the prognostic utility of mutation analysis, for example in the *VHL* gene, is currently being investigated [25–28, 30–32, 101]. Nucleic acid-based microarrays and differential display approaches have also facilitated the search for markers/targets [66, 102–108]. These developments are exciting and illustrate the power of examining multiple genes to classify tumors, although not yet in routine clinical use. However proteomics-based approaches also offer the benefits of multiplex approaches, measure the protein *per se* rather than the transcript, provide information about post-translational modifications which are likely to change with disease and may result in an easier translation to the clinic in some cases given the potential to accurately measure circulating markers in biological fluids without the need for extraction and laborious analysis.

Although historically the first tumor marker, Bence-Jones protein, was identified on the basis of the serendipitous observation by Dr. William McIntyre of an aggregate formation in urine samples [109–111], most protein markers or targets

have been identified from the identification of antigens reactive with specific antibodies raised against tumor material or cell lines, for example CEA, CA-125, and, as illustrated above, G250. Relatively few new tumor markers have entered clinical practice in recent years but the improvements in proteomic technology allowing systematic higher throughput profiling may alter this in the future. The variety of approaches being used in RCC to identify potential markers or targets, which will also increase our understanding of the underlying pathogenesis are illustrated in the following sections.

5.2.1 Tissue-based Studies

The cell complement and level of cellular complexity of normal tissues or organs vary considerably, reflecting their disparate functions. The kidney comprises at least 17 different cell types including proximal tubule epithelial cells, distal tubule epithelial cells, collecting duct epithelial cells, endothelial cells, mesangial cells, fibroblasts and muscle cells. Lysates prepared from normal kidney samples for analysis of the renal proteome therefore reflect numerous cell types. In tissue from renal tumors, the main cell type may be the malignant cell but additional cells such as endothelial cells and infiltrating lymphocytes will be present. Additionally within the tumor, there may be cells with varying degrees of genetic evolution and reflecting different microenvironments, for example normoxic and hypoxic areas. Such tissue heterogeneity will undoubtedly impact on comparative studies, introducing more "noise" to the analysis than would be present if the malignant cells alone were compared with the proximal epithelial tubule cells which are widely thought to be the normal counterpart in the case of clear cell RCC.

To overcome this, strategies such as immuno-isolation and laser-assisted or laser-capture microdissection have been investigated and are described below. Such selection processes may greatly facilitate comparison of specific cell types or areas and generate less noisy and more easily interpretable and readily applicable data. However they involve processing steps which ultimately may introduce *in vitro* artefacts. Set against this, analysis of whole-tissue lysates may be more difficult to interpret but ultimately they involve minimal manipulation of the *in vivo* state. Additionally interplay between different elements of the tumor such as stromal-epithelial and endothelial-epithelial interactions are undoubtedly important in the biology of the tumor and therefore analysis of whole-tissue lysates may ultimately give increased information about the overall changes in the tumor as a whole rather than solely the epithelial elements. The relative merits of the different approaches in producing clinically relevant findings remains to be determined.

5.2.1.1 Whole-tissue Lysates

Only four studies have been published examining the proteome of renal cell carcinoma based on two-dimensional polyacrylamide gel electrophoresis (2-D PAGE) analysis of material from tissue samples, and these have largely focused on the

clear cell type. The first two of these used samples that had been subjected to initial processing in terms of depletion of contaminating lymphocytes and as such are described in detail in the following section. In the third study, 12 RCC samples were examined with four proteins identified out of 12 being found to be reduced in all samples, namely aminoacylase I, aldehyde dehydrogenase I, enoyl-CoA-hydratase, and α-glycerol-3-phosphate [112]. In the most recent study, an analysis of matched normal and malignant samples from six patients with grades 2–4 clear cell RCC with subsequent validation of some results by immunohistochemistry and Western blotting [113] found 41 proteins to be decreased in all six malignant samples with 32 proteins being increased in at least 4 of 6 patients (Fig. 5.1).

Many of the identified protein changes were in metabolic enzymes with a coordinate demonstration of the Warburg effect with increased expression of the final six enzymes of glycolysis and parallel decrease in three of the enzymes catalyzing the reverse reactions in gluconeogenesis, together with a parallel decrease in several mitochondrial enzymes involved in oxidative phosphorylation including aminoacylase I and aldehyde dehydrogenase I found in the previous study [112]. In

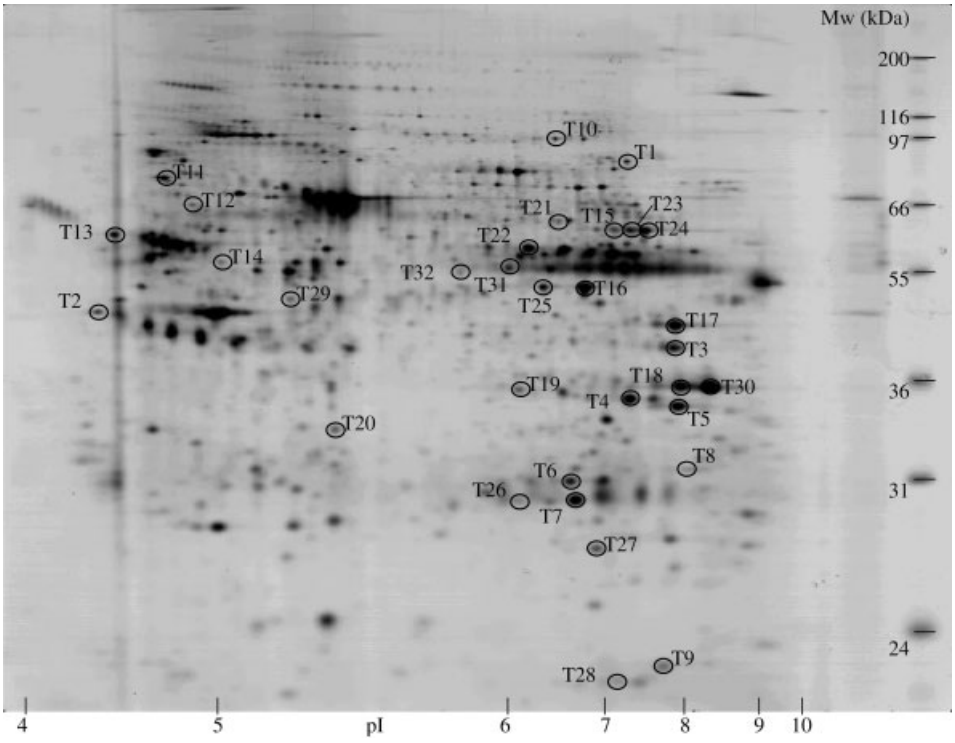


Figure 5.1 Example of a silver-stained analytical 2-D gel (pH 3–10, non-linear) containing 30 µg human RCC tissue displaying locations of 32 protein spots which are present in significantly higher amounts in RCC tissue when compared with normal kidney cortex (reproduced from [113]).

addition, the enzymes phosphoglucomutase, which catalyzes the final step in the conversion of glycogen to glucose-6-phosphate for entry into glycolysis, and lactate dehydrogenase, which converts pyruvate to lactate under low oxygen concentrations, were also shown by 2-D PAGE to be upregulated. Some of these changes have been previously reported in RCC on the basis of enzyme activity analysis [114] but the protein expression data in this study [113] confirms and provides further more comprehensive evidence of the "Warburg effect" [115] representing impaired mitochondrial function and increased dependence on glycolysis as a source of energy even in aerobic conditions. This phenomenon is widespread in cancers although the mechanism is still poorly understood.

In addition to supporting a down-regulation of mitochondrial enzymes involved in oxidative phosphorylation, enzymes involved in other cycles including fatty acid and amino acid metabolism and the urea cycle were also found to be down-regulated [113] and such metabolic alterations may underlie the cytoplasmic accumulation of lipids and glycogen characteristic of clear cell RCC. The extent of the decreases in mitochondrial enzyme content and oxidative phosphorylation complexes of clear cell and chromophilic RCC have previously been reported to be correlated with tumor aggression [116].

Several other changes previously described in RCC on the basis of enzymatic or immunohistochemical studies were found, including increases in the angiogenic factor thymidine phosphorylase, vimentin, and specific isoforms such as pyruvate kinase M2, aldolases A and C, and lactate dehydrogenase A. Additionally novel changes were seen, for example changes in expression of three members of the annexin family. With the demonstration of binding sites for the transcription factor hypoxia inducible factor 1 (HIF-1) in the promoters of many of the enzymes or specific isoforms, including aldolases A and C, enolase 1, LDH-A, phosphoglycerate kinase 1, pyruvate kinase, GAPDH, and proteins such as VEGF [117–123], the biological explanation for some findings becomes more apparent. VHL functions as the substrate recognition unit of an E3 ubiquitin ligase complex, with HIF family members being currently recognized substrates. Hence, under either hypoxic conditions or under normoxic conditions as a result of specific mutations in the *VHL* gene in RCC, HIF is not ubiquitinated and is stabilized within the cell, resulting in the transcription of several genes [124–126].

Changes in isozyme pattern with cancer have been recognized for many years [127] with particular forms representing adaptation for different functions, for example anerobic conditions. Several of the isoform switches described above have been previously found enzymatically in RCC, for example the switch to predominantly the LDH-A form [128], the increase in aldolase A form (fetal form present largely in muscle) and C form (brain form) [129] with elevated serum concentrations of aldolase A reported in 37–75% of patients with RCC indicating a potential use as a marker [129–131]. Similarly the increased expression of PK-M2 has been reported for a variety of tumor types including RCC [132], with the dimeric tumor-specific form assayed in serum providing clinically promising utility [133].

The findings of increased expression of both thymidine phosphorylase (TP) and phosphoglycerate kinase (PGK) may be related to the vascularity of RCC, as both

have been implicated in the regulation of angiogenesis. In mice bearing fibrosarcomas, elevated serum PGK was found to act as a disulfide reductase, cleaving plasmin to form angiostatin, an angiogenesis inhibitor [134]. Potentially the elevation in tissue PGK in RCC could translate to an increase in serum levels and hence modulate angiostatin activity. TP has been reported to be correlated with microvessel density, grade, and prognosis of RCC [85] and with restricted expression in normal tissues [135], may represent a potential therapeutic target.

Several proteins involved in structural functions were found to be up-regulated [113]. The intermediate filament protein vimentin has previously been found to be overexpressed in 51% of clear cell RCC when analyzed using cDNA microarrays, with expression correlating with disease outcome [107] and subsequently confirmed at the protein level [83]. The annexins, a family of calcium-dependent phospholipid-binding proteins implicated in exocytosis and endocytosis although their exact physiological roles are not yet clear, have been found to be up-regulated in other proteomic studies of cancers including colorectal [136] and bladder [137], although annexin I was found to be down-regulated in prostate cancer [138]. IL-6, postulated to be an autocrine growth factor in RCC [139, 140], is known to up-regulate annexins [141] and thus may be implicated in the changes seen in RCC. Immunohistochemistry confirmed the tumor localization of several of these proteins with one exception, namely cofilin which was localized to infiltrating T lymphocytes.

A complementary technique in terms of the optimum mass range of proteins examined is surface-enhanced laser desorption ionisation mass spectrometry (SELDI), which is essentially a mass spectrometer with time-of-flight detector into which can be inserted ProteinChips containing samples and resulting in mass-charge profiles of proteins/peptides present in the immobilized sample [142]. The spectrum of molecules profiled by the SELDI and selectivity of this system is generated by different ProteinChips being coated with different surfaces similar to those found in conventional chromatography. Preliminary results using SELDI analysis of tissue lysates have shown the presence of differentially expressed proteins in RCC compared with normal tissue, in particular, two peaks at 11.95 and 12.0 kDa, the identities of which were not known [143]. Similarly, using SELDI to ascertain whether distinct protein "fingerprints" could be generated from archival fine needle aspirate material from five RCC and nine metastatic malignant melanoma samples, promising initial results were obtained with the subsequent "blind" analysis of 15 samples resulting in the correct diagnosis in 87% of cases [144]. Whether this can be extended to further similar samples and indeed other cancer types has yet to be explored.

The results using 2-D PAGE on whole-tissue lysates validate the approach, essentially confirming previously described findings and/or producing results which can be explained on the basis of underlying genetic changes. The need for validation, both quantitatively and qualitatively in terms of isoform and cellular localization, is also illustrated by these studies. The overlap of the 2-D PAGE findings for RCC with those of other cancer types, such as the increases in annexins and GRP-78, but also α-enolase, peptidylprolyl isomerase A (cyclophilin), cofilin, elon-

gation factor 2, Mn-superoxide dismutase (Mn-SOD), vimentin, GAPDH, and lamin B1 [136, 137, 145] indicate the common aspects of some of the underlying pathologies. Future work needs to address the issues of examining under-represented proteins such as membrane proteins or those present in lower abundance, and the issue of differential post-translational modifications such as glycosylation.

5.2.1.2 **Enriched Cell Populations**

Immunoisolates Positive immunoselection of specific cell types has been successfully used in proteomic analysis of tissues such as breast with subsequent 2-D PAGE analysis showing different protein profiles with differing cell populations [146]. No similar studies have been performed in renal cancer although immunomagnetic isolation of specific segments such as proximal and distal tubular epithelial cells is possible [147]. With an antibody such as G250 it should theoretically be possible to enrich for malignant cells in clear cell RCC, although the effect of sample processing such as enzymatic digestion to yield cell suspensions would require evaluation.

Negative selection, that is depletion of contaminating cells, has been used successfully in the first study using 2-D PAGE to analyze changes in renal cancer [148]. Tissue samples were snap-frozen, subsequently scraped to form single-cell suspensions which were then immunodepleted of lymphocytes. Using samples from 10 patients with various subtypes of RCC, two out of four proteins found to be lost in tumors were identified as the mitochondrial enzymes ubiquinol cytochrome *c* reductase and NADH-ubiquinone oxidoreductase complex I (24 kDa) involved in respiratory chain electron transport [148]. Both of these were also identified in the previously described study using whole-tissue lysates [113] but as different subunits present in the complexes. In a follow-up study [149], a further two proteins were identified as being down-regulated in RCC, namely isoforms of plasma glutathione peroxidase. The results of these studies have been used as the basis for a normal kidney map on the Swiss 2-D PAGE website (http://ca.expasy.org/cgi-bin/map2/def?KIDNEY-HUMAN).

Two proteins identified by this group as being expressed only in the RCC samples were multimeric and reduced forms of Mn-SOD [149]. A single form of Mn-SOD was also subsequently found to be up-regulated in the whole-tissue lysate study [113]. Elevated Mn-SOD activity in RCC has been described previously [150] although immunolabeling of clear cell tumors for Mn-SOD has shown weak staining compared with normal proximal tubules [151] and it may be that the major change is in the precise forms of Mn-SOD found. Pivotal in defense against radical oxygen species, Mn-SOD has been implicated in resistance to radiotherapy and chemotherapy [152, 153], which is of particular interest in RCC where resistance to these conventional therapies is a major clinical issue. Mn-SOD has been suggested as a potential drug target, with inhibition in tumor cells potentially leading to increased levels of damaging oxygen radicals and apoptosis [154].

Laser-capture Microdissection With the advent of laser-assisted microdissection methods, studies have examined their utility for procuring material suitable for subsequent proteomic analysis. The vast majority of proteomic studies employing such approaches have used the laser-capture microdissection (LCM) system [155]. The system is based on an inverted microscope with a near-infrared laser mounted above it, a manipulator arm, and a computer link to save images. The section to be dissected is fixed/stained and the slide placed on the microscope platform. A Perspex cap with a 6-mm diameter transparent thermo-labile ethylene vinyl acetate film on its lower surface is placed on the section with the film in direct contact with the tissue. The laser beam, of diameter 7.5–30 μm, is fired through the cap at the cells of interest, causing the film to melt and fuse with the cells below. When the cap is lifted from the section, the dissected cells fused with the cap are also removed and can be solubilized using an appropriate extraction buffer.

The other two laser-assisted systems developed more recently and which function using semi-automated non-contact laser-assisted dissection by cutting round the areas of interest, have not yet been used greatly in proteomic approaches although this will undoubtedly change [156, 157].

An initial "proof of principle" demonstration using kidney and cervical tissue showed that material collected by LCM could be used to generate 2-D PAGE protein profiles with minimal effects of processing and subsequent satisfactory sequencing of a range of proteins [158], with further definition of effects of different tissue processing regimens indicating critical factors such as type of fixative and stain [159]. Laser-assisted microdissection has been used widely for the study of DNA and RNA in selected cell populations, largely due to use of PCR amplification so that sensitivities down to the single cell level can be achieved. The main limitation for its routine use in proteomic studies is the length of dissection required to procure sufficient cells for analysis with normal kidney requiring tens of thousands of shots and several days of dissection to obtain sufficient material. Even with tissues that are relatively easy to dissect, only the most abundant protein species can be examined. The degree of purity of the microdissected cells is likely to be compromised by the extensive microdissection needed and the degree of enrichment afforded in the protein profile relative to the starting material also varies between tissue types [159].

The approach can be used successfully, as clearly shown by the comparative analysis of normal and malignant esophageal cells dissected from two patient samples where dissection of approximately 50 000 cells allowed visualization of almost 700 proteins with 98% similarity between normal and malignant protein profiles and identification of annexin 1 as one of the proteins lost in tumors [160]. However we conclude that the use of LCM with 2-D PAGE will not be a worthwhile option for normal kidney and RCC comparisons.

Where we have found the LCM to be useful in RCC is in its use with Western blotting to confirm the 2-D PAGE results of whole tissue [113, 161]. As few as 200–2500 cells may be sufficient depending on the antibody used and the relative expression level. This complements immunohistochemistry, being able to confirm

both the cellular compartment in which a particular protein is expressed and the form of protein in terms of its molecular size. Several studies have confirmed this utility in other cancers, for example in confirming either total loss or dramatic reduction of annexin 1 in both oesophageal and prostate cancers [138]. Other possible combinations of LCM with techniques requiring small amounts of material include immunoassay, protein arrays, and SELDI. Although successfully applied in investigations involving other cancers such as prostate [162–164], these approaches have not yet been used specifically in RCC but undoubtedly will be in the future with initial results using RCC tissue showing enriched profiles on SELDI following LCM (our unpublished data).

5.2.2
Primary and Established Cell Lines

In addition to lysates prepared from tissue samples, established or short-term primary cell lines which represent a pure cell population continue to be a valuable adjunct in proteomic studies, particularly in functional studies with the ability to manipulate conditions and thereby unravel biological pathways. An important caveat is that studies have shown that primary lines do undergo changes essentially as a result of culture and such *in vitro* artefacts need to be borne in mind when interpreting the data. In renal cancer, this is exemplified by RAGE where, although the antigen is expressed in many cell lines, its expression in RCC tumors is very limited [71, 72]. In bladder primary cultures compared with parent tissue samples following methionine labeling, although strikingly similar, a number of components including gelsolin, keratin 18, HSP-28, GST-P, annexin V, and fatty acid-binding proteins have been found by 2-D PAGE to be differentially synthesized in lines compared with the parent tissue [165]. Although for renal cancer no such studies have yet been published, our experience with both normal and malignant renal primary cultures is similar in that although similarities with parent tissue are apparent, clearly expression of some proteins does change (unpublished data).

One potential use of cell lines is in unravelling the mechanisms of chemoresistance and this approach is already being exploited in several cancers using 2-D PAGE [166–169]. This work is discussed in detail in Chapter 11. This is of particular interest in renal cancer given its marked resistance to all agents examined. Although the transporter protein P-glycoprotein (MDR1/PgP) has been implicated in RCC and indeed other cancers on the basis of its immunohistochemical demonstration of up-regulation, there may be other mechanisms and this is the focus of some of our current studies using the anti-tubulin agents as model agents together with established and primary renal lines.

Cell lines also represent a valuable tool in unravelling the downstream pathways perturbed by particular genetic changes and their involvement in tumorigenesis. As part of a collaborative study we are currently studying a range of RCC lines transfected with either full-length or mutant forms of the *VHL* gene to investigate the feasibility of this approach. Initial results have shown that <50 pro-

teins are significantly changed between wild-type and transfected lines with only one of these being an on-off difference and these are now being characterized.

Based on the immune responsiveness of RCC, a number of studies have examined proteins involved in antigen presentation and therefore of critical importance in ensuring the stimulation of T cell responses. Serial immunoblotting of 2-D PAGE western blots of an established RCC cell line for components of the MHC class I antigen-presenting machinery has enabled the mapping and identification of the chaperones calreticulin, protein disulfide isomerase, and tapasin as well as several heat shock proteins hsp70, hsp90, grp96, and grp78/Bip [170] and examined the effects of IFNγ treatment on protein expression profiles. Such studies offer exciting opportunities to investigate proteins implicated in the successful immunotherapeutic treatment of RCC and may offer new insights into design and effectiveness of therapies.

5.2.3 Biological Fluids

5.2.3.1 Serum and Plasma

No systematic analysis of plasma or serum samples from patients with renal cancer has yet been published although this is ongoing in laboratories including our own utilizing both 2-D PAGE and SELDI-based approaches. Inevitably with the dominance of particular abundant proteins such as albumin and immunoglobulins and the challenge of covering such a wide concentration range, depletion or prefraction strategies are necessary and are currently being developed by several groups [171–174].

Several studies have, however, examined serum antibody reactivity. Circulating anti-tumor antibodies have been described in the serum of patients with various types of cancer recognizing a range of proteins including p53 [175], Her2/neu [176], and Ras [177]. The first development to systematically exploit the existence of such antibodies, based on the premise that the presence of antibodies to specific proteins may indicate a humoral response due to either overexpression, mutation, or aberrant forms of proteins in tumors and therefore indicate potential therapeutic targets or markers, was the SEREX approach (serological analysis of recombinant cDNA expression). This involves the production of a cDNA expression library from a tumor tissue or cell line and probing with patient serum antibodies [178]. Several hundred potential antigens have so far been identified in this way with approximately one-third appearing to be novel, although many remain to be sequenced [179]. Included amongst these are a number of potential renal cancer antigens such as carbonic anhydrase XII [178], LUCA-15, gene21, the kinases LKB1/STK11 and bcr, and the potential tumor suppressor SNC6 [180], although their utility has yet to be validated.

Subsequently, several studies have pursued a similar approach but using blots of 2-D PAGE-separated lysates of tumor samples or cell lines as the source of potential antigens, with the advantage of also potentially identifying antigens whose reactivity is dependent on post-translational modification. The cancers studied in-

clude neuroblastoma [181], lung cancer [182], and several studies in renal cancer, variably termed SPEAR (serological and proteomic evaluation of antibody response) [161], SERPA (serological proteomic analysis) [183], or PROTEOMEX [184]. In the first study [183], using tissue lysates from seven patients, five RCC-specific spots were found, reactive only with patient sera, two of which were identified as smooth muscle protein 22-alpha (SM22-α) and carbonic anhydrase I. Using lysates from normal kidney and RCC cell lines, reactivity to a number of proteins including cytokeratin 8, γ-actin, strathmin, tropomyosin, α- and β-tubulin and vimentin was found [184] with further studies using cell lines identifying nine metabolic enzymes as reacting with patient sera [185], including aldose reductase, GST-P, superoxide dismutase 1, thioredoxin, and triosephosphate isomerase A and reactivity to heat shock proteins in both RCC patients and controls although dependent on the cell line used as antigenic source [186]. In the most recent study using eight patient-matched normal and tumor tissue lysates probed with autologous sera and sera from healthy controls [161], nine reactive proteins were found in 2/8 patients with none being found in more than two. Proteins recognized included annexins I and IV, triosephosphate isomerase, thymidine phosphorylase, carbonic anhydrase I, Mn-SOD, and major vault protein, several of which confirmed other studies and were also shown to be up-regulated in tumor tissue samples.

There is apparent overlap between the findings of these studies and it is noteworthy that several of the proteins such as thymidine phosphorylase, Mn-SOD, heat shock proteins, and vimentin have previously been found in either proteomic or transcriptomic studies to be overexpressed in renal cancer. Some of the autoreactive antibodies have also been described in other cancers, for example circulating antibodies to glycosylated annexins I and/or II have been described in 60% of patients with lung adenocarcinoma and 33% of patients with squamous cell lung carcinoma [182]. The finding of antibodies against carbonic anhydrase I in two of the above studies is interesting as two other carbonic anhydrase family members, IX and XII, have been shown to be RCC-tumor antigens [62, 187] the latter by SEREX. Further such studies are now needed but the potential of the approach is apparent.

5.2.3.2 **Urine**

A potentially rich source of tumor antigens or markers in urological malignancies where tumor-derived products may be released directly into the urine, analysis of urine offers multiple technical challenges such as the several hundred- to thousand-fold more dilute overall protein concentration compared with serum, the effect of hydration state on protein concentration, the presence of proteases, and the potential for contamination from other sources such as seminal or vaginal fluids. Several 2-D PAGE studies and more recently liquid chromatography-tandem mass spectrometry approaches have systematically examined urine composition, describing the presence of several hundred proteins including intact and many cleaved forms of plasma proteins such as retinol-binding protein, trans-

ferrin, albumin, and β_2-microglobulin, renal- or urogenital-tract-derived proteins such as erythropoietin, urokallikrein, epidermal growth factor, E-cadherin, basement membrane antigens and Tamm-Horsfall protein, and viral or bacterial associated proteins [188–194].

Specific tumor-associated proteins in urine include PCA-1 forms in prostate cancer [195], psoriasin (S-100A7) in bladder squamous cell carcinoma [196, 197], tumor-associated trypsin inhibitor (TATI) in many cancers [198] and several potential urinary markers for transitional cell bladder cancer including nuclear matrix protein (NMP)-22, BTA, and fibrin-fibrinogen degradation products (FDPs) [199, 200]. Preliminary studies indicate a possible use of NMP-22 [201, 202] and beta-glucuronidase [203] in renal cancer.

SELDI offers an attractive alternative approach to the analysis of urinary proteins, particularly with its inherent optimal range being <20 kDa, and analysis of samples from 30 patients with transitional cell carcinoma (TCC) of the bladder using an anion-exchange chip surface (SAX2) found five potential novel biomarkers [204]. Sensitivity ranged from 43 to 70% and specificity from 70 to 86% based on individual peaks but the presence of particular clusters of peaks increased sensitivity to 87% with a specificity of 66%, with the detection of low-grade tumors being superior to conventional cytologic approaches. In a larger more recent study, samples from 48 patients prior to undergoing nephrectomy for clear cell RCC, 38 normal volunteers, and 20 outpatients with benign urological conditions were profiled on a cation-exchange chip (WCX2) and used to successfully train various neural network models. In an initial "blind" group of samples from 12 patients with RCC, 11 healthy controls and 9 patients with benign diseases, sensitivities and specificities of 81.8–83.3% were achieved [205]. However sensitivities and specificities declined markedly when tested on a larger "blind" group of samples analyzed almost a year later with several factors such as changing laser performance with time being implicated. This is being pursued in addition to the identification of several peaks which differed significantly between the normal and RCC groups.

5.3 Conclusions

There are a wealth of studies published using proteomics-based approaches in cancer with the aims of understanding the underlying pathological changes, describing new markers for diagnostic or prognostic use, or discovering new therapeutic targets. The number of these examining renal cancer is relatively small, probably reflecting the lower incidence of this cancer. However this is an ideal cancer in which to demonstrate how proteomics-based approaches can address several clinical issues in oncology, encompassing as it does issues of earlier diagnosis, prognosis, mechanisms of chemoresistance, cytokine and growth factor-mediated growth and immune responsiveness, immunogenicity of tumors and corresponding strategies for identification of therapeutic targets, and also, of

course, the relationship of particular genetic events with protein changes and the clinical implications. This chapter has illustrated how such issues may be addressed. The emphasis now must be to build on the foundations set by the initial studies to produce results of real clinical utility. We also need to expand the range of approaches to extremely important, yet technically challenging areas, such as the analysis of post-translational modifications (e.g. glycosylation) and the membrane proteome of normal and malignant cells.

5.4 References

1 Pantuck, A. J., Zisman, A., Belldegrun, A. S. *J. Urol.* **2001**, *166*, 1611–1623.
2 CancerStats. Cancer Research UK, London, **2002**.
3 Dhote, R., Pellicer-Coeuret, M., Thiounn, N., Debre, B., Vidal-Trecan, G. *Br. J. Urol. Int.* **2000**, *86*, 20–27.
4 McLaughlin, J. K., Lindblad, P., Mellemgaard, A., McCredie, M. et al. *Int. J. Cancer* **1995**, *60*, 194–198.
5 Bergstrom, A., Hsieh, C. C., Lindblad, P., Lu, C. M. et al. *Br. J. Cancer* **2001**, *85*, 984–990.
6 Lindblad, P., Wolk, A., Bergstrom, R., Persson, I., Adami, H. O. *Cancer Epidemiol. Biomarkers Prev.* **1994**, *3*, 631–639.
7 Mellemgaard, A., Lindblad, P., Schlehofer, B., Bergstrom, R. et al. *Int. J. Cancer* **1995**, *60*, 350–354.
8 Lindblad, P., Mellemgaard, A., Schlehofer, B., Adami, H. O. et al. *Int. J. Cancer* **1995**, *61*, 192–198.
9 McLaughlin, J. K., Chow, W. H., Mandel, J. S., Mellemgaard, A. et al. *Int. J. Cancer* **1995**, *63*, 216–221.
10 Kreiger, N., Marrett, L. D., Dodds, L., Hilditch, S., Darlington, G. A. *Cancer Causes Control* **1993**, *4*, 101–110.
11 Anttila, A., Pukkala, E., Sallmen, M., Hernberg, S., Hemminki, K. *J. Occup. Environ. Med.* **1995**, *37*, 797–806.
12 Lynge, E., Anttila, A., Hemminki, K. *Cancer Causes Control* **1997**, *8*, 406–419.
13 Mandel, J. S., McLaughlin, J. K., Schlehofer, B., Mellemgaard, A. et al. *Int. J. Cancer* **1995**, *61*, 601–605.
14 Wolk, A., Gridley, G., Niwa, S., Lindblad, P. et al. *Int. J. Cancer* **1996**, *65*, 67–73.
15 Lindblad, P., Wolk, A., Bergstrom, R., Adami, H. O. *Cancer Epidemiol. Biomarkers Prev.* **1997**, *6*, 215–223.
16 Augustsson, K., Skog, K., Jagerstad, M., Dickman, P. W., Steineck, G. *Lancet* **1999**, *353*, 703–707.
17 Prineas, R. J., Folsom, A. R., Zhang, Z. M., Sellers, T. A., Potter, J. *Epidemiology* **1997**, *8*, 31–36.
18 Yuan, J. M., Gago-Dominguez, M., Castelao, J. E., Hankin, J. H. et al. *Int. J. Cancer* **1998**, *77*, 211–216.
19 Takahashi, S., Shirai, T., Ogawa, K., Imaida, K. et al. *Acta Pathol. Jpn* **1993**, *43*, 674–682.
20 Maher, E. R., Yates, J. R. *Br. J. Cancer* **1991**, *63*, 176–179.
21 Turner, K. J. *Br. J. Urol. Int.* **2000**, *86*, 155–164.
22 Tomlinson, I. P., Alam, N. A., Rowan, A. J., Barclay, E. et al. *Nature Genet.* **2002**, *30*, 406–410.
23 Iliopoulos, O., Eng, C. *Semin. Oncol.* **2000**, *27*, 138–149.
24 Zbar, B., Glenn, G., Lubensky, I., Choyke, P. et al. *J. Urol.* **1995**, *153*, 907–912.
25 Gnarra, J. R., Tory, K., Weng, Y., Schmidt, L. et al. *Nature Genet.* **1994**, *7*, 85–90.
26 Yao, M., Yoshida, M., Kishida, T., Nakaigawa, N. et al. *J. Natl. Cancer Inst.* **2002**, *94*, 1569–1575.
27 Kondo, K., Yao, M., Yoshida, M., Kishida, T. et al. *Genes Chromosom. Cancer* **2002**, *34*, 58–68.
28 Hamano, K., Esumi, M., Igarashi, H., Chino, K. et al. *J. Urol.* **2002**, *167*, 713–717.

29 Ma, X., Yang, K., Lindblad, P., Egevad, L., Hemminki, K. *Oncogene* **2001**, *20*, 5393–5400.

30 Brauch, H., Weirich, G., Brieger, J., Glavac, D. et al. *Cancer Res.* **2000**, *60*, 1942–1948.

31 Schraml, P., Struckmann, K., Hatz, F., Sonnet, S. et al. *J. Pathol.* **2002**, *196*, 186–193.

32 Brauch, H., Weirich, G., Hornauer, M.A., Storkel, S. et al. *J. Natl. Cancer Inst.* **1999**, *91*, 854–861.

33 Gudbjartsson, T., Jonasdottir, T.J., Thoroddsen, A., Einarsson, G.V. et al. *Int. J. Cancer* **2002**, *100*, 476–479.

34 Motzer, R.J., Bander, N.H., Nanus, D.M. *N. Engl. J. Med.* **1996**, *19*, 335, 865–875.

35 Thoenes, W., Storkel, S., Rumpelt, H.J. *Pathol. Res. Pract.* **1986**, *181*, 125–143.

36 Kovacs, G., Akhtar, M., Beckwith, B.J., Bugert, P. et al. *J. Pathol.* **1997**, 183, 131–133.

37 Storkel, S., Eble, J.N., Adlakha, K., Amin, M. et al. *Cancer* **1997**, *80*, 987–989.

38 Zambrano, N.R., Lubensky, I.A., Merino, M.J., Linehan, W.M., Walther, M.M. *J. Urol.* **1999**, *162*, 1246–1258.

39 Gelb, A.B. *Cancer* **1997**, *80*, 981–986.

40 Guinan, P., Sobin, L.H., Algaba, F., Badellino, F. et al. *Cancer* **1997**, *80*, 992–993.

41 Javidan, J., Stricker, H.J., Tamboli, P., Amin, M.B. et al. *J. Urol.* **1999**, *162*, 1277–1281.

42 Kinouchi, T., Saiki, S., Meguro, N., Maeda, O. et al. *Cancer* **1999**, *85*, 689–695.

43 Tsui, K.H., Shvarts, O., Smith, R.B., Figlin, R.A. et al. *J. Urol.* **2000**, *163*, 1090–1095.

44 Marcus, S.G., Choyke, P.L., Reiter, R., Jaffe, G.S. et al. *J Urol.* **1993**, *150*, 463–466.

45 Elhilali, M.M., Gleave, M., Fradet, Y., Davis, I. et al. *Br. J. Urol. Int.* **2000**, *86*, 613–618.

46 Oliver, R.T.D. In *Renal and Adrenal Tumors* (A. Belldegrun, A.W.S. Ritchie, R.A. Figlin, R.T.D. Oliver, E.D. Vaughan, Eds.). Oxford University Press, New York, **2003**, pp 165–177.

47 MRC Renal Cancer Collaborators. *Lancet* **1999**, *353*, 14–17.

48 Decatris, M., Santhanam, S., O'Byrne, K. *BioDrugs* **2002**, *16*, 261–281.

49 Atzpodien, J., Kirchner, H., Illiger, H.J., Metzner, B. et al. *Br. J. Cancer* **2001**, *85*, 1130–1136.

50 Rosenberg, S.A. *J. Clin. Oncol.* **1992**, *10*, 180–199.

51 Law, T.M., Motzer, R.J., Mazumdar, M., Sell, K.W. et al. *Cancer* **1995**, *76*, 824–832.

52 Bukowski, R.M., Sharfman, W., Murthy, S., Rayman, P. et al. *Cancer Res.* **1991**, *51*, 4199–4205.

53 Figlin, R.A., Pierce, W.C., Kaboo, R., Tso, C.L. et al. *J. Urol.* **1997**, *158*, 740–745.

54 Figlin, R.A., Thompson, J.A., Bukowski, R.M., Vogelzang, N.J. et al. *J. Clin. Oncol.* **1999**, *17*, 2521–2529.

55 Childs, R., Chernoff, A., Contentin, N., Bahceci, E. et al. *N. Engl. J. Med.* **2000**, *343*, 750–758.

56 Gitlitz, B.J., Figlin, R.A., Pantuck, A.J., Belldegrun, A.S. *Curr. Urol. Rep.* **2001**, *2*, 46–52.

57 Holtl, L., Zelle-Rieser, C., Gander, H., Papesh, C. et al. *Clin. Cancer Res.* **2002**, *8*, 3369–3376.

58 Marten, A., Flieger, D., Renoth, S., Weineck, S. et al. *Cancer Immunol. Immunother.* **2002**, *51*, 637–644.

59 Oosterwijk-Wakka, J.C., Tiemessen, D.M., Bleumer, I., De Vries, I.J. et al. *J. Immunother.* **2002**, *25*, 500–508.

60 Oosterwijk, E., Ruiter, D.J., Hoedemaeker, P.J., Pauwels, E.K. et al. *Int. J. Cancer* **1986**, *38*, 489–494.

61 Oosterwijk, E., Bander, N.H., Divgi, C.R., Welt, S. et al. *J. Clin. Oncol.* **1993**, *11*, 738–750.

62 Uemura, H., Nakagawa, Y., Yoshida, K., Saga, S. et al. *Br. J. Cancer* **1999**, *81*, 741–746.

63 Ivanov, S., Liao, S.Y., Ivanova, A., Danilkovitch-Miagkova, A. et al. *Am. J. Pathol.* **2001**, *158*, 905–919.

64 Opavsky, R., Pastorekova, S., Zelnik, V., Gibadulinova, A. et al. *Genomics* **1996**, *33*, 480–487.

65 Grabmaier, K., Vissers, J.L., De Weijert, M.C., Oosterwijk-Wakka, J.C. et al. *Int. J. Cancer* **2000**, *85*, 865–870.

66 Ivanov, S.V., Kuzmin, I., Wei, M.H., Pack, S. et al. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 12596–12601.

67 McKiernan, J.M., Buttyan, R., Bander, N.H., de la Taille, A. et al. *Cancer* **1999**, *86*, 492–497.

68 Divgi, C.R., Bander, N.H., Scott, A.M., O'Donoghue, J.A. et al. *Clin. Cancer Res.* **1998**, *4*, 2729–2739.

69 Steffens, M.G., Boerman, O.C., De Mulder, P.H., Oyen, W.J. et al. *Clin. Cancer Res.* **1999**, *5*, 3268s–3274s.

70 Hernandez, J.M., Bui, M.H., Han, K.R., Mukouyama, H. et al. *Clin. Cancer Res.* **2003**, *9*, 1906–1916.

71 Gaugler, B., Brouwenstijn, N., Vantomme, V., Szikora, J.P. et al. *Immunogenetics* **1996**, *44*, 323–330.

72 Neumann, E., Engelsberg, A., Decker, J., Storkel, S. et al. *Cancer Res.* **1998**, *58*, 4090–4095.

73 Ross, J.S., Fletcher, J.A., *Semin. Cancer Biol.* **1999**, *9*, 125–138.

74 Brossart, P., Stuhler, G., Flad, T., Stevanovic, S. et al. *Cancer Res.* **1998**, *58*, 732–736.

75 Latif, Z., Watters, A.D., Bartlett, J.M. Underwood, M.A., Aitchison, M., *Br. J. Urol. Int.* **2002**, *89*, 5–9.

76 Stumm, G., Eberwein, S., Rostock-Wolf, S., Stein, H. et al. *Int. J. Cancer* **1996**, *69*, 17–22.

77 Seliger, B., Rongcun, Y., Atkins, D., Hammers, S. et al. *Int. J. Cancer* **2000**, *87*, 349–359.

78 Koeppen, H.K., Wright, B.D., Burt, A.D., Quirke, P. et al. *Histopathology* **2001**, *38*, 96–104.

79 Rioux-Leclercq, N., Epstein, J.I., Bansard, J.Y., Turlin, B. et al. *Hum. Pathol.* **2001**, *32*, 1209–1215.

80 Suzuki, K., Tokue, A. *Urol. Int.* **2002**, *69*, 57–62.

81 Fujioka, T., Hasegawa, M., Suzuki, Y., Suzuki, T. et al. *Int. J. Urol.* **2000**, *7*, 16–21.

82 Kruslin, B., Kopjar, A., Dimanovski, J., Belicza, M. *Acta Med. Croatica* **1999**, *53*, 11–14.

83 Sabo, E., Miselevich, I., Bejar, J., Segenreich, M. et al. *Br. J. Urol.* **1997**, *80*, 864–868.

84 Young, A.N., De Oliveira Salles, P.G., Lim, S.D., Cohen, C. et al. *Am. J. Surg. Pathol.* **2003**, *27*, 199–205.

85 Suzuki, K., Morita, T., Hashimoto, S., Tokue, A. *Urol. Res.* **2001**, *29*, 7–12.

86 Hirano, Y., Takayama, T., Kageyama, S., Ushiyama, T. et al. *Urol. Res.* **2002**, *30*, 112–115.

87 Imazano, Y., Takebayashi, Y., Nishiyama, K., Akiba, S. et al. *J. Clin. Oncol.* **1997**, *15*, 2570–2578.

88 Kinsui, H., Ueda, T., Suzuki, H., Isaka, S. et al. *Jpn: J. Cancer Res.* **2002**, *93*, 340–345.

89 Wada, S., Yoshimura, R., Naganuma, T., Yoshida, N. et al. *Br. J. Urol. Int.* **2003**, *91*, 105–108.

90 Hemmerlein, B., Kugler, A., Ozisik, R., Ringert, R.H. et al. *Virchows Arch.* **2001**, *439*, 645–652.

91 Blancher, C., Moore, J.W., Robertson, N., Harris, A.L. *Cancer Res.* **2001**, *61*, 7349–7355.

92 Brieger, J., Weidt, E.J., Schirmacher, P., Storkel, S. et al. *J. Mol. Med.* **1999**, *77*, 505–510.

93 Fang, J., Yan, L., Shing, Y., Moses, M.A. *Cancer Res.* **2001**, *61*, 5731–5735.

94 Gunningham, S.P., Currie, M.J., Han, C., Turner, K. et al. *Cancer Res.* **2001**, *61*, 3206–3211.

95 Igarashi, H., Esumi, M., Ishida, H., Okada, K. *Cancer* **2002**, *95*, 47–53.

96 Lee, J.S., Kim, H.S., Jung, J.J., Park, C.S., Lee, M.C. *J. Surg. Oncol.* **2001**, *77*, 55–60.

97 Nicol, D., Hii, S.I., Walsh, M., Teh, B. et al. *J. Urol.* **1997**, *157*, 1482–1486.

98 Relf, M., LeJeune, S., Scott, P.A., Fox, S. et al. *Cancer Res.* **1997**, *57*, 963–969.

99 Takahashi, A., Sasaki, H., Kim, S.J., Tobisu, K. et al. *Cancer Res.* **1994**, *54*, 4233–4237.

100 Tomisawa, M., Tokunaga, T., Oshika, Y., Tsuchida, T. et al. *Eur. J. Cancer* **1999**, *35*, 133–137.

101 Gallou, C., Longuemaux, S., Delomenie, C., Mejean, A. et al. *Pharmacogenetics* **2001**, *11*, 521–535.

102 Young, A.N., Amin, M.B., Moreno, C.S., Lim, S.D. et al. *Am. J. Pathol.* **2001**, *158*, 1639–1651.

103 Takahashi, M., Rhodes, D.R., Furge, K.A., Kanayama, H. et al. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 9754–9759.

104 Stassar, M.J., Devitt, G., Brosius, M., Rinnab, L. et al. *Br. J. Cancer* **2001**, *85*, 1372–1382.

105 Skubitz, K.M., Skubitz, A.P. *J. Lab. Clin. Med.* **2002**, *140*, 52–64.

106 Rae, F.K., Stephenson, S.A., Nicol, D.L., Clements, J.A. *Int. J. Cancer* **2000**, *88*, 726–732.

107 Moch, H., Schraml, P., Bubendorf, L., Mirlacher, M. et al. *Am. J. Pathol.* **1999**, *154*, 981–986.

108 Higgins, J.P., Shinghal, R., Gill, H., Reese, J.H. et al. *Am. J. Pathol.* **2003**, *162*, 925–932.

109 Heppell-Parton, A.C., Daly, M.C., Drabkin, H.A., Rabbitts, P.H. *Cytogenet. Cell Genet.* **1993**, *63*, 64–65.

110 Ortaldo, J.R., Glenn, G.M., Young, H.A., Frey, J.L. *J. Natl. Cancer Inst.* **1992**, *84*, 1897–1903.

111 Freedman, S.F., Amedee, R.G., Molony, T. *Ear Nose Throat J.* **1992**, *71*, 655–658.

112 Balabanov, S., Zimmermann, U., Protzel, C., Scharf, C. et al. *Eur. J. Biochem.* **2001**, *268*, 5977–5980.

113 Unwin, R.D., Craven, R.A., Harnden, P., Hanrahan, S. et al. *Proteomics* **2003**, *3*, 1620–1632.

114 Steinberg, P., Storkel, S., Oesch, F., Thoenes, W. *Lab. Invest.* **1992**, *67*, 506–511.

115 Warburg, O. *Science* **1956**, *123*, 309–314.

116 Simonnet, H., Alazard, N., Pfeiffer, K., Gallou, C. et al. *Carcinogenesis* **2002**, *23*, 759–768.

117 Semenza, G.L., Jiang, B.-H., Leung, S.W., Passantino, R. et al. *J. Biol. Chem.* **1996**, *271*, 32529–32537.

118 Firth, J.D., Ebert, B.L., Ratcliffe, P.J. *J. Biol. Chem.* **1995**, *270*, 21021–21027.

119 Ebert, B.L., Gleadle, J.M., O'Rourke, J.F., Bartlett, S.M. et al. *Biochem. J.* **1996**, *313*, 809–814.

120 Semenza, G.L., Roth, P.H., Fang, H.M., Wang, G.L. *J. Biol. Chem.* **1994**, *269*, 23757–23763.

121 Wang, G.L., Semenza, G.L. *J. Biol. Chem.* **1995**, *270*, 1230–1237.

122 Firth, J.D., Ebert, B.L., Pugh, C.W., Ratcliffe, P.J. *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 6496–6500.

123 Graven, K.K., Yu, Q., Pan, D., Roncarati, J.S., Farber, H.W. *Biochim. Biophys. Acta* **1999**, *1447*, 208–218.

124 Tanimoto, K., Makino, Y., Pereira, T., Poellinger, L. *EMBO J.* **2000**, *19*, 4298–4309.

125 Krieg, M., Haas, R., Brauch, H., Acker, T. et al. *Oncogene* **2000**, *19*, 5435–5443.

126 Cockman, M.E., Masson, N., Mole, D.R., Jaakkola, P. et al. *J. Biol. Chem.* **2000**, *275*, 25733–25741.

127 Schapira, F. *Isozymes. Curr. Top. Biol. Med. Res.* **1981**, *5*, 27–75.

128 Matsuda, M., Osafune, M., Nakano, E., Kotake, T. et al. *Urol. Res.* **1980**, *8*, 201–206.

129 Zhu, Y.Y., Takashi, M., Miyake, K., Kato, K. *J. Urol.* **1991**, *146* , 469–472.

130 Takashi, M., Zhu, Y., Nakano, Y., Miyake, K., Kato, K. *Urol. Res.* **1992**, *20*, 307–311.

131 Takashi, M., Haimoto, H., Koshikawa, T., Kato, K. *Am. J. Clin. Pathol.* **1990**, *93*, 631–636.

132 Brinck, U., Eigenbrodt, E., Oehmke, M., Mazurek, S., Fischer, G. *Virchows Arch.* **1994**, *424*, 177–185.

133 Oremek, G.M., Sapoutzis, N., Kramer, W., Bickeboller, R., Jonas, D. *Anticancer Res.* **2000**, *20*, 5095–5098.

134 Lay, A.J., Jiang, X.M., Kisker, O., Flynn, E. et al. *Nature* **2000**, *408*, 869–873.

135 Fox, S.B., Moghaddam, A., Westwood, M., Turley, H. et al. *J. Pathol.* **1995**, *176*, 183–190.

136 Stulik, J., Hernychova, L., Porkertova, S., Knizek, J. et al. *Electrophoresis* **2001**, *22*, 3019–3025.

137 Orntoft, T.F., Thykjaer, T., Waldman, F.M., Wolf, H., Celis, J.E. *Mol. Cell. Proteomics* **2002**, *1*, 37–45.

138 Paweletz, C.P., Ornstein, D.K., Roth, M.J., Bichsel, V.E. et al. *Cancer Res.* **2000**, *60*, 6293–6297.

139 Chang, S.G., Lee, S.J., Lee, S.J., Kimi, J.I. et al. *Anticancer Res.* **1997**, *17*, 113–115.

140 Takenawa, J., Kaneko, Y., Fukumoto, M., Fukatsu, A. et al. *J. Natl. Cancer Inst.* **1991**, *83*, 1668–1672.

141 Solito, E., Coupade, C., Parente, L., Flower, R.J., Russo-Marie, F. *Cytokine* **1998**, *10*, 514–521.

142 Merchant, M., Weinberger, S.R., *Electrophoresis* **2000**, *21*, 1164–1177.

143 von Eggeling, F., Junker, K., Fiedle, W., Wollscheid, V. et al. *Electrophoresis* **2001**, *22*, 2898–2902.

144 Fetsch, P.A., Simone, N.L., Bryant-Greenwood, P.K., Marincola, F.M. et al. *Am. J. Clin. Pathol.* **2002**, *118*, 870–876.

145 Lim, S.O., Park, S.J., Kim, W., Park, S.G. et al. *Biochem. Biophys. Res. Commun.* **2002**, *291*, 1031–1037.

146 Page, M.J., Amess, B., Townsend, R.R., Parekh, R. et al. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 12589–12594.

147 Baer, P.C., Nockher, W.A., Haase, W., Scherberich, J.E. *Kidney Int.* **1997**, *52*, 1321–1331.

148 Sarto, C., Marocchi, A., Sanchez, J.C., Giannone, D. et al. *Electrophoresis* **1997**, *18*, 599–604.

149 Sarto, C., Frutiger, S., Cappellano, F., Sanchez, J.C. et al. *Electrophoresis* **1999**, *20*, 3458–3466.

150 Yang, A.H., Oberley, T.D., Oberley, L.W., Schmid, S.M., Cummings, K.B. *In Vitro Cell Dev. Biol.* **1987**, *23*, 546–558.

151 Oberley, T.D., Sempf, J.M., Oberley, M.J., McCormick, M.L. et al. *Virchows Arch.* **1994**, *424*, 155–164.

152 Motoori, S., Majima, H.J., Ebara, M., Kato, H. et al. *Cancer Res.* **2001**, *61*, 5382–5388.

153 Hirose, K., Longo, D.L., Oppenheim, J.J., Matsushima, K. *FASEB J.* **1993**, *7*, 361–368.

154 Huang, P., Feng, L., Oldham, E.A., Keating, M.J., Plunkett, W. *Nature* **2000**, *407*, 390–395.

155 Emmert-Buck, M.R., Bonner, R.F., Smith, P.D., Chuaqui, R.F. et al. *Science* **1996**, *274*, 998–1001.

156 Schutze, K., Lahr, G. *Nature Biotechnol.* **1998**, *16*, 737–742.

157 Kolble, K. *J. Mol. Med.* **2000**, *78*, B24–B25.

158 Banks, R.E., Dunn, M.J., Forbes, M.A., Stanley, A. et al. *Electrophoresis* **1999**, *20*, 689–700.

159 Craven, R.A., Totty, N., Harnden, P., Selby, P.J., Banks, R.E. *Am. J. Pathol.* **2002**, *160*, 815–822.

160 Emmert-Buck, M.R., Gillespie, J.W., Paweletz, C.P., Ornstein, D.K. et al. *Mol. Carcinog.* **2000**, *27*, 158–165.

161 Unwin, R.D., Harnden, P., Pappin, D., Rahman, D. et al. *Proteomics* **2003**, *3*, 45–55.

162 Simone, N.L., Remaley, A.T., Charboneau, L., Petricoin, E.F. et al. *Am. J. Pathol.* **2000**, *156*, 445–452.

163 Paweletz, C.P., Charboneau, L., Bichsel, V.E., Simone, N.L. et al. *Oncogene* **2001**, *20*, 1981–1989.

164 Wright Jr., G.L., Cazares, L.H., Leung, S.-M., Nasim, S. et al. *Prostate Cancer Prostatic Dis.* **1999**, *2*, 264–276.

165 Celis, A., Rasmussen, H.H., Celis, P., Basse, B. et al. *Electrophoresis* **1999**, *20*, 355–361.

166 Poland, J., Schadendorf, D., Lage, H., Schnolzer, M. et al. *Clin. Chem. Lab. Med.* **2002**, *40*, 221–234.

167 Sinha, P., Hutter, G., Kottgen, E., Dietel, M. et al. *Electrophoresis* **1999**, *20*, 2961–2969.

168 Sinha, P., Kohl, S., Fischer, J., Hutter, G. et al. *Electrophoresis* **2000**, *21*, 3048–3057.

169 Hutter, G., Sinha, P. *Proteomics* **2001**, *1*, 1233–1248.

170 Lichtenfels, R., Ackermann, A., Kellner, R., Seliger, B. *Electrophoresis* **2001**, *22*, 1801–1809.

171 Anderson, N.L., Anderson, N.G. *Mol. Cell Proteomics* **2002**, *1*, 845–867.

172 Lollo, B.A., Harvey, S., Liao, J., Stevens, A.C. et al. *Electrophoresis* **1999**, *20*, 854–859.

173 Pieper, R., Su, Q., Gatlin, C.L., Huang, S.T. et al. *Proteomics* **2003**, *3*, 422–432.

174 Wang, Y.Y., Cheng, P., Chan, D.W. *Proteomics* **2003**, *3*, 243–248.

175 Soussi, T. *Cancer Res.* **2000**, *60*, 1777–1788.

176 Disis, M.L., Knutson, K.L., Schiffman, K., Rinn, K., McNeel, D.G. *Breast Cancer Res. Treat.* **2000**, *62*, 245–252.

177 Takahashi, M., Chen, W., Byrd, D.R., Disis, M.L. et al. *Clin. Cancer Res.* **1995**, *1*, 1071–1077.

178 Sahin, U., Tureci, O., Schmitt, H., Cochlovius, B. et al. *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 11810–11813.

179 Old, L.J., Chen, Y.-T. *J. Exp. Med.* **1998**, *187*, 1163–1167.

180 Scanlan, M.J., Gordan, J.D., Williamson, B., Stockert, E. et al. *Int. J. Cancer* **1999**, *83*, 456–464.

181 Prasannan, L., Misek, D.E., Hinderer, R., Michon, J. et al. *Clin. Cancer Res.* **2000**, *6*, 3949–3956.

182 Brichory, F.M., Misek, D.E., Yim, A.M., Krause, M.C. et al. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 9824–9829.

183 Klade, C.S., Voss, T., Krystek, E., Ahorn, H. et al. *Proteomics* **2001**, *1*, 890–898.

184 Kellner, R., Lichtenfels, R., Atkins, D., Bukur, J. et al. *Proteomics* **2002**, *2*, 1743–1751.

185 Lichtenfels, R., Kellner, R., Atkins, D., Bukur, J. et al. *Biochim. Biophys. Acta* **2003**, *1646*, 21–31.

186 Lichtenfels, R., Kellner, R., Bukur, J., Beck, J. et al. *Proteomics* **2002**, *2*, 561–570.

187 Tureci, O., Sahin, U., Vollmar, E., Siemer, S. et al. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 7608–7613.

188 Anderson, N.G., Anderson, N.L., Tollaksen, S.L. *Clin. Chem.* **1979**, *25*, 1199–1210.

189 Edwards, J.J., Tollaksen, S.L., Anderson, N.G. *Clin. Chem.* **1982**, *28*, 941–948.

190 Marshall, T., Williams, K.M., Vesterberg, O. *Electrophoresis* **1985**, *6*, 47–52.

191 Büeler, M.R., Wiederkehr, F., Vonderschmitt, D.J. *Electrophoresis* **1995**, *16*, 124–134.

192 Grover, P.K., Resnick, M.I. *J. Urol.* **1993**, *150*, 1069–1072.

193 Spahr, C.S., Davis, M.T., McGinley, M.D., Robinson, J.H. et al. *Proteomics* **2001**, *1*, 93–107.

194 Pang, J.X., Ginanni, N., Dongre, A.R., Hefta, S.A., Opiteck, G.J. *J. Proteome Res.* **2002**, *1*, 161–169.

195 Edwards, J.J., Anderson, N.G., Tollaksen, S.L., von Eschenbach, A.C., Guevara, J., Jr. *Clin. Chem.* **1982**, *28*, 160–163.

196 Ostergaard, M., Wolf, H., Orntoft, T.F., Celis, J.E. *Electrophoresis* **1999**, *20*, 349–354.

197 Rasmussen, H.H., Orntoft, T.F., Wolf, H., Celis, J.E. *J. Urol.* **1996**, *155*, 2113–2119.

198 Stenman, U.H. *Clin. Chem.* **2002**, *48*, 1206–1209.

199 Konety, B.R., Getzenberg, R.H. *J. Urol.* **2001**, *165*, 600–611.

200 Ross, J.S., Cohen, M.B. *Adv. Anat. Pathol.* **2001**, *8*, 37–45.

201 Huang, S., Rhee, E., Patel, H., Park, E., Kaswick, J. *Urology* **2000**, *55*, 227–230.

202 Ozer, G., Altinel, M., Kocak, B., Yazicioglu, A., Gonenc, F. *Urology* **2002**, *60*, 593–597.

203 Rodriguez-Cuartero, A., Barcelona-Martin, F., Perez-Blanco, F.J. *Clin. Nephrol.* **2000**, *53*, 288–290.

204 Vlahou, A., Schellhammer, P.F., Mendrinos, S., Patel, K. et al. *Am. J. Pathol.* **2001**, *158*, 1491–1502.

205 Rogers, M.A., Clarke, P., Noble, J., Munro, N.P. et al. *Cancer Res.* **2003**, *63*, 6971–6983.

206 Kovacs, G. In *Renal and Adrenal Tumors* (A. Belldegrun, A.W.S. Ritchie, R.A. Figlin, R.T.D. Oliver, J.E.D. Vaughan, Eds). Oxford University Press, New York, **2003**, pp 90–97.

6
Heat Shock Protein 27 in Cancer

Cecilia Sarto, Fulvio Magni, Cristina Valsecchi, and Paolo Mocarelli

6.1
Introduction

The heat shock proteins (HSP), also termed molecular chaperones, are a family of several ubiquitous proteins that have a role in increased resistance in damaged cells by heat shock. They are separated into different groups according to their approximate molecular weight. The term heat shock protein only partially explains their function. Indeed these housekeeping proteins can be induced by a range of cellular insults, including increased temperature, oxidative stress, nutritional deficiencies, ultraviolet irradiation, exposure to chemicals such as ethanol, viral infection, and ischemia-reperfusion. Whilst the HSPs with high molecular weight, such as HSP70 and HSP90 have been largely studied and their chaperone activity documented, so far too little attention has been given to the smaller forms, including HSP27. Similar to HSPs with a higher molecular weight, HSP27 and the other small HSPs such as HSP40, α-crystallin, HSP17, and myotonic dystrophy protein kinase binding protein (MKBP) [1] participate in protein folding under normal and stress conditions causing cellular damage. This chapter will focus on HSP27 and, in particular, on the relationships known to date between HSP27 structure, post-translational modifications, functions, and expression in human cancer.

6.2
Genomic Aspects

Human HSP27 is encoded by an inducible gene mapping at chromosome 7q. General features of heat shock protein gene transcription and the DNA bond with different members of the heat shock factor (HSF) family have been elucidated without specific reference to HSP27 [2, 3]. Actually, ubiquitous HSFs are genetic regulatory switches of HSPs in response to different stress stimuli, and have been studied with attention to synthesis modulation of HSPs with high molecular weight such as HSP90 and HSP70. In contrast, most of reports of HSP27 describe downstream regulation pathways of its different functions. HSF1 is the principal factor in vertebrates present in the cytoplasm as a latent monomeric

Biomedical Application of Proteomics
Edited by J.-C. Sanchez, G. L. Corthals, D. F. Hochstrasser

ISBN: 3-527-30807-5

molecule that is unable to bind DNA. Following stress conditions, a Ras-dependent phosphorylation converts latent monomeric HSF1 to active trimers, which translocate to the nucleus. The trimers bind heat shock element (HSE) in a promoter region of the HSP gene consisting of arrays of the 5-bp unit nGAAn consensus motif arranged as inverted repeats [4].

Although molecular triggers for small HSPs are largely unknown, the regulatory action of HSE is also believed to play a role in transcriptional activation for these shock proteins, because the three inverted repeats of consensus motif sequence have been found in the small HSP genes of all species examined [5, 6]. Even if stress stimuli are prolonged, the generation of HSPs is short lasting in order to prevent adverse influence on protein homeostasis and cellular function. Therefore the HSF1 activity is down-regulated by HSP70 and heat shock protein-binding factor 1 [7, 8].

6.3 Structure

The primary structure of HSP27, as well as the other small HSPs, is characterized by a conserved sequence of 88–100 residues, termed the α-crystallin domain, located in the C-terminal region [9] flanked by a variable hydrophobic N-terminal region and a short C-terminal extension (Fig. 6.1). Sequence modeling and physical analyses show that the secondary structure of small heat shock/α-crystallin proteins is predominantly a β-sandwich with two antiparallel β-sheets. The flexible C-terminal extension contributes to chaperone activity by enhancing the solubility of small heat shock/α-crystallin proteins. Crystallography, site-directed spin-labeling, and yeast two-hybrid selection have demonstrated that during oligomer assembly dimers and tetramers are formed by interactions of disulfide bond of cysteine 137 and regions within the α-crystallin domain with the N-terminus [10, 11]. The quaternary structure is one of two prerequisites to HSP27 function as a molecular chaperone.

Up to 40 monomers can be assembled to form large cytosolic aggregates that vary in size from 150 kDa to 1000 kDa and in quaternary structure with a wide range of diameters. Formation of large oligomers is dependent on the variable WDPF-domain located in the very terminal part of the N-terminal sequence (Fig. 6.1) [1]. The involvement of a flexible C-terminus in formation of large oligomeric species is less obvious: it is possible that participates in the interactions of the small HSPs with the target proteins affecting general hydrophobicity and in the stabilization of oligomer formation [12]. Furthermore, the phosphorylation of HSP27 provokes dissociation from multimers to small oligomers, probably as tetrameric forms, which are active at conditions where other proteins are denaturing, so they can down-regulate chaperone activity and influence actin polymerization [13–15]. Phosphorylation levels influence the biological activity via the MAP kinase pathway. Oligomer size is regulated by a stress-responsive cascade including MAPKAP kinase 2/3 and p38, but it is not clear if the total amount of HSP27 rather than un- or monophosphorylated protein is mainly responsible for protec-

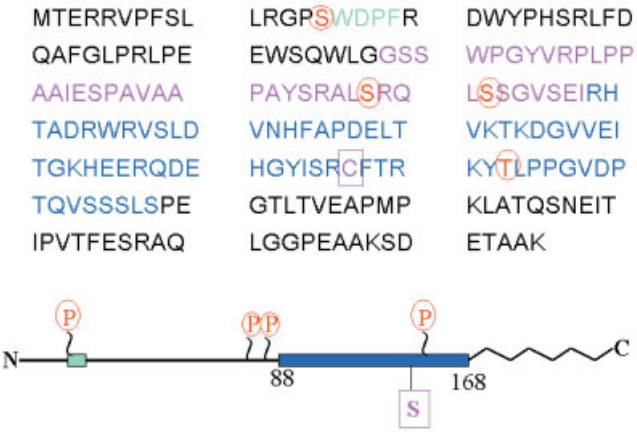


Figure 6.1 Primary sequence of HS27 from www.expasy.org. Circles and squares indicate possible phosphorylation and *S*-thiolation sites. Green letters indicate the N-terminal domain responsible of large oligomers formation, while the pink and the blue letters indicate the sequence involved in interaction with cytochrome *c* and the α domain, respectively.

tive effects on stressed cells. Dynamic equilibrium of different levels of the quaternary structure, depending on several signaling pathways, seems to be one reason for regulating the performance of chaperone activity of HSP27 and all other small HSPs [16, 17].

6.4 Functions

Like other members of the chaperone family, HSP27 participates in thermotolerance in mammalian cells [18] and contributes to the balance between cell survival and cell death by preventing protein aggregation during several types of stress and interacting with wide substrate range [19]. Under physiological conditions, its abundance depends on tissue type and development, cellular cycle, differentiation, and it can be strongly induced in oncogenic status of cells. As a chaperone protein, it plays a role in several cellular functions including cell proliferation, immune response, intracellular protein movement, cytoskeletal dynamics, and drug sensitivity. Contrary to heat shock proteins with high molecular weight that are regulated by ATP, the small HSPs and 40 kDa HSP do not have an ATP-binding site, although ATP somehow affects their structure and their interaction with partially denatured proteins to prevent aggregation and to complex them with chaperones possessing ATPase activity [20]. Post-translational modifications including oligomerization and phosphorylation characterizing HSP27 and all small HSPs are essential for their chaperone activity.

The main cellular function that is widely documented is related to actin action in all kind of cells. During normal cellular events HSP27 interacts directly with actin and intermediate filaments by phosphorylation to prevent specifically non-covalent filament/filament interactions of the intermediate filaments [21, 22]. Data from a recent publication suggests that in response to various stress elements phosphorylation of HSP27 protects actin filaments from fragmentation and preserves their focal contacts fixed at the cell membrane [23]. Moreover, interaction of HSP27 with actin appears specific in renal epithelial cells subjected to energy depletion, preserving architecture in specific intracellular domains. In renal epithelial cells actin-capping activity aimed to prevent actin disruption and chaperone activity through interactions with aggregates of non-native proteins are two of the multifunctional properties of HSP27 that have been determined [24].

Enhanced resistance of actin stress fibers to disaggregation by hyperthermia and cytochalasin D has been obtained by artificially increasing HSP27 expression in cells by gene transfection, thus suggesting that a major role of HSP27 is regulation of actin filament dynamics [25]. This protein is rapidly phosphorylated in response to growth factors localized in lamellopodia of fibroblastic cells and its overexpression corresponds to accumulation of cortical F-actin, a stress fiber actin, and increased fluid phase of pinocytotic activity in quiescent cells. Furthermore, the function of HSP27 is activated by phosphorylation via activation of p38, the major physiological activator of MAPKAP kinase 2/3 during stress [26], maintaining myofibrilla integrity against oxidative and mechanical stresses. It has been reported that phosphorylation of pre-existing HSP27 seems to be a first phase of stress response while HSP27 overexpression seems to correspond to a second phase, at a time when phosphorylation is already down-regulated [27]. Nevertheless, until now it is not clear if HSP27 displays different cellular functions in the several phases at different stages of stress response.

Among the protein components of molecular machinery of cell death, chaperone proteins may directly affect the apoptotic signaling pathway that is physiologically involved in cell death regulation to balance cell number, size, and differentiation in embryonic development.

Two different regulatory functions have been suggested to explain the modulation mechanisms of apoptosis. Whereas HSP60 and HSP10 support the execution of apoptosis, promoting the proteolytic maturation of precursor caspases [28, 29], HSP70 [30, 31] and HSP27 [32, 33] have been shown to exert a negative influence on apoptosis, either inhibiting the activation of procaspase-9 or preventing association *in vivo* and *in vitro* of Daxx protein with Fas and Ask1 [34]. The constant equilibrium between expression and action of the different HSPs can modulate the apoptotic pathway and determine the fate of stressed cells. Although HSP27 is localized in the cytosol it can interact with the outer mitochondrial membranes to interfere with apoptosis and to inhibit formation of the apoptosome. In addition it moves to the perinuclear region in response to a plethora of stress stimuli [35].

HSP27 does not interact directly with procaspase-9, but rather specifically interferes with the activation of cytochrome *c*/Apaf-1/dATP complex. In particular, amino acids 57–88 from N-terminal domain, and at least cysteine 137 of the

α-crystallin domain are involved in the HSP27 interaction with cytochrome *c* (Fig. 6.1). In addition, HSP27 interacts with serine-threonine kinase Akt, p38 MAPK, MAPK-activated protein kinase-2 to regulate neutrophil apoptosis inducing a prolonged inflammatory response [36]. Finally, only the phosphorylated HSP27 form interacts and inhibits Daxx apoptotic protein that is associated with nuclear substructures. Although in recent years numerous studies have been carried out to clarify different aspects of small HSP functions, a comprehensive picture is still lacking.

6.5 HSP27 Expression in Cancer

To date, the precise mechanisms of action of HSP27 are not yet known, but protection from apoptosis was reported as an important role of this small stress protein. Uncontrolled proliferation and unregulated signaling apoptosis are mechanisms strongly associated with neoplastic transformation and tumor progression. Apoptosis has emerged as the major mechanism by which anticancer agents eliminate the preneoplastic and cancer cells [37]. Moreover the overexpression of HSP27 was found to confer resistance against actin fragmentation, mediating an adaptive response to oxidative stress including carcinogens, anticancer drugs, and other xenobiotics [38]. Therefore it is not surprising that a large number of cancers show increased levels of HSP27 associated with different expression of phosphorylated isoforms compared with normal cells [25]. High expression of HSP27 in breast, endometrial, and gastric cancer has been associated with an inverse relation with cell proliferation, metastasis, poor prognosis, and resistance to chemo- or radiotherapy [39]. High expression of HSP27 has also been detected in the serum of patients with breast cancer using two-dimensional electrophoresis (2-DE) followed by mass spectrometry (MS) identification [40]. Based on these results, HSP27 has been proposed as a screening or diagnostic marker for breast cancer. Moreover, a 2-DE-based proteomic study of human breast cancer MCF-7 cell line has revealed down-regulation effect of doxorubicin (DOX) on expression and phosphorylation status of HSP27, inducing decreased rate of cell proliferation and differentiation [41].

A recent work has shown that high levels of HSP27 prevent apoptosis induced by curcumin in human colon cancer cell lines derived from primary tumor. Curcumin is an ingredient of turmeric, possessing antiproliferative, antimutagenic, anticarcinogenic properties that induce apoptosis via the caspase-9 pathway, retaining cytochrome *c* within mitochondria. It has been shown that low expression of HSP27 is not efficient at inhibiting apoptosis [42]. In addition, the HT-29 colon cancer cell line has shown a relative resistance to butyrate-induced apoptosis following a simultaneous up-regulation of pro-apoptotic as well as anti-apoptotic proteins such as HSP27 [43]. Despite the discrepancies between these data, the common feature is the anti-apoptotic interaction of HSP27, suggesting that this protein might be a potential therapeutic target for chemotherapy.

Proteomic analysis (2-DE) of human astrocytes and glioblastoma cell lines show differential expression of HSP27 among high-, low-grade, and non-malignant cell types that seems to correlate with the degree of glial tumor malignancy [44]. Furthermore, in human high-grade astrocytomas, HSP27 was observed to be coexpressed with p-Jun, AP-1 (activator protein 1), and c-Jun and c-Fos oncoproteins. Depolymerization of the cytoskeleton in glial cells is activated by induction of AP-1; while HSP27 expression activates p-Jun and p38 MAPK and induces c-Jun expression via tyrosine kinase signaling pathway. Therefore it is possible to speculate that actin filament action is regulated by p38 which is an upstream activator of HSP27 phosphorylation [45]. In contrast, genomic and proteomic analyses using 2-DE of primary neuroblastoma tumors showed that overexpression of HSP27 was associated with tumor differentiation and favorable outcome [46]. In response to tumor necrosis factor alpha (TNFα), phosphorylated HSP27 isoforms were increased and they underwent changes in their intracellular distribution and structural organization. TNFα induced the formation of large HSP27 aggregates concomitantly with increasing in phosphorylation. In particular, using 2-DE western blot immunodetection, one isoform of HSP27 phosphorylation was found to be induced early by TNFα, although differentially phosphorylated forms were not confirmed, unlike in the majority of studies, because of the difficulty of analyzing them [47].

Studies were also performed to test pattern expression of HSP27 in urological tumors including human prostate cancer [48, 49], renal cell carcinoma, and testicular cancer [50], confirming that HSP27 overexpression correlates with proliferation, poor differentiation, and resistant therapeutic outcome.

6.6
Post-translational Modification and Proteomic Tools

Proteomic studies based on MS analysis have been performed on various cancers, such as kidney, breast, lung cancer, ovarian, and prostate cancer [51]. Peptide mass fingerprinting (PMF) is one of the several strategies developed for the identification of proteins by MS and it is based on the ability to accurately determine the mass of peptides after specific protein cleavage by endoproteolytic enzymes. Figure 6.2 shows a typical matrix-assisted laser desorption ionization/time-of-flight (MALDI-TOF) mass spectrum of HSP27 after tryptic digestion. Here accurate masses of MS-analyzed peptides are correlated with calculated masses of peptides from proteins in sequence databases, with the same cleavage specificity of the endoproteases used during the experiment (see Fig. 6.3). For a detailed review on MS techniques see Corthals et al. [52].

Tandem MS (MS/MS) can also be used for protein identification and for *de novo* sequencing. With MS/MS single peptides are selected for collision with an inert gas after an initial MS scan (survey scan). During this collision process peptides fragment into smaller fragments mostly along the peptide backbone. Subsequently the masses of these smaller fragments are measured again by MS (hence

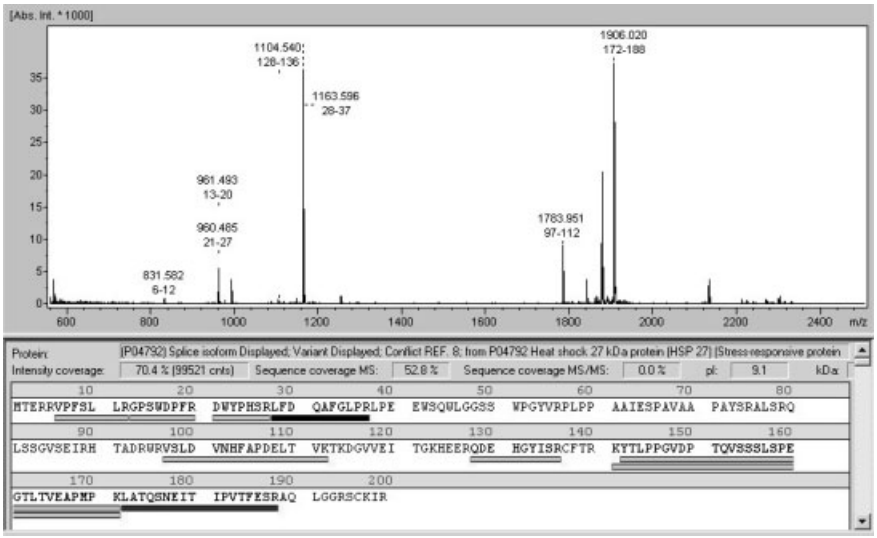


Figure 6.2 Typical MALDI-TOF mass spectrum of one isoform of HSP27 deriving from renal cancer tissue. Sequence labels in red indicate those tryptic peptides which were matched at the selected search criteria.

MS/MS). This fragmentation data, more commonly known as product ion spectra or daughter ion spectra, can also be correlated with sequence database information to yield a protein identification. This process has been automated in almost all MS/MS instruments and is usually referred to as collision-induced dissociation (CID). MS/MS measurements typically give higher identification rates as the primary sequence is captured during an experiment, as opposed to only peptide masses with PMF. The MS/MS spectra can be obtained with either electrospray ionization-MS (ESI-MS) or with matrix-assisted laser desorption/ionization MS (MALDI-MS).

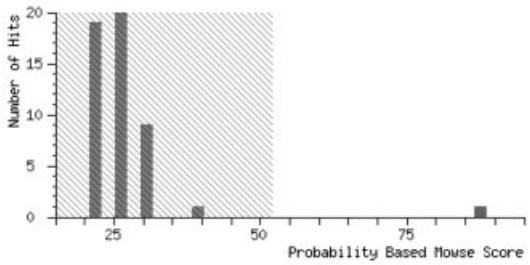
Mass spectrometry has also been applied to heat shock protein studies in order to define their structure and their importance in several pathologies. Despite the fact that HSP27 is reported to be overexpressed in a wide number of human cancers [53] few studies are focused on HSP27 identification and on characterization of its post-translational modifications by MS, and only a few articles describe the relationship of HSP27 isoforms with human cancer. Nevertheless some important results regarding this protein have been obtained in cancer research, in particular those related to liver cancer [52], human colon cancer [43], renal cell carcinoma [54], breast tumors [55], in urothelial papillomas [56], and thyroid tissue [57].

Two-dimensional electrophoresis studies comparing human hepatoma cell line BEL-7404 and normal liver cell line L-02 proteomes have shown an altered level of only one isoform of HSP27 present in human hepatocellular carcinoma cells [52]. The amount of HSP27 in BEL-7404 cells was 1.8-fold of that in L-02. The HSP27 protein was identified by in-line coupled liquid chromatography/ESI-MS/MS (LC/

User : Fulvio
Email : FULVIO.MAGNI@UNIMIB.IT
Search title : hsp27
Database : Sprot 20030511 (126147 sequences; 46451541 residues)
Taxonomy : Homo sapiens (human) (9483 sequences)
Timestamp : 23 Jun 2003 at 13:13:42 GMT
Top Score : 87 for HS27_HUMAN, (P04792) Heat shock 27 kDa protein (HSP 27) (Stress-response)

Probability Based Mowse Score

Score is -10*Log(P), where P is the probability that the observed match is a random event.
Protein scores greater than 52 are significant (p<0.05).



Concise Protein Summary Report

Switch to full Protein Summary Report

To create a bookmark for this report, right click this link: Concise Summary Report (hsp27)

Re-Search All | Search Unmatched

1. HS27_HUMAN Mass: 22826 Total score: 87 Peptides matched: 7
(P04792) Heat shock 27 kDa protein (HSP 27) (Stress-response)

Figure 6.3 Identification of HSP27 by database search based on Swiss-Prot, as database, and on the algorithm present in Mascot program.

ESI-MS/MS). Two isoforms of HSP27 were observed for the human hepatoma cell (JX-0) proteome [58]. The peptide VSLDVNHFAPDELTVK was used for the identification of the proteins because this sequence is unique to human HSP27. However, the structural difference between the two isoforms was not clarified. A possible different phosphorylation modification was suggested although it was not observed.

Alterations of the HSP27 expression level might be the result of differential chemical modifications, suggesting a possible dynamic role of post-translational modifications of proteins on the growth of hepatoma cells [52]. However other chemical modifications may be applied to HSP27 besides phosphorylation [52, 58]. A study using functional proteomics to profile human colon cancer cells (HT-29) before and after treatment with sodium butyrate has also been reported [43].

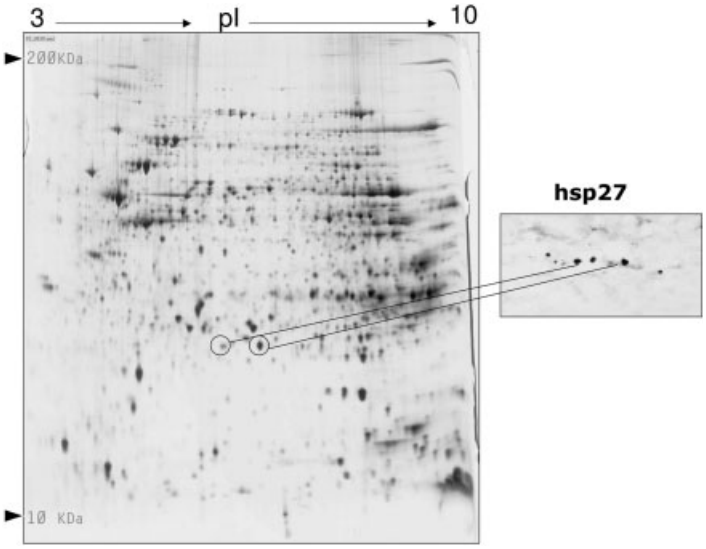


Figure 6.4 Silver-stained 2-D PAGE of normal kidney cortex tissue showing two isoforms of HSP27 indicated by circles and identified by MALDI-TOF and immunodetection using anti-HSP27 antibody. Zoomed region of a 2-DE western blot showing several isoforms of HSP27 after immunodetection by specific anti-HSP27 antibody.

Butyrate-treated cells exhibited growth inhibition and apoptosis, accompanied by proteome alterations within the cell. HSP27 was found to be strongly up-regulated in these butyrate-treated HT-29 cells. Identification was conferred by PMF and database searching was carried out using parameters such as possible oxidation of methionine, N-terminal acetylation, carboxyamido-methylation of cysteine and phosphorylation of serine, threonine, and tyrosine. Positive identification was based on the following criteria: MS match for at least four peptide mass accuracy lower than 50 ppm, at least 20% coverage, with M_r and pI matching the estimates or published values. Detection of more than two isoforms of HSP27 was reported in various cancer tissues. Three isoforms of HSP27 were observed in renal cell carcinoma [54] and in breast cells (one of these isoforms seem to be blocked at the N-terminus) [55], only one isoform in a neuroblastoma cell line [59], four isoforms in breast carcinoma cells after Newcastle disease virus (NDV) infection (presence of HSP27 phosphorylated forms were detected using a specific antiserum) [60] and two isoforms in neoplastic thyroid tissues [57].

Identification of all these proteins as HSP27 has been carried out by MS either by PMF or MS/MS. Figure 6.4 shows a silver-stained 2-D PAGE gel of normal kidney cortex tissue showing two isoforms of this protein identified by MALDI-TOF analysis (see Fig. 6.2). It is possible to visualize more than two isoforms by 2-DE western blot immunodetection using a specific anti-HSP27 antibody (unpublished data), but only two of these isoforms have been detected by MS so far.

6.6.1 Phosphorylation

The most important post-translational modification reported for HSP27 is its phosphorylation. It is well accepted that phosphorylation can occur at a site containing the consensus sequence RXXS which is common to several kinases. In the HSP27 sequence there are three serines in position 78, 82, and 15 that have been reported to be modified by kinases in several conditions [61–63]. Characterization of these phosphorylated forms of HSP27 has been done using different techniques. None of them is based on mass spectrometry. Digestion of ^{32}P-labeled HSP27 with trypsin and fractionation of the peptides by reverse-phase high-performance liquid chromatography and detection of radioactive peptides followed by chemical microsequencing has been used in human HeLa cells study [61]. Phosphorylation of HSP27 was detected by autoradiography after SDS-PAGE separation [62] or by immunodetection with specific antisera for the phosphorylated forms of HSP27 [55, 60] and only the identity of these proteins as HSP27 was ensured by MALDI-TOF or ESI-MS/MS [59].

To date only one report [44] has revealed the detection of phosphorylated peptides by MS, following *in-gel* enzymatic digestion of HSP27. These samples were obtained from fetal human astrocytes. Figure 3 in [44] shows the ESI-MS/MS spectra of the peptides QLSSGVSEIR (parent ion $[M+2H]^{2+}$ at m/z 538.33) and of the corresponding phosphorylated isoform (parent ion $[M+2H]^{2+}$ at m/z 578.32). In the latter spectrum the ion due to the loss of H_3PO_4 (m/z 816.5) is clearly distinguishable. Recently an additional putative phosphorylation site of HSP27 for cAK and cGK at threonine 143 has been reported in human platelets incubated with [^{32}P]orthophosphate [62]. This result was proved by the absence of incorporation of P-32 when Thr143 was substituted with glutamic acid.

6.6.2 *S*-Thiolation, Oxidation, and Others

Other possible post-translational modifications of HSP27 have been reported to occur in human myocardial tissue [64]. Fifty-nine spots deriving from 2-DE separation of human myocardial tissue were reacted with anti-HSP27 antibody. Identification of all these proteins was carried out by MALDI-TOF and in some cases by post source decay (PSD)-MALDI-TOF analyses (see review by Corthals et al. [52]). It is interesting to note that only one of these HSP27 isoforms was postulated to be phosphorylated at Ser82 based on PSD-MALDI-TOF analysis. In the PSD mass spectrum of the peptide deriving from the tryptic cleavage of HSP27 with a molecular mass of 1139.6 Da two product ions at m/z 1059.6 and at m/z 1041.0 were observed. The authors interpreted these ions as a loss of phosphoric acid (–98 Da) by β-elimination (ion at 1041.0 Da) followed by partial rehydration that originated the ion at 1059.6 Da. All other forms of HSP27 were evaluated by immuno-staining using anti-phosphothreonine, anti-phosphotyrosine, and anti-phosphoserine and not found to be phosphorylated.

Nevertheless several HSP27 were found to be modified when interpretation of the MALDI-TOF spectra were done considering other modifications like methionine and tryptophan oxidation and ammonia elimination of N-terminal glutamine yielding pyroglutamic acid. Degradation either by oxidation or by enzymatic cleavage originating from smaller forms of HSP27 can occur *in vivo*. Lower molecular masses of the HSP27 derived from degradation were detected and identified in cardiomyopathic heart tissues [65].

Recently it has been shown that HSP27 can also undergo to *S*-thiolation during cardiac oxidative stress in rat [63]. *S*-Thiolation of rat heart HSP27 occurred at the only cysteine residue present in this protein. Therefore we may speculate that under oxidative stress conditions HSP27 could also be expected to be modified in humans at Cys137, which may also be a target for *S*-thiolation.

6.7 Perspectives

In recent years the number of investigations focused on the characterization of the genomic, functional, and proteomic aspects of HSP27 and on understanding its multiple actions in diseases including cancers has grown. As previously described, altered expression and post-translational modifications in various cancers are probably related to the different HSP27 functions, but to date neither transcription regulation nor post-translational modifications have been precisely characterized. A huge amount of work still needs to be done in order to identify all post-translational modifications and to explain the complex role of this protein in cancer. Currently available proteomic methods to separate and to identify proteins have high sensitivity and specificity, but up to now results have failed to characterize details of modifications of this protein. Certainly the sensitivity of mass spectrometers is not sufficient to detect small amounts of proteins and to solve the characterization of the numerous isoforms pointed out by immunodetection.

We anticipate that in the immediate future with the progress of technology more information will be uncovered about chemical and structure modifications as well as the role of HSP27 in biological systems. Knowledge of the relationship between overexpression of several isoforms of HS27 and the development and progression of cancer will be very useful in studying human cancer.

6.8 Acknowledgements

This paper has been supported by grants COFIN 2001 and FIRB 2001.

6.9 References

1 A. Suzuki, Y. Sugiyama, Y. Hayashi et al. *J. Cell Biol.* **1998**, *140*, 1113–1124.
2 R. I. Morimoto. *Science* **1993**, *259*, 1409–1410.
3 J. Lis, C. Wu. *Cell* **1993**, *74*, 1–4.
4 M. Fernandes, H. Xiao, J. T. Lis. *Nucleic Acids Res.* **1994**, *22*, 167–173.
5 E. Frohli, A. Aoyama, R. Klemenz. *Gene* **1993**, *128*, 273–277.
6 R. Klemenz, E. Frohli, R. H. Steiger, R. Schafer, A. Aoyama. *Proc. Natl Acad. Sci. USA* **1991**, *88*, 3652–3656.
7 Y. Shi, D. D. Mosser, R. I. Morimoto. *Genes Dev.* **1998**, *12*, 654–666.
8 S. H. Satyal, D. Chen, S. G. Fox, J. M. Kramer, R. I. Morimoto. *Genes Dev.* **1998**, *12*, 1962–1974.
9 T. D. Ingolia, E. A. Craig. *Proc. Natl Acad. Sci. USA* **1982**, *79*, 525–529.
10 R. L. van Montfort, E. Basha, K. L. Friedrich, C. Slingsby, E. Vierling. *Nature Struct. Biol.* **2001**, *8*, 1025–1030.
11 T. H. MacRae. *Cell. Mol. Life Sci.* **2000**, *57*, 899–913.
12 W. C. Boelens, Y. Croes, M. de Ruwe, L. de Reu, W. W. de Jong. *J. Biol. Chem.* **1998**, *273*, 28085–28090.
13 R. Benndorf, K. Hayess, S. Ryazantsev, M. Wieske, J. Behlke, G. J. Lutsch. *J. Biol. Chem.* **1994**, *269*, 20780–20784.
14 T. Rogalla, M. Ehrnsperger, X. Preville et al. *J. Biol. Chem.* **1999**, *274*, 18947–18956.
15 J. L. Martin, E. Hickey, L. A. Weber, W. H. Dillmann, R. Mestril. *Gene Expr.* **1999**, *7*, 349–355.
16 M. P. Bova, H. S. McHaourab, Y. Han, B. K. Fung. *J. Biol. Chem.* **2000**, *275*, 1035–1042.
17 D. A. Haley, M. P. Bova, Q. L. Huang, H. S. Mchaourab, P. L. Stewart. *J. Mol. Biol.* **2000**, *298*, 261–272.
18 P. Chretien, J. Landry. *J. Cell. Physiol.* **1988**, *137*, 157–166.
19 U. Jacob, M. Gaestel, K. Engel, J. Buchner. *J. Biol. Chem.* **1993**, *268*, 1517–1520.
20 X. Y. Wang, X. Chen, H. J. Oh, E. Repasky, L. Kazim, J. Subjeck. *FEBS Lett.* **2000**, *465*, 98–102.
21 Y. Zhu, S. O'Neill, J. Saklatvala, L. Tassi, M. E. Mendelsohn. *Blood* **1994**, *84*, 3715–3723.
22 M. D. Perng, L. Cairns, P. van den Ijssel, A. Prescott, A. M. Hutcheson, R. A. Quinlan. *J. Cell Sci.* **1999**, *112*, 2099–2112.
23 C. Schafer, P. Clapp, M. J. Welsh, R. Benndorf, J. A. Williams. *Am. J. Physiol.* **1999**, *277*, C1032–1043.
24 S. K. van Why, A. S. Mann, T. Ardito et al. *J. Am. Soc. Nephrol.* **2003**, *13*, 98–106.
25 J. N. Lavoie, E. Hickey, L. A. Weber, J. Landry. *J. Biol. Chem.* **1993**, *268*, 24210–24214.
26 J. Guay, H. Lambert, G. Gingras-Breton, J. N. Lavoie, J. Huot, J. Landry. *J. Cell Sci.* **1997**, *110*, 357–368.
27 J. Landry, P. Chretien, A. Laszlo, H. Lambert. *J. Cell. Physiol.* **1991**, *147*, 93–101.
28 S. Xanthoudakis, S. Roy, D. Rasper et al. *EMBO J.* **1999**, *18*, 2049–2056.
29 A. Samali, T. G. Cotter. *Exp. Cell Res.* **1996**, *223*, 163–170.
30 H. M. Beere, B. B. Wolf, K. Cain et al. *Nature Cell Biol.* **2000**, *2*, 469–475.
31 A. Saleh, S. M. Srinivasula, L. Balkir, P. D. Robbins, E. S. Alnemri. *Nature Cell Biol.* **2000**, *2*, 476–483.
32 C. Garrido, J. M. Bruey, A. Fromentin, A. Hammann, A. P. Arrigo, E. Solary. *FASEB J.* **1999**, *13*, 2061–2070.
33 J. M. Bruey, C. Ducasse, P. Bonniaud et al. *Nature Cell. Biol.* **2000**, *2*, 645–652.
34 S. J. Charette, J. N. Lavoie, H. Lambert, J. Landry. *Mol. Cell. Biol.* **2000**, *20*, 7602–7612.
35 D. Neumann, U. zur Nieden, R. Manteuffel, G. Walter, K.-D. Scharf, L. Nover. *Eur. J. Cell Biol.* **1987**, *43*, 71–81.
36 M. J. Rane, Y. Pan, S. Singh et al. *J. Biol. Chem.* **2003**, *278*, 27828–27835.
37 A. Samali, B. Zhivotovsky , D. Jones, S. Nagata, S. Orrenius. *Cell Death Differ.* **1999**, *6*, 495–496.
38 J. Huot, F. Houle, D. R. Spitz, J. Landry. *Cancer Res.* **1996**, *56*, 273–279.

39 L. M. Vargas-Roig, M. A. Fanelli, L. A. Lopez et al. *Cancer Detect. Prev.* **1997**, *21*, 441–451.

40 Z. Rui, J. Jian-Guo, T. Yuan-Peng, P. Hai, R. Bing-Gen. *Proteomics* **2003**, *3*, 433–439.

41 S. T. Chen, T. L. Pan, Y. C. Tsai, C. M. Huang. *Cancer Lett.* **2002**, *181*, 95–107.

42 R. Rashmi, T. R. Santhosh Kumar, D. Karunagaran. *FEBS Lett.* **2003**, *538*, 19–24.

43 S. Tan, T. K. Seow, R. C. M. Y. Liang et al. *Int. J. Cancer* **2002**, *98*, 523–531.

44 R. Zhang, T. L. Tremblay, A. McDermid, P. Thibault, D. Stanimirovic. *Glia* **2003**, *42*, 194–208.

45 M. Assimakopoulou, J. Varakis. *J. Cancer Res. Clin. Oncol.* **2001**, *127*, 727–732.

46 K. Wimmer, R. Kuick, D. Thoraval, S. M. Hanash. *Electrophoresis* **1996**, *17*, 1741–1751.

47 P. Mehlen, A. Mehlen, D. Guillet, X. Preville, A. P. Arrigo. *J. Cell Biochem.* **1995**, *58*, 248–259.

48 P. A. Cornford, A. R. Dodson, K. F. Parsons et al. *Cancer Res.* **2000**, *60*, 7099–7105.

49 L. Bubendorf, M. Kolmer, J. Kononen et al. *J. Natl Cancer Inst.* **1999**, *91*, 1758–1764.

50 M. Takashi, S. Katsuno, T. Sakata, S. Ohshima, K. Kato. *Urol. Res.* **1998**, *26*, 395–399.

51 A. A. Alaya, B. Franzen, G. Auer, S. Linder. *Electrophoresis* **2000**, *21*, 1210–1271.

52 G. L. Corthals, S. Gygi, R. Aebersold, S. P. Patterson. In *Proteome Research: 2D Gel Electrophoresis and Detection Methods* (Rabilloud, T., Ed.). Springer, Heidelberg, **1999**, pp 197–231.

53 C. Sarto, P.-A. Binz, P. Mocarelli. *Electrophoresis* **2000**, *21*, 1218–1226.

54 R. Lichtenfels, R. Kellner, J. Bukur et al. *Proteomics* **2002**, *2*, 561–570.

55 S.-T. Chen, T.-L. Pan, Y.-C. Tsai, C.-M. Huang. *Cancer Lett.* **2002**, *181*, 95–107.

56 J. E. Celis, P. Celis, H. Palsdottir et al. *Mol. Cell. Proteomics* **2002**, *1*, 269–279.

57 C. Srisomsap, P. Subhasitanont, A. Otto et al. *Proteomics* **2002**, *2*, 706–712.

58 L.-H. Yu, X.-X. Shao, W.-L. Jiang et al. *Electrophoresis* **2001**, *22*, 3001–3008.

59 B. K. Shin, H. Wang, A. M. Yim et al. *J. Biol. Chem.* **2003**, *278*, 7607–7616.

60 L. Bai, J. Koopmann, C. Fiola, P. Fournier, V. Schirrmacher. *Int. J. Oncol.* **2002**, *21*, 685–694.

61 J. Landry, H. Lambert, M. Zhou et al. *J. Biol. Chem.* **1992**, *267*, 794–803.

62 E. Butt, D. Immler, H. E. Meyer, A. Kotlyarov, K. Laab, M. Gaestel. *J. Biol. Chem.* **2001**, *276*, 7108–7113.

63 P. Eaton, W. Fuller, M. J. Shattock. *J. Biol. Chem.* **2002**, *277*, 21189–21196.

64 C. Scheler, E.-C. Muller, J. Stahl, U. Muller-Werdan, J. Salnikow, P. Jungblut. *Electrophoresis* **1997**, *18*, 2823–2831.

65 C. Scheler, X.-P. Li, J. Salnikow, M. P. R. Jungblut. *Electrophoresis* **1999**, *20*, 3623–3628.

7
Proteomic Approaches for Biomarker Discovery in Colorectal Cancer

Richard J. Simpson and Donna S. Dorow

7.1
Introduction and Background to Colorectal Cancer

Colorectal cancer (CRC) is a leading cause of cancer death in the western world [1–3]. By the age of 70, at least 50% of people in western populations develop a colorectal tumor, and for about 10% of these individuals, progression to malignancy ensues. CRC is thought to arise from the cumulative effects of multiple mutations within colonic epithelial cells, allowing escape from growth and regulatory control mechanisms. This stepwise progression of mutations facilitates the histological transition from normal colonic mucosa to adenoma to carcinoma (Fig. 7.1). While the majority of cases of CRC are sporadic, a significant minority (5% of all CRCs) occur as a result of an inherited genetic mutation [4]. The most common hereditary syndromes are familial adenomatous polyposis (FAP) and hereditary non-polyposis colon cancer (HNPCC) [5, 6]. Patients with these syndromes usually have a family history of CRC presenting at an early age. Similarly, sporadic tumors can be divided into at least two major subtypes, one resembling FAP found predominantly in the distal region (left side) of the colon, and the other more like HNPCC, mainly located on the proximal region (right side) of the colon (Fig. 7.2). Of these, left-sided tumors are more aggressive, but if diagnosed early, both types respond to treatment.

In order to improve outcomes for CRC patients there is a pressing need to identify biomarkers for the early detection (diagnostic markers), prognosis (prognostic indicators), tumor responses (predictive markers), and disease recurrence (monitoring markers). Despite recent advances in the use of genomic analysis for risk assessment, in the area of biomarker identification genomic methods have yet to produce reliable candidate markers for CRC. For this reason, attention is now being directed towards protein chemistry or proteomics as an analytical tool for biomarker identification. Here we will review the technologies of proteomics with reference to how they may contribute to this vital search for ways to combat CRC.

Biomedical Application of Proteomics
Edited by J.-C. Sanchez, G. L. Corthals, D. F. Hochstrasser

ISBN: 3-527-30807-5

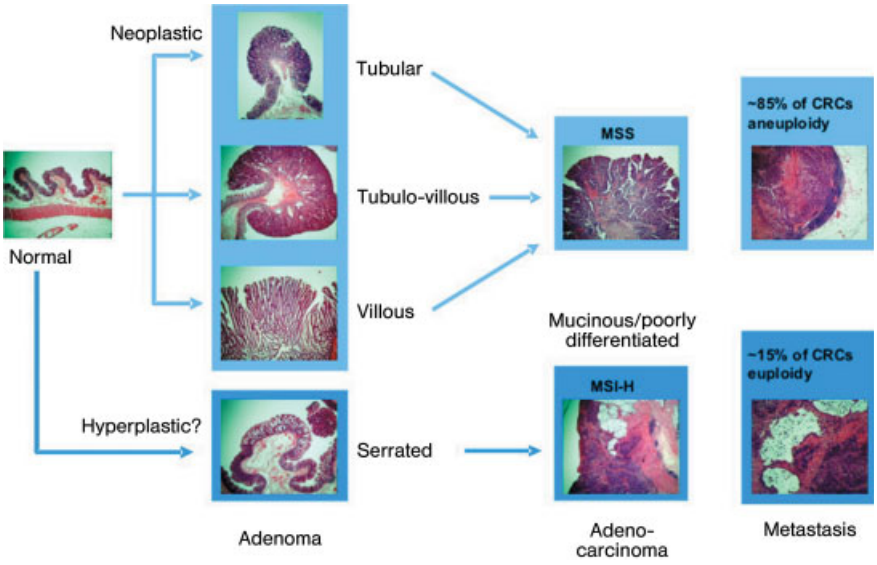


Figure 7.1 There are two major pathways for CRC development and progression. The 'traditional' pathway progresses through neoplastic adenomas (which can be tubular, villous or both) which progress to well- or moderately differentiated adenocarcinomas. Tumors arising from this pathway are usually aneuploid and show chromosomal instability (CIN). Recently, a second pathway has been recognized, which accounts for 15% of colorectal cancers. It probably develops from so-called serrated adenomas, which are thought to progress to mucinous/poorly differentiated adenocarcinomas. These tumors are typically diploid and exhibit microsatellite instability (MIN). The familial counterparts of these two pathways are represented by familial adenomatous polyposis (FAP) and hereditary non-polyposis colon cancer (HNPCC), respectively. Pathology slides kindly provided by Dr. Paul Waring of the Peter MacCallum Institute of Cancer Research.

7.2
Molecular Basis of Colorectal Cancer: Tumorigenesis is a Multistep Process

Cancer can be viewed as the endpoint of a series of genetic and epigenetic alterations in a cell or group of cells [7, 8]. Like many cancers, CRC is a heterogeneous disease resulting from mutations to genes such as *p53*, *Ras* or *Apc*, but which can also contain a variety of other mutations to genes such as *MLH1* [9–11] or *B-raf* [12] capable of modulating tumor phenotype and outcome [6, 13] (Fig. 7.2). Two forms of genetic instability, chromosomal instability (CIN), and microsatellite instability (MIN), have been described in CRC [14].

CIN tumors have a wide variation in chromosome number (i.e. their karyotypes are aneuploid with whole or large pieces of chromosomes continuously gained or lost) [15]. Although thought to involve a single dominant event (e.g. mutation in a gene important or chromosomal segregation [16]), the molecular events underpinning CIN are not fully understood.

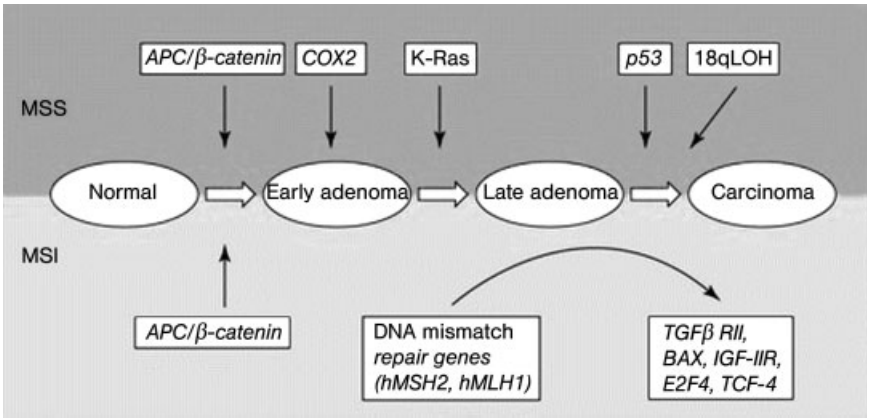


Figure 7.2 Colorectal cancer (CRC) is a multistep process that typically develops over decades and appears to require multiple genetic events for completion. The stepwise accumulation of mutations affect growth control, differentiation, and cell survival [81]. In this figure, the crucial genes involved in the development of colorectal polyps and cancer are shown in boxes [82]. CRC is usually initiated by mutations in the adenomatous polyposis coli (*Apc*) tumor suppressor gene, which leads to activation of *β*-catenin and downstream targets, such as c-myc and cyclin D1. The specific timing of each genetic event is crucial in tumor pathogenesis. CRCs can be grouped into two broad categories, as illustrated. The first category (~85% of tumors), shown in blue, exhibit CIN but stability of microsatellite DNA – these are referred to as microsatellite stable (MMS) tumors (also referred to as CIN tumors). After mutations of genes early in the *Apc/β*-catenin pathway, further mutations often occur in the oncogene *K-ras* (in ~50% of CRCs) and the tumor suppressor gene, *p53*. These latter genetic lesions drive tumor progression towards the final stages of cancer. The second category (shown in gray) represents ~15% of colorectal tumors and exhibits instability of microsatellite DNA – referred to as microsatellite instable (MSI) tumors (or MIN tumors). This population of CRCs is associated with mutations in DNA mismatch repair genes, which lead to an accumulation of widespread mutations in genes, such as the genes encoding type II TGF*β* receptor, the apoptosis regulator BAX, the insulin-like growth factor II receptor, the cell cycle-regulated transcription factors EF2–4, and the homeobox factor CDX2, to name a few (adapted from [82] with permission).

In contrast, MIN was first described in sporadic CRC [17, 18] and is considered to be one of the best understood forms of genetic instability, arising from inactivation of DNA mismatch repair (MMR) genes such as *MSH2* or *MLH1* [19, 20]. MIN tumors have nucleotide mutation rates two to three orders of magnitude higher than normal cells or MMR-proficient cancers of the same cell type [21–23]. Because MIN and CIN instabilities rarely coexist in colorectal tumors, it has been postulated that a single form of instability is sufficient to drive tumorigenesis [14] (Fig. 7.2). The usual inactivating mechanism for the DNA MMR gene *hMLH1* found in MIN cancers involves hypermethylation of the promoter region of the gene [24–26]. Aberrant cytosine methylation of gene promoter-region CpG islands and associated alterations in histone acetylation appear to be primary mediators of epigenetic inheritance in cancer cells [27, 28]; CRCs demonstrating hypermethylation and functional silencing of

multiple genes are described as CIMP (cytosine phosphoguanosine (CpG) island methylated phenotype) positive [29]. Genes other than *hMHL1* that may be methylated in human tumors include *ER, p16, p14, HPP1/TPEF, MGMT, THBS1, Apc, COX-2, CDH1, RIZ1,* and *RASSF1A*. Interestingly, loss of imprinting, an epigenetic alteration associated with DNA hypomethylation, affecting the insulin-like growth factor II gene (*IGF2*) has been recently found in normal colonic mucosa of about 30% of CRC patients, but only 10% of healthy individuals [30].

7.3 The Case for Early Detection: CRC is Treatable if Detected Early

Early detection is the single most important factor influencing outcome for CRC patients. If CRC can be identified while still localized it is more amenable to treatment, thereby preventing not only mortality, but also reducing morbidity and overall associated health costs [31]. For example, the Surveillance, Epidemiology, and End Results (SEER) study for CRC cases diagnosed between 1990 and 1999 inclusive indicates that patient survival is excellent when early-stage disease is treated with existing therapies (Fig. 7.3). According to the SEER study, projected 5-year survival rate if all tumors were localized when detected is ~90% (National Cancer Institute Surveillance Epidemiology and End Results Program http://seer.cancer. gov/2002).

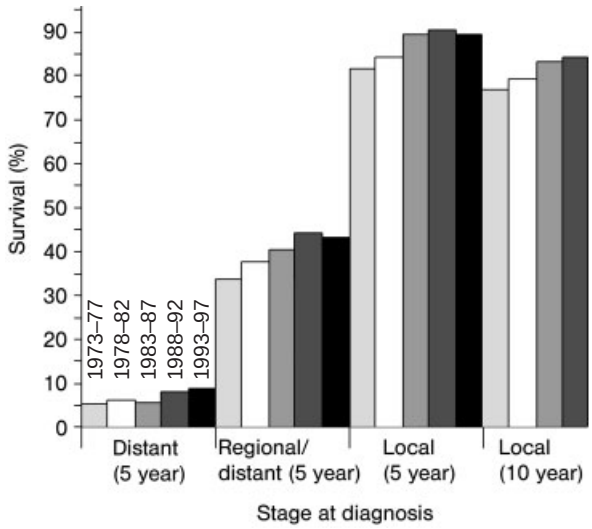


Figure 7.3 Colorectal cancer is treatable if detected early. Relative survival (5-year or 10-year) among colorectal cancer cases diagnosed with distant, regional, or distant, and localized disease by year of diagnosis. Source: Surveillance, Epidemiology, and End Results (SEER) http://seer.cancer.gov/(2002) (reproduced from [31] with permission).

Current community-based methods (e.g. annual fecal occult blood test (FOBT) and 5-yearly sigmoidoscopy) for detecting CRC are inadequate. While the FOBT is sensitive for fecal occult blood it is not specific for CRC (only 40% of patients with positive tests will have neoplasia). However, prospective controlled trials of FOBT in Minnesota, the UK, and Denmark revealed an overall 15–33% reduction of CRC mortality with annual FOBT [32]. Although the FOBT in conjunction with flexible sigmoidoscopy has the potential to reduce CRC mortality, these approaches have not been found to be acceptable for large-scale population screening trials and very few national governments have been persuaded to support them financially. Currently available molecular markers of CRC such as serum carcinoembryonic antigen (CEA) [33] and CA-19.9 [34, 35] are useful to monitor disease progression in patients after primary therapy or to detect disease recurrence at an early stage [36]. CEA in particular has proved useful for detection of liver metastasis from CRCs [37], however, only a small proportion of CRCs express elevated CEA and CA-19.9 levels at the time of diagnosis. In the context of screening for CRC, more reliable serum markers that meet the criteria outlined by the Tumor Marker Utility Grading System (TMUGS) [38] are needed to improve diagnosis and to follow cancer progression.

Table 7.1 Putative biomarkers for detection of colorectal cancer.

Gene[a]	*Chromosomal location*	*Specimen*[b]
Apc (~70%)	5q21	Stool DNA
p53 (~50–70%)	17p13	Stool DNA, tissue
K-ras (~50%)	12p	Stool DNA, colonic effluent, colonic DNA
CTNNB (~4–15%)	3p22	Tis sue
Src (2%)	20p22	Tissue
SMAD4 (~16%)	18q21	Tissue
SMAD2 (~6%)	18q21	Tissue
DCC (~3%)	18q21	Tissue
hMSH2	2p21	Tissue
hMLH1	3p21	Tissue
hMSH6	2p21	Tissue
hPMS1	2q31–33	Tissue
hPMS2	7p22	Tissue
MSI		Tissue, stool DNA, plasma DNA
Ploidy		Tissue
IGF-I		Serum
CEA		Serum

a Numbers in parentheses represent prevalence of genetic mutations reported by [144].
b Specimens listed are those for which the assay for mutations can be performed (reproduced from [145]).

In addition to proteins, gene mutations and gene instability can also serve as specific biomarkers for CRC [39, 40]. For example, analysis of free DNA in serum of cancer patients [41] has led to today's DNA-based blood tests for oncogene mutations (Apc, p53, Ras, B-raf, etc.), microsatellite instability and hypermethylation of promoter regions (e.g. MLH1, CDKN2A and, more recently, loss of imprinting (LOI) in the *IGF2* gene [42]) for predicting the risk of developing CRC [43]. However, since non-tumor cells also shed DNA into the serum, some cancer-specific changes can be difficult to detect above wild-type background DNA. A list of putative biomarkers (by no means complete) for detection of CRC is given in Table 7.1.

7.4 Approaches to Biomarker Discovery

Two philosophically different but parallel approaches can be utilized for biomarker discovery research. On the one hand there are hypothesis-driven approaches; logical step-by-step approaches that build upon what we have already learnt about cancer biology. A recent case in point is loss of imprinting of the *IGF2* gene, thought to be involved in the biology of CRC [42]. Recently, more open-ended 'discovery-based research' [44] has gained wide acceptance because of the availability of high-throughput techniques that allow simultaneous assessment of tens of thousands of genes or proteins in relatively short periods of time.

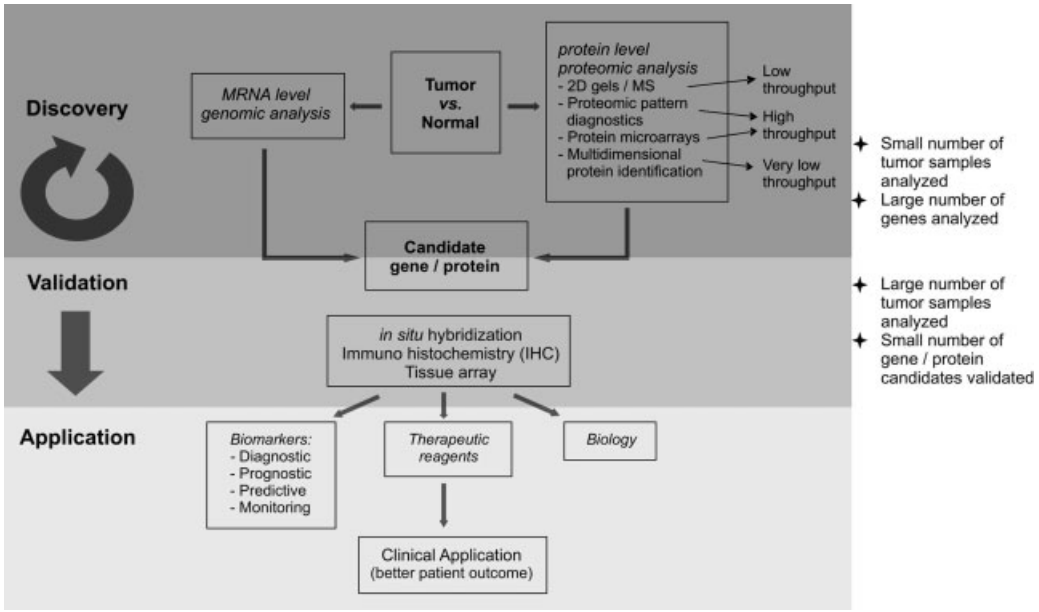


Figure 7.4 Strategies for new biomarker discovery.

Table 7.2 Comparison of proteomics technologies and their contribution to biomarker discovery.

ELISA	***2-D PAGE***	***Multidimensional protein identification technologies***	***Proteomic pattern diagnostics***	***Protein microarrays***
Sensitivity				
Highest	Overall low, particularly for less-abundant proteins; sensitivity limited by detection method; LCM can improve specificity via enrichment of selected cell populations	High	Medium sensitivity with diminishing yield at higher molecular weights; will improve with new MS instrumentation	Medium/high
Direct identification of markers				
N/A	Yes	Yes	Yes for MALDI-TOF MS analysis of 1-mm tissue sections. Newer MS technologies might make this possible for SELDI-MS approaches	Possible when coupled with MS technologies
Use				
Detection of single, specific well-characterized analyte in body fluid or tissue; gold standard of clinical assays	Means for discovery and identification of biomarkers, not a direct means of early detection in itself	Detection and identification of potential biomarkers	Diagnostic pattern analysis in body fluids and tissues; potential biomarker identification	Multiparametric analysis of many analytes simultaneously

Table 7.2 (cont.)

ELISA	*2-D PAGE*	*Multidimensional protein identification technologies*	*Proteomic pattern diagnostics*	*Protein microarrays*
Throughput				
Moderate	Low	Very low	Highest	High
Advantages/drawbacks				
Very robust; well-established use in clinical assays; requires well-characterized antibody for detection and extensive validation; not amenable to direct discovery (strictly measurement based)	All IDs require validation and testing before clinical use; tried and true methodology, reproducible and more quantitative combined with fluorescent dyes	Significantly higher sensitivity than 2-D PAGE (much larger coverage of the proteome for biomarker discovery)	Protein IDs not necessary for diagnostic pattern analysis; reproducibility issues need to be addressed; need for validation; coupling to adaptive informatics tools might revolutionize the field of clinical chemistry	Format is flexible: can be used to assay for multiple analytes in a single specimen or a single analyte in a large number of specimens; requires prior knowledge of analyte being measured; limited by antibody sensitivity and specificity; requires use of an amplified tag detection system

LCM, laser-capture microdissection; MS, mass spectrometry; MALDI-TOF-MS (matrix-assisted laser desorption/time-of-flight mass spectrometry [146]; SELDI-MS, surface-enhanced laser desorption ionization/time-of-flight mass spectrometry (adapted from [147] with permission).

Using this approach there is no a priori reason to understand the biological processes or identify individual targets, as large portions of the genome or proteome may be screened simultaneously. For example, RNA expression microarrays can be used to predict a prognosis of cancer [45–47] – or the entire proteome of serum may be scanned by mass spectrometry using pattern recognition to discriminate among persons with and without a specific type of tumor [48]. It is expected that new candidate biomarkers discovered by these high-throughput methods can then be assessed individually in a large number of tumor types or tissues with conventional techniques such as ELISA or immunohistochemistry (Fig. 7.4). A comparison of various proteomic technologies and their contribution to biomarker discovery is given in Table 7.2.

There are a wide range of proteomic approaches that can be applied to the early detection of colorectal cancer (see Wulfkuhle et al. for a review [49]). These include ELISA, two-dimensional gel electrophoresis (2-DE), multidimensional protein identification technologies, proteomic pattern diagnostics, and protein arrays (Fig. 7.4). A brief outline of the most important proteomic techniques currently in use is given below.

7.4.1 Use of ELISAs to Detect CRC

The risk of recurrence and subsequent death of CRC is closely related to the stage of disease at the time of initial diagnosis. There is a clear correlation between improved survival rate of CRC and early diagnosis when the cancer is still localized (Fig. 7.3). CEA, one of the most extensively studied serological tumor markers, is commonly analyzed in serum by ELISA, but has low sensitivity for either early-stage CRC or local recurrence [50, 51]. Nevertheless, serum CEA is still the most widely used tumor marker in the diagnosis and follow-up of CRC. Although other potential serum markers such as CA-242, CA-19.9, and CA-50 have also been studied, none of these has the high sensitivity and high specificity required for routine use for detection of early-stage CRC. Recently, the levels of tissue inhibitor of metalloproteinases (TIMP) have been measured in plasma and shown to be significantly elevated in CRC, including early-stage disease [52]. Further, a combination of CEA and TIMP-1 measurements increased the sensitivity obtained from TIMP-1 alone (Fig. 7.5).

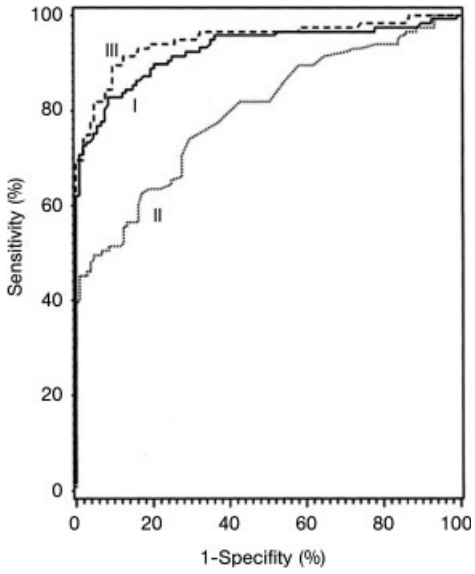


Figure 7.5 Two biomarkers are better than one. Combining CEA and TIMP-1 measurements increased the sensitivities obtained from TIMP-1 alone. ROC curves for total plasma TIMP-1 (I), serum CEA (II), and TIMP-1/CEA combined (III) in right-sided CRC patients. AUC for each ROC curve: I, 0.93 (SE=0.02); II, 0.80 (SE=0.03); and III, 0.95 (SE=0.02) (reproduced from [45, 52] with permission).

In recent years, the role of cytokines in cancer immunity has been revealed and Kaminska and colleagues report that levels of C-reactive protein (CRP) and the proinflammatory cytokines tumor necrosis factor α (TNFα), interleukin 6 (IL-6), and IL-8 are elevated in CRC patients prior to tumor removal, regardless of the stage of the disease [53]. In addition, McMillan et al. report that in advanced cancer patients the presence of a systemic inflammatory response (indicated by elevated serum levels of CRP) and the magnitude of that response predicts the duration of cancer-specific and non-cancer survival [54]. In another study, De Vita and colleagues report that IL-6 and IL-10 serum levels correlate with overall survival in patients with advanced gastrointestinal cancer but are not independent prognostic indicators [55, 56].

7.4.2 Two-dimensional Gel Electrophoresis

Protein expression profiling of whole-cell lysates or enriched subcellular fractions, involving the combination of 2-DE with mass spectrometry (MS) identification, has been the primary technique for biomarker discovery in conventional proteomic analysis [57]. This approach is ideally suited for the direct comparison of protein expression profiles because 2-DE fractionates proteins by charge in the first dimension and molecular weight in the second dimension, permitting the resolution of several thousand proteins in the one experiment (Fig. 7.6). (For a recent overview of 2-DE technology, see Gorg and colleagues [58].) This technique has been used to identify proteins that are differentially expressed between normal and tumor tissues in colorectal cancer, as well as derived CRC call lines (Table 7.3). However, despite its utility for protein separation, 2-DE has several limitations for identification of tumor markers. First, it requires a large amount of protein as starting material (approximately, 10^8 cells are required to analyze proteins of medium abundance using modern MS identification methods. (For a discussion of the challenges that dynamic range of protein expression presents in 2-DE see Corthals et al. [59].) Nevertheless, advances in 2-DE methodology, such as the introduction of immobilized pH gradient (IPG) sample application strips for separations up to pH 12, narrow range IPGs to amplify regions of 1–1.5 pH units, and increased loading capacity [60, 61], have kept 2-DE at the forefront of proteomics separation technology. Significant advances in fluorescent [62–64] and non-fluorescent [65, 66] techniques, including differential in-gel electrophoresis (DIGE), have improved reproducibility, sensitivity, and quantitative aspects of 2-DE [67, 68].

Using the DIGE approach [67], cellular protein mixtures are differentially labeled with fluorescent dyes (Cy2, Cy3, and Cy5), mixed and electrophoresed on a single 2-D gel. The gel is then scanned to generate a map for each labeled protein pool. The images can then be compared for differences in fluorescent intensities between labels for any given protein spot. To overcome the problem of dynamic range there has been much effort directed towards simplifying protein mixtures – by confining analysis to specific proteins or cellular compartments – the 'subpro-

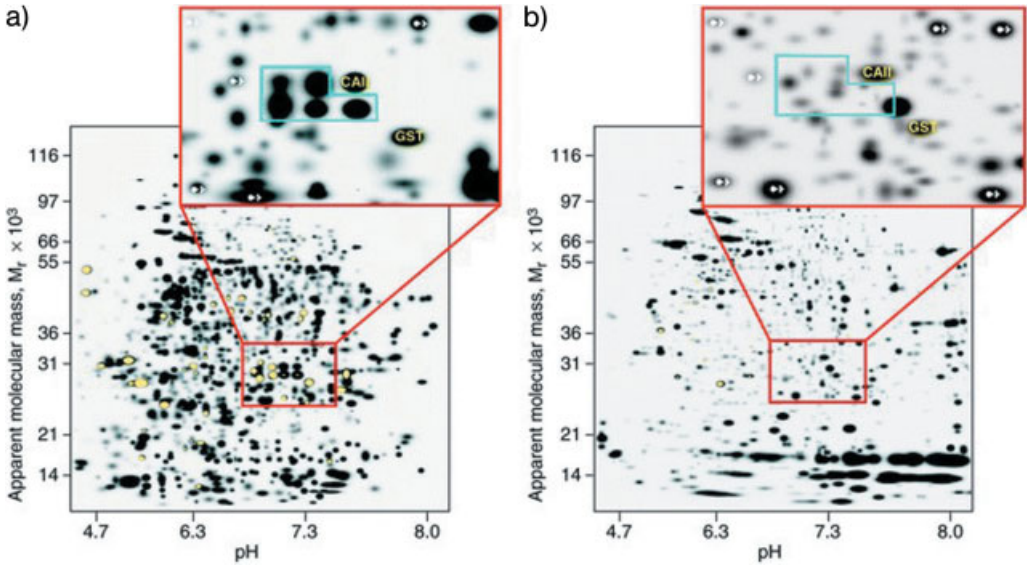


Figure 7.6 2-DE gel of proteins from a whole-cell extract stained with Coomassie brilliant blue. (a) Wild-type C57Black/6J murine colonic crypts versus (b) polyps from multiple intestinal neoplasia (MIN) mice. The synthetic gel images were generated using PDQuest software. Differentially expressed proteins are marked in yellow. The inserts show that carbonic anhydrase (CAII) and glutathione-*S*-transferase (GST) are both highly expressed in wild-type crypts and MIN polyps, whereas expression of several CAII isoforms (outlined box) is dramatically reduced in MIN polyps (adapted from [24] with permission).

teome'. For example, coupling of 2-DE with methods for enrichment of specific groups of cellular proteins by fractionation of subcellular organelles can be useful to reduce complexity and increase identification capability [69].

2-DE can also be used in conjunction with immunological detection techniques to identify novel antigens for which a suitable antibody is available. However, conventional 2-DE, which ordinarily requires reducing conditions and the presence of chaotropes, may be unsuitable for use with monoclonal antibodies (mAb) that recognize conformational epitopes. In the case of the A33 mAb, which selectively localizes to metastatic lesions in the colon [70], immunoblotting of non-reducing 2-DE gels was necessary to visualize the A33 antigen [68, 71], leading to its purification [72] and subsequent biochemical characterization [73, 74].

A critical consideration in any proteomic profiling study (and, indeed, mRNA profiling) is rigorous sample preparation. Parameters such as time interval between surgical removal and sample processing, type of fixative and embedding medium, length of fixation time, and temperature of tissue processing all impact significantly on the integrity of proteins in the final sample. The development of laser-capture microdissection (LCM) technology [75] has greatly improved the specificity of 2-DE for biomarker discovery, because it enables the procurement of homogeneous tumor tissue, thereby enriching for the proteome of interest and

Table 7.3 2-D gel electrophoresis approaches for colorectal cancer biomarker discovery and multidrug resistance.

Specimen and/or cell line	*Comments*	*References*
Human colonic crypts, carcinoma cell lines LIM1215, LIM1863 and LIM1899	Protein expression profiles of normal colonic crypts isolated from different regions of the large intestine were compared using 2-DE with several colorectal cell lines. Although protein profiles for crypts from different regions of the colon were 95% identical, only 75–85% of the proteins expressed by the cell lines could be matched with those from crypts. This study raises the issue in cancer research as to whether primary tumors can be accurately represented by tumor-derived cell lines. A similar observation was noted by Ornstein et al. when comparing prostate cancer-derived cell lines with prostate epithelium [148]	[71, 149, 150][a]
Murine colonic crypts, polyps	2-DE protein profiles from crypts from normal, multiple intestinal neoplasia (MIN) or p53 knockout mice reveal no significant differences. However, when compared with MIN polyps, ~64 differentially expressed proteins (including calreticulin, carbonic anhydrase I, and a member of the glutathione S-transferase theta family) were observed (see Figure 7.6)	[151][b]
Colorectal carcinoma, fresh tumors	Stulik and co-workers report down-regulation of liver fatty acid-binding protein (L-FABP), smooth muscle protein 22-alpha, and cyclooxygenase 2 in colon cancer tissue when compared with normal tissue; novel variant of heat shock protein 70 (hsp70) and several members of the S-100 protein family of calcium-binding proteins. In 23 matched sets of colon carcinoma and normal colon mucosa, IHC revealed S100A8 and S100A9 expression levels were increased in macrophages and polymorphonuclear leukocytes along the invasive margin of colorectal carcinoma, suggesting that these proteins might be involved in cancer regression. Immunohistochemistry confirmed the disease association of calgranulin B	[152–154]
Human colonic polyps	Numatrin (nucleophosphine/B23), hsp70 and hsp60 protein expression levels are increased while those of L-FABP, 14-3-3 σ, cytokeratin 20, cytochrome c oxidase polypeptide Va, β- and γ-actins, and Rho GDP-dissociation inhibitor (Rho GDI) were decreased in colonic polyps when compared with normal colon specimens	[155]
Colon carcinoma cell line HT 29	Using 2-DE, adenine phosphoribosyl transferase and breast cancer-specific gene 1 (BCSG1) were found to be overexpressed in the multidrug resistance (mitoxantrone) cell line variant of HT-27	[156]

Table 7.3 (cont.)

Specimen and/or cell line	*Comments*	*References*
Murine tissue (mutated Apc gene)	Minowa and co-workers show that several proteins are overexpressed in the mutant mice polyps compared with normal tissue (e.g. truncated β-tubulin)	[157]
Human nuclear matrix proteins in colon tumors	Using 2-DE, nuclear matrix protein patterns found in human colon tumors were compared with those from normal colon epithelia. At least six nuclear matrix proteins that were tumor specific were characterized. Similar findings have been reported by Szymczyk et al.	[158, 159]
Murine azoxymethane induced colon tumors (fresh sections)	Using MALDI-TOF-MS, a comparison of the protein expression profiles of normal and cancerous mouse colon tissue revealed three potential tumor-specific protein markers – calgranulin B (S-100A9), calgranulin A (S-100A8) and calgizzarin (S-100A11)	[160]
Human carcinoma cell line LoVo	Membrane proteins of viable, intact cells were biotinylated then affinity-captured and purified on an avidin column. Biotinylated proteins were separated by 2-DE and identified by MS. Amongst others, several proteins with chaperone function were identified	[82]

a http://www.ludwig.edu.au/jpsl/jpslhome.html
b http://www.ludwig.edu.au/jpsl/jpslhome.html

reducing contamination from extraneous tissue [68, 76]. However, due to the labor-intensive nature of LCM, the expense of the equipment involved and the small quantities of sample obtained, this technique has yet to gain wide acceptance.

7.4.3
One-dimensional Sodium Dodecyl Sulfate-Polyacrylamide Gel Electrophoresis (SDS-PAGE) Plasma Membrane Proteome

The proteomic definition of plasma membrane proteins is a critical step in searching for biomarkers of CRC. Because of charge heterogeneity and poor solubility, membrane-associated proteins are often refractory to conventional 2-DE. (Although great strides have now been made in methods for solubilization of membrane proteins [45, 52, 77–80] this class of molecules still presents a technical challenge for 2-DE.) Recently, we described a non-2-DE method for identifying membrane proteins from the CRC cell line LIM1215 [81]. As outlined schematically in Figure 7.7, proteins from an enriched membrane preparation were initially separated by SDS-PAGE (4–20%) and the unstained gel was cut into slices and each slice digested with trypsin. Generated sets of tryptic peptides were subjected to capillary column reverse phase high-performance liquid chromatography (RP-HPLC) coupled to elec-

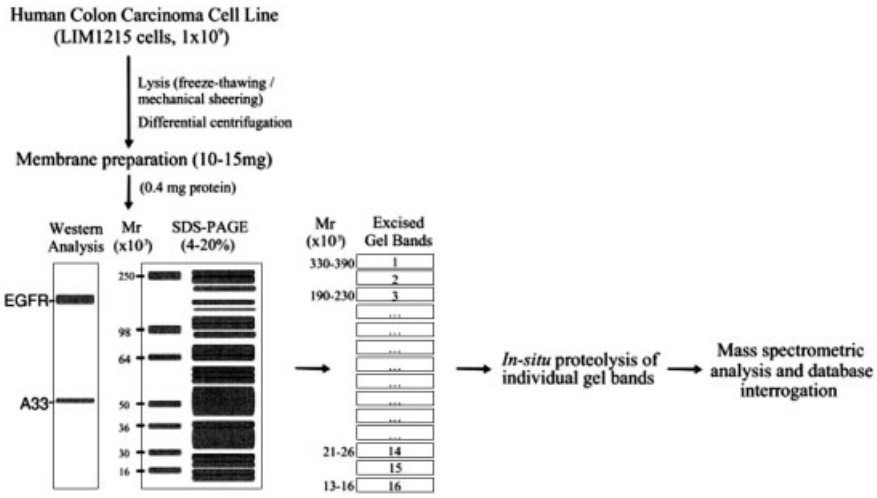


Figure 7.7 Schematic representation of the strategy used for the proteomic analysis of an enriched membrane preparation from colorectal LIM1215 cells. The one-dimensional SDS-PAGE gel was cut into 16×3-mm slices. Each slice was digested *in situ* with trypsin and MS analysis/database interrogation was then performed on each of the 16 peptide extracts. Using this approach, over 300 membrane-associated proteins were identified, many of which had not been previously identified on 2-DE gels (reproduced from [81] with permission).

trospray ionization ion-trap MS. Using this approach, ~300 proteins (including 92 membrane proteins) were identified. Among the proteins identified were many integral membrane proteins not previously identified by 2-DE, many proteins seen at the genomic level only, as well as several proteins previously known solely from expressed sequence tags (ESTs). Moreover, several peptides that had been identified by *de novo* MS sequence analysis, were completely novel and not present in any publicly available database.

Recently, Hanash and colleagues applied a global surface protein biotinylation strategy, coupled with the use of 2-DE and MS, to the colorectal cell line LoVo; this study uncovered an abundance of proteins with chaperone function [82].

7.4.4 Multidimensional Protein Identification Technologies

Because of the inherent limitations of 2-DE [57, 68, 71, 83, 84] much effort has been directed towards the development of separation techniques to enable a more comprehensive description of tumor cell proteomes. In this pursuit, a number of multiplexed technologies have been developed. These include 2-D liquid chromatography of complex mixtures of proteins with a combination of non-porous RP-HPLC, ion-exchange chromatography (IEX) (pioneered by Isobe and colleagues [85, 86]), and liquid-phase isoelectric focusing (IEF) [87]. (For a review of these techniques, see [57] and references therein.)

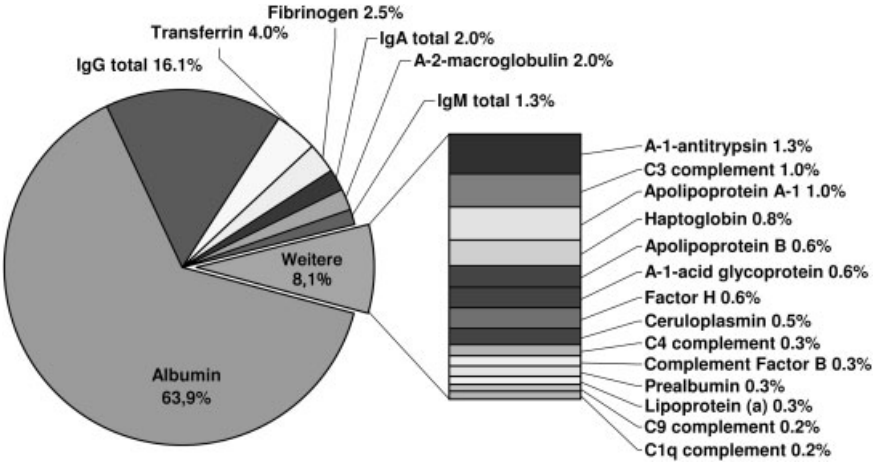


Figure 7.8 Plasma protein abundances, shown as percentages of total plasma proteins, calculated from the information given in Figure 7.3 (reproduced from [161] with permission).

Another approach involves direct proteolysis (typically, with trypsin) of complex protein mixtures and resolution of generated peptides by a combination of RP-HPLC/IEX prior to online MS identification [88–90].

Several other recently described IEF techniques are now being applied to proteome analysis of cancer. These include the liquid-based IEF procedure of free flow electrophoresis (FFE) [91], as well as multicompartment electrolyzers that utilize IEF membranes of varying pH ranges [92–95] and a novel electrophoresis device that separates proteins based upon both pI values and molecular weight [96]. Unlike the gel-based IPG strips used in 2-DE methods, the liquid-based IEF techniques are not restricted by sample load and are thus more amenable to the study of low-abundance proteins where larger amounts of starting material are available. While these multiplexed methods may someday replace traditional 2-DE, they do have major drawbacks, most notably a requirement for relatively large amounts of protein that preclude their use with specimens such as clinical biopsies. However, these non-2-DE multidimensional separation technologies are gaining much attention in the quest for discovering novel serum-based biomarkers. So, the main caveat in this case is that given the dynamic range of protein concentrations in plasma/serum (Fig. 7.8), these technologies will need to be coupled to methods for depleting major proteins in blood prior to analysis [97–99].

7.4.5
Proteomic Pattern Diagnostics

The application of high-throughput MS for pattern recognition in cancer is a further area of recent interest. This technology, based upon surface-enhanced laser desorption ionization/time-of-flight (SELDI-TOF) mass spectrometry technology, has the potential for the rapid identification of cancer-specific biomarkers and

proteomic patterns, especially in body fluids such as blood. Recently, pattern recognition algorithms coupled to high-throughput MS have been developed and applied to the problem of ovarian cancer diagnostics [100–103]. This new approach to biomarker discovery is very much 'black box' at this stage. However, with improved MS identification capability, pattern recognition has much potential given that the 'patterns' themselves may be used as the test – in much the same way as 'gene expression signatures' may be used without the need to understand precisely which molecules comprise the 'pattern' [45]. Nevertheless, this approach must be validated rigorously in order to avoid problems of data overfitting and bias [104, 105].

Recently, Caprioli and colleagues developed another high-throughput proteomics approach based upon MALDI-TOF-MS that allows direct mapping of protein expression in 1-mm regions of single frozen tissue samples from surgically resected tumor tissues [45, 106–108]. At present, this technology permits the profiling of proteins up to 50 kDa in size. A frozen tissue section is placed directly in the MALDI-TOF-MS instrument and ions (m/z values) from several regions of the tissue section obtained. A 'rasterized image' of peak intensities for any given m/z value can be overlayed on the biopsy section to generate a 3-D expression map that is indicative of the protein expression profile [109] over the section. These protein profiles obtained can contain several thousand data points, necessitating the development of sophisticated algorithms to extract peak ion clusters enabling classification of proteins related to cancer histology [110]. Using this approach to study non-small-cell lung cancer, Yanagisawa and colleagues report a proteomic expression pattern comprising 15 distinct MS peaks that allowed the researchers to distinguish tumors with good prognosis from those with poor prognosis [110].

7.4.6 Protein Microarrays

In recent years, DNA microarrays have become a standard tool for the molecular analysis of cancer, providing global profiles of transcription that may reflect development, progression, or invasive properties of cancer [111, 112]. The ability to complement these molecular diagnostic approaches with methods that analyze the proteome has the potential to revolutionize the analysis of cancer. Protein microarray technologies have been explored for high-throughput screening of changes in protein expression, including both antigen and antibody arrays [113–124]. Many of these protein microarray methods are capable of screening for enzymatic activity [125] including protein kinase activity [126] or serine hydrolase activity [127]. A major limitation of protein microarrays is that in cellular systems proteins undergo a variety of post-translational modifications, some of which may be important in the transformation of a normal protein into an oncoprotein, which in turn may lead to cancer development (e.g. phosphorylation and acetylation of p53 [128]). To overcome this problem, recent biomarker discovery efforts have utilized native proteins, for example, tissue microarrays [129–131], and the separation of proteins extracted from cells using either 2-DE or multidimensional IEF/chromatography [132].

A remaining challenge for the full implementation of protein microarrays is the acquisition of large sets of high-affinity and highly specific protein capture reagents.

7.4.7
Proteomic Approaches for Identifying Tumor-specific Autoantigens

There is increasing evidence for a humoral immune response to cancer-related molecules in humans. This is demonstrated, in part, by identification of tumor-specific autoantigens [133–135]. Hence, tumor antigens that elicit a humoral response may have utility as biomarkers for the detection, progression, and invasive nature of cancers, as well as immunotherapy against the disease. Several approaches have been described for the identification of tumor-specific autoantigens, most notably the screening of cDNA expression libraries derived from tumors with autologous serum (SEREX) [136]. Recently, several approaches for identifying tumor antigens that combine serology and proteomics have been developed. In one approach, tumor-associated antigens (TAAs) were identified by comparing the reactivity of proteins resolved by 2-DE with sera from disease patients versus that of a healthy donor ('serological proteome analysis', SERPA [137] or PROTEOMEX (combination of proteome analysis and SEREX) [138, 139]). This approach employs SDS-PAGE to separate individual cellular proteins from both tumor and adjacent normal tissue. (A variation on this approach utilizes cultured cancer cells as a source of antigens [140].) Separated proteins are transferred onto membranes, which are subsequently incubated with sera from normal donors or disease patients. Detection of tumor-specific antigens utilized a second antibody either directed against human IgM or IgG followed by autoradiography. Identification of the proteins reacting specifically with sera from cancer patients was then accomplished by MS analysis of the protein spot from a matching 2-DE gel that had been stained with Coomassie blue or silver [57]. (For a review of this technique see Seliger and Kellner [141].)

Instead of 2-DE, two-dimensional liquid chromatography can be employed where intact proteins from cell lysates are separated using chromatofocusing in the first dimension, and non-porous RP-HPLC in the second dimension [142]. The fractionated proteins can then be spotted onto a microarray slide (nitrocellulose) and assayed for TAAs by exposing the slide to sera from cancer patients or normal donors. Another approach utilizes phage display to select peptides recognized by autologous antibodies purified from the serum of cancer patients [143].

7.5
Conclusions

It has now become clear that processes of tumor formation and growth are accompanied by many physiological changes that, although individually small and dispersed, will produce molecular signatures that contain clues to the type and lo-

cation of the cryptic tumor lesion. A wide range of proteomic and genomic technologies are being applied to defining the signatures and these have begun to produce huge amounts of data within which the threads of the molecular clues can be picked up. The major challenge for the future will be the task of analyzing these data to bring together the molecular threads and find ways to use them to improve outcomes for cancers such as CRC.

7.6 References

1 Parkin, D.M. *Lancet Oncol.* **2001**, *2*, 533–543.
2 Parkin, D.M., Bray, F., Ferlay, J., Pisani, P. *Int. J. Cancer* **2001**, *94*, 153–156.
3 Parkin, D.M., Bray, F.I., Devesa, S.S. *Eur. J. Cancer* **2001**, *37 Suppl. 8*, S4–66.
4 Fearnhead, N.S., Wilding, J.L., Bodmer, W.F. *Br. Med. Bull.* **2002**, *64*, 27–43.
5 Fearnhead, N.S., Britton, M.P., Bodmer, W.F. *Hum. Mol. Genet.* **2001**, *10*, 721–733.
6 Fishel, R., Kolodner, R.D. *Curr. Opin. Genet. Dev.* **1995**, *5*, 382–395.
7 Breivik, J., Gaudernack, G. *Adv. Cancer Res.* **1999**, *76*, 187.
8 Tomlinson, I., Bodmer, W. *Nature Med.* **1999**, *5*, 11–12.
9 Jass, J.R. *Ann. NY Acad. Sci.* **2000**, *910*, 62–73.
10 Jass, J.R. *Lancet Oncol.* **2000**, *1*, 220–226.
11 Mori, Y., Selaru, F.M., Sato, F. et al. *Cancer Res.* **2003**, *63*, 4577–4582.
12 Davies, H., Bignell, G.R., Cox, C. et al. *Nature* **2002**, *417*, 949–954.
13 Fearnhead, N.S., Britton, M.P., Bodmer, W.F. *Hum. Mol. Genet.* **2001**, *10*, 721–733.
14 Lengauer, C., Kinzler, K.W., Vogelstein, B. *Nature* **1997**, *386*, 623–627.
15 Aaltonen, L.A., Peltomaki, P., Leach, F.S. et al. *Science* **1993**, *260*, 812–816.
16 Cahill, D.P., Lengauer, C., Yu, J., Riggins, G.J. et al. *Nature* **1998**, *392*, 300–303.
17 Ionov, Y., Peinado, M.A., Malkhosyan, S., Shibata, D., Perucho, M. *Nature* **1993**, *363*, 558–561.
18 Thibodeau, S.N., Bren, G., Schaid, D. *Science* **1993**, *260*, 816–819.
19 Peltomaki, P., de la Chapelle, A. *Adv. Cancer Res.* **1997**, *71*, 93–119.
20 Thibodeau, S.N., French, A.J., Cunningham, J.M. *et al. Cancer Res.* **1998**, *58*, 1713–1718.
21 Parsons, R., Li, G.M., Longley, M.J. et al. *Cell* **1993**, *75*, 1227–1236.
22 Bhattacharyya, N.P., Skandalis, A., Ganesh, A., Groden, J., Meuth, M. *Proc. Natl Acad. Sci. USA* **1994**, *91*, 6319–6323.
23 Eshleman, J.R., Lang, E.Z., Bowerfind, G.K. et al. *Oncogene* **1995**, *10*, 33–37.
24 Kane, M.F., Loda, M., Gaida, G.M. et al. *Cancer Res.* **1997**, *57*, 808–811.
25 Herman, J.G., Umar, A., Polyak, K. et al. *Proc. Natl Acad. Sci. USA* **1998**, *95*, 6870–6875.
26 Cunningham, J.M., Christensen, E.R., Tester, D.J. et al. *Cancer Res.* **1998**, *58*, 3455–3460.
27 Issa, J.P. *J. Nutr.* **2002**, *132*, 2388S–2392S.
28 Jubb, A.M., Bell, S.M., Quirke, P. *J. Pathol.* **2001**, *195*, 111–134.
29 Ahuja, N., Mohan, A.L., Li, Q. et al. *Cancer Res.* **1997**, *57*, 3370–3374.
30 Cui, H., Cruz-Correa, M., Giardiello, F.M. et al. *Science* **2003**, *299*, 1753–1755.
31 Etzioni, R., Urban, N., Ramsey, S. et al. *Nature Rev. Cancer* **2003**, *3*, 243–252.
32 Burt, R.W. *Gastroenterology* **2000**, *119*, 837–853.
33 Gold, P., Freedman, S.O. *J. Exp. Med.* **1965**, *122*, 467–481.
34 Koprowski, H., Steplewski, Z., Mitchell, K., Herlyn, M., Herlyn, D., Fuhrer, P. *Somat. Cell Genet.* **1979**, *5*, 957–971.
35 Kuusela, P., Jalanko, H., Roberts, P. et al. *Br. J. Cancer* **1984**, *49*, 135–139.

36 Benson, A. B., III, Choti, M. A., Cohen, A. M., Doroshow, J. H., Fuchs, C., Kiel, K., Martin, E. W., Jr., McGinn, C., Petrelli, N. J. et al. *Oncology (Huntington)* **2000**, *14*, 203–212.

37 Duffy, M. J. *Clin. Chem.* **2001**, *47*, 624–630.

38 Hayes, D. F., Bast, R. C., Desch, C. E. et al. *J. Natl Cancer Inst.* **1996**, *88*, 1456–1466.

39 Srinivas, P. R., Kramer, B. S., Srivastava, S. *Lancet Oncol.* **2001**, *2*, 698–704.

40 Sidransky, D. *Nature Rev. Cancer* **2002**, *2*, 210–219.

41 Lecomte, T., Berger, A., Zinzindohoue, F. et al. *Int. J. Cancer* **2002**, *100*, 542–548.

42 Cui, H., Cruz-Correa, M., Giardiello, F. M. et al. *Science* **2003**, *299*, 1753–1755.

43 Laird, P. W. *Nature Rev. Cancer* **2003**, *3*, 253–266.

44 Stears, R. L., Martinsky, T., Schena, M. *Nature Med.* **2003**, *9*, 140–145.

45 van de Vijver, M. J., He, Y. D., van't Veer, L. J. et al. *N. Engl. J. Med.* **2002**, *347*, 1999–2009.

46 Petricoin, E. F., III, Hackett, J. L., Lesko, L. J. *et al. Nature Genet.* **2002**, *32 Suppl*, 474–479.

47 Velculescu, V. E., Vogelstein, B., Kinzler, K. W. *Trends Genet.* **2000**, *16*, 423–425.

48 Petricoin, E. F., Ardekani, A. M., Hitt, B. A. et al. *Lancet* **2002**, *359*, 572–577.

49 Wulfkuhle, J. D., Liotta, L. A., Petricoin, E. F. *Nature Rev. Cancer* **2003**, *3*, 267–275.

50 Dhar, P., Moore, T., Zamcheck, N., Kupchik, H. Z. *JAMA* **1972**, *221*, 31–35.

51 Hine, K. R., Dykes, P. W. *Br. J. Cancer* **1984**, *49*, 689–693.

52 Holten-Andersen, M. N., Christensen, I. J., Nielsen, H. J. et al. *Clin. Cancer Res.* **2002**, *8*, 156–164.

53 Kaminska, J., Kowalska, M. M., Nowacki, M. P., Chwalinski, M. G., Rysinska, A., Fuksiewicz, M. *Pathol. Oncol. Res.* **2000**, *6*, 38–41.

54 McMillan, D. C., Elahi, M. M., Sattar, N., Angerson, W. J., Johnstone, J., McArdle, C. S. *Nutr. Cancer* **2001**, *41*, 64–69.

55 De Vita, F., Romano, C., Orditura, M. et al. *J. Interferon Cytokine Res.* **2001**, *21*, 45–52.

56 Galizia, G., Orditura, M., Romano, C. et al. *Clin. Immunol.* **2002**, *102*, 169–178.

57 Simpson, R. J. Cold Spring Harbor Laboratory Press, New York, 2003.

58 Gorg, A., Obermaier, C., Boguth, G. et al. *Electrophoresis* **2000**, *21*, 1037–1053.

59 Corthals, G. L., Wasinger, V. C., Hochstrasser, D. F., Sanchez, J. C. *Electrophoresis* **2000**, *21*, 1104–1115.

60 Gorg, A., Obermaier, C., Boguth, G. et al. *Electrophoresis* **2000**, *21*, 1037–1053.

61 Hanash, S. M. *Electrophoresis* **2000**, *21*, 1202–1209.

62 Patton, W. F., Schulenberg, B., Steinberg, T. H. *Curr. Opin. Biotechnol.* **2002**, *13*, 321–328.

63 Patton, W. F. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* **2002**, *771*, 3–31.

64 Rabilloud, T., Strub, J. M., Luche, S., van Dorsselaer, A., Lunardi, J. *Proteomics* **2001**, *1*, 699–704.

65 Rabilloud, T. *Anal. Chem.* **2000**, *72*, 48A–55A.

66 Sinha, P., Poland, J., Schnolzer, M., Rabilloud, T. *Proteomics* **2001**, *1*, 835–840.

67 Unlu, M., Morgan, M. E., Minden, J. S. *Electrophoresis* **1997**, *18*, 2071–2077.

68 Zhou, G., Li, H., DeCamp, D. et al. *Mol. Cell Proteomics* **2002**, *1*, 117–124.

69 Jung, E., Heller, M., Sanchez, J. C., Hochstrasser, D. F. *Electrophoresis* **2000**, *21*, 3369–3377.

70 Welt, S., Scott, A. M., Divgi, C. R. et al. *J. Clin. Oncol.* **1996**, *14*, 1787–1797.

71 Ji, H., Moritz, R. L., Reid, G. E. et al. *Electrophoresis* **1997**, *18*, 614–621.

72 Moritz, R. L., Ritter, G., Catimel, B. et al. *J. Chromatogr. A* **1998**, *798*, 91–101.

73 Ritter, G., Cohen, L. S., Nice, E. C. et al. *Biochem. Biophys. Res. Commun.* **1997**, *236*, 682–686.

74 Heath, J. K., White, S. J., Johnstone, C. N. et al. *Proc. Natl Acad. Sci. USA* **1997**, *94*, 469–474.

75 Emmert-Buck, M. R., Bonner, R. F., Smith, P. D. et al. *Science* **1996**, *274*, 998–1001.

76 Emmert-Buck, M. R., Gillespie, J. W., Paweletz, C. P. et al. *Mol. Carcinog.* **2000**, *27*, 158–165.

77 Santoni, V., Kieffer, S., Desclaux, D., Masson, F., Rabilloud, T. *Electrophoresis* **2000**, *21*, 3329–3344.

78 Santoni, V., Molloy, M., Rabilloud, T. *Electrophoresis* **2000**, *21*, 1054–1070.

79 Rabilloud, T. *Nature Biotechnol.* **2003**, *21*, 508–510.

80 Luche, S., Santoni, V., Rabilloud, T. *Proteomics* **2003**, *3*, 249–253.

81 Simpson, R.J., Connolly, L.M., Eddes, J.S., Pereira, J.J., Moritz, R.L., Reid, G.E. *Electrophoresis* **2000**, *21*, 1707–1732.

82 Shin, B.K., Wang, H., Yim, A.M. et al. *J. Biol. Chem.* **2003**, *278*, 7607–7616.

83 Gygi, S.P., Corthals, G.L., Zhang, Y., Rochon, Y., Aebersold, R. *Proc. Natl Acad. Sci. USA* **2000**, *97*, 9390–9395.

84 Corthals, G.L., Wasinger, V.C., Hochstrasser, D.F., Sanchez, J.C. *Electrophoresis* **2000**, *21*, 1104–1115.

85 Isobe, T., Uchida, K., Taoka, M., Shinkai, F., Manabe, T., Okuyama, T. *J. Chromatogr.* **1991**, *588*, 115–123.

86 Matsuoka, K., Taoka, M., Isobe, T., Okuyama, T., Kato, Y. *J. Chromatogr.* **1990**, *515*, 313–320.

87 Wall, D.B., Kachman, M.T., Gong, S.Y. et al. *Anal. Chem.* **2000**, *72*, 1099–1111.

88 Washburn, M.P., Wolters, D., Yates, J.R., III. *Nature Biotechnol.* **2001**, *19*, 242–247.

89 Washburn, M.P., Ulaszek, R., Deciu, C., Schieltz, D.M., Yates, J.R., III. *Anal. Chem.* **2002**, *74*, 1650–1657.

90 Wolters, D.A., Washburn, M.P., Yates, J.R., III. *Anal. Chem.* **2001**, *73*, 5683–5690.

91 Hoffmann, P., Ji, H., Moritz, R.L. et al. *Proteomics* **2001**, *1*, 807–818.

92 Candiano, G., Musante, L., Bruschi, M. et al. *Electrophoresis* **2002**, *23*, 292–297.

93 Righetti, P.G., Castagna, A., Herbert, B. *Anal. Chem.* **2001**, *73*, 320A–326A.

94 Herbert, B., Righetti, P.G. *Electrophoresis* **2000**, *21*, 3639–3648.

95 Zuo, X., Speicher, D.W. *Anal. Biochem.* **2000**, *284*, 266–278.

96 Locke, V.L., Gibson, T.S., Thomas, T.M., Corthals, G.L., Rylatt, D.B. *Proteomics* **2002**, *2*, 1254–1260.

97 Steel, L.F., Trotter, M.G., Nakajima, P.B., Mattu, T.S., Gonye, G., Block, T. *Mol. Cell Proteomics* **2003**, *2*, 262–270.

98 Adkins, J.N., Varnum, S.M., Auberry, K.J. et al. *Mol. Cell Proteomics* **2002**, *1*, 947–955.

99 Wang, Y.Y., Cheng, P., Chan, D.W. *Proteomics* **2003**, *3*, 243–248.

100 Petricoin, E.F., Ardekani, A.M., Hitt, B.A. et al. *Lancet* **2002**, *359*, 572–577.

101 Conrads, T.P., Zhou, M., Petricoin, E.F., III, Liotta, L., Veenstra, T.D. *Expert. Rev. Mol. Diagn.* **2003**, *3*, 411–420.

102 Petricoin, E.F., Liotta, L.A. *J. Nutr.* **2003**, *133*, 2476S–2484S.

103 Liotta, L.A., Petricoin, E.F., III, Ardekani, A.M. et al. *Gynecol. Oncol.* **2003**, *88*, S25–S28.

104 Simon, R., Mirlacher, M., Sauter, G. *Expert. Rev. Mol. Diagn.* **2003**, *3*, 421–430.

105 Simon, R., Radmacher, M.D., Dobbin, K., McShane, L.M. *J. Natl Cancer Inst.* **2003**, *95*, 14–18.

106 Chaurand, P., Dague, B.B., Pearsall, R.S., Threadgill, D.W., Caprioli, R.M. *Proteomics* **2001**, *1*, 1320–1326.

107 Chaurand, P., Schwartz, S.A., Caprioli, R.M. *Curr. Opin. Chem. Biol.* **2002**, *6*, 676–681.

108 Chaurand, P., Caprioli, R.M. *Electrophoresis* **2002**, *23*, 3125–3135.

109 Stoeckli, M., Chaurand, P., Hallahan, D.E., Caprioli, R.M. *Nature Med.* **2001**, *7*, 493–496.

110 Yanagisawa, K., Shyr, Y., Baogang J. et al. *Lancet* **2003**, *362*, 433–439.

111 Ramaswamy, S., Tamayo, P., Rifkin, R. et al. *Proc. Natl Acad. Sci. USA* **2001**, *98*, 15149–15154.

112 Ross, D.T., Scherf, U., Eisen, M.B. et al. *Nature Genet.* **2000**, *24*, 227–235.

113 Haab, B.B. *Curr. Opin. Drug Discov. Dev.* **2001**, *4*, 116–123.

114 Haab, B.B., Dunham, M.J., Brown, P.O. *Genome Biol.* **2001**, *2*, RESEARCH 0004.

115 MacBeath, G., Schreiber, S.L. *Science* **2000**, *289*, 1760–1763.

116 Cahill, D.J. *J. Immunol. Methods* **2001**, *250*, 81–91.

117 Madoz-Gurpide, J., Wang, H., Misek, D.E., Brichory, F., Hanash, S.M. *Proteomics* **2001**, *1*, 1279–1287.

118 de Wildt, R.M., Mundy, C.R., Gorick, B.D., Tomlinson, I.M. *Nature Biotechnol.* **2000**, *18*, 989–994.

119 Walter, G., Bussow, K., Cahill, D., Lueking, A., Lehrach, H. *Curr. Opin. Microbiol.* **2000**, *3*, 298–302.

120 Sreekumar, A., Chinnaiyan, A.M. *Curr. Opin. Mol. Ther.* **2002**, *4*, 587–593.

121 Sreekumar, A., Chinnaiyan, A.M. *Biotechniques* **2002**, Suppl., 46–53.

122 Templin, M.F., Stoll, D., Schrenk, M., Traub, P.C., Vohringer, C.F., Joos, T.O. *Trends Biotechnol.* **2002**, *20*, 160–166.

123 Mitchell, P. *Nature Biotechnol.* **2002**, *20*, 225–229.

124 Li, M. *Nature Biotechnol.* **2000**, *18*, 1251–1256.

125 Arenkov, P., Kukhtin, A., Gemmell, A., Voloshchuk, S., Chupeeva, V., Mirzabekov, A. *Anal. Biochem.* **2000**, *278*, 123–131.

126 Zhu, H., Klemic, J.F., Chang, S. et al. *Nature Genet.* **2000**, *26*, 283–289.

127 Jessani, N., Liu, Y., Humphrey, M., Cravatt, B.F. *Proc. Natl Acad. Sci. USA* **2002**, *99*, 10335–10340.

128 Minamoto, T., Buschmann, T., Habelhah, H. et al. *Oncogene* **2001**, *20*, 3341–3347.

129 Kallioniemi, O.P. *Ann. Med.* **2001**, *33*, 142–147.

130 Kallioniemi, O.P., Wagner, U., Kononen, J., Sauter, G. *Hum. Mol. Genet.* **2001**, *10*, 657–662.

131 Moch, H., Kononen, T., Kallioniemi, O.P., Sauter, G. *Adv. Anat. Pathol.* **2001**, *8*, 14–20.

132 Yan, F., Sreekumar, A., Laxman, B., Chinnaiyan, A.M., Lubman, D.M., Barder, T.J. *Proteomics* **2003**, *3*, 1228–1235.

133 Scanlan, M.J., Chen, Y.T., Williamson, B. et al. *Int. J. Cancer* **1998**, *76*, 652–658.

134 Stockert, E., Jager, E., Chen, Y.T. et al. *J. Exp. Med.* **1998**, *187*, 1349–1354.

135 Old, L.J., Chen, Y.T. *J. Exp. Med.* **1998**, *187*, 1163–1167.

136 Sahin, U., Tureci, O., Schmitt, H. et al. *Proc. Natl Acad. Sci. USA* **1995**, *92*, 11810–11813.

137 Klade, C.S., Voss, T., Krystek, E. et al. *Proteomics* **2001**, *1*, 890–898.

138 Kellner, R., Lichtenfels, R., Atkins, D. et al. *Proteomics* **2002**, *2*, 1743–1751.

139 Lichtenfels, R., Kellner, R., Bukur, J. et al. *Proteomics* **2002**, *2*, 561–570.

140 Le Naour, F., Brichory, F., Misek, D. E., Brechot, C., Hanash, S.M., Beretta, L. *Mol. Cell Proteomics* **2002**, *1*, 197–203.

141 Seliger, B., Kellner, R. *Proteomics* **2002**, *2*, 1641–1651.

142 Yan, F., Sreekumar, A., Laxman, B., Chinnaiyan, A.M., Lubman, D.M., Barder, T.J. *Proteomics* **2003**, *3*, 1228–1235.

143 Mintz, P.J., Kim, J., Do, K.A. et al. *Nature Biotechnol.* **2003**, *21*, 57–63.

144 Chung, D.C. *Gastroenterology* **2000**, *119*, 854–865.

145 Srivastava, S., Verma, M., Henson, D.E. *Clin. Cancer Res.* **2001**, *7*, 1118–1126.

146 Chaurand, P., Caprioli, R.M. *Electrophoresis* **2002**, *23*, 3125–3135.

147 Wulfkuhle, J.D., Liotta, L.A., Petricoin, E.F. *Nature Rev. Cancer* **2003**, *3*, 267–275.

148 Ornstein, D.K., Gillespie, J.W., Paweletz, C.P. et al. *Electrophoresis* **2000**, *21*, 2235–2242.

149 Ji, H., Whitehead, R.H., Reid, G.E., Moritz, R.L., Ward, L.D., Simpson, R.J. *Electrophoresis* **1994**, *15*, 391–405.

150 Ji, H., Reid, G.E., Moritz, R.L., Eddes, J.S., Burgess, A.W., Simpson, R.J. *Electrophoresis* **1997**, *18*, 605–613.

151 Cole, A.R., Ji, H., Simpson, R.J. *Electrophoresis* **2000**, *21*, 1772–1781.

152 Jungblut, P.R., Zimny-Arndt, U., Zeindl-Eberhart, E. et al. *Electrophoresis* **1999**, *20*, 2100–2110.

153 Stulik, J., Osterreicher, J., Koupilova, K. et al. *Electrophoresis* **1999**, *20*, 1047–1054.

154 Stulik, J., Koupilova, K., Osterreicher, J. et al. *Electrophoresis* **1999**, *20*, 3638–3646.

155 Melis, R., White, R. *Electrophoresis* **1999**, *20*, 1055–1064.

156 Sinha, P., Hutter, G., Kottgen, E., Dietel, M., Schadendorf, D., Lage, H. *Electrophoresis* **1999**, *20*, 2961–2969.

157 Minowa, T., Ohtsuka, S., Sasai, H., Kamada, M. *Electrophoresis* **2000**, *21*, 1782–1786.

158 Szymczyk, P., Krajewska, W.M., Jakubik, J. et al. *Tumori* **1996**, *82*, 376–381.

159 Keesee, S.K., Meneghini, M.D., Szaro, R.P., Wu, Y.J. *Proc. Natl Acad. Sci. USA* **1994**, *91*, 1913–1916.

160 Chaurand, P., Dague, B.B., Pearsall, R.S., Threadgill, D.W., Caprioli, R.M. *Proteomics* **2001**, *1*, 1320–1326.

161 Anderson, N.L., Anderson, N.G. *Molecular & Cellular Proteomics* **2002**, *1(11)*, 845–867.

8
Clinical Proteomics: Ovarian Cancer

Ayodele A. Alaiya

8.1
Introduction

Approximately 90% of all human malignancies are carcinomas, i.e. they are derived from epithelial cells. Ovarian cancer is a major cause of morbidity and mortality in women in many parts of the world. Despite advances made in cancer treatment, the overall mortality rates for most solid tumors, including ovarian cancer, remain unchanged. For instance, it was estimated that well over 23 000 new cases of ovarian cancer would be diagnosed in the United States in the year 2001 and that approximately 60% of the women would die of the disease [1]. The incidence and mortality of ovarian cancer in Europe is similar to that in the USA (Fig. 8.1). Malignant ovarian tumors are heterogeneous in their biological and clinical behavior and a greater understanding of how they develop and progress is a prerequisite to successful early detection, screening programs, and treatment modalities.

Molecular analytical tools are now available that can be applied to clinical materials, increasing the potential for relating molecular alterations to clinical behavior and response to therapy. There are, however, several problems in evaluating the results of such analyses. The main problem is the lack of detailed understanding of the biological behavior of cancer with regard to parameters such as metastatic propensity. Another problem is tumor heterogeneity, as most solid tumors are heterogeneous with respect to proliferative activity and different cell types. Unpredictable biological behavior is today a major challenge in early diagnosis and management of ovarian malignancies. The benign and less aggressive tumors of low malignant potential show a very favorable prognosis with more than 95% with a 5-year survival rate. By contrast, biologically aggressive ovarian cancers are frequently diagnosed at a well-advanced stage and often rapidly lead to mortality; less than 35% of the patients live longer than 5 years [2].

Tumor-associated mechanisms that regulate cell cycle progression, apoptosis, and metastatic properties are complex. Although years of molecular analytical biology have been extremely fruitful and a large number of molecules and their involvement in these processes have been described, from the point of tumor diagnosis, this complexity has masked our ability to select which molecules represent

Biomedical Application of Proteomics
Edited by J.-C. Sanchez, G. L. Corthals, D. F. Hochstrasser

ISBN: 3-527-30807-5

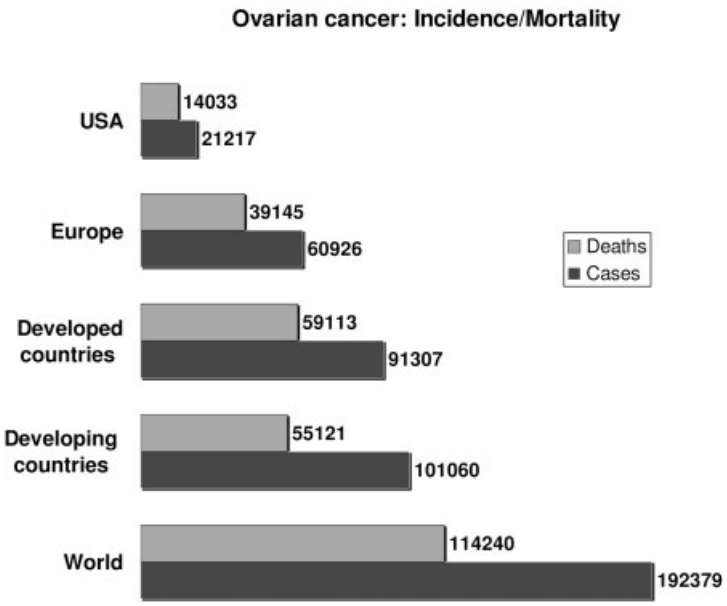


Figure 8.1 Global trends in ovarian cancer: Number of new cases and deaths per year [127].

the presence of tumors. The task ahead is not simple, as the finding of single molecules/markers is unlikely to yield accurate diagnostic and prognostic information.

Nevertheless, discovery of markers for early detection of ovarian cancer has the potential to significantly reduce the high mortality associated with this disease as well as improving the quality of life of ovarian cancer patients. Menon and Jacobs recently gave a concise account of recent developments in ovarian cancer screening since the release in 1998 of a comprehensive review of all ovarian cancer prospective screening studies conducted by the UK National Health Service Center for Reviews and Dissemination (NHSCR). A key message from this review is to work towards a better understanding of ovarian cancer screening and to define premalignant lesions. The role of ultrasonography alone or in combination with CA-125 needs further investigation in their ability to detect different stages of ovarian cancer [3].

New analytical tools in proteomics are emerging that give new insights into biological processes. These tools are powerful in that they allow for high-throughput analysis, which may speed up the discovery of potential biomarkers. Quantitative molecular variations may be used for the development of methods for tumor classification based on large amounts of gene expression data generated by two-dimensional electrophoresis (2-DE) analysis of proteins [4, 5]. In recent years surface-enhanced laser desorption/ionization (SELDI) time-of-flight mass spectrometry (TOF-MS) has been introduced for the rapid profiling of serum proteins. This

approach has the potential to be adapted for use in the screening and classification of ovarian cancer patients on a large scale. An initial success has been achieved using this approach in pattern-recognition analysis of serum patterns in ovarian cancer [6]. Indeed, with adequate knowledge of bioinformatics, SELDI has the potential to be used for the accurate diagnosis and early detection of preclinical lesions in various disease conditions, including cancer [7–10].

SELDI analysis does not provide identities of proteins directly, although the phenotypic fingerprint may provide sufficient information for diagnosis. The fact that phenotypic characteristics of tumors manifest at the protein level requires full characterization of the protein expression profiles in human carcinomas using proteomic approaches. Cancer proteomics is an aspect of biomedical research that promises to make an important contribution to our understanding of tumor biology. This chapter reviews the current state of the art as well as potential clinical applications of proteomics in human tumors and in particular ovarian cancer. It also demonstrates that proteomics is offering new tools for the study of complex biomedical problems. Finally, the problem areas that need to be addressed before these methods could have an impact on patient care are discussed.

8.2 General Background

8.2.1 Ovarian Cancer

Epithelial carcinoma of the ovary is one of the most common gynecologic malignancies with varying histological characteristics [11] (Fig. 8.2). It is the leading cause of cancer deaths among all gynecological malignancies in women and the fifth most common cancer type among women in the western world [12, 13]. The majority of malignant ovarian tumors occur in women over 65 years [14], while the benign tumors are generally more common in younger women between 25 and 45 years old.

Ovarian cancer is an insidious and aggressive disease because it is frequently asymptomatic and there is no sensitive screening method. Furthermore, no precursor lesion has been established and clinical manifestation often occurs at an advanced stage of the disease. The etiology of ovarian cancer is poorly understood, but studies have shown that some women are at a higher risk of developing ovarian tumor than others. Age, nulliparity, and family history of ovarian cancer and personal history of breast, colon, or endometrial cancer are among the widely studied associated risk factors in ovarian cancer [15]. More recently, genetic linkage analysis has resulted in the identification of the *BRCA1* and *BRCA2* genes. Women that have mutations in the *BRCA1* and *BRCA2* genes may have increased risk of developing ovarian cancer [16]. Mutations in *BRCA2* produce a relatively lower lifetime risk of between 10 and 20% compared with approximately 40–60% risk of developing ovarian cancer in women carrying mutations in *BRCA1* [17].

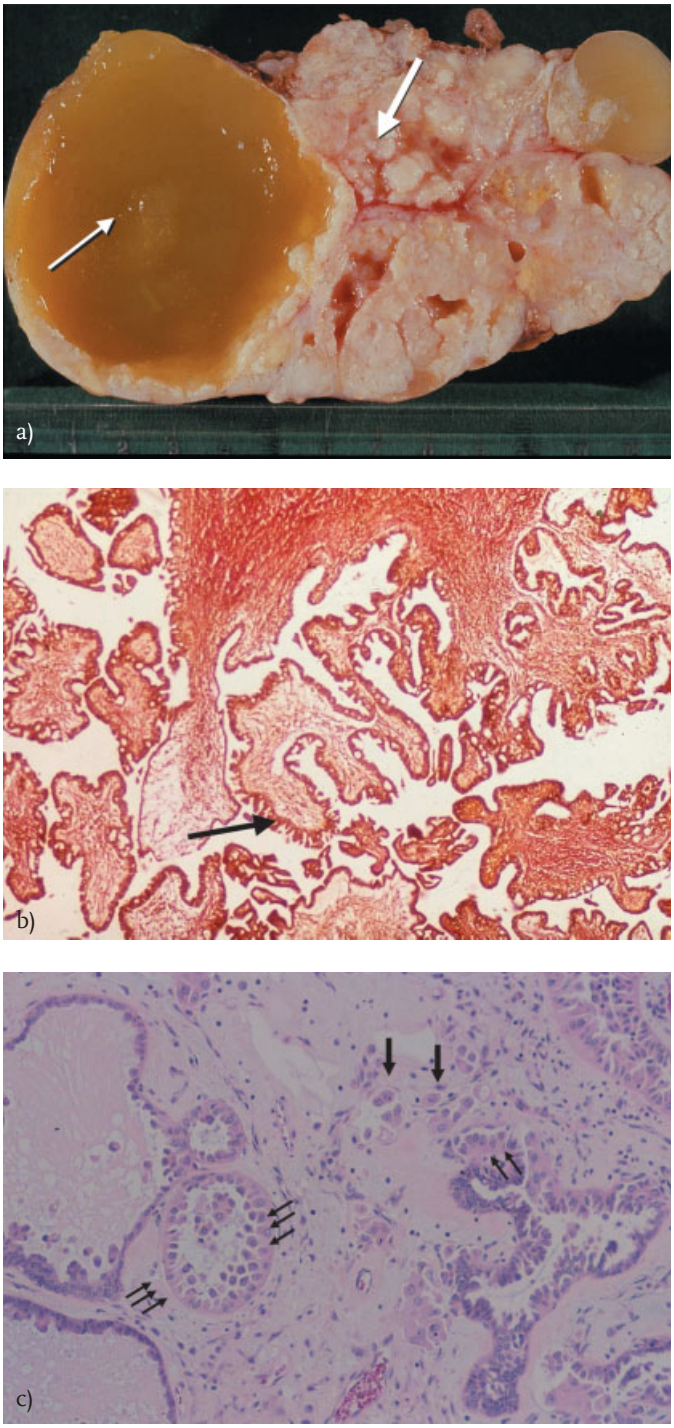


Figure 8.2 a–c

The three main types of ovarian tumors have been named after their cell of origin: (a) epithelial; (b) sex-cord; and (c) germ-cell tumor. Epithelial ovarian tumor is derived from the surface epithelium and is the most common, representing between 80 and 90% of all ovarian cancers. Amongst the different subtypes of common epithelial tumors are the serous, mucinous, endometrioid, and clear-cell types, depending on different cell and tissue features [18]. The sex-cord (stromal) and germ-cell tumors are relatively uncommon and represent approximately 6% and 3% of the cases respectively.

8.2.2
Ovarian Tumor Markers

Tissue markers such as enzymes, receptors, and various structural components can be used in the diagnosis and prognosis of neoplasms. Immunohistochemical analysis of hormone receptors, proliferation markers, proteases, and markers of angiogenesis are used for the routine diagnosis of cancer. Because of the multifactorial nature of cancer, it is very likely that only a combination of several markers (instead of a single marker) will effectively predict the biological behavior of different tumors.

CA-125 is a serum marker currently being used in the diagnosis and particularly in the prognosis of ovarian cancer. It has been validated and has reached widespread acceptance as an ovarian cancer tumor marker to aid in clinical diagnosis and monitoring of response to treatment. Recent attempts have been focused on improving the diagnostic accuracy of CA-125, either alone or in combination with new novel ovarian cancer biomarkers that are being discovered [19, 20]. CA-125 is found in breast milk and amniotic fluid in healthy women. However it is also present in women with gynecological conditions such as uterine leiomyoma (fibroid) and endometriosis, decreasing its specificity. CA-125 levels may vary with age, as preoperative measurement of CA-125 was found to be a better discriminator of benign and malignant ovarian tumors in postmenopausal women than in (younger) women in their reproductive years [21, 22]. The lack of a high predictive value of CA-125 assay may be due to the small number of times the assay has been used.

The use of statistical modeling and serial measurements of serum CA-125 levels over a period of time have been suggested in the past and it has been shown that this can improve the positive predictive power of CA-125 assay in the diagnosis and screening of ovarian cancer [23].

Figure 8.2 (a) Macroscopic view of a seropapillary ovarian cancer: Serous cystadenocarcinoma of the ovary are relatively common. The tumors are often unilateral although patients with tumors present in both ovaries also occur. Thin arrow shows gelatinous material and thick arrow show papillary excrescences. (b) A microscopic view of seropapillary ovarian tumor of borderline type. Arrow indicates papillary structure of the tumor that is lined by neoplastic cells. (c) A microscopic view of serous cystadenocarcinoma of the ovary. The epithelium shows eosinophilic cytoplasm (small arrows) and invasion of the ovarian stroma (thick arrow).

Table 8.1 Examples of marker proteins for ovarian tumor.

Name	*Abbreviation*	*References*
Ovarian carcinoma-associated antigen	OCA	[128, 129]
Macrophage colony-stimulating factor	M-CSF	[130]
Lysophosphatidic acid	LPA	[131]
Serum fragment of cytokeratin 19	CYF RA 21-1	[132–135]
Inhibin		[136–141]
Prostasin		[142]
Lewis X mucin determinant	OVX1	[143]
Lipid-associated sialic acid	LASA	[19, 144, 145]
Cancer-associated antigen 15-3	CA15-3	[146]
Cancer-associated antigen 19-9	CA 19-9	[27]
Cancer-associated antigen 72-4	CA72-4	[27, 147]
Cancer-associated antigen 125 plus D-dimer 10	CA-125 + D-dimer 10	[148–151]
Osteoponin		[116]
Human kallikrein	HK	[25, 152, 153]
Tumor-associated trypsin inhibitor	TATI	[154]
Cancer-associated serum antigen	CASA	[155–157]

Serum hK6 concentration has also been reported to decrease after surgical removal of ovarian cancer, indicating its potential as a prognostic marker. Recent work by Yousef and Diamandis described the differential analysis of some members of the human kallikrein (KLK) gene family in ovarian cancer using the SAGE (serial analysis of gene expression [24]) method [25]. The mRNA transcripts of each KLK were compared with the human expressed sequence tag (EST) databases of the Cancer Genome Anatomy Project (CGAP, http://cgap.nci.nih.gov/). The libraries screened included 9 normal, 3 premalignant, and 17 cancerous tissues. In this study, seven KLK genes were up-regulated in ovarian cancer compared with normal samples.

Other studies have shown that the combination of panels of associated markers have better potential to improve diagnostic efficacy than the use of single biomarkers. Diamandis and colleagues have reported that combined measurements of human kallikrein 6 (hK6) and serum CA-125 resulted in a 20% increase in diagnostic sensitivity compared with measurement of hKG alone [26].

Several other potential new biomarkers for ovarian cancer have been described recently (Table 8.1). However, the diagnostic sensitivity and specificity of these markers, either alone or in combination with CA-125, need further validation in order to establish their usefulness for disease diagnosis and prognosis. Among some of the reported serum markers for ovarian cancer are prostasin, OVX1, LASA, CA-15.3, CA-72.4, and inhibin [27, 28].

8.2.3
Screening Methods and Diagnostic Difficulties in Ovarian Tumors

Ovarian cancer is usually diagnosed by bimanual pelvic examination, transvaginal sonography, and measurement of serum CA-125 antigen [29]. None of these methods, singly or in combination, have a high specificity. The overall sensitivity in detecting early stage ovarian cancer was estimated to be below 30% [30]. Several studies have reported the use of transabdominal ultrasound (TAU) and transvaginal sonography (TVS) as the only ovarian cancer-screening tools [3, 31]. Although, TVS has been demonstrated to give a better image status of the ovary than TAU, some of the reported larger studies have shown a less than 10% positive predictive value of the use of ultrasonography in screening for ovarian cancer [32]. Consequently the diagnosis and staging of ovarian cancer is often done by an invasive laparotomy procedure.

Cancer diagnosis in general largely depends on morphologic alterations of cells and tissues [33]. In most tumors the distinction between benign and malignant tumors is possible; however, in a substantial number of tumors, more detailed examination, extensive experience, and in some cases additional markers are required before an accurate diagnosis can be made [34]. Early detection of preinvasive and invasive carcinomas of the ovary remains elusive with the present screening methods. This problem is compounded by the lack of well-defined ovarian cancer precursor lesions and inadequate knowledge about at what stage the tumors give rise to metastasis. It is not very clear whether or not stage I disease should be classified as precursor lesion.

In a recent review by Bast on the status of ovarian cancer screening, an estimation of the duration of preclinical ovarian tumor was made using serial measurements of CA-125. The results showed a mean duration of 1.9 ± 0.4 years for a preclinical ovarian cancer [32]. This therefore calls for the discovery of markers or a panel of markers with high sensitivity and specificity in order to achieve early detection and more efficient screening of ovarian cancer.

Of great diagnostic interest are certain ovarian epithelial tumors, the morphologic features of which do not clearly indicate whether they are malignant or not. These are referred to as tumors of low malignant potential (LMP tumors) or ovarian borderline tumors. LMP tumors proliferate slowly and exhibit low invasive potential. The distinction between borderline and malignant or between benign and borderline is often difficult. It is not clear whether borderline tumors represent intermediate steps in tumor progression or whether they should be considered as a separate group [35]. Therefore, common epithelial ovarian tumors are classified as benign, borderline, and malignant, illustrating the diagnostic problems [29, 36].

8.2.4
Treatment and Prognosis of Ovarian Tumors

Over the past two decades, significant progress has been achieved in the treatment of patients with advanced stage disease. Successful clinical trials have con-

tributed to global adoption of combination chemotherapy. Patients with advanced ovarian cancer are now commonly being treated using more aggressive surgery followed by combined chemotherapy (paclitaxel (Taxol)/carboplatin) [37]. Benign and early stage ovarian cancers are often cured by surgical removal of the tumors.

Prognosis in ovarian cancer is influenced by several factors, including younger age, tumor size, low stage disease, and degree of tumor differentiation, presence of ascites and cell type other than mucinous and clear cell [38, 39]. The 5-year survival rates varied significantly from benign to malignant tumors. Patients with a stage I ovarian carcinoma have a 5-year survival rate of approx. 80%, while the 5-year survival rate for patients with stage III disease is about 30% [2]. However, the advanced stage ovarian cancers accounts for a greater fraction of all the total diagnosed cases. The delay in the early diagnosis of the disease has been attributed to a poor overall survival rate of about 30% across all the disease stages [40].

8.3 Cancer Proteomics

8.3.1 Protein Profiling and Cancer

Decades of cancer research have been devoted to analysis of single or a few genes that are differentially expressed between normal and transformed cells. The complete sequence of billions of base pairs that make up the human genes has been a milestone achievement of the Human Genome Project [41–43]. Contrary to earlier expectations, there are fewer genes (approx. 30 000) in humans than in other lower organisms. However, we do know that the average human gene makes far more proteins/protein complexes and partially explains the complexity of the human proteome. The exact number of proteins in human cells is not known, but is estimated that 50 000 to >100 000 proteins are encoded, as each gene is thought to produce anywhere between 2 and 10 proteins. This increase in the estimated number of cellular proteins has been attributed to the fact that some genes produce multiple variants of the same protein through events such as alternative transcriptional processing and RNA splicing [44]. The biomolecules of the genome that are of immediate interest are the protein-coding genes whose sequences are transcribed into messenger RNA (mRNA) and subsequently translated into functional proteins.

The global analysis of differential changes in protein expression in a defined state has been termed “expression proteomics” [45]. Such analysis of differential protein expression between benign and malignant tumor cells may elucidate the potential roles of proteins in the development and progression of cancer. Cancer proteomics focuses on the expression profile and identification of cancer-related gene products at the protein level. Differential expression of various proteins is a reflection of the phenotypic changes resulting from altered genetic events. The functional characterization or pattern of expression may lead to a better under-

standing of pathogenesis and identification of novel markers, may have diagnostic use and may open new targets for therapeutic development.

8.3.2
RNA Expression Analysis in Cancer Cells: Promises and Pitfalls

Over the past few years several gene expression-profiling methods have been described and successfully used in life science research, including cancer. Differential screening of cDNA libraries has been commonly used to detect differences in gene expression between cell types [46, 47]. This technique is tedious and usually does not detect rare transcripts. With the advent of rapid DNA-sequencing techniques, investigators have begun to sequence randomly picked cDNA clones from libraries and to compare the spectrum of sequenced clones between different libraries (see Cancer Gene Anatomy Project, described below). The differential display method is based on polymerase chain reaction (PCR) amplification of DNA fragments from cDNA [48, 49]. A number of different primer combinations are used to display hundreds of cDNA fragments. By comparing different samples, changes in the relative abundance of nucleic acids can be detected.

SAGE is a powerful method capable of identification of a pattern of gene expression using cDNA sequencing information [24, 50]. Some advantages of this method over other methods such as cDNA microarrays are that SAGE is able to detect transcripts expressed at low copy numbers and is independent of prior knowledge of transcript information. However, it requires complex sample preparation protocols and large sequence information data. Another method of differential gene analysis is RNA modification of representational difference analysis (RDA). More recently, the use of microarrays [51–53] have become popular because they are relatively easy to use and require less DNA sequencing. This method uses cellular cDNAs as probes and thousands of individual cDNAs are immobilized and simultaneously quantified. The technique requires large amounts of PCR products or cDNAs in order to be used to monitor large genes of interest. The expressed sequence tag (EST) for the relative abundance of mRNA transcripts can be evaluated using the random sequencing of cDNA clones.

In our studies we have compared the abundance of mRNA transcripts and the corresponding proteins using CGAP data and quantitative 2-DE data. The relative abundance of some gene products, notably EF-2, calreticulin, and SOD, were much higher at the protein level than at the mRNA level [54]. Differences between protein and mRNA abundance levels are likely to be caused by variations in the half-life of the proteins. Andersson and Seilhamer [55] were first to report a correlation coefficient of less than 0.5 when mRNA and protein abundance of some gene products were compared in human liver. This result was further supported when expression of only one protein (GST-π) in 60 different human cell lines was studied by HPLC and quantitative northern blot analysis methods for protein and mRNA measurement respectively [56]. The implication of the few studies showing a poor correlation of mRNA and protein abundance is that it is necessary to study not only mRNA expression, but also protein expression, in or-

der to achieve a larger picture of the gene expression profile of a particular cell type. In addition to variations in protein turnover, post-translational events such as protein modifications will escape detection at the mRNA level.

8.3.3 Potentials and Limitations of Current Protein Profiling Technologies

Two-dimensional gel electrophoresis is the most powerful method currently available to resolve complex protein mixtures from cells, tissues, or other biological samples. The two main components of the classical proteomics combines the separation of polypeptides using high-resolution 2-DE together with MS or tandem MS (MS/MS) protein identification. 2-DE differential expression allows both qualitative and quantitative analysis of proteins and provides clues about possible post-translationally modified proteins that will elude detection using any of the nucleic acid-based methods such as the DNA sequencing or cDNA arrays. The unique potential of 2-DE has been clearly demonstrated in studies of global protein expression in clinical human tumors, which have resulted in the identification of several polypeptides with potential for application in clinical diagnostics [5, 57–59].

Advances in 2-DE technology have recently been made, especially in the resolution of proteins using immobilized pH gradient (IPG)-based isoelectric focusing. The robustness of IPG allows almost the entire protein spectrum to be resolved, including proteins with more basic isoelectric points [60, 61]. This technology has made 2-DE gels more reproducible and allows result comparisons between different laboratories. Further improvement allows high sample loading of up to milligram amounts of protein, which provides sufficient protein in various detection methods and subsequent protein identification. However, many proteins of functional significance occur in only trace amounts (< 1000 copies/cell). Many of these proteins may therefore elude detection using many of the presently available staining methods. More sensitive methods to resolve and visualize these low-abundance proteins are warranted to determine their contribution in disease development. The IPG strips exist in different ranges from ultra short to very wide including overlapping pH in between.

When different overlapping pH windows are used for the same sample, singly or in combination with sample fractionation, it may be possible to separate and analyze as many as 20 000 proteins with inclusion of many of the low-abundance proteins [62–65]. However, the resources, cost and, more importantly, insufficient amount of clinical samples for running several gels from the same samples precludes the routine clinical use of the so-called cyber gels.

A significant drawback of classical protein separation using 2-DE for many researchers is the slowness of the expression analysis. Alternative methodologies, more sophisticated data analysis software, and bioinformatics are key issues to address before 2-DE transforms become a rapid characterization technique. Fluorescence-based differential gel electrophoresis (DIGE) is a new development in proteomics, which allows the quantitative analysis of protein profiles in single gels [66, 67]. Here two or three samples can be labeled covalently with different fluo-

rescent dyes prior to isoelectric focusing; this mixture of samples is then run in a single 2-DE gel. Three fluorescent dyes currently commercially available are Cy2, Cy3, and Cy5 (Amersham Biosciences, San Francisco, CA, USA). Use of these dyes allows analysis of multiple samples with increased throughput compared with the conventional 2-DE method. This method also eliminates the problem of gel matching, since the two or three samples are run under identical conditions. When a large series of samples need to be analyzed, a third Cy dye may be used as a reference sample (e.g. a pool of all samples).

8.3.4
Unravelling Tumor Complexity Prior to Proteome Analysis

The quality and reproducibility of 2-DE gels depends largely on adequate sample preparation. The cells or tissue contents must be completely solubilized in order to fully extract the representative protein population required for analysis. The solubility of some membrane proteins may be improved by applying sequential extraction or, in the case of hydrophobic proteins, the use of a cocktail of different detergents [68]. Sample handling and preservation or storage is of paramount importance in proteome analysis. For example, the quality of protein extracts derived from formalin-fixed or paraffin-embedded tissue samples is not suitable for most protein analysis [69]. Sample heterogeneity and in particular tumor tissue heterogeneity need to be carefully addressed before analysis results can be seen as potential markers. The reliability and interpretation of results derived from whole-tissue samples significantly depends on the proportions of target cell populations. The size of tissue sample available for analysis is also a reflection of the extent of contamination of relevant cell population by other surrounding cell types. An admixture of normal cells, pathologic cells, stromal and connective tissue constitutes a disproportional representation of disease cells and reflects the extent of complexity of whole-tissue samples. Earlier work from our group indicated that the use of fresh tissue samples is superior to fresh frozen tissue samples, with clear evidence of enrichment of some protein species [70]. The successful purification of tumor epithelial cells using antibody-coated magnetic beads or Dynabeads® has also been recently described [71, 72].

In addition to purification of classes of cells, sample complexity can be further reduced by prefractionation of cellular organelles. For 2-DE, this improves the resolution and considerably large numbers of protein spots can be separated. In addition, it allows for identification of more proteins because a large quantity of a subset of proteins (organelle-specific) can be loaded and subsequently analyzed. The knowledge of the subcellular localization of a protein ultimately helps to assign specific functions to those proteins.

The use of culture cells can be seen as an alternative to the isolation of homogeneous cell populations. Although, short primary cell culture resembles the parent tissue, it is impossible to replace the natural environment of the primary source. A comparison of short primary culture of human prostate epithelial cells with the tumor cells from the tissue showed significantly altered protein profiles [73]. The

use of short-term primary cell culture from clinical material has also been used. Primary cultured cells derived from fresh, superficial transitional cell carcinomas (TCCs) were subjected to 2-D PAGE. Comparison of the 2-D gels from fresh tumors and their corresponding primary cultures showed a strikingly similar protein expression patterns. However, because of short-term culturing, a number of proteins whose rate of synthesis was differentially regulated between pairs of tumor and culture cells were observed [74].

A more recent approach capable of achieving high sample homogeneity is the use of laser-capture microdissection (LCM). This allows efficient procurement of microscopic and highly representative subpopulations of cells from complex heterogeneous tissue samples [75]. There has been wide application of LCM in proteomics. LCM was used to study the progression of prostate cancer by comparative protein analysis of normal, premalignant, and malignant prostate epithelial cells combined with SELDI-TOF analysis [76, 77]. Using this approach it was possible to precisely select homogeneous cell population of normal, carcinoma *in situ*, and malignant cells from the same tissue sample, thus allowing adequate assessment of intra-sample variability and extent of tissue complexity. Fend and colleagues have demonstrated the presence of bi-clonality in non-Hodgkin's lymphomas in the same tumor using cells obtained by LCM [78]. Suarez-Quian et al. described the use of LCM for the isolation of single cells [79], which could then be used for a more sophisticated gene expression analysis such as cDNA arrays [80, 81]. Other studies have used this method to procure tumor cells sufficient to run standard 2-D gel [82–84].

8.3.5 The Future of Clinical Proteomics: Challenges and Opportunities

New tools in proteomics such as laser-capture microdissection, SELDI-TOF, integrated affinity chromatography procedures, protein arrays, and other micro fluidic technologies are emerging and are capable of efficient and rapid protein analysis [85]. In the past 10 years the major breakthrough in proteomics was the development of highly sensitive protein identification techniques. Peptide mass fingerprinting (PMF) via MALDI-TOF and electrospray ionization (ESI)-MS-MS have now become standard methods in proteomic workflows. Their success is partially due to the continued development of comprehensive and non-redundant sequence databases and EST databases that allow protein identification by correlation of database sequence information with MS-generated data [86–95].

The recently introduced SELDI protein profiling method mentioned above is promising although it is apparent that medical proteomics requires protein expression patterns with subsequent protein identification.

The application of appropriate knowledge algorithms on proteomic data and adequate result interpretation represents two main focus areas that need to be addressed in order for proteome analysis to have a place in routine clinical use. Effective integration of clinical and pathological information into protein expression patterns will enhance the power of a multivariate analysis and prevent loss of vital

biological information. Such integrated networks may be further developed to serve as a "proteome scanner", i.e. an artificial intelligence tool capable of assisting and or complementing clinical judgements in establishing a more accurate diagnosis and prognosis based on an individual protein fingerprint.

8.4 Short Overview of Ovarian Cancer Proteomics

Ovarian cancer is a complex biological process that goes through several phases over a time course and that lacks well-defined precursor lesions. This makes screening and detection of early-stage ovarian cancer very difficult. In addition, many tumors show marked variations in response to treatment modalities. The discovery of biomolecules capable of detecting early events in cellular transformation to cancerous would provide an important opportunity to control cancer development and progression.

Proteome analysis as discussed above provides approaches to define markers useful for tumor diagnosis. Summaries of studies of protein expression in ovarian tumors and potential clinical applications, especially in tumor diagnosis and prognosis are given below. For reviewed studies on general cancer proteomics, the reader is referred to the reviewed articles and references therein [96–99].

8.4.1 The Promise of Proteomics in Ovarian Cancer Diagnostics

The main weakness of routine histopathological assessment of neoplasias is that tumor diagnosis is based on semi-quantitative evaluation by conventional microscopy of cells or tissue biopsy materials [33, 100, 101]. Because of large inter- or intra-observer variability, the method does not allow for reproducible accurate diagnosis. Immunostaining can be used to determine the expression of the various diagnostic markers and has a higher reproducibility. However, this method is not efficient for assessment of multiple markers at the same time. Expression profiling technologies such as DNA microarrays and proteomics do generate large sets of quantitative data that allow simultaneous differential analysis of genes and proteins of normal and disease states and at different stages of disease progression. Eisen and colleagues described a method of cluster analysis that is based on similarity in pattern of gene expression [102]. Using cDNA array-generated data, different methods have now been described for molecular classification of human disease including various forms of cancer such as leukemia, breast cancer, melanoma, lymphoma, and lung cancer [103–107].

There have been several interesting studies published on expression profiling of ovarian cancers [108–115]. Kim and colleagues reported osteoponin to be a potential biomarker that may aid in the diagnosis and prognosis of ovarian cancer [116]. Zhang and co-workers described the use of artificial neural network analysis on multiple ovarian cancer biomarkers. They were able to distinguish benign pel-

vic lesions from malignant ovarian cancer. This approach has the potential to detect early stage ovarian cancer [117].

8.4.2 Analysis of Tissue Samples

Borderline ovarian tumors (low malignant potential ovarian tumors) are pathologically characterized by their inability to invade the underlying tissue and show variability in clinical outcome compared with malignant tumors. Very often, the distinction between borderline and truly malignant invasive tumors may present with diagnostic ambiguity. This problem may eventually be solved by the discovery of more specific biomarkers and the use of novel classification methods. Jones and colleagues used expression data from 2-DE analysis of LCM cells derived from two tumors of low malignant potential and three ovarian cancers. The authors described 13 and 10 protein spots that were uniquely over- or under-expressed in malignant and borderline ovarian tumors respectively. Among the identified spots that are uniquely expressed in the malignant samples are a 52 kDa FK506 binding protein, Rho G-protein dissociation inhibitor (RhoGDI), and glyoxalase [118]. Although few patient samples were studied, these results are encouraging and deserve further evaluation.

We have reported earlier that ovarian tumors are relatively heterogeneous based on their global protein expression profiles [119]. We examined nine ovarian samples including benign, borderline, and malignant tumors. These tumors were large and different areas of the same tumor were examined by 2-DE. Interestingly, different areas of the same primary tumors showed similar protein profiles. In contrast, the differences in gene expression between pairs of malignant carcinomas were slightly larger than the observed differences between pairs of benign tumors. Studies from Alaiya and co-workers have resulted in proteomic definition of ovarian tumors as benign, borderline, and malignant [57]. Ovarian carcinomas showed high abundance of cell cycle-related proteins and members of the heat shock protein family. The expression of some cytoskeletal proteins and some high molecular weight tropomyosins were increased in the benign lesions compared with the invasive carcinomas. The borderline tumors showed intermediate expression of these proteins. Some of these proteins are indicated in the 2-DE map in Figure 8.3.

An attempt to apply artificial learning strategies using quantitative "2-dimensional array" data for ovarian tumor diagnosis was first presented by Alaiya et al. [5]. We have used the statistical concepts of principal components analysis (PCA) combined with partial least squares analysis (PLS) of a subset of variables in the 2-DE pattern, based on 170 polypeptides. A training model was constructed as a platform for classification into three groups (benign/borderline/malignant) using 22 tumors. The model was tested using 18 new tumors and we observed a correct classification of the majority of the cases (11/18). A clear separation between carcinomas and benign/borderline tumors was observed. Not surprisingly, benign and borderline tumors were more difficult to separate. If, however, the benign + borderline cases are defined as one group, 18/18 cases were correctly classified. This

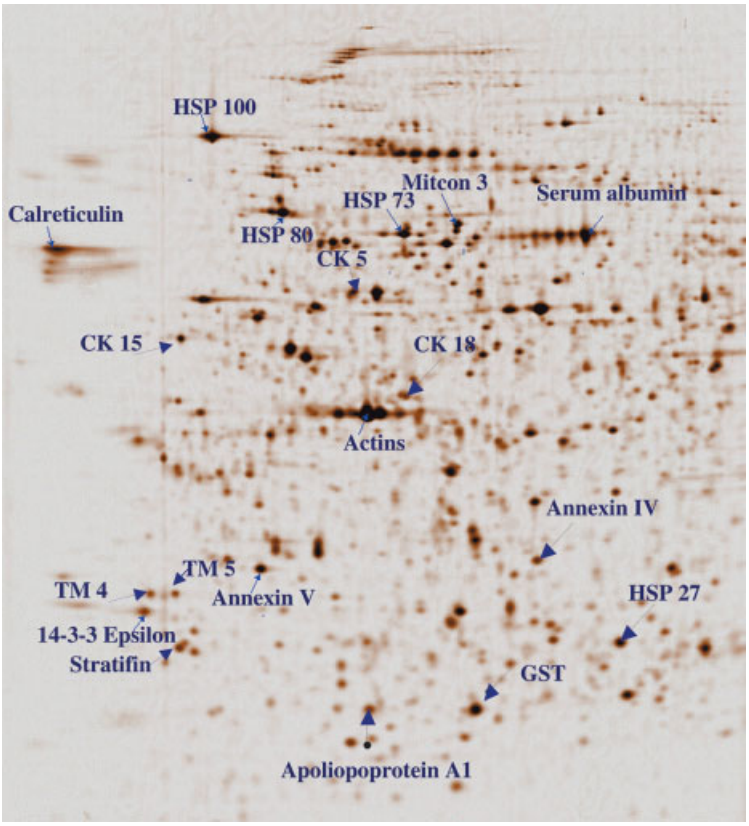


Figure 8.3 Silver-stained 2-DE map of ovarian carcinoma: Extracted proteins were resolved by linear IPG strip pH 4–7 followed by 10–13% gradient SDS-PAGE. Some polypeptides of known identity that differ in expression between benign, borderline and malignant ovarian tumors are indicated: HSP (heat shock proteins), TM (tropomyosin), GST (glutathione-S-transferase), CK (cytokeratin).

study represents the first using partial least squares/discriminant analysis (PLS-DA) and proteomics data and it may therefore be possible to further improve the classification of tumors by their constitutive gene expression profile using multivariate analysis.

8.4.3 Analysis of Serum Samples

More recently, initiatives from different leading proteomics research groups have been made on the analysis of blood serum protein profiles. The goal is to identify biomarkers in human blood samples for early detection of individuals with (early stage) cancer. One of the qualities of a biomarker is its ability to be measured in blood and other body fluids, and it is anticipated that subtle pathologic changes

taking place within a diseased tissue or organ may well be reflected in biomarker patterns measurable in the serum. SELDI-TOF is highly suited for biomarker discovery and as potential screening or diagnostic tool because it requires minimal sample preparation and only very small amounts of sample (microliters) are needed for the analysis. Paweletz and colleagues described SELDI-TOF analysis of normal and representative tumor cells procured by LCM. Characteristic sets of defined protein patterns were identified in the transition from normal cells through pre-neoplastic and to malignant prostate cancer [76].

Now differential expression analyses using protein data from plasma samples are emerging. Rai and colleagues have investigated plasma protein profiles of normal and ovarian cancer patients using SELDI protein chip analysis. They reported a set of seven "features" that were used to separate the cancer from normal samples. The diagnostic accuracy of the samples was improved when these features were measured in combination with serum CA-125 [120]. Two groups have both independently reported that analyses of complex spectra derived from SELDI-TOF MS experiments on nipple fluid aspirates resulted in the identification of several potential biomarkers for early detection of breast cancer [121, 122].

Several other studies have used the SELDI technology to identify potential biomarkers in premalignant and malignant tumors cells [123, 124]. Petricoin and colleagues recently described the SELDI approach as potential diagnostic tool for the early detection of ovarian cancer [6]. In this study, generic pattern recognition algorithms were used to analyze complex spectra generated by SELDI-TOF analysis of sera from normal healthy individual and diagnosed ovarian cancer patients. Using a set of selected discriminatory spectra, 50 cancer and 50 unaffected non-cancer serum samples were used to establish a learning classification model. All the samples were correctly classified as either cancer or non-cancer based on their characteristic proteomic patterns. One hundred and sixteen additional serum samples, including 50 ovarian cancer and 66 control unaffected women with non-gynecological diseases were then used to test the classification model. The algorithm correctly classified all ovarian cancers, including 18 samples with stage I disease, and 63 of the 66 control samples were classified as non-cancer. This study resulted in 100% sensitivity, 95% specificity, and 94% positive predictive value. The potential of this method as a possible screening tool for ovarian cancer especially in high-risk group justifies further investigation.

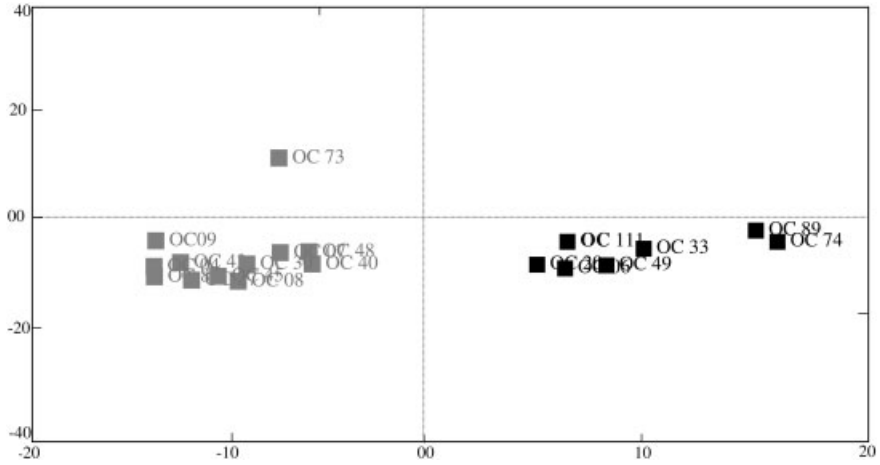
In keeping with the earlier observed 2-DE-based ovarian cancer profile, this finding suggests that it may be possible to classify tumors according to their constitutive protein expression profile using multivariate analysis, thus making classification by artificial intelligence a future possibility [5].

8.4.4 Disease Prognosis and Protein Expression Data

Despite the reports reviewed above and the tremendous achievements in the rate of identification of gel-separated proteins, most differential changes in protein expression across different samples do not correlate well with disease outcome.

Serial no	Pathologic diagnosis	Follow-up (year)	Status	Sample no.
1	Serous papillary adenocarcinoma IC	6	Alive	OC06
2	Serous papillary adenocarcinoma IC	5	Alive	OC20
3	Serous papillary adenocarcinoma IC	5	Alive	OC111
4	Serous papillary adenocarcinoma IC	5	Alive	OC33
5	Serous papillary adenocarcinoma IC	5	Alive	OC49
6	Serous papillary adenocarcinoma IC	5	Alive	OC74
7	Serous papillary adenocarcinoma IC	5	Alive	OC89
8	Mixed tumor	6	Dead	OC04
9	Clear cell tumor IVC	6	Dead	OC30
10	Clear cell tumor IVC	5	Dead	OC27
11	Endometrioid carcinoma IIIC	5	Dead	OC43
12	Serous papillary adenocarcinoma IC	4	Dead	OC07
13	Endometrioid carcinoma IIIC	4	Dead	OC8
14	Endometrioid carcinoma IIIC	4	Dead	OC09
15	Clear cell tumor IVC	4	Dead	OC40
16	Endometrioid carcinoma IIIC	4	Dead	OC45
17	Serous papillary adenocarcinoma IC	4	Dead	OC48
18	Serous papillary adenocarcinoma IC	3	Dead	OC73
19	Serous papillary adenocarcinoma IC	3	Dead	OC84

a)



b)

Figure 8.4 (a) Prognosis of ovarian cancers: Hierarchical cluster analysis of 19 ovarian carcinomas using 2-DE data from 142 protein spots to classify the samples in relation to patient survival. All tumor samples from surviving patients could be distinctly clustered. Included are the pathologic diagnosis and clinical follow–up periods. Analysis details are described elsewhere [4]. (b) Correspondence analysis plot of same data set as in (a). All tumor samples from surviving patients could be distinctly separated; black boxes = alive, open boxes = deceased (over 2–6 years clinical follow-up).

Only a small fraction of the identified polypeptides have so far been linked to changes in different disease processes. However, it should be kept in mind that using expression profiling in predicting disease prognosis will require a reasonable follow-up period, analysis of a large number of samples and development of efficient data analysis algorithms. Voss and co-workers described the potential of proteome data in disease prognosis in a recent study where 24 patients diagnosed with chronic lymphocytic leukemia (B-CLL) were studied using 2-D gel electrophoresis. The expression of 12 proteins, including redox enzymes, heat shock protein 27, and protein disulfide isomerase were observed to correlate well with patient survival [125].

We have also investigated whether prognostic information regarding patient outcome could be achieved using protein expression patterns [4]. A total of 40 ovarian tumor samples were studied consisting of 10 benign, 11 borderline, and 19 malignant tumors. The majority (12/19) of patients with ovarian carcinomas had a disease-related fatal clinical outcome during a 2- to 6-year clinical follow-up period. All of the benign and borderline patients were still alive during the same follow-up period. Hierarchical cluster analysis was used to evaluate protein expression profiles based on 142 (yet unidentified) proteins. We could not group tumors with a similar outcome when the data set comprising all 40 patients was evaluated. However, when only the 19 patients with ovarian carcinoma were evaluated, tumors from patients with similar outcome distinctly clustered together (Fig. 8.4) [126]. Analysis of larger numbers of patients with tumors of similar subtype and clinical stage will be necessary to validate this observation. Analysis of sufficient number of clinical samples to draw significant conclusions from many of the existing protein profiling studies is a recognized limiting factor in translational research. This situation may be improved if researchers can make their observed data sets publicly available in a central proteome database. Development of a common protocol of sample handling and data mining will be necessary to facilitate reliable interlaboratory results comparison. It may, therefore, be possible to predict disease prognosis based on their constitutive protein expression profiles using well-defined standard multivariate data analysis.

8.5 Acknowledgements

I thank Dr. Bo Franzén for comments and critical reading of the manuscript. The help of Dr. Claes Silfversvärd with some histologic slides is kindly acknowledged. Finally, the author thanks Ms Ingeborg May for photographic assistance.

8.6 References

1 R. T. Greenlee, M. B. Hill-Harmon, T. Murray, M. Thun. *CA Cancer J. Clin.* **2001**, *51*, 15.
2 R. Hand, A. Fremgen, J. S. Chmiel et al. *JAMA* **1993**, *269*, 1119.
3 U. Menon, I. J. Jacobs. *Curr. Opin. Obstet. Gynecol.* **2000**, *12*, 39.
4 A. A. Alaiya, B. Franzen, A. Hagman et al. *Int. J. Cancer* **2002**, *98*, 895.
5 A. A. Alaiya, B. Franzen, A. Hagman et al. *Int. J. Cancer* **2000**, *86*, 731.
6 E. F. Petricoin, A. M. Ardekani, B. A. Hitt et al. *Lancet* **2002**, *359*, 572.
7 E. F. Petricoin, K. C. Zoon, E. C. Kohn, J. C. Barrett, L. A. Liotta. *Nature Rev. Drug Discov.* **2002**, *1*, 683.
8 P. R. Srinivas, M. Verma, Y. Zhao et al. *Cancer Detect. Prev.* **2002**, *26*, 249.
9 C. M. Michener, A. M. Ardekani, E. F. Petricoin III, L. A. Liotta, E. C. Kohn. *Cancer Detect. Prev.* **2002**, *26*, 249.
10 P. C. Herrmann, L. A. Liotta, I. E. Petricoin. *Dis. Markers* **2001**, *17*, 49.
11 P. Russell. In *Blaustein's Pathology of the Female Genital Tract* (R. J. Kurman, A. Blaustein, Eds). Springer, New York, **1994**, p 705.
12 E. C. Grendys, Jr. *Curr. Opin. Obstet. Gynecol.* **1998**, *10*, 1.
13 A. P. M. Heintz. *Curr. Opin. Obstet. Gynecol.* **1997**, *9*, 1.
14 R. Yancik. *Cancer* **1993**, 71, 517.
15 C. I. Amos, J. P. Struewing. *Cancer* **1993**, *71*, 566.
16 S. A. Gayther, B. A. Ponder. *Mol. Med. Today* **1997**, *3*, 168.
17 D. F. Easton, D. Ford, D. T. Bishop. *Am. J. Hum. Genet.* **1995**, *56*, 265.
18 M. P. Boente, J. Hurteau, G. C. Rodriguez, R. C. Bast, Jr., A. Berchuck. *Curr. Opin. Oncol.* **1993**, *5*, 900.
19 R. P. Woolas, M. R. Conaway, F. Xu et al. *Gynecol. Oncol.* **1995**, *59*, 111.
20 R. P. Woolas, F. J. Xu, I. J. Jacobs et al. *J. Natl Cancer Inst.* **1993**, *85*, 1748.
21 H. W. de Bruijn, A. G. van der Zee, J. G. Aalders. *Curr. Opin. Obstet. Gynecol.* **1997**, *9*, 8.
22 D. H. Oram, A. R. Jeyarajah. *Br. J. Obstet. Gynaecol.* **1994**, *101*, 939.
23 S. J. Skates, F. J. Xu, Y. H. Yu et al. *Cancer* **1995**, *76*, 2004.
24 V. E. Velculescu, L. Zhang, B. Vogelstein, K. W. Kinzler. *Science* **1995**, *270*, 484.
25 G. M. Yousef, E. P. Diamandis. *Tumour Biol.* **2002**, *23*, 185.
26 E. P. Diamandis, A. Scorilas, S. Fracchioli et al. *J. Clin. Oncol.* **2003**, *21*, 1035.
27 T. Meyer, G. J. Rustin, *Br. J. Cancer* **2000**, *82*, 1535.
28 J. S. Berek, R. C. Bast, Jr. *Cancer* **1995**, *76*, 2092.
29 M. L. Friedlander. *Semin. Oncol.* **1998**, *25*, 305.
30 J. O'Rourke, S. M. Mahon. *Clin. J. Oncol. Nurs.* **2002**, *7*, 41.
31 S. Campbell, V. Bhan, P. Royston, M. I. Whitehead, W. P. Collins. *BMJ* **1989**, *299*, 1363.
32 R. C. Bast, Jr., *J. Clin. Oncol.* **2003**, *21*, 200.
33 R. Fox, M. Hull. *Ann. NY Acad. Sci.* **1993**, *687*, 217.
34 L. W. Dalton, D. L. Page, W. D. Dupont. *Cancer* **1994**, *73*, 2765.
35 C. J. Link, Jr., E. Reed, G. Sarosy, E. C. Kohn. *Am. J. Med.* **1996**, *101*, 217.
36 C. Bergman, M. Boente. *Curr. Opin. Oncol.* **1998**, *10*, 434.
37 R. F. Ozols. *Semin. Oncol.* **2000**, *27*, 3.
38 G. A. Omura, M. F. Brady, H. D. Homesley et al. *J. Clin. Oncol.* **1991**, *19*, 1138.
39 T. Thigpen, M. F. Brady, G. A. Omura et al. *Cancer* **1993**, *71*, 606.
40 D. Gertig, D. Hunter. In *Textbook of Cancer Epidemiology* (H.-O. Adami, D. Hunter, D. Trichopoulos, Eds). Oxford University Press, New York, **2002**, p 378.
41 J. C. Venter, M. D. Adams, E. W. Myers et al. *Science* **2001**, *291*, 1304.
42 E. S. Lander, L. M. Linton, B. Birren et al. *Nature* **2001**, *409*, 860.
43 G. Subramanian, M. D. Adams, J. C. Venter, S. Broder, *JAMA* **2001**, *286*, 2296.
44 A. A. Gooley, N. H. Packer. In *Proteome Research: New Frontiers in Functional Genomics (Principles and Practice)* (M. R.

Wilkins, K.L. Williams, R.D. Appel, D.F. Hochstrasser, Eds). Springer, Berlin, **1997**, p 65.
45 R. Simpson, D. Dorow. *Trends Biotechnol.* **2001**, *2001*, S40.
46 C.M. Perou, S.S. Jeffrey, M. van de Rijn et al. *Proc. Natl Acad. Sci. USA* **1999**, *96*, 9212.
47 D.T. Ross, U. Scherf, M.B. Eisen et al. *Nature Genet.* **2000**, *24*, 227.
48 P. Liang, A.B. Pardee. *Science* **1992**, *257*, 967.
49 Q. Cheng, W.M. Lau, S.K. Tay, S.H. Chew, T.H. Ho, K.M. Hui. *Int. J. Cancer* **2002**, *98*, 419.
50 V.E. Velculescu, B. Vogelstein, K.W. Kinzler. *Trends Genet.* **2000**, *16*, 423.
51 V.R. Iyer, M.B. Eisen, D.T. Ross et al. *Science* **1999**, *283*, 83.
52 S. Granjeaud, F. Bertucci, B.R. Jordan. *Bioessays* **1999**, *21*, 781.
53 S. Granjeaud, P. Naquet, F. Galland. *Immunogenetics* **1999**, *49*, 964.
54 A. Alaiya, U. Roblick, L. Egevad et al. *Anal. Cell Pathol.* **2000**, *21*, 1.
55 L. Anderson, J. Seilhamer. *Electrophoresis* **1997**, *18*, 533.
56 K. Anderson, R. Andrews, L. Yin et al. *Hum. Exp. Toxicol.* **1998**, *17*, 131.
57 A.A. Alaiya, B. Franzen, K. Fujioka et al. *Int. J. Cancer* **1997**, *73*, 678.
58 R.E. Banks, M.J. Dunn, D.F. Hochstrasser et al. *Lancet* **2000**, *356*, 1749.
59 J.E. Celis, H. Wolf, M. Ostergaard. *Electrophoresis* **2000**, *21*, 2115.
60 A. Gorg, C. Obermaier, G. Boguth et al. *Electrophoresis* **2000**, *21*, 1037.
61 A. Gorg, C. Obermaier, G. Boguth, W. Weiss. *Electrophoresis* **1999**, *20*, 712.
62 M.P. Molloy, B.R. Herbert, B.J. Walsh et al. *Electrophoresis* **1998**, *19*, 837.
63 T. Rabilloud. *Electrophoresis* **1996**, *17*, 813.
64 G.L. Corthals, V.C. Wasinger, D.F. Hochstrasser, J.C. Sanchez. *Electrophoresis* **2000**, *21*, 1104.
65 R.M. Levenson, G.M. Anderson, J.A. Cohn, P.J. Blackshear. *Electrophoresis* **1990**, *11*, 269.
66 R. Tonge, J. Shaw, B. Middleton et al. *Proteomics* **2001**, *1*, 377.
67 M. Unlu, M.E. Morgan, J.S. Minden. *Electrophoresis* **1997**, *18*, 2071.
68 M.P. Molloy. *Anal. Biochem.* **2000**, *280*, 1.
69 M. Ahram, M.J. Flaig, J.W. Gillespie et al. *Proteomics* **2003**, *3*, 413.
70 B. Franzen, S. Linder, K. Okuzawa, H. Kato, G. Auer. *Electrophoresis* **1993**, *14*, 1045.
71 M.J. Page, B. Amess, R.R. Townsend et al. *Proc. Natl Acad. Sci. USA* **1999**, *96*, 12589.
72 M.A. Reymond, J.C. Sanchez, G.J. Hughes et al. *Electrophoresis* **1997**, *18*, 2842.
73 D.K. Ornstein, J.W. Gillespie, C.P. Paweletz et al. *Electrophoresis* **2000**, *21*, 2235.
74 A. Celis, H.H. Rasmussen, P. Celis et al. *Electrophoresis* **1999**, *20*, 355.
75 M.R. Emmert-Buck, R.F. Bonner, P.D. Smith et al. *Science* **1996**, *274*, 998.
76 C.P. Paweletz, J.W. Gillespie, D.K. Ornstein et al. *Drug Dev. Res.* **2000**, *49*, 34.
77 E.F. Petricoin III, D.K. Ornstein, C.P. Paweletz et al. *J. Natl Cancer Inst.* **2002**, *94*, 1576.
78 F. Fend, L. Quintanilla-Martinez, S. Kumar et al. *Am. J. Pathol.* **1999**, *154*, 1857.
79 C.A. Suarez-Quian, S.R. Goldstein, T. Pohida et al. *Biotechniques* **1999**, *26*, 328.
80 W. Dietmaier, A. Hartmann, S. Wallinger et al. *Am. J. Pathol.* **1999**, *154*, 83.
81 L. Luo, R.C. Salunga, H. Guo et al. *Nature Med.* **1999**, *5*, 117.
82 L. Mouledous, S. Hunt, R. Harcourt, J.L. Harry, K.L. Williams, H.B. Gutstein. *Electrophoresis* **2003**, *24*, 296.
83 J.D. Wulfkuhle, K.C. McLean, C.P. Paweletz et al. *Proteomics* **2001**, *1*, 1205.
84 R.A. Craven, R.E. Banks. *Proteomics* **2001**, *1*, 1200.
85 H. Zhou, S. Roy, H. Schulman, M. Natan. *Trends Biotechnol.* **2001**, *2001*, S34.
86 E. Mortz, P.B. O'Connor, P. Roepstorff et al. *Proc. Natl Acad. Sci. USA* **1996**, *93*, 8264.
87 K.R. Clauser, S.C. Hall, D.M. Smith et al. *Proc. Natl Acad. Sci. USA* **1995**, *92*, 5072.
88 J.K. Eng, A.L. McCormack, J.R. Yates III. *J. Am. Soc. Mass Spectrometry* **1994**, *5*, 976.

89 W. J. Henzel, T. M. Billeci, J. T. Stults, S. C. Wong, C. Grimley, C. Watanabe. *Proc. Natl Acad. Sci. USA* **1993**, *90*, 5011.

90 P. James, M. Quadroni, E. Carafoli, G. Gonnet. *Biochem. Biophys. Res. Commun.* **1993**, *195*, 58.

91 M. Mann, P. Hojrup, P. Roepstorff. *Biol. Mass Spectrom.* **1993**, *22*, 338.

92 M. Mann, M. Wilm. *Anal. Chem.* **1994**, *66*, 4390.

93 A. L. McCormack, D. M. Schieltz, B. Goode et al. *Anal. Chem.* **1997**, *69*, 767.

94 D. J. C. Pappin, P. Hojrup, A. J. Bleasby. *Curr. Biol.* **1993**, *13*, 327.

95 J. R. Yates III, S. Speicher, P. R. Griffin, T. Hunkapiller. *Anal. Biochem.* **1993**, *214*, 397.

96 S. M. Hanash, M. P. Bobek, D. S. Rickman et al. *Proteomics* **2002**, *2*, 69.

97 V. E. Bichsel, L. A. Liotta, E. F. Petricoin III. *Cancer J.* **2001**, *7*, 69.

98 A. A. Alaiya, B. Franzen, G. Auer, S. Linder. *Electrophoresis* **2000**, *21*, 1210.

99 G. Chambers, L. Lawrie, P. Cash, G. I. Murray. *J. Pathol.* **2000**, *192*, 280.

100 P. Kronqvist, R. Montironi, T. Kuopio, Y. U. Collan. *Anal. Quant. Cytol. Histol.* **1997**, *19*, 423.

101 H. Fox. *Genet. Diagn. Pathol.* **1997**, *143*, 117.

102 M. B. Eisen, P. T. Spellman, P. O. Brown, D. Botstein. *Proc. Natl Acad. Sci. USA* **1998**, *95*, 14863.

103 T. R. Golub, D. K. Slonim, P. Tamayo et al. *Science* **1999**, *286*, 531.

104 C. M. Perou, T. Sorlie, M. B. Eisen et al. *Nature* **2000**, *406*, 747.

105 M. Bittner, P. Meltzer, Y. Chen et al. *Nature* **2000**, *406*, 536.

106 A. A. Alizadeh, M. B. Eisen, R. E. Davis et al. *Nature* **2000**, *403*, 503.

107 M. E. Garber, O. G. Troyanskaya, K. Schluens et al. *Proc. Natl Acad. Sci. USA* **2001**, *98*, 13784.

108 M. Schummer, W. V. Ng, R. E. Bumgarner et al. *Gene* **1999**, *238*, 375.

109 K. Wang, L. Gan, E. Jeffery et al. *Gene* **1999**, *229*, 101.

110 T. S. Furey, N. Cristianini, N. Duffy, D. W. Bednarski, M. Schummer, D. Haussler. *Bioinformatics* **2000**, *16*, 906.

111 R. S. Ismail, R. L. Baldwin, J. Fang et al. *Cancer Res.* **2000**, *60*, 6744.

112 A. M. Martoglio, B. D. Tom, M. Starkey, A. N. Corps, D. S. Charnock-Jones, S. K. Smith. *Mol. Med.* **2000**, *6*, 750.

113 K. Ono, T. Tanaka, T. Tsunoda et al. *Cancer Res.* **2000**, *60*, 5007.

114 J. B. Welsh, P. P. Zarrinkar, L. M. Sapinoso et al. *Proc. Natl Acad. Sci. USA* **2001**, *98*, 1176.

115 I. Haviv, I. G. Campbell. *Mol. Cell Endocrinol.* **2002**, *191*, 121.

116 J. H. Kim, S. J. Skates, T. Uede et al. *JAMA* **2002**, *287*, 1671.

117 Z. Zhang, S. D. Barnhill, H. Zhang et al. *Gynecol. Oncol.* **1999**, *73*, 56.

118 M. B. Jones, H. Krutzsch, H. Shu et al. *Proteomics* **2002**, *2*, 76.

119 A. A. Alaiya, B. Franzen, B. Moberger, C. Silfversward, S. Linder, G. Auer. *Electrophoresis* **1999**, *20*, 1039.

120 A. J. Rai, Z. Zhang, J. Rosenzweig et al. *Arch. Pathol. Lab. Med.* **2002**, *126*, 1518.

121 C. P. Paweletz, B. Trock, M. Pennanen et al. *Dis. Markers* **2001**, *17*, 301.

122 E. R. Sauter, W. Zhu, X. J. Fan, R. P. Wassell, I. Chervoneva, G. C. Du Bois. *Br. J. Cancer* **2002**, *86*, 1440.

123 G. W. Jr, L. H. Cazares, S. M. Leung et al. *Prostate Cancer Prostatic Dis.* **1999**, *2*, 264.

124 A. Vlahou, P. F. Schellhammer, S. Mendrinos et al. *Am. J. Pathol.* **2001**, *158*, 1491.

125 T. Voss, H. Ahorn, P. Haberl, H. Dohner, K. Wilgenbus. *Int. J. Cancer* **2001**, *91*, 180.

126 A. A. Alaiya, U. J. Roblick, B. Franzen, H. P. Bruch, G. Auer. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* **2003**, *787*, 207.

127 F. B. J. Ferlay, P. Pisani, D. M. Parkin. In *IARC CancerBase* No. 5. IARC Press, Lyon, 2001.

128 Y. Valladares. *Rev. Esp. Oncol.* **1985**, *32*, 363.

129 M. A. McGuckin, G. T. Layton, M. J. Bailey, T. Hurst, S. K. Khoo, B. G. Ward. *Gynecol. Oncol.* **1990**, *37*, 165.

130 F. J. Xu, S. Ramakrishnan, L. Daly et al. *Am. J. Obstet. Gynecol.* **1991**, *165*, 1356.

131 Y. Xu, Z. Shen, D. W. Wiper et al. *JAMA* **1998**, *280*, 719.

132 H. Bodenmuller. *Scand. J. Clin. Lab. Invest. Suppl.* **1995**, *221*, 60.

133 A. Gadducci, M. Ferdeghini, S. Cosio, A. Fanucchi, R. Cristofani, A. R. Genazzani. *Int. J. Gynecol. Cancer* **2001**, *11*, 277.

134 C. Tempfer, L. Hefler, H. Heinzl et al. *Br. J. Cancer* **1998**, *78*, 1108.

135 R. Molina, C. Agusti, X. Filella et al. *Tumour Biol.* **1994**, *15*, 318.

136 D. M. Robertson, T. Stephenson, E. Pruysers et al. *Mol. Cell Endocrinol.* **2002**, *191*, 97.

137 H. G. Burger, P. J. Fuller, S. Chu et al. *Mol. Cell Endocrinol.* **2001**, *180*, 145.

138 D. M. Robertson, N. Cahir, H. G. Burger et al. *Clin. Chem.* **1999**, *45*, 651.

139 I. Cooke, M. O'Brien, F. M. Charnock, N. Groome, T. S. Ganesan. *Br. J. Cancer* **1995**, *71*, 1046.

140 R. E. Lappohn, H. G. Burger, J. Bouma, M. Bangah, M. Krans, H. W. de Bruijn. *N. Engl. J. Med.* **1989**, *321*, 790.

141 M. Rishi, L. N. Howard, G. L. Bratthauer, F. A. Tavassoli. *Am. J. Surg. Pathol.* **1997**, *21*, 583.

142 S. C. Mok, J. Chao, S. Skates et al. *J. Natl Cancer Inst.* **2001**, *93*, 1458.

143 F. J. Xu, Y. H. Yu, L. Daly et al. *J. Clin. Oncol.* **1993**, *11*, 1506.

144 J. H. Nam, L. A. Cole, J. T. Chambers, P. E. Schwartz. *Gynecol. Oncol.* **1990**, *36*, 383.

145 A. M. Dnistrian, M. K. Schwartz. *Clin. Chem.* **1981**, *27*, 1737.

146 G. Scambia, P. Benedetti Panici, G. Baiocchi et al. *Oncology* **1988**, *45*, 263.

147 F. Guadagni, M. Roselli, M. Cosimelli et al. *Cancer Invest.* **1995**, *13*, 227.

148 A. Gadducci, U. Baicchi, R. Marrai, M. Ferdeghini, R. Bianchi, V. Facchini. *Gynecol. Oncol.* **1996**, *60*, 197.

149 C. Rella, M. Coviello, N. De Frenza et al. *Tumori* **1993**, *79*, 347.

150 S. S. Mirshahi, E. Pujade-Lauraine, C. Soria et al. *Cancer* **1992**, *69*, 2289.

151 R. Schrock, R. Hafter, H. Graeff, L. Schmid. *Onkologie* **1985**, *8*, 260.

152 E. P. Diamandis, G. M. Yousef, A. R. Soosaipillai, P. Bunting. *Clin. Biochem.* **2000**, *33*, 579.

153 L. Y. Luo, P. Bunting, A. Scorilas, E. P. Diamandis. *Clin. Chim. Acta* **2001**, *306*, 111.

154 U. H. Stenman, H. Alfthan, J. Vartiainen, P. Lehtovirta. *Ann. Med.* **1995**, *27*, 115.

155 M. K. Oehler, M. Sutterlin, H. Caffier. *Anticancer Res.* **1999**, *19*, 2513.

156 O. Kierkegaard, O. Mogensen, B. Mogensen, A. Jakobsen. *Gynecol. Oncol.* **1995**, *59*, 251.

157 U. Hasholzner, L. Baumgartner, P. Stieber, W. Meier, K. Hofmann, A. Fateh-Moghadam. *Anticancer Res.* **1994**, *14*, 2743.

9
Protein Expression Profiling Analysis in Hematopoietic Stem Cells: Phenotypic Characterization of Mesenchymal Stem Cells

Juan Antonio López, Antonio Bernad, and Juan Pablo Albar

9.1
Introduction

When adult mesenchymal tissues containing undifferentiated cells are cultured, the cells retain their capacity to differentiate to osteocyte-, adipocyte-, myocyte-, or chondrocyte-derived lineages when induced by the appropriate signals. These isolated cell populations should serve as ideal candidates in tissue engineering applications for repairing and regeneration of damaged tissues. Another application of mesenchymal stem cells (MSCs) might be their use as vehicles for gene therapy when engineered to express therapeutic proteins and then returned to the patient to have these proteins secreted. The molecular mechanisms underlying the signals that induce the differentiation of MSCs into distinctive mature lineages are not fully understood. A major goal of our work is to identify not only the proteins responsible for maintaining the multipotential mesenchymal state, but also the factors that induce the specific differentiation programs of MSCs into osteocytes, adipocytes, and chondrocytes. Applied to the different stages of our work, proteomics, the global study of proteins expressed by a cell, tissue, or organelle in a given time, should provide us the basis for a more comprehensive understanding of the processes involved in differentiation.

9.2
Mesenchymal Stem Cells

It is known that a zygote, after implantation in the right conditions and environment, develops its immense biological potential, generating a full assemblage of cell types that constitute a new human being. During many of the embryonic stages totipotential stem cells exist (a term used for cells capable of generating all cell lineages present in an adult). These embryonic totipotential cells have been termed "embryonic stem cells" (ES) [1, 2] and are found in the embryonic inner cell mass in early stages of development [3]. A single ES cell is capable of generating a mouse indistinguishable from their congeners, i.e. their original donor mouse strain. Genetic manipulation of ES cells is the key step for the develop-

Biomedical Application of Proteomics
Edited by J.-C. Sanchez, G. L. Corthals, D. F. Hochstrasser

ISBN: 3-527-30807-5

ment of a mouse model for human diseases and a critical tool in the *in vivo* analysis of the functional role of specific genes.

The idea to use stem cells in adult organisms was established after the pioneering work done on bone marrow transplants. The bone marrow produces all the required blood cells through a complex and pyramidal developmental program. Seminal for the perfect functioning of the system during the whole life of the animal is the maintenance of a small number (0.01–0.1% of the total bone marrow cells) of stem cells, allocated in the long bones of humans. The isolation and transplantation of a sufficient number of these stem cells reconstitutes and maintains the fully functional lympho-haematopoietic system of the recipient organism. In the last 10 years this initial concept has been extended to many other organs in the adult organism.

In the adult organism the immense majority of cells are in a post-mitotic stage, in which they rarely divide and multiply. At this stage they execute programmed functions and only in exceptional situations (insults of very different origin) will they react and repair the local lesions, to the extent that the environment and their proper potential permits. It was thought that only four organs divert to this general role because they require a very high turnover of cells. These organs are: (a) the bone marrow, responsible for the continuous generation of blood and immune cells, (b) the gonads, for germinal cells, (c) the liver, with a high capacity for regeneration, and (d) the epithelia, mainly the intestinal and epidermis, in continuous regeneration. In these four organs tissue-specific cell populations exist with characteristics of stem cells that supply their renovation potential. Recently, it has been demonstrated that the brain also presents some capacity for the generation of new neurons. Previously it was assumed that the number of neurons and the brain structure was established during the first years of life, and that with the posterior development only restructuring and loss of connections would occur. Nowadays, however, it has been estimated that as much as about 10^5 cells/day could be produced in the brain. These cells must be generated from neural stem cells allocated in the specific brain areas.

9.2.1 Stem Cell Definition Criteria

Stem cells, as opposed to the majority of somatic cells, seem to be regulated by a conservative division mechanism (asymmetric mitosis) where a cell equivalent to the original is created, and one of these assumes the rest of the developmental/differentiation programs. This mechanism is called "self-renewal" due to the fact that it seems to be capable of strictly controling the stem cell population present in a specific organ.

Although this mechanism has still not been formally demonstrated in mammals, based on data obtained in model organisms such as *Drosophila melanogaster* and *Arabidopsis thaliana*, it is generally accepted that this would also occur in mammals. In addition, at least in the adult organism, not all stem cells present in a determined organ participate simultaneously in the functional process of main-

tenance and regeneration; only few contribute simultaneously at particular moments. The great majority of the stem cells appear to be in a particular physiological state known as "quiescence", which is interpreted as a state that protects the cells both from external insults, physical or chemical, and aging. When the actual contributing stem cells clones lose their proliferative potential they disappear and are progressively replaced by the progeny of other cells that activate; this phenomena is known as "clonal succession".

In summary, a stem cell is defined, functionally, as a cell capable of executing "self-renewal" and with the potential to generate (through the controlled differentiation of their descendent cells) several cell lineages (multipotential) or whole organisms (totipotent). In addition a stem cell, in its natural environment, presents the tendency to enter a quiescent state [4].

9.2.2
New Horizons in Stem Cell Biology

The past few years have provided strong evidence that the classic concepts of stem cells in the adult organism need revision. Originally it was thought that the adult stem cells were only implicated in the physiology of the highly demanding organs, as previously cited. Now we know that stem cells populations have been characterized in muscle and brain [5–8], and progenitors for endothelium in bone marrow [9].

Most importantly, the established dogma indicating that the stem cell populations would only be capable of contributing to the generation or regeneration of the lineage, tissue, or organ in which they are allocated, has been outmoded. By refinement of the analysis methods, it has been possible to demonstrate that stem cells isolated from bone marrow can be used for the repair of damage induced in heart, muscle, brain, liver, epithelia, and blood vessels [5, 9–16]. In a similar way, stem cell lines derived from central nervous system (CNS) and muscle (satellite cells) can also contribute to the regeneration of the lympho-hematopoietic system [6, 17, 18]. These solid evidences conform a new frame for a novel concept denominated "stem cell plasticity" cemented by the fact that a generic stem cell possesses the intrinsic capacity for executing a variety of genetic developmental programs, depending on signal (secreted factors, cell/cell contacts, etc.) provided by the tissue environment [4]. However, what we now know represents only the first scientific evidence indicating that a "trans-differentiation" process can occur or can be induced, and that we are still far from a realistic clinical application (Fig. 9.1).

9.2.3
Current and Future Applications of the Stem Cell Technology

The therapeutic potential of stem cell is huge. Only consider the large numbers of people who die while waiting for the transplant of a vital organ, knowing that if a compatible donor could be found they would encounter a new opportunity

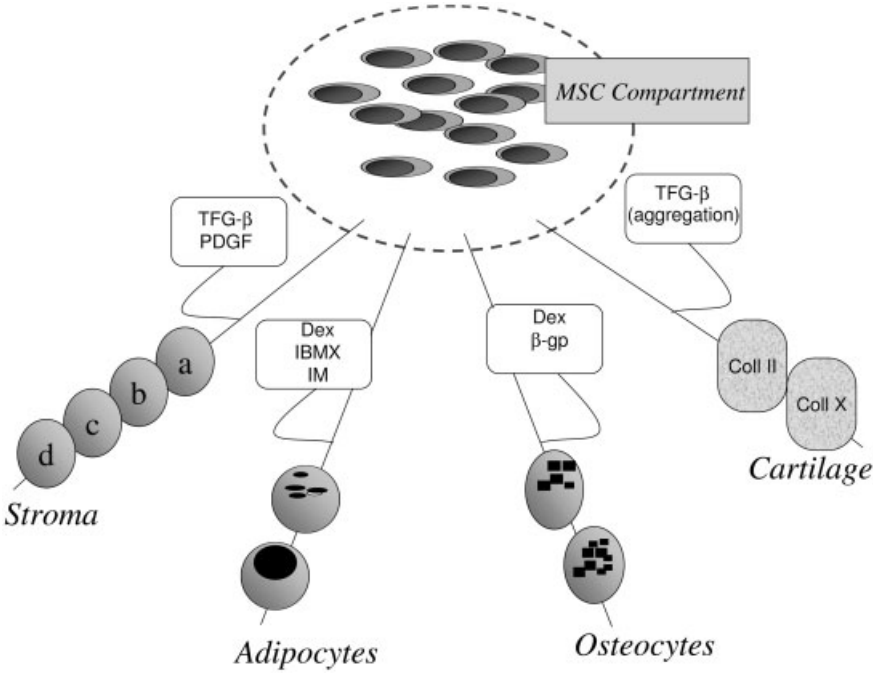


Figure 9.1 Simplified representation of mesenchymal stem cells pluripotency. Pluripotential stem cells can be differentiated to specific lineages by using appropriate cytokine cocktails, TFGβ and PDGF for stroma phenotype, Dex, IBMX, and IM for adipocyte, Dex and β-gp for osteocyte lineage and TGFβ for cartilage. Dex, Dexamethasone; β-gp, beta-glycerophosphate; IBMX, isobutyl-methyl-xantine; IM, indometacine.

with sufficient quality of life. In other cases, specific organs suffer damage due to traumatic, pathological processes or caused by unhealthy habits that could be only alleviated by complex clinical and/or surgical procedures. Stem cell technology starts to offer a new future in all these fields.

Taking as the first consideration that the majority of the results are still in the preclinical phase, the last few years have generated very promising and spectacular results. Using stem cells isolated from bone marrow it has been possible to improve considerably the pathologic condition of a mouse line suffering from cirrhosis [14]. In a similar fashion a mouse in which an infarct was surgically induced has been treated [12]. Nervous system repair is also giving promising results. Using olfactory glial cells it has been possible to regenerate and functionalize a spinal cord lesion in rats that provoked the permanent paralysis of the extremities [19]. This approach is currently being validated in primates as well.

9.2.4
Stemness and Stem Cell-associated Genetic Programs

The cellular intrinsic (genetic program) or extrinsic (environment-derived) mechanisms that govern the biology of stem cells are only just beginning to be uncovered [4]. The main problem that the clinician or the researcher encounters in the use of stem cells for therapy is that isolated from their natural niches, these scarce cell populations tend to differentiate, rapidly losing their characteristic stem cell properties, although there are notable exceptions to this. The real problem is the current "vague knowledge" about the genetic programs that control their regulation. Ultimately, when critical information is known about stem cell regulation, it will be possible to maintain and expand the stem cells *ex vivo* in optimal conditions for their further re-implantation in patients.

Recently human embryonic stem cell lines have been obtained (hES) [20, 21], although their culture conditions and expansion potential are not as well defined as in the mouse model. In spite of that, due to the wide experience in the manipulation of murine ES, just their existence has opened a big social, ethical, and scientific debate on the matters and topics that could be studied, and on the reform of the current laws to regulate the economic and general activity on this probable therapeutic potential. Issues concerning ethical considerations have been discussed in Chapter 1.

In this wide scenario, new cells have come continuously to the fore in recent years. Among them, the most established therapeutic strategy is based on the use of MSCs. These cells were first isolated and characterized from bone marrow and seem to be responsible for the creation and maintenance of the stromal compartment. Studies initiated in 1992 have demonstrated that MSCs can be expanded and manipulated *ex vivo*, presenting the capacity to differentiate into numerous cell types including adipocytes, myoblasts, osteoblasts, and chondrocytes [22, 23]. Therefore, the use of MSCs has opened a new front in the development of future strategies for tissue engineering [24]. A particular limitation of this strategy may lie in the scarce representation of these cells in the bone marrow (about 1 in 10^5 stromal cells are MSCs). New data suggest that this limitation could probably be bypassed using MSCs obtained from alternative richer sources. Hedrich and colleagues (University of California, LA, USA), have demonstrated that the adipose tissue discarded during liposuction procedures seems to be a rich source of MSC-like cells [25]. Although it has not been confirmed that the MSC-like cells derived from adipose tissue are in fact stem cells, the multilineage-differentiation results obtained clearly point in this direction. In the case of positive confirmation, it is clear that the high accessibility of the adipose tissue and the autologous use of these cells have opened a new avenue for the potential treatment of human illness by cell therapy.

Taking in consideration all the above, it is clear that the stem cell biotechnology field using adult stem cells will become an important therapeutic tool in future medicine. From this perspective it is extremely important that after the completion of the human genome era, an exhaustive study is developed for the under-

standing of the genetic interconnected programs that play critical roles in the maintenance and control of the stem cell signature. Recently, the first results in this direction have been published with the global aim of comparing the transcriptional expression profiles of distinct highly enriched embryonic and adult stem cell populations [26, 27]. The main conclusion indicates that a group of about 200 genes are associated with the "stemness" property (the molecular signature defined by stem cell behavior), although more genetic functions may be related to the specific stem cell populations when they are compared in pairs. Whilst these data are the first steps and are interesting, there is still little overall knowledge of the expression profiles of the genes and in particular the proteins. For the extension of current knowledge on stem cell biology, in order to define rational procedures for their *ex vivo* expansion and manipulation for therapeutic purposes, more thorough and comprehensive studies must be carried out, including the definition of the stem cell proteome and the modifications that occur in the commitment decisions.

9.3 Proteomics

Genomics projects have produced a large number of DNA sequences from a wide range of organisms, including humans and other mammals [28]. Even if a complete set of genes is known, the function of many of these new genes remains unclear; furthermore, the pivotal role of protein/protein interactions in living organisms is largely unknown [29].

Proteomics, one step beyond functional genomics approaches, together with the study of RNA expression levels, is associated with the analysis of global protein expression in cells, organisms, tissues, organelles, etc. [30, 31]. A deep quantitative study of protein expression must take into account the influence of varying conditions, as well as the occurrence of post-translational modifications, interactions with other molecules, etc. to gain a holistic view of expressed proteins, i.e. the proteome. According to this definition, the proteome is a dynamic entity, so usually only one of the possible snapshots of all possible cell proteomes is obtained in any single experiment.

To accomplish this work, proteomics used to involve a separation step, traditionally high-resolution two-dimensional gel electrophoresis (2-DE), followed by an identification step, mainly mass spectrometry [32]. Proteins isolated by 2-DE could be identified by in-gel trypsin digestion and peptide mass fingerprinting (PMF) using mass spectrometry (MS) (Fig. 9.2) or tandem mass spectrometry (MS/MS) [33].

New approaches enable protein identification in complex mixtures without the use of electrophoresis gels by separating tryptic peptides by two- or multidimensional liquid chromatography (MuD-LC) coupled to electrospray ionization (ESI)-MS/MS [34]. There is a concomitant expansion of databases that describe protein expression in organisms, tissues, cells, and organelles by using protein identifica-

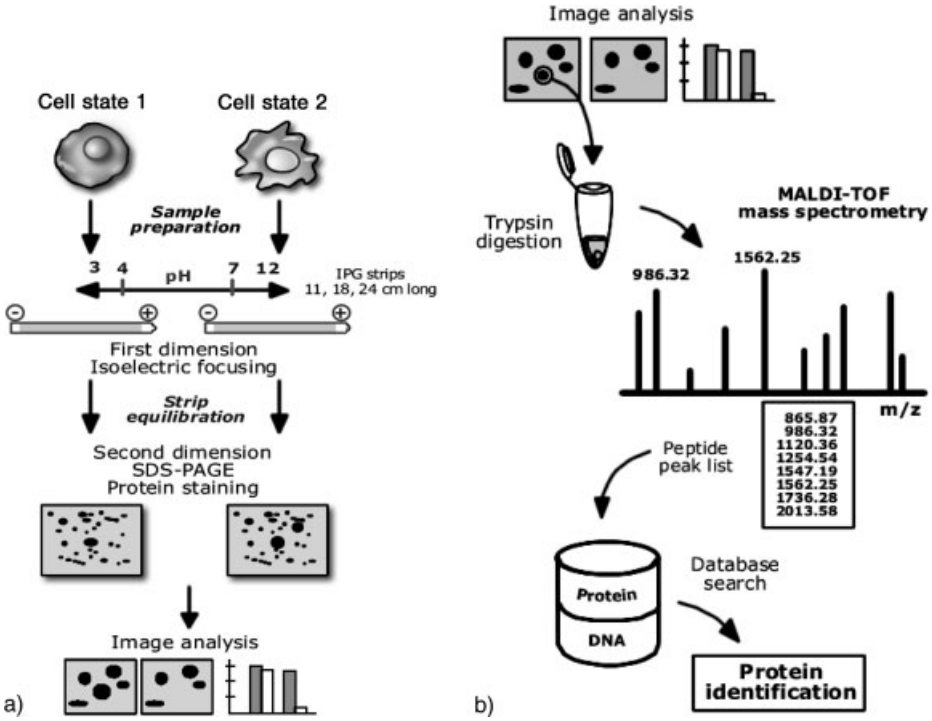


Figure 9.2 Schematic diagram of our proteomics strategy for protein identification. (a) Separation of proteins from cell extracts is carried out by 2-D PAGE. Protein extracts are first separated according to their isoelectric points using immobilized pH gradients (IPG strips), then strips are transferred to a second dimension and separated according to their size on SDS-PAGE. Proteins are visualized with a variety of staining protocols and the images are analyzed and compared with appropriate image analyzer programs. (b) Spots selected by image analysis are excised, digested, and mass analyzed by MALDI-TOF-MS for peptide mass fingerprinting. If sequenced genomes are available, this alone may be sufficient to identify proteins in complex mixtures. When necessary, additional methods such as Edman degradation, PSD-MALDI or electrospray mass spectrometry can be used to generate peptide sequence tags for searching in protein and DNA databases.

tion data from 2-DE and/or MuD-LC-MS. In these databases, expression profiles are normally described under different experimental conditions or disease stages, thus enabling the identification of up- or down-regulated proteins or modified protein species that can be used as diagnostic, prognostic, or therapeutic targets. Furthermore, the study of protein interactions networks in the complex cell environment will improve our understanding of cell biology and the characterization of disease progression. The critical importance and role of databases has been discussed by Yip and colleagues in Chapter 21.

9.3.1 Differential Display via Two-dimensional Gel Electrophoresis

Traditionally, proteomics projects make a combined use of protein separation by 2-DE and mass spectrometric protein identification. Complex mixtures of proteins, such as proteomes expressed by organisms or eukaryotic cells sum up thousands to tens of thousands of proteins at any given time, and therefore the identification of single protein species needs to decrease this complexity. One-dimensional gel electrophoresis, SDS-PAGE, can only separate a low number of protein fractions according to their molecular weight. Because only around 100 proteins are visualized by this technique, additional separation steps are necessary in proteomic studies. 2-DE is an orthogonal separation method: the first separation is based on the protein isoelectric point (pI) using isoelectric focusing (IEF) under gradient pH strips, whereas the second separation is based on molecular weight in SDS-PAGE gels. After separation, proteins are visualized by silver or fluorescent staining techniques. With 2-DE, hundreds of proteins can be separated into a single, reproducible gel.

Current separation by IEF is done on immobilized pH gradient strips (IPG). In these strips (Immobilines^{TM}), buffering acrylamide derivatives that contain free carboxylic acids or tertiary amino groups are copolymerized with acrylamide to yield a stable, reproducible pH gradient. Using the appropriate combination of immobilines, it is possible to generate any pH gradient, although narrow basic pH ranges are difficult to obtain due to polyacrylamide instability at high pH. Despite the resolving power of 2-DE gels, this technique has some limitations, namely low detection sensitivity and linearity, poor solubility of membrane proteins in cell lysis buffer, and limited loading capacity of gradient pH strips. Proteins are expressed in cells over a wide range of concentrations, probably exceeding 8–10 logs of difference [35], and the linearity of silver staining is limited, thus preventing detection of low concentration proteins when mixed with high concentration proteins. This drawback is lessened with fluorescent stains, with similar sensitivity (about 2 ng per spot) but higher linearity. Staining with fluorescent dyes such as Sypro Ruby, that binds non-covalently SDS molecules attached to proteins, is accomplished in a few reproducible steps with little protein variability. However, protein spots in gels represent only the most abundant proteins even with high sample loads [35]. The amount is limited by the loading capacity of strips (typically below 2 mg of proteins with narrow pH strips) due to complications during the focusing process. An additional limitation of 2-DE is the low number of membrane proteins it can resolve. Most of them are probably low-copy proteins that remain undetected by traditional staining methods, but most problems with membrane proteins arise from their poor solubility in the aqueous lysis buffer compatible with IEF.

Sample preparation for IEF is a key point in 2-DE-based proteomics projects. Proteins from cells, tissues, etc. can be solubilized, denatured, and reduced. The lysis buffer breaks intra- and intermolecular interactions of proteins, yielding soluble proteins during the process, provided that the lysis buffer is compatible with

focusing requirements. Only clean, unmodified protein samples yield high-resolution 2-DE maps. Samples must be free of DNA, lipids, salts, or ionic detergents which disturb the separation process. Furthermore, the sample preparation protocol should render unmodified proteins, preventing artifactual modifications of proteins and non-specific ruptures by inactivating proteases and phosphatases. A lysis buffer should contain the chaotropic agents such as urea, that disturbs the hydrogen bond network in water breaking the weak, non-covalent forces that maintain protein structure and solubilizing most of the proteins.

Other necessary reagents are surfactants, currently non-ionic or zwitterionic detergents, that remove lipids and solubilize some additional membrane proteins, as well as reducing reagents such as DTT to break disulfide bonds. DTT molecules may migrate out of the pH gradient, lowering the reducing power in the strip, which results in the loss of some proteins when their disulfide bonds are re-arranged. Replacing the thiol reagent DTT with non-charged reducing agents such as tributyl phosphine (TBP) enhances the solubility of specific proteins. However, despite these advances in sample preparation transmembrane proteins are never to rarely seen. Perhaps new surfactants or detergents to be developed and the use of chaotropes other than urea, such as mixtures of urea and thiourea, will raise protein solubility.

Contaminating compounds, usually DNA and lipids, render samples viscous, reducing the entry of proteins into IPG strips and lowering the resolution of protein spots in gels. To clean samples, proteins are precipitated with acetone, TCA, or TCA/acetone, or some commercial cleaning kit. If only desalting is required, dialyzing against lysis buffer can be appropriate. This cleaning process is essential when working with basic or narrow pH gradients, whose loading capacity is higher.

Because of the complications arising from the relative abundance of proteins and the low number of membrane proteins in gels, a further protein fractionation step is required to reduce complexity prior to IEF. This prefractionation can be achieved under different approaches: (1) differential solubility, applying sequential extraction procedures with increased solubility power; (2) separation of complex mixtures by established procedures like liquid chromatography, free flow electrophoresis or the recently described solid IEF method; (3) the use of narrow, overlapping pH-gradient strips, increasing the amount of protein loaded; and (4) the isolation of organelles, cell compartments, or membrane fractions (for review see [35]).

The isolation of organelles reduces complexity, allowing higher loading and visualization of low-copy-number proteins, and allowing the assessment of the location of proteins in the cell. There are numerous protocols aimed at fractionating subcellular compartments on the basis of differential centrifugation of cell extracts. Commercial kits are increasingly being employed to enrich protein samples in particular fractions, but most of these methods still show high cross-contamination levels with proteins from other fractions; it is for this reason that they are more suitable for protein enrichment rather than for organelle isolation.

The combined use of several prefractionation methods increases the number of proteins visualized in gels to map a proteome, but large numbers of gels are re-

quired and, consequently, a large initial number of cells, thus currently preventing the wide application of these approaches to minute precious samples.

9.3.2 Protein Identification

Following protein separation, stained protein spots are in-gel digested with a highly specific proteolytic enzyme, usually trypsin. Upon digestion, tryptic peptides are extracted from the gel matrix and the peptide masses are measured by matrix-assisted laser desorption ionization/time-of-flight (MALDI-TOF)-MS. In a process known as PMF, experimental masses are matched against theoretical masses from *in-silico* digested proteins using sophisticated algorithms, which results in a list of candidate proteins that meet the matching criteria (for more information on protein identification methods, see [33]).

Because of the limitations arising from classical proteomics approaches (2-DE followed by MS), other approaches such as MuD-LC or multidimensional protein identification technology (MuDPIT), on-line with ESI-MS/MS have gained popularity [34, 36]. In this approach complex mixtures of proteins are digested in solution, and the resulting peptide mixture is fractionated in a strong cation-exchange column by increasing salt concentration in a stepwise manner. The peptides eluted in each fraction are separated again in a reversed-phase capillary chromatography, and analyzed in a data-dependent manner by MS/MS. This approach does not share the same limitations as 2-DE for dynamic range of analysis or the detection of low-concentration proteins. Furthermore it is capable of identifying proteins with extreme pI and/or molecular weight not seen on 2-DE gels. The main limitation though is the inability to perform differential display analysis as only proteins are identified and relative quantitative analysis is not possible. Promising new approaches (see below) lie in using combined isotope-coded affinity tags [37] and MuDPIT for proteome-wide screening, but this has not yet been applied to MSC-related research.

9.3.3 Differential Proteomics

Differential proteomics, the comparison of different proteomes (e.g. normal versus diseased cells, diseased cells versus treated cells, etc.) is of extreme importance. Several approaches now exist and typically involve either electrophoresis or chromatography combined with chemical labeling. Electrophoresis is a relatively simple visual method for mapping differences in protein expression. Certain limitations exist and have been discussed above (low dynamic range, no low-copy proteins, no proteins with extreme pI, and/or mass), but others exist as well such as the reproducibility of gels and low linear range of visualization (staining) procedures. 2-DE also has a relatively low throughput due to the fact that current software for 2-D gel analysis requires time-consuming manual operation verification. However it remains the quickest method currently for directly targeting differences in protein expression.

Some drawbacks mentioned above can be circumvented by using the differential in-gel electrophoresis (DIGE) system developed by Amersham Biosciences (http://www.amershambiosciences.com). With this chemistry cellular protein levels can be compared by covalently labeling cell extracts with one of the three fluorescent dyes (Cy2, Cy3, or Cy5). This is performed prior to IEF, and the labeled mixture is run in a single gel that is visualized at the three wavelengths with the appropriate scanning equipment. Spot patterns can be automatically compared with software programs such as DeCyder (Amersham Biosciences), and the results are a table of statistically relevant differences.

An alternative method for differential analysis is to use isotopically coded affinity tags (ICAT reagents, http://www.appliedbiosystems.com). The workflow consists of two sections: (1) proteins are covalently tagged with ICAT reagents followed by proteolysis of the combined labeled protein samples; isolation, identification, and quantification of the tagged peptides by MuD-LC-MS/MS; and (2) computational analysis of the data is performed. Sample complexity is diminished by affinity chromatography on avidin columns, which selectively retains peptides bonded to ICAT molecules.

9.3.4
Protein Profiling

Surface-enhanced laser desorption ionization (SELDI)-TOF-MS [38] enables the analysis of complex protein mixtures separated by on-chip retentate chromatography. This method constitutes a rapid, reproducible, and robust analytical tool that enables the comparative analysis of protein expression profiles in the low fmol range. The SELDI process has been elegantly explained in Chapter 16.

9.4
Proteomic Analysis of MSCs

Proteomic strategies have been implemented for a wide range of cancer studies, as mentioned in other chapters (see, for example, Chapter 5 on renal cancer, Chapter 6 on HSP27 in cancer, Chapter 7 on colon cancer, Chapter 8 on ovarian cancer, and Chapter 10 on lymphoblastoid cells and lymphomas). It has also been widely reported in the literature for: liver carcinoma [39], fibrosarcoma [40], bladder cancer [41, 42], breast cancer [43], lung cancer [44], kidney cancer [45], and ovary cancer [46]. While several tumor markers have been reported [47], only a few have been useful in clinical diagnosis and in the prognosis of survival and recurrence [48]. Hence the characterization and description of new markers is still of importance. Such markers will be applied to address a variety of new biological questions such as: pathogenicity and antibiotic resistance determinants in microorganisms [49, 50]; disorders and degeneration of the nervous system [51, 52]; and organ-specific human pathologies [53].

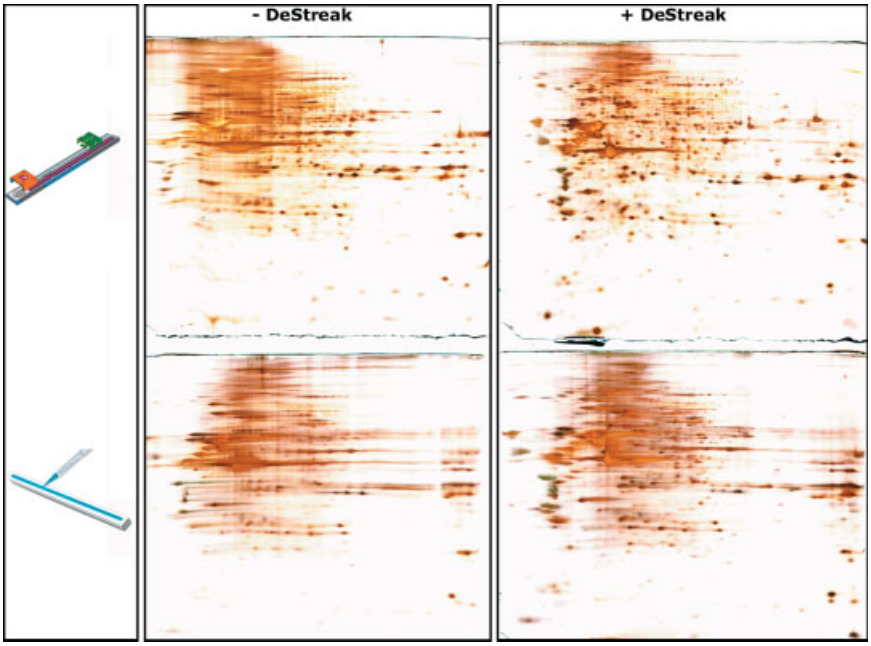


Figure 9.3 Effects of DeStreak reagent treatment and anodic cup loading sample application of cell total lysates. Cell lysates from human adult mesenchymal stem cells (hMSC) corresponding to 40 µg protein were loaded on each 11 cm IPG strip. Protein extracts were cleared by centrifugation at 14 000 rpm and cleared supernatants, either treated or non-treated with the DeStreak reagent, were loaded for IEF by cup-loading or by in-gel rehydration. 2-D PAGE was performed using precast immobilized 3–10 pH gradient (IPG) strips, 11 cm length. The positive effect in terms of spot protein localization and resolution is clearly shown by using DeStreak reagent and anodic cup-loading.

To date very little work focused on stem cell proteomics has been published. In one study by Colter et al. [54] surface epitopes and other proteins were potentially associated to distinguish the small (spindle-shaped) from the large (flat) cells. The results suggested the relevance of distinguishing the major subpopulations of marrow stromal cells in defining their biology and their potential for cell and gene therapy. A deep proteomic approach to the study of these issues will provide a more comprehensive view of the underlying processes.

Work carried out in our laboratory has focused on a 2-DE-based approach aimed at analyzing the differential protein pattern of mesenchymal stem cells extracts. An overview of this process can be viewed in Figure 9.2. Basically, MSC cultures are taken from human adult MSCs derived from human bone marrow (supplied by Cambrex, NJ, USA). Cells are grown undifferentiated in Cambrex MSCGM medium. From these cells protein extraction is performed followed by 2-DE using a range pH gradients. We have noticed that for the first dimension (IEF with IPG strips) the DeStreak solution (Amersham Biosciences) improves spot patterning in the basic pH range (Fig. 9.3). Its reaction with cysteine residues is

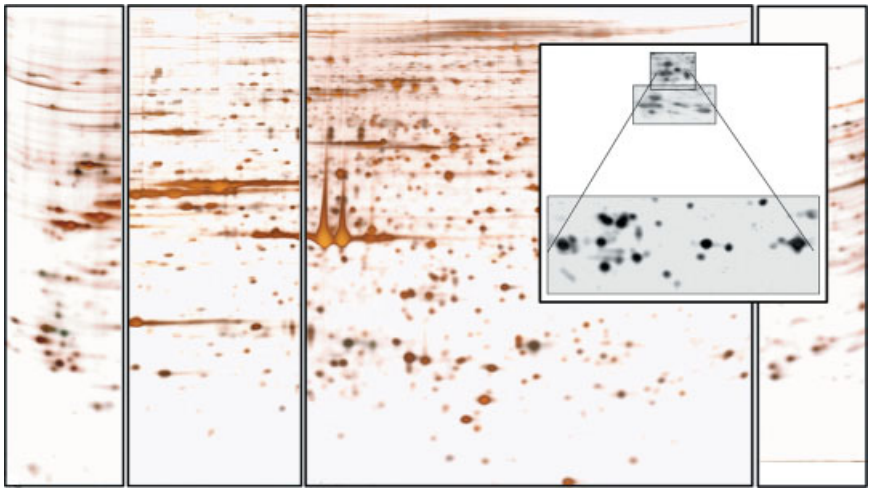


Figure 9.4 Composition of 2-D silver-stained gels of cell proteins from hMSC extracts. First dimension, IEF with IPG strips pH 4–5, 4.5–5.5, 5–6, and 4–7, basic end on the right. Second dimension, SDS-PAGE with 10% acrylamide. Overlapping pH regions are omitted. Selected regions are amplified below for comparison. Insert: Three similar regions of silver-stained 2-D PAGE images that illustrate the increased resolution when using narrower pH gradients (3–10, 4–7, and 5–6) and larger formats (11–18 cm) in the first dimension.

reversible and has no effect on the subsequent protein identification by MS. For protein staining we rely on silvernitrate, which is compatible with MS and allows the visualization of up to thousands of proteins. An example of an MSC spot pattern in the pH 4–7 range and improved resolution with narrow pH range is displayed in Figure 9.4.

For protein identification, spots from the silver-stained 2-DE gels are excised using a scalpel and in-gel enzymatically digested, usually with trypsin. MALDI-TOF analysis of the extracted peptides yields a PMF of the protein (Fig. 9.5), which is matched against all sequence entries in a known database. We can currently identify about 85% of silver-stained protein spots by PMF. Interestingly, no relation has been found between silver staining intensity, protein molecular mass (except for low mass proteins) or pI with the sequence coverage or the percentage of proteins identified.

Another approach in our laboratory includes the use of the ProteinScape software program (Bruker Daltonics) which is a bioinformatics software package for proteomics data warehousing, data analysis, and protein identification. Briefly, data from different origins are linked within the ProteinScape database. Linked information can relate to most types of data generated in the standard workflow, such as: spots detected in 2-D gel images; spots manually picked from gels and automatically digested; spots manually applied to MALDI targets; and finally, MALDI spectra taken automatically or manually. The ProteinScape program performs protein identification through the Mascot search engine which interrogates

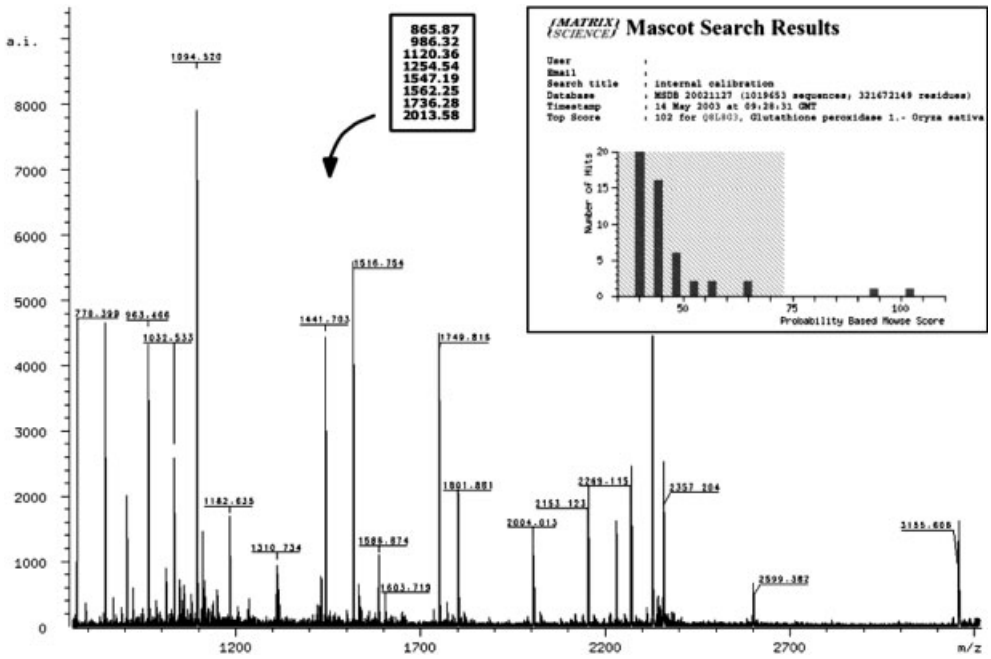


Figure 9.5 Peptide mass fingerprinting of an hMSC protein spot isolated by 2-D PAGE and in-gel digested with trypsin. To achieve the highest mass accuracy, the equipment is first externally calibrated employing protonated mass signals from a peptide mixture covering the 1000–3200 *m*/*z* range and thereafter every spectrum is internally calibrated using selected signals arising from trypsin autoproteolysis to reach a typical mass measurement accuracy of ±30 ppm. The measured tryptic peptide masses are transferred through MS BioTools (Bruker Daltonics) program as inputs to search the NCBI nr database using the Mascot program (Matrix Science, London, UK).

databases according to user-defined parameters. In addition it can evaluate automatically the search results based on preset confidence levels. The software has additional tools for advanced data analysis, such as recalibration of peaks, or comparison of the theoretical and experimental pI and molecular weight values for every protein (Fig. 9.6).

The use of software and existing and new biochemical approaches will expand our horizons in this important field of research. The data obtained from the preliminary global study of proteins expressed by human MSCs, in combination with the alternative differential expression profiling analyses will allow us to define with more certainty the molecular circuits involved in the maintenance of stem cell properties and the critical proteins modulated during the lineage-commitment processes.

Figure 9.6 Identified hMSC proteins (see labels on the 2-D PAGE image). ProteinScape (Bruker Daltonics) software allowed us to link spot position on the gel with protein identification results from different databanks, as well as to track the whole process (separation on 2-D PAGE, spot excision, spot digestion, MS acquisition, DB search, and protein identification).

9.5 Acknowledgements

We are very grateful to L. Emilio Camafeita for critical review of the manuscript. The Department of Immunology and Oncology was founded and it is partially supported by the Spanish Research Council (CSIC) and Pharmacia Corporation.

9.6 References

1 Martin, G.R. *Proc. Natl Acad. Sci. USA* **1981**, *78*, 7634–7638.
2 Evans, M.J., Kaufman, M. *Nature* **1981**, *292*, 151–156.
3 Bongso, A., Fong, C.Y., Ng, S.C., Ratnam, S. *Hum. Reprod.* **1994**, *9*, 2110–2117.
4 Zon, L.I. *Nature Immunol.* **2001**, *2*, 142–143.
5 Ferrari, G., Cusella-De Angelis, G., Coletta, M. et al. *Science* **1998**, *279*, 1528–1530.
6 Bjornson, C.R., Rietze, R.L., Reynolds, B.A., Magli, M.C., Vescovi, A.L. *Science* **1999**, *283*, 534–537.
7 Doetsch, F., Garcia-Verdugo, J.M., Alvarez-Buylla, A. *Proc. Natl Acad. Sci. USA* **1999**, *96*, 11619–11624.
8 Uchida, N., Buck, D.W., He, D. et al. *Proc. Natl Acad. Sci. USA* **2000**, *97*, 14720–14725.
9 Takahashi, T., Kalka, C., Masuda, H. et al. *Nature Med.* **1999**, *5*, 434–438.
10 Asahara, T., Murohara, T., Sullivan, A. et al. *Science* **1997**, *275*, 964–967.
11 Gómez-Navarro, J., Contreras, J.L., Arafat, W. et al. *Gene Ther.* **2000**, *7*, 43–52.
12 Orlic, D., Kajstura, J., Chimenti, S. et al. *Nature* **2001**, *410*, 701–705.
13 Petersen, B.E., Bowen, W.C., Patrene, K.D. et al. *Science* **1999**, *284*, 1168–1170.
14 Lagasse, E., Connors, H., Al-Dhalimy, M. et al. *Nature Med.* **2000**, *6*, 1229–1234.
15 Gussoni, E., Soneoka, Y., Strickland, C.D. et al. *Nature* **1999**, *401*, 390–394.
16 Brazelton, T.R., Rossi, F.M., Keshet, G.I., Blau, H.M. *Science* **2000**, *290*, 1775–1779.
17 Mezey, E., Chandross, K.J., Harta, G., Maki, R.A., McKercher, S.R. *Science* **2000**, *290*, 1779–1782.
18 Jackson, K.A., Mi, T., Goodell, M.A. *Proc. Natl Acad. Sci. USA* **1999**, *96*, 14482–14486.
19 Ramon-Cueto, A., Cordero, M.I., Santos-Benito, F.F., Avila, J. *Neuron* **2000**, *25*, 425–435.
20 Thomson, J.A., Itskavitz-Eldor, J., Shapiro, S.S. et al. *Science* **1998**, *282*, 1145–1147.
21 Reubinoff, B.E., Pera, M.F., Fong, C.-Y., Trounson, A., Bongso, A. *Nature Biotechnol.* **2000**, *18*, 399–404.
22 Pittenger, M.F., Mackay, A.M., Beck, S.C. et al. *Science* **1999**, *284*, 143–147.
23 Huss, R., Lange, C., Weissinger, E.M., Kolb, H.J., Thalmeier, K. *Stem Cells* **2000**, *18*, 252–260.
24 Deng, W., Obrocka, M., Fischer, I., Prockop, D.J. *Biochem. Biophys. Res. Commun.* **2001**, *282*, 148–152.
25 Zuk, P.A., Zhu, M., Mizuno, H. et al. *Tissue Eng.* **2001**, *7*, 211–228.
26 Ramalho-Santos, M., Yoon, S., Matsuzaki, Y., Mulligan, R.C., Melton, S. *Science* **2002**, *298*, 597–600.
27 Ivanova, N.B., Dimos, J.T., Schaniel, C., Hackney, J.A., Moore, K.A., Lemischka, I.R. *Science* **2002**, *298*, 601–604.
28 Venter, J.C., Adams, M.D., Myers, E.W. et al. *Science* **2001**, *291*, 1304–1351.
29 Pawson, T., Nash, P. *Genes Dev.* **2000**, *14*, 1027–1047.
30 Anderson, N.G., Anderson, N.L. *Electrophoresis* **1996**, *17*, 443–453.
31 Pandey, A., Mann, M. *Nature* **2000**, *405*, 837–846.
32 Andersen, J.S., Mann, M. *FEBS Lett.* **2000**, *480*, 25–31.
33 Corthals, G.L., Gygi, S., Aebersold, R., Patterson, S.P. In *Proteome Research, 2D Gel Electrophoresis and Detection Methods* (Rabilloud, T. Ed.). Springer, Heidelberg, **1999**, pp 197–231.
34 Link, A.J., Eng, J., Schieltz, D.M. et al. *Nature Biotechnol.* **1999**, *17*, 676–682.
35 Corthals, G.L., Wasinger, V.C., Hochstrasser, D.F., Sanchez, J.-C. *Electrophoresis* **2000**, *21*, 1104–1115.
36 Washburn, M.P., Wolters, D., Yates, J.R. III. *Nature Biotechnol.* **2001**, *19*, 242–247.
37 Gygi, S.P., Rist, B., Gerber, S.A., Turecek, F., Gelb, M.H., Aebersold, R. *Nature Biotechnol.* **1999**, *17*, 994–999.

38 Merchant, M., Weinberger, R. S. *Electrophoresis* **2000**, *21*, 1164–1177.

39 Jungblut, P. R., Zimny-Arndt, U., Zeindl-Eberhardt, E. et al. *Electrophoresis* **1999**, *20*, 2100–2110.

40 Sinha, P., Hutter, G., Kottgen, E. et al. *Electrophoresis* **1999**, *20*, 2961–2969.

41 Celis, J. E., Ostergaard, M., Basse, B. et al. *Cancer Res.* **1996**, *56*, 4782–4790.

42 Celis, J. E., Ostergaard, M., Rasmussen, H. H. et al. *Electrophoresis* **1999**, *20*, 300–309.

43 Giometti, C. S., Williams, K., Tollaksen, S. L. *Electrophoresis* **1997**, *18*, 573–581.

44 Okuzawa, K., Franzen, B., Lindholm, J. et al. *Electrophoresis* **1994**, *15*, 382–390.

45 Sarto, C., Marocchi, A., Sanchez, J.-C. et al. *Electrophoresis* **1997**, *18*, 599–604.

46 Lawson, S. R., Latter, G., Miller, D. S. et al. *Gynecol. Oncol.* **1991**, *41*, 22–27.

47 Carney, W. *Immunol. Today* **1988**, *9*, 363–364.

48 Robbins, J. *J. Endocrinol. Invest.* **1995**, *18*, 159–160.

49 Niimi, M., Cannon, R. D., Monk, B. C. *Electrophoresis* **1999**, *11*, 2299–2308.

50 Cash, P., Argo, E., Ford, L. et al. *Electrophoresis* **1999**, *11*, 2259–2268.

51 Johnston-Wilson, N. L., Bouton, C. M., Pevsner, J. et al. *Int. J. Neuropsychopharmacol.* **2001**, *1*, 83–92.

52 Rohlff, C. *Electrophoresis* **2000**, *6*, 1227–1234.

53 Cutler, P., Bell, D. J., Birrell, H. C. et al. *Electrophoresis* **1999**, *18*, 3647–3658.

54 Colter, D. C., Sekiya, I., Prockop, D. J. *Proc. Natl Acad. Sci. USA* **2001**, *98*, 7841–7845.

10
Lymphoblastoid and Lymphoma Cells

Raymonde Joubert-Caron, Didier Lutomski, and Michel Caron

10.1
Introduction

Basic science research in biomedicine for the past 50 years has been determining the functions of gene products using classical approaches that typically involve studying one or a few genes at a time. In response to the availability of complete genome sequence of various organisms including a draft of the human genome, the challenging task now is to decipher the molecular organization of cellular networks. This requires documentation on proteins, which are the functional units of cellular molecular machinery. Over the past decade the field of proteomics has emerged with the goals of developing and applying methodologies that accelerate the functional analysis of proteins. Biomedical research is in the center stage of the application field of proteomics. Disruption of the physiological balance between cell proliferation and death is a universal feature of all cancers. A major challenge in cancer therapy is the identification of drugs that kill tumor cells while preserving normal cells. Proteomics allows the direct determination of the effects of drugs at the molecular level as well as the side effects drugs have on a biological system. Moreover, proteomics allows a better knowledge of cell-signaling pathways mediated primarily by changes of protein phosphorylation states detectable on two-dimensional gel electrophoresis (2-DE). The study of phosphorylation is very important for better knowledge of receptor/ligand interactions during processes such as apoptosis.

Our approach for understanding variations in lymphoblastoid and lymphoma cells at the molecular level following growth modulation, consisted of investigations on protein modifications related to cell treatment either with a demethylating drug 5′-azacytidine (AZC), which has been used clinically for treatment of patients with malignant hemopathies [1, 2], or with galectin 1 (Gal1), a molecule that is a master regulator of immune cell homeostasis via apoptosis induction [3]. Various appropriate cell models were selected to perform these studies. Finally, the obtained information was stored in a 2-DE database [4], enhancing the efficiency of proteome research on lymphoblastoid and lymphoma cells.

Biomedical Application of Proteomics
Edited by J.-C. Sanchez, G. L. Corthals, D. F. Hochstrasser

ISBN: 3-527-30807-5

10.2 Experimental Models

In recent years, we have attempted to identify and develop a database of lymphoid proteins detectable by 2-DE as a prerequisite for analysis of drug effects and lymphocyte cell diseases [4, 5]. Human lymphoid B and T cell lines are widely used models for biomedical applications in immunology and hematology [6], and in the experimental studies presented in this chapter different cell lines derived from Burkitt's lymphomas (BL) were used.

BL is a tumor derived from proliferating centroblasts of the germinal center. All BLs carry reciprocal chromosomal translocations that activate the c-*myc* oncogene through juxtaposition to one of the immunoglobulin loci [7]. Many BL tumors carry point mutation in the *p53* tumor suppressor gene [8] or other defects in the p14ARF-MDM2-p53 pathway, and inactivation of the *p16INK4a* gene by promoter methylation or homozygous deletion. This indicates that disruption of both the pRb and p53 tumor suppressor pathways is critical for BL development [9]. Alterations of other genes, including *Bax*, *p73*, and *BCL-6*, may provide further growth stimulation and apoptosis protection. BLs are characterized by an enhanced proliferative activity, as a result of the deregulation of oncogenes with cell cycle regulatory functions [10], such as c-*myc*. BLs are able to accumulate other alterations in cell cycle regulation, most frequently involving tumor suppressor genes such as p16 (*INK4a*) and p27 (*KIP1*). As a consequence, these tumors behave as highly aggressive lymphomas.

Thus, BL development involves multiple genetic and epigenetic changes that drive cell cycle progression and avert cell death by apoptosis [9]. For these reasons BL cell lines constitute a suitable model for our investigations. Two BL cell lines in particular were used for these studies: DG 75 and BL 36. DG 75 is a human Epstein-Barr virus (EBV)-negative BL cell line. It was established from the pleural effusion of a 10-year-old boy with BL in 1975 [11]. BL 36 is a human EBV-infected BL cell line established by Lenoir (CIRC, Lyon, France).

The data concerning AZC effects were obtained on DG 75 cells that responded well to AZC treatment in terms of viability and growth. BL 36 cell lines was used to evidence the cascade of cellular events related to the binding of Gal1 to its receptor on the cell membrane and inducing cell apoptosis.

10.2.1 Experimental Procedures

For the cell culture conditions for the experiments discussed in this chapter we have used the following experimental outline. Lymphoid cell lines were treated with AZC or Gal1 as described previously [5, 12]. Protein extracts from cells were performed on soluble proteins extracted by saline buffer and on proteins soluble in the presence of detergent.

Soluble proteins were extracted as previously described [13]. The proteins from both these solutions were then analyzed by 2-DE. Analytical gels were stained

with silver nitrate [14], and the preparative gels with either Colloidal Coomassie [15] or silver nitrate.

We have used an affinity-based purification of Gal1 receptor as it constitutes a generic approach for the identification of cell receptors. It was adapted from a method described in [16], modified according to [12]. Briefly, recombinant Gal1 is immobilized on agarose beads according to [17]. A cell membrane pellet from BL 36 cells is solubilized and incubated with 1 mg of lysate (with 200 µL agarose-Gal1). The resulting complex is then resuspended and incubated with NHS-biotin. The biotinylated proteins are finally eluted by incubation of the complex with a Tris buffer containing NaCl and lactose (a specific inhibitor of Gal1) [18].

Affinity capture is currently used in our laboratory for the study of either a single protein [12] or a family of proteins (subproteomics, e.g. phosphoproteomics) [19]. With this approach antibodies are generally coupled to UltraLink Biosupport Medium (e.g. 10 µL of antibody/10 mg of support, for the affinity capture of Lyn described in this chapter), as described by the manufacturer (Perbio Science France, Bezons, France). The interaction with the protein to be captured is performed in solution with orthovanadate for 2 hours at room temperature followed by an overnight incubation at 4 °C. Elution from the affinity support of the immuno-captured proteins is obtained with the isoelectric focusing (IEF) buffer used for the rehydration of the immobilized pH gradients (IPG) strips (except DTT). The reducing agent is finally added to the eluted proteins.

All proteins have been analyzed by mass spectrometry using either matrix-assisted laser desorption ionization/time-of-flight (MALDI-TOF) or microcapillary liquid chromatography coupled with electrospray ionization/tandem mass spectrometry (µLC/ESI-MS/MS). For MALDI-TOF-MS, analysis was performed as described elsewhere [13]. The peptide mass profiles produced by MALDI-MS were analyzed using PeptIdent (http://www.expasy.org/tools/peptident.html), Profound (http://prowl.rockefeller.edu/cgi-bin/ProFound), or Mascot (http://www.matrixsciences.com/). Peptide masses were compared with the theoretical masses derived from the sequence contained in the Swiss-Prot/TrEMBL database. For µLC/ESI-MS/MS, analysis was carried out using an Ultimate capillary LC system (LC Packings, The Netherlands) coupled to a quadrupole-TOF (Q-TOF) mass spectrometer (Waters, UK) equipped with a nano-electrospray source. The tryptic peptides were concentrated and desalted on a 75 µm id/150 mm length C18 PepMap column (LC Packings, The Netherlands). The eluted peptide mixture was analyzed by automated MS/MS. The database search was performed with the Mascot search tool available on the Matrix Science site (http://www.matrixsciences.com/) screening Swiss-Prot and NCBI nr.

10.3
Studies of Protein-pattern Changes Following Treatment with AZC

There is substantial evidence that DNA methylation plays an important role in silencing specific genes during development and cell differentiation [20, 21]. Hyper-

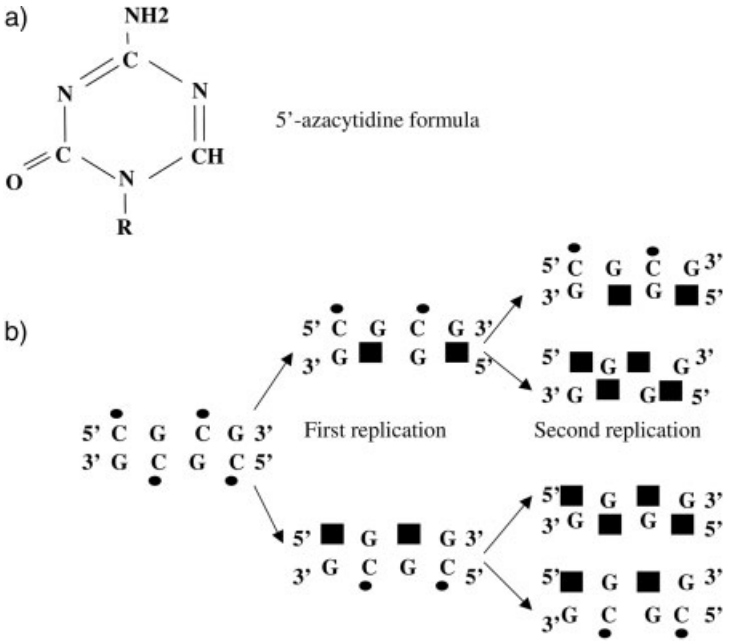


Figure 10.1 DNA modification of 5′-azacytidine (AZC). AZC is a cytosine analogue (a), which incorporates into DNA (b). One generation in the presence of the drug is sufficient to cause much of the DNA to become hemimethylated, whereas double-strand demodification can be observed after a second replication.

methylation is one of various mechanisms for repression of tumor suppressor gene in cancer [22]. Silencing of one of the most important cell cycle regulatory proteins, p16 (INK4a), by methylation of the CpG islands in the promoter region has been found to be a common event in tumors [23]. Protein p16 suppresses S-phase entry by antagonizing the cyclin-dependent kinases CDK4 and CDK6 [21]. The increased DNA-methyltransferase (DNA-Mtase) activity seen in multiple cancers has prompted targeting of the inhibition of this enzyme as an anticancer strategy. In this context, the DNA-Mtase inhibitors, AZC, and 5-azadeoxycytidine drugs have been used clinically for treatment of patients with malignant hemopathies [1, 2]. A major problem, as noted, is that these drugs have effects in addition to inducing demethylation and have many clinical side effects.

We have determined the changes in protein expression following a treatment of DG 75 cells with AZC. Treatment with AZC was conducted as follows: on day 1, 10 μM of AZC were added to 5 mL of complete medium. The cells were then fed for 7 days with complete medium with or without 10 μM AZC. On indicated days, cells were harvested and counted. The effects of the treatment in terms of cell viability and growth have been described elsewhere [24]. Application of proteomics to our purpose conveniently simplifies the identification of modified proteins by

comparing the protein expression levels in untreated and treated cells. Figure 10.1 shows the principle of action of AZC on DNA.

10.3.1 Soluble Protein-pattern Changes Induced by AZC Treatment

Subtractive analysis using Melanie 3.7 software (http://www.genebio.com) on a set of six gels for each condition was used to compare different protein profiles. Experimental variations such as stain intensities were determined via "scatter analysis" and further software-based spot normalization. Matching a set of gels to a reference gel automatically creates groups of spots. To facilitate the comparison of protein abundance two classes were defined, namely class A for control samples and class B for AZC-treated ones.

Based on our statistical tests, AZC treatment altered the abundance of 36 proteins from cytosol and mitochondria. Mass spectrometry allowed identification of 23 proteins (Table 10.1). Caspase 10 D, IF2A, tubulin-folding cofactor B, BET3, malate dehydrogenase, lactate dehydrogenase H-chain, two isoforms of transaldolase, and GARS protein were up-regulated. GARS was only known as a gene until now, and its function remains unknown. The cellular response to AZC treatment included key enzymes regulating energy metabolism. Transaldolase is an enzyme that catalyzes the reversible transfer of a three-carbon ketol unit from sedoheptulose 7-phosphate to glyceraldehyde 3-phosphate to form erythrose 4-phosphate and fructose 6-phosphate. Malate dehydrogenase is an oxidoreductase. It adds NAD to *S*-malate to product oxaloacetate. It also oxidizes some other 2-hydroxydicarboxylic acids. It is implicated in different pathways such as pyruvate metabolism, glyoxylate, and dicarboxylate metabolism, carbon fixation, reductive carboxylate cycle (CO_2 fixation), and the citrate cycle [25, 26]. L-Lactate dehydrogenase is also an oxidoreductase acting on *S*-lactate to product pyruvate. Various implications of L-lactate dehydrogenase have been described, including glycolysis/gluconeogenesis and cysteine metabolism.

The effects of AZC treatment were extended to proteins involved in cytoskeletal structure, such as tubulin-folding cofactor. This protein enters the pathway leading from newly synthesized tubulin (the major constituent of microtubules) to properly folded heterodimer (Fig. 10.2) or in the early stages of protein synthesis (IF2A). Finally, caspase 10 D is involved in the activation of the cascade of caspases responsible for apoptosis execution [27].

Surprisingly, some of the polypeptides were down-regulated following AZC treatment. The decrease in protein level observed may due to size effects of AZC not directly related to the demethylating activity of the drug. MS identified 15 down-regulated polypeptides corresponding to 13 proteins (Table 10.1). Down-regulated proteins belong to various pathways such as those involved with the biosynthesis of porphyrins, cholesterol metabolism (HMG-CoA synthetase) and cell detoxification (glutathione-*S*-transferase P). The various pathways affected by AZC treatment are shown in Figure 10.3. Up- and down-regulated proteins were taken together and placed in their respective pathways for obtaining much information

Table 10.1 Soluble proteins identified with changed expression following AZC treatment.

Protein identity	*Accession number in Swiss-Prot*	*Apparent pI/M,*[a)]	*Identification method and sequence coverage*	*Regulation*	*N-fold*
ADA3-like protein	O75528	6.09/42514	MALDI (26.4%)	down	2.0
BET3	O43617	4.95/19507	MALDI (40.0%)	up	1.8
Beta tubulin	PO7437	4.93/56268	MALDI (43.9%)	down	3.0
Caspase 10D precursor	Q9Y2U7	6.10/57645	MALDI (29.7%)	up	2.0
Copine 1	Q99829	5.63/61982	MALDI (34.6%)	down	4.5
GARS	Q15374	6.04/55784	MALDI (57.5%)	up	2.0
Glutathione-*S*-transferase P	P09211	5.0 8/24655	MALDI (56.5%)	down	1.7
GMP synthetase	P49915	6.17/74784	MALDI (59.3%)	down	1.9
Hydrolase AP4A	P50583	5.35/1 8617	MALDI (46.6%)	down	2.2
Hydroxymethylglutaryl-CoA synthetase	Q01581	5.40/57545	MALDI (25.2%)	down	3.2
L-Lactate dehydrogenase H chain	P07195	5.63/36368	MALDI (36.9%)	up	3.6
Malate dehydrogenase	P40925	6.09/36131	MALDI (53.2%)	up	1.9
Nucleoside diphosphate kinase A	P15531	5.70/21004	MALDI (47.4%)	down	4.0
Peroxyredoxin 3	P30048	5.90/24782	MALDI (45.4%)	down	1.6
Proteasome iota chain	P34062	5.88/26450	MALDI (24.4%)	down	2.3
Stathmin	P16949	5.51/18258	MALDI (39.9%)	down	1.8
Transaldolase	P37837	5.99/39 223	MALDI (50.7%)	down	1.8
Transaldolase	P37837	5.80/39 223	MALDI (26.7%)	down	1.8
Transaldolase	P37837	6.01/39 449	MALDI (51.9%)	up	2.7
Transaldolase	P37837	6.10/39 223	MALDI (54.3%)	up	1.8
Tropomyosin	P06468	4.74/38110	MALDI (78.8%)	down	2.0

Table 10.1 (cont.)

Protein identity	*Accession number in Swiss-Prot*	*Apparent pI/M_r* [a)]	*Identification method and sequence coverage*	*Regulation*	*N-fold*
Tubulin cofactor B	Q99426	5.09/32687	MALDI (29.9%)	up	1.6
Uroporphyrinogene decarboxylase	P06132	5.78/41189	MALDI (40.9%)	down	1.8

a) Apparent pI and M_r correspond to the experimental values calculated from the position of the spot on the 2-D map.

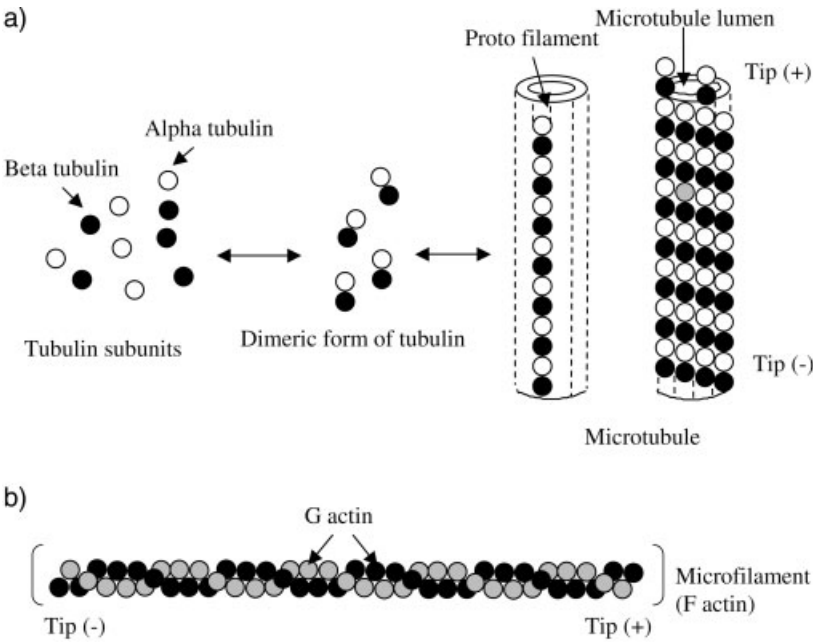


Figure 10.2 Organization of microtubules and microfilaments. Microtubules are constituted by a definite assembling of dimeric tubulin (a). F actin microfilament formations involve the association of two chains of G actin (b).

on AZC effect at the molecular level. We demonstrated that the proteomics data were completely in accordance with the biological results. A major consequence of AZC treatment was a decrease in cell growth with a blockage in the G_0/G_1 phase of the cell cycle [24]. The growth inhibition observed after 24 hours can be explained by dramatic changes observed in puric nucleotide biosynthesis [5].

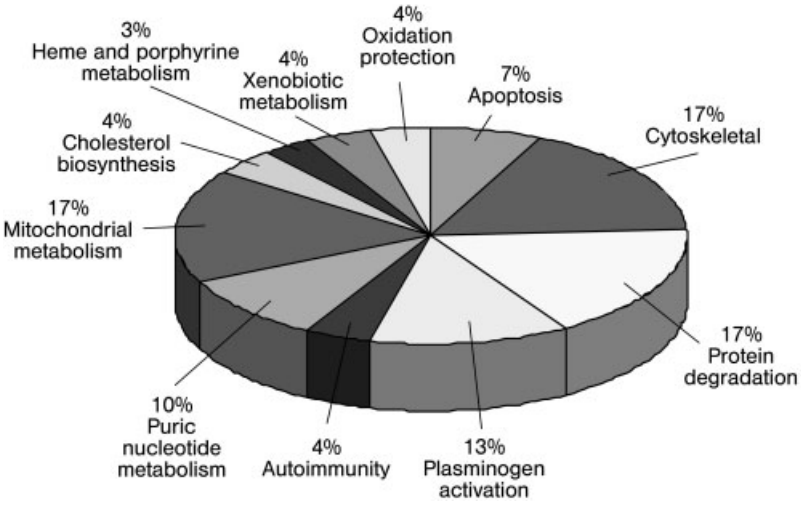


Figure 10.3 Pie chart of the protein functions affected by AZC, which gives rise to suggestions of possible biological effects of the drug. For instance several modifications concern the cytoskeletal organization, the mitochondrial metabolism, and the regulation of protein degradation.

More interesting was the decreased level of the enzyme glutathione-*S*-transferase P that catalyzes glutathione conjugation to a wide variety of exogenous and endogenous hydrophobic electrophiles. Catalysis of this conjugation is a step for elimination of compounds that become less toxic and more hydrosoluble than the start compounds [28]. Glutathione conjugate transporters then pump the foreign chemical out of the cell. It was previously reported that the effects of anticancer drugs are greatly reinforced when employed in combination to AZC for hemopathies treatment [22]. It is attractive to hypothesize that the decrease expression of glutathione-*S*-transferase P may be in part responsible for this effect.

10.3.2
Membrane-associated Protein Pattern Changes Induced by AZC Treatment

The analysis was also performed on membrane-associated proteins. AZC treatment altered the abundance of 49 polypeptides (Table 10.2). As described above for soluble proteins, AZC treatment affected protein profiles in two different ways: (1) the level of seven proteins was decreased; (2) 42 proteins were expressed or up-regulated. Presently, 34 identified polypeptides are located on the gel, corresponding to 27 individual proteins. Two of the proteins identified have been previously identified in the soluble extract, but with significant differences in pI. Interestingly, different isoforms of the same proteins (caldesmon, glutathione transferase omega 1) were found to be either down- or up-regulated, suggesting that they are modified at a post-translational level. Even if most of the affected pro-

Table 10.2 Membrane-associated proteins identified with changed expression following AZC treatment.

Protein identity	*Accession number in Swiss-Prot* [a)]	*Apparent pI/M_r* [b)]	*Identification method and sequence coverage* [c)]	*Regulation*	*N-fold*
Alpha enolase	P06733	6.95/52949	MALDI (32.8%)	up	
Alpha enolase	P06733	7.18/52478	MALDI (30.3%)	up	6.6
Alpha enolase	P06733	7.22/52478	MALDI (45.3%)	up	9.4
Alpha enolase	P06733	7.56/51912	MALDI (49.2%)	up	2.6
Bridging integrator-3	Q9NQY0	6.22/27763	MALDI (31.8%)	down	2.2
Caldesmon	Q05682	6.59/89010	MALDI (24.2%)	down	2.8
Caldesmon (splice isoform)	Q05682	5.88/61996	MALDI (24.7%)	up	2.6
CDNA FLJ20550 FIS, CLONE KAT11636	Q9NWX3	6.37/17636	MALDI (36.2%)	up	6.7
cDNA FLJ20760 FIS, clone HEP01082	Q9NWK9	5.83/51715	MALDI (35.3%)	up	11.3
EIF-3 theta	Q14152	7.15/165628	MALDI (24.2%)	down	3.5
Glutathione transferase omega 1	P78417	5.86/25259	MALDI (23.2%)	down	2.4
Glutathione transferase omega 1	P78417	6.01/27628	MALDI (25.3%)	up	19
Glutathione transferase omega 1	P78417	6.10/26979	MALDI (22.0%)	up	
Hypothetical protein CGI-99	Q9Y224	6.88/29594	MALDI (29.1%)	up	27.4
IDN4-GGTR7 protein	Q9Y6Y7	6.88/27673	MALDI (26.2%)	up	2.1
Islet cell autoantigen 1	Q05084	6.04/54065	MALDI (28.4%)	up	6
Lamin B1	P20700	5.69/67693	MALDI (48.0%)	up	257.2
Lamin B2 (fragment)	Q03252	5.91/65940	MS/MS (1)	up	
L-Lactate dehydrogenase B chain	P07195	6.19/36998	MALDI (53.5%)	up	2.3
Mago Nashi protein homologue	P50606	6.16/17799	MALDI (54.8%)	up	25.1
Nucleoside diphosphate kinase A	P15531	6.33/21937	MALDI (44.1%)	up	31.6
Phospholipid hydroperoxide glutathione Px	P36969	7.25/27398	MALDI (25.1%)	up	230

Table 10.2 (cont.)

Protein identity	*Accession number in Swiss-Prot*[a)]	*Apparent pI/M_r*[b)]	*Identification method and sequence coverage*[c)]	*Regulation*	*N-fold*
Proteasome activator PA28 alpha	Q06323	6.24/32252	MALDI (62.2%)	up	2.4
Proteasome activator PA28 alpha	Q06323	6.40/32153	MALDI (58.7%)	up	1.8
Proteasome activator PA28 beta	Q9UL46	5.87/32714	MALDI (52.3%)	up	2
Proteasome beta chain	P28070	6.14/28082	MALDI (53.4%)	up	2.2
Proteasome beta chain	P28070	5.77/27127	MALDI (25.6%)	up	12.5
Proteasome subunit alpha type 2	P25787	7.43/28174	MALDI (38.2%)	up	2.5
PtdIns transfer protein beta	P48739	7.34/31179	MALDI (35.2%)	up	8.9
RAS-related protein RAB-14	P35287	6.52/27628	MALDI (42.3%)	up	
Thioredoxin-dependent peroxide reductase	P30048	6.56/27763	"MALDI (36.1%), MS/MS (1)	up	
TIP47	O60664	5.68/48119	MS/MS (4)	up	3.7
TNF receptor superfamily member 12	Q93038	6.60/29000	MALDI (21.7%)	down	2.3
XAP-5 protein	Q14320	6.38/39187	MALDI (36.0%)	up	

a) Bold characters indicate proteins uniquely expressed in treated samples.
b) Apparent pI and M_r correspond to the experimental values calculated from the position of the spot on the 2-D map.
c) The percentage of sequence coverage and the number of identifying peptides are indicated for MALDI or LC-MS/MS identifications, respectively.

teins were up-regulated, it does not mean that this result is directly due to the demethylating activity of AZC. Among the identified up-regulated polypeptides, only 6 of 29 were uniquely found in treated samples. Moreover, some of them do not actually correspond to the expression of new proteins but of new isoforms. For instance, an isoform of α-enolase with a pI of 6.95 was only detected after AZC treatment, whereas three less acidic isoforms were detected at different levels without or with a treatment. These results indicate that the global changes observed after the drug treatment cannot be explained only by its demethylating activity. Or, in other words, AZC also regulates metabolic pathways.

As for soluble proteins, membrane-associated proteins were then grouped with regard to their functional classification based on the following categories: mem-

brane receptors, signaling molecules, cytoskeletal, nuclear protein, protein processing, metabolic enzymes, cell protection, and hypothetical [29]. Relevant information on protein function was retrieved from the Swiss-Prot and GeneBank databases (http://www.expasy.org/sprot/ and http://www.ncbi.nlm.nih.gov/Genbank/index.html). The difficulties of this functional classification are actually increased by the fact that many proteins sustain various functions. For instance, the presence of α-enolase, a key glycolytic enzyme, in a membrane fraction may seem astonishing. In fact enolase is a multifunctional protein that, in addition to its intracellular function, belongs to a class of surface proteins that do not possess classical machinery for surface transport, yet are transported on the cell surface through an unknown mechanism. Its ability to serve as a plasminogen receptor on the surface of a variety of hematopoietic, epithelial and endothelial cells, with which it is strongly associated, suggest that it may play an important role in the intravascular and pericellular fibrinolytic system [30, 31].

Nevertheless, within its limits, the functional classification of modified proteins is a way to advance hypothesis concerning the networks involved in a peculiar phenomenon.

10.4 Proteomic Study of Gal1-mediated B Cell Apoptosis

Galectin-1 (Gal1) is the earliest described member of a family of endogenous lectins showing affinity for β-galactosides [32]. During proteomic studies, it is one of the intracellular proteins whose expression is currently described as modified in cancer cells. Even if it is generally described as a plurifunctional protein, it is now known to play a role as a regulator of the cell cycle or as a pro-apoptotic factor. However, the molecular basis of these observations is far from known.

Gal1 has been shown to be secreted by a non-conventional pathway and to bind to cell surface receptors in an autocrine or paracrine manner, in different cells or tissues [33, 34]. In earliest studies, we used the erythroleukemia cytokine-dependent TF-1 cells as a model system to evidence the externalization of Gal1 and its autocrine binding to the cell membrane in hematopoietic cells. Two inducers of differentiation were used: aphidicolin, a chemical inhibitor of DNA polymerase α, and erythropoietin, a physiological inducer of erythroid differentiation. The modifications due to these inducers were analyzed by 2-DE on soluble protein extracts. Both treatments induced a synthesis of hemoglobin, confirming the acquisition of an erythroid phenotype by TF-1 cells. The differentiation apparently induced a clear decrease of Gal1. However, this decrease or lack of cytosolic Gal1 was not related to a decrease in synthesis, as no modification of expression was observed on northern blots performed in parallel, but to its externalization and autocrine binding on the cell surface. Moreover, the detection of the leukocyte common antigen CD45 by flow cytometry was strongly increased when membrane-bound Gal1 was eluted by a specific competitor (thiodigalactoside), suggesting that an effect of the competitor was to free masked receptors, and that the protein tyrosine phospha-

tase CD45 could be a major receptor for Gal1 at least for TF-1 cells. The signaling pathway governed by the binding of Gal1 to the cell membrane is not understood; the next steps toward the elucidation of Gal1-initiated events were to confirm the identity of its membrane receptor on another cell model, and to characterize the signal resulting from the binding of Gal1 to this receptor.

10.4.1 Identification of the Major Gal1-binding Membrane Glycoprotein

In light of the previous results, and the demonstration that in hematopoietic cells Gal1 can be externalized and then bind to surface receptors, it seemed reasonable to postulate that Gal1 is responsible for CD45 ligation associated with the regulation of signal transduction. The BL 36 cell line was used to verify whether CD45 was actually a receptor for Gal1 in B cells, and whether it was the major or the only receptor. Preliminary experiments showed that during SDS-PAGE of BL 36 cell membrane extracts, the proteins recognized by Gal1 migrated with a mobility similar to that of the proteins detected by a specific antibody directed against the tyrosine phosphatase glycoprotein CD45 [35].

To confirm this identity, Gal1 receptors were adsorbed on immobilized Gal1, biotinylated, and specifically eluted (Fig. 10.4). The eluted protein fractions were resolved by 8.5% SDS-PAGE, transferred to Immobilon-P membrane, and detected either by incubation with specific antibodies or by probing with streptavidin-horseradish peroxidase complex (Strep-HRP). The eluate resolved by SDS-PAGE was identified as CD45. No additional bands were detected by the Strep-HRP that detected all the proteins bound to Gal1. Moreover, an aliquot of Gal1-binding proteins was immunodepleted by immunoprecipitation with anti-CD45 antibody. The bands detected by Strep-HRP disappeared after this depletion. Therefore, it was clearly demonstrated that CD45 was the leading candidate as the major Gal1 membrane receptor on B cells.

10.4.2 Kinetics of Modification of Phosphorylation of the Protein Tyrosine Kinase Lyn

CD45 is well known to regulate the activity of protein tyrosine kinases (PTKs) of the Src family by dephosphorylating a regulatory phosphotyrosyl residue found at their C-terminus [36]. In B cells, the Src-family kinases are represented mostly by Blk, Fyn, and Lyn. In these cells Lyn is one of the most abundant Src family PTKs and is known to be associated with the antigen receptor and the CD19 co-receptor, as well as with CD45. Moreover, the C-terminal regulatory tyrosine is hyperphosphorylated on Lyn in CD45-deficient B cells, suggesting that it is a target of CD45 [37]. Our hypothesis was that in B cells the binding of Gal1 to CD45 induces its ligation, the result being a blockage of its phosphatase activity, resulting in a hyperphosphorylation of Lyn and therefore an inhibition of its kinase activity. To show if the binding of Gal1 to CD45 actually induced a modification of the

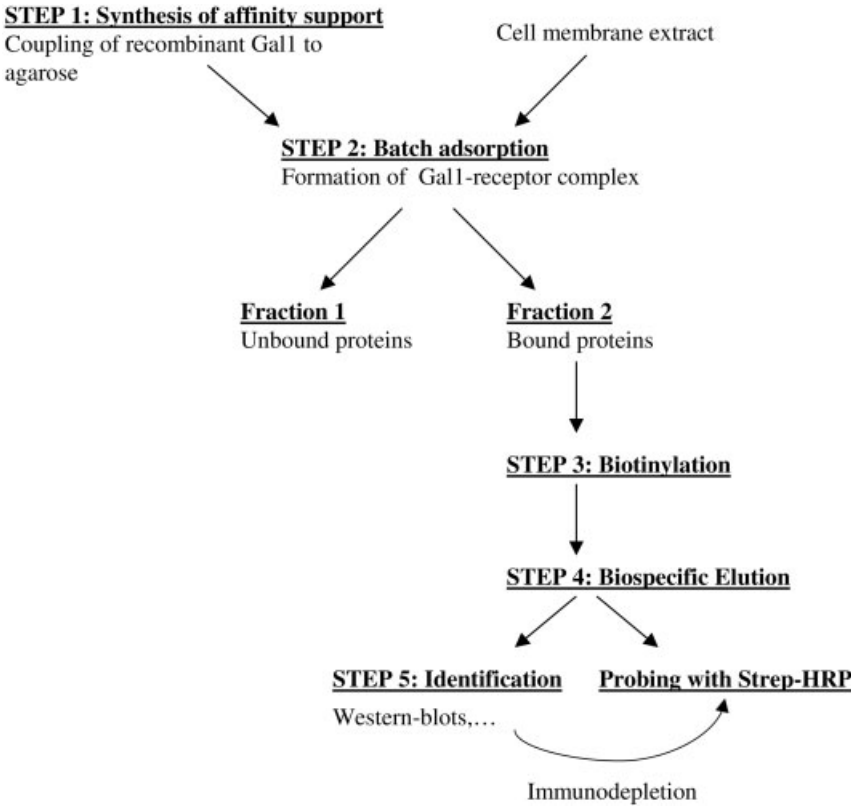


Figure 10.4 Flowchart of the experimental procedure for the identification of Gal1-binding membrane proteins. Gal1-binding proteins are first separated by batch adsorption of immobilized Gal1. Subsequently, the bound proteins are biotinylated and eluted by a specific inhibitor.

phosphorylation of Lyn, BL 36 cells were stimulated with 700 nM recombinant Gal1 for various periods and soluble cell extracts were prepared for these different conditions [12]. Immunoblotting studies showed that the phosphotyrosine staining of Lyn doubled in 2 min of the establishment of the Gal1/CD45 interaction, while the amount of Lyn protein remained constant. In addition, functional tests showed that this increased phosphorylation was accompanied by a decreased *in vitro* kinase activity.

Then, cell lysates of untreated and Gal1-treated BL 36 cells were immunoprecipitated with anti-Lyn antibody. For each experimental condition, 200 µL of cell extract were incubated with 10 mg of Ultralink Biosupport (Pierce) coupled with anti-Lyn antibodies. After several washes, the protein was eluted in the buffer used for the first dimension of 2-DE, and its isoforms were analyzed by 2-DE. Lyn was resolved in several spots, indicating the presence of post-translational modifi-

cations in addition to an alternative splicing. After the binding of Gal1 to CD45, only two spots were modified, these two spots showing a same molecular weight of about 58 kDa and a difference in pI of 0.3 units: 5.3 and 5.6, this shift corresponding to the addition of a phosphoryl group of the protein. The binding of Gal1 to CD45 led to a decrease of the volume of the less phosphorylated spot with a pI of 5.6, and to an increase of the more phosphorylated with a pI of 5.3. After 5 min, where the kinase activity of the protein was abolished, only the hyperphosphorylated form was detected. The modification was rapidly reversible, and after 10 min both the equilibrium between the two forms and the kinase activity were recovered.

The regulation of Lyn kinase by phosphorylation is complex in that there are two identified tyrosine phosphorylation sites: C-terminal phosphorylation is inhibitory and autophosphorylation is stimulatory. Our results suggested that Lyn is phosphorylated at the activation autophosphorylation site before the binding of Gal1 to CD45, but at both sides after this binding.

10.5
Lymphoblastoid and Lymphoma Cells 2-DE Database

In line with our wish to allow the scientific community to access our extensive work on the identity of the proteins on the lymphoblastoid and lymphoma 2-DE maps, we have created an online database that provides an interactive way to query the annotated reference maps included in the BPP (Biochimie des Protéines et Protéomique) database. The 2-DE data collected in this database are being made publicly available through a World Wide Web (WWW) site that links the protein data with databases containing protein sequence and biological information, as described in detail previously [4]. Further information can be obtained by directly examining the database online: http://www-smbh.univ-paris13.fr/lbtp/Biochemistry/Biochimie/bque.htm.

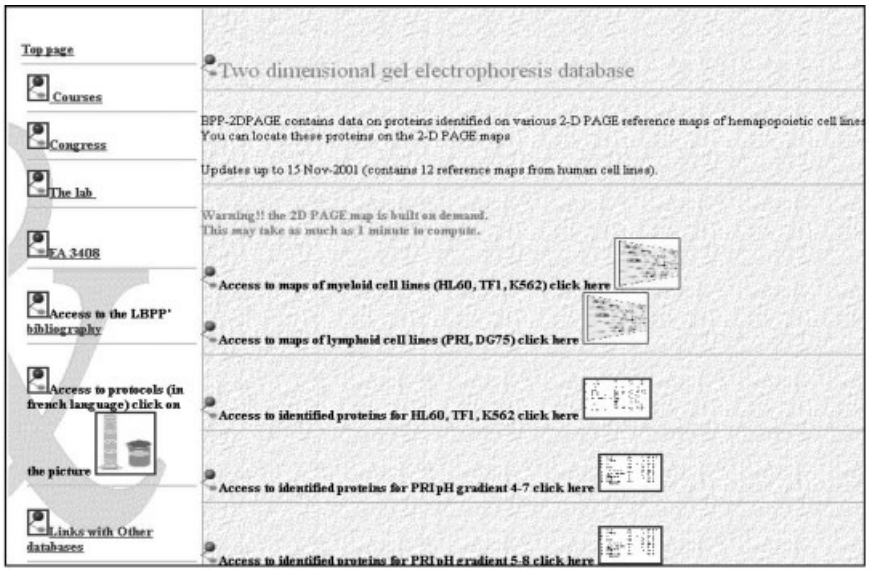
The BPP 2-DE database contains data on soluble proteins identified on various 2-DE maps of hematopoietic cell lines. Reference maps for the following lymphoblastoid and lymphoma cells are currently available: Burkitt lymphoma DG75 cells (pH gradient: 4–7) treated or not with AZC, lymphoblastoid cell line PRI (pH gradients: 4–7 and 5–8) during exponential growth phase.

There are two ways to query the BPP database (Fig. 10.5 a):

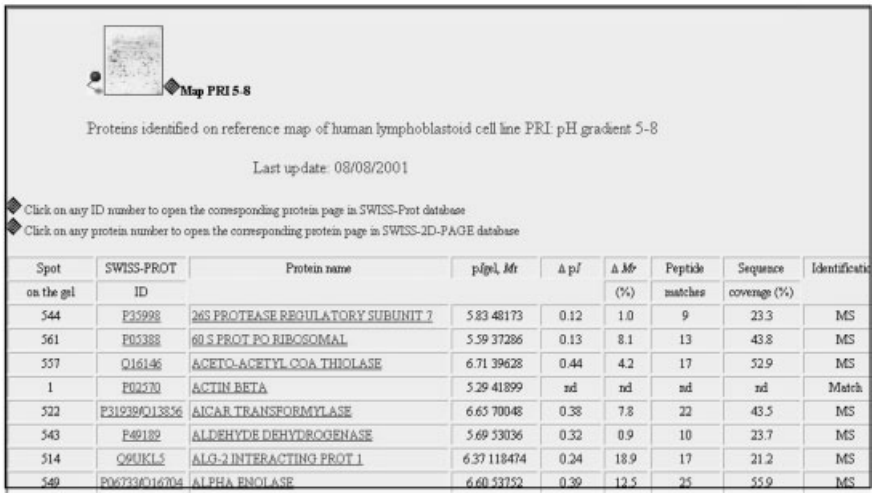
- Option 1: a view of any map with labels of identified proteins is obtained by clicking on the gel icon, while a view of the gel without labels is obtained by clicking on “without index” inside the window. Spots flagged with a label correspond to identified polypeptides. The labels correspond to the accession number in Swiss-Prot/TrEMBL.
- Option 2: a file listing the information entered for the polypeptides identified for each gel is displayed in the format shown in Figure 10.5 b, by clicking on a link: “Access to identified proteins for…”. For any given polypeptide find in the

file is listed the mapping procedure (MS, MS/MS, match with another gel), information on the theoretical MW and pI, experimental MW and pI, the number of matches and sequence coverage obtained by MS. In addition, links are given to the Swiss-Prot (http://us.expasy.org/sprot/) and Swiss 2-D PAGE (http://us.expasy.org/ch2d/) information pages, for the proteins indexed in these databases.

a)



b)



Spot on the gel	SWISS-PROT ID	Protein name	p*I*/gel, *M*r	Δ p*I*	Δ *M*r (%)	Peptide matches	Sequence coverage (%)	Identificatic
544	P35998	26S PROTEASE REGULATORY SUBUNIT 7	5.83 48173	0.12	1.0	9	23.3	MS
561	P05388	60 S PROT PO RIBOSOMAL	5.59 37286	0.13	8.1	13	43.8	MS
557	Q16146	ACETO-ACETYL COA THIOLASE	6.71 39628	0.44	4.2	17	52.9	MS
1	P02570	ACTIN BETA	5.29 41899	nd	nd	nd	nd	Match
522	P31939/Q13836	AICAR TRANSFORMYLASE	6.65 70048	0.38	7.8	22	43.5	MS
543	P49189	ALDEHYDE DEHYDROGENASE	5.69 53036	0.32	0.9	10	23.7	MS
514	Q9UKL5	ALG-2 INTERACTING PROT 1	6.37 118474	0.24	18.9	17	21.2	MS
549	P06733/Q16704	ALPHA ENOLASE	6.60 53752	0.39	12.5	25	55.9	MS

Figure 10.5 LBPP 2-D gel electrophoresis database main page (a), and protein selection page (b).

10.6 References

1 L. R. Silverman, J. F. Holland, R. S. Weinberg et al. *Leukemia* **1993**, *7 Suppl. 1*, 21–29.
2 D. J. Richel, L. P. Colly, J. C. Kluin-Nelemans, R. Willemze. *Br. J. Cancer* **1991**, *64*, 144–148.
3 G. A. Rabinovich, N. Rubinstein, M. A. Toscano. *Biochim. Biophys. Acta* **2002**, *1572*, 274–284.
4 M. Caron, N. Imam-Sghiouar, F. Poirier, J. P. Le Caër, V. Labas, R. Joubert-Caron. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* **2002**, *771*, 197–209.
5 F. Poirier, V. Labas, J.-P. Le Caër, M. Caron, R. Joubert-Caron. In *2001 Congress Functional Proteomics* (P. Palagi, J.-C. Sanchez, R. Stöcklin, Eds). Fontis Media, Lausanne, **2001**, pp 66–70.
6 T. Toda, M. Sugimoto. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* **2003**, *787*, 197–206.
7 A. Wennborg, M. Classon, G. Klein, A. von Gabain. *Biol. Chem. Hoppe Seyler* **1995**, *376*, 671–680.
8 T. Soussi, P. Jonveaux. *Nouv. Rev. Fr. Hematol.* **1991**, *33*, 477–480.
9 M. S. Lindstrom, K. G. Wiman. *Semin. Cancer Biol.* **2002**, *12*, 381–387.
10 M. Sanchez-Beato, A. Sanchez-Aguilera, M. A. Piris. *Blood* **2003**, *101*, 1220–1235.
11 H. Ben-Bassat, N. Goldblum, S. Mitrani et al. *Int. J. Cancer* **1977**, *19*, 27–33.
12 M. Fouillit, F. Poirier, E. Monostori, et al. *Electrophoresis* **2000**, *21*, 275–280.
13 R. Joubert-Caron, J. P. Le Caër, F. Montandon et al. *Electrophoresis* **2000**, *21*, 2566–2575.
14 T. Rabilloud, L. Vuillard, C. Gilly, J. J. Lawrence. *Cell Mol. Biol.* **1994**, *40*, 57–75.
15 N. L. Anderson, N. G. Anderson. *Electrophoresis* **1998**, *19*, 1853–1861.
16 A. Cosma. *Anal. Biochem.* **1997**, *252*, 10–14.
17 J. D. Cornillot, M. Pontet, C. Dupuy et al. *Glycobiology* **1998**, *8*, 425–432.
18 D. Bladier, R. Joubert, V. Avellana-Adalid, J. L. Kémény, C. Doinel, J. Amouroux, M. Caron. *Arch. Biochem. Biophys.* **1989**, *269*, 433–439.
19 N. Imam-Sghiouar, I. Laude-Lemaire, V. Labas et al. *Proteomics* **2002**, *2*, 828–838.
20 A. D. Riggs. *Cytogenet. Cell Genet.* **1975**, *14*, 9–25.
21 R. Singal, G. D. Ginder, *Blood* **1999**, *93*, 4059–4070.
22 M. Lubbert. *Curr. Top. Microbiol. Immunol.* **2000**, *249*, 135–164.
23 S. B. Baylin, J. G. Herman, J. R. Graff, P. M. Vertino, J. P. Issa. *Adv. Cancer Res.* **1998**, *72*, 141–196.
24 F. Poirier, M. Pontet, V. Labas et al. *Electrophoresis* **2001**, *22*, 1867–1877.
25 A. Lombardi, L. Beneduce, M. Moreno et al. *Endocrinology* **2000**, *141*, 1729–1734.
26 I. Schepens, K. Johansson, P. Decottignies et al. *J. Biol. Chem.* **2000**, *275*, 20996–21001.
27 P. W. Ng, A. G. Porter, R. U. Janicke. *J. Biol. Chem.* **1999**, *274*, 10301–10308.
28 S. Chatterjee, S. Battacharya, *Toxicol. Lett.* **1984**, *22*, 187–198.
29 F. Poirier, R. Joubert-Caron, V. Labas, M. Caron. *Proteomics* **2003**, *3*, 1028–1036.
30 S. Moscato, F. Pratesi, A. Sabbatini et al. *Eur. J. Immunol.* **2000**, *30*, 3575–3584.
31 F. Pratesi, S. Moscato, A. Sabbatini, D. Chimenti, S. Bombardieri, P. Migliorini. *J. Rheumatol.* **2000**, *27*, 109–115.
32 K. I. Kasai, J. Hirabayashi. *J. Biochem.* **1996**, *119*, 1–8.
33 D. N. W. Cooper, S. H. Barondes. *J. Cell Biol.* **1990**, *110*, 1681–1691.
34 D. Lutomski, M. Fouillit, P. Bourin et al. *Glycobiology* **1997**, *7*, 1193–1199.
35 M. Fouillit, R. Joubert-Caron, F. Poirier et al. *Glycobiology* **2000**, *10*, 413–419.
36 I. Tamir, J. C. Cambier. *Oncogene* **1998**, *17*, 1353–1364.
37 S. Yanagi, H. Sugawara, M. Kurosaki, H. Sabe, H. K. Yamamura. Urosaki, T. *J. Biol. Chem.* **1996**, *271*, 30487–30492.

Part IV
Pharmaco-toxicology

Biomedical Application of Proteomics
Edited by J.-C. Sanchez, G. L. Corthals, D. F. Hochstrasser

ISBN: 3-527-30807-5

11
Chemoresistance in Cancer Cells

Julia Poland, Dirk Schadendorf, Hermann Lage, and Pranav Sinha

11.1
Introduction

Palliative treatment including chemotherapy and other modes of treatment, e.g. radiotherapy and/or hyperthermia, is often the only remaining option in the management of certain solid tumors. Unfortunately its efficacy is poor due to low tumor sensitivity and the development of therapy resistance. Many different malignancies, including gastric and pancreatic cancers as well as malignant melanoma, show a particularly high level of intrinsic drug resistance to chemotherapeutic agents. Additionally, these tumors develop acquired drug resistance, which includes the classical multidrug resistance (MDR) phenomenon accompanied by the synthesis of P-glycoprotein (Pgp). Furthermore, atypical MDR phenotypes are mediated by several different mechanisms, some of which are still unknown.

Cell culture models represent a useful tool to study resistance phenomena. In order to find candidate proteins that are potentially associated with chemoresistance, we analyzed the differential protein expression *in vitro* in previously established cancer cell lines using proteomics. Proteome analysis is most commonly accomplished by the combination of two-dimensional polyacrylamide gel electrophoresis (2-D PAGE) and mass spectrometry (MS) and allows the identification and quantification of the proteins expressed by cells, tissues, or an organism in a high throughput manner. A model system for studying chemoresistance of melanoma cells has been established, utilizing four commonly used cytotoxic drugs for the generation of stable drug-resistant sublines of the human melanoma cell line MeWo [1]. These four anticancer agents – vindesine, cisplatin, fotemustine, and etoposide – differ in their mode of action: vindesine inhibits the formation of microtubules and thus leads to cell cycle arrest during mitosis, the cytotoxic effect of fotemustine is based on its alkylating activity, the antitumor effect of cisplatin issues from the drug-mediated formation of DNA adducts, whereas etoposide is a potent DNA-topoisomerase II (Topo II) inhibitor. Since the cytostatic drugs used belong to four different substance classes and exhibit completely different modes of drug action, it appears obvious that exposure to these drugs results in the activation of different mechanisms of resistance. Additionally, we established model systems for classical and atypical MDR using the human pancreatic cancer cell

Biomedical Application of Proteomics
Edited by J.-C. Sanchez, G. L. Corthals, D. F. Hochstrasser

ISBN: 3-527-30807-5

line EPP85-181-P and the human gastric cancer cell line EPG85-257-P to study drug resistance [2–4].

In the following, we describe fractionation of these cell lines with conventional two-dimensional electrophoresis (2-DE) in the pH range 4–8 using the ISO-DALT system (for materials and methods see refs. [5–12]). The results are given for each type of tumor independently. For melanoma, we have also used special fractionation techniques with immobilized pH gradients (IPG) in very acidic (pH 2.8–5.0) and basic pH ranges (pH 8.0–11). For details of these studies the reader is referred to [12] and [13]. These results obtained with proteome analysis are fundamental to the elucidation of the molecular mechanisms of chemoresistance that may assist the therapy of inoperable cancers. Subsequently, other studies directly focusing on these proteins can be performed based on the proteomics data, including confirmation studies (e.g. western blotting) and functional studies (e.g. transfection experiments) to prove the validity of the results.

11.2 Two-dimensional Electrophoresis Maps of Gastric Cancer, Pancreatic Cancer, and Melanoma

We have noticed in the past that typical spot patterns obtained after conventional 2-DE show around 1500 protein spots altogether, independent of the cell line used. Below we discuss three individual gels representing the three cell cultures grown under identical conditions.

11.2.1 Gastric Cancer

The human gastric carcinoma cell line EPG85-257 and its drug-resistant sublines were established in our laboratory previously [14]. The two chemoresistant cell lines EPG85-257-RNOV (exhibiting atypical MDR after selection against mitoxantrone) and EPG85-257-RDB (exhibiting classical MDR after selection against daunorubicin) have been compared with the parental, drug-sensitive cell line EPG85-257-P. Figure 11.1 compares representative electropherograms of these cell lines. Differentially expressed proteins are marked and named. Overexpressed proteins are underlined, while underexpressed proteins are in italics and marked in broken lines. Computer-assisted analysis revealed differential protein expression of 14 proteins in EPG85-257-RNOV in comparison with EPG85-257-P. Differentially up-regulated were seven proteins, namely a 14-3-3 related protein, alkaline phosphatase, BiP (glucose regulated protein 78), G-beta 1 and 2 (β-subunit of G-proteins), glutathione transferase M3 (GSTM3), peroxiredoxin 1, and transgelin 2 (TAGLN2). Down-regulated were the following seven protein spots: aldehyde dehydrogenase 1 (ALDH1A1), calnexin, cyclin D2, FK506-binding protein 4 (FKBP4), HSP27, nucleolar protein B 23 (numatrin, nucleophosmin), and vimentin. In the classical MDR subline EPG85-257-RDB five proteins were found to be differen-

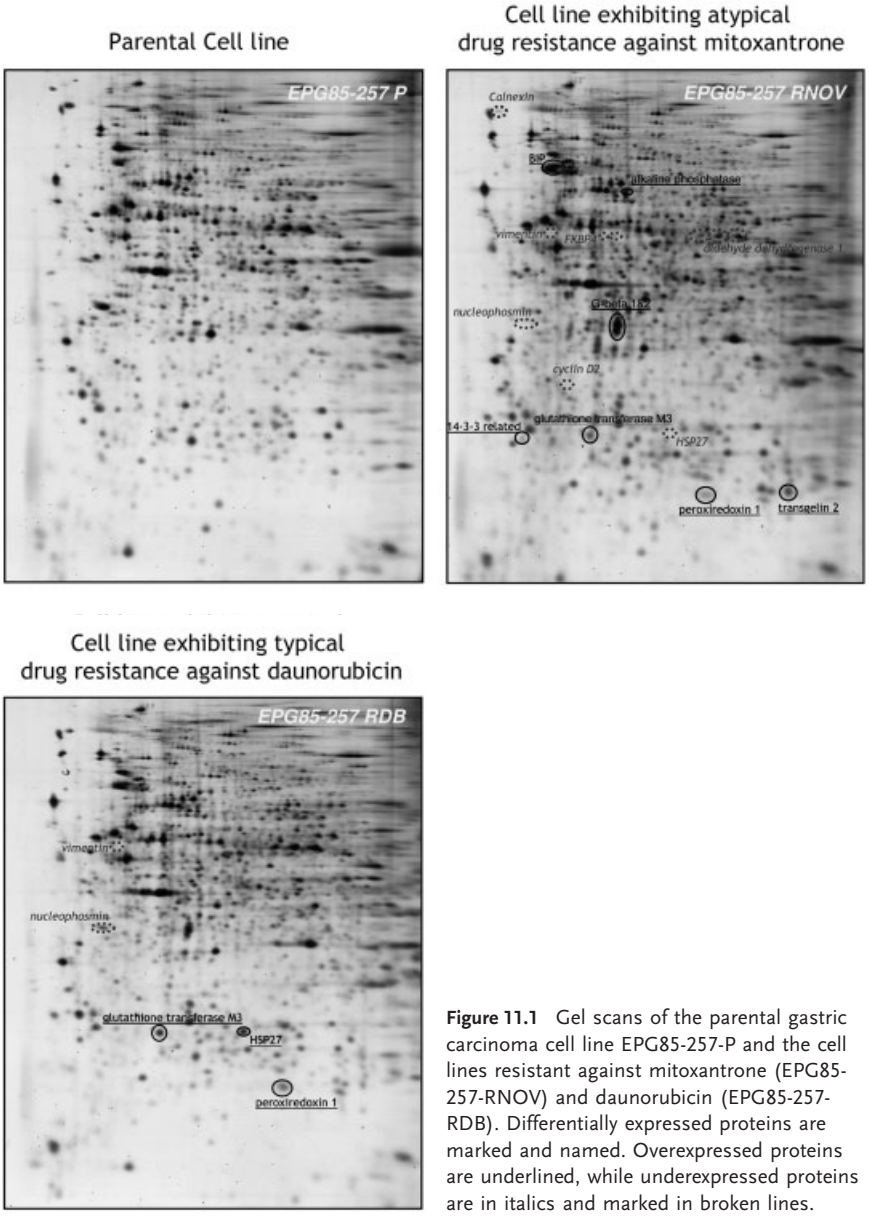


Figure 11.1 Gel scans of the parental gastric carcinoma cell line EPG85-257-P and the cell lines resistant against mitoxantrone (EPG85-257-RNOV) and daunorubicin (EPG85-257-RDB). Differentially expressed proteins are marked and named. Overexpressed proteins are underlined, while underexpressed proteins are in italics and marked in broken lines.

Table 11.1 Differential expression of proteins in gastric cancer cells.

Protein	*Differential overexpression*	*Differential underexpression*
14-3-3 related	EPG85-257-RNOV	
Aldehyde dehydrogenase 1 (ALDH1A1)		EPG85-257-RNOV
Alkaline phosphatase	EPG85-257-RNOV	
BiP, glucose-regulated protein 78 (grp 78) (HSPA5)	EPG85-257-RNOV	
Calnexin (CANX)		EPG85-257-RNOV
Cyclin D2 (CCND2)		EPG85-257-RNOV
FK506-binding protein 4 (FKBP4)		EPG85-257-RNOV
G-beta 1 and 2 (β-subunit of G-proteins)	EPG85-257-RNOV	
Glutathione transferase M3 (GSTM3)	EPG85-257-RNOV EPG85-257-RDB	
HSP27 (HSPB1) (HSPB2)	EPG85-25 7-RDB	EPG85-257-RNOV
Nucleolar protein B 23, numatrin, nucleophosmin (NPM1)		EPG85-257-RNOV EPG85-257-RDB
Peroxiredoxin 1	EPG85-257-RNOV EPG85-257-RDB	
Transgelin 2 (TAGLN2)	EPG85-257-RNOV	
Vimentin (VIM)		EPG85-257- RNOV EPG85-257-RDB

tially expressed. Overexpressed were the three proteins glutathione transferase M3 (GSTM3), HSP27, and peroxiredoxin 1; down-regulated were nucleolar protein B 23 and vimentin. The results of the differential protein expression studies are summarized in Table 11.1.

11.2.2 Pancreatic Cancer

The human pancreatic carcinoma cell line EPP85-181-P and its drug-resistant (classical MDR subline EPP85-181-RDB and atypical MDR subline EPP85-181-RNOV) sublines have previously been established, analogous to the gastric cancer cell lines [15]. Figure 11.2 shows comparison of the spot patterns obtained from the parental cell line EPP85-181-P and the two chemoresistant counterparts. In the atypical MDR subline EPP85-181-RNOV, eight proteins were found to be differentially expressed in comparison to the parental cell line. Cytokeratin 7 and 8, enoyl CoA hydratase, HSP70-related protein and nucleolar protein B 23 were up-regulated while BiP, G-beta 1 and 2, HSP27 and a proteasomal protein were down-regulated. In the case of the classical MDR subline EPP85-181-RDB, there was differential expression of the following nine proteins: 14-3-3 σ, calnexin, cancer oncogene, HSP70-related protein, nucleolar protein B 23, thioredoxin peroxidase 1 were overexpressed and BiP, HSP27, and a proteasomal protein were underexpressed. The results for differential protein expression are summarized in Table 11.2.

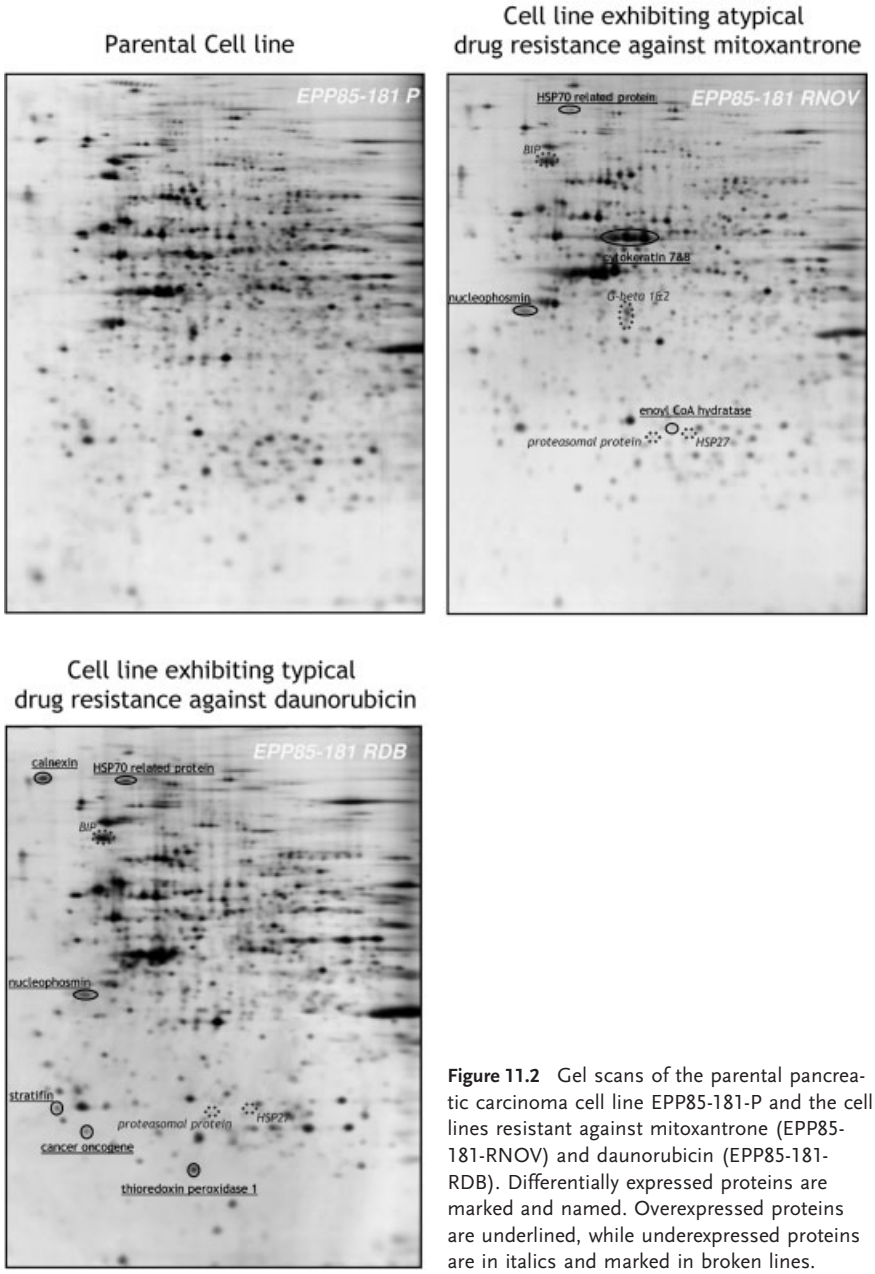


Figure 11.2 Gel scans of the parental pancreatic carcinoma cell line EPP85-181-P and the cell lines resistant against mitoxantrone (EPP85-181-RNOV) and daunorubicin (EPP85-181-RDB). Differentially expressed proteins are marked and named. Overexpressed proteins are underlined, while underexpressed proteins are in italics and marked in broken lines.

Table 11.2 Differential expression of proteins in pancreatic cancer cells.

Protein	*Differential overexpression*	*Differential underexpression*
14-3-3 σ (Stratifin) (SFN)	EPP85-181-RDB	
BiP, glucose-regulated protein 78 (grp 78)		EPP85-181-RDB
(HSPA5)		EPP85-181-RNOV
Calnexin (CANX)	EPP85-181-RDB	
Cancer oncogene	EPP85-181-RDB	
Cytokeratin 7 and 8	EPP85-181-RNOV	
Enoyl CoA hydratase	EPP85-181-RNOV	
G-beta 1 and 2 (-subunit of G-proteins)		EPP85-181-RNOV
HSP27 (HSPB1)		EPP85-181-RDB
(HSPB2)		EPP85-181-RNOV
HSP70-related protein	EPP85-181-RDB	
	EPP85-181-RNOV	
Nucleolar protein B 23, numatrin, nucleophos-	EPP85-181-RDB	
min (NPM1)	EPP85-181-RNOV	
Proteasomal protein		EPP85-181-RDB
		EPP85-181-RNOV
Thioredoxin peroxidase 1 (PRDX2)	EPP85-181-RDB	

11.2.3 Melanoma

The human melanoma cell line MeWo was derived from a metastatic lymph node of a patient with malignant melanoma established at the Memorial Sloan Kettering Cancer Centre (New York, USA). The drug-resistant variants were grown by adding twice per week different concentrations of the drugs (cisplatin, vindesine, etoposide, and fotemustine) to the medium. Details of this procedure can be found in [1].

The spot pattern obtained from the chemosensitive cancer cell line MeWoP was compared with the patterns obtained from the drug-resistant counterparts, selected in the presence of different concentrations of cisplatin, etoposide, fotemustine, and vindesine. A matchset has been established using PDQuest software 6.1 consisting of 10 member gels each, namely $MeWo_P$ and the drug-resistant descendants $MeWo_{Cis0.01}$ and $MeWo_{Cis1}$, $MeWo_{Eto0.1}$, $MeWo_{Eto0.5}$, and $MeWo_{Eto1}$, $MeWo_{Fote4}$, and $MeWo_{Fote40}$ as well as $MeWo_{Vin0.5}$ and $MeWo_{Vin5}$.

A comparison of the gel spot images showed that 14 proteins were differentially expressed in the chemoresistant descendants of $MeWo_P$. Eleven proteins, namely HSX70 variant, HSP60, HSP60 variant, HSP27, HSP27 isoform, β_1- and β_2-subunit of G-proteins, peroxiredoxin 1, a mixture of proteasome subunit beta type 3 and peroxiredoxin 3, nicotinamide *N*-methyltransferase as well as a hypothetical protein DKFZ p566J2046 were differentially overexpressed. Hydroxyacyl-coenzyme A dehydrogenase type II comigrated in our gels together with the HSP27 isoform. The remaining three proteins, nucleophosmin, HSP90, and phosphoglyceromu-

Table 11.3 Differentially expressed proteins in chemoresistant melanoma cell lines.

	Cis 0.01	*Cis 1*	*Eto 0.1*	*Eto 0.5*	*Eto 1*	*Fote 4*	*Fote 40*	*Vin 0.5*	*Vin 5*
HSX70 variant	(↑)	(↑)	(↑)	–	↑	↑	–	↑	(↑)
HSP60 and HSP60 variant	↑	–	–	↑	(↑)	(↑)	(↑)	(↑)	(↑)
HSP27	↑	↑	↑	↑	(↑)	↑	(↑)	↑	↑
HSP27 variant/hydroxyacyl-coenzyme A dehydrogenase, type II	↑	↑	–	↑	(↑)	↑	–	↑	–
β_1/β_2-subunit of G-proteins	(↑)	(↑)	(↑)	(↑)	↑	↑	–	↑	–
Nicotinamide *N*-methyltransferase	–	–	–	(↑)	–	–	–	(↑)	↑
Peroxiredoxin 1	–	(↑)	–	–	↑	↑	–	↑	–
Proteasome subunit beta type 3/peroxiredoxin 3	↑	–	–	(↑)	(↑)	–	↑	↑	–
Hypothetical protein DKFZp566J2046	↑	(↑)	–	(↑)	↑	↑	(↑)	↑	–
HSP90	↓	(↓)	–	–	–	–	(↓)	–	–
Nucleophosmin	–	–	↓	–	↓	↓	–	(↓)	–
Phosphoglyceromutase	↓	(↓)	–	–	↓	↓	–	↓	–

– No change
(↑) Slightly overexpressed (spot intensity > 2-but < 4-fold increased)
(↓) Slightly underexpressed (spot intensity > 2-but < 4-fold decreased)
↑ Clearly overexpressed (spot intensity > 4-fold increased)
↓ Clearly underexpressed (spot intensity > 4-fold decreased)

tase, were underexpressed in at least three chemoresistant cell lines. The results are summarized in Table 11.3.

11.3 Evaluation of the 2-DE Protein Maps

11.3.1 Chemoresistance Overview

Chemotherapy of cancer may fail for various reasons. Among these, drug resistance is the most important one. According to the figures for Germany, provided by the Robert Koch Institute in Berlin from 1997, around 338 300 patients suffer from different cancers each year (http://www.rki.de/GBE/KREBS/KREBS.HTM). The average 5-year survival rate in men is about 31% and in women about 44%. Mortality due to cancer affects 215 000 patients every year. Cancer mortality constitutes the second most common cause of death in western countries.

Solid tumors, e.g. tumors of the breast, pancreas, stomach, colon, and melanoma, as well as fibrosarcomas, constitute about a third of all new cases of cancer

per year. During treatment, resistance towards therapy, especially resistance towards anticancer drugs, develops in a third of these cases, i.e. roughly affecting around 40 000 patients. However, this number is probably much higher than official records maintain.

Resistance may be primary (intrinsic), i.e. the tumor cells do not respond from the start; or it may be secondary (acquired), i.e. the tumor initially responds to therapy, but eventually tumor growth resumes and the patient relapses.

11.3.2 Mechanisms of Drug Resistance

It is likely that the biochemical mechanisms involved in primary and secondary resistance are largely similar, but we know far more of secondary resistance, since it is easier to study in the laboratory. Resistance can be induced in cultured cells and one can then determine the differences between the resistant cells and the parental cells from which they were derived. Usually this allows the investigator to determine which difference is responsible for resistance. In contrast, in primary resistance there is no sensitive cell for comparison that is identical in all properties to the resistant one, but for the resistance. Acquired drug resistance in cultured cells is nearly always caused by a genetic change in these cells, i.e. resistant cells are mutant cells.

11.3.2.1 Classical Multidrug Resistance

One of the best-investigated mechanisms of drug resistance is the phenomenon of drug efflux with a non-specific carrier, such as the P-glycoprotein, called classical multidrug resistance [16–18]. It occurs because of the overexpression of P-glycoprotein (Pgp), a transmembrane protein with a molecular weight of 170 kDa. Pgp is coded by the *mdr-1* gene on chromosome 7q21 and belongs to the ATP-binding cassette (ABC) transporter family. Synonyms for Pgp are MDR-1 and PGY1, and according to the HGNC gene family nomenclature proposed in 1999, Pgp is also called ABCB1 (http://www.gene.ucl.ac.uk/users/hester/abc.html). ABC transporters are a large superfamily of integral membrane proteins involved in ATP-dependent transport of chemotherapeutic drugs across biological membranes. Most ABC transporters consist of four domains – two membrane-spanning domains and two cytoplasmic domains. The latter contain conserved nucleotide-binding motifs. Chemoresistance caused by the overexpression of Pgp exhibits a typical resistance pattern towards anthracyclins, epipodophyllotoxins, vinca alkaloids, and taxol [19].

Although the clinical relevance of classical multidrug resistance is understood, the impact on clinical practice is disappointing. This may be attributed to the lack of therapeutic agents effective in inhibiting PGP or other MDR-related factors, but even more to the ignorance of a large number of other strategies by which the cell survives potentially lethal damage. In consideration of the complexity of cellular processes it is clear that only the combination of several factors results in

chemoresistance [20]. These factors include intracellular activation or blocking of signal pathways, elevation of drug detoxification, modification of drug targets, cell arrest, increased DNA repair, and circumvention of apoptosis.

11.3.2.2 Atypical Multidrug Resistance

ABC Transporters Chemoresistant cells that do not overexpress Pgp are said to be atypically resistant – i.e. they overexpress alternative ABC transporters, e.g. "multidrug resistance-associated protein" (MRP) [21, 22] now designated as MRP1 or ABCC1, MRP2, or ABCC2 (cMOAT, "canalicular multispecific organic anion transporter") [23], MRP3 or ABCC3 (MOAT-D) [24], MRP5 or ABCC5 (MOAT-C) [25], and "breast cancer resistance protein" (BCRP) [26] also designated as ABCP (from placenta-specific ABC transporter) [27] or MXR (from "mitoxantrone resistance") [28] or ABCG2 belonging to the white subfamily G of ABC transporters. MRP, like Pgp, causes resistance towards anthracyclins, epipodophyllotoxins, vinca alkaloids but not towards colchicine and taxol. MRP can transport unconjugated drugs and drugs that are conjugated to glutathione, a process catalyzed by glutathione *S*-transferase. MRP2 or cMOAT also transports GSH-conjugated drugs and is especially important for the transport of platinum-containing drugs and perhaps also anthracyclins. A summary of the mechanisms can be found in [29].

Detoxification by Glutathione Transferases Another form of drug resistance in tumor cells arises from the intracellular detoxification of anticancer agents by the glutathione *S*-transferases, a group of intracellular enzymes consisting of three different isoenzymes, μ, π, σ [30–33]. Increased activities of one of the isoforms may be responsible for the development of chemoresistance towards cisplatin, 1,3-bis(2-chloroethyl)-1-nitrosourea (BCNU), doxorubicin, and daunorubicin especially in ovarian and gastrointestinal tumors. Altered enzyme activity, e.g. of topoisomerase I (hydrolysis of one DNA strand) and topoisomerase II (hydrolysis of both DNA strands) or mutant topoisomerases can also lead to multidrug resistance in several cell types, e.g. melanomas [1, 34].

Lung Resistance Protein A protein called lung resistance protein (LRP) was shown to be overexpressed in several atypically resistant cancer cells [35]. LRP has been identified as the major vault protein (MVP), the main component of multimeric vault particles. With the recent identification of the two minor vault proteins as telomerase-associated protein (TEP1) and vault-poly (ADP-ribose) polymerase (VPARP), and with high-resolution three-dimensional imaging, the composition of vaults is almost unravelled [36]. Although the first direct evidence for a causal relationship between LRP/MVP expression and drug resistance has been obtained, many functional aspects of vaults in normal physiology and in MDR still need to be clarified. The current clinical data on LRP/MVP detection indicate that LRP/MVP expression can be of high clinical value to predict the response to chemotherapy of several tumor types [37–40].

Modulation of Apoptosis Modulation of the apoptotic pathways is a mechanism that has attracted increased attention in the past. There are two well-characterized pathways leading to terminal caspase activation. The first is initiated by cell surface death receptor (Fas, Fas ligand, death domains); recruitment of adapter molecules and initiator caspases (caspase 8 and 10) followed by activation of terminal caspases 3 and 7. The intrinsic pathway is initiated by changes in the mitochondrial transmembrane potential caused by changes in membrane-associated proteins (Bcl-2, bax, bclxl) and release of cytochrome *c*. This is followed by the formation of the apoptotic protease-activating factor (Apaf-1), recruitment of caspase 9 in the complex, activation of terminal caspases and DNA fragmentation. Chemostatic drugs are supposed to cause an inhibition or dysregulation of apoptosis [30].

Other Mechanisms Other mechanisms that have been implicated in the development of chemoresistance in cancers cells are modulation of enzymes involved in DNA repair [41–43], intracellular sequestration with metallothioneins [44], exocytosis in which proteins such as kinesins and dynines may be involved and modification of protein kinase C activity. These mechanisms are summarized in [45].

11.3.3 Differentially Expressed Proteins in Drug-resistant Cancer Cells

We show that a comparatively small number of proteins were found to be differentially expressed in gastric and pancreatic cancer cells as well as in melanoma cells and their chemoresistant variants in the investigated pH range 4–8. This may be due to the fact that the resulting 2-D pattern mainly consists of high-abundance proteins and we did not apply subcellular fractionation (e.g. membrane fractionation) in this study. In melanoma cells, we also used very acidic and basic pH ranges with immobilized pH gradients to analyze the differential protein expression over an expanded range [12, 13].

Summarizing the results, we show that the following groups of proteins are involved: proteins involved in signal transduction, heat shock proteins and other chaperones, enzymes involved in metabolic pathways, calcium-binding proteins, and proteins involved in drug detoxification. In the investigated pH range, no plasma membrane transporters (e.g. ABC transporters) were found to be differentially expressed using the above-mentioned protein extraction procedure. Furthermore, only some of these proteins have been previously linked to chemoresistance and this may be an advantage of the proteomic approach in studying chemoresistance. Two groups of proteins turn out to be differentially expressed in all three types of cancer and are furthermore interesting from the functional point of view; these are the molecular chaperones and the proteins involved in drug detoxification. The results for both groups will be summarized briefly in the following.

11.3.3.1 Group of Molecular Chaperones

One of the major protein groups found to be overexpressed in certain chemoresistant gastric and pancreatic cancer as well as melanoma cells is the group of chaperones, including heat shock proteins (HSPs). Each member of the HSP superfamily appears to have a distinct set of functions within the cell. Some of these functions are well documented, such as the modulatory effects of HSPs on apoptosis, others need further elucidation, for example the role of special HSPs in thermo-tolerance. Our results obtained from the chemoresistant cancer cell lines exhibit differential expression of several major HSPs. In the case of melanoma, in nearly every chemoresistant variant there was an increased expression of the small stress protein HSP27, and most cell lines showed up-regulation of either HSX70 variant, a HSP70 family member, or a HSP60 isoform. Furthermore, proteome analysis revealed down-regulation of HSP90 in cisplatin- and fotemustine-resistant cells. In chemoresistant gastric cancer cell lines, HSP27 was overexpressed in the daunorubicin-resistant cell line, but underexpressed in mitoxantrone-resistant cells. For some other proteins shown to be differentially expressed (partly up-regulation, partly down-regulation), their chaperone activity has also been reported, namely calnexin (calcium-dependent chaperone), nucleophosmin, and BiP. In the case of pancreatic cancer, the small HSP27 was down-regulated in both drug-resistant variants. In contrast, in both resistant cell lines there was up-regulation of an HSP70-related protein. In addition, nucleophosmin was overexpressed in both mitoxantrone- and daunorubicin-resistant cell lines, while calnexin was up-regulated only in the daunorubicin-resistant cell line.

Summarizing these partially conflictive results of differential protein expression of chaperones, it is important to note that there is an increased expression of one of the major HSPs in every chemoresistant cell line, independent of the type of cancer. Interestingly, we have obtained similar results for thermo-resistant cancer cells [11].

HSP70 and HSP27, besides their putative role in thermo-tolerance, are of special clinical interest because they are thought to be involved in the MDR phenotype. HSP27 is a major target of phosphorylation upon cell stimulation and has been suggested to have a phosphorylation-regulated function [46]. HSP70 has long been recognized as one of the primary heat shock proteins in mammalian cells, and its structure is highly conserved during evolution. Both proteins have been repeatedly demonstrated to inhibit apoptosis induced by different chemotherapeutics, especially drugs that target topoisomerase II enzymes, such as anthracyclins and etoposide. Studies have shown that the production of reactive oxygen species (ROS) by these drugs has a major role in their apoptosis induction, while treatment with antioxidants increases cellular resistance to these agents [47]. Details about potential mechanisms for inhibition of apoptosis by HSPs can be found elsewhere [48, 49].

HSP90 chaperones play a role in many immunological processes. They are involved in antigen peptide transfer between proteasomes, function as transporters associated with antigen-processing molecules, and participate in steroid hormone responses [50–52]. Interestingly, HSP90 has been implicated in the induction of apoptosis in response to TNF in combination with cycloheximide [53]. This may be linked to the chaperone function of HSP90, assuming that the effect could be

due to the blockage of an unknown apoptotic inhibitor. It may be possible that down-regulation of HSP90 in drug-resistant melanoma cell lines indirectly results in the inhibition of apoptotic events. Data concerning these events, however, are very limited and the role of HSP90 in regulation of apoptosis induced by anticancer drugs has not been investigated yet.

11.3.3.2 Group of Proteins Involved in Drug Detoxification

Antioxidants govern intracellular redox status. Inside cells, glutathione (GSH), glutaredoxin, and thioredoxin represent the major reducing agents.

The glutathione-related detoxification pathway is commonly discussed as one of the major mechanisms of MDR. GSH and the associated enzymes glutathione *S*-transferases, peroxidases, and reductase, have frequently been implicated in chemoresistance [54–56]. Glutathione reductase was differentially overexpressed in several chemoresistant variants of the human melanoma cell line MeWo. This enzyme maintains high levels of reduced GSH in the cytosol and may therefore act as an indirect marker for GSH concentration.

In mitoxantrone- and daunorubicin-resistant gastric cancer cell lines, another protein belonging to the glutathione system was overexpressed, namely glutathione transferase M3. This protein has also been found to be up-regulated in certain thermo-resistant cell lines but there is no evidence for a connection to chemoresistance in pancreatic cancer, where GST M3 is not even expressed.

Apart from that, members of the thioredoxin system, peroxiredoxin 1 and peroxiredoxin 3 (comigration with proteasome subunit beta type 3 in melanoma cells), were shown to be overexpressed in chemoresistant cells of all three types of cancer.

Thioredoxin reductase (TR), thioredoxin (Trx), and thioredoxin peroxidase (Tpx)/peroxiredoxin (Prx) are three linked components in a redox chain that couples peroxide reduction to NADPH oxidation. The possible functions of these proteins have already been reviewed previously by us [12].

11.4 Conclusions

In conclusion, proteome analysis of chemoresistant cancer cell lines represents a powerful tool in finding candidate proteins that are potentially involved in the drug-resistant phenotype. It is interesting to note that a large number of differentially expressed proteins in drug-resistant gastric and pancreatic cancer and melanoma cells exhibit chaperone activities, while others are involved in drug intoxication. For some proteins participation in inhibition of apoptosis is already reported, but this role needs to be proved in the context of therapy resistance. Therefore, further functional studies are necessary to investigate the role of individual proteins and sets of proteins in chemoresistance. Proteomics is only the first step to analyze global protein expression and to find candidate proteins that may be involved in certain pathways.

11.5 References

1 M.A. Kern, H. Helmbach, M. Artuk, D. Karmann, K. Jurgowsky, D. Schadendorf. *Anticancer Res.* **1997**, *17*, 4359–4370.

2 M. Dietel, H. Arps, D. Gerding, M. Trapp, A. Niendorf. *Klin. Wochenschr.* **1987**, *65*, 507–512.

3 M. Dietel, A. Seidel. *Cancer Treat. Rev.* **1990**, *17 Suppl, A*, 3–10.

4 M. Dietel, H. Arps, H. Lage, A. Niendorf. *Cancer Res.* **1990**, *50*, 6100–6106.

5 T. Rabilloud, C. Adessi, A. Giraudel, J. Lunardi. *Electrophoresis* **1997**, *18*, 307–316.

6 T. Rabilloud. *Electrophoresis* **1998**, *19*, 758–760.

7 C.K. Cheung, Y.T. Mak, R. Swaminathan. *Ann. Clin. Biochem.* **1987**, *24*, 140–144.

8 J.E. Celis, G. Ratz, B. Basse et al. High-resolution Two-dimensional Gel Electrophoresis of Proteins: Isoelectric Focusing (IEF) and Nonequilibrium pH Gradient Electrophoresis (NEPHGE). Academic Press, San Diego, **1998**, pp 375–385.

9 V. Neuhoff, N. Arnold, D. Taube, W. Ehrhardt. *Electrophoresis* **1988**, *9*, 255–262.

10 J. Poland, D. Schadendorf, H. Lage, M. Schnolzer, J.E. Celis, P. Sinha. *Clin. Chem. Lab. Med.* **2002**, *40*, 221–234.

11 P. Sinha, J. Poland, M. Schnölzer, J.E. Celis, H. Lage. *Electrophoresis* **2001**, *22*, 2990–3000.

12 P. Sinha, J. Poland, S. Kohl et al. *Electrophoresis* **2003**, *24(14)*, 2386–2404.

13 P. Sinha, S. Kohl, J. Fischer et al. *Electrophoresis* **2000**, *21*, 3048–3057.

14 H. Lage, A. Jordan, R. Scholz, M. Dietel. *Int. J. Hypertherm.* **2000**, *16*, 291–303.

15 H. Lage, M. Dietel. *J. Cancer Res. Clin. Oncol.* **2002**, *128*, 349–357.

16 T. Tsuruo, M. Naito, A. Tomida et al. *Cancer Sci.* **2003**, *94*, 15–21.

17 M. Kuwano, T. Uchiumi, H. Hayakawa et al. *Cancer Sci.* **2003**, *94*, 9–14.

18 O. van Tellingen. *Toxicol. Lett.* **2001**, *120*, 31–41.

19 C. Meisel, I. Roots, I. Cascorbi, U. Brinkmann, J. Brockmoller. *Clin. Chem. Lab. Med.* **2000**, *38*, 869–876.

20 S.P. Cole, R.G. Deeley. *Cancer Treat. Res.* **1996**, *87*, 39–62.

21 S.P. Cole, R.G. Deeley. *Bioessays* **1998**, *20*, 931–940.

22 S.P. Cole. *J. Natl Cancer Inst.* **1999**, *91*, 888–889.

23 K. Taniguchi, M. Wada, K. Kohno et al. *Cancer Res.* **1996**, *56*, 4124–4129.

24 M. Kool, M. van der Linden, M. de Haas et al. *Proc. Natl Acad. Sci. USA* **1999**, *96*, 6914–6919.

25 J. Wijnholds, C.A. Mol, L. van Deemter et al. *Proc. Natl Acad. Sci. USA* **2000**, *97*, 7476–7481.

26 L.A. Doyle, W. Yang, L.V. Abruzzo et al. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 15665–15670.

27 R. Allikmets, W.W. Wasserman, A. Hutchinson et al. *Gene* **1998**, *215*, 111–122.

28 K. Miyake, L. Mickley, T. Litman et al. *Cancer Res.* **1999**, *59*, 8–13.

29 P. Borst, R. Evers, M. Kool, J. Wijnholds. *J. Natl Cancer Inst.* **2000**, *92*, 1295–1302.

30 H. Helmbach, E. Rossmann, M. Kern, D. Schadendorf. *Int. J. Cancer* **2001**, *93(5)*, 617–622.

31 J. Benard, O. Rixe. *Presse Med.* **1996**, *25*, 1724–1730.

32 G.H. Mickisch. *Urologe A* **1996**, *35*, 370–377.

33 D. Schadendorf, B.M. Czarnetzki. *Hautarzt* **1994**, *45*, 678–684.

34 H. Lage, H. Helmbach, M. Dietel, D. Schadendorf. *Br. J. Cancer* **2000**, *82*, 488–491.

35 M.L. Slovak, J.P. Ho, S.P. Cole et al. *Cancer Res.* **1995**, *55*, 4214–4219.

36 G.L. Scheffer, A.B. Schroeijers, M.A. Izquierdo, E.A. Wiemer, R.J. Scheper. *Curr. Opin. Oncol.* **2000**, *12*, 550–556.

37 P. Sonneveld. *Pathol. Biol. (Paris)* **1999**, *47*, 182–187.

38 P. Sonneveld. *J. Intern. Med.* **2000**, *247*, 521–534.

39 B. Tan, D. Piwnica-Worms, L. Ratner. *Curr. Opin. Oncol.* **2000**, *12*, 450–458.

40 M.M. van den Heuvel-Eibrink, P. Sonneveld, R. Pieters. *Int. J. Clin. Pharmacol. Ther.* **2000**, *38*, 94–110.

41 H. Lage, M. Christmann, M.A. Kern et al. *Int. J. Cancer* **1999**, *80*, 744–750.

42 H. Lage, M. Christmann, M.A. Kern et al. *Int. J. Cancer* **1999**, *80*, 744–750.

43 M. Christmann, M. Pick, H. Lage, D. Schadendorf, B. Kaina. *Int. J. Cancer* **2001**, *92*, 123–129.

44 H. Eid, L. Geczi, A. Magori, I. Bodrogi, E. Institoris, M. Bak. *Anticancer Res.* **1998**, *18*, 3059–3064.

45 H. Lage. *Arzneimitteltherapie* **1999**, *17*, 39–44.

46 C. Sarto, P.A. Binz, P. Mocarelli. *Electrophoresis* **2000**, *21*, 1218–1226.

47 E.M. Creagh, D. Sheehan, T.G. Cotter. *Leukemia* **2000**, *14*, 1161–1173.

48 P. Mehlen, K. Schulze-Osthoff, A.P. Arrigo. *J. Biol. Chem.* **1996**, *271*, 16510–16514.

49 J. Buchner. *Trends Biochem. Sci.* **1999**, *24*, 136–141.

50 L.H. Pearl, C. Prodromou. *Curr. Opin. Struct. Biol.* **2000**, *10*, 46–51.

51 P.K. Srivastava, H. Udono, N.E. Blachere, Z. Li. *Immunogenetics* **1994**, *39*, 93–98.

52 J. Galea-Lauri, A.J. Richardson, D.S. Latchman, D.R. Katz. *J. Immunol.* **1996**, *157*, 4109–4118.

53 R. Krishna, L.D. Mayer. *Eur. J. Pharm. Sci.* **2000**, *11*, 265–283.

54 J.A. Moscow, K.H. Dixon. *Cytotechnology* **1993**, *12*, 155–170.

55 D. Nielsen, C. Maare, T. Skovsgaard. *Gen. Pharmacol.* **1996**, *27*, 251–255.

56 D.M. van der Kolk, E. Vellenga, M. Muller, E.G. de Vries. *Adv. Exp. Med. Biol.* **1999**, *457*, 187–198.

12
Diabetes Mellitus: Complex Molecular Alterations

Gerhard Schmid and Jean-Charles Sanchez

12.1
Introduction

The term diabetes mellitus describes a complex metabolic disorder of multiple origin characterized by chronic hyperglycemia with disturbances of carbohydrate, fat, and protein metabolism resulting from defects in insulin secretion and action.

An estimated 16 million Americans, about 6% of the US population, have diabetes. The World Health Organization estimates that 120–140 million people suffer from diabetes worldwide. This number may well double by the year 2025. The increase will mostly be due to population aging, unhealthy diets, obesity, and a sedentary lifestyle. The rapid rise of obesity among children in recent years is an alarming trend. Elevated blood glucose levels cause chronic complications in the form of eye and kidney diseases, nervous system damage, and increased risk of coronary artery disease and stroke. The costs for the treatment of both diabetes and its complications are enormous ($\sim$US$ 1 billion per year in Switzerland).

There are two main variants of diabetes mellitus, type 1 and type 2. Type 1 diabetes mellitus (T1DM) is an organ-specific autoimmune disease with destruction of the insulin producing pancreatic beta-cells in the islets of Langerhans. The disease commonly occurs in childhood and has a relatively acute onset.

In type 2 diabetes mellitus (T2DM) impaired insulin secretion by the beta-cells and reduced sensitivity of the target tissues (muscle, fat, and liver cells) towards the action of insulin act together. Overt diabetes arises when beta-cells can no longer compensate for the insulin resistance. T2DM accounts for more than 90% of the diabetic patients. The disease usually occurs later in life and has a more insidious onset. As with other common diseases (cancer or cardiovascular diseases), multiple genes are involved, contributing to both impaired insulin secretion and insulin resistance. An increasing number of T2DM susceptibility genes have been discovered over the past decade. Collectively, they are estimated to account for only 10–20% of total susceptibility to T2DM [1]. Only for certain rare monogenic subtypes of T2DM, such as maturity onset diabetes of the young (MODY), the genetic basis is increasingly well understood. Considerable evidence indicates that environmental and other factors, including diet, stress, physical activity, obesity, and aging, also play important roles in the development of the disease.

Biomedical Application of Proteomics
Edited by J.-C. Sanchez, G. L. Corthals, D. F. Hochstrasser

ISBN: 3-527-30807-5

In conclusion, both T1DM and T2DM are complex heterogeneous multifactorial and polygenic diseases, resulting from gene-environment interactions [2].

This chapter reviews the current knowledge on the molecular alterations associated with the development of impaired insulin secretion and action. Gene expression-based technologies are ideally suited to study such alterations in disease. Proteomic approaches in particular are very promising since altered protein expression is not necessarily predictable from a genomic analysis [3]. Proteomic analysis allows the possibility of characterizing the protein expression profile in a cell or tissue at a given time point, reflecting the metabolic and functional status at that exact time. In contrast to genomics, proteomics takes into account the dynamic nature of the proteome of a cell or a tissue.

12.1.1 Glucose Homeostasis

Normal glucose homeostasis in the human body represents a balance between glucose appearance and tissue glucose uptake and utilization. This balance is tightly regulated. Therefore, plasma glucose concentrations are maintained within a narrow range: 60–110 mg/dl for the normal fasting state and $< 140\ \mathrm{mg\ dl^{-1}}$ for 2-hour post-glucose-load. Three physiologic processes need to interact in a highly coordinated way to maintain glucose homeostasis: beta-cell insulin secretion, tissue glucose uptake, and hepatic glucose production.

In the fasting state, hepatic glucose production via glycogenolysis and gluconeogenesis is almost exclusively responsible for the plasma glucose levels. During fasting, tissue glucose uptake is predominantly insulin-independent and occurs mostly in tissues that require glucose such as the central nervous system. In order to maintain fasting glucose concentrations constant, the rate of tissue glucose uptake dictates the rate of hepatic glucose production.

Following a meal (absorptive state), plasma glucose levels rise and promote hepatic glucose uptake. Elevated glucose levels stimulate the release and production of insulin by the beta-cell, which acts as a glucose sensor. Meaningfully, hepatic glucose production is suppressed by increased plasma insulin concentrations. Peripheral glucose uptake, mostly in the muscle, is also stimulated by insulin. In this way, hyperglycemia is minimized and the return of mealtime glycemic levels to pre-meal values is ensured [4].

Free fatty acid (FFA) metabolism also plays an important role in maintaining glucose homeostasis in the fasting and absorptive state. Elevated rates of fat breakdown (lipolysis) lead to an increased release of FFAs. These have a detrimental action on the uptake of insulin by the liver, which in turn results in increased hepatic gluconeogenesis. In addition, elevations of FFA inhibit glucose uptake and utilization in insulin-sensitive tissues. Increased plasma insulin levels after a meal inhibit lipolysis and FFA release from adipose tissue.

12.1.2 The Islets of Langerhans

Molecular alterations in the pancreatic islets are of special interest, since the insulin-producing beta-cell obviously plays a central role in the development of both T1DM and T2DM. Therefore, the pancreatic islets of Langerhans and the pancreatic beta-cell are described here briefly.

More than 100 years ago, the islets were first described by Paul Langerhans as "heap and clusters of cells, entangled in a very vascularized network, rather surprisingly in the abdominal salivary gland". Nowadays, our knowledge about the complex organization of the islets of Langerhans has increased drastically. The pancreas is an elongated organ nestling next to the first part of the small intestine. The bulk of the pancreas is composed of pancreatic exocrine cells and their associated ducts, which secrete enzymes into the digestive tract. The islets of Langerhans are scattered amidst the exocrine tissue and constitute the endocrine portion of the pancreas [5].

Four types of islet endocrine cells have been described according to the main hormonal content of their secretory granules: beta-cells (insulin) accounting for 70–80% of total islet cells, alpha-cells (glucagon) accounting for 15–20%, delta-cells (somatostatin) accounting for 5%, and PP-cells (pancreatic polypeptide) accounting for 5% of total islet cells. These cells are derived from the epithelia of embryonic pancreatic ducts. In addition to their specific hormones, they also express numerous other peptides whose role is often unclear. Recently, a novel developmentally regulated islet cell, the ghrelin cell, was identified in islets of human pancreas [6]. This cell is named according to the peptide it secretes. Importantly, ghrelin was not coexpressed with any known islet hormone. Ghrelin secretion is most prominent in the stomach and the peptide has been suggested to act as a hormone to release growth hormone from the pituitary [7].

The heterogeneity of the islet cellular composition depends on the location of the islets within the pancreas and is the result of the embryonic development of the organ. The mammalian pancreas originates from two separate endodermal buds, one ventral and one dorsal. The ventral bud turns dorsally and approaches the dorsal bud, with which it fuses. The ventral bud becomes part of the head of the pancreas, which represents the PP-rich duodenal lobe. The dorsal bud gives rise to the body and the tail and comprises the glucagon-rich splenic lobe.

Interestingly, the different cell types within an islet are not randomly distributed. The centrally located polyhedral beta-cells are surrounded by a mantle of non-beta-cells (alpha-, delta-, and PP-cells) 1–3 cells thick. This pattern of organization is based on intrinsic qualities of the cell adhesion molecules expressed on the surfaces of the different cell types, as shown by reaggregation studies of single dispersed islet cells.

Islets are highly vascularized. Although islets comprise only 1–2% of the pancreatic mass, they receive 10–20% of the pancreatic blood flow. The microvascularization of the islets is also heterogeneous and related to their size. In the rat 8–10 beta-cells are arranged in the form of a rosette around a central venous capillary. The insulin-containing secretory granules were shown to be polarized toward this

central venous capillary. The outer face of each beta-cell in the rosette is against an arterial capillary which is arranged perpendicularly to the central capillary. The lateral interfaces of the beta-cells build a canalicular system that extends from the arterial to the venous capillary. The canaliculi contain microvilli that are enriched in glucose transporters (Glut-2). Thus, these canaliculi may serve as the initial interface for glucose sensing by the beta-cell.

Data obtained by administration of exogenous islet hormones to isolated islets or to perfused pancreas preparations indicate that islet hormones influence the other islet cells in feedback loops. Thus, insulin can inhibit both alpha- and delta-cells, similarly, somatostatin can inhibit both alpha- and beta-cells, and glucagon can stimulate both beta- and delta-cells. These data have suggested that islet hormone secretion is regulated by specific intra-islet interactions [8]. These interactions can occur either directly from cell to cell via junctional communication, blood-borne via the vasculature, or in a paracrine manner via the interstitium.

Islet cell mass increases considerably from the embryo to the adult. Despite impressive progress that has been made in the clarification of the molecular events that determine the fate of pluripotent cells in the embryo and direct them toward specific functions, the knowledge regarding pancreatic development remains limited [9]. Indeed, many tissues formerly thought to be finally differentiated and therefore impossible to be replaced in the adult are plastic and renewable.

We can summarize that the islets of Langerhans are composed of various components and represent well-organized micro-organs. Islet mass is dynamic and changes both with growth and development and with functional challenges. Islets function both singly and in concert. Much needs still to be learned to develop more meaningful interpretations of the physiologic and pathophysiologic characteristics of islet function.

12.1.3 The Pancreatic Beta-Cell

Within the islets of Langerhans, the beta-cells are by far the most abundant. Through secretion of appropriate insulin quantities in response to the actual glucose levels, the beta-cells are mainly responsible for the maintenance of the body's glucose levels within a very narrow range. Therefore, the pancreatic beta-cell is central in the pathophysiology of diabetes mellitus.

Like the other pancreatic endocrine cells, beta-cells develop from endodermal progenitor cells. The most significant growth of beta-cell mass occurs at a period of late fetal life. The identity of the progenitor cells and of the transcription factors involved in the generation of functional beta-cells is of enormous interest and considerable research effort is spent in this field [10]. Throughout the lifetime of a mammal, low levels of beta-cell replication (2–3% per day) seem to balance beta-cell loss by apoptosis [11]. This results in a slowly increasing mass [12]. Unfortunately, assessment of beta-cell mass is not a straightforward method and caution must be taken when looking at such data. Any limitations on beta-cell replacement are thought to have direct consequences for glucose homeostasis.

12.2
Molecular Alterations in the Pathogenesis of T1DM and T2DM

12.2.1
Type 1 Diabetes Mellitus

T1DM results from an absolute deficiency of insulin due to the autoimmunological attack and destruction of the beta-cells. The present concept of the pathogenesis of T1DM is mainly based on evidence from studies on the animal models of human T1DM. The diabetes-prone BioBreeding (BB-DP) rat spontaneously develops a diabetic syndrome with characteristics similar to the human T1DM. In the non-obese diabetic mice model (NOD), autoimmune diabetes can be transferred from a diabetic to a non-diabetic animal via T lymphocytes. This confirms that the T lymphocytes are the mediators of autoimmune diabetes in NOD mice. It seems that both glutamate decarboxylase-(GAD) and insulin-reactive T cells are diabetogenic in the NOD mice. Antibodies to the above antigens can be observed long before the manifestation of hyperglycemia. The major histocompatibility complex class II allele, HLA-DQ, is the major genetic element, conferring susceptibility to or protection from human T1DM. However, the mechanisms determining the contribution of the HLA-DQ allele remain to be elucidated [13]. Accumulating evidence suggests that the regulation of the gut immune system may be aberrant in T1DM, possibly due to environmental factors [14].

Many autoimmune diseases are characterized by an overproduction of cytokines. In the case of T1DM, interferon-γ (IFNγ), tumor necrosis factor α (TNFα) and interleukin 1β (IL-1β) have been shown to contribute to the pathology. In addition, the quality of the immune reactivity is an important factor in beta-cell destruction. In the animal model, only infiltration of the islets by helper T cells of the subtype Th1, which secret high levels of INFγ, caused beta-cell damage [15]. According to the so-called "Copenhagen model" of T1DM , beta-cells are destroyed by cytokine-induced free radical formation before cytotoxic helper T cells and/or autoantibody-mediated cytolysis [16, 17]. A consequence of this model is that proinflammatory cytokines induce both protective and deleterious mechanisms in all cells expressing cytokine receptors. In genetically predisposed individuals, T1DM development occurs because the deleterious events in the beta-cell prevail.

12.2.2
Type 2 Diabetes Mellitus

The glucose homeostasis dysregulation in T2DM is mainly due to a combination of two physiological defects: impaired insulin secretion and reduced peripheral and hepatic insulin sensitivity [18] (Fig. 12.1). The development of T2DM is a gradual process, and the relative roles of the two defects in the etiology of the disease remain controversial. However, it is now generally agreed that overt diabetes arises when beta-cells can no longer compensate for the insulin resistance. In the following, impaired insulin secretion and insulin resistance, which are physiologically interconnected will be analyzed in more detail.

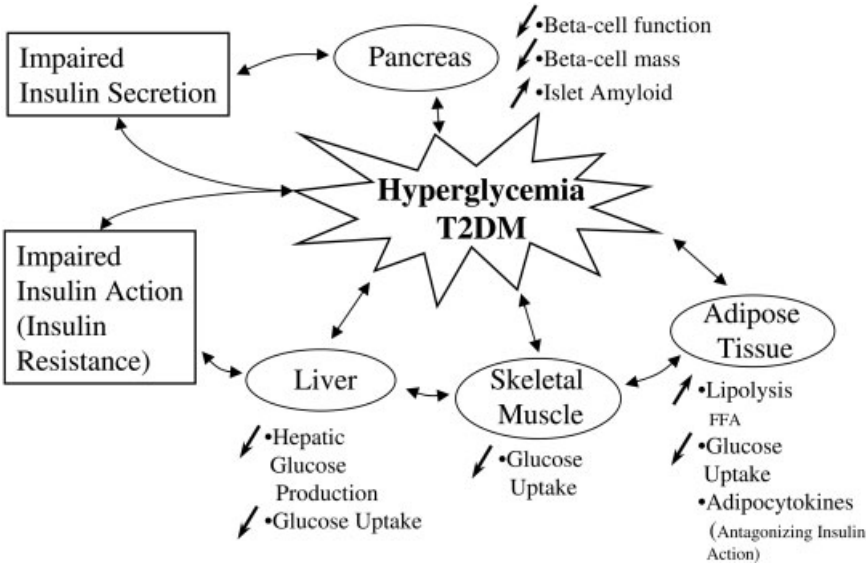


Figure 12.1 Pathophysiology of type 2 diabetes mellitus. T2DM is characterized by two physiological defects: impaired insulin secretion and impaired insulin action (insulin resistance). Whether the primary defect initiating the glucose intolerance resides in the beta-cell or in the peripheral tissues remains controversial. Abnormalities in the four major organ systems involved in glucose homeostasis contribute to the hyperglycemia observed in T2DM. Knowledge of the precise molecular alterations associated with these abnormalities will be of great help for a successful disease prevention and treatment.

12.2.2.1 **Impaired Insulin Secretion**

The pancreatic beta-cell has the ability to sense glucose levels in order to adapt insulin secretion to blood glucose fluctuations. This is accomplished by the expression of a particular profile of carbohydrate transporters and enzymes in the beta-cell [19]. There is substantial knowledge about the molecular mechanism of glucose metabolism in the beta-cell and how this metabolism is coupled to insulin secretion. However, our understanding is still incomplete. For example, the precise role of mitochondrial metabolism and mitochondrial messengers in the stimulation of insulin exocytosis is only starting to emerge [20].

T2DM is characterized by a progressive loss of beta-cell function. An initial defect is the near-absent first-phase glucose-induced insulin secretion, followed by impaired second-phase insulin secretion and increased release of pro-insulin relative to insulin (hyperproinsulinemia). Last, defective basal insulin secretion develops, leading to complete beta-cell failure.

Any defect in the uptake of glucose, its metabolism or in the coupling to insulin secretion could be responsible for the reduced beta-cell response to the glucose stimulus. In addition, there is evidence that beta-cell dysfunction is also implicated in the reduction of the beta-cell mass which is frequently observed in T2DM (Fig. 12.2).

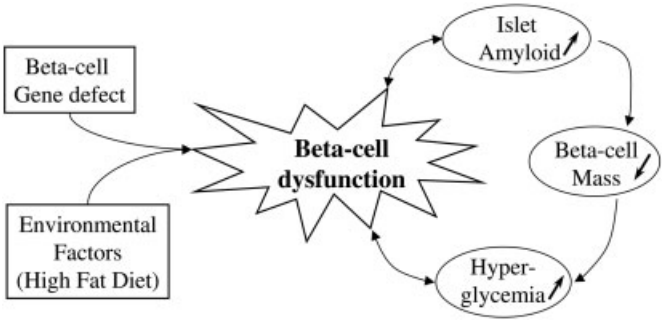


Figure 12.2 Model of the development of beta-cell dysfunction. A combination of beta-cell gene defects with environmental factors such as high-fat diet with chronically increased FFA levels leads to an initial beta-cell dysfunction. This dysfunction is associated with altered insulin secretion resulting in the development of hyperglycemia. In addition, altered beta-cell function is critical for the initial formation of islet amyloid fibrils. Progressive amyloid deposition and the eventual replacement of beta-cell mass by amyloid result in increased hyperglycemia. This in turn further aggravates beta-cell function.

Candidate Genes In some rare cases, the specific cause for the impaired insulin secretion is known.

For some MODY subtypes (rare monogenic forms of T2DM), the disease could be linked to mutations in the beta-cell glucose sensor glucokinase (in MODY2) [21] or to mutations in the transcription factor hepatocyte nuclear factor 1 alpha, which controls the expression of insulin and the glucose transporter GLUT2 (in MODY3) [22]. Similarly, in mitochondrial diabetes, mutations in mitochondrial DNA were shown to cause a defect in the glucose-induced insulin secretion [23, 24]. The effects of these mutations on insulin secretion are so severe that diabetes can develop in people with normal or relatively normal insulin sensitivity and without a major contribution of environmental factors. However, these and other reported mutations [1, 25] are rare and thus not common causes of impaired insulin secretion and subsequent development of T2DM.

Beta-cell Dysfunction In common T2DM, more moderate abnormalities of insulin secretion are thought to be present. They cause glucose intolerance only if insulin resistance is also present. Complex interactions between genes and environment determine whether diabetes will develop.

In the pre-diabetic state, there is an inverse relationship between insulin sensitivity and insulin secretion [26]. The hyperinsulinemia associated with insulin resistance results from a combination of an increased insulin secretion and a reduced insulin clearance rate. The compensatory hypersecretion of insulin is due to an expansion of the beta-cell mass [27] and changes in the expression of key enzymes in the beta-cell glucose metabolism [28]. However, the mechanisms underlying this compensatory hypersecretion remain unknown [29].

In T2DM, an overall reduction in insulin secretion is observed since beta-cells can no longer secrete sufficient insulin to maintain normal blood glucose levels.

Development of beta-cell dysfunction and progression to complete beta-cell failure was attributed to a combination of beta-cell gene defects and environmental factors such as high fat diet leading to chronically increased FFA levels [26]. According to this model, beta-cell dysfunction is associated with the development of islet amyloid and subsequent reduction of the beta-cell mass. The resulting hyperglycemia further aggravates beta-cell function (Fig. 12.2).

Autopsy studies revealed a reduction of the beta-cell mass of up to 50% in T2DM. Beta-cell loss has been associated with amyloid deposition [30]. Islet amyloid consists of fibrils formed by islet-amyloid polypeptide (IAPP or amylin), which is co-secreted with insulin. Altered beta-cell function results in a change in the production, processing, and/or secretion of IAPP, which in turn is critical for the initial formation of islet amyloid fibrils. Patients with islet cell tumors are characterized by relative hyperproinsulinemia and pancreatic islet amyloid deposits, supporting a causative role of impaired IAPP processing in amyloid formation [31]. Initial fibril deposition then allows the progressive accumulation of IAPP-containing fibrils and the eventual replacement of beta-cell mass by amyloid, resulting in hyperglycemia. Beta-cell dysfunction seems to be a central point in this model since no major clinical hyperglycemia would be expected from the beta-cell mass loss alone. The fibrils themselves may be toxic to beta-cells and contribute to the dysfunction. Hyperglycemia *per se* alters beta-cell function and stimulates IAPP production.

Once established, hyperglycemia exerts a deleterious effect on insulin secretion. Many studies have provided evidence that an acquired defect in insulin secretion is related to glucose toxicity. It is hypothesized that chronic oxidative stress is an important mechanism for glucose toxicity [32]. The increased expression of antioxidant and antiapoptotic genes in islets may contribute to beta-cell survival during chronic hyperglycemia [33]. In addition, antioxidant treatment protected beta-cell mass in diabetic mice [34]. It has been shown that chronic exposure to a high level of glucose impaired potassium-channel function [35], reduced glucose transporter (GLUT-2) expression [36] or led to loss of part of the beta-cell's differentiated phenotype [37]. However, glucose stimulates growth of the islet beta-cell mass and glucose protects against apoptosis [38]. This implies that glucose not only regulates the phenotype of existing beta-cells, but also regulates cell number. One can conclude that glucose has a pleiotropic effect on beta-cells. Its molecular understanding represents a major challenge in diabetes research.

12.2.2.2 **Impaired Insulin Action: Insulin Resistance**

As outlined in the section on glucose homeostasis, in the healthy state, insulin promotes glucose uptake in the peripheral target tissues skeletal muscle and adipose tissue. In addition, insulin inhibits glucogenolysis and gluconeogenesis in the liver. In the case of insulin resistance, the ability of insulin to activate its signaling pathways is reduced. Consequently less glucose is taken up by the target tissues and basal hepatic glucose production is increased. Interestingly, insulin sensitivity varies widely in the normoglycemic population. For normal beta-cell

function, insulin resistance may be present for many years before the onset of overt hyperglycemia, because of a compensatory secretory response of the beta-cell. However, in individuals predisposed to T2DM, beta-cells either fail to compensate the developing insulin resistance or fail after compensating for some time due to exhaustion. Any defect in the control of hepatic glucose production and/or peripheral glucose uptake could result in insulin resistance.

Candidate Genes Indeed, in some cases, the specific cause for the impaired insulin action is known. The insulin receptor (IR), a ligand-stimulated tyrosine kinase, mediates the first step in insulin action. Consequently, the IR gene has been considered as the primary candidate gene contributing to insulin resistance in T2DM [39]. Another strong candidate gene for association with T2DM is the insulin receptor substrate (IRS), which is phosphorylated by the activated receptor. IRS phosphorylation leads to activation of multiple downstream signaling pathways and finally culminates in the translocation of glucose transporters to the cell membrane and in the stimulation of glycogen synthesis. It has to be said that this complex signaling pathway is still far from being completely understood [40]. Four IRS isoforms are currently known. The prevalence of mutations in IRS-1 was found to be higher in T2DM patients than in control subjects. Interestingly, disruption of IRS-2 in mice impaired both peripheral insulin signaling and beta-cell function [41]. Unfortunately, results from studies on IRS-2 in humans were rather controversial, and IRS-2 polymorphisms do not seem to be a major genetic susceptibility factor for T2DM [42]. The same holds true for mutations in other candidate genes reported so far. These include glucose transporters (GLUTs), glycogen synthase (GSY), beta-3-adrenergic receptor (β-3AR), hexokinase II, fatty acid-binding protein (FABP2), leptin, peroxisome proliferator-activated receptor gamma (PPARγ), prohormone convertase 2 (PC-2), tumor necrosis factor α (TNFα) or Ras associated with diabetes (RAD) [25]. Therefore, the genetic basis for insulin resistance may derive from the multiplicative effects of common polymorphisms affecting several steps in the insulin signaling pathway, each causing only a modest functional impairment individually.

The Role of Adipose Tissue Besides the genetic component, many environmental factors, including obesity, physical activity, diet, and age, also influence sensitivity to insulin.

Obesity is defined by a body mass index (weight divided by square of the height) of 30 kg m^{-2} or greater. About 60–90% of all patients with T2DM are or have been obese. The reverse, however, is not true, and the majority of obese patients do not become diabetic [43]. But, the risk of diabetes was shown to be increased 5-fold for those with a BMI of 25, 28-fold for those with BMI of 30, and 93-fold for those women with a BMI of 35 or greater, compared with women with a BMI of less than 21 [44]. Insulin resistance is the dominating factor in obesity.

Free fatty acids (FFAs) are released from adipose triglyceride stores into the circulation by lipolysis. They are often elevated in obese individuals and have emerged as a major link between obesity and insulin resistance and T2DM [45].

In skeletal muscle, it is hypothesized that FFAs interfere with normal insulin signaling by activation of the protein kinase C (PKC) [46, 47]. In the liver, an FFA-induced increase in glucose production can be observed. The mechanisms responsible for such a diabetogenic effect are complex and incompletely defined [48]. In the beta-cell however, FFAs potentiate glucose-stimulated insulin secretion. It is thought that this prevents the development of T2DM in the majority of obese insulin-resistant people.

In addition to the release of FFAs, adipocytes also secrete a variety of proteins into circulating blood that affect whole-body homeostasis. These proteins, collectively known as adipocytokines, include, for example, TNF*α*, leptin, resistin, and adiponectin [49–51].

TNF*α*: Expression of TNF*α* was found to be increased in adipocytes and muscle tissue of obese individuals, where the degree of expression positively correlated with the degree of obesity and insulin resistance. TNF*α* is considered as a major contributor to the development of the peripheral tissue resistance to insulin [52, 53]. It induces insulin resistance directly by interfering with the tyrosine kinases activity of the insulin receptor and by down-regulating the glucose transporter 4 (GLUT4) and indirectly by stimulating lipolysis of fat cells resulting in FFA release [54, 55]. TNF*α* as well as insulin stimulate the secretion of leptin from adipocytes.

Leptin: The product of the *ob* gene serves as a signal that provides information about the size of the energy reserves to systems that are regulating feeding, substrate utilization, and energy balance. Therefore, leptin levels are higher in obese individuals and increase with overfeeding. Leptin-deficient (*lep/lep*) and leptin receptor-deficient (*db/db*) mice are obese and exhibit severe insulin resistance, which can be reversed in the *lep/lep* mouse by leptin administration. To what extent the net hypoglycemic effect of leptin is direct and therefore independent of its weight-reducing effects remains to be elucidated.

Adiponectin: Another circulating protein specifically secreted by the adipose tissue is adiponectin (additional names include AdipoQ, apM1, GBP28, Acrp30). Originally, a reduction in adiponectin gene expression was observed in adipose tissue of obese mice and humans compared with lean controls [56]. More recently, plasma adiponectin levels were shown to be significantly decreased in obese [57] and T2DM [58] subjects. Low plasma adiponectin may, therefore, contribute to the pathogenesis of insulin resistance and T2DM. However, plasma concentrations of most proteins from adipose tissue are increased in obesity because of an increase in the total body fat mass. TNF*α*, which is increased in obesity, was proposed as a candidate molecule responsible for decreased adiponectin levels in obesity. The expression and secretion of adiponectin from adipose tissue were found to be significantly reduced by TNF*α* [59, 60].

In animal models of obesity and T2DM, recombinant adiponectin was shown to work as an enhancer of insulin action in muscle and liver [61, 62]. The hypoglycemic effect was not associated with an increase in insulin levels. Endogeneous adi-

ponectin gene expression correlated with insulin sensitivity. Adiponectin was proposed to decrease insulin resistance in mice by increasing tissue fat oxidation, which leads to decreased triglyceride content in muscle and liver and decreased FFA levels in plasma [62].

The receptor for adiponectin and the downstream signaling pathway are currently unknown. Moreover, the role of adiponectin in insulin resistance in humans remains to be clarified [63].

Resistin: Expression of resistin (product of the *Retn* gene) was found to be induced during adipogenesis *in vitro*. Consequently, serum resistin levels were raised in *lep/lep* and *db/db* and in diet-induced models of obesity and T2DM [64]. Interestingly, *Retn* mRNA levels are reduced in adipose tissue from obese mice, suggesting that net increase in resistin is due to increased number of adipocytes in obesity. As for adiponectin, the nature of the resistin receptor is still unknown and a causative association of resistin with insulin resistance in humans has not been shown yet [65, 66]. A putative human homolog of resistin shows only 59% amino acid identity to the mouse protein and the expression in human adipose tissue is much less than that observed for the *Retn* gene in mouse adipose tissue [67, 68].

12.3 The Treatment of Diabetes Mellitus

The goal of a diabetes treatment seems simple, a tight control of the blood glucose levels is sufficient. However, consideration of the complex regulation of blood glucose in humans clearly indicates that a tight control of blood glucose is not an easy task.

In T1DM plasma insulin levels are low or absent and treatment with insulin is always necessary. Since it is not yet possible to mimic the normal function of a beta-cell, which precisely adjusts the rate of insulin secretion in response to biological needs, the treatment of T1DM is complicated by episodes of hypoglycemia and hyperglycemia. Hypoglycemia can cause coma, and hyperglycemia increases the risk for diabetic complications in the form of eye disease, kidney disease, nervous system damage, and coronary artery disease and stroke [69].

In T2DM there is a wide spectrum in the degree of impaired insulin secretion and action. First, patients are separated into an obese and non-obese group, because weight loss is a crucial factor [70]. Then three categories can be established for both groups. First, patients with sufficient beta-cell mass and insulin sensitivity, who can maintain normal glucose levels through the control of energy intake and expenditure via diet and physical activity. Second, patients who require oral antidiabetic drugs. Third, patients who need exogeneous insulin to control their glycemia. The oral antidiabetic drugs can be separated according to their mechanism of action. Sulfonylureas stimulate insulin secretion by acting directly on beta-cells [71]. Glucosidase inhibitors delay the absorption of glucose in the small

intestine and therefore attenuate mealtime hyperglycemia [72]. Both biguanides and thiazolidinediones (TZDs) increase insulin sensitivity, although their mechanism and site of action are different. These agents may be used alone or in combination with each other or with insulin [73]. In the following we will focus on the TZDs, since they seem to be most promising to provide a long-term glycemic control. In addition, they are less likely than sulfonylureas to provoke hypoglycemia [74].

The commercialized TZDs include troglitazone (Rezulin), pioglitazone (Actos), and rosiglitazone (Avandia) [75]. TZDs are synthetic activators of the PPARγ. This nuclear receptor heterodimerizes with the retinoid X receptor (RXR) and regulates transcription of a number of genes involved in lipid metabolism and adipocyte differentiation.

To date, the precise mechanisms underlying the actions of TZDs are largely unknown [76]. PPARγ is expressed at a much higher level in adipose tissue than in muscle and liver. However, TZDs enhance insulin sensitivity also in muscle and liver. This suggests that the primary effect of TZDs occurs in adipose tissue and is transmitted to muscle and liver via mediators, such as TNFα, leptin, adiponectin, and FFAs [50]. Both TNFα and leptin expression are reduced upon PPARγ activation. In turn, PPARγ activation induces lipoprotein lipase, thereby increasing triglyceride uptake by adipose tissue and reducing circulating FFAs, which then reduces insulin resistance in the liver and the muscle. In addition, rosiglitazone treatment significantly increased plasma adiponectin levels in T2DM subjects [77]. Interestingly, TZDs seem to reduce insulin resistance even in the absence of adipose tissue [78]. Enhanced expression of GLUT4 may be another mechanism contributing to the improved insulin action after treatment with TZDs [79].

In addition to their role as insulin sensitizers, TZDs have been shown to have a protective effect on pancreatic islets cell structure in mice, which seems to be potentially important to the long-term efficacy of rosiglitazone [80]. TZDs do not directly stimulate insulin secretion. Recent studies in animals and humans suggest that TZD treatment may influence pancreatic beta-cell function [81–83]. Although the mechanisms underlying these effects remain to be elucidated, it seems that they may be mediated indirectly by a decrease in insulin resistance, by reduction in FFAs, or by a decrease in islet triglyceride content [81].

12.4 Proteomics: a Global Approach to the Study of Diabetes Mellitus

Proteomic approaches aim at the qualitative and quantitative comparison of proteomes under different conditions. A proteome is the ensemble of proteins expressed by a genome of a species, an organ, or a cell at a particular moment under particular conditions. Approaches include comparative analysis of global protein expression in normal and disease tissues to identify aberrantly expressed proteins. In the recent years, several groups took advantage of the power of a proteomic approach for their studies in the field of diabetes. Most studies have been car-

ried out in rodent animal models of diabetes. Because of species differences between humans and rodents, especially in adipocyte function [50], an increased use of human samples from diabetic patients would be desirable.

12.4.1 Type 1 Diabetes Mellitus

A series of proteomic studies were performed aiming at the identification of proteins with possible deleterious and protective roles in the initial cytokine-induced beta-cell damage in T1DM. The cytokine IL-1β inhibits insulin release and is selectively cytotoxic to beta-cells in isolated pancreatic rat islets. IL-1β induces nitric oxide (NO) via the enzyme inducible nitric oxide synthase (iNOS) and NO seems to be a major effector molecule in IL-1β-induced beta-cell damage. For the proteomic studies, pancreatic islets isolated from neonatal Wistar Furth (WF) rats were incubated in the presence and absence of IL-1β. In addition, inhibitors of NO production as well as chemically generated NO were used. Since *de novo* protein synthesis was shown to be necessary for the deleterious effect of IL-1β, islets were metabolically labeled with [^{35}S]methionine. Finally, islet proteins were separated by two-dimensional (2-D) gel electrophoresis. Overall, 105 protein spots were found to be modulated by IL-1β exposure of islets in culture [84, 85]. Protein identification could be made for 60 of them by peptide mass fingerprinting [86].

In a similar study, islets from diabetes-prone BioBreeding (BB-DP) rats were exposed to IL-1β. Overall, 82 protein spots were found to be modulated by IL-1β exposure [87]. Finally, 45 different proteins could be identified [88]. Comparison of IL-1β-exposed BB-DB and WF islets showed common changes in 14 proteins. In addition, several proteins previously not described in islets of Langerhans were identified.

The complexity of the effects of IL-1β on islets protein expression support the hypothesis that the development of T1DM is the result of a collective, dynamic instability, rather than the result of a single factor [17].

12.4.2 Type 2 Diabetes Mellitus

A mouse Swiss 2-D PAGE database has been established [89]. It contains a number of 2-D gel electrophoresis reference maps with protein identifications from mouse white and brown adipose tissue, pancreatic islets, liver and skeletal muscle. These maps can be accessed through the internet (http://us.expasy.org/cgi-bin/map1) and represent a valuable tool for further proteomic studies of diabetes and obesity.

In the following, recent proteomic studies in the field of T2DM will be presented. The studies are grouped according to the tissue that was analyzed.

12.4.2.1 Muscle

In order to investigate the molecular alterations associated with insulin resistance in muscle tissue, human skeletal muscle biopsies from patients with T2DM and from healthy controls were compared by 2-D gel electrophoresis [90]. Eight potential protein markers were identified. ATP synthase β-subunit was found to be down-regulated in skeletal muscle of patients with T2DM. In addition, the levels of a phosphoisoform of the down-regulated ATP synthase β-subunit correlated inversely with fasting plasma glucose levels in diabetic muscle. Therefore, the authors suggested a role for phosphorylation of ATP synthase β-subunit in the regulation of ATP synthesis. Alterations in this regulation may contribute to skeletal muscle insulin resistance in T2DM.

In a proteomic comparison analysis between obese (*lep/lep*) and lean mice, four proteins were found to be underexpressed in obese versus lean [103]. Among them were pyruvate carboxylase (PC) and 2-oxoglutarate dehydrogenase E1 component (OGDH). Underexpression of PC and OGDH in muscle in combination with the stimulation of glycolysis by hyperinsulinemia may lead to an increased availability of pyruvate and result in increased lipogenesis from glutamate.

12.4.2.2 Liver

In order to gain further insight into the molecular mechanism underlying the therapeutic action of PPAR activators, two studies recently investigated the global protein expression changes in liver following administration of PPAR activators. Edvardsson et al. [91] analyzed hepatic protein expression of lean and obese (*lep/lep*) mice treated with rosiglitazone (PPARγ agonist) and WY14643 (PPARα agonist). They found that obese mice display higher levels of enzymes involved in fatty acid oxidation and lipogenesis than lean mice. These differences were further amplified by treatment with both PPAR activators. Their approach allowed them to discriminate between effects caused by treatment with agonists of the closely related PPARα and PPARγ.

White et al. [92] looked at protein changes in rat liver upon treatment with troglitazone (PPARγ agonist), WY14643, and a compound with mixed PPARα/γ agonist activity.

Fourteen proteins were found to be underexpressed in liver of obese (*lep/lep*) versus lean mice [103]. Seven of them were modulated back to levels of lean by rosiglitazone treatment of the obese mice. The association of the reactivation of proteins such as alcohol dehydrogenase, HSP60, or glycine *N*-methyltransferase by rosiglitazone treatment remained unclear.

12.4.2.3 Adipose Tissue

A proteomic approach was used for the identification of secreted factors during the differentiation of preadipocytes (3T3-L1) to adipocytes *in vitro* [93]. Besides conventional gel-based protein separation combined with tandem mass spectrome-

try, a liquid chromatography-based separation followed by automated tandem mass spectrometry was also used. Several additional secreted proteins, including resistin, were identified by this method.

In white and brown adipose tissue of mice, four and five underexpressed and three and two overexpressed polypeptides were observed in obese (*lep/lep*) relative to lean [103]. Rosiglitazone treatment decreased the expression of an albumin fragment and of galectin 1 in obese mice to the level of controls in white and brown adipose tissue, respectively.

12.4.2.4 **Pancreatic Islets**

In a proteomic analysis of normal mouse pancreatic islets, proteins implicated in Alzheimer's disease (AD) were found to be highly expressed [94]. Although the relevance of AD-related proteins in islets remained unclear, the authors speculated about a potential parallel between the pathophysiology of T2DM and AD.

In pioneering studies, glucose-responsive proteins in pancreatic islets from rats were analyzed using 2-D gel electrophoresis [95, 96]. *In vitro* labeling with [^{35}S]methionine allowed the detection of many proteins that showed a differential expression between islets cultured at low and high glucose levels. However, no protein identifications could be made. The same is true for a more recent study, where fluorescence-activated cell sorter (FACS)-purified beta-cells from rats were analyzed [97].

The identification of new molecular targets associated both with islet cell dysfunction and protection is an important goal in diabetes research. Therefore, the effect of rosiglitazone on the differential expression of diabetes-associated proteins in pancreatic islets of obese mice was investigated using a proteomic approach [98]. In a first step, diabetes-associated proteins were identified by comparison of islet 2-D gels from lean and obese (*lep/lep*) mice (C57Bl/6J). In a second step, islets from mice treated with rosiglitazone were included in the analysis. Four of the nine diabetes-associated proteins were found to be modulated by the drug treatment, i.e. their expression levels in obese were back to normal lean levels after treatment. Increased secretion of proinsulin relative to insulin is a common feature of T2DM. Rosiglitazone treatment increased carboxypeptidase B expression in both obese and lean mice. The processing of proinsulin has been associated with the activity of carboxypeptidase E, whereas carboxypeptidase B is thought to be associated with exocrine secretion. Nevertheless, the two enzymes have the same enzymatic properties, suggesting that a primary effect of rosiglitazone may be the increase of the conversion of proinsulin to insulin through an increased expression of carboxypeptidase B.

In an attempt to confirm some of the differentially expressed proteins, islets from C57Bl/6J mice were compared with islets from C57Bl/Ks mice [98]. The two strains share 95% of their genome content. However, C57Bl/Ks mice have only half the islet cell mass of C57Bl/6J mice and are significantly more susceptible to diabetogenic factors. Interestingly, tropomyosin 1 and profilin showed similar

changes in expression in C57Bl/Ks mice and in C57Bl/6J lep/lep mice, supporting the idea that these proteins may be important to normal islet cell function.

12.5 Conclusions

Designing successful disease prevention and treatment requires both knowledge of the pathogenesis of the disease and the ability to predict which individuals are at high risk of developing it. The discovery of new genes and pathways involved in the pathogenesis of T1DM and T2DM is of major importance. However, the identification of genes that contribute to the risk of developing these diseases represents a significant challenge as both of them are complex diseases with many genetic and environmental causes.

A number of diverse approaches have been used to discover and validate potential new targets. To date, DNA-based approaches using candidate gene and genome-wide linkage analysis have had limited success in identifying genomic regions or genes involved in the development of these diseases. Multiple interactions between genes and between genes and the environment influence the development of diabetes, indicating that DNA-based approaches alone are not the ideal tool [99].

Gene expression-based technologies may prove to better suit the aim of identifying diabetes candidate genes. RNA-based technologies have been applied to various target tissues and cell lines in attempts to identify differences in specific mRNA expression and genes of relevance for diabetes [33, 100–102]. However, there is not necessarily a strict linear relationship between mRNA and protein levels. Importantly, in contrast to mRNA expression studies, protein expression studies allow the assessment of post-translational modifications, which are often necessary for protein function and may also be of relevance in the pathogenicity.

Proteomics offers a considerable potential as a hypothesis-generating tool in biomedical applications. Of course, looking at global alterations in protein expression in disease is one thing; extraction of the relevant data and understanding of the physiologically important biochemical pathways is another. Obviously, their biological significance should be addressed.

The series of proteomic studies in T1DM should soon allow a more detailed picture of the molecular processes leading to specific beta-cell destruction to be obtained. This will then offer the perspective of making the beta-cell more resistant to immunological mediators, and may increase survival time of transplanted islets with a reduced need for immunosuppressive drugs. To date, there is a less widespread application of proteomics in T2DM. Most studies have focused on protein modulation in different target tissues upon drug treatment. Nevertheless, such information can be of great value in the molecular understanding of the mechanism of action of the drug. Moreover, it may guide in developing compounds with the best therapeutic profile, and with the least toxicologic effects possible.

In future, the combined genomic and proteomic approach should lead to a better understanding of the biochemical mechanisms underlying diseases in general

and diabetes in particular. The field for the application of proteomics in diabetes research is wide open. For example, the identification of all molecular players involved in stimulus-secretion coupling pathways in the beta-cell is a major challenge in diabetes research. This will be essential for a detailed understanding of the beta-cell function and for the characterization of the molecular basis of beta-cell dysfunction in pathophysiological conditions, including diabetes. Comparative proteomic analysis of glucose-responsive and glucose-unresponsive beta-cells certainly represents a promising approach. Corresponding beta-cell lines or pancreatic islets isolated from diabetic animal models or patients could be used. Analysis of subcellular fractions such as mitochondria or insulin secretory granules might be of particular interest.

Using a similar approach, the poorly understood biochemical mechanisms in the beta-cell that account for the compensatory hyperfunction with insulin resistance (so-called beta-cell adaptation) could be studied.

12.6 References

1 Gloyn, A. L., McCarthy, M. I. *Best Pract. Res. Clin. Endocrinol. Metab.* **2001**, *15*, 293–308.
2 Bartsocas, C. S., Leslie, R. D. *Am. J. Med. Genet.* **2002**, *115*, 1–3.
3 Hanash, S. *Nature* **2003**, *422*, 226–232.
4 Porte, D., Jr. *Diabetes Metab. Res. Rev.* **2001**, *17*, 181–188.
5 Bonner-Weir, S., Smith F. E. *Joslin's Diabetes Mellitus*, 13th edn. Williams & Wilkins, Philadelphia, pp 15–27, **1994**.
6 Wierup, N., Svensson, H., Mulder, H., Sundler, F. *Regul. Pept.* **2002**, *107*, 63–69.
7 Kojima, M., Hosoda, H., Matsuo, H., Kangawa, K. *Trends Endocrinol. Metab.* **2001**, *12*, 118–122.
8 Weir, G. C., Bonner-Weir, S. *J. Clin. Invest.* **1990**, *85*, 983–987.
9 Slack, J. M. *Development* **1995**, *121*, 1569–1580.
10 Edlund, H. *Diabetes* **2001**, *50* Suppl 1, S5–9.
11 Habener, J. F., Stoffers, D. A. *Proc. Assoc. Am. Physicians* **1998**, *110*, 12–21.
12 Bonner-Weir, S. *Diabetes* **2001**, *50 Suppl. 1, S20–S24.*
13 Moustakas, A. K., Papadopoulos, G. K. *Am. J. Med. Genet.* **2002**, *115*, 37–47.
14 Akerblom, H. K., Vaarala, O., Hyoty, H., Ilonen, J., Knip, M. *Am. J. Med. Genet.* **2002**, *115*, 18–29.
15 Healey, D., Ozegbe, P., Arden, S. et al. *J. Clin. Invest.* **1995**, *95*, 2979–2985.
16 Nerup, J., Mandrup-Poulsen, T., Molvig, J. et al. *Diabetes Care* **1988**, *11 Suppl. 1, 16–23.*
17 Freiesleben De Blasio, B., Bak, P., Pociot, F., Karlsen, A. E., Nerup, J. *Diabetes* **1999**, *48*, 1677–1685.
18 Bajaj, M., Defronzo, R. A. *J. Nucl. Cardiol.* **2003**, *10*, 311–323.
19 Newgard, C. B., McGarry, J. D. *Annu. Rev. Biochem.* **1995**, *64*, 689–719.
20 Maechler, P. *Cell Mol. Life Sci.* **2002**, *59*, 1803–1818.
21 Froguel, P., Zouali, H., Vionnet, N. et al. *N. Engl. J. Med.* **1993**, *328*, 697–702.
22 Yamagata, K., Oda, N., Kaisaki, P. J. et al. *Nature* **1996**, *384*, 455–458.
23 Suzuki, S., Hinokio, Y., Hirai, S. et al. *Diabetologia* **1994**, *37*, 818–825.
24 Maassen, J. A. *Am. J. Med. Genet.* **2002**, *115*, 66–70.
25 Jun, H., Bae, H. Y., Lee, B. R. et al. *Adv. Drug Deliv. Rev.* **1999**, *35*, 157–177.
26 Kahn, S. E. *J. Clin. Endocrinol. Metab.* **2001**, *86*, 4047–4058.

27 Pick, A., Clark, J., Kubstrup, C. et al. *Diabetes* **1998**, *47*, 358–364.
28 Cockburn, B.N., Ostrega, D.M., Sturis, J. et al. *Diabetes* **1997**, *46*, 1434–1439.
29 Bell, G.I., Polonsky, K.S. *Nature* **2001**, *414*, 788–791.
30 Kahn, S.E., Andrikopoulos, S., Verchere, C.B. *Diabetes* **1999**, *48*, 241–253.
31 Porte, D., Jr., Kahn, S.E. *Diabetes* **1989**, *38*, 1333–1336.
32 Robertson, R. P., Harmon, J., Tran, P.O., Tanaka, Y., Takahashi, H. *Diabetes* **2003**, *52*, 581–587.
33 Laybutt, D.R., Kaneto, H., Hasenkamp, W. et al. *Diabetes* **2002**, *51*, 413–423.
34 Kaneto, H., Kajimoto, Y., Miyagawa, J. et al. *Diabetes* **1999**, *48*, 2398–2406.
35 Purrello, F., Vetri, M., Vinci, C. et al. *Diabetes* **1990**, *39*, 397–399.
36 Thorens, B., Wu, Y. J., Leahy, J.L., Weir, G.C. *J. Clin. Invest.* **1992**, *90*, 77–85.
37 Laybutt, D.R., Sharma, A., Sgroi, D.C. et al. *J. Biol. Chem.* **2002**, *277*, 10912–10921.
38 Hoorens, A. Van de Casteele, M., Kloppel, G., Pipeleers, D. *J. Clin. Invest.* **1996**, *98*, 1568–1574.
39 Taylor, S.I. *Diabetes* **1992**, *41*, 1473–1490.
40 Holman, G.D., Kasuga, M. *Diabetologia* **1997**, *40*, 991–1003.
41 Withers, D. J., Gutierrez, J.S., Towery, H. et al. *Nature* **1998**, *391*, 900–904.
42 D'Alfonso, R., Marini, M.A., Frittitta, L. et al. *J. Clin. Endocrinol. Metab.* **2003**, *88*, 317–322.
43 Felber, J. P., Golay, A. *Int. J. Obes. Relat. Metab. Disord.* **2002**, *26 Suppl. 2*, S39–S45.
44 Colditz, G.A., Willett, W.C., Rotnitzky, A., Manson, J.E. *Ann. Intern. Med.* **1995**, *122*, 481–486.
45 Boden, G. *Curr. Opin. Clin. Nutr. Metab. Care* **2002**, *5*, 545–549.
46 Griffin, M.E., Marcucci, M. J., Cline, G.W. et al. *Diabetes* **1999**, *48*, 1270–1274.
47 Itani, S.I., Ruderman, N.B., Schmieder, F., Boden, G. *Diabetes* **2002**, *51*, 2005–2011.
48 Lam, T.K., Carpentier, A., Lewis, G.F. et al. *Am. J. Physiol. Endocrinol. Metab.* **2003**, *284*, E863–E873.
49 Ahima, R.S., Flier, J.S. *Trends Endocrinol. Metab.* **2000**, *11*, 327–332.
50 Arner, P. *Trends Endocrinol. Metab.* **2003**, *14*, 137–145.
51 Steppan, C.M., Lazar, M.A. *Trends Endocrinol. Metab.* **2002**, *13*, 18–23.
52 Saghizadeh, M., Ong, J.M., Garvey, W.T., Henry, R.R., Kern, P.A. *J. Clin. Invest.* **1996**, *97*, 1111–1116.
53 Hotamisligil, G.S., Arner, P., Caro, J.F., Atkinson, R.L., Spiegelman, B.M. *J. Clin. Invest.* **1995**, *95*, 2409–2415.
54 Peraldi, P., Spiegelman, B. *Mol. Cell Biochem.* **1998**, *182*, 169–175.
55 Le Roith, D., Zick, Y. *Diabetes Care* **2001**, *24*, 588–597.
56 Hu, E., Liang, P., Spiegelman, B.M. *J. Biol. Chem.* **1996**, *271*, 10697–10703.
57 Arita, Y., Kihara, S., Ouchi, N. et al. *Biochem. Biophys. Res. Commun.* **1999**, *257*, 79–83.
58 Hotta, K., Funahashi, T., Arita, Y. et al. *Arterioscler. Thromb. Vasc. Biol.* **2000**, *20*, 1595–1599.
59 Kappes, A., Loffler, G. *Horm. Metab. Res.* **2000**, *32*, 548–554.
60 Fasshauer, M., Klein, J., Neumann, S., Eszlinger, M., Paschke, R. *Biochem. Biophys. Res. Commun.* **2002**, *290*, 1084–1089.
61 Berg, A.H., Combs, T. P., Du, X., Brownlee, M., Scherer, P.E. *Nature Med.* **2001**, *7*, 947–953.
62 Yamauchi, T., Kamon, J., Waki, H. et al. *Nature Med.* **2001**, *7*, 941–946.
63 Diez, J.J., Iglesias, P. *Eur. J. Endocrinol.* **2003**, *148*, 293–300.
64 Steppan, C.M., Bailey, S.T., Bhat, S. et al. *Nature* **2001**, *409*, 307–312.
65 Janke, J., Engeli, S., Gorzelniak, K., Luft, F.C., Sharma, A.M. *Obes. Res.* **2002**, *10*, 1–5.
66 Smith, U. *Obes. Res.* **2002**, *10*, 61–62.
67 Savage, D.B., Sewter, C. P., Klenk, E.S. et al. *Diabetes* **2001**, *50*, 2199–2202.
68 Nagaev, I., Smith, U. *Biochem. Biophys. Res. Commun.* **2001**, *285*, 561–564.
69 Taylor, S.I. *Cell* **1999**, *97*, 9–12.
70 Williams, K.V., Bertoldo, A., Kinahan, P., Cobelli, C., Kelley, D.E. *Diabetes* **2003**, *52*, 1619–1626.
71 Panten, U., Schwanstecher, M., Schwanstecher, C. *Exp. Clin. Endocrinol. Diabetes* **1996**, *104*, 1–9.

72 Campbell, L. K., White, J. R., Campbell, R. K. *Ann. Pharmacother.* **1996**, *30*, 1255–1262.

73 Scheen, A. J., Lefebvre, P. J. *Drugs* **1998**, *55*, 225–236.

74 Del Prato, S., Aragona, M., Coppelli, A. *Diabetes Nutr. Metab.* **2002**, *15*, 444–450; discussion 450–451.

75 Wagstaff, A. J., Goa, K. L. *Drugs* **2002**, *62*, 1805–1837.

76 Diamant, M., Heine, R. *Drugs* **2003**, *63*, 1373–1405.

77 Yang, W. S., Jeng, C. Y., Wu, T. J. et al. *Diabetes Care* **2002**, *25*, 376–380.

78 Burant, C. F., Sreenan, S., Hirano, K. et al. *J. Clin. Invest.* **1997**, *100*, 2900–2908.

79 Yonemitsu, S., Nishimura, H., Shintani, M. et al. *Diabetes* **2001**, *50*, 1093–1101.

80 Jones, N. P., Charbonnel, B., Lannqvist, F., Owen, S., Patwardhan, R. *Diabetologia* **1999**, *42*, 859.

81 Shimabukuro, M., Zhou, Y. T., Lee, Y., Unger, R. H. *J. Biol. Chem.* **1998**, *273*, 3547–3550.

82 Higa, M., Zhou, Y. T., Ravazzola, M. et al. *Proc. Natl Acad. Sci. USA* **1999**, *96*, 11513–11518.

83 Prigeon, R. L., Kahn, S. E., Porte, D., Jr. *J. Clin. Endocrinol. Metab.* **1998**, *83*, 819–823.

84 Andersen, H. U., Fey, S. J., Larsen, P. M. et al. *Electrophoresis* **1997**, *18*, 2091–2103.

85 John, N. E., Andersen, H. U., Fey, S. J. et al. *Diabetes* **2000**, *49*, 1819–1829.

86 Larsen, P. M., Fey, S. J., Larsen, M. R. et al. *Diabetes* **2001**, *50*, 1056–1063.

87 Christensen, U. B., Larsen, P. M., Fey, S. J. et al. *Autoimmunity* **2000**, *32*, 1–15.

88 Sparre, T., Christensen, U. B., Mose Larsen, P. et al. *Diabetologia* **2002**, *45*, 1550–1561.

89 Sanchez, J. C., Chiappe, D., Converset, V. et al. *Proteomics* **2001**, *1*, 136–163.

90 Hojlund, K., Wrzesinski, K., Larsen, P. M. et al. *J. Biol. Chem.* **2003**, *278*, 10436–10442.

91 Edvardsson, U., Brockenhuus Von Lowenhielm, H., Panfilov, O. et al. *Proteomics* **2003**, *3*, 468–478.

92 White, I. R., Man, W. J., Bryant, D. et al. *Proteomics* **2003**, *3*, 505–512.

93 Kratchmarova, I., Kalume, D. E., Blagoev, B. et al. *Mol. Cell Proteomics* **2002**, *1*, 213–222.

94 Nicolls, M. R., D'Antonio, J. M., Hutton, J. C. et al. *J. Proteome Res.* **2003**, *2*, 199–205.

95 Collins, H., Najafi, H., Buettger, C. et al. *J. Biol. Chem.* **1992**, *267*, 1357–1366.

96 Guest, P. C., Bailyes, E. M., Rutherford, N. G., Hutton, J. C. *Biochem. J.* **1991**, *274* (Pt 1), 73–78.

97 Schuit, F., Flamez, D., De Vos, A., Pipeleers, D. *Diabetes* **2002**, *51 Suppl 3*, S326–S332.

98 Sanchez, J. C., Converset, V., Nolan, A. et al. *Mol. Cell Proteomics* **2002**, *1*, 509–516.

99 Collier, G., Walder, K., De Silva, A. et al. *Ann. NY Acad. Sci.* **2002**, *967*, 403–413.

100 Lilla, V., Webb, G., Rickenbach, K. et al. *Endocrinology* **2003**, *144*, 1368–1379.

101 Nadler, S. T., Stoehr, J. P., Schueler, K. L. et al. *Proc. Natl Acad. Sci. USA* **2000**, *97*, 11371–11376.

102 Busch, A. K., Cordery, D., Denyer, G. S., Biden, T. J. *Diabetes* **2002**, *51*, 977–987.

103 Sanchez, J. C., Converset, V., Nolan, A., Schmid, G., Wang, S. et al. *Proteomics* **2003**, *3*, 1500–1520.

Part V
Infectious Diseases

Biomedical Application of Proteomics
Edited by J.-C. Sanchez, G. L. Corthals, D. F. Hochstrasser

ISBN: 3-527-30807-5

13
Proteome Approach to Infectious Diseases: Acute-phase Proteins and Antibody Profiles as Diagnostic Indicators in Human Plasma

Luca Bini, Sabrina Liberatori, and Vitaliano Pallini

13.1
Introduction

The pathology of an infectious disease reflects the outcome of the battle between the host genome and the pathogen genome [1]. Continuing the metaphor, one can also conclude that most weapons of this battle are proteins. In fact, the pathogenic strategies of bacteria and viruses are based on protein expression and protein/protein interactions, which underlie the entering into host's cells, the subversion of host's signaling system, the escape from host's defences, the finding of a niche and eventually multiplication and infection of further hosts. If pathogenic strategies of bacteria and viruses are extremely various [2], the host can rely upon two basic defensive processes: immunity (innate and acquired, cellular and humoral), and inflammation (a dynamic homeostatic process whose earliest reactions are named "acute phase") (for reviews see [3, 4]). Both acute-phase proteins and antibodies occur in the human plasma proteome and it is there that the hallmarks of victory and defeat are to be found.

Early after infection, inflammation signals determine impressive variations in the plasma composition, with a group of proteins (positive reactants) increasing, and a different group (negative reactants) decreasing in concentration. The acute-phase response (APR) is not highly specific, but it has long been known that different primary inducers may evoke responses different to the reactant pattern [3, 5, 6]. Then, circulating antibodies to molecular structures of the pathogen can be detected in the host plasma. Antibody variations are astounding; millions of different sequences and specificities are estimated to occur in the serum of a normal adult. Antibody responses are highly specific and are directed by the antigenic structures of the pathogen, but the genetic susceptibility of individual hosts may determine the different antibody repertoires to the same pathogens. These differences may give rise to antibodies involved in protection and in pathogenesis, both affecting the outcome of the disease [3].

At present, clinical monitoring of protein concentrations and of antibody titer and specificity are carried out by immunometry. Within the last 10 years proteome techniques have been applied to research on acute-phase proteins [7–11] and, far more extensively, on the human antibody responses to pathogen pro-

Biomedical Application of Proteomics
Edited by J.-C. Sanchez, G. L. Corthals, D. F. Hochstrasser

ISBN: 3-527-30807-5

teomes, as reported in a recent dedicated issue of the *Proteomics* journal (Vol. 1, Issue 4, 2001). The word "immunoproteome" has been coined to define the antigenic components of a proteome [12].

In this chapter we review our data, providing further evidence for the specificity of the APR in infection diseases. An electrophoretic map of human APR proteins is also presented, in which protein separation, identification, and quantitation have been carried out by two-dimensional electrophoresis (2-DE) followed by silver-staining and spot assignment by matching with reference maps [13, 14]. In addition, we highlight differences in antibody responses of two strains of mice (BALB/c and C3H) upon experimental infection by *Chlamydia trachomatis*, and compare mouse anti-*C. trachomatis* antibody profiles with those found in human patients with *Chlamydia*-induced chronic pelvic diseases. To achieve this, proteins of *C. trachomatis* have been fractionated by 2-DE followed by western blotting with serum from individual mice and patients in order to provide individual antibody profiles.

13.2 Electrophoretic Map of Acute-phase Response Proteins

The electrophoretic (2-DE) display of plasma proteins from a patient affected by *H. influenzae* type b infection is presented in Figure 13.1. By matching to the image of normal human plasma in Swiss 2-D PAGE (http://www.expasy.org/ch2d/) we could assign spots corresponding to well-known positive reactants (ceruloplasmin, complement factor B, α-1B glycoprotein, α_1-antichymotrypsin, α_1-antitrypsin, haptoglobin α and β, orosomucoid 1, complement C4) and negative reactants (albumin, α_2-HS-glycoprotein, transthyretin, serum retinol-binding protein). Spots of serum amyloid A protein (SAA) and C-reactive protein (CRP) were identified in our acute-phase serum map by immunoblotting with specific antibodies [8]. They are notably absent from the Swiss 2-D PAGE reference gel, as it represents a healthy plasma electropherogram.

To ascertain the behavior of reactants during the disease, quantitative variations of proteins were monitored as ratios between volumes of spots in samples taken during acute phase and after recovery. Correction in variability of silver staining, relative volumes to the sum of the volume of all spots in each gel were calculated by the Melanie software (http://www.genebio.com). The overall experimental variability in the densitometric quantitation was assessed by measuring ratios between relative volume of spots in 12 gels, each loaded with 0.75 µL of serum. Coefficients of variation (standard deviations divided by the mean $\times 100$) for proteins relevant in this study were between ca. 10% (for haptoglobin, α_1-antitrypsin and transthyretin) and ca. 20% (for serum retinol-binding protein). The average coefficient of variation for all relevant proteins (Table 13.1) was about 14%. Experiments with increasing loads of the same serum (0.075–4.5 µL) indicated that variations in the volume of the serum albumin spot cannot be determined and that variations in α_1-antitrypsin and haptoglobin spots tend to be underestimated presum-

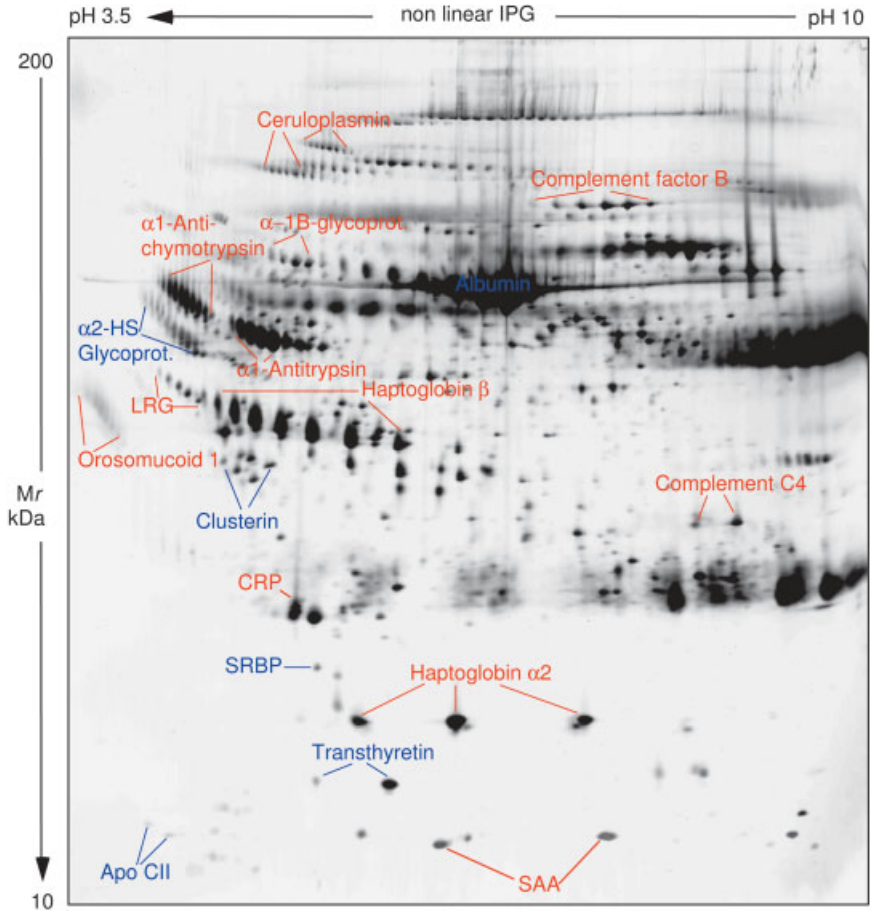


Figure 13.1 Silver-stained 2-D electropherogram of acute-phase human serum from a patient infected by *Haemophilus influenzae* type b. Colours indicate negative reactants (blue) and positive reactants (red), respectively.

ably because of stain saturation. The streak related to CRP could not be usefully quantitated by densitometry.

Following the above described protocol we have undertaken several follow-up studies on APR in infectious diseases. Plasma samples from patients on the onset of the disease (acute phase) were compared to samples taken after recovery or during the evolution of the disease. Such comparisons were performed within samples from the same patient in order to rule out genetic differences. As a final validation, data obtained by silver-stained gel densitometry compared well with data obtained by nephelometry [15]. Under these conditions we could observe that spots of known reactants behaved as expected. Most other spots exhibited statistically inconsistent quantitative variations, being close to the average coefficient of variation. More interesting was the

Table 13.1 APR protein variations in 2-D electropherograms of patient sera infected with *Haemophilus influenzae* type b.

	Protein name	***SwissProt AC***	***A/R*** [a)]
Positive acute-phase reactants	Ceruloplasmin	P00450	1.5
	Complement factor B	P00751	1.3
	α-1B-glycoprotein	P04217	4. 2
	α-1-antichymotrypsin	P01011	4.0
	α-1-antitrypsin	P01009	2.0
	Leucine-rich α_2-glycoprotein (LRG)	P02750	3.2
	Haptoglobin β- and α-chains	P00737 P00738	2.2
	Orosomucoid 1	P02763	4.1
	Complement C4	P01028	1.3
	C-reactive protein (CRP)	P02741	+
	Serum amyloid A protein (SAA)	P02735	+
Negative acute-phase reactants	Albumin	P02768	nq
	α_2-HS-glycoprotein	P02765	0.6
	Clusterin	P10909	–
	Serum retinol-binding protein (SRBP)	P02753	0.2
	Transthyretin	P02766	0.4
	Apolipoprotein C-II (Apo CII)	**P02655**	0.3

a) Average relative volume of spots in gels from acute-phase (A) and recovery (R) sera.
+ spots present only in the acute-phase serum.
– spots not always detectable in the acute-phase serum.
nq = not quantifiable because of silver-stain saturation.

finding that these proteome procedures, used to screen high numbers of proteins, allowed the identification of new APRs, such as leucine-rich α_2-glycoprotein (positive) and apolipoprotein C-II and clusterin (negative) (Table 13.1 and Fig. 13.1).

13.3 Clinical Monitoring of APR Proteins by 2-DE

APR proteins have long been reputed to be unspecific markers of inflammation. Measurements of some of these proteins (most frequently CRP) were carried out to evaluate the severity of the inflammation and the effects of the therapy, which is more accurate than the erythrocyte sedimentation rate test. The resolution power of 2-DE allows numerous reactants simultaneously to be monitored instead of one or just a few. Using 2-DE we have been able to distinguish bacterial from viral infections [8]. Our conclusion was based on the higher incidence of SAA which was present in all 18 cases of bacterial diseases, whereas it was only pre-

sent in 6 out of 16 cases for viral diseases. Moreover the concentration of other reactants (α_1-antitrypsin, haptoglobin, leucine-rich α_2-glycoprotein, transthyretin and serum retinol-binding protein) varied simultaneously during the course of bacterial diseases, but variations occured in a staggered fashion in cases of viral diseases for these same proteins. These data provide evidence for the specificity of APR, which can be exploited for diagnostic purposes. More recently our results have been confirmed in immunometric procedures. The distinction between bacterial and viral infections was immunometrically confirmed for different APR, such as CRP, SAA, and procalcitonin [16] and by CRP and a newly discovered marker, 2′-5′ oligoadenylate synthetase [17]. It is interesting that procalcitonin and 2′-5′ oligoadenylate synthetase have been recently recognized to be APR proteins.

Further research from our laboratory has demonstrated that SAA and haptoglobin α- and β-chains are markers for sepsis in perinatal plasma, at developmental stages in which most other APR protein variations are not consistently related with the onset and duration of the disease. The under-sialylation of haptoglobin β-chain, and the different phenotypes of α-chains exclude the maternal origin of haptoglobin in perinatal plasma [9].

Five different spots of SAA have been characterized by tandem mass spectrometry [10, 18]. Among these, two derive from the gene 1α and are the most intense and amenable to densitometry. The primary gene product is known to undergo a post-translational extracellular cleavage of the N-terminal arginine [19]. As a result, two spots appear in the 2-D electropherogram, one at pI 6.15 (primary product) and the other at pI 5.69 (des-Arg product). In Figure 13.2 it is shown that the abundance ratio between SAA1α and des-Arg SAA1α increase over time. One can imagine that the time of inversion of the ratio may mark the beginning of the decline of pathogenic processes. Here, 2-DE separation would be the technique of

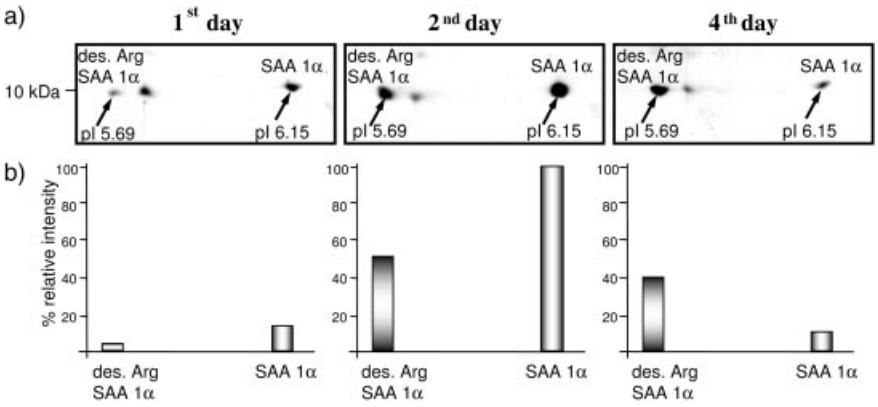


Figure 13.2 Time-course abundance variations of SAA1α isoforms (see text): silver-stained spot in the relevant window (a) and quantification of SAA1α spots (b). In the ordinate the spot intensity is expressed in arbitrary units normalized to the highest value.

choice to carry out this investigation because antibodies specific for one of the two SAA isoforms are not available.

13.4 *Chlamydia trachomatis* Immunoproteome

After infection, the presence in the host organism of non-self pathogen structures elicit a polyclonal antibody response. One can characterize such responses by defining the specificity of the antibodies generated in the response. Only a subset of the pathogen proteome elicits antibody responses which are responsible for the 'immunoproteome'; i.e. they give rise to antibody responses and bind to these antibodies. With respect to the infected host, a genetic susceptibility may occur, which could favor or inhibit the expression of particular antibodies. In short, immunoproteome component proteins are produced by the pathogen as potential antigens and later are selected by the host to raise antibodies against. The sequencing of pathogen genomes has made possible the characterization of their corresponding proteomes, and in turn, this knowledge opens new ways leading to the discovery of specific antigenic proteins. Two-dimensional electrophoresis and mass spectrometry are currently the tools of choice in this endeavor.

The genus *Chlamydia* is composed of a group of obligate intracellular gram-negative bacteria, and two main species, *C. trachomatis* and *C. pneumoniae*, are associated with human pathologies. *Chlamydia trachomatis* causes prevalent infections of the mucosal tissue of the eyes and the urogenital tract, and is a major cause of sexually transmitted disease worldwide. Infections are insidious and, though often asymptomatic, can have serious consequences, particularly for women. Left untreated, genital chlamydial infections are chronic, and repeated infections are common. There is a close reciprocal interaction between antibody and antigen. If one wants to monitor an antibody, a purified antigen is needed, and vice versa. Ideally if we want to dissect an antibody response we need to have all proteins expressed by the infecting pathogen, the whole proteome. The genome of *C. trachomatis* contains 1042 kb, corresponding to 897 predicted proteins (http://www.ebi.ac.uk/cgi-bin/genomes.cgi?genomes=Bacteria), a number amenable to be separated and characterized by proteomic technologies.

The first report on this type of application was by our group where we described the proteome of *C. trachomatis* L2 [20]. In this work we used 2-DE and immunoblotting to dissect (i.e. to define the specificity) antibody responses of infected patients. In the following paragraphs we present and discuss data on the specificity of human antibody response to *C. trachomatis* [21] and we also report studies on the genetic susceptibility in the response of inbred mouse to the same pathogen, to evidence the role of individual antibody species as to the outcome of the disease.

Technically, 2-D unstained electrophoretic patterns of *C. trachomatis*, prepared by standard procedures [13, 22] are transferred (western blot) to replicas on nitrocellulose and probed with immune plasma. Spots stained by labelled antibodies

are indicative of the specificity of the antibody response of an individual patient. Immunoblotting patterns can be interpreted in order to find potential vaccines, but also to acquire information on the outcome of the disease. In the latter case, immune sera must be related with an individual patient. A typical experiment is performed as follows: 12 2-D gels of total protein extracts from elementary bodies (EBs, the chlamydial extracellular infectious form) of *C. trachomatis* serotype L2 were run in parallel, under identical conditions. After the run, two gels were stained with silver nitrate and the other 10 were electrotransferred onto nitrocellulose to be used for subsequent immunodetection with patient sera. Briefly, electroblotting of gels was performed according to Towbin et al. [23] and further processed according to standard procedures, modified as described by Bini et al. [24]. Immunoreactive spots were detected by overnight incubation at room temperature with patient sera (1500- to 5000-fold dilution), followed by incubation with rabbit anti-human IgGs conjugated with peroxidase and detected with chemiluminescence reaction. The chemiluminescence technology is about 100-fold more sensitive than colorimetric detection and allows high dilution of the patient sera, preventing the risk of unspecific reactions.

In this study 17 sera were obtained from four cases of lower genital tract infection, and 13 cases of pelvic inflammatory disease (PID), including two cases of secondary sterility. All sera were positive for a standard microimmunofluorescence test (MIF) with purified *C. trachomatis* L2 EBs. An additional group of 10 seronegative control sera from healthy blood donors was tested by immunoblotting in the same way, and using the same dilutions as for the patient sera, in order to exclude the occurrence of non-specific reactions.

This approach can hide some problems such as cross-reaction with antibodies elicited from other bacteria, especially with highly conserved proteins. To overcome this difficulty some authors have analyzed the immune response in mouse models infected by human pathogens [25, 26], demonstrating an high agreement between the immune response of infected mice and of human patients. Animal models may have an important role in understanding the immunopathology of human chlamydial diseases, providing a means to study the immune response to defined chlamydial components. Some studies demonstrate the possibility to infect genitally different strains of inbred mice with human pathogenic strains of *C. trachomatis*. These infected mice can be used to define protective immune mechanisms leading to different disease susceptibilities as the different strains of mice show different severity of the disease. For instance, C3H mice infected with *C. trachomatis* serovar F by intrauterine injection develop chronic inflammation and severe diseases, while BALB/c mice infected with the same human strain of *C. trachomatis* present only an acute inflammation [27]. The study of immune responses to chlamydial infection in mice has the advantages of using inbred mice to obtain pathogen-free animals and to follow the immune response before and during the outcome of the disease.

We reported an experiment in which five C3H ($H\text{-}2^0$) and five BALB/c ($H\text{-}2^k$) mice were immunized three times, approximately one week apart, with *C. trachomatis* serovar F, via the intraperitoneal route. Five serum samples were collected

to follow the growth of the immune response in the two different strains of mice. The first sample was collected before the start of the infection (0 days) and represented the preimmune serum; the other four were collected at 15, 23, 37 and 53 days after the first immunization. EBs of *C. trachomatis* serovar F were separated by 2-DE and subjected to immunoblotting with mice sera as described above. All the animals were pathogen-free females.

The above described approach, in which 2-DE is combined with immunoblotting with patient sera, is now widely used to identify immunogenic proteins in different pathogens. For example, some studies have been performed to investigate immunorelevant *Borrelia garinii* antigens in patients affected by Lyme disease [28], to identify possible vaccine candidates in infections of *Staphylococcus aureus*, using pooled sera from different patients [29] or for mapping immunoreactive antigens in *Francisella tularensis* [30]. (See Chapter 16 for reference maps and comparative analysis of *F. tularensis*.) A number of studies have been performed to investigate the antigenicity of *Helicobacter pylori*, some using 2-DE and immunoblotting with a pooled sera from patients [31, 32], others using sera from individual patients to evaluate the frequency of the antigens during different gastroduodenal pathologies [33, 34], or analyzing only the proteins present on the cell surface [35]. Other humoral immune responses to parasitic organisms that cause infections in humans have been evaluated with this proteomic approach, such as the serological response to *Toxoplasma gondii* infections [36] and definition of immunogenic proteins in fungal diseases caused by *Candida albicans* [37, 38].

13.5
Human Humoral Immune Response to *Chlamydia trachomatis* Infections

Infectious, extracellular elementary bodies are endowed with adhesion molecules, which permit the invasion of host cells and are reported to elicit host immune response. With the methodological approach reported above, 55 EB's antigens have been found and all the immunoreactive spots have been reported in the silver-stained 2-D gel image of *C. trachomatis* serovar L2 (Fig. 13.3). It is important to note that patient blots showed individually different patterns comprising a number of antigenic spots which varied from 2 to 28. In Figure 13.3 we have represented with different colors the frequencies of antigens found in the 17 patients. The high-frequency antigens (17/17–11/17) are shown in red; of these it is interesting to observe that the cysteine-rich outer membrane protein 2 (OMP2) is the only one recognized by all the patient sera used. The antigens shown in blue had a medium frequency ranging from 4 to 10 out of 17 patients and those in green are all the patient-specific antigens that are probably due to a different immune response of each individual (frequency 3/17 to 1/17). This latter group of antigens represents about 51% of all the immunreactive proteins, indicating that the individual component of immune response plays a key role in the battle of the host against the infective agent and could be correlated with the outcome of the disease.

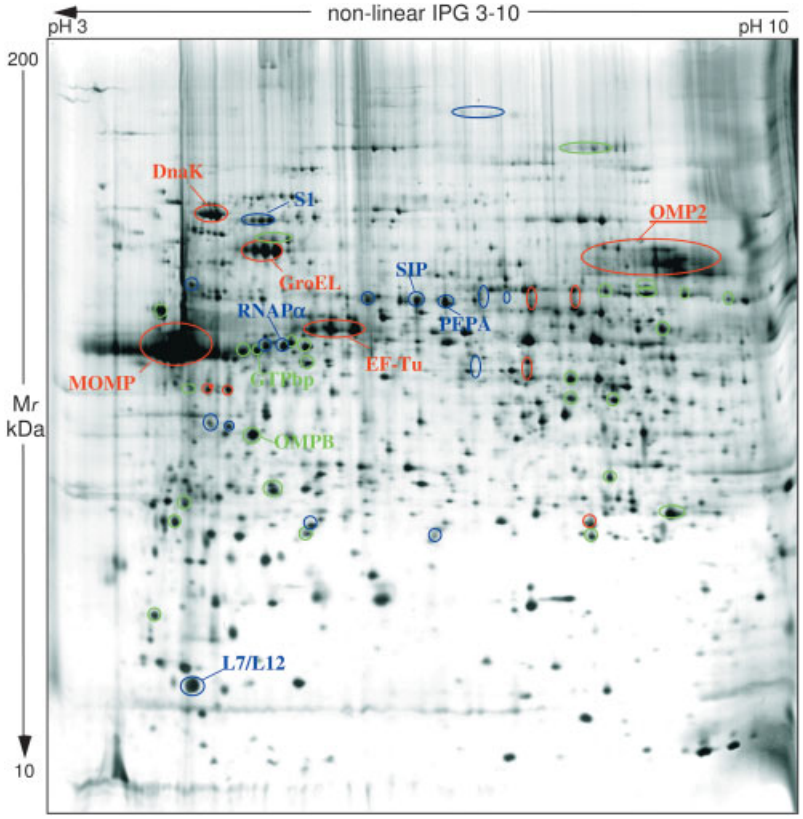


Figure 13.3 Silver-stained gel of *Chlamydia trachomatis* serovar L2 showing all the 55 antigens immunostained with patient sera. Spots circled in red represent high frequency antigens (17/17–11/17); spots circled in blue represent medium frequency antigens (10/17–4/17); spots circled in green represent patient-specific antigens (3/17–1/17). Identified proteins are marked by abbreviations: MOMP, major outer membrane protein; OMP2, outer membrane protein 2; OMPB, outer membrane protein B; GroEL, GroEL-like protein; DnaK, DnaK-like protein; EF-Tu, elongation factor Tu; S1, ribosomal proteins S1; L7/L12, ribosomal protein L7/L12; RNAPα, RNA polymerase α-subunit; GTPbp, GTP binding protein; PEPA, leucine peptidase; SIP, putative stress-induced protease of the HtrA family. Note that OMP2 has been underlined to point out its presence in all the 17 sera analyzed.

Some of these antigens have been identified by computer matching with a previously published *C. trachomatis* L2 reference map [20], available through the Siena 2-D PAGE database (http://www.bio-mol.unisi.it/2d/2d.html) and correspond to some of the antigens already described in the literature, like MOMP, OMP2, GroEL-like protein, and DnaK-like protein. Some others, not yet described as antigens, have been identified in the study, and among them we can mention the outer membrane protein B (OMPB), and seven conserved bacterial proteins such as elongation factor Tu (EF-Tu), ribosomal proteins S1 and L7/L12, RNA polymerase

α-subunit (RNAPα), GTP-binding protein (GTPbp), leucine peptidase (PEPA), and a putative stress-induced protease of the HtrA family (SIP).

Immune reactions are known to play an important role in disease evolution. Antibodies against the antigens involved in adhesion may neutralize the parasite, preventing cellular infection, but it is also reported that host immune response to chlamydial infection is a "double-edged sword", which can elicit both protective and pathological immune responses; for instance the major outer membrane protein (MOMP) is considered a protective antigen whereas heat shock protein 60 (GroEL-like) might contribute to pathogenesis [39]. *Chlamydia trachomatis*, an intracellular parasite, may overcome the host humoral immune response, and often causes asymptomatic infections. Some patients who cannot efficiently resolve the infection, or become sensitized to chlamydial antigens following repeated, often asymptomatic, infections, elicit pathogenic immune responses which lead to tissue scarring and serious sequelae, such as blindness and sterility [40].

It is interesting to note that our proteomic approach is needed to assess the real antigens recognized by the host humoral immune system. The antigenic frequency in different patients may also be able to be related to the outcome of the disease and to be used to define the role of single antigens as prognostic and/or diagnostic markers. In the end, the role, the frequency and the type of immunoresponse elicited (protective or pathogenic), may be of crucial importance in the choice of candidates to be used in vaccine planning. One must also consider that the antigens thus identified may include true, chlamydia-specific "primary" immunogens, but also proteins from different bacterial species which cross-react with antibodies elicited during different infections. A particular case could be represented by the presence of antibodies to *Chlamydia pneumoniae*, which were not assessed in our sera panel, but for which a high prevalence has (often) been reported in healthy populations. In fact, healthy controls frequently harbor chlamydial infection without clinical symptoms, rendering discrimination difficult among antibodies induced by chlamydial infections and by cross-reacting antigens. Moreover, this bacteria causes chronic infections in humans and it is impossible to obtain the control serum from the same patient in the early stage of infection before any immune response. These problems might be solved using animal models infected with human pathogenic strains of *C. trachomatis*.

13.6 Genetic Susceptibility to *Chlamydia trachomatis* Determines the Outcome of the Disease: Data from a Mouse Model

Immunoblotting obtained with sera of BALB/c and C3H mice collected after immunizations with *C. trachomatis* are displayed in Figure 13.4. Blots with pre-immune sera were blank (data not shown); on the other hand, blots with immune sera collected at 15, 23, 37, and 53 days post-infection exhibited reproducible patterns of immunoreactive spots. It has also become evident that in the two groups of mice, which have developed different diseases, humoral immune response was

different. In BALB/c mice, which had a transient salpingitis without infertility, the immune response started very fast, especially against the MOMP. On the other hand in C3H mice, which become chronically ill and infertile, the response was quite slow, with, at the end, a different pattern of immunostained antigens, such as OMP2, GroEL-like protein and EF-Tu.

These results suggest that a different immune response can be correlated with different outcome of the disease. In particular we have focused attention on the time course of two of the well-studied antigens, such as MOMP and OMP2. As reported in Figure 13.5 a the histogram associated with the immunoreactivity time course of MOMP in the two different strains of mice showed different immune responses. Indeed in BALB/c the immunostaining of MOMP revealed a rapid increase in expression reaching a high intensity, whereas in C3H the increase was slower and reached an intensity of immunostaining lower than in BALB/c. Instead, OMP2 was immunostained only in the C3H mice as shown in the histogram of Figure 13.5 b. From these last results it is clear that different antibodies may be associated with protective or pathological immune responses. For example, a very fast and high response to MOMP is associated with healing, as BALB/c mice do not develop chronic salpingitis. On the other hand, antibodies against OMP2, GroEL-like protein, and EF-Tu play an important role in directing the disease in a chronic state. This observation has been already reported for GroEL-like protein which was associated with chlamydial disease pathogenesis [41, 42]. In the two strains of mice, genetically different as to susceptibility to the disease [43], we also found different antibody responses to pathogen proteins.

Animal models allow comparison of experimental data with data on human immune responses due to natural infections. Actually, in our study, the antibody immune responses in mice are very similar to those observed in patients. As reported in Figure 13.6, C3H mice, which become chronically sick, show a similar immune response to that in humans with chronic salpingitis. Both human and C3H sera immunoreacted with OMP2, GroEL-like protein, and EF-Tu. On the other hand, BALB/c mice, which develop a mild form of disease, raised a different immune response, comprising antibodies specific for different chlamydial proteins.

In conclusion the host immune response to infectious agents may be of primary importance. The description of the immunoproteome, most frequently exploited in vaccine discovery, also correlates the host's antibody responses, either protective or pathogenic, with the outcome of the disease. This kind of research is aimed at the development of new diagnostic/prognostic markers.

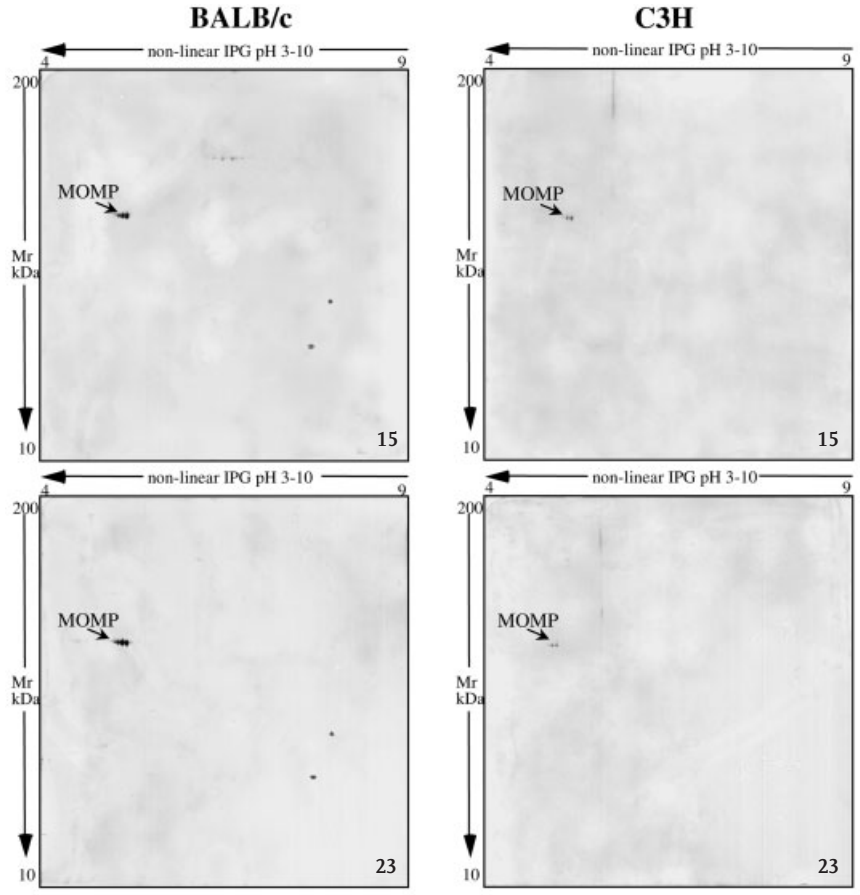


Figure 13.4 Immunoblots with mouse sera collected at 15, 23, 37 and 53 days post-immunization in the two different strains of mice analyzed (BALB/c and C3H).

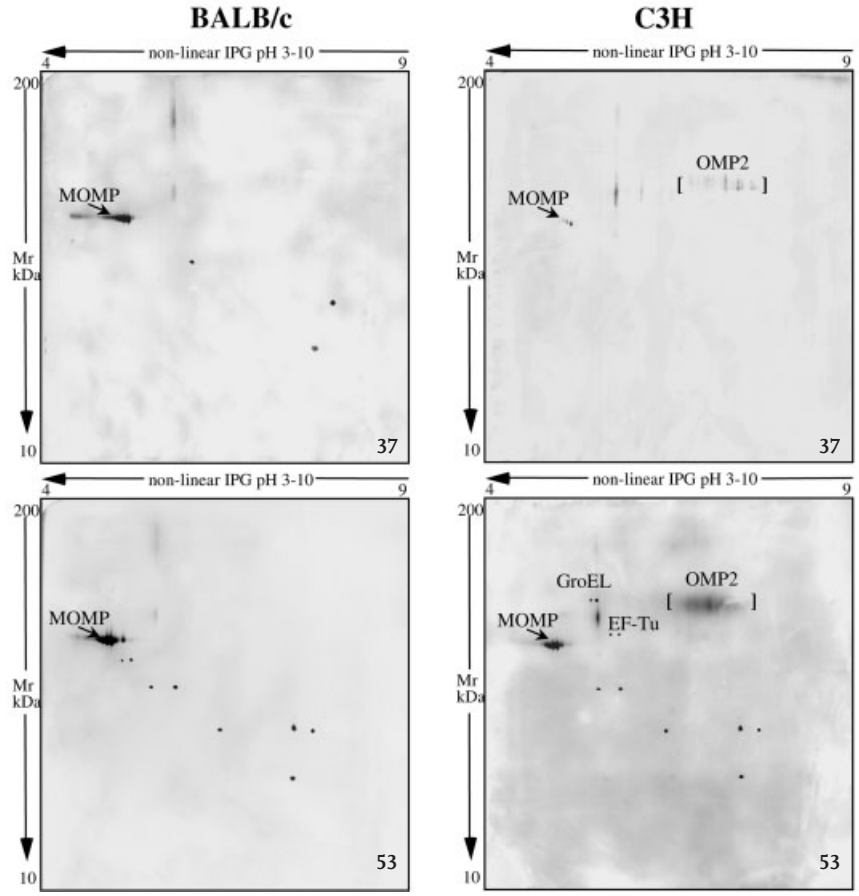


Figure 13.4 (continued)

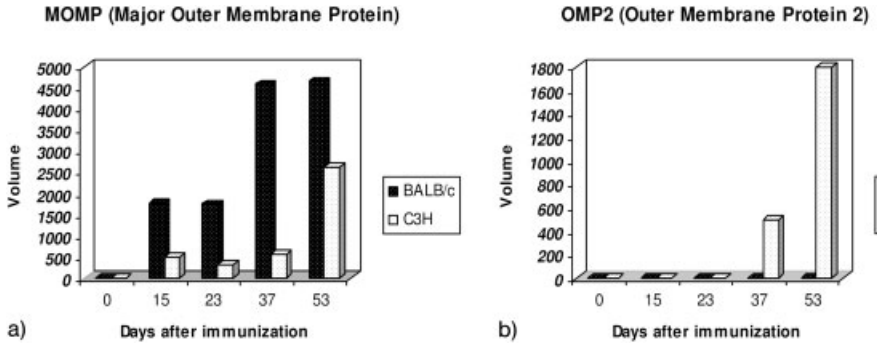


Figure 13.5 Time-course variations of MOMP (a) and OMP2 (b) immunoreactivity in sera of BALB/c and C3H mice. Note: BALB/c mice developed a very fast and high response against MOMP. On the other hand, OMP2 was recognized only by C3H mice.

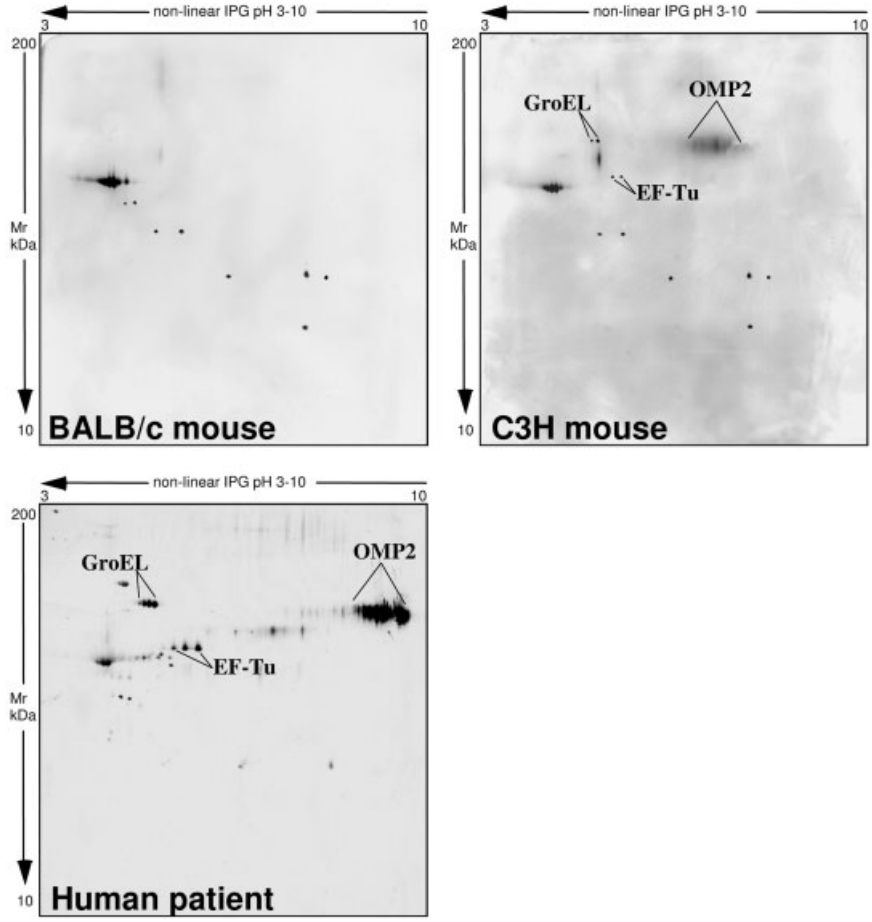


Figure 13.6 Comparison of immunoblots obtained from 53 days post-infection of BALB/c and C3H mice and a typical human patient with a chronic salpingitis. C3H mice and humans develop a similar immune response and disease.

13.7 Conclusions

We have presented data on acute phase and on humoral antibody responses because they are the main defenses of the host against infections and they can be assessed in plasma, the acute phase starting first followed by the antibody response. The interest in these markers has recently increased with the simultaneous monitoring of APR protein concentrations and the titer of antibodies against infective agents in studies on the relationship between infectious and coronary diseases [44, 45].

Proteome technologies can play an important role in the discovery of protein markers for the diagnosis and prognosis. Using 2-DE, silver-staining and image analysis we could establish that:

- bacterial and viral infections can be distinguished by monitoring a group of APR markers (SAA, haptoglobin, α_1-antitrypsin, leucine-rich α_2-glycoprotein, transthyretin and serum retinol-binding protein);
- haptoglobin can be used as a sepsis marker in perinatal ages, when several other plasma proteins do not show consistent variations in relation with the disease evolution;
- post-translationally modified SAA1α isoforms can be used in the prognostic follow-up of the disease.

Using the same tools combined with western blotting, we established that:

- antibody responses of all patients with chronic diseases induced by *C. trachomatis* stain characteristic immunoreactive profiles of *C. trachomatis* proteins, different from those produced by healthy controls;
- such profiles are very similar to those stained by sera from C3H mice experimentally infected with the same pathogen. Interestingly, immunoreactive profiles from BALB/c mice, which develop a mild disease, are noticeably different. Therefore, the resulting profiles can be correlated with the severity of the disease.

In order to reach these conclusions, we had to consider the behavior of a large group of APR proteins (30 positive/negative reactants) and to select out subsets forming significant profiles for diagnosis. In a parallel fashion we noted that subsets of the global *C. trachomatis* immunoproteome (a total of 55 antigenic spots) formed profiles that could be correlated to the different responses.

Modern diagnosis tends to be based on the monitoring of many proteins (multiparameter indicators) instead of one or a few [46]. Net-fishing is better than line-fishing [47]. Due to its large-scale character, proteomic research can monitor large numbers of proteins simultaneously and select out panels of indicators. We envisage that proteomic results will be increasingly used in routine analyses, especially when integrated with new fast screening technologies. Among them peptide and protein microarrays [48] provide very rapid quantitative assessment of large sets of proteins.

In our work protein separation, quantitation, and antigen/antibody interaction by western blotting have been performed using 2-DE. Important improvements can be introduced in this area of medical research. As to quantitation, innovative procedures have been introduced, including the mass spectrometry-based ICAT method [49] and the 2-D gel-based fluorescence difference gel electrophoresis (DIGE) [50]. Separation can be improved by mass spectrometry in several versions, such as MALDI-TOF mass spectrometry of bacteria [51] and SELDI-TOF mass spectrometry [52].

With these tools, we expect to improve weapons against the challenge of emerging infectious diseases due to global transmission spread by person to person, microbial adaptation, and, among others, bioterrorism.

13.8 References

1 Adams, L.G. *Vet. Microbiol.* **2002**, *90*, 553–561.
2 Bloom, B.R. *Nature* **2000**, *406*, 760–761.
3 Gallagher, R.B., Miller, L.J. *Science* **1996**, *272*, 13.
4 Ceciliani, F., Giordano, A., Spagnolo, V. *Protein Pept. Lett.* **2002**, *9*, 211–223.
5 Baumann, H., Gauldie, J. *Immunol. Today* **1994**, *15*, 74–80.
6 Kushner, I. *Methods Enzymol.* **1988**, *163*, 373–383.
7 Bini, L., Magi, B., Cellesi, C., Rossolini, A., Pallini, V. *Electrophoresis* **1992**, *13*, 743–746.
8 Bini, L., Magi, B., Marzocchi, B. et al. *Electrophoresis* **1996**, *17*, 612–616.
9 Liberatori, S., Bini, L., De Felice, C. et al. *Electrophoresis* **1997**, *18*, 520–526.
10 Bruun, C.F., Sanchez, J.C., Hochstrasser, D.F., Marhaug, G., Husby, G. *Electrophoresis* **1998**, *19*, 776–781.
11 Wait, R., Miller, I., Eberini, I. et al. *Electrophoresis* **2002**, *23*, 3418–3427.
12 Jungblut, P.R., Bumann, D. *Methods Enzymol.* **2002**, *358*, 307–316.
13 Bjellqvist, B., Pasquali, C., Ravier, F., Sanchez, J.C., Hochstrasser, D. *Electrophoresis* **1993**, *14*, 1357–1365.
14 Hughes, G.J., Frutiger, S., Paquet, N. et al. *Electrophoresis* **1992**, *13*, 707–714.
15 Meucci, D., Messina, M., Garzi, A. et al. *Pediatr. Med. Chir.* **1996**, *18*, 601–606.
16 Shimetani, N., Shimetani, K., Mori, M. *Scand. J. Clin. Lab. Invest.* **2001**, *61*, 567–574.
17 Sasaki, K., Fujita, I., Hamasaki, Y., Miyazaki, S.J. *Infect. Chemother.* **2002**, *8*, 76–80.
18 Ducret, A., Bruun, C.F., Bures, E.J., Marhaug, G., Husby, G., Aebersold, R. *Electrophoresis* **1996**, *17*, 866–876.
19 Malle, E., Steinmetz, A., Raynes, J.G. *Atherosclerosis* **1993**, *102*, 131–146.
20 Bini, L., Sanchez-Campillo, M., Santucci, A. et al. *Electrophoresis* **1996**, *17*, 185–190.
21 Sanchez-Campillo, M., Bini, L., Comanducci, M. et al. *Electrophoresis* **1999**, *20*, 2269–2279.
22 Görg, A., Postel, W., Gunther, S. *Electrophoresis* **1988**, *9*, 531–546.
23 Towbin, H., Staehelin, T., Gordon, J. *Proc. Natl Acad. Sci. USA* **1979**, *76*, 4350–4354.
24 Bini, L., Liberatori, S., Magi, B., Marzocchi, B., Raggiaschi, R., Pallini, V. In Proteome Research: Two-dimensional Gel Electrophoresis and Identification Methods (Rabilloud, T., Ed.). Springer, Berlin Heidelberg New York, 2000, pp 127–141.
25 Pitarch, A., Diez-Orejas, R., Molero, G. et al. *Proteomics* **2001**, *1*, 550–559.
26 Bumann, D., Holland, P., Siejak, F. et al. *Infect. Immun.* **2002**, *70*, 6494–6498.
27 Tuffrey, M., Alexander, F., Woods, C., Taylor-Robinson, D. *J. Reprod. Fertil.* **1992**, *95*, 31–38.
28 Jungblut, P.R., Grabher, G., Stoffler, G. *Electrophoresis* **1999**, *20*, 3611–3622.
29 Vytvytska, O., Nagy, E., Bluggel, M. et al. *Proteomics* **2002**, *2*, 580–590.
30 Havlasova, J., Hernychova, L., Halada, P. et al. *Proteomics* **2002**, *2*, 857–867.
31 McAtee, C.P., Fry, K.E., Berg, D.E. *Helicobacter* **1998**, *3*, 163–169.
32 Lock, R.A., Coombs, G.W., Mc Williams, T.M. et al. *Helicobacter* **2002**, *7*, 175–182.
33 Kimmel, B., Bosserhoff, A., Frank, R., Gross, R., Goebel, W., Beier, D. *Infect. Immun.* **2000**, *68*, 915–920.
34 Haas, G., Karaali, G., Ebermayer, K. et al. *Proteomics* **2002**, *2*, 313–324.
35 Utt, M., Nilsson, I., Ljungh, A., Wadstrom, T. *J. Immunol. Methods* **2002**, *259*, 1–10.
36 Geissler, S., Sokolowska-Kohler, W., Bollmann, R., Jungblut, P.R., Presber, W. *FEMS Immunol. Med. Microbiol.* **1999**, *25*, 299–311.
37 Pitarch, A., Pardo, M., Jimenez, A. et al. *Electrophoresis* **1999**, *20*, 1001–1010.
38 Pardo, M., Ward, M., Pitarch, A. et al. *Electrophoresis* **2000**, *21*, 2651–2659.
39 Morrison, R.P., Manning, D.S., Caldwell, H.D. In *Sexually Transmitted Diseases* (Quinn, T., Ed.), Raven Press, New York, **1992**, pp 57–84.

40 Ward, M.E. *APMIS* **1995**, *103*, 769–796.

41 Morrison, R.P., Belland, R.J., Lyng, K., Caldwell, H.D. *J. Exp. Med.* **1989**, *170*, 1271–1283.

42 Peeling, R.W., Kimani, J., Plummer, F. et al. *J. Infect. Dis.* **1997**, *175*, 1153–1158.

43 Stagg, A.J., Tuffrey, M., Woods, C., Wunderink, E., Knight, S.C. *Infect. Immun.* **1998**, *66*, 3535–3544.

44 Biasucci, L.M., Liuzzo, G., Grillo, R.L. et al. *Circulation* **1999**, *99*, 855–860.

45 Rupprecht, H.J., Blankenberg, S., Bickel, C. et al. *Circulation* **2001**, *104*, 25–31.

46 Anderson, N.L., Anderson, N.G. *Mol. Cell. Proteomics* **2002**, *1*, 845–867.

47 Fields, S. *Science* **2001**, *291*, 1221–1224.

48 Bacarese-Hamilton, T., Bistoni, F., Crisanti, A. *Biotechniques* **2002**, Dec Suppl, 24–29.

49 Zhou, H., Ranish, J.A., Watts, J.D., Aebersold, R. *Nature Biotechnol.* **2002**, *20*, 512–515.

50 Yan, J.X., Devenish, A.T., Wait, R., Stone, T., Lewis, S., Fowler, S. *Proteomics* **2002**, *2*, 1682–1698.

51 Lay, J.O. Jr. *Mass Spectrom. Rev.* **2001**, *20*, 172–194.

52 Issaq, H.J., Veenstra, T.D., Conrads, T.P., Felschow, D. *Biochem. Biophys. Res. Commun.* **2002**, *292*, 587–592.

14
Proteomic Studies of Human Lymphocytes: New Insights into HIV Lymphocyte Infection?

Françoise Vuadens, David Crettaz, Amalio Telenti, Manfredo Quadroni, Michel A. Duchosal, Philippe Schneider, and Jean-Daniel Tissot

14.1
Introduction

The achievements made over the years in proteomic sciences have given us a unique opportunity to study various human disease processes [1] and the complex relations between pathogens and the cells or tissues that participate in the pathogenesis of diseases or that are specifically targeted by infectious agents. Impressive developments have been achieved in the past few years in proteomics, allowing the study of proteins in a given cell, the protein isoforms and modifications as well as protein/protein interactions [2]. The ability of mass spectrometry (MS), in association with bioinformatics, to identify small amounts of proteins from complex mixtures is the cornerstone of proteomics [3, 4]. One of our research project is to study a model of HIV infection of CD4 T lymphocytes that shows significant components of the *in vivo* process leading to immunosuppression and development of acquired immunodeficiency syndrome (AIDS). However, before this task can be carried out, it is necessary to gain insight into the proteome of normal lymphocyte populations. Then it will be possible to dissect the mechanisms involved in HIV infection of T lymphocytes.

14.1.1
The Lymphocytes

Peripheral blood cells derive from a common hematopoietic stem cell through a tightly regulated process comprising several differentiation steps, each controlled by intrinsic and extrinsic factors. The latter factors are both produced by maturing hematopoietic cells and by the non-hematopoietic environment such as stromal cells and are notably trapped by the extracellular matrix [5]. There are evidences that the lymphoid lineages, consisting of B, T and natural killer cells, are generated from a common lymphoid progenitor [6]. Subsequent B cell commitment appears to be dependent on Pax5 transcription factor, whereas notch signaling is critical for T cell fate (reviewed in [7]). Human T and B cell precursors continuously mature in the thymus, and in the bone marrow, respectively. B and T cells are key components of the adaptative immune responses, where they play roles in

Biomedical Application of Proteomics
Edited by J.-C. Sanchez, G. L. Corthals, D. F. Hochstrasser

ISBN: 3-527-30807-5

both its initiation and regulation as well as its effects [8, 9]. Mature, antigen-naive lymphocytes circulate continually from the blood to the lymph through transendothelial migration initiated by specialized interactions occurring at the level of post-capillary veinules within secondary lymphoid organs, and return to the blood via the lymphatic vessels.

B cell differentiation is a highly regulated process with pathways and steps that have been well delineated [10]. B cells express uniquely CD19, a 95-kDa protein present throughout the B cell lineage until plasma cell differentiation. During their recirculation process, naïve B lymphocytes may encounter antigen in peripheral lymphoid organs. The initial activation of B cells and their fate following antigen contact are intimately linked to their stimulation by cytokines and to costimulation of cell surface receptors. Main costimulation on B cells is mediated through CD40, a glycoprotein belonging to the tumor necrosis factor α family, which plays a critical role in B cell proliferation, survival, and in their preferential differentiation in memory cells rather than plasma cells [11]. CD40 plays an exquisite role in thymus-dependent humoral response, where its interaction with CD154 on activated T cells is accompanied by antibody affinity maturation and isotype switching from IgM to IgG, IgA, and IgE [12]. The intracellular pathways leading to plasma cell formation are complex and their characterization is still ongoing. Expression of B lymphocyte-induced maturation protein 1 (Blimp-1) and of transcription factor XBP-1 appear critical for plasma cells commitment [13]. Plasma cells and plasmablasts home mainly in the bone marrow, where the latter mature into plasma cells, and where these terminally differentiated cells persist for a long time [14]. Plasma cells are effector cells that synthesize immunoglobulins.

T lymphocyte precursors originate in the bone marrow and mature in the thymus. Mature T cells are separated in two main subsets, which can be differentiated according to the mutually exclusive expression of two polypeptides, named CD4 and CD8. Both proteins are cell surface molecules related to T cell subset functions that interact with determinants on major histocompatibility complex proteins and that are able of signal transduction through their cytoplasmic tail capable of interacting with lck, a Src family protein-tyrosine kinase molecule. The T helper/inducer lymphocytes express the CD4 peptide, whereas the T cytotoxic/suppressor lymphocytes harbor the surface CD8 molecule. These latter cells mediate cytotoxicity towards cells expressing on their surface antigen epitopes associated with class I major histocompatibility complex. Activation of naïve CD4 positive helper T (Th) cells is initiated through their T cell receptor binding to antigen epitopes coupled to class II major histocompatibility complexes. Additional costimulation through CD28, a homodimeric heavily glycosylated protein that binds to CD80 and CD86 on antigen-presenting cells, causes these cells to proliferate and differentiate into effector T helper cells. Two main subsets of effector Th cells, namely Th1 and Th2 cells, have been defined on the basis of their distinct cytokine secretion patterns and their different immunomodulatory effects [15]. Differentiation of uncommitted T cell towards Th1 and Th2 cells is dependent on several factors, including principally interleukin 12 (IL-12) for Th1 differentiation and IL-4 for Th2 fate (reviewed in [15]). Th1 cells produce primarily interferon-

gamma and tumor necrosis factor, which are required for cell-mediated inflammatory reactions, whereas Th2 cells secrete IL-4, IL-5, IL-10, and IL-13, which primarily mediate B cell activation and antibody production [16]. In general, an efficient clearance of intracellular pathogens is based on Th1 cell activation, while antibody responses triggered by Th2 cells are best suited for neutralizing extracellular infections. The commitment of naïve CD4 T cells to become Th1 and Th2 has therefore important consequences towards the success of an immune response to eradicate pathogens and influence therefore the progression of their associated diseases [17].

14.1.2 Human Immunodeficiency Virus and the Lymphocytes

The human immunodeficiency virus (HIV) was first identified in 1983 and, in 1984, it was shown to be the cause of AIDS [18–22]. HIV specifically targets CD4 lymphocytes, and these cells provide the necessary environment to sustain viral replication. AIDS is characterized by the depletion of the CD4 helper/inducer subset of T lymphocytes, leading to severe immunosuppression and development of opportunistic infections and neoplasms. During the acute HIV syndrome that follows infection, a burst of viremia can be detected and the level of CD4 cells declines but subsequently returns to near-normal levels. The levels of CD3, CD8, and B cells also drop during the acute HIV syndrome [23]. The median time between primary HIV infection and the development of AIDS is approximately 11 years. During this period, the number of CD4 cells usually decreases gradually until it reaches a level at which the risk of opportunistic diseases is high. However, some seropositive persons, infected for more than 10 years without treatment, have still apparently normal CD4 cell populations. The progression from initial infection to the diagnosis of AIDS is influenced mainly by three different factors: (1) the viral strain, (2) the host immune system, and (3) the genetic differences between individuals [24].

On the surface of the virus is the glycoprotein gp120/gp41. This protein binds to the CD4 protein that leads to the fusion between the viral envelope and the membrane of the lymphocyte. The striking depletion of CD4 lymphocytes observed in AIDS patients, combined with the extraordinary sensitivity of CD4 cells to infection by HIV *in vitro*, led to the discovery that the CD4 peptide is a high-affinity receptor for HIV [25]. The addition of soluble recombinant CD4 molecules to cultures of CD4 lymphocytes and HIV was shown to inhibit HIV infection of these cells. However, CD4 is not sufficient to allow HIV infection. Chemokine receptors CCR5 and CXCR4 are needed for optimal infection [26, 27], and it has been shown that deletion of 32 base pairs in the CCR5 gene leads to protection against HIV infection, without altering T cell activity [28, 29]. In addition, the HIV life cycle is characterized by interactions with a number of human proteins playing a role in the viral replication [30], and many host proteins have been proposed to directly interact with HIV (http://pim.hybrigenics.com). Certain host proteins are necessary for infection and for sustaining viral replication, while others repre-

sent antiviral factors. A number of host proteins are incorporated into the nascent virion. These numerous interactions with host cellular proteins, and polymorphism in the corresponding cellular genes probably constitute the basis for inter-individual variation in susceptibility to infection [24, 30].

14.2 Proteomics of Lymphocytes

14.2.1 Isolation of Lymphocytes

Methods used to purify lymphocytes have been described in detail elsewhere [31]. About 450 mL of blood, obtained from volunteer donors, was collected in plastic bags, and white blood cells as well as platelets were collected by filtration. Leukocytes were recovered by injecting, in the reverse sense of the filters, 3×10 mL of phophate-buffered saline (PBS) containing 10% citric acid-dextrose-adenin. Platelets, monocytes as well as lymphocytes were recovered and separated from residual red blood cells and granulocytes using Ficoll-Paque gradient centrifugation. From this cellular preparation, monocytes were separated from lymphocytes over a Percoll gradient. Positive selection of lymphocytes was then performed with beads coated with specific monoclonal antibodies (anti-CD19, anti-CD8 and anti-CD4, respectively). After washing, the purified lymphocytes were recovered from the beads. The purity of the three lymphocyte populations was >95%, as determined by cytofluorimetric analyses [31].

14.2.2 Two-dimensional Electrophoresis and Mass Spectrometry

The cellular proteins were solubilized in a solution composed of 8 M urea, 4% CHAPS, 40 mM Tris, and 65 mM 1,4-dithioerythritol, and isoelectric focusing (IEF) was performed using non-linear (NL) immobilized pH gradients (IPG) using pH range from 3 to10, or from 4 to 7. After equilibration, strips were placed on the top of 9–16% T gradient polyacrylamide second-dimensional gels. The silver-stained gel images were captured with a high-resolution densitometer, and analyzed using the software Melanie, version 3.0 [32, 33]. Heuristic clustering analysis was performed using a previously described algorithm [34, 35].

Coomassie blue stained spots were excised from the 2-D gels, and in-gel proteolytic cleavage with sequencing grade trypsin was performed. Supernatants containing proteolytic peptides were analyzed either by MALDI-TOF or by LC-MS/MS on a hybrid quadrupole-TOF instrument equipped with a nanoelectrospray source and interfaced to a low flow (200 nl min^{-1}) reversed-phase HPLC system (LC-Packings, Amsterdam, The Netherlands).

Peptide mass lists or tandem mass spectra of peptides were directly used for database search procedures using the ProFound peptide mass fingerprinting search

engine or MASCOT [36]. The databases employed were NCBInr, a compilation of several databases including, Swiss Prot, TrEMBL PIR, PRF, PDB, and GenBank CDS translations.

14.3 Results and Discussion

14.3.1 A (Preliminary) Reference Lymphocyte Map (IPG: 3–10 NL)

After 2-DE of T lymphocytes (CD4 and CD8) and B (CD19) lymphocytes, 1411 ± 73 spots (mean ± SD; $N = 9$) were detected [31]. No significant differences in the number of spots were found between the three populations of lymphocytes ($P = 0159$; one-way analysis of variance). Mass fingerprinting analysis was performed on 125 spots retrieved from preparative gels derived from a pool of CD8 lymphocytes, collected from three different donors. Sixty-four different proteins were identified (Fig. 14.1). A large variety of structural proteins and of enzymes is represented (Table 14.1). The actualized lymphocyte map is available at the following internet address: http://www.expasy.org/cgi-bin/map2/def?LYMPHOCYTE_HUMAN.

The identification of the other spots, and particularly those situated in the higher molecular area of the gel is in progress.

14.3.2 Differential Expression of Proteins in CD4, CD8, and CD 19 Lymphocytes

When 2-D gels from CD4, CD8, and CD19 lymphocytes were compared using stringent criteria (presence versus absence and large spot sizes), a restricted number of spots appeared to be population specific. Using this approach, five spots were highlighted, and analyzed by liquid chromatography-tandem mass spectrometry. Their characteristics are presented in Table 14.2. Figure 14.2 shows the discriminating spots, from which some appeared to be T cell related (present only in CD8 and CD4 gels), whereas, as shown in Figure 14.3, two appeared to be CD8 associated [31]. The cluster of spots, characterized by an apparent M_r of 50 100, corresponded to a mixture of proteins containing vimentin, tubulin α, and desmin, respectively. The difference of the expression of vimentin or tubulin α, between T and B cells, was confirmed by western blotting, using specific antibodies (Fig. 14.4). This observation is of possible functional importance, because the intermediate filament proteins have key specific functions in signal transduction (reviewed in [37]). Desmin and vimentin belong to the intermediate filament protein family and play an essential part in the rigidity of circulating lymphocytes. Vimentin is found in a complex with plectrin and fodrin in peripheral blood T lymphocytes. These three proteins are involved in the migration of T lymphocytes [38]. Tubulin α is a protein of the microtubule that takes part in the structure of the cell as well as in its motility [39]. T lymphocytes are dynamic cells, continually

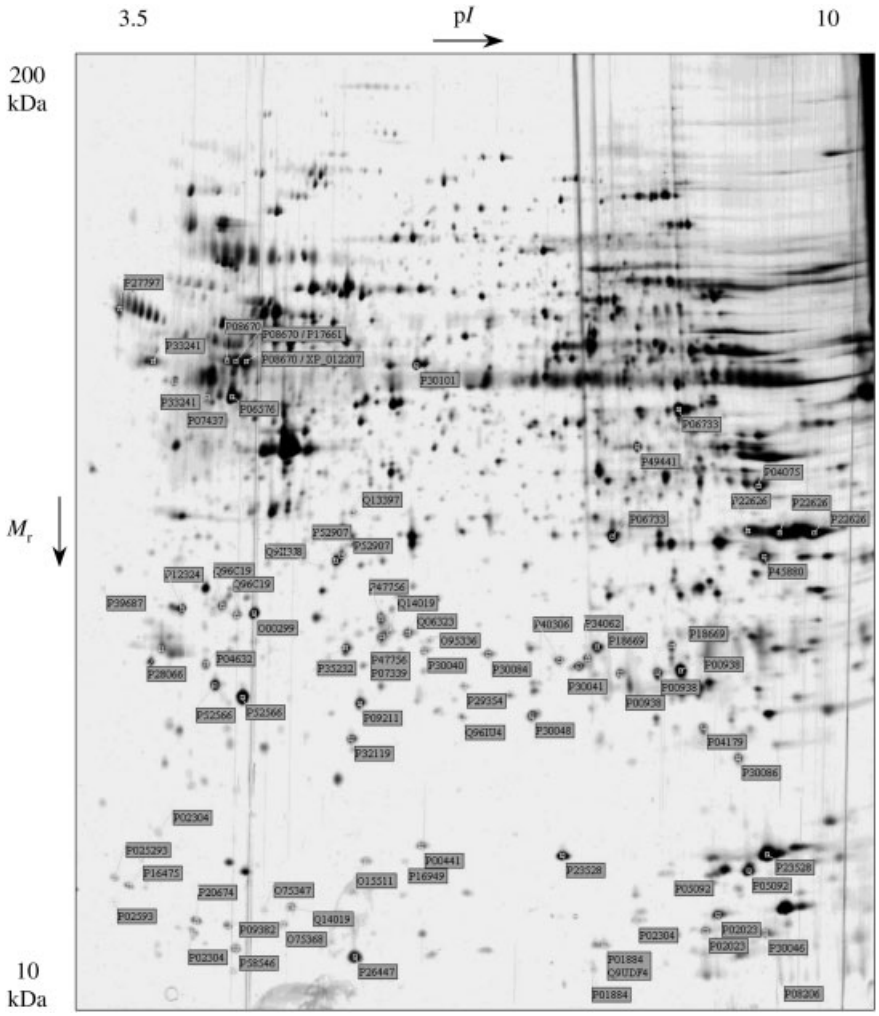


Figure 14.1 Silver-stained 2-DE gel of a pool of purified (>95%) CD8 lymphocytes isolated from healthy blood donors. First dimension: immobilized 3–10 NL pH gradient, second dimension: 9–16% T polyacrylamide gel electrophoresis. Proteins were identified by mass spectrometry and are designated using their SwissProt accession number. This map is also available at the following internet address: http://www.expasy.org/cgi-bin/map2/def?LYMPHOCYTE_HUMAN.

seeking contacts and connections with their environment. These cells migrate towards sites of inflammation. The contortions necessary for their extravasation require coordination between extracellular receptors, and intracellular cytoskeletal networks composed notably of actin and tubulin [40]. Therefore, reorganization of tubulin is a crucial event implied in T cell mobility [41]. In addition, efficient killing by cytotoxic T cells requires translocation of the microtubule-organizing center

Table 14.1 Proteins identified by mass spectrometry.

Identification	*SwissProt*	*pI*[a)]	M_r[a)]	*Function*
Acidic leucine-rich nuclear phosphoprotein 32	P39687	3.99	28585	Inhibitor protein of protein phosphatase 2A
Actin β[b)]	P02570	5.29	41736	Involved in cell motility
Adenine phosphoribosyltransferase	P07741	5.79	19476	Catalyzes a reaction resulting in the formation of AMP
Antioxydant protein	P30041	6.02	24904	Involved in redox regulation of the cell
ARP2/3 complex 16 kDa	O15511	5.47	16320	Impli cated in the control of actin polymerization
ATP synthase β-chain	P06576	5.00	51769	Produces ATP from ADP (catalytic subunit)
β_2-Microglobulin precursor	P01884	6.06	13714	Subunit of the major histocompatibilty complex class I
Calcium-dependent protease, small subunit	P04632	5.05	28316	Involved in cytoskeletal remodeling and signal transduction
Calmodulin	P02593	4.09	16706	Mediates the control of a number of enzymes
Calpactin I light chain	P08206	7.30	11071	Induces the dimerization of annexin II
Calreticulin precursor	P27797	4.29	48141	Calcium binding protein
Cathepsin D precursor	P07339	6.10	44552	Protease active in intracellular protein breakdown
Chloride intracellular channel protein 1	O00299	5.09	26922	Is active as a chloride ion channel
Cofilin	P23528	8.22	18502	Controls the reversibility of actin polymerization
Cytochrome C oxidase polypeptid VA	P20674	4.88	12513	Oxydase in the mitochondria
D-Dopachrome tautomerase	P30046	7.25	12580	Tautomerization with decarboxylation
Desmin[b)]	P17661	5.21	53404	Intermediate filament
Elongation factor Tu, mitochondrial precursor	P49411	7.26	49541	Promotes the GTP binding of aminoacyl-tRNA to the A-site of ribosomes
Endoplasmic reticulum protein ERp29 precursor	P30040	6.77	29993	Plays a role in the processing of secretory proteins
Enolase α	P06733	6.99	47037	Enzyme
Enoyl-CoA hydratase, mitochondrial precursor	P30084	8.34	37371	Enzyme
Eukaryotic translation initiation factor 3 subunit	Q13347	5.38	36501	Binds to the 40S ribosome
F-actin capping protein α_1-subunit	P52907	5.45	32923	Binds to the fast growing ends of actin filaments

Table 14.1 (cont.)

Identification	*SwissProt*	*pI*[a)]	M_r[a)]	*Function*
F-actin capping protein β-subunit	P47756	5.36	31350	Binds to the fast growing ends of actin filaments
Fructose-bisphosphate aldolase A	P04075	8.39	39289	Enzyme
Galectin-1	P09382	5.34	14584	May regulate cell apoptosis and cell differentiation
Glutathione *S*-transferase P	P09211	5.44	23224	Reduces glutathion
Growth factor receptor-bound protein 2	P29354	5.89	26206	Associates with epidermal growth factors receptors
Hemoglobin β-chain	P02023	6.81	15867	Red cell contaminant of the preparation
Heterogeneous nuclear ribonucleoproteins A2/B1	P22626	8.97	37430	Involved in pre mRNA processing
Histone H4	P02304	11.36	11236	Plays a central role in nucleosome formation
Hypothetical protein FLJ90553	Q96IU4	5.94	22345	Unknown
Lymphocyte-specific protein 1	P33241	4.69	37191	Unknown
My027 protein	Q9H3J8	5.40	33232	Unknown
Myosin light chain alkali, non-muscle isoform	P16475	4.56	16799	Unknown
Myotrophin	P58546	5.28	12763	Unknown
Peptidyl-prolyl *cis-trans* isomerase A	P05092	7.82	17881	Accelerates the folding of proteins
Peroxiredoxin 2	P32119	5.66	21892	Involved in redox regulation of the cell
Phosphatidylethanolamine-binding-protein	P30086	7.43	20925	Binds ATP opioids and phosphatidylethanolamine
6-Phosphogluconolactonase	O95336	5.70	27546	Enzyme
Phosphoglycerate mutase 1	P18669	6.75	28672	Enzyme
Placental calcium-binding protein	P26447	5.85	11728	Potential calcium binding protein
Prohibitin	P35232	5.57	29804	Inhibits DNA synthesis
Proteasome activator complex subunit 1	Q06323	5.78	28723	Implicated in proteasome assembly
Proteasome subunit α type 5	P28066	4.69	26469	Multicatalytic proteinase complex
Proteasome subunit α type 6	P34062	6.35	27399	Multicatalytic proteinase complex
Proteasome subunit β type 10	P40306	7.69	28936	Multicatalytic proteinase complex
Protein disulfide isomerase	P07237	4.76	56116	Rearrangement of disulfide bonds in proteins
Protein disulfide isomerase A3 precursor	P30101	5.99	56782	Rearrangement of disulfide bonds in proteins

Table 14.1 (cont.)

Identification	*SwissProt*	*pI*[a)]	M_r[a)]	*Function*
Rho GDP-dissociation inhibitor 2	P52566	5.1	22988	Regulates the GDP/GTP exchange reaction of the Rho proteins
SH3 domain-binding glutamic acid-rich like protein	O75368	5.24	12774	Unknown
Stathmin	P16949	5.77	17171	Prevents assembly and promotes disassembly of microtubules
Superoxide dismutase	P00441	5.70	15804	Destroys toxic radicals within the cells
Superoxide dismutase mitochondrial precursor	P04179	8.50	24722	Destroys toxic radicals within the cells
Swiprosin 1[c)]	Q96C19	5.15	26697	Potential calcium binding protein
Thioredoxin-dependent peroxide reductase	P30048	7.68	27962	Involved in redox regulation of the cell
Triosephosphate isomerase	P00938	6.51	26538	Plays roles in several metabolic pathways
Tropomyosin, cytoskeletal type	P06753	4.69	28995	Implicated in stabilizing cytoskeleton actin filaments
Tubulin α[b)]	P05209	4.94	50151	Major constituent of microtubules
Tubulin β	P07437	4.75	49759	Major constituent of microtubules
Tubulin-specific chaperone A	O75347	5.25	12724	Involved in the early step of the tubulin folding pathway
Vimentin[b)]	P08670	4.80	53554	Intermediate filaments found in various cells
Voltage-dependent anion-selective channel protein 2 (isoform 4)	P45880	7.51	33000	Forms a channel through the mitochondrial membrane

a) Calculated values.
b) Expressed predominantly in CD8 and CD4 lymphocytes.
c) Expressed predominantly in CD8 lymphocytes.

Table 14.2 Major spot differences between populations of lymphocytes (3–10 NL pH gradients).

% *Vol*			*pI* [a)]	M_r [a)]	*Identification*	*SwissProt*
CD8	*CD4*	*CD19*				
0.502 ± 0.377	0.854 ± 0.214	Absent	5.2	50 100	Vimentin	P08670
					Tubulin α	P05209
0.311 ± 0.043	0.157 ± 0.013	Absent	5.3	50 100	Vimentin	P08670
					Desmin	P17661
0.964 ± 0.317	0.576 ± 0.148	Absent	5.4	50 100	Vimentin	P08670
0.167 ± 0.047	Absent	Absent	4.9	30 400	Swiprosin 1	Q96C19
0.128 ± 0.016	Absent	Absent	5.1	29 800	Swiprosin 1	Q96C19

a) pI and M_r were calculated using Melanie 3.0, taken the values of known proteins such as actin, cathepsin D, glyceraldehyde 3-phosphate dehydrogenase, superoxide dismutase, and glutathione *S*-transferase P that were obtained by comparison with the map of a lymphoma available in the Swiss 2-D PAGE database (http://www.expasy.ch/ch2d/). (Adapted from [31], with permission of the publisher).

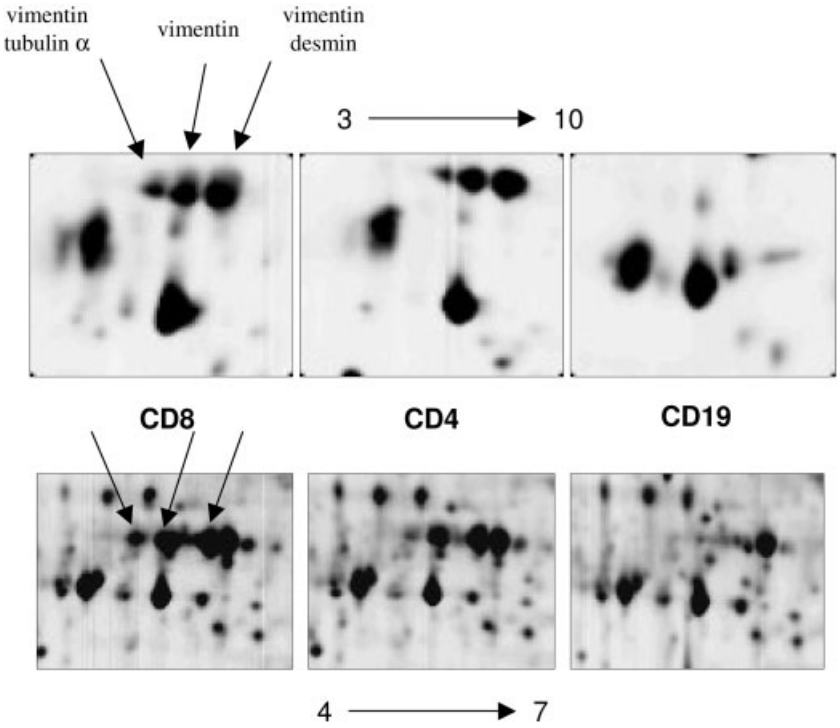


Figure 14.2 Details of silver-stained 2-DE gel of three different populations of lymphocytes showing a cluster of discriminating spots, present in CD8 and CD4, but absent in CD19 lymphocytes. First dimension: immobilized 3–10 N pH gradient (upper panels), or immobilized 4–7 NL pH gradient (lower panels). Second dimension: 9–16% T polyacrylamide gel electrophoresis. (Adapted from [31], with permission of the publisher).

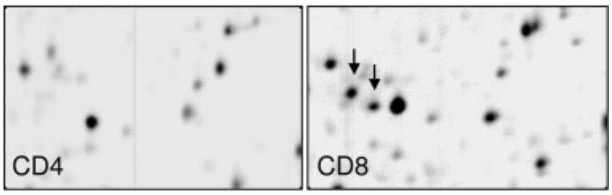


Figure 14.3 Details of high-resolution silver-stained 2-D gels of CD4 and CD8 lymphocytes showing the spots corresponding to swiprosin 1 (Q96C19) that are highlighted by arrows. First dimension: immobilized 3–10 pH gradient, second dimension: 9–16% T polyacrylamide gel electrophoresis. (Adapted from [31], with permission of the publisher).

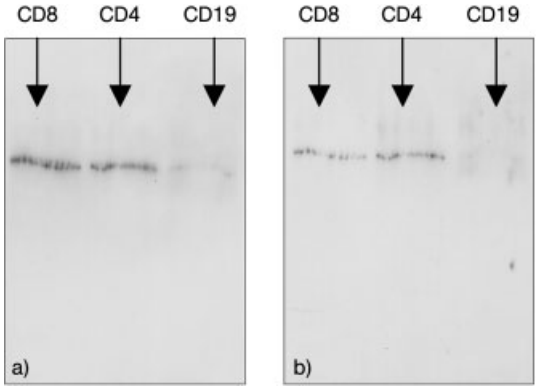


Figure 14.4 Western blots of CD8, CD4, and CD19 lymphocyte proteins separated by SDS-PAGE. (a) Monoclonal anti-vimentin antibody. (b) Monoclonal anti-tubulin α antibodies. The antigen/antibody reaction was revealed by chemiluminescence. The amount of protein loaded for electrophoresis (20 µg) was identical for the three samples.

	1	11	21	31	41	51
1	MATDELATKL	SRRLQMEGEG	GGETPEQPGL	NGAAAAAAGA	PDEAAEALGS	ADCELSAKLL
61	RRADLNQGIG	EPQSPSRRVF	NPYTEFKEFS	RKQIKDMEKM	FKQY*DAGRDG*	*FIDLME*LKLM
121	MEKLGAPQTH	LGLKNMIKEV	***DEDFDSKLSF***	***RE***FLLIFRKA	AAGELQEDSG	LCVLARLSEI
181	DVSSEGVKGA	KSFFEAKVQA	INVSSRFEEE	IKAEQEERKK	QAEEMKQRKA	AFKELQSTFK

Figure 14.5 Sequence of swiprosin 1 (Q96C19). This protein is characterized by the presence of two EF-hand domains (in italics and bold), that may confer Ca^{2+} binding activity, whereas the amino acids identified by mass spectrometry (LC-MS/MS) are underlined. The sequence of the gene coding to this protein was reported in 2002 [44].

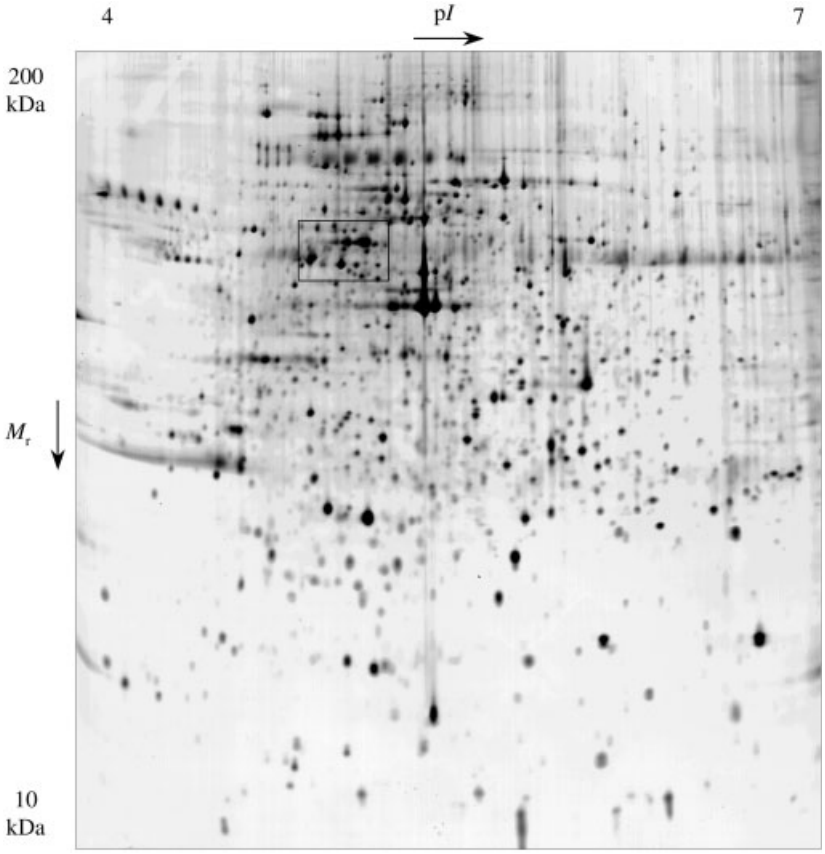


Figure 14.6 Silver-stained 2-DE gel of a pool of purified CD8 lymphocytes isolated from healthy blood donors. First dimension: immobilized 4–7 NL pH gradient, second dimension: 9–16% polyacrylamide gel electrophoresis. The frame highlights the cluster of proteins related to vimentin, tubulin α and desmin. The number of spots detected (1447 ± 198; mean ± SD, $N = 12$) is similar to that detected when using immobilized 3–10 NL pH gradients (1411 ± 73, $N = 9$), indicating an important enrichment of the proteins resolved in these electrophoretic conditions.

[42]. Finally, the microtubule organizing center also appears to be involved in the regulation of the tyrosine kinase PYK2 in T cells [43].

The identification of the apparently CD8-specific spots (Fig. 14.3) revealed that they corresponded to a new protein, named swiprosin 1 (Q96C19). This protein is characterized by the presence of two EF-hand domains (Fig. 14.5), that may confer Ca^{2+}-binding activity. The sequence of the gene coding to this protein was reported in 2002 [44].

CD4, CD8, and CD19 lymphocytes were further analyzed using 4–7 NL pH gradient for IEF (Fig. 14.6). By this approach, 1447 ± 198 (mean ± SD, $N = 12$) spots

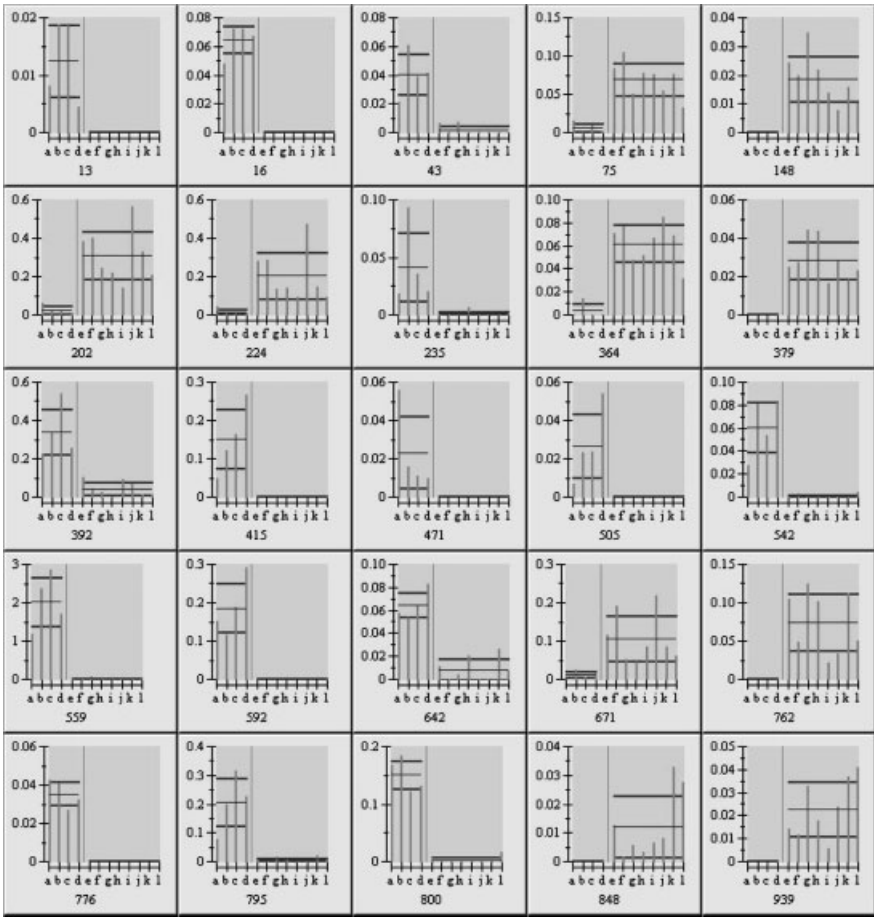


Figure 14.7 Examples of histograms showing discriminating spots (expressed as % volume) between B and T lymphocyte populations. (a–d) CD19 lymphocytes, (e–h) CD4 lymphocytes, and (i–l) CD8 lymphocytes.

were detected, without significant differences between the three populations of lymphocytes ($P=0.311$; one-way analysis of variance). With the four gels of each lymphocyte population, three synthetic gels were created that contained 728 spots (CD19), 655 spots (CD8), and 872 spots (CD4), respectively. Then, a master gel containing 1150 spots was constructed, and these spots were matched with those of the 12 gels. Fifty-four spots segregated B (CD19) from T (CD4 and CD8) cells (Fig. 14.7). The identification of these spots is currently in progress, but appears to be quite difficult, most of them being low-abundance proteins.

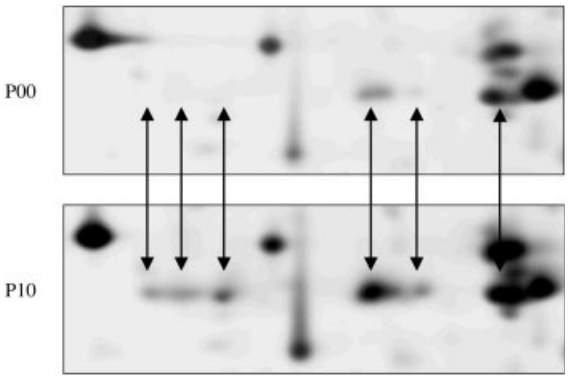


Figure 14.8 Isoforms of cyclophilin A in fetal skin cells in culture at passage zero (P00) and at passage 10. Note the appearance of acidic isoforms (related to oxidation of cytein residues into cysteic acid) after cell passaging. (Adapted from [49], with permission of the publisher).

14.3.3 Applications in HIV Physiopathology

Among the proteins identified on the lymphocyte 2-D map, one has been implicated in the pathogenesis of HIV infection. This protein is peptidyl-prolyl *cis-trans* isomerase A (P05092; also known under the names of cyclophilin A or cyclosporin A-binding protein) [45–47]. Various isoforms of the protein have been identified, and each appeared to play a critical role either in the attachment of HIV to the surface of target cells or in the regulation of the conformation of the HIV capsid protein [48]. Isoforms of cyclophilin A can be readily identified after 2-DE (Fig. 14.8). As observed in a study of the plasticity of protein expression of fetal skin cells in cultures, additional isoforms of the protein may be observed during the time of culture [49]. This finding may have implications for further studies on the role of cyclophilin A in cellular HIV inflection.

We used proteomics to assess the characteristics of lymphocytes from selected individuals, differing in their susceptibility to HIV infection. CD4 lymphocyte populations were selected according to their permissiveness to HIV infection [50, 51], and the two groups of cells were evaluated according to either their hyper-susceptibility or to their resistance phenotype. Protein pattern modifications were observed according to the susceptibility to HIV infection, as shown in Figure 14.9. The identification of the implicated protein is still in progress. Nevertheless, these preliminary results clearly show that proteomics may be useful to characterize some key host lymphocytic proteins participating in the viral life cycle.

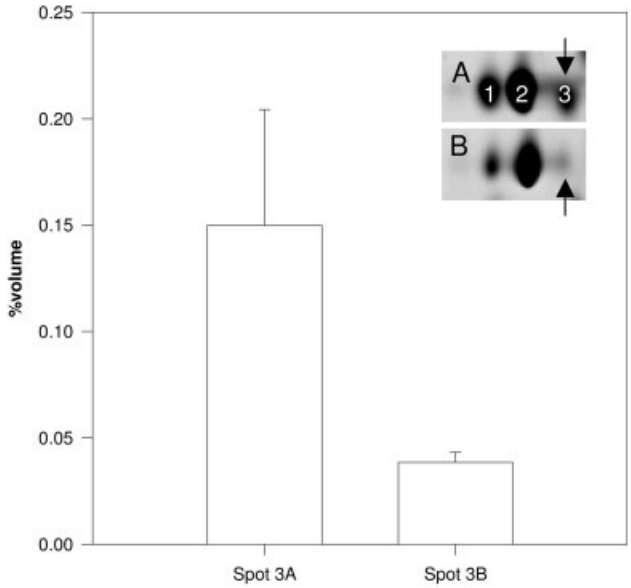


Figure 14.9 Spot pattern modifications of CD4 lymphocytes either hyperpermissive (inset A) or hypopermissive (inset B) to HIV infection. The discriminating spot is arrowed. The % volume of spot No. 3 show statistically significant different results between the two groups ($P=0.006$; *t*-test, $N=4$ in each groups).

14.4 Summary

Proteomic studies have been already applied to the study of hematopoietic cell lines and blood cells (reviewed in [52]). Since the pioneer 2-DE investigations of lymphocyte proteins [53–68], a number of proteomic studies about lymphocytes have been published [69, 70], including B cell chronic leukemia or lymphoma [69, 71, 72]. Finally, remarkable studies on lymphoblastoid proteins, including phosphoproteins, were published [69, 73–80]. More than 60 different proteins were identified, and several phosphorylated peptides were characterized [73].

In this chapter, we have briefly summarized our experience in proteomics of human lymphocytes collected from healthy blood donors, and have shown that the global protein pattern of the three main lymphocyte populations is very similar, most likely because all these cells derive from a common progenitor. In addition, it must be mentioned that with the technique of protein separation used in this study, low-abundance proteins as well as several membrane proteins were either poorly or not detected. Nevertheless, by heuristic analysis, it was possible to show that B cells segregate from T cells. Some discriminating spots were identified, and a novel protein, named swiprosin 1, was discovered. Swiprosin 1 is a

member of a new family of proteins detected in human (Q96C19, Q9BUP0), mouse (Q9D8Y0, Q9D4J1), and *Drosophila melanogaster* (Q9VJ26).

Finally, it must be pointed out that the lymphocytic map presented in this chapter is the first showing proteins that are common to B, CD4 and CD8 lymphocytes isolated from healthy individuals. The establishment of such a 2-D map is certainly a cornerstone in the study of diseases affecting the lymphocytes, such as HIV infection. All these studies open up new areas of rapidly moving research concerning hematologists, virologists, and immunologists.

14.5 References

1 Hanash, S. *Nature* **2003**, *422*, 226–232.
2 Tyers, M., Mann, M. *Nature* **2003**, *422*, 193–197.
3 Aebersold, R., Mann, M. *Nature* **2003**, *422*, 198–207.
4 Boguski, M.S., McIntosh, M.W. *Nature* **2003**, *422*, 233–237.
5 Verfaillie, C.M. *Trends Cell Biol.* **2002**, *12*, 502–508.
6 Kondo, M., Weissman, I.L., Akashi, K. *Cell* **1997**, *91*, 661–672.
7 Busslinger, M., Nutt, S.L., Rolink, A.G. *Curr. Opin. Immunol.* **2000**, *12*, 151–158.
8 Delves, P.J., Roitt, I.M. *N. Engl. J. Med.* **2000**, *343*, 37–49.
9 Delves, P.J., Roitt, I.M. *N. Engl. J. Med.* **2000**, *343*, 108–117.
10 Hardy, R.R., Hayakawa, K. *Annu. Rev. Immunol.* **2001**, *19*, 595–621.
11 Duchosal, M.A. *Semin. Hematol.* **1997**, *34*, 2–12.
12 Banchereau, J., Bazan, F., Blanchard, D. et al. *Annu. Rev. Immunol.* **1994**, *12*, 881–922.
13 Angelin-Duclos, C., Cattoretti, G., Lin, K.I., Calame, K. *J. Immunol.* **2000**, *165*, 5462–5471.
14 Tew, J.G., DiLosa, R.M., Burton, G.F. et al. *Immunol. Rev.* **1992**, *126*, 99–112.
15 Murphy, K.M., Reiner, S.L. *Nature Rev. Immunol.* **2002**, *2*, 933–944.
16 Chinen, J., Shearer, W.T. *J. Allergy Clin. Immunol.* **2003**, *111*, S813–S818.
17 Diehl, S., Rincon, M. *Mol. Immunol.* **2002**, *39*, 531–536.
18 Montagnier, L. *Science* **2002**, *298*, 1727–1728.
19 Gallo, R.C., Montagnier, L. *Science* **2002**, *298*, 1730–1731.
20 Sleasman, J.W., Goodenow, M.M. *J. Allergy Clin. Immunol.* **2003**, *111*, S582–S592.
21 Prusiner, S.B. *Science* **2002**, *298*, 1726.
22 Gallo, R.C. *Science* **2002**, *298*, 1728–1730.
23 Chinen, J., Shearer, W.T. *J. Allergy Clin. Immunol.* **2002**, *110*, 189–198.
24 Chun, T.W., Justement, J.S., Lempicki, R.A. et al. *Proc. Natl Acad. Sci. USA* **2003**, *100*, 1908–1913.
25 McMichael, A.J., Rowland-Jones, S.L. *Nature* **2001**, *410*, 980–987.
26 O'Brien, S.J., Moore, J.P. *Immunol. Rev.* **2000**, *177*, 99–111.
27 Cilliers, T., Nhlapo, J., Coetzer, M. et al. *J. Virol.* **2003**, *77*, 4449–4456.
28 Paxton, W.A., Kang, S., Liu, R. et al. *Immunol. Lett.* **1999**, *66*, 71–75.
29 Paxton, W.A., Neumann, A.U., Kang, S. et al. *J. Infect. Dis.* **2001**, *183*, 1678–1681.
30 Greene, W.C., Peterlin, B.M., *Nature Med.* **2002**, *8*, 673–680.
31 Vuadens, F., Gasparini, D., Deon, C. et al. *Proteomics.* **2002**, *2*, 105–111.
32 Appel, R.D., Palagi, P.M., Walther, D. et al. *Electrophoresis* **1997**, *18*, 2724–2734.
33 Appel, R.D., Vargas, J.R., Palagi, P.M., Walther, D., Hochstrasser, D.F. *Electrophoresis* **1997**, *18*, 2735–2748.
34 Appel, R., Hochstrasser, D., Roch, C., Funk, M., Muller, A.F., Pellegrini, C. *Electrophoresis* **1988**, *9*, 136–142.
35 Appel, R.D., Hochstrasser, D.F. *Methods Mol. Biol.* **1999**, *112*, 363–381.

36 Perkins, D. N., Pappin, D. J., Creasy, D. M., Cottrell, J. S. *Electrophoresis* **1999**, *20*, 3551–3567.

37 Paramio, J. M., Jorcano, J. L. *Bioessays* **2002**, *24*, 836–844.

38 Brown, M. J., Hallam, J. A., Liu, Y., Yamada, K. M., Shaw, S. *J. Immunol.* **2001**, *167*, 641–645.

39 Hubbert, C., Guardiola, A., Shao, R. et al. *Nature* **2002**, *417*, 455–458.

40 Punt, J., Owen, J. *Trends Immunol.* **2001**, *22*, 419–420.

41 Volkov, Y., Long, A., McGrath, S., Ni, E. D., Kelleher, D. *Nature Immunol.* **2001**, *2*, 508–514.

42 Kuhn, J. R., Poenie, M. *Immunity* **2002**, *16*, 111–121.

43 Rodriguez-Fernandez, J. L., Sanchez-Martin, L., de Frutos, C. A. et al. *J. Leukoc. Biol.* **2002**, *71*, 520–530.

44 Strausberg, R. L., Feingold, E. A., Grouse, L. H. et al. *Proc. Natl Acad. Sci. USA* **2002**, *99*, 16899–16903.

45 Misumi, S., Fuchigami, T., Takamune, N., Takahashi, I., Takama, M., Shoji, S. *J. Virol.* **2002**, *76*, 10000–10008.

46 Braaten, D., Luban, J. *EMBO J.* **2001**, *20*, 1300–1309.

47 Thali, M., Bukovsky, A., Kondo, E. et al. *Nature* **1994**, *372*, 363–365.

48 Sherry, B., Zybarth, G., Alfano, M. et al. *Proc. Natl Acad. Sci. USA* **1998**, *95*, 1758–1763.

49 Vuadens, F., Crettaz, D., Scelatta, C. et al. *Electrophoresis* **2003**, *24*, 1281–1291.

50 Spira, A. I., Ho, D. D. *J. Virol.* **1995**, *69*, 422–429.

51 Eisert, V., Kreutz, M., Becker, K. et al. *Virology* **2001**, *286*, 31–44.

52 Tissot J. D., Schneider P. In *Biomedical and Pharmaceutical Applications of Proteomics* (Hondermarck, H., Eds). Kluwer Academic Publisher, Amsterdam, in press.

53 Pasquariello, A., Ferri, C., Moriconi, L. et al. *Am. J. Nephrol.* **1993**, *13*, 300–304.

54 Hanash, S. M., Strahler, J. R., Chan, Y. et al. *Proc. Natl Acad. Sci. USA* **1993**, *90*, 3314–3318.

55 Lester, E. P., Lemkin, P., Lipkin, L. *Clin. Chem.* **1982**, *28*, 828–839.

56 Hochstrasser, D., Roux, P., Dayer, J. M. et al. *Schweiz. Med. Wochenschr.* **1986**, *116*, 1773–1775.

57 Willard, K. E., Anderson, N. G. *Clin. Chem.* **1981**, *27*, 1327–1334.

58 Willard, K. E. *Clin. Chem.* **1982**, *28*, 1031–1035.

59 Willard, K. E. *Clin. Chem.* **1982**, *28*, 1074–1083.

60 Willard-Gallo, K. E., Houck, D. W., Loken, M. R. *Eur. J. Immunol.* **1988**, *18*, 1453–1461.

61 Willard, K. E., Thorsrud, A. K., Munthe, E., Jellum, E. *Clin. Chem.* **1982**, *28*, 1067–1073.

62 Willard-Gallo, K. E., Humblet, Y., Symann, M. *Clin. Chem.* **1984**, *30*, 2069–2077.

63 Willard-Gallo, K. E. *Ann. NY Acad. Sci.* **1984**, *428*, 201–222.

64 Gemmell, M. A., Anderson, N. L. *Clin. Chem.* **1982**, *28*, 1062–1066.

65 Kuick, R. D., Hanash, S. M., Strahler, J. R. *Electrophoresis* **1988**, *9*, 192–198.

66 Giometti, C. S., Willard, K. E., Anderson, N. L. *Clin. Chem.* **1982**, *28*, 955–961.

67 Hanash, S. M., Baier, L. J., Welch, D., Kuick, R., Galteau, M. *Am. J. Hum. Genet.* **1986**, *39*, 317–328.

68 Nyman, T. A., Rosengren, A., Syyrakki, S., Pellinen, T. P., Rautajoki, K., Lahesmaa, R. *Electrophoresis* **2001**, *22*, 4375–4382.

69 Caron, M., Imam-Sghiouar, N., Poirier, F. et al. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* **2002**, *771*, 197–209.

70 Antonucci, F., Chilosi, M., Santacatterina, M., Herbert, B., Righetti, P. G. *Electrophoresis* **2002**, *23*, 356–362.

71 Hanash, S. M., Teichroew, D. *Electrophoresis* **1998**, *19*, 2004–2009.

72 Voss, T., Ahorn, H., Haberl, P., Dohner, H., Wilgenbus, K. *Int. J. Cancer* **2001**, *91* , 180–186.

73 Toda, T., Sugimoto, M., Omori, A., Matsuzaki, T., Furuichi, Y., Kimura, N. *Electrophoresis* **2000**, *21*, 1814–1822.

74 Feuillard, J., Schuhmacher, M., Kohanna, S. et al. *Blood* **2000**, *95*, 2068–2075.

75 Fouillit, M., Joubert-Caron, R., Poirier, F. et al. *Glycobiology* **2000**, *10*, 413–419.

76 Imam-Sghiouar, N., Laude-Lemaire, I., Labas, V. et al. *Proteomics* **2002**, *2*, 828–838.

77 Joubert-Caron, R., Le Caer, J.P., Montandon, F. et al. *Electrophoresis* **2000**, *21*, 2566–2575.
78 Poirier, F., Pontet, M., Labas, V. et al. *Electrophoresis* **2001**, *22*, 1867–1877.
79 Poirier, F., Bourin, P., Bladier, D., Joubert-Caron, R., Caron, M. *Cancer Cell Int.* **2001**, *17*, 1–12.
80 Poirier, F., Imam, N., Pontet, M., Joubert-Caron, R., Caron, M. *J. Chromatogr. B Biomed. Sci. Appl.* **2001**, *753*, 23–28.

15
Modifications of Host Cell Proteome Induced by Herpes Simplex Virus Type 1

Anna Greco, Yohann Couté, Stéphane Giraud, and Jean-Jacques Diaz

15.1
Introduction

Herpes simplex viruses type 1 and type 2 (HSV-1 and HSV-2 respectively) are human herpes viruses constituted by an outer envelope, a tegument, a capsid and a linear, double-stranded DNA. These two viruses are responsible for a variety of different clinical manifestations in humans. HSV-1 and HSV-2 are distributed worldwide, and are transmitted from infected to susceptible individuals during close personal contact. Herpes infections are generally localized in the face and in the trunk in the case of HSV-1, and in the genital area in the case of HSV-2, although an increasing proportion of genital infection could also be due to HSV-1.

During primary infection, the virus enters sensory nerves that innervate the mucosal membranes. After primary infection, which may eventually be asymptomatic, the virus persists in the host, generally for its lifetime, in an apparently inactive state called the latent state. Most probably, latent viruses reside in the ganglia that innervate the sites of inoculation: the trigeminal ganglia after oro-facial infection and sacral ganglia after genital infection. Once latency has been established, the latent virus can be periodically reactivated by numerous stimuli, and is transported back to peripheral tissues usually at or near the site of primary infection [1].

Herpes infections have been recognized since ancient times. They are very frequent and genital herpes is at present epidemic in the United States [2]. Currently, the prevalence of herpes simplex infection is between 70 and 90% of the world's adult population. It is estimated that over one-third of the world's population have recurrent HSV infections between 3 and 5 times a year, and, therefore, the capability of transmitting HSV during episodes of productive infection. Viral reactivation occurs generally following a fever, stress, exposure to ultraviolet light, menstrual cycle, pregnancy, etc. However, the rhythm and the intensity of the reactivation vary from one individual to another.

The clinical manifestations of the infection and of the reactivation are numerous; some of them are severe and can lead to death. The epidemiology and clinical manifestations of herpes infections have been described in detail in [1]. For example, oropharyngeal primary infection of children or young adults can induce a

Biomedical Application of Proteomics
Edited by J.-C. Sanchez, G. L. Corthals, D. F. Hochstrasser

ISBN: 3-527-30807-5

gingivostomatitis, pharyngitis, or a mononucleosis-like syndrome. Primary genital infection induces the appearance of macules and papules, followed by vesicles, pustules, and ulcers; often, primary infection is associated with fever, dysuria, localized inguinal adenopathy, and malaise. The most severely affected population are newborn children, who are infected at the time of birth and who present irreversible neurological damage in spite of antiviral therapy. This population is going to increase because of the increase in genital herpes infections in recent decades. It has been shown that 61% of children less than one year old contaminated at the time of delivery and who present a spread HSV infection, die as a consequence of this infection.

Herpes infections are the most common causes of genital ulcers in industrialized countries. The prevalence of genital herpes is between 20% and 90% of the world's adult population, according to socio-cultural context. Finally, patients compromised by immunosuppressive therapy (drug therapy treatment, bone marrow, or organ transplantation) or after infection by HIV develop severe HSV infections with extensive morbidity.

For more than 15 years, acyclovir (ACV) and a few related drugs have been accepted as the treatment of choice for the management of HSV infections. ACV is an analogue of deoxyguanosine whose antiherpes activity was described for the first time in the late 1970s [3]. It is a prodrug specifically phosphorylated by viral thymidine kinase of infected cells. Incorporation of the nucleotide analogue within the viral DNA prevents further polymerization. Since the discovery of ACV, several nucleoside analogues have been launched. Today, the nucleosides acyclovir (Zovirax), valacyclovir (Valtrex), penciclovir (Vectavir, Denavir), and famciclovir (Famvir) represent the standard therapy for herpes simplex disease. Valacyclovir and famciclovir are prodrugs of acyclovir, which offer a more convenient dosing schedule due to improved oral bioavailability and pharmacokinetics. A second generation of nucleosidic drugs, pyrimidine analogues such as brivudine (Helpin) and sorivudine (Usevir), which mimic dTTP have been developed, but further development were halted due to the narrow antiviral spectrum, mutagenic potential and drug-to-drug interactions of these molecules.

The widespread use of ACV and its derivatives has led to the emergence of HSV strains that are resistant to ACV. The prevalence of resistance to ACV in immunocompetent patients is low, below 1% [4], but the prevalence of resistance in immunocompromised patients can be very high, depending on the type of immunosuppression. According to several reports, prevalence of ACV resistance among these patients is about 5%. However resistance has also been detected among 2.8% of solid organ transplant patients, 3.5% of HIV-infected patients, and 14% of bone marrow transplant patients. The incidence of resistance in this last population can vary from 2% for autologous transplant patients up to 29% for allogenic transplant patients [5]. Mechanisms of resistance have been well decrypted. Mutations of the two viral genes coding for the enzymes involved in ACV metabolism account for resistance to ACV. These mutations induce modifications of the activity of these two enzymes. The more frequent ACV-resistant viral strains, which occur in approximately 95–96% of cases, display either a complete or partial loss of

thymidine kinase activity or an alteration of thymidine kinase substrate specificity. Alternatively, a few ACV-resistant viruses display mutations in the DNA sequences coding for the viral DNA polymerase, which result in alterations of their DNA polymerase activity [5].

All these considerations point out the crucial need for the development of new antiherpetic drugs that can not only inhibit replication of wild-type HSV viruses, but also inhibit replication of HSV viruses resistant to conventional antiherpetic drugs. This is why many efforts have been made these last years to set up different experimental strategies allowing the identification of new potential antiherpetic drugs. New technologies like X-ray crystallography or NMR, rational drug design and high-throughput screening have been used to identify novel inhibitors of viral replication. Ribonucleotide reductase (RR) has been the focus of drug discovery for at least a decade, however, the irreversible RR inhibitors identified, such as thiocarbonohydrazone, inhibit the viral as well as the human RR and thus only topical treatment was considered [6]. Herpes virus proteases were also considered as targets for antiviral chemotherapy. Several inhibitors like cysteine-specific, sulfonamide type [7] or natural products were discovered, but the therapeutic index was too low and no *in vivo* activity in animal models has been reported.

Recently, a really potent class of thiazolylphenyl compound directed against helicase-primase was published. BAY 57-1293 possesses pharmalogical profiles superior to ACV, pencyclovir, and famcyclovir, and is especially efficient when treatment is delayed or the viral load is increased, features essential for the treatment of herpes encephalitis and disseminated disease [8]. Like BAY 57-1293, the lead compound Bils 179 is as potent as acyclovir and more efficient when treatment is delayed [9]. The discovery of these kinds of inhibitors is encouraging for the development of a new class of drugs.

Nevertheless, conventional antiviral drugs are all designed to target viral proteins in order to avoid toxicity and to ensure selectivity. In general, antiviral drugs that target viral proteins are active against only a relatively narrow range of viruses and often need to be replaced with novel drugs because the viral genes that encode these protein targets mutate spontaneously, allowing the emergence of resistant viruses. This is why an alternative to this traditional drug design approach is to target cellular proteins. Indeed, antiviral drugs that target cellular proteins could be, in theory, active against a broad range of viruses, and they would be active against mutant viruses resistant to conventional drugs. Targeting cellular gene products might result in cytotoxicity and other undesirable side-effects, although from different areas of drug discovery like AIDS or cancer, we now know that many host functions can be blocked without any obvious effects on the host. Hopefully, in the future, alternative use of these latter drugs together with conventional drugs for the management of herpetic infections will reduce the emergence and dissemination of drug-resistant viruses.

Unfortunately today, the only evidence that cellular proteins can be relevant targets for antiviral drugs relates to the cyclin-dependent kinases (cdks). Cdks are required for the replication of HSV-1 as well as other viruses such as papillomaviruses and human cytomegalovirus. Several purine-derived cyclin-dependent ki-

nase inhibitors like roscovitine and olomucine have been found to inhibit replication of HSV-1. Roscovitine belongs to the class of specific cdks inhibitors, which inhibit cdk1, 2, 5, and 7. The inhibition of HSV-1 replication by roscovitine occurs by various mechanisms. It has been demonstrated that roscovitine inhibits the post-translational modification and particularly the phosphorylation of ICP0 as well as the transactivating activity of this protein [10]. Roscovitine also affects replication of HSV-1 DNA by a mechanism that remains to be elucidated [11]. Although the pathways used by roscovitine to inhibit HSV-1 infection are still unclear, anti-cdks are the first promising cellular protein targets in the design of antiviral drugs.

In our laboratory we have developed several different proteomic approaches to unravel cellular proteins or cellular metabolic pathways whose activities are absolutely required for the outcome of infection, in order to evaluate whether these proteins could be considered as potential non-conventional antiherpetic therapeutic targets. An overview of the main results we have obtained in recent years is presented in the next section, followed by a promising perspective in the search for original targets, which could be mainly located within the host cell nucleus.

15.2 Modifications of Host Cell Gene Expression: a Proteomic Approach

HSV-1 gene expression is sequentially and temporally regulated through complex regulatory mechanisms. Viral genes can be divided into immediate-early (α), early (β), and late (γ) genes according to their kinetics of expression. It is well established that during lytic HSV-1 infection, host cell protein synthesis is rapidly and severely repressed. This shut-off of protein synthesis takes place in two stages. The primary shut-off occurs very early after infection and does not require *de novo* protein synthesis. The secondary shut-off is concomitant with the synthesis of the early and late viral proteins. It concerns both host and viral proteins, and requires viral gene expression [12, 13].

However, despite the virally induced inhibition of host protein synthesis, a small number of cellular proteins continue to accumulate upon infection [14–18]. We speculated that this particular class of cellular proteins could play a major role in virus and cell interactions, and in determining the outcome of infection. As a first step to test this hypothesis, we undertook a functional proteomic approach to characterize cellular proteins whose synthesis is stimulated or sustained during the course of infection.

HSV infection also induces modifications of the host cell morphology and metabolism. Some of these modifications concern nucleoli – which are the site of ribosome synthesis – and more particularly ribosomes, which appear strongly modified after infection [17, 19–23]. We hypothesized that ribosomes themselves could be directly involved in the post-transcriptional regulation of cellular and viral gene expression. Therefore, we undertook the identification of ribosomal proteins that are modified and of non-ribosomal proteins that associate with ribosomes upon

infection, as a prerequisite for the study of their potential involvement in the control of gene expression.

15.2.1 Cellular Proteins Escape the Virally Induced Shut-off of Protein Synthesis

Because data about cellular proteins that escape the virally induced inhibition of synthesis were poorly documented and fragmented, we conducted experiments in order to evaluate the proportions of these host proteins and to characterize them [24].

For this, the rate of synthesis of a subset of cellular proteins was determined individually during the course of infection. HeLa cells were infected with HSV-1, and proteins were metabolically labeled with [^{35}S]methionine for one hour at different times of infection just prior harvesting. Total labeled proteins were first submitted to an acetic acid extraction in order to obtain a preparation of proteins free of nucleic acid. This method, originally developed for viral protein purification, was efficient in getting rid of nucleic acids and in purifying proteins with pI ranging from less than 4.0 up to 13.46 [25, 26]. Labeled proteins were then separated by two-dimensional polyacrylamide gel electrophoresis (2-D PAGE) using a linear immobilized pH gradient (IPG) ranging from pH 4.0 to 7 for the first dimension and the Laemmli-SDS system for the second dimension [24]. Radioactivity incorporated into individual proteins was then measured by scanning densitometry of the dried gels using a PhosphorImager. In these experimental conditions, the amount of radioactivity incorporated into each protein reflects their individual synthesis rates.

Two host proteins and one viral protein were used as controls of the validity of the experimental strategy. Their kinetics of synthesis was already known during the course of infection (Fig. 15.1): the synthesis of β-actin is shut off during HSV-1 infection, whereas the synthesis of the heat shock protein 70 (hsp70) is stimulated [18, 28, 29]. The control viral protein encoded by the UL42 early gene is a polymerase-associated processivity factor whose synthesis is induced during the early phase of the replicative viral cycle.

As expected, the amount of radioactivity incorporated into cellular proteins, including β-actin, decreased dramatically and progressively during infection. This reflects the well-documented virally induced inhibition of host protein synthesis. However, 28 proteins out of the 183 proteins analyzed behave in a very different way. Indeed, in cells infected for 3 or 6 hours, the amount of radioactivity incorporated into these 28 proteins was either unchanged or increased compared with uninfected cells. They include hsp70, whose synthesis is strongly stimulated from 6 hours post infection. Individually, these 28 proteins exhibited very different synthesis rates. Their synthesis rates were sustained or increased up to 3 hours of infection and remained relatively high up to late time of infection. For some of them, the increase was very drastic. Simultaneously, new labeled proteins appeared in gels corresponding to proteins extracted from infected cells, while they were absent in gels obtained from uninfected cells [24]. They correspond either to

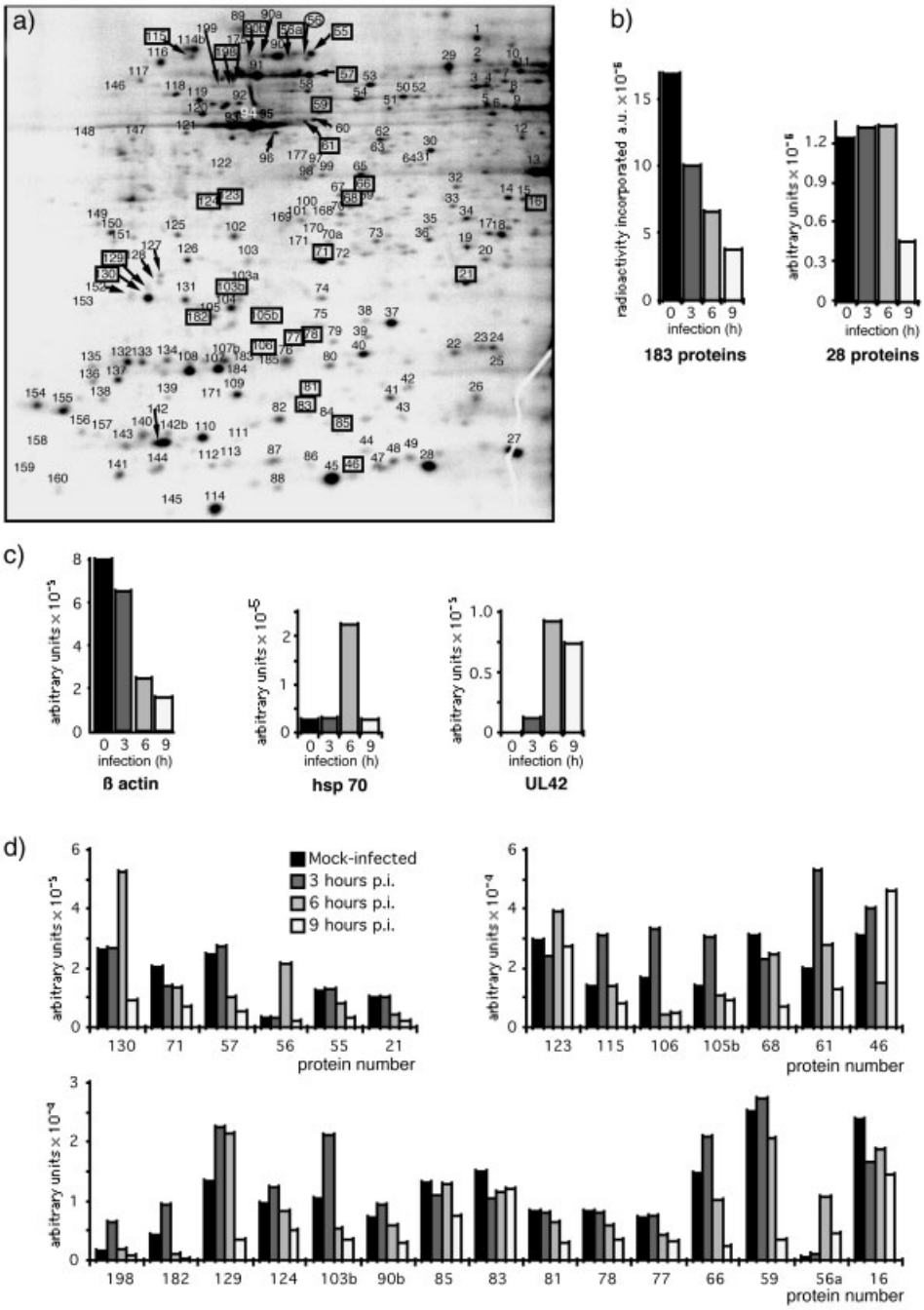


Figure 15.1 a–d

cellular proteins, which are modified after infection, or to viral proteins, which appeared with different kinetics. The kinetics of appearance of UL42 protein is typical of the early class of viral proteins, since its synthesis rate was maximum at 6 hours post infection.

These data revealed that an unexpectedly high proportion of host proteins continue to be efficiently synthesized upon HSV infection: these 28 cellular proteins represent about 15% of the analyzed subset of cellular proteins. Except for hsp70, these proteins remain to be identified. Therefore, following our hypothesis, these proteins probably represent a type of proteins that could be considered as a new class of antiherpetic therapeutic targets.

15.2.2 Ribosome Biogenesis Persists in HSV-1-infected Cells

The HSV-1-induced modifications of ribosomes include the phosphorylation of some ribosomal proteins, and the progressive association of several non-ribosomal proteins to ribosomes.

15.2.2.1 Modifications of Ribosomes upon HSV-1 Infection

We analyzed the HSV-1-induced alterations of ribosomal protein maps to characterize the modifications of ribosomes in infected cells. Ribosomal proteins purified from ribosomes from mock-infected and HSV-1-infected cells were extracted by the acetic acid procedure and alkylated with iodoacetamide [26]. They were then separated using a specific 2-D PAGE method called the "four corner method" dedicated to the separation of ribosomal proteins [27]. In this electrophoretic system, the position of each individual ribosomal protein is known and is extremely reproducible.

Figure 15.1 Synthesis rates of a set of 183 host proteins during the course of HSV-1 infection. This figure corresponds to Figure 2 of [24]. HeLa cells were infected with HSV-1. At different times after infection, cells were incubated with a mixture of [^{35}S]methionine and [^{35}S]cysteine for 1 hour just before harvesting. Acetic acid-soluble proteins were extracted and separated by 2-D PAGE. The first dimensional separation was performed using a linear pH 4–7 IPG. The second dimension was performed using the Laemmli-SDS system. Relative synthesis rates of each 183 host protein were estimated after scanning densitometry of the gels corresponding to mock-infected cells, and to cells infected for 3, 6, and 9 hours, using a PhosphorImager. (a) PhosphorImager image of the gel of mock-infected cells. The 183 proteins chosen randomly are indicated and numbered. Numbers of proteins whose synthesis is maintained during the course of infection are boxed. (b) Variation of the total radioactivity incorporated into the 183 and into the 28 proteins in mock-infected cells and in infected cells. (c) Variation of the radioactivity incorporated into two cellular and one viral control proteins, i.e. β-actin (numbers 93, 94, and 95 indicated in white in (a), hsp70 (spot 56) and viral UL42 early protein. (d) Variation of the radioactivity incorporated individually into the 28-boxed proteins. Names or numbers of each protein are indicated below the corresponding histograms.

Phosphorylation of Ribosomal Proteins The phosphorylation of ribosomal proteins was investigated by 2-D PAGE analysis of proteins from mock-infected and HSV-1-infected cells after *in vivo* labeling with ^{32}P for 90 min at times corresponding to the synthesis of α, β, or γ proteins.

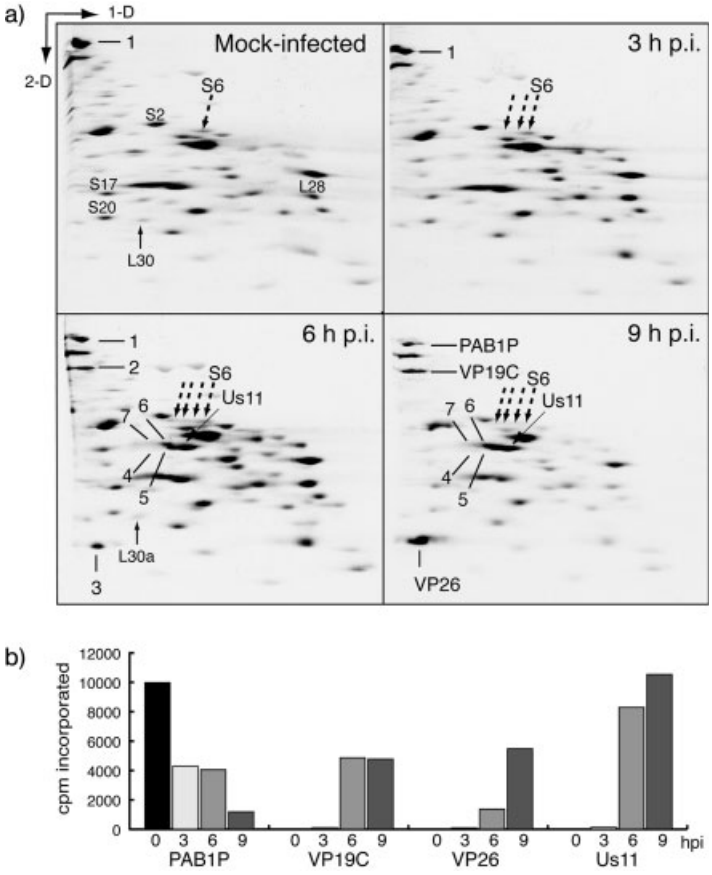


Figure 15.2 Kinetics of association of non-ribosomal proteins to the ribosomal fraction during the course of infection. This figure is adapted from Figure 2 of [23]. Proteins from HeLa cells were metabolically labeled as indicated in the legend of Figure 15.1. Proteins of the ribosomal fraction were extracted and separated by 2-D PAGE. (a) Autoradiograms of the dried gels. The non-ribosomal proteins are numbered from 1 to 7. Spot positions of some ribosomal proteins are indicated in the left panel. They are specified according to the proposed uniform nomenclature for mammalian ribosomal proteins [111]. Hashed lines on the right of S6 and protein L30a correspond to phosphorylated derivatives of ribosomal proteins S6 and L30 respectively. Non-ribosomal proteins are indicated by arrows and are numbered from 1 to 7. (b) Histograms representing the amount of radioactivity incorporated into the indicated proteins during the course of infection. Fragments of gels containing proteins of interest were cut out of the gels and the radioactivity was measured by scintillation counting of the solubilized proteins.

The overall experiments performed demonstrated that HSV-1 induces the phosphorylation of five ribosomal proteins. Four of them, S2, Sa, S3a, and S6, belong to the 40S subunit and one protein, L30, belongs to the 60S subunit. Ribosomal proteins, Sa, S2, S3a, and L30, are unusually phosphorylated, and phosphorylation of Sa, S3a, and L30 appears specific to HSV-1 infection [20, 21].

Further experiments were conducted to analyze the kinetics of phosphorylation of these ribosomal proteins during the course of infection. They revealed that S6 undergoes an irreversible hyperphosphorylation starting immediately after infection, even in the absence of serum, while *de novo* phosphorylation of S2, S3a, Sa, and L30 appear later, during the early stage of infection (see Fig. 15.2 for spot position) [17, 19–23].

At present, sequences of almost all mammalian ribosomal proteins are known [30]. However, their structures and functions within ribosomes remain poorly documented. This is particularly true for Sa, S2, and L30. However, one could speculate that L30 might be involved in the modulation of ribosome assembly after HSV-1 infection [31, 32].

Several ribosomal proteins are bifunctional. S3a has been shown recently to play a role in cell death and transformation [33]. It is not known whether the functions of S3a may be regulated by phosphorylation. However, because HSV-1 modifies the host cells' proliferation and death programs, one may postulate that the phosphorylation of S3a induced by HSV-1 participates in this mechanism.

The function of ribosomal protein S6 is quite well elucidated. S6 plays a crucial role in the control of the translation of mRNAs coding for many components of the translational apparatus, including mRNAs coding for ribosomal proteins [34]. Nevertheless, at present, the role of S6 during HSV-1 infection is not clear.

Association of Non-ribosomal Proteins with Infected Cell Ribosomes In parallel with the above proteome analysis of the subset of cellular proteins, metabolically ^{35}S-labeled proteins from uninfected cells and from cells infected for various periods of time were extracted from cellular fractions containing ribosomes. Proteins from the ribosomal fractions were separated by 2-D PAGE using the specific electrophoretic system II, which is one of the "four corner methods". It allows the separation of very basic proteins with pI values ranging from 9.0 to 13.5 [27]. The first separation was carried out in polyacrylamide tube gels containing 8 M urea at pH 8.6; the second separation was performed in polyacrylamide slab gels containing 6 M urea at pH 6.75 in the presence of sodium dodecyl sulfate (SDS). These experiments allowed us to identify non-ribosomal proteins that appear to be tightly associated with ribosomes.

Several non-ribosomal proteins were present in ribosomal fractions from mock-infected as well as from infected cells. This is the case for protein 1 in Figure 15.2. The other non-ribosomal proteins were absent from mock-infected cells. These latter might be either viral proteins or cellular proteins, which associate with ribosomes only upon infection (Fig. 15.2, proteins 2–7).

Us11 was the first identified non-ribosomal protein found to be associated with ribosomes of HSV-1-infected cells (SWISS-PROT entry P56958) [35]. It is synthe-

sized and phosphorylated during the late phase of the HSV-1 replication cycle, packaged in the tegument of the native virion, and delivered into cells after infection [35, 36]. Different experiments allowed us to identify proteins 5–7 as the phosphorylated derivatives of Us11 protein [20, 22, 23, 35]. US11 is a plurifunctional protein whose functions will be discussed below.

The N-terminal sequence of protein 2 matches that of HSV-1 protein VP19C (SWISS-PROT entry P32888), and peptide mass fingerprints and quadrupole time-of-flight (Q-TOF) analysis obtained for protein 3 identify this protein as VP26 from HSV-1 (SWISS-PROT entry P10219) [23]. VP19C and VP26 are two major components of the HSV-1 capsid.

Although VP19C and VP26 are components of the capsid and Us11 protein resides in the tegument, their presence in the ribosomal fractions of HSV-1-infected cells seems due rather to specific interactions of these proteins with ribosomes than to co-purification of viral particles with ribosomes during the cell fractionation procedure. Indeed, they were found to be associated with the ribosomes as early as 6 hours post infection [23], whereas capsids are first assembled in the nucleus of infected cells and packaged with the viral DNA before appearing in the cytoplasm at a very late time of infection. VP26 protein may be involved in linking the capsid to the surrounding tegument and envelope at a late stage of viral assembly (for review see [1]). Furthermore, VP19C, VP26, and Us11 proteins associate with ribosomes with different kinetics (Fig. 15.2) [23], and only a fraction of Us11 protein might bind specifically to the 60S ribosomal subunits [35, 36]. However, the role, if any, played by VP19C and VP26 on the translational apparatus remains to be elucidated.

Protein 4, which is present in small amounts, did not correspond to any known protein as revealed by its peptide mass fingerprinting [23].

Protein 1 was revealed to be one of the two poly(A)-binding proteins, PABP1 or PABP2 (SWISS-PROT entries P11940 and Q15097 respectively) [23]. Results from peptide mass analysis did not allow us to discriminate between the two proteins. Both proteins are very close in sequence but differ in length. PABP1 is longer than PABP2 and the apparent molecular mass of protein 1 estimated by electrophoresis is closer to that of PABP1 than to that of PABP2. This is in good agreement with the following published data. PABP1 plays a major role in mRNA turnover as well as in regulation of translation [37]. Its interaction with the eIF4G translational eukaryotic initiation allows the circularization of mRNAs and promotes translation initiation [38]. In contrast, PABP2 interacts slightly with the poly(A) tail of mRNAs and does not interact significantly with eIF4G [39]. The homologous Pab1p in yeast associates with polyribosomes [40]. In addition, we have shown that association of protein 1 with ribosomes is specific and is lost only after RNA digestion. Altogether, these results support an identity of protein 1 with PABP1. The newly synthesized PABP1 continues to associate with ribosomes despite the inhibition of host protein synthesis and the virally induced desegregations of most of the polyribosomes into 80S monoribosomes (Fig. 15.2 b) [23].

15.2.2.2 Ribosome Biogenesis Persists in HSV-1-infected Cells

HSV-1-induced modifications of ribosomes include the irreversible phosphorylation of ribosomal protein S6. It is known that increase in S6 phosphorylation is always correlated to the preferential recruitment and translation of ribosomal protein mRNAs [34]. Therefore, we analyzed the synthesis of ribosomal proteins in cells infected with HSV-1 under different experimental conditions [17, 18, 21, 24]. Proteins were metabolically labeled with [^{35}S]methionine for 1 hour just prior harvesting as indicated above. We estimated the rate of synthesis of proteins present in the ribosomal fraction of mock-infected cells and of cells infected for various periods of time. The rate of synthesis of all the basic ribosomal proteins separated in system II remained very efficient: the synthesis of ribosomal proteins was sustained until 9 hours post infection (Fig. 15.2a) and finally decreased for most of them at 15 hours post infection [23, 24]. Nevertheless, the synthesis of the four ribosomal proteins S11, L10, L21, and L28 persisted during the late time of infection, when the synthesis of the other ribosomal proteins was shut off.

15.2.2.3 HSV-1-induced Differential Translational Regulation of Gene Expression

We have shown that the synthesis of several host proteins, including ribosomal proteins, is sustained or stimulated upon infection. Concomitantly, the synthesis of β-actin and glyceraldehyde-3-phosphate dehydrogenase (GAPDH) is drastically inhibited, together with that of most host proteins. The molecular mechanisms that regulate differential expression of cellular genes remained to be elucidated. However, mRNAs coding for β-actin, GAPDH, and ribosomal proteins are submitted to a progressive non-specific degradation during infection [17, 18, 41]. Therefore, the synthesis of these cellular proteins should be under differential virally induced translational control.

In order to verify this hypothesis, we have analyzed the distribution of different cellular mRNAs among polyribosomes including GAPDH, β-actin, and various ribosomal protein mRNAs. GAPDH and β-actin mRNA molecules were associated with polyribosomes of large size in uninfected cells; they were shifted to polyribosomes of smaller size, and to free inactive particles in infected cells (Fig. 15.3). This indicates that the efficiency of these mRNAs translation is reduced by the decrease of the number of ribosomes translating the same mRNA molecule. Conversely, the ribosomal protein mRNAs are redistributed into larger polyribosomes in infected cells compared to uninfected cells. Therefore, translation of GAPDH and β-actin mRNAs became progressively inhibited at the initiation step, before the binding of the 60S ribosomal subunit. In contrast, translation of ribosomal protein mRNAs became more efficient after infection than before [17, 18, 41].

These results demonstrate that HSV-1 infection induces a selective translational control of host protein. Because ribosomes play a central role in cellular and viral gene expression, and because their biogenesis persists and they are severely modified during HSV-1 infection, we propose that the ribosomal proteins and the other ribosome-associated proteins whose metabolism is modified during infection could be evaluated as potential antiherpetic therapeutic targets.

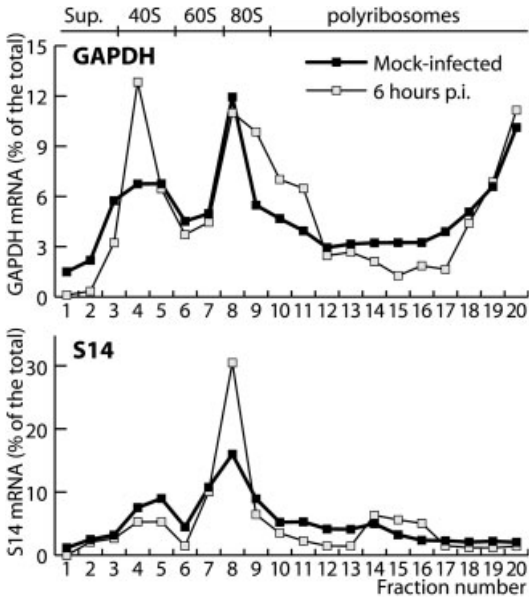


Figure 15.3 Distribution of GAPDH and S14 mRNAs among ribosomal fractions during the course of HSV-1 infection. Post-mitochondrial supernatants of control mock-infected HeLa cells and of HeLa cells infected with HSV-1 for 6 hours were layered onto 10–40% linear sucrose gradients. After centrifugation, 20 fractions of identical volume were collected. RNA was extracted from each fraction, and the distribution of GAPDH and S14 mRNAs among the different fractions was analyzed by northern blot with the relevant ^{32}P-labeled specific DNA probes. Fraction numbers are indicated under the profiles. Position of the supernatant (sup.), 40S, 60S, 80S, and polyribosomes are specified at the top of the figure. The amounts of radioactivity of the hybridized probes were quantified by scanning densitometry of the membranes using a PhosphorImager. Results are expressed in percent of the radioactivity in each fraction versus the sum of the radioactivity in the 20 fractions.

15.3
Nucleus and Lytic HSV-1 Infection

15.3.1
Functional Organization of the Cell Nucleus

Early microscopists first described the nucleus of eukaryotic cells about 200 years ago. This organelle is physically separated from the cytoplasm by a lipid bilayer that acts as a regulator of import and export of molecules through nuclear pores. The progresses made in the techniques of observation, especially in electron and in fluorescent microscopy, have led to the vision of the modern nucleus: a highly organized structure in which molecules are not distributed randomly [42]. For example, it has been well demonstrated that, in the interphase nucleus, chromo-

somes occupy defined spaces that are called chromosome territories [43]. In fact, more generally, within the nucleus, molecules – nucleic acids and proteins – accumulate into distinct regions to give rise to numerous different supramolecular structures called nuclear domains that are not surrounded by any membrane. These domains are the support of the major biological activities that take place within nucleus. The number and the presence or absence of most of the nuclear domains are dependent on the species, the cell type, and the physiological or pathological state of the cell.

At present, about 30 different nuclear domains have been characterized. Some of these domains have been studied more extensively than the others. This is the case of the nucleolus, the more prominent nuclear domain, a characteristic that permitted its description early in the nineteenth century. Nucleoli are the site of ribosome biogenesis, a very complex process that give rise to three distinct nucleolar compartments: the fibrillar center, the dense fibrillar component, and the granular component [44]. During the past decade, numerous studies have demonstrated that nucleoli not only carry out ribosome biogenesis but are also involved in other biological functions such as transfer RNA maturation, maturation and assembly of ribonucleoprotein complexes [45], cell cycle regulation [46], and cellular aging [47], leading to the notion of a plurifunctional nucleolus [48–50].

Biological functions fulfilled by other nuclear domains have been proposed, for example, for the splicing-factor compartments in which splicing factors accumulate [51] or the Cajal bodies in which small nuclear and nucleolar ribonucleoprotein particles [52] and transcription machineries [53] probably mature. However, the full composition and the precise biological functions fulfilled by the vast majority of the nuclear domains remain to be discovered. This is why numerous proteomic studies have been undertaken recently in order to gain information about various nuclear complexes. For example, two independent proteomic analyses of nucleoli purified from human cells have led to the identification of about 350 different proteins [49, 50]. Another proteomic analysis has been carried out for another nuclear domain, the interchromatin granule clusters and led to the identification of 36 proteins [54]. Some nuclear complexes have also been studied using proteomic approaches such as yeast and mammalian pore complexes [55, 56], yeast spindle pole [57], mammalian nuclear envelope [58], or human spliceosome [59, 60]. All these studies gave molecular data reinforcing the understanding of the regulation of metabolic activities that occur within the cell nucleus. Furthermore, the identification of unexpected proteins or of proteins with unknown functions within defined nuclear complexes provides some basis for their analysis at the molecular level.

Progress made in fluorescent time-lapse microscopy have allowed the observation that, accompanying its very high degree of organization, the nucleus is a very dynamic structure where molecules constantly associate and dissociate [61]. This has been demonstrated for transcription factors, splicing factors, DNA repair enzymes, chromatin-binding proteins, and nucleolar proteins [62].

Another interesting aspect in the mobility of molecules within the cell nucleus is the communication between nuclear domains. In fact, some factors are ex-

changed between domains. This is the case for Nopp140, a protein that shuttles between the nucleolus and the Cajal bodies [63]. Intranuclear trafficking pathways of newly synthesized molecules have also been described within nucleus. Fluorescent-tagged Sm proteins are first imported in the nucleus, accumulate in Cajal bodies, pass through nucleoli and finally reach the splicing-factor compartments [64]. Recently, another intranuclear pathway has been described for a nucleolar protein that accumulates in splicing-factor compartments before reaching nucleoli [65].

In various diseases, alterations and/or modifications of several nuclear domains have been described. For example, the size and the number of nucleoli is a relevant independent prognostic factor in numerous types of human tumors [66]. The number of Cajal bodies is also increased in transformed cells [67]. A well-described alteration of a nuclear domain in pathological conditions is the case of promyelocytic leukemia (PML) bodies. The biological functions of PML bodies have never been demonstrated clearly, but this domain is of particular interest since, at least in its native form, it is absent in cells from patients suffering from promyelocytic leukemia. The disappearance of PML bodies is due to a chromosomal translocation leading to the production of a chimeric protein between the PML proteins, the core of the PML bodies, and the retinoic acid receptor α [68, 69]. In this context, the fusion protein loses its ability to form PML bodies. Interestingly, administration of retinoic acid results in the reappearance of PML bodies and could lead to the remission of the cancer [70, 71].

15.3.2 HSV-1-induced Modifications of the Host Cell Nucleus

To enter a cell, HSV-1 binds to its surface on heparan sulfate receptors using different envelope glycoproteins. The viral envelope then fuses with the plasma membrane of the cell [72]. The viral capsid is released into the cytoplasm of the host cell and reaches the nucleus, the site of its replication. Migration of capsids is performed via the motor protein dynein and its cofactor dynactin, which allow capsids to travel along microtubules from the cell periphery to nuclear pore complexes [73]. After docking at the nuclear pore complex, capsids undergo conformational changes resulting in the expulsion of the viral genome into the nucleus [74]. Then, the productive viral cycle begins and leads to the production of virions. Viral transcription, DNA replication, assembly of new capsids, and packaging of viral DNA occur in the host nucleus and are accompanied by dramatic modifications of its architecture.

After its entry in the host nucleus, the viral DNA is deposited in the interchromosomal space, adjacent to PML bodies where viral transcription begins [75]. Viral transcription is accompanied by the appearance of regions of low contrast called viral replication compartments (VRCs) as visualized by electron microscopy analyses. Host chromatin is excluded from these compartments and becomes confined to the border of nuclei [76]. Time-lapse microscopy on infected cells has shown that VRCs are developed from parental HSV-1 genomes, preferentially

from PML body-associated genomes [77]. VRCs are the exclusive site of replication, transcription, and encapsidation of HSV-1 genomes [78] and expand during infection until they represent about the half of the nuclear volume. Quantitative three-dimensional image analysis performed with about 150 nuclei permitted a doubling of the nucleus volume during infection to be measured. This 2-fold increase is mainly due to VRC expansion without any decrease in chromatin volume [79]. Later in the infection, chromatin volume decreases to reach one-third of its initial volume.

A really surprising and yet not completely understood effect of HSV-1 infection is the elimination of PML bodies very soon after nuclear entry of the viral genome. In fact, viral genomes are deposited against PML bodies and, before replication begins, these nuclear domains are disrupted. This spectacular effect on PML bodies is due to the effect of the HSV-1 immediate-early regulatory protein ICP0 that induces proteasome-dependent degradation of PML body proteins [80]. The biological reasons for the elimination of PML bodies remain poorly understood. However, the fact that PML body formation is stimulated by interferon and that ICP0-deficient viruses are hypersensitive to interferon suggest that PML bodies could exert some antiviral activities that HSV-1 counters by disrupting these domains.

Other nuclear domains undergo alterations during HSV-1 infection. For example, it has been shown that the number of coiled bodies decreases during infection [81]. Splicing-factor compartments are redistributed from their diffuse speckled pattern to a highly punctuate organization [82]. Otherwise, nucleoli are modified during infection but they remain perfectly identifiable [83]. Their characteristic structure is loosened and finally consists of networks of granules joined by putative proteinaceous filaments. Transcription of the pre-rRNA continues during infection [78, 84]. However the finding that the 18S rRNA accumulates within nucleoli during infection raises the question of the outcome of the processing of rRNAs and assembly of ribosomes within modified nucleoli. Another exciting observation made by the group of Francine Puvion-Dutilleul is the accumulation of rRNA molecules within VRCs and interchromatin granule clusters of infected cells [84]. From these observations, it could be argued that ribosomal subunits could be there in contact with mRNAs for a coupled export into the cytoplasm or alternatively, for their translation within the nucleus. However studies remain to be performed to assess these hypotheses and gain insight into the possibility of a nuclear translation for viruses with an intranuclear replication cycle.

Other modifications of the host cell nuclear structure appear during infection, although the profound nature of these modifications is not yet elucidated. One of the characteristic morphological changes in the nucleus of HSV-1-infected cells is the appearance of dense bodies. The biological function fulfilled by these nuclear domains of viral origin is still unknown. A few of their proteins have been identified, however: the viral proteins UL11 [85], ICP22, UL3, UL4 [86], and two major nucleolar proteins, nucleolin [87] and fibrillarin [88]. Dense bodies seem to be devoid of nucleic acids [84]. Other nuclear domains of viral origin called assemblons appear. They are formed by aggregation of immature viral capsids and tegument

proteins [89]. Obviously the modifications of host cell nuclear domains are mainly due to the presence of viral proteins within these domains. This is why the next section will focus on the peculiar nuclear localization of various viral proteins during lytic infection.

15.3.3 Distribution of HSV-1 Proteins Within Different Nuclear Domains

15.3.3.1 ICP0

The HSV-1 immediate-early regulatory protein ICP0 is required for the initiation of lytic infection and for reactivation of the virus [90]. ICP0 has been described as a non-specific activator of viral and cellular gene expression. The first striking results on the behavior of the nuclear architecture with ICP0 were the discovery that the viral protein disrupts PML bodies very soon during lytic infection [91]. *In vitro* experiments have shown that ICP0 exhibits an ubiquitin E3 ligase activity that could explain a direct implication in the proteasome-dependent degradation of proteins contained within PML bodies [92]. However ICP0 does not localize only within PML bodies. Indeed, ICP0 is also found in centromeres, where it exerts the same degradation inducer effects as those in PML bodies. Two centromeric proteins, CENP-A and CENP-C, are degraded by the proteasome before the disruption of centromeres [80].

15.3.3.2 ICP27

ICP27 (also known as IE63) is an immediate-early protein. It is an essential protein involved in every step of mRNA metabolism [93]. It is well known that ICP27 stimulates gene expression at both the transcriptional and post-transcriptional level. Recent data indicate that it associates with cellular RNA polymerase II [94], although its role in post-transcriptional regulation of gene expression has been more extensively studied. ICP27 binds RNA via its RGG motif, regulates 3′ processing of viral pre-mRNAs, stabilizes the labile 3′ ends of mRNAs, and inhibits splicing of viral and cellular transcripts by causing the hyperphosphorylation of SR proteins that block spliceosome assembly [95]. This induces nuclear retention of intron-containing viral transcripts and ICP27 exports intronless mRNAs [93]. ICP27 seems to allow the coupling of transcription, processing, and export of RNAs, leading to comparisons with certain cellular factors like heterogeneous nuclear ribonucleoproteins or RNA helicase A [94]. These different activities of ICP27 are linked to its distribution within the infected cells. ICP27 shuttles from the nucleus to the cytoplasm [93]. Within the nucleus, ICP27 is found in splicing-factor compartments and is responsible for their reorganization [81]. During infection, ICP27 is also distributed in VRCs [96] and nucleoli [97] but its functions in these domains remain elusive.

15.3.3.3 US11

The US11 protein is a 17-kDa highly basic phosphoprotein. It is encoded by a non-essential gene of HSV-1 and is produced at late stages of infection. US11 is present in the tegument of the virion in 600–1000 copies that are released when HSV-1 infects a cell [98]. Within the host cell, US11 is localized in the cytoplasm, possibly associated with ribosomes. It is also abundant in the cell nucleus where it accumulates in nucleoli, particularly in the dense fibrillar and granular components, and, sometimes, US11 can be found within splicing-factor compartments [99]. At the functional level, different roles have been assigned to US11. It seems to play a major role in the blockade of activation of the cellular double-stranded RNA-dependent protein kinase R (PKR) during HSV-1 infection [100]. PKR is a serine/threonine kinase activated by autophosphorylation that notably possesses antiviral activities. Its activation notably leads to an inhibition of translation that the virus must prevent. PKR could be activated by dsRNA through an interaction mediating a change in its conformation leading to its activation. Alternatively, PKR could be activated by protein-protein interactions, such as its interaction with PKR activator (PACT) [101]. US11 prevents a possible activation of PKR by these two different ways [102, 103]. The inhibition of PKR activation by US11 and its accompanying maintenance of translation can be linked to the activity of US11 after heat shock. Indeed, it has been shown that US11 increases recovery of protein synthesis in cells that have undergone a heat shock [104]. Coupled with the fact that US11 is found associated with ribosomes, these elements suggest that US11 could be involved in the maintenance of protein synthesis when host cells undergo various kinds of stress. Thus, if US11 appears linked to the translation of mRNAs, some results have clearly shown that it is more generally involved in the post-transcriptional regulation of gene expression. In fact, unlike ICP27, US11 activates the export of unspliced mRNAs and their translation into the cytoplasm [105]. However, this has been demonstrated in a heterologous experimental system, showing that US11 was able to bind unspliced mRNAs from human T cell leukemia virus type I (HTLV-I) and human immunodeficiency virus type 1 (HIV-1) and to activate their export into the cytoplasm where they were translated. Interestingly, recently, it has been shown that US11 binds directly three different transcripts from HSV-1, those derived from UL12, UL13, and UL14 genes in a sequence-specific manner and that it down-regulates the expression of the UL13 protein [106].

15.3.3.4 Other Nuclear HSV-1 Proteins

The particular intracellular localization of other proteins encoded by the HSV-1 genome has been determined. This is the case for tegument proteins VP22 and VP13/14 that accumulate in nuclear dots rapidly after the beginning of infection. An intriguing but not understood finding is that these nuclear dots are juxtaposed to those formed by ICP0 at PML bodies [107]. The same observation has been made for the essential immediate-early protein ICP4 in living infected cells [108]. ICP4 that is required to activate the transcription of early and late genes of HSV-1

localizes adjacent to PML bodies. This could reflect its recruitment to the viral genome very soon after its deposition in the nucleus. Finally, the $\gamma_1$34.5 protein of HSV-1 localizes in both the nucleus and cytoplasm of host cells. Like US11, it inhibits the shut-off of protein synthesis mediated by PKR [109]. It is interesting to note that, like US11, the $\gamma_1$34.5 protein localizes in the nucleolus and shuttles between the nucleus and the cytoplasm [110]. However, as for almost all the proteins described above, the links between their localization within the nucleus and their biological functions are not fully elucidated.

All the observations reported above clearly indicate that HSV-1, which replicates entirely within nucleus, induces a dramatic reorganization of the host cell nucleus and in particular of the many pre-existing nuclear domains. Clearly much work remains to be done to understand the role of the modified domains in the control of the outcome of infection. However, although this work is not yet achieved, one can assume that many of the proteins contained within these particular nuclear domains play a crucial role in determining the efficiency of HSV-1 infection. As a consequence many of these proteins can be considered as potential antiherpetic therapeutic targets.

15.4 Conclusions

Profound modifications of the host cell occur during HSV-1 infection. In particular, the translation apparatus undergoes important modifications and the nuclear architecture is markedly reorganized, with the appearance of new nuclear domains, disappearance or alterations of others. These events must reflect changes in the functions of the translational machinery and in biochemical activities occurring within the various nuclear domains. However, at present, all these observations have been made protein by protein, particularly for the nucleus, and a comprehensive view of the HSV-1-induced modifications of the host cell is dramatically missing. There is therefore an urgent need to develop new proteomic approaches to pursue the analyses of the extensive modifications of the host cell proteome induced by HSV-1 and in particular the modifications of the host cell nucleus, since several of the clue events determining the outcome of infection take place within this highly complex, organized and dynamic structure. Without doubt, these analyses will provide essential elements for the discovery of new original therapeutic targets for the development of antiherpetic molecules.

15.5 References

1 B. Roizman, R. J. Whitley, C. Lopez. *The Human Herpes Viruses*. Raven Press, New York, **1993**.

2 M. Reyes, N. Shaik, J. Graber, R. Nisenbaum, N. Wetherall, K. Fukuda, W. Reeves. *Arch. Intern. Med.* **2003**, *13*, 76–80.

3 H. J. Schaeffer, L. Beauchamp, P. de Miranda, G. B. Elion, D. J. Bauer, P. Collins. *Nature* **1978**, *272*, 583–585.

4 J. Christophers, J. Clayton, J. Craske et al. *Antimicrob. Agents Chemother.* **1998**, *42*, 868–872.

5 F. Morfin, D. Thouvenot. *J. Clin. Virol.* **2003**, *26*, 29–37.

6 C. G. Bridges, S. P. Ahmed, P. S. Sunkara, J. R. McCarthy, A. S. Tyms. *Antiviral Res.* **1995**, *27*, 325–334.

7 C. T. Supuran, A. Casini, A. Scozzafava. *Med. Res. Rev.* **2003**, *23*, 535–558.

8 G. Kleymann, R. Fischer, U. A. Betz et al. *Nature Med.* **2002**, *8*, 392–398.

9 J. J. Crute, C. A. Grygon, K. D. Hargrave et al. *Nature Med.* **2002**, *8*, 386–391.

10 D. J. Davido, D. A. Leib, P. A. Schaffer. *J. Virol.* **2002**, *76*, 1077–1088.

11 L. M. Schang, A. Bantly, M. Knockaert et al. *J. Virol.* **2002**, *76*, 7874–7882.

12 Y. Nishioka, S. Silverstein. *Proc. Natl. Acad. Sci. USA* **1977**, *74*, 2370–2374.

13 M. L. Fenwick, J. Clark. *J. Gen. Virol.* **1982**, *61*, 121–125.

14 L. M. Kemp, D. S. Latchman. *Virology* **1988**, *166*, 258–261.

15 E. L. Notarianni, C. M. Preston, *Virology* **1982**, *123*, 113–122.

16 R. D. Everett. *EMBO J.* **1985**, *4*, 1973–1980.

17 A. Greco, A. Laurent, J.-J. Madjar. *Mol. Gen. Genet.* **1997**, *256*, 320–327.

18 D. Simonin, J.-J. Diaz, T. Massé, J.-J. Madjar. *J. Gen. Virol.* **1997**, *78*, 435–443.

19 I. M. Kennedy, W. S. Stevely, D. P. Leader. *J. Virol.* **1981**, *39*, 359–366.

20 T. Massé, D. Garcin, B. Jacquemont, J.-J. Madjar. *Mol. Gen. Genet.* **1990**, *220*, 1–12.

21 D. Simonin, J.-J. Diaz, K. Kindbeiter, L. Denoroy, J.-J. Madjar. *Electrophoresis* **1995**, *16*, 854–859.

22 D. Simonin, J.-J. Diaz, K. Kindbeiter, P. Pernas, J.-J. Madjar. *Electrophoresis* **1995**, *16*, 1317–1322.

23 A. Greco, W. Bienvenut, J.-C. Sanchez et al. *Proteomics* **2001**, *1*, 545–549.

24 A. Greco, N. Bausch, Y. Couté, J.-J. Diaz. *Electrophoresis* **2000**, *21*, 2522–2530.

25 H. Fraenkel-Conrat. *Virology* **1957**, *4*, 1–4.

26 A. Greco, J.-J. Madjar. *Cell Biology: A Laboratory Handbook*, Vol. 2. J. E. Celis, London, **1998**, pp 135–139.

27 J.-J. Madjar, S. Michel, A. Cozzone, J. Reboud. *Anal. Biochem.* **1979**, *92*, 174–182.

28 B. A. Mayman, Y. Nishioka. *J. Virol.* **1985**, *53*, 1–6.

29 B. Phillips, K. Abravaya, R. I. Morimoto. *J. Virol.* **1991**, *65*, 5680–5692.

30 I. G. Wool, Y.-L. Chan, A. Glück. In *Translational Control* (J. W. B. Hershey, M. B. Mathews, N. Sonenberg, Eds). Cold Spring Harbor Press, New York, **1996**, pp 685–732.

31 H. Mao, S. A. White, J. R. Williamson. *Nature Struct. Biol.* **1999**, *6*, 1139–1147.

32 J. Vilardell, P. Chartrand, R. H. Singer, J. R. Warner. *RNA* **2000**, *6*, 1773–1780.

33 H. Naora. *Immunol. Cell Biol.* **1999**, *77*, 197–205.

34 O. Meyuhas. *Eur. J. Biochem.* **2000**, *267*, 6321–6330.

35 J.-J. Diaz, D. Simonin, T. Massé et al. *J. Gen. Virol.* **1993**, *74*, 397–406.

36 R. J. Roller, B. Roizman. *J. Virol.* **1992**, *66*, 3624–3632.

37 C. Grosset, C. Y. Chen, N. Xu, N. Sonenberg, H. Jacquemin-Sablon, A. B. Shyu. *Cell* **2000**, *103*, 29–40.

38 S. E. Wells, P. E. Hillner, R. D. Vale, A. B. Sachs. *Mol. Cell* **1998**, *2*, 135–140.

39 H. Imataka, A. Gradi, N. Sonenberg. *EMBO J.* **1998**, *17*, 7480–7489.

40 D. Mangus, N. Ambrani, A. Jacobson. *Mol. Cell. Biol.* **1998**, *18*, 7383–7396.

41 A. Laurent, J.-J. Madjar, A. Greco. *J. Gen. Virol.* **1998**, *79*, 2765–2775.
42 M. Dundr, T. Misteli. *Biochem. J.* **2001**, *356*, 297–310.
43 T. Cremer, C. Cremer. *Nature Rev. Genet.* **2001**, *2*, 292–301.
44 T. Melese, Z. Xue. *Curr. Opin. Cell Biol.* **1995**, *7*, 319–324.
45 T. Pederson, J.C. Politz. *J. Cell Biol.* **2000**, *148*, 1091–1095.
46 R. Visintin, A. Amon. *Curr. Opin. Cell Biol.* **2000**, *12*, 752.
47 F.B. Johnson, R.A. Marciniak, L. Guarente. *Curr. Opin. Cell Biol.* **1998**, *10*, 332–338.
48 T. Pederson. *Nucleic Acids Res.* **1998**, *26*, 3871–3876.
49 J.S. Andersen, C.E. Lyon, A.H. Fox et al. *Curr. Biol.* **2002**, *12*, 1–11.
50 A. Scherl, Y. Coute, C. Deon et al. *Mol. Biol. Cell.* **2002**, *13*, 4100–4109.
51 T. Misteli. *J. Cell Sci.* **2000**, *113*, 1841–1849.
52 S.C. Ogg, A.I. Lamond. *J. Cell Biol.* **2002**, *159*, 17–21.
53 J.G. Gall. *FEBS Lett.* **2001**, *498*, 164–167.
54 P.J. Mintz, S.D. Patterson, A.F. Neuwald, C.S. Spahr, D.L. Spector. *EMBO J.* **1999**, *18*, 4308–4320.
55 M.P. Rout, J.D. Aitchison, A. Suprapto, K. Hjertaas, Y. Zhao, B.T. Chait. *J. Cell Biol.* **2000**, *148*, 635–651.
56 J.M. Cronshaw, A.N. Krutchinsky, W. Zhang, B.T. Chait, M.J. Matunis. *J. Cell Biol.* **2002**, *158*, 915–927.
57 P.A. Wigge, O.N. Jensen, S. Holmes, S. Soues, M. Mann, J.V. Kilmartin. *J. Cell Biol.* **1998**, *141*, 967–977.
58 M. Dreger, L. Bengtsson, T. Schoneberg, H. Otto, F. Hucho. *Proc. Natl Acad. Sci. USA* **2001**, *98*, 11943–11948.
59 J. Rappsilber, U. Ryder, A.I. Lamond, M. Mann. *Genome Res.* **2002**, *12*, 1231–1245.
60 Z. Zhou, L. J. Licklider, S.P. Gygi, R. Reed. *Nature* **2002**, *419*, 182–185.
61 T. Misteli. *Science* **2001**, *291*, 843–847.
62 J. Roix, T. Misteli. *Histochem. Cell Biol.* **2002**, *118*, 105–116.
63 C. Isaac, Y. Yang, U.T. Meier. *J. Cell Biol.* **1998**, *142*, 319–329.
64 J.E. Sleeman, A.I. Lamond. *Curr. Biol.* **1999**, *9*, 1065–1074.
65 A.K. Leung, A.I. Lamond. *J. Cell Biol.* **2002**, *157*, 615–629.
66 A. Pich, L. Chiusa, E. Margaria. *Micron* **2000**, *31*, 133–141.
67 D.L. Spector, G. Lark, S. Huang. *Mol. Biol. Cell* **1992**, *3*, 555–569.
68 A. Kakizuka, W.H. Miller, Jr., K. Umesono et al. *Cell* **1991**, *66*, 663–674.
69 H. de The, C. Lavau, A. Marchio, C. Chomienne, L. Degos, A. Dejean. *Cell* **1991**, *66*, 675–684.
70 M.H. Koken, F. Puvion-Dutilleul, M.C. Guillemin *et al. EMBO J.* **1994**, *13*, 1073–1083.
71 K. Weis, S. Rambaud, C. Lavau et al. *Cell* **1994**, *76*, 345–356.
72 P.G. Spear. *Semin. Virol.* **1993**, *4*, 167–180.
73 K. Dohner, A. Wolfstein, U. Prank, C. Echeverri, D. Dujardin, R. Vallee, B. Sodeik. *Mol. Biol. Cell* **2002**, *13*, 2795–2809.
74 B.R. Cullen. *Cell* **2001**, *105*, 697–700.
75 A.M. Ishov, G.G. Maul. *J. Cell Biol.* **1996**, *134*, 815–826.
76 F. Puvion-Dutilleul, S. Besse. *Chromosoma* **1994**, *103*, 104–110.
77 G. Sourvinos, R.D. Everett. *EMBO J.* **2002**, *21*, 4989–4997.
78 F. Puvion-Dutilleul, E. Puvion. *J. Electron Microsc. Tech.* **1991**, *18*, 336–353.
79 K. Monier, J.C. Armas, S. Etteldorf, P. Ghazal, K.F. Sullivan. *Nature Cell Biol.* **2000**, *2*, 661–665.
80 R.D. Everett. *Oncogene* **2001**, *20*, 7266–7273.
81 A. Phelan, M. Carmo-Fonseca, J. McLaughlan, A.I. Lamond, J.B. Clements. *Proc. Natl Acad. Sci. USA* **1993**, *90*, 9056–9060.
82 T.E. Martin, S.C. Barghusen, G.P. Leser, P.G. Spear. *J. Cell Biol.* **1987**, *105*, 2069–2082.
83 F. Puvion-Dutilleul, J. Pedron, M. Laithier, P. Sheldrick. *Biol. Cell* **1982**, *44*, 249–260.
84 S. Besse, F. Puvion-Dutilleul. *J. Cell Sci.* **1996**, *109* (Pt 1), 119–129.
85 J.D. Baines, R.J. Jacob, L. Simmerman, B. Roizman. *J. Virol.* **1995**, *69*, 825–833.

86 N. S. Markovitz, B. Roizman. *J. Virol.* **2000**, *74*, 523–528.

87 F. Puvion-Dutilleul, E. Pichard, P. Sheldrick, F. Amalric, E. Puvion. *Eur. J. Cell Biol.* **1986**, *39*, 458–468.

88 C. Lopez-Iglesias, F. Puvion-Dutilleul, J. Cebrian, M. E. Christensen. *Eur. J. Cell Biol.* **1988**, *46*, 259–269.

89 P. L. Ward, W. O. Ogle, B. Roizman. *J. Virol.* **1996**, *70*, 4623–4631.

90 R. D. Everett. *Bioessays* **2000**, *22*, 761–770.

91 G. G. Maul, H. H. Guldner, J. G. Spivack. *J. Gen. Virol.* **1993**, *74* (Pt 12), 2679–2690.

92 C. van Sant, R. Hagglund, P. Lopez, B. Roizman. *Proc. Natl Acad. Sci. USA* **2001**, *98*, 8815–8820.

93 A. Phelan, J. B. Clements. *Semin. Virol.* **1998**, *8*, 309–318.

94 C. Zhou, D. M. Knipe. *J. Virol.* **2002**, *76*, 5893–5904.

95 K. S. Sciabica, Q. J. Dai, R. M. Sandri-Goldin. *EMBO J.* **2003**, *22*, 1608–1619.

96 A. de Bruyn Kops, S. L. Uprichard, M. Chen, D. M. Knipe. *Virology* **1998**, *252*, 162–178.

97 W. E. Mears, V. Lam, S. A. Rice. *J. Virol.* **1995**, *69*, 935–947.

98 R. J. Roller, B. Roizman. *J. Virol.* **1992**, *66*, 3624–3632.

99 S. Besse, J. J. Diaz, E. Pichard, K. Kindbeiter, J. J. Madjar, F. Puvion-Dutilleul. *Chromosoma* **1996**, *104*, 434–444.

100 K. A. Cassady, M. Gross. *J. Virol.* **2002**, *76*, 2029–2035.

101 R. C. Patel, G. C. Sen. *EMBO J.* **1998**, *17*, 4379–4390.

102 D. Khoo, C. Perez, I. Mohr. *J. Virol.* **2002**, *76*, 11971–11981.

103 G. A. Peters, D. Khoo, I. Mohr, G. C. Sen. *J. Virol.* **2002**, *76*, 11054–11064.

104 C. Diaz-Latoud, J. J. Diaz, N. Fabre-Jonca, K. Kindbeiter, J. J. Madjar, A. P. Arrigo. *Cell Stress Chaperones* **1997**, *2*, 119–131.

105 J. J. Diaz, M. D. Dodon, N. Schaerer-Uthurralt, D. Simonin, K. Kind beiter, L. Gazzolo, J. J. Madjar. *Nature* **1996**, *379*, 273–277.

106 H. L. Attrill, S. A. Cumming, J. B. Clements, S. V. Graham. *J. Virol.* **2002**, *76*, 8090–8100.

107 I. Hutchinson, A. Whiteley, H. Browne, G. Elliott. *J. Virol.* **2002**, *76*, 10365–10373.

108 R. D. Everett, G. Sourvinos, A. Orr. *J. Virol.* **2003**, *77*, 3680–3689.

109 B. He, M. Gross, B. Roizman. *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 843–848.

110 G. Cheng, M. E. Brett, B. He. *J. Virol.* **2002**, *76*, 9434–9445.

111 E. H. McConkey, H. Bielka, J. Gordon et al. *Mol. Gen. Genet.* **1979**, *169*, 1–6.

16
Francisella tularensis

Jirí Stulík, Martin Hubálek, Lenka Hernychová, Jana Havlasová, Juraj Lenco, Ales Macela, Igor Golovliov, and Anders Sjöstedt

16.1
Intracellular Pathogen *Francisella tularensis*

The facultative intracellular pathogen *Francisella tularensis* is the causative agent of the zoonotic disease tularemia. This pathogen was discovered as the agent of a plague-like disease among squirrels in Tulare County, California, in 1911 [1]. *Francisella tularensis* is highly virulent for human beings and the most important vectors are hares, rabbits, and rodents [2]. The inoculation or inhalation of 10–25 bacteria suffices to cause a severe disease [3]. Infection with *F. tularensis* targets predominantly the liver and spleen and is accompanied by negligible inflammatory reaction [4]. Two clinical forms of the disease predominate: ulceroglandular or glandular tularemia is caused by a contact with infected animals, while pneumonic tularemia is induced by the inhalation of dust contaminated by bacteria [5]. *In vitro* infection of monocyte/macrophage cell lines leads to host cell death via apoptosis but the mechanism remains unclear [6]. The high infectivity of bacteria together with its ease of aerosol dissemination, and the ability to cause a fatal disease are the main reasons why this bacterium is on the list of potential biological warfare category A agents [7]. Furthermore, *F. tularensis* belongs to the most common air-borne laboratory-acquired infections, therefore, the laboratory work has to be carried out by trained personnel, and in the case of fully virulent strains in safety cabinets under bio-safety level 3 conditions [8, 9].

16.1.1
Subtypes of *F. tularensis*

Two major biotypes of *F. tularensis* can be discriminated according to their virulence for humans. Type A or *F. tularensis* subspecies *tularensis* (*F.t. tularensis*), which occurs almost exclusively in North America, represents the most virulent variant. The outcome of the disease caused by type A infection is often fatal if treatment with appropriate antibiotics is not provided. Type B or *F. tularensis* subspecies *holarctica* (*F.t. holarctica*), found in Europe, Central Asia and Scandinavia, causes disease that is clinically indistinguishable from disease induced by subtype A but it is less severe and with only rare mortality [10, 11]. In addition, there are

Biomedical Application of Proteomics
Edited by J.-C. Sanchez, G.L. Corthals, D.F. Hochstrasser

ISBN: 3-527-30807-5

two other less important subspecies, *F. tularensis* subspecies *novicida* that has been isolated only in North America and is closely related to *F. t. tularensis* and, further, *F. tularensis* subspecies *mediaasiatica*, which occurs only in Central Asian parts of the former Soviet Union [12].

In 1950, Eigelsbach and Downs produced the *F. tularensis* live vaccine strain, designated *F. tularensis* LVS, after re-isolation of individual colonies originally derived from a Russian strain of subspecies *holarctica* [13]. This vaccine strain can impart protective immunity against systemically initiated infection by virulent strains of the pathogen and, therefore, it was used for vaccination in USSR and Eastern Europe, and for protection of laboratory workers at USAMRIID, Fort Detrick, Frederick, MD, USA [8]. Nevertheless, LVS strain appears to be less protective against aerosol challenge with virulent type A strains of the pathogen. In mice, this strain is as virulent as wild *holarctica* isolate, hence, it has been broadly used as the model for the study of pathogenesis of tularemia *in vivo* and *in vitro* [14].

16.1.2 The Major Objectives of Proteome Studies of *F. tularensis*

As mentioned above, there is serious threat that *F. tularensis* can be weaponized. The accessibility of rapid typing methods for sensitive detection of type A strains will be required in this case. However, the conserved nature of *F. tularensis* species and, additionally, the serious risk of laboratory-acquired infection hinder the rapid progress in this area. It is anticipated that the identification of type A and B unique protein markers will soon lead to a breakthrough in the development of new efficient typing tools.

The other unsolved problem concerns the molecular mechanisms of *F. tularensis* infection. Despite long-term intensive research there is little information available on the virulence factors of *F. tularensis*. Unlike other intracellular pathogens, *F. tularensis* has non-toxic LPS and does not produce any known toxin. To date, five genetic loci essential for intracellular growth of *Francisella tularensis* in host macrophage cells have been detected using transposon methodology [15]. Of them, three interrupted loci showed similarity at the deduced amino acid level to alanine racemase, the ClpB heat shock protease, and glutamine phosphoribosylpyrophosphate amidotransferase. The other inactivated loci corresponded to the *F. tularensis* gene encoding the 23-kDa protein that is prominently expressed after engulfment of bacteria by phagocyte [16]. As with the virulence factors of *F. tularensis*, there also is paucity of information about *F. tularensis* immunodominant antigens. It was reported that T lymphocytes collected from both naturally infected individuals and vaccinees recognize a multitude of membrane antigens. However, so far only a 17-kDa lipoprotein has been cloned and tested as a potential vaccine candidate [17, 18]. The lack of knowledge of immunodominant antigens hampers the efforts aimed at the development of new types of subcellular vaccines against tularemia.

16.2
Construction of Two-dimensional Electrophoresis (2-DE) Reference Protein Maps of Non-virulent and Highly Virulent *F. tularensis* Strains

The construction of reference protein maps of both non-virulent and fully virulent *F. tularensis* strains is the basic precondition for any subsequent comparative protein profile studies. The existence of annotated DNA sequence databases is fundamental to such studies. Regarding the sequencing of *F. tularensis* genomes, the project set up to annotate the genome sequence of SCHU S4 strain (subspecies *tularensis*) is nearly finished, while the analysis of the LVS genome (subspecies *holarctica*) is ongoing (September 2003). The assembled nucleotide sequence of the virulent SCHU4 strain provides 1.83 Mb of the genome sequence and a total of 1289 potential coding open reading frames (ORFs) have been identified in the data set [19]. The protein database formed by translation of all possible SCHU4 strain ORFs has been used in our laboratory for protein identification.

The total cellular proteins extracted from either *F. tularensis* LVS (ATCC 29684, American Type Culture Collection, Rockville, MD, USA) or from SCHU4 strain, isolated from human ulcer (Ohio, USA, 1994), have been used for construction of 2-DE reference maps of non-virulent and fully virulent *F. tularensis* strains. A suspension of microbes was washed, centrifuged, and the pellets were then immediately homogenized in lysis buffer containing 137 mM NaCl, 10% glycerol, 1% *p*-octyl-β-D-glucopyranoside, 50 mM NaF, 1 mM Na_3VO_4, and protease inhibitors. Extracted proteins were precipitated overnight [20] and then solubilized in buffer for isoelectric focusing-IEF buffer (9 M urea, 4% CHAPS, 70 mM DTT, and 2% carrier ampholytes). The equivalent of 150 µg of protein was loaded either by in-gel rehydration onto broad range non-linear immobilized pH 3–10 gradient (IPG) strips or by cup-loading at the anode in the case of basic range pH 6–11 IPG strips. In the second dimension, separation on gradient 9–16% sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) was performed. Using broad-range pH 3–10 IPG strips, more than 1000 spots were detected (Fig. 16.1 a, b).

The application of an overlapping gradient pH 6–11 (Fig. 16.2) allowed the separation of about 500 spots involving either spots already detected on a wide-range gradient but much better resolved, or proteins missing on the 3–10 non-linear pH gradient. The accession numbers indicate the identified proteins in 2-DE reference maps. Since there are only 58 entries for *F. tularensis* in Swiss-Prot and TrEMBL, the identification of individual proteins was mostly based on the finding of a particular ORF followed by the BLAST search for known proteins with similar amino acid sequences in other organisms listed in the NCBI nr database. Classical peptide mass fingerprinting (PMF) performed by matrix-assisted laser desorption ionization/time-of-flight mass spectrometry (MALDI-TOF-MS) was mostly sufficient for reliable identification of translated tularemic ORFs. Figure 16.3 shows a high-quality mass spectrum that led to unambiguous protein identification.

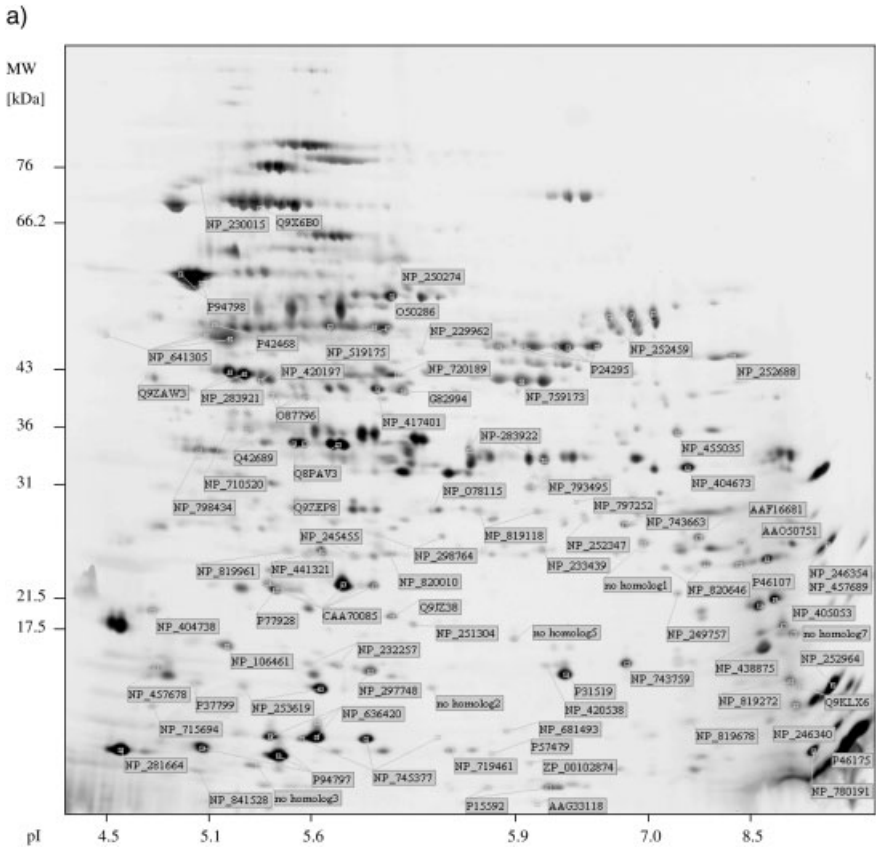


Figure 16.1 Silver-stained 2-DE reference maps of whole-cell lysates of (a) *Francisella tularensis* LVS and (b) *Francisella tularensis* subsp. *tularensis* SCHU4. Cell lysate extract (150 µg) was loaded onto IPG strip pH 3–10 followed by SDS-PAGE (9–16% gradient). Accession numbers indicate the identified spots. The names of tularemic proteins or their protein homologues as determined by NCBI nr BLAST search are in Table 16.1.

In cases where the results from PMF alone did not allow conclusive ORF identification, owing to the presence of few tryptic peptides, the partial amino acid sequence obtained from post-source decay (PSD) analysis of chemically modified peptides has been used for ORF database searching. This approach, called chemically assisted fragmentation MALDI (CAF-MALDI), is based on the introduction of a negatively charged group to the N-terminus of peptides generated by trypsin digestion. The negative charge at the N-terminus prevents the detection of the b-fragments resulting in a spectrum in which only y-ions become visible [21]. Figure 16.4 a–c depicts successful fragmentation of three modified peptides resulting in complete y-ion series of two of them and nearly complete y-ion series of a third, larger peptide.

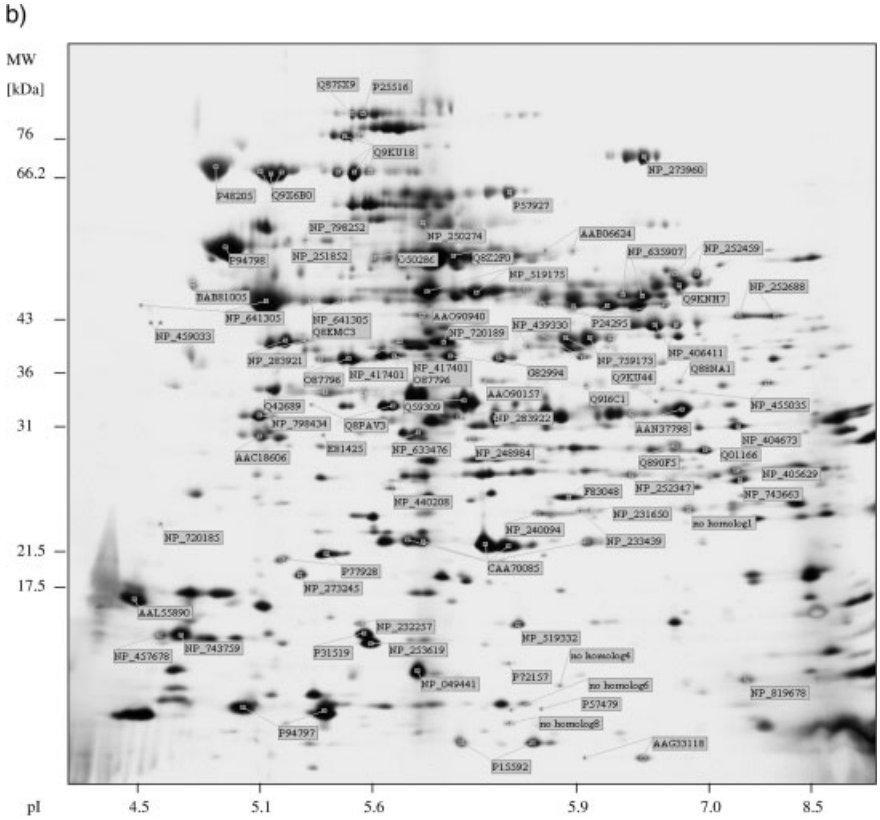


Figure 16.1 b

Globally, 220 different proteins have been already identified by the MALDI-TOF approach, that is about 17% of predicted ORFs. Of them, 143 proteins are indicated in the representative 2-DE maps of *F. tularensis* LVS and *F. tularensis* subsp. *tularensis* SCHU; their basic characteristics together with the sources of their homologues are listed in Table 16.1. Among all identified *F. tularensis* proteins, the 23 spots were classified as hypothetical proteins, and 11 additional ones were classified as unknowns. According to genome sequence data, about 413 ORFs should encode proteins with no homology to known proteins [19]. The 2-DE reference maps, with description of identified proteins, are publicly available on the website http://www.mpiib-berlin.mpg.de/2D-PAGE/.

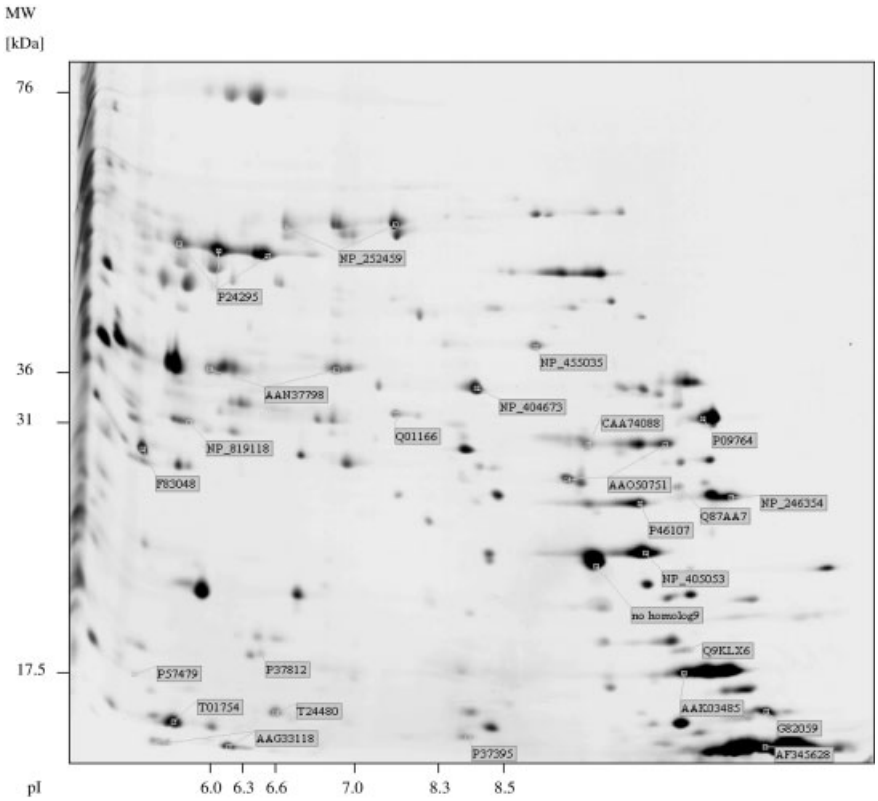


Figure 16.2 Silver-stained 2-DE reference map of basic proteins of *Francisella tularensis* LVS. Cell lysate extract (150 µg) was loaded onto IPG strip pH 6–11 followed by SDS-PAGE (9–16% gradient). Accession numbers indicate the identified spots. The names of tularemic proteins or their protein homologues as determined by NCBI nr BLAST search are in Table 16.1.

16.3 Comparative Proteome Analysis of *F. tularensis* Subspecies

We conducted a comparative study with four different isolates of *F. t. holarctica* (Russia – Norwegian rat, Sweden – human ulcer, USA – beaver, and reference strain obtained from the Veterinary Institute in Brno, Czech Republic) and five different isolates of *F. t. tularensis* (USA – twice human ulcer, human pleural fluid, Slovakia – mite, and reference strain obtained from Veterinary Institute in Brno, Czech Republic). First, we compared the protein profiles of whole-cell lysates resolved in the first dimension on broad pH 3–10 IPG strips. The computer analysis was performed using Melanie 3 software package and separated proteins were quantitated in terms of their relative volumes. We wanted to know whether

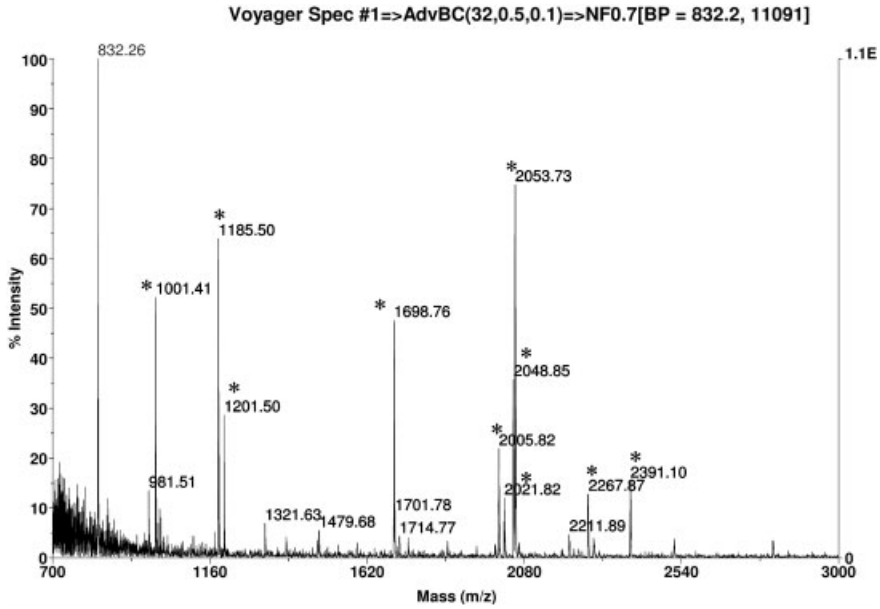


Figure 16.3 MALDI-TOF-MS peptide mass fingerprint of a tryptic digest of unknown protein. The list of masses unambiguously identified 30-kDa protein in the protein database with 10 indicated peptides matching the calculated tryptic peptide masses.

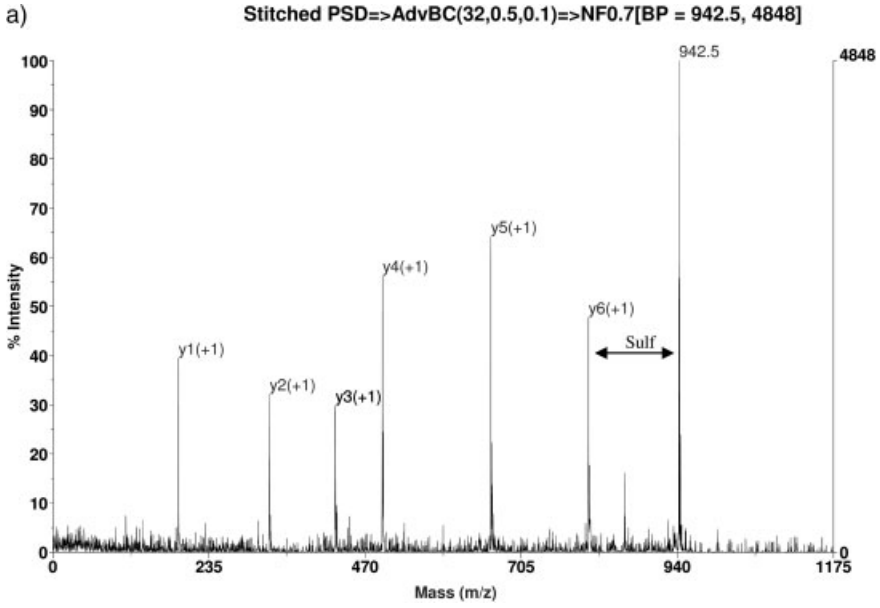


Figure 16.4 a Legend see page 292

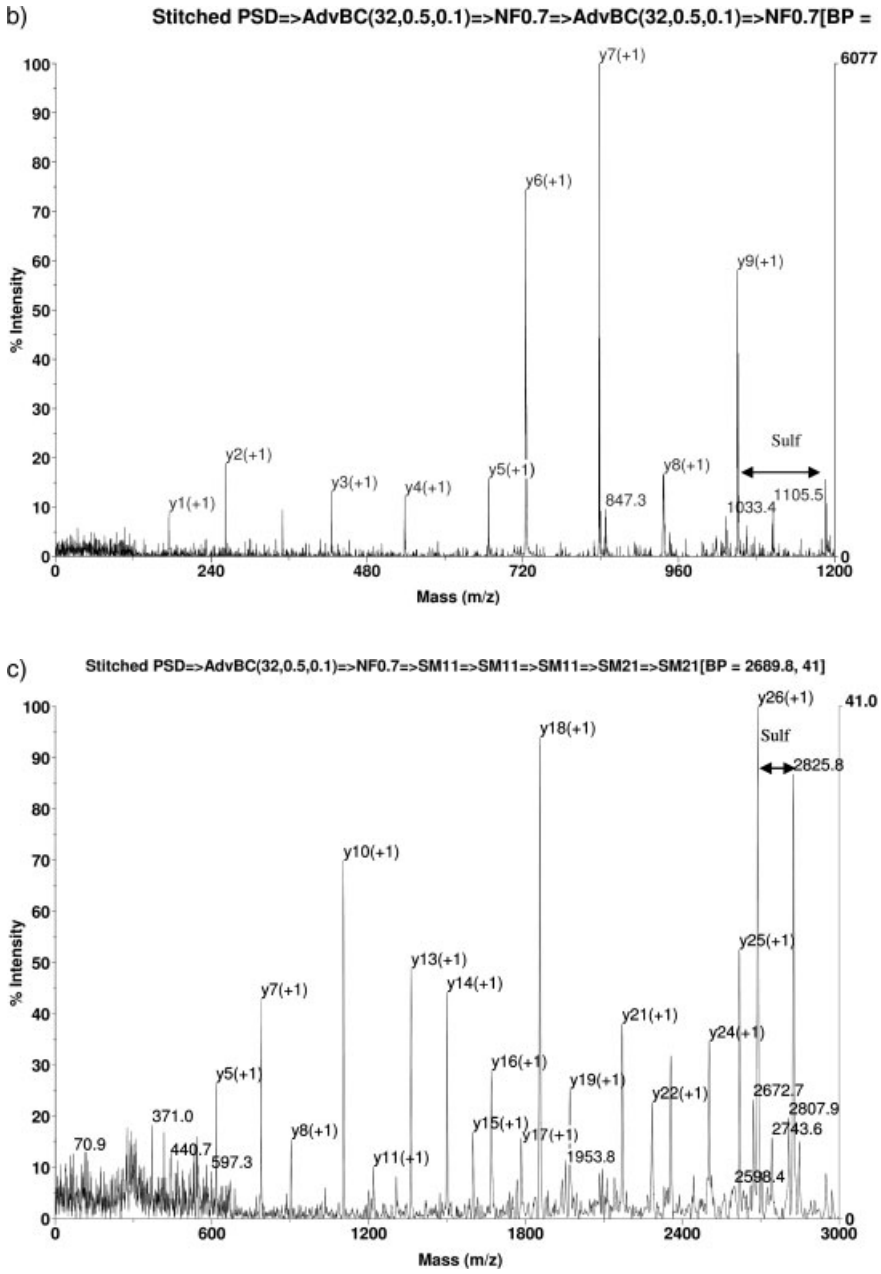


Figure 16.4 CAF-MALDI analysis of the three selected derivatized peptides from 16-kDa tularemic protein. (a) Analysis of the peptide of *m/z* 942.5; (b) analysis of the peptide of *m/z* 1186.2; (c) analysis of the peptide of *m/z* 2825.8. Two spectra (a, b) yielded complete y-ion series, the third spectrum (c) covered the bigger part of the predicted sequence.

Table 16.1 List of proteins and protein homologues identified on all gels shown in the article.

Accession no.	*Protein name*	*Organism*	*E-value (BLAST)*	*MW (kDa) (ORF)*	*pI (ORF)*
AAB00857	Outer membrane protein	*Francisella tularensis*	1E-149	41 286	5.35
AAB06624	Acid phosphatase	*Francisella tularensis* subsp. *novicida*	0	57 750	6.07
AAC18606	Hypothetical protein IP1	*Francisella tularensis*	1E-115	30 818	5.08
AAF16681	Efflux pump regulator SrpS	*Pseudomonas putida*	2E-32	26 427	7.54
AAG33118	Putative peptidyl-prolyl *cis-trans* isomerase	*Francisella tularensis* subsp. *tularensis*	2E-44	10 241	6.22
AAK03485	RpS8	*Pasteurella multocida*	6E-37	14 400	9.40
AAL55890	Omp1-like protein	*Actinobacillus actinomycetemcomitans*	2E-6	20 945	5.32
AAN37798	Malate dehydrogenase	*Francisella tularensis* subsp. *tularensis*	1E-172	36 065	6.05
AAO50751	Hypothetical protein	*Dictyostelium discoideum*	5E-09	27 667	8.89
AAO90157	UDP-3-*O*-[3-hydroxymyristoyl. glucosamine *N*-acyltransferase]	*Coxiella burnetii*	3E-48	21 357	6.51
AAO90940	NADH dehydrogenase I, F subunit	*Coxiella burnetii*	1.8E-145	46 252	5.50
AF345628	Histone-like protein HU form B (hupB)	*Pseudomonas putida*	3E-27	9 500	9.80
CAA70085	Hypothetical 23-kDa protein	*Francisella tularensis*	1E-102	22 433	5.65
CAA74088	Succinate dehydrogenase putative iron sulfur subunit	*Shewanella frigidimarina*	1E-93	26 595	7.88
E81425	Probable periplasmic protein	*Campylobacter jejuni*	1E-10	35 811	5.63
F83048	Probable two-component response regulator	*Pseudomonas aeruginosa*	4E-41	25 506	6.02
G82059	Ribosomal protein S10	*Vibriocholerae*	3E-45	11 900	9.80
G82994	Glycine-cleavage system protein T1	*Pseudomonas aeruginosa*	4E-99	39 584	5.61
NP_049441	Stress response homologue Hsp	*Bacillus subtilis*	8E-19	16 711	5.58

Table 16.1 (cont.)

Accession no.	*Protein name*	*Organism*	*E-value (BLAST)*	*MW (kDa) (ORF)*	*pI (ORF)*
NP_078115	Purine-nucleoside phosphorylase	*Ureaplasma urealyticum*	2E-39	26 892	5.63
NP_106461	Peroxiredoxin 2 family protein	*Mesorhizobium loti*	3E-56	19 698	4.98
NP_229962	Transcription termination factor Rho	*Vibrio cholerae*	1E-168	47 129	5.67
NP_230015	Elongation factor G	*Vibrio cholerae*	0	77 698	4.94
NP_231650	Thymidylate kinase	*Vibrio cholerae*	2E-44	23 726	6.13
NP_232257	Shikimate kinase	*Vibrio cholerae*	1E-41	19 737	5.28
NP_233439	Oxidoreductase, short-chain dehydrogenase/ reductase family	*Vibrio cholerae*	2E-84	25 989	5.87
NP_240094	Orotidine 5′-phosphate decarboxylase	*Buchnera aphidicola*	2E-17	23 369	5.98
NP_245455	Grx2	*Pasteurella multocida*	1E-24	25 211	5.54
NP_246340	RpS8 (ribosomal protein S8)	*Pasteurella multocida*	6E-35	14 397	9.38
NP_246354	RpL3 (50S ribosomal protein L3)	*Pasteurella multocida*	2E-73	22 494	9.65
NP_248984	Probable hydratase	*Pseudomonas aeruginosa*	6E-89	32 430	5.73
NP_249757	Probable short-chain dehydrogenase	*Pseudomonas aeruginosa*	1E-45	21 705	5.56
NP_250274	Succinate dehydrogenase (A subunit)	*Pseudomonas aeruginosa*	0	65 871	5.77
NP_251304	Periplasmic chaperone LolA	*Pseudomonas aeruginosa*	7E-23	23 496	7.27
NP_251852	30S ribosomal protein S1	*Pseudomonas aeruginosa*	1E-172	61 491	5.09
NP_252347	Methionine aminopeptidase	*Pseudomonas aeruginosa*	1E-102	28 360	6.39
NP_252459	Inosine-5′-monophosphate dehydrogenase	*Pseudomonas aeruginosa*	1E-172	52 078	7.00
NP_252688	D-ala-D-ala-carboxypeptidase	*Pseudomonas aeruginosa*	2E-67	48 100	8.62
NP_252964	50S ribosomal protein L11	*Pseudomonas aeruginosa*	2E-43	15 256	9.49
NP_253619	50S ribosomal protein L9	*Pseudomonas aeruginosa*	3E-25	16 061	5.55

Table 16.1 (cont.)

Accession no.	*Protein name*	*Organism*	*E-value (BLAST)*	*MW (kDa) (ORF)*	*pI (ORF)*
NP_273245	Ribosome recycling factor	*Neisseria meningitidis*	7E-53	20544	5.42
NP_273960	Isocitrate dehydrogenase	*Neisseria meningitidis*	0	82254	6.42
NP_281664	50S ribosomal protein L7/L12	*Campylobacter jejuni*	3E-25	12848	4.54
NP_283921	Putative succinyl-CoA synthetase beta subunit	*Neisseria meningitidis*	1E-149	41542	5.17
NP_283922	Putative succinyl-CoA synthetase alpha subunit	*Neisseria meningitidis*	1E-108	30095	6.10
NP_297748	Nucleoside diphosphate kinase	*Xylella fastidiosa*	2E-54	15527	5.66
NP_298764	ABC transporter ATP-binding protein	*Xylella fastidiosa*	3E-83	27447	5.63
NP_404673	Acetyl-coenzyme A carboxylase carboxyl transferase subunit alpha	*Yersinia pestis*	1E-104	35464	7.69
NP_404738	Peptidoglycan-associated lipoprotein Pal	*Yersinia pestis*	7E-19	23262	5.04
NP_405053	Conserved hypothetical protein	*Yersinia pestis*	1E-15	20974	8.50
NP_405629	Septum site-determining protein	*Yersinia pestis*	2E-77	30121	6.85
NP_406411	Serine hydroxymethyltransferase	*Yersinia pestis*	1E-161	45282	6.47
NP_417401	Phosphoglycerate kinase	*Escherichia coli*	1E-122	41937	5.42
NP_420197	GlpX protein	*Caulobacter crescentus*	4E-91	34814	5.19
NP_420538	Glutathione peroxidase	*Caulobacter crescentus*	4E-43	17786	5.98
NP_438875	Transcription antitermination protein (NusG)	*Haemophilus influenzae*	2E-27	12720	5.24
NP_439330	*S*-Adenosylmethionine synthetase	*Haemophilus influenzae*	1E-151	42133	5.98
NP_440208	Hypothetical protein	*Synechocystis* sp.	2E-54	25172	5.58
NP_441321	Adenine phosphoribosyltransferase	*Synechocystis* sp.	7E-41	18795	5.24
NP_455035	Conserved hypothetical protein	*Salmonella enterica* subsp. *enterica* serovar *typhi*	2E-3	39891	8.53

Table 16.1 (cont.)

Accession no.	*Protein name*	*Organism*	*E-value (BLAST)*	*MW (kDa) (ORF)*	*pI (ORF)*
NP_457678	Transcription elongation factor	*Salmonella enterica* subsp. *enterica* serovar *typhi*	8E-41	17 703	4.73
NP_457689	Possible exported protein	*Salmonella enterica* subsp. *enterica* serovar *typhi*	4E-08	25 318	9.58
NP_459033	Putative thiol-disulfide isomerase	*Salmonella typhimurium*	6E-20	38 721	5.11
NP_519175	Probable 3-oxoacyl-(acyl-carrier-protein) synthase II	*Ralstonia solanacearum*	1E-125	44 080	5.62
NP_519332	Probable nucleoside-diphosphate kinase protein	*Ralstonia solanacearum*	3E-37	15 562	5.94
NP_562215	Enolase	*Clostridium perfringens*	1E-156	49 538	4.70
NP_633476	Universal stress protein	*Methanosarcina mazei*	2E-18	30 187	5.40
NP_635907	Biotin carboxylase subunit of acetyl CoA carboxylase	*Xanthomonas campestris* pv. *campestris*	1E-176	49 966	6.33
NP_636420	Bacterioferritin	*Xanthomonas campestris* pv. *campestris*	8E-4	18 508	5.22
NP_641305	Elongation factor Tu	*Xanthomonas axonopodis* pv. *citri*	1E-165	41 836	4.95
NP_681493	Asparaginase	*Thermosynechococcus elongatus*	5E-55	30 941	5.76
NP_710520	Glycerophosphoryl diester phosphodiesterase	*Leptospira interrogans* serovar *lai*	9E-19	39 043	5.29
NP_715694	Protein-export protein SecB	*Shewanella oneidensis*	5E-33	16 900	4.72
NP_719461	Ribosomal protein S6	*Shewanella oneidensis*	4E-32	14 948	6.64
NP_720185	Enhancing lycopene biosynthesis protein	*Shewanella oneidensis*	5E-47	23 692	4.49
NP_720189	2-amino-3-ketobutyrate coenzyme A ligase	*Shewanella oneidensis*	1E-137	44 575	5.69
NP_743663	Adenylate kinase	*Pseudomonas putida*	1E-67	24 362	7.54
NP_743759	(3*R*)-hydroxymyristoyl-(acyl-carrier-protein) dehydratase	*Pseudomonas putida*	3E-33	18 150	6.38
NP_745377	Heat shock protein, HSP20 family	*Pseudomonas putida*	1E-15	16 712	5.41

Table 16.1 (cont.)

Accession no.	*Protein name*	*Organism*	*E-value (BLAST)*	*MW (kDa) (ORF)*	*pI (ORF)*
NP_759173	Citrate synthase	*Vibrio vulnificus*	1E-124	47 039	6.22
NP_780191	50S ribosomal protein L25	*Xylella fastidiosa*	3E-21	10 846	9.47
NP_793495	Enoyl-(acyl-carrier-protein) reductase	*Pseudomonas syringae* pv. *tomato*	9E-86	27 805	5.77
NP_797252	Septum site-determining protein MinD	*Vibrio parahaemolyticus*	2E-76	30 121	6.45
NP_798252	Fumarate hydratase, class I	*Vibrio parahaemolyticus*	1E-177	54 974	5.26
NP_798434	Malonyl CoA-acyl carrier protein transacylase	*Vibrio parahaemolyticus*	3E-75	33 502	5.09
NP_819118	Rhodanese domain protein	*Coxiella burnetii*	4E-32	27 865	6.01
NP_819272	Ribosomal protein L10	*Coxiella burnetii*	9E-40	18 719	8.65
NP_819678	Riboflavin synthase, beta subunit	*Coxiella burnetii*	5E-36	16 290	7.45
NP_819961	Hypothetical protein	*Coxiella burneti*	6E-5	25 730	7.27
NP_820010	Dethiobiotin synthetase	*Coxiella burnetii*	2E-36	25 189	5.08
NP_820646	CBS domain protein	*Coxiella burnetii*	2E-7	22 653	7.73
NP_841528	Conserved hypothetical protein	*Nitrosomonas europaea*	1E-12	10 959	4.90
O50286	Dihydrolipoamide dehydrogenase	*Vibrio parahaemolyticus*	1E-163	50 485	5.54
O87796	Fructose-bisphosphate aldolase	*Pseudomonas stutzeri*	1E-170	38 146	5.26
P09764	50S ribosomal protein L1	*Serratia marcescens*	9E-69	24 628	9.50
P15592	Probable sigma (54) modulation protein	*Pseudomonas putida*	3E-23	11 182	5.71
P24295	NAD-specific glutamate dehydrogenase	*Clostridium symbiosum*	1E-162	49 108	6.33
P25516	Aconitrate hydratase 1	*Escherichia coli*	0	102 616	5.40
P31519	Cationic 19-kDa outer membrane protein precursor	*Yersinia enterocolitica*	2E-11	18 766	7.05
P37395	Thioredoxin	*Cyanidium caldarium*	1E-22	12 200	7.70
P37799	Biotin carboxyl carrier protein of acetyl CoA carboxylase	*Pseudomonas aeruginosa*	7E-23	16 494	4.88

Table 16.1 (cont.)

Accession no.	*Protein name*	*Organism*	*E-value (BLAST)*	*MW (kDa) (ORF)*	*pI (ORF)*
P37812	ATP synthase epsilon chain	*Bacillus subtilis*	1E-14	15 800	6.20
P42468	ATP synthase beta chain	*Burkholderia cepacia*	0	49 801	5.02
P46107	Pyrrolidone-carboxylate peptidase	*Bacillus amyloliquefaciens*	2E-45	24 389	8.98
P46175	30S ribosomal protein S17	*Acyrthosiphon kondoi endosymbiont*	3E-19	9 863	9.57
P48204	GrpE protein	*Francisella tularensis*	1E-103	22 052	4.81
P48205	Chaperone protein dnaK	*Francisella tularensis*	0	69 255	4.86
P57479	3-Dehydroquinate dehydratase	*Buchnera aphidicola*	4E-28	16 330	5.77
P57927	Transke tolase 1	*Pasteurella multocida*	2.6E-209	73 322	5.82
P72157	Phosphoribosylamino-imidazole carboxylase catalytic subunit	*Pseudomonas aeruginosa*	2.9E-53	17 196	5.51
P77928	Superoxide dismutase	*Pseudomonas putida*	1E-69	21 940	5.39
P94797	10-kDa chaperonin	*Francisella tularensis*	1E-46	10 245	5.14
P94798	60-kDa chaperonin	*Francisella tularensis*	0	57 429	4.92
Q01166	Beta-lactamase precursor	*Yersinia enterocolitica*	3E-44	31 983	7.42
Q42689	Glutamine synthetase	*Chlamydomonas reinhardtii*	4E-97	38 285	5.34
Q59309	Glyceraldehyde 3-phosphate dehydrogenase C	*Neisseria meningitidis*	1E-126	37 284	6.13
Q87AA7	Dephospho-CoA kinase	*Xylella fastidiosa*	5E-21	23 434	9.26
Q87SX9	Sulfate adenyltransferase subunit 1	*Vibrio parahaemolyticus*	1E-144	52 926	6.68
Q88NA1	Tryptophanyl-tRNA synthetase, putative	*Pseudomonas putida*	1.4E-90	37 844	6.66
Q890F5	Choloylglycine hydrolase	*Lactobacillus plantarum*	1.9E-28	37 175	8.26
Q8PAV3	Elongation factor Ts	*Xanthomonas campestris* pv. *campestris*	3E-60	30 988	5.45
Q8Z2F0	2,3-Bisphosphoglycerate-independent phosphoglycerate mutase	*Salmonella typhi*	1E-175	57 600	5.77
Q9I6C1	Signal recognition particle receptor FtsY	*Pseudomonas aeruginosa*	9E-99	36 137	5.96

Table 16.1 (cont.)

Accession no.	*Protein name*	*Organism*	*E-value (BLAST)*	*MW (kDa) (ORF)*	*pI (ORF)*
Q9JZ38	ATP-dependent Clp protease proteolytic subunit	*Neisseria meningitidis*	5E-70	22 150	5.49
Q9KLX6	Peptide methionine sulfoxide reductase msrA/msrB	*Vibrio cholerae*	5E-52	19 633	9.40
Q9KNH7	UDP-*N*-acetylglucosamine pyrophosphorylase	*Vibrio cholerae*	7.4E-126	50 985	6.90
Q9KU18	ClpB protein	*Vibrio cholerae*	0	95 930	5.41
Q9KU44	IspH protein	*Vibrio cholerae*	2E-38	21 048	5.63
Q9X6B0	Peroxidase/catalase	*Yersinia pestis*	0	82 501	5.32
Q9ZAW3	Cell division protein ftsZ	*Francisella tularensis*	1E-164	39 746	4.71
Q9ZEP8	3-Methyl-2-oxobutanoate hydroxymethyltransferase	*Pseudomonas fluorescens*	4E-54	31 362	5.42
T01754	Hypothetical protein 102	*Pseudomonas putida*	2E-23	11 200	5.90
T24480	Hypothetical protein T04H1.4	*Caenorhabditis elegans*	2E-02	16 100	6.60
ZP_00102874	Hypothetical protein	*Desulfitobacterium hafniense*	1E-13	12 729	5.70
No homologue 1	–	–	–	25 683	8.49
No homologue 2	–	–	–	19 081	5.66
No homologue 3	–	–	–	13 745	5.13
No homologue 4	–	–	–	11 010	6.70
No homologue 5	–	–	–	16 706	8.17
No homologue 6	–	–	–	16 134	6.39

Table 16.1 (cont.)

Accession no.	*Protein name*	*Organism*	*E-value (BLAST)*	*MW (kDa) (ORF)*	*pI (ORF)*
No homologue 7	–	–	–	20975	9.23
No homologue 8	–	–	–	15700	6.82
No homologue 9	–	–	–	19786	9.10

Proteins were identified by peptide mass fingerprinting (PMF). MALDI mass data were matched against preliminary *Francisella* ORFs database. These preliminary sequence data were provided by *Francisella tularensis* strain SCHU4 genome sequencing consortium. Similarity search of identified ORF was carried out using BLAST algorithm against proteins listed in the nr database of NCBI or Swall database of EBI. Only protein spots with sequence coverage bigger than 20% of ORF sequence have been listed in the table.

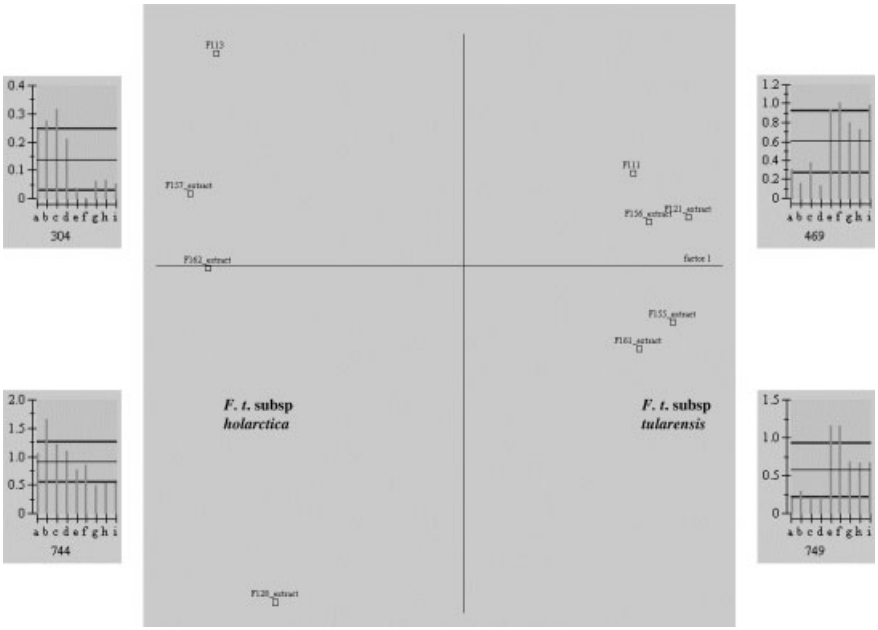


Figure 16.5 Plot of correspondence analysis for nine gels encompassing whole-cell lysates of *F. t.* subsp. *holarctica* (F113, F157, F162, F128) and *F. t.* subsp. *tularensis* (F111, F121, F155, F156, F161). Histograms (% vol.) show abundance alterations of the most important spots representative of a particular class of gels: (a–d) *F. t.* subsp. *holarctica*, (e–i) *F. t.* subsp. *tularensis*.

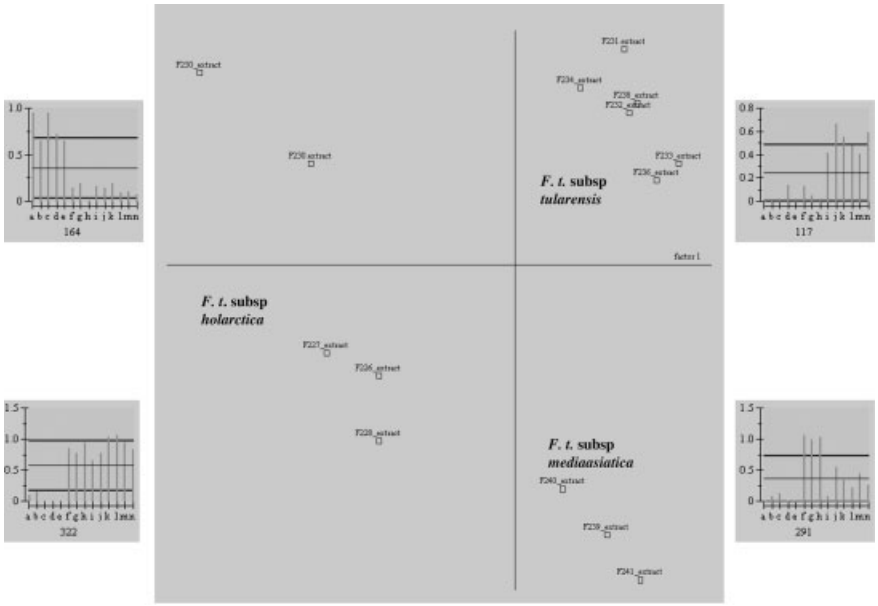


Figure 16.6 Plot of correspondence analysis for 14 gels encompassing basic protein patterns of *F. tularensis* LVS (F250), *F.t.* subsp. *holarctica* (F226, F227, F228, F230), *F.t.* subsp. *mediaasiatica* (F239, F240, F241) and *Francisella t.* subsp. *tularensis* (F231, F232, F233, F234, F236, F238). Histograms (% vol.) show abundance alterations of the most important spots representative of a particular class of gels: (a) *F.t.* LVS, (b–e) *F.t.* subsp. *holarctica*, (f–h) *F.t.* subsp. *mediaasiatica*, (i–n) *F.t.* subsp. *tularensis.*

both types of *F. tularensis* strains could be distinguished into two different classes according to their protein patterns. We have used correspondence analysis, implemented in Melanie that is able to automatically create sets of gels and display significant groups of spots. Figure 16.5 shows the results from correspondence analysis documenting clear discrimination of both *F. tularensis* subspecies along the first factorial axis, including the histograms of the most significant spots representative of a particular class of gels, as well. Furthermore, a higher degree of protein pattern heterogeneity was observed among the strains in the *holarctica* cluster, than among the strains in the *tularensis* cluster.

We have also applied the same approach for protein profiles of basic proteins. In this case, besides *holarctica* and *tularensis* subspecies, the protein profiles of basic proteins of three different isolates of *F.t. mediaasiatica* were included. Additionally, the *holarctica* set of gels was enlarged about 2-DE gel of *F.t.* LVS strain. Figure 16.6 shows the plot of correspondence analysis together with selected histograms of differentially expressed proteins. Likewise in whole-cell lysates, *holarctica* strains, including LVS, were distinguished from *tularensis* ones along the first factorial axis. Furthermore, the set of *mediaasiatica* gels formed the third class that exhibited both *mediaasiatica* unique markers and markers common with *holarctica*

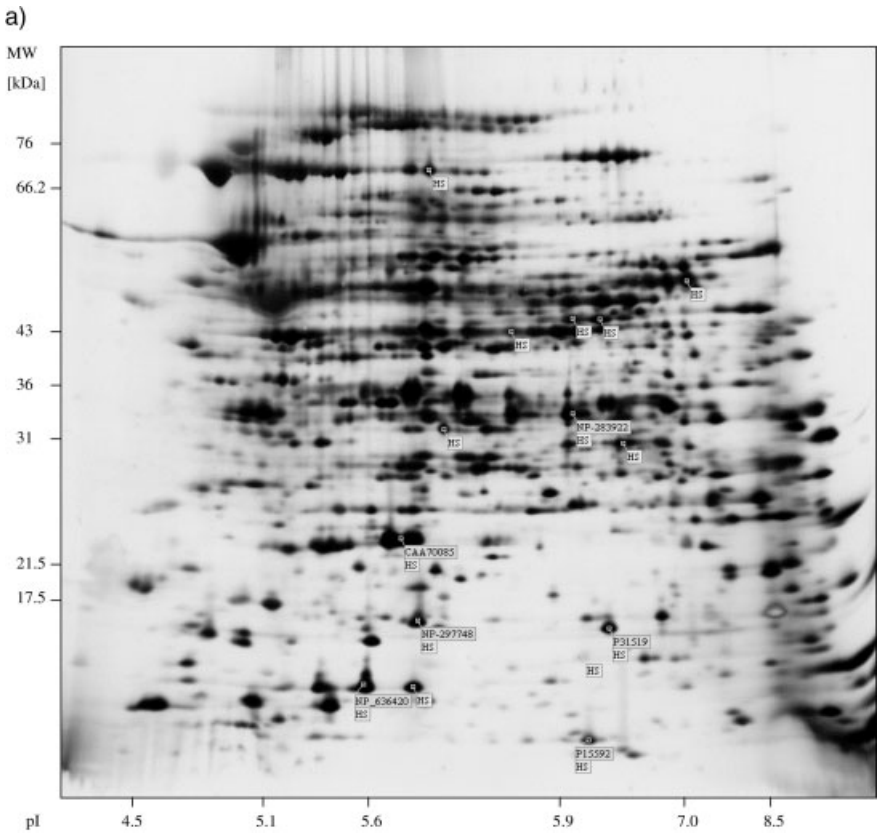


Figure 16.7 Representative silver-stained 2-DE maps of whole-cell extracts and basic proteins of (a, b) *Francisella t.* subsp. *holarctica* and (c, d) *Francisella t.* subsp. *tularensis* SCHU4. Proteins characteristic for *Francisella t.* subsp. *holarctica* or *Francisella t.* subsp. *tularensis* SCHU4 are indicated as HS and TS, respectively. Identified proteins are labelled with accession numbers.

or *tularensis* strains. In the next step, Student's *t*-test was performed on the two given classes (*holarctica* and *tularensis*) of gels. All proteins differentially expressed in the two classes of gels are indicated in Figure 16.7 a–d. The spots marked with accession numbers have been already identified. The histograms showing the abundance alterations of identified proteins comparing *holarctica* and *tularensis* strains are in Figure 16.8.

Among the identified proteins, nucleoside diphosphate kinase and acid phosphatase were implicated as virulence factors of intracellular pathogens. Some pathogenic bacteria release nucleoside diphosphate kinase into the extracellular environment where this enzyme acts as a cytotoxic factor inducing P2Z receptor-mediated death of infected macrophages [22]. Figure 16.8 shows that nucleoside diphosphate kinase occurs predominantly in less virulent *holarctica* strains, however, our latest

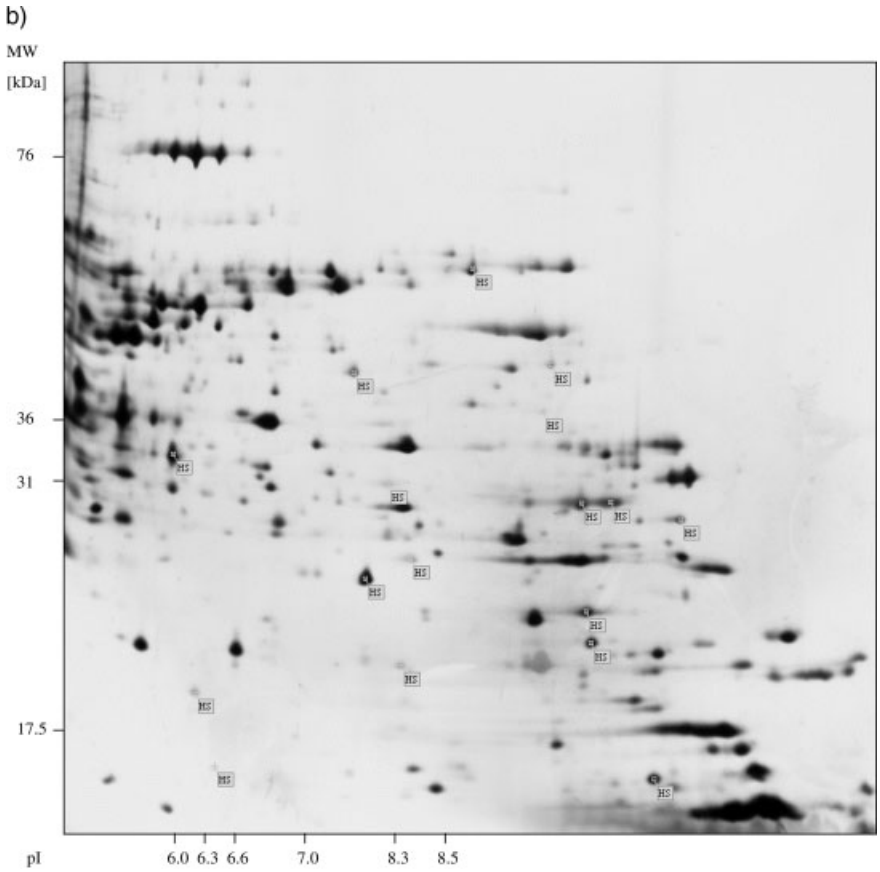


Figure 16.7 b Legend see page 302

results proved that *tularensis* strains express this enzyme, as well, but with a more basic pI value. In contrast, acid phosphatase was found only in highly virulent *tularensis* strains. This enzyme, which can also be released from bacterial cells, positively affects bacteria multiplication via its respiratory burst-inhibiting activity [23].

The other interesting difference between *holarctica* and *tularensis* subspecies concerns the expression of the 23-kDa protein. This protein is highly abundant in whole-cell lysates, in which four major charge variants occur with different abundance distributions in *holarctica* and *tularensis* strains. As can be seen in Figure 16.9 while the more acidic 23-kDa protein variants predominate in *holarctica* subspecies, the more basic ones are characteristic for *tularensis* strains. The biological function of the 23-kDa protein is hard to predict since no protein homologue or similar structural motifs can be found in the public databases. The gene encod-

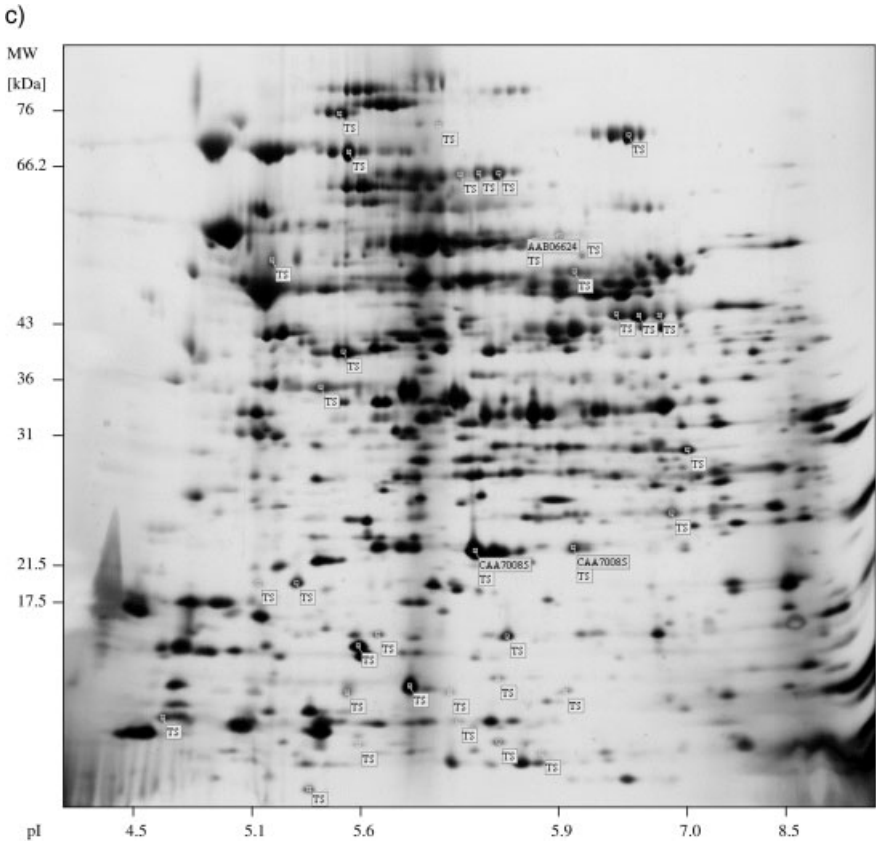


Figure 16.7 c Legend see page 302

ing the 23-kDa protein resides, together with additional three genes, in an *igl* (intracellular growth locus) operon that exists in two copies in *F. tularensis* LVS. Recently, *F. tularensis* LVS mutants were prepared in which either one (*ΔiglC*1) or both copies (*ΔiglC*1+2) of the 23-kDa gene were deleted [24]. In comparison to wild *F. tularensis* LVS, the *ΔiglC*+2 mutant proliferates poorly in macrophages and does not inhibit Toll-like signaling pathway and secretion of tumor necrosis factor α and interleukin 1 β of murine macrophages [25]. The underlying molecular mechanism is not known but there is speculation that a deletion of the 23-kDa gene can influence the expression of downstream genes of the operon. Therefore, the extensive comparative proteome study of wild *F. tularensis* LVS and its mutants have been started. Figure 16.10 depicts the enlarged areas of 2-DE gels of wild and both mutants of *F. tularensis* LVS documenting the loss of the 23-kDa protein in *ΔiglC*1+2 mutant strain.

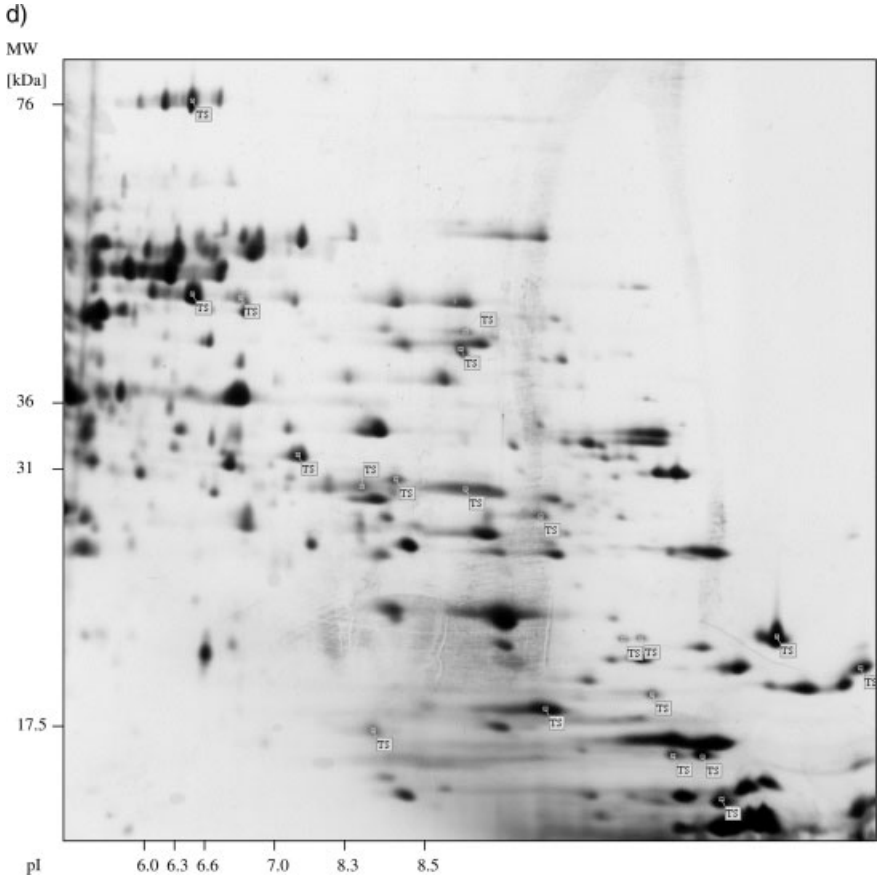


Figure 16.7 d Legend see page 302

16.4 Application of MALDI-TOF Mass Spectrometry for Typing of *F. tularensis*

Rapid and accurate identification of type A and B *F. tularensis* subspecies is crucial in diagnosing disease and recognizing potential warfare threat. However, current molecular methods are mostly capable of discriminating the *F. tularensis* species but not the subspecies. The development of a polymerase chain reaction (PCR) approach based on the existence of a one-base variability in the 16S ribosomal DNA sequences of *F.t. tularensis* and *F.t. holarctica* strains enabled the discrimination of these two subspecies. Nevertheless, there are serious doubts about a consistency at this position among all individual isolates of each subspecies [26].

MALDI-TOF mass spectrometry of intact or corona plasma discharge-treated cells represents a new perspective tool for rapid identification of microorganisms [27]. Using this approach the variety of cell components (biomarkers) such as lipids, lipopolysaccharides, oligosaccharides, proteins, and DNA may be used for

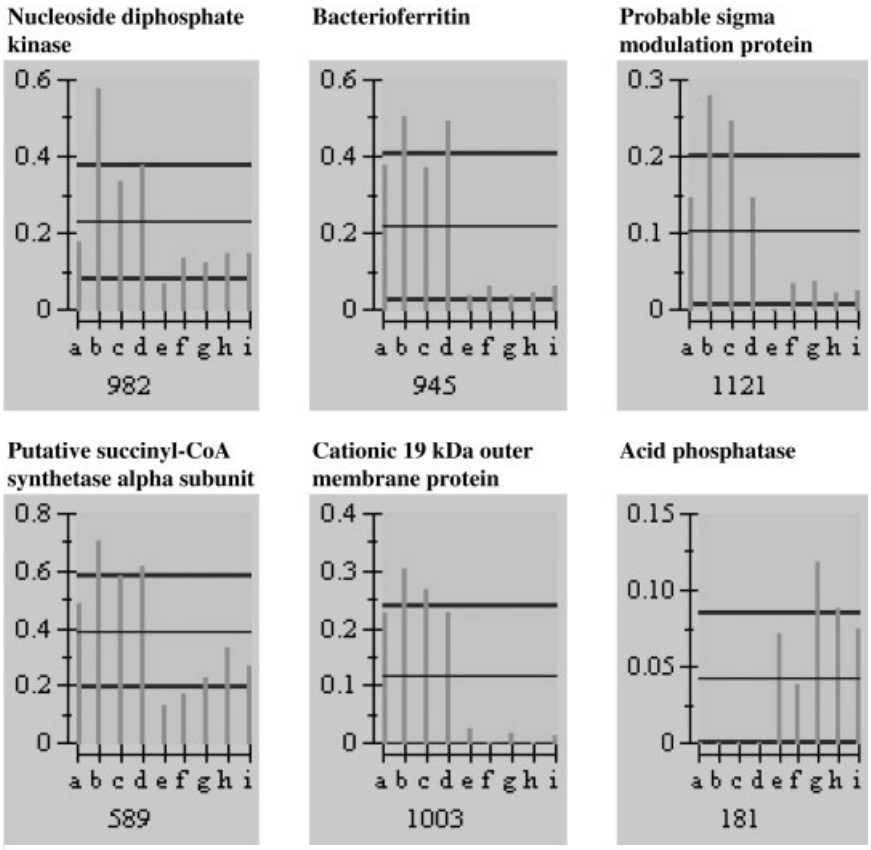


Figure 16.8 Histograms (% vol.) showing the abundance alterations of identified proteins in *holarctica* (a–d) and *tularensis* (e–i) strains.

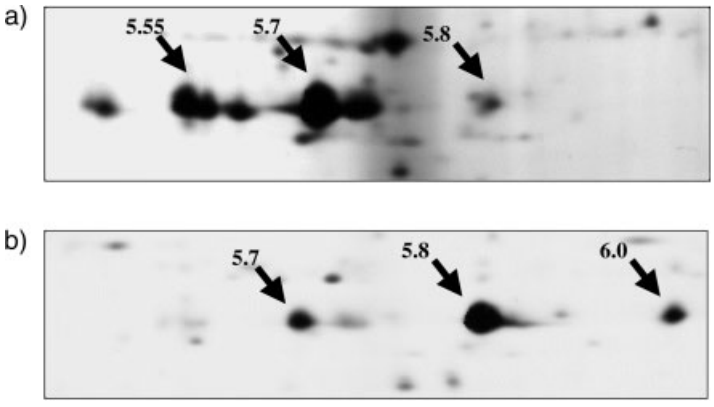


Figure 16.9 Distribution of charge variants of the 23-kDa protein in (a) *holarctica* and (b) *tularensis* subspecies.

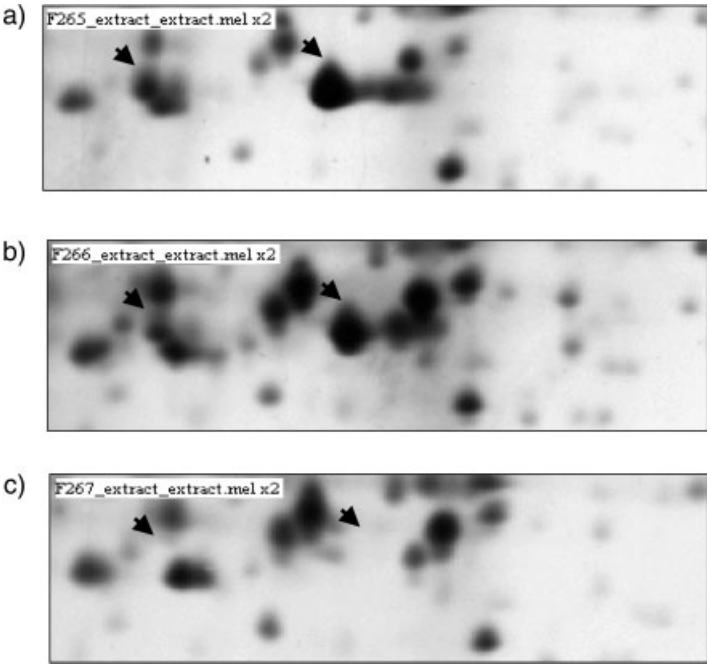


Figure 16.10 The enlarged regions of 2-DE gels of whole-cell extracts of *F. tularensis* LVS (a) and two *F. tularensis* LVS mutants – *ΔiglC*1 (b) and *ΔiglC*1 + 2 (c). The positions of the 23-kDa protein are indicated.

identification purposes. Among them protein molecules attract specific attention because some of them are related to the virulence of bacteria and, as such, will be specific for particular microorganism. The discrimination of individual microbial species is then based on differences in protein profiles that can be observed in the patterns of mass spectra peaks. In our case, we have tested this procedure with the aim of distinguishing *F.t. tularensis* and *F.t. holarctica* subspecies. The harvested bacterial cultures were suspended in a 70 : 30 solution of CH_3CN/0.1% trifluoroacetic acid, centrifuged, and 1 μL of supernatant mixed with equal amount of sinapinic acid was deposited on the target [28]. Figure 16.11 shows positive ion mass spectra of acetonitrile extracts from both *tularensis* and *holarctica* strains. As can be seen from this figure subspecies unique markers were detected, demonstrating the feasibility of this approach for *F. tularensis* typing.

Along with mass spectrometry analysis, the classical proteome approach was adopted for identification of proteins in acetonitrile extracts. Since the acetonitrile extraction does not kill the microbes and only the membrane structure seems to be changed (Fig. 16.12), we suppose that the extract should be a mixture of low molecular weight proteins loosely attached to cell membrane and fragments of outer membrane proteins. In order to resolve properly low molecular weight proteins, tricine SDS-PAGE has been used in the second dimension. The tricine

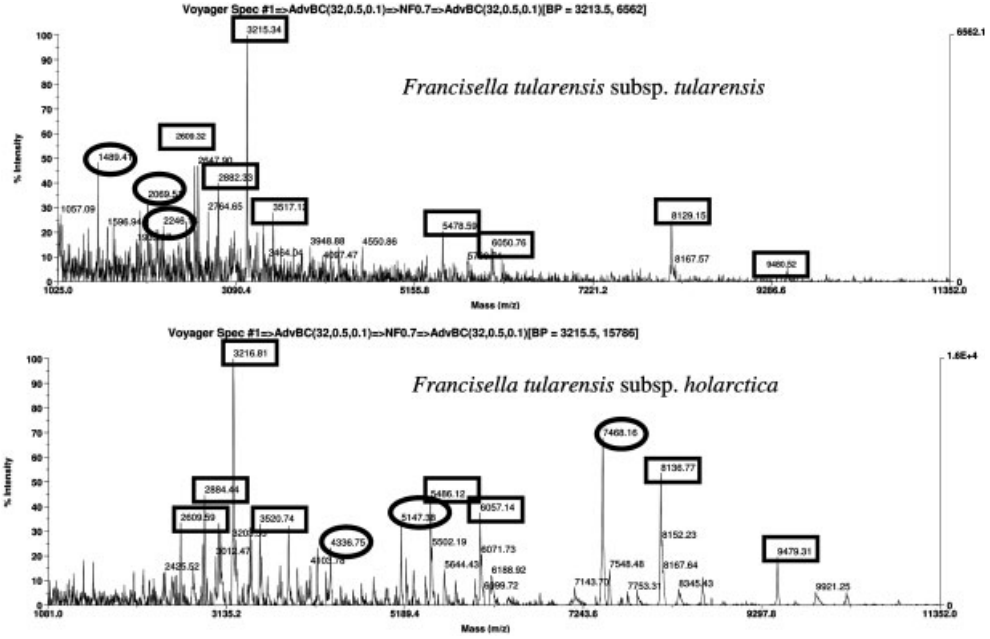


Figure 16.11 Linear-mode MALDI spectra of acetonitrile extracts of *Francisella t.* subsp. *tularensis* and *Francisella t.* subsp. *holarctica*. Subspecies unique markers are indicated by ellipses.

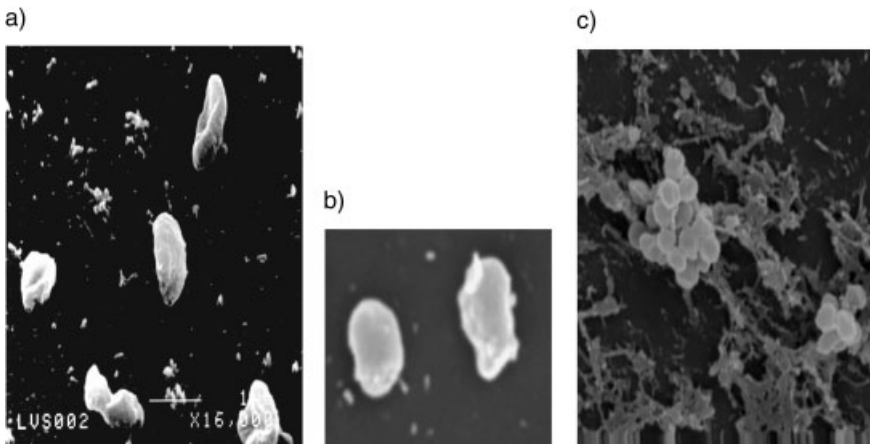


Figure 16.12 (a) Scanning electron microscopy (SEM) of *Francisella tularensis* LVS harvested from solid cultivation plates and resuspended in phosphate-buffered saline (PBS), magnification 16 000×. (b, c) SEM of *Francisella tularensis* LVS after acetonitrile extraction; (b) magnification 16 000×, (c) magnification 15 000×.

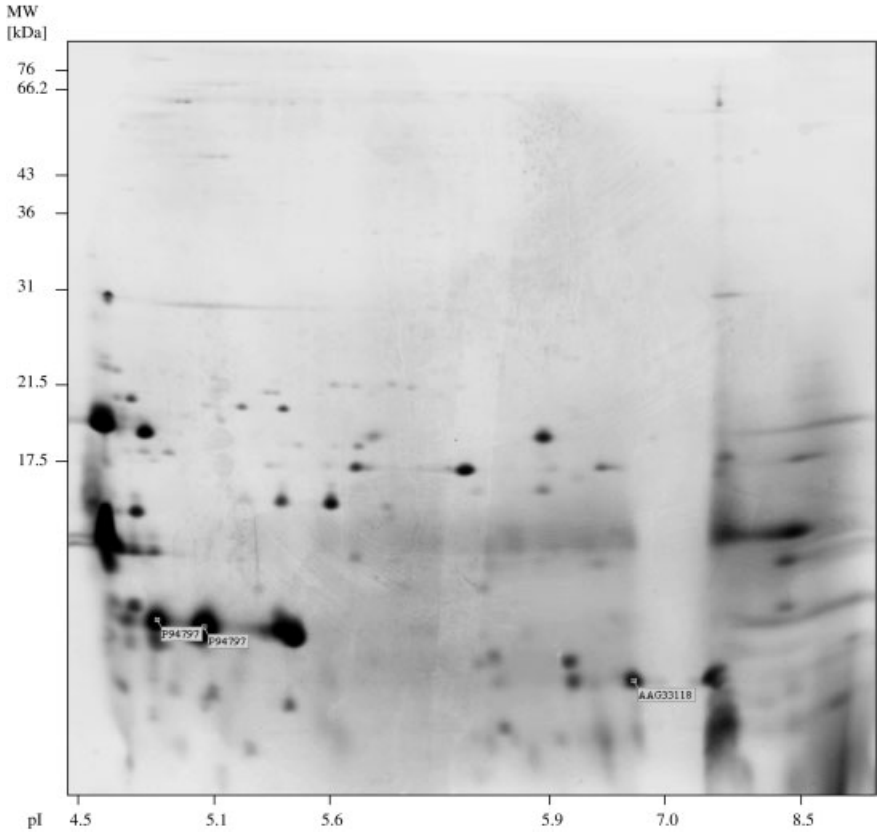


Figure 16.13 Silver-stained 2-DE map of acetonitrile extracts of *Francisella t.* subsp. *holarctica*. Extracted proteins were resolved by IEF in the pH range 3–10, followed by tricine SDS-PAGE gel (16.5% T, 6% C). The accession numbers indicate identified spots.

2-DE gel of acetonitrile extracts obtained from type B *holarctica* strain is shown in Figure 16.13. The 2-DE gel contains 190 spots that are mostly located in the gel region resolving proteins with molecular weights ranging from 4 to 21 kDa. Of them, three proteins, two charge variants of 10-kDa chaperonin and putative peptidyl-prolyl *cis-trans* isomerase, were identified. The secretion of these molecules by other bacteria was observed and their chaperone-like function on maturation of outer membrane proteins has been suggested [29, 30].

16.5 Identification of Tularemic Antigens Recognized by Sera Collected from Naturally Infected Individuals

It is generally accepted that host defense against facultative intracellular pathogens, including *F. tularensis*, is based on cell-mediated immunity [31]. Nevertheless, both natural infection and vaccination with *F. tularensis* LVS induce strong simultaneous production of all three immunoglobulin classes [32]. Furthermore, it was proved for *F. tularensis* and for other intracellular pathogens like *Ehrlichia chaffeensis* and *Salmonella enterica*, that transfer of immune serum could induce some degree of protection [33–35]. Additionally, some tularemic antigens activate both cellular as well as antibody immune response in naturally infected individuals [36]. Therefore, knowledge about the tularemic antigens that trigger humoral immune response would be useful not only for diagnostic purposes but also for identification of candidates for vaccine development.

Serological proteome analysis, based on a combination of 2-DE and immunoblotting with sera from infected individuals, is a suitable approach for the detection of immunoreactive antigens. In our study, first, the sera collected from 44 patients with tularemia were examined by 1-DE immunoblotting against whole-cell lysates. Sera from 20 healthy blood donors served as controls and the reaction with 16 sera from Lyme disease patients was tested as well, because both borreliosis and tularemia are tick-borne diseases. Then the sera reacting with the characteristic profile of tularemic antigens were selected and analyzed by 2-DE immunoblotting. Figure 16.14 shows all 25 identified immunoreactive antigens together with information about the specificity of their immunorecognition. The names of identified spots with their further characteristics are listed in Table 16.1. The 14 antigens provided a highly specific reaction only with human sera collected from patients suffering from tularemia. These antigens involve: chaperones, outer membrane proteins (OMPs), oxidoreductases, different metabolic enzymes, protein participating in transcriptional regulation, and hypothetical protein. Besides OMPs, membrane localization or secretion for an additional four proteins (hypothetical protein, peroxidase/catalase, pyrrolidone-carboxylate peptidase, oxidoreductase) was suggested by an algorithm providing information about topological protein classification [37, 38].

It is interesting to note that of the group of chaperones, 10-kDa chaperonin, ClpB, GrpE, and pyrrolidone-carboxylate peptidase seem to be promising candidates for *Francisella*-specific markers. In contrast, both the chaperone protein dnaK and 60-kDa chaperonin provided cross-reactions with patients suffering from borreliosis, and dnaK was also recognized with sera from healthy blood donors (Fig. 16.14).

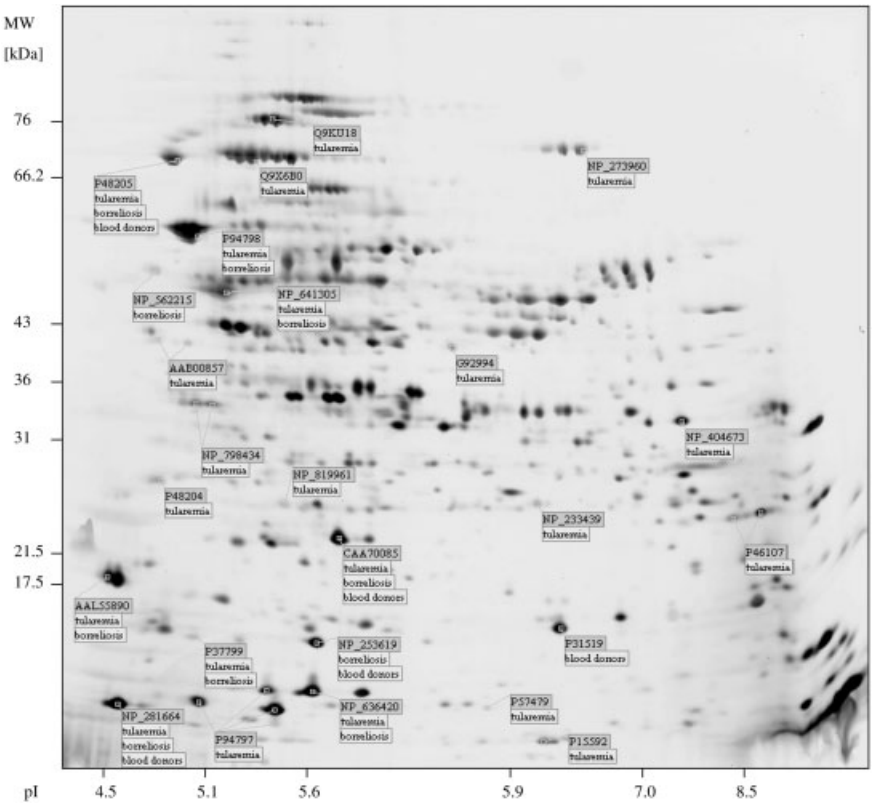


Figure 16.14 Silver-stained 2-DE map of whole-cell lysate of *Francisella tularensis* LVS. Extracted proteins were resolved by IEF in the pH range 3–10, followed by SDS-PAGE (9–16% gradient). The accession numbers indicate the identified immunoreactive spots. The reactivity of identified tularemic antigens with the sera collected from patients suffering from tularemia or borreliosis or from healthy blood donors is indicated.

16.6 Conclusions

Francisella tularensis is not a pathogen of prime clinical interest. The distribution of this microbe is limited to the Northern Hemisphere and there are about 200 new cases per year in the United States. On the other hand, recent bio-terror attacks and other world events have focused the medical community's attention on agents like *F. tularensis* that might be used in biological warfare. From this point of view, it is alarming that we know virtually nothing about the virulence mechanisms of the highly virulent type A *F. tularensis* strains, reliable typing methods do not exist, and, additionally, a safe and efficient vaccine is not available. This chapter describes the first systemic proteomic research performed on this human intracellular pathogen. Further proteomics-based studies should con-

centrate on the analyses of subcellular proteomes (cytosol-soluble, membrane, and secreted proteins), protein alterations associated with cultivation of microbes under stressful conditions or after their engulfment by macrophages. There is an assumption that the combination of results from ongoing proteome and genomic studies together with the genetic manipulation of microbes could bring significant breakthrough in these areas.

16.7 Acknowledgements

The authors wish to acknowledge the excellent technical support provided by Jana Michalickova and Alena Firychova. This work was supported by Grant LN00A033 from the Ministry of Education, Youth and Sport, Czech Republic.

16.8 References

1 G.W. McCoy. *Public Health Bull.* **1911**, *43*, 53.
2 A. Sjöstedt. *Bergey's Manual of Systematic Bacteriology.* Springer, New York, **2003**.
3 S. Saslaw, H. Eigelsbach, H. Wilson, J. Prior, S. Carhart. *Arch. Intern. Med.* **1961**, *107*, 121–133.
4 A. Sjöstedt. *Curr. Opin. Microbiol.* **2003**, *6*, 66–71.
5 M. Ericsson, I. Golovliov, G. Sandström, A. Tärnvik, A., Sjöstedt. *Infect. Immun.* **2001**, *65*, 1284–1829.
6 X.H. Lai, I. Golovliov, A. Sjöstedt. *Infect. Immun.* **2001**, *69*, 4691–4694.
7 A.S. Khan, S. Morse, Lillibridge. *Lancet* **2000**, *356*, 1179–1182.
8 D.S. Burke. *J. Infect. Dis.* **1977**, *135*, 55–60.
9 R.M. Pike. *Health Lab. Sci.* **1976**, *13*, 105–114.
10 N.G. Olsufjev, O.S. Emelyanova, T.N. Dunayeva. *J. Hyg. Epidemiol. Microbiol. Immunol.* **1959**, *3*, 138–149.
11 W.L. Jellison. *Tularemia in North America 1930–1974.* University of Montana, University of Montana Foundation, Misooula, Montana, **1974**.
12 N.G. Olsufjev, I.S. Meshcheryakova. *J. Hyg. Epidemiol. Microbiol. Immunol.* **1982**, *26*, 291–299.
13 H.T. Eigelsbach, C.M. Downs. *J. Immunol.* **1961**, *87*, 415–425.
14 L.S.D. Anthony, P.A.L. Kongshavn. *Microb. Pathog.* **1987**, *2*, 3–14.
15 C.G. Gray, S.C. Cowley, K.K.M. Cheung, F.E. Nano. *FEMS Microbiol. Lett.* **2002**, *215*, 53–56.
16 I. Golovliov, M. Ericsson, G. Sandström, A. Tärnvik, A., Sjöstedt. *Infect. Immun.* **1997**, *65*, 2183–2189.
17 A. Tärnvik. *Rev. Infect Dis.* **1989**, *11*, 440–451.
18 A. Sjöstedt, A. Tärnvik, G. Sandström. *Infect. Immun.* **1991**, *59*, 3163–3168.
19 R.G. Prior, L. Klasson, P. Larsson et al. *J. Appl. Microbiol.* **2001**, *91*, 614–620.
20 A. Görg, C. Obermaier, G. Boguth, A. Csordas, J.J. Diaz, J.J. Madjar. *Electrophoresis* **1997**, *18*, 328–337.
21 U. Hellman, R. Bhikhabhai. *Rapid Commun. Mass Spectrom.* **2002**, *16*, 1851–1859.
22 P. Chopra, A. Singh, A. Koul et al. *Eur. J. Biochem.* **2003**, *270*, 625–634.
23 S. Das, A.K. Saha, A.T. Remaley et al. *Mol. Biochem. Parasitol.* **1986**, *20*, 143–153.
24 I. Golovliov, A. Sjöstedt, A. Mokrievich, V. Pavlov. *FEMS Microbiol. Lett.* **2003**, *28*, 273–280.
25 M. Telepnev, I. Golovliov, T. Grundström, A. Tärnvik, A, Sjöstedt. *Cell. Microbiol.* **2003**, 5, 41–51.

26 M. Forsman, G. Sandström, A. Sjöstedt. *Int. J. Syst. Bacteriol.* **1994**, *44*, 38–46.

27 Y. Hathout, P.A. Demirev, Y.P. Ho et al. *Appl. Environ. Microbiol.* **1999**, *65*, 4313–4319.

28 P.A. Demirev, Y.P. Ho, V. Ryzhov, C. Fenselau. *Anal. Chem.* **1999**, *15*, 2732–2738.

29 M. Harboe, H.G. Wiker, G. Ulvund et al. *Infect. Immun.* **1998**, *66*, 289–296.

30 E. Wahlstrom, M. Vitikainen, V.P. Kontinen, M. Sarvas. *Microbiology* **2003**, *149*, 569–577.

31 J.L. Claflin, C.L. Larson. *Infect. Immun.* **1972**, *5*, 311–318.

32 L. Bevanger, J.A. Maeland, A.L. Naess. *J. Clin. Microbiol.* **1988**, *26*, 433–437.

33 G.M. Winslow, E. Yager, K. Shilo, E. Volk, A. Reilly, F.K. Chu. *Infect. Immun.* **2000**, *68*, 2187–2195.

34 S.J. McSorley, M.K. Jenkins. *Infect. Immun.* **2000**, *68*, 3344–3348.

35 J.J. Drabick, R.B. Narayanan, J.C. Williams, J.W. Leduc, C.A. Nacy. *Am. J. Med. Sci.* **1994**, *308*, 83–87.

36 A. Sjöstedt, G. Sandström, A. Tärnvik, B. Jaurin. *J. Immunol.* **1990**, *145*, 311–317.

37 M. Hubalek, L. Hernychova, J. Havlasova et al. *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.* **2003**, *787*, 149–177.

38 M. Gomez, S. Johnson, M.L. Gennaro. *Infect. Immun.* **2000**, *68*, 2323–2327.

Part VI
Central Nervous System

Biomedical Application of Proteomics
Edited by J.-C. Sanchez, G. L. Corthals, D. F. Hochstrasser

ISBN: 3-527-30807-5

17
Proteomics in Clinical Neuroscience

Pierre R. Burkhard and Jean-Charles Sanchez

17.1
Introduction

Over the last two decades, proteomics-based approaches have been used in both basic and clinical neurosciences with significant success. The number of studies in the field is currently growing exponentially [1–4]. The partial proteome of many central nervous system (CNS) organs, tissues, cells, subcellular structures, organelles, and biological fluids of animals and humans is about to be delineated and compared between normal and pathological conditions. While still in progress, this line of work has already generated a tremendous amount of information regarding many molecular and biological mechanisms underlying highly complex neural functions and pathological processes. Such areas of research include axon growth and nerve regeneration, synapse formation, plasticity, and synaptic transmission, the neurobiology of learning, memory, and other domains of cognition or behavior, the development and aging of the brain, degeneration and related neurological conditions, as well as tumoral, vascular, immune-mediated, traumatic, and neuropsychiatric conditions affecting the CNS. An example, among many others, of such an achievement is provided by the recent study of Husi and Grant [5] in which the multiprotein complexes forming the glutamate *N*-methyl-D-aspartate (NMDA) receptor have been deciphered, using a typical proteomics approach of highly enriched samples. Although a clear understanding of the physiological and pathophysiological pathways underlying all aspects of CNS function and dysfunction remain yet unmet goals, there is little doubt that proteomics will stand as a pivotal technology for this purpose in the years to come.

In this chapter, we provide a review of proteomics research currently available in the field of clinical neuroscience, by addressing the current knowledge of the relevant proteomes, defining their alterations in some neurological diseases and discussing the present and future lines of research that may, in our opinion, benefit enormously from proteomics.

Biomedical Application of Proteomics
Edited by J.-C. Sanchez, G. L. Corthals, D. F. Hochstrasser

ISBN: 3-527-30807-5

17.2 Nervous System-related Samples

The proteomics characterization of the CNS and, to a lesser extent, the peripheral nervous system (PNS) in normal and neurological disease states is mostly and directly determined by the samples under investigation. There are as many proteomes as different sample types such as tissues, cells, cell fractions, or fluids (in this context, mainly the cerebrospinal fluid – CSF). Furthermore, we anticipate an increased complexity with further more specific and fractionated samples. Comparisons between all these CNS-related proteomes may show that they overlap substantially, clarifying the origin and significance of some key proteins. It is therefore of major importance to accurately determine the various physiological proteomes relevant to neuroscience and subsequently how they are altered by pathological processes. In theory, samples can be obtained from the entire brain, gray or white matters, anatomically defined brain structures (cortex, basal ganglia, brainstem, cerebellum, spinal cord, nerve, etc.), functional pathways and tracts (pyramidal, spino-thalamic, visual pathway, etc.), neurotransmitter-related systems (dopaminergic, serotoninergic, noradrenergic, glutamatergic, etc.), cultured or highly purified cell fractions (neurons, astrocytes, oligodendrocytes, microglia, ependymal cells, etc.), cell parts, subcellular structures or organelles (dendrites, axons, synaptic complexes, receptors, nucleus, etc.) and, finally, biological fluids (CSF, aqueous humor, perilymph, nasal mucus, etc.). So far, proteomic studies have been applied to some but not all of these possible samples and a clear integrated overview of these proteomes and their interactions are missing.

We should underline that proteomic analysis of human biological samples, in particular comparisons between normal and disease states, are difficult, since protein expression may be modified by factors unrelated to the disease. These include age of subjects, gender, drug therapies, co-morbidity, multiplicity of normal protein isoforms and technical drawbacks, such as sample preparation flows, diversity of two-dimensional electrophoresis (2-DE) methods used, and post-mortem alterations when autopsy-derived samples are used [6]. A number of limitations are particularly relevant for the study of neural tissues. The number of spots detectable on a single gel generally does not exceed 2000–3000, while the theoretical number of brain-related gene products is probably 10-fold higher. Most of the highly relevant proteins are generally present at low or very low concentrations and tend to escape detection by 2-DE. This is particularly apparent when studying body fluids such as CSF. A polypeptide that can be seen on a gel will depend on the copy number, on the quantity loaded on the gel, and on the method of detection. Protein copy number is spread out from 7 to 8 orders of magnitude in cells and probably up to 12 orders of magnitude in CSF. Albumin and most of the neuropeptides already have 9 orders of magnitude differences in concentration, whereas classical CSF silver-stained 2-DE cannot display more than 4 orders of magnitude. A number of proteins are still unknown and therefore not mentioned in most available protein databases, limiting the capacity of the method to identify all spots of interest. Finally, some hydrophobic or very basic proteins are poorly

separated with the usual 2-DE approach. While all efforts have to be made to minimize these undesirable and challenging problems, they should be kept in mind when critically analyzing neuroscience results.

Here we review the proteomic studies of the most relevant nervous system-related samples – brain tissue and CSF.

17.2.1 Brain Tissue

In a series of successive papers on the topic, Comings et al. were the first to study human brain proteins by 2-DE in the early 1980s [7–12] using very large polyacrylamide gels for the second dimension (20×86 cm). These authors analyzed neocortical gray matter, white matter, cerebellum, and embryonic brain samples obtained from a tissue bank where samples were kept frozen up to several years. Gel images were divided into 20 sections of different sizes based on the molecular mass of known standard proteins, allowing a topographical nomenclature of groups of spots [7]. Proteins were identified as actin, α and β tubulin, glial fibrillary acidic protein (GFAP), and calmodulin as well as neurofilament-, synaptosome-, and myelin-associated proteins. Comparing tissue samples from 145 normal and neurologically affected brains, mostly by degenerative diseases, these authors were able to show that protein patterns were rather consistent across gels from the same and from different individuals, although considerable fluctuation of expression was noted, possibly related to the underlying pathologies [9]. They identified a number of non-genetic variations, in the form of one or more additional spots found in the vicinity of the common train of spots representing the protein, and a few genetic variants, including one of GFAP, called GFAP-Duarte [10], and one of myelin basic protein (MBP), called MBP-Duarte [11]. These mutant proteins, characterized by the presence of a single consistent spot at the basic tail of the protein, could not be assigned to a specific pathology but rather to rare (9 out of 145 brains, including a pair of twins, for GFAP-Duarte) polymorphisms. One the other hand, these authors showed that a massive increase of the expression of some proteins could be found in neurodegenerative diseases such as GFAP in the case of gliosis.

This enormous work was followed, a few years later, by that of Eckerskorn et al. [13] in which adult mouse whole-brain samples were analyzed using 2-DE, electroblotting, amino acid sequencing by Edman degradation, and comparison with the NBRF-PIR protein sequence database and amino acid composition analysis. Spots of both 2-DE gels and siliconized glass fiber membranes were visualized with Coomassie blue staining. From 12 randomly selected, highly expressed spots submitted to further analysis, seven were identified by both methods, including albumin, hemoglobin α_1- and β-chain, glyceraldehyde 3-phosphate dehydrogenase, creatine kinase B-chain, and triose-phosphate isomerase (two separated spots), and one was identified as aldehyde dehydrogenase only by amino acid composition analysis. Mainly methodological, this study was followed a few years later by others from the same group [14, 15] in which, using a similar protocol imple-

mented with a homemade identification software, authors were able to identify 32 additional spots with high or intermediate level of confidence. Those spots corresponded to the following proteins: carbonic anhydrase, α-enolase, actin, fructose bisphosphatase, fructose bisphosphate aldolase, pyruvate dehydrogenase, myoinositol-1-monophosphatase, phosphoglycerate mutase, ubiquitin-C-terminal hydrolase, GTP-binding protein Rab 3b, thymidine kinase, glutathione transferase, glycosyl asparaginase, Wnt-5a protein, malate dehydrogenase, histidine decarboxylase, phosphoprotein phosphatase 2A, serine esterase, cytotoxic T-lymphocyte proteinase 3, transforming protein N-Ras, superoxide dismutase, and an uncertain major histocompatibility complex (MHC) class 1 molecule.

Perhaps the most extensive contribution to the reference map of brain tissue was carried out by Fountoulakis et al., Gauss et al., and Langen et al. a few years ago in both rodents [16, 17] and humans [18]. In the study by Gauss et al., the whole C57BL/6 mouse brain proteome was studied, after fractionation, using very large 2-DE gels (40×30 cm), allowing about 8600 spots to be visualized. From them, they identified 166 spots corresponding to 90 different proteins and showed, by comparisons with another mouse species, that 27 of them displayed genetic variations, in the form of different electrophoretic mobility, spot intensity, or additional spots that appeared strain-specific and possibly related to different protein production or degradation mechanisms [16].

Of particular interest for the study of human brain diseases is the study by Langen et al. in which samples of human parietal cortex were obtained at autopsy from a neurologically intact subject and studied by means of immobilized pH gradient (IPG) strip-based 2-DE and matrix-assisted laser desorption/ionization (MALDI) mass spectrometry (MS) [18]. These researchers were able to identify about 400 spots corresponding to 180 different proteins, the majority of which corresponded to cytoplasmic, mitochondrial and nuclear structural proteins (tubulin chains, synaptosomal proteins, and neurofilaments), and enzymes, in particular those involved in glucose metabolism. Other identified proteins included heat shock proteins, transporters, channels, cholesterol- and lipid-binding proteins, growth factors, enhancers, and transcription factors. Interestingly, some proteins were found to be brain-specific, including the brain form of fatty acid-binding protein (B-FABP), MBP, myelin P2 protein, synaptosomal associated protein 25 (SNAP-25), synaptotagmin, synapsin, 14-3-3 protein, the brain isoform of vacuolar ATP synthase, α-synuclein, GFAP, and glia maturation factor β (GMFβ).

Furthermore, to improve the detection of low-abundance proteins, this group used heparin chromatography to enrich the brain protein extract submitted to 2-DE and MALDI-MS [19], which allowed 18 additional proteins, undetectable on routine 2-DE gels, to be identified for the first time, including T-complex protein 1 subunits, proteasome components and members of the dihydropyrimidinase-related protein (DRP-1 to -3) family.

Taken together, these reference maps covered nearly all abundant proteins detected and represent the most exhaustive human brain proteome to date. This work is still in progress, the latter authors having further identified over 700 gene products so far [3]. In addition, the recent study by Davidsson et al. [20], using a

modified, preparative 2-DE protocol for the analysis of human frontal cortex, produced additional brain protein identifications, including ubiquitin, synaptophysin, syntaxin-binding protein 1, neural specific protein (NSP) 60, Na^+/K^+ATPase, and Cu-Zn superoxide dismutase, among others.

Besides these global analyses, proteomic maps of selected mammalian brain areas have been recently established, including the cerebellum [21–25], hippocampus [24, 25], and pituitary gland [26]. Moreover, Tsugita et al. [27], who analyzed and compared the proteome of five different regions of the mouse brain (cerebellum, cortex, hippocampus, striatum, and cervical spinal cord), observed significant quantitative, rather than qualitative, variability of several proteins between these brain areas, suggesting that protein expression might be regulated with some degree of regional specificity throughout the brain. This hypothesis was further supported by Chen et al. who compared 2-DE-separated proteins from right and left temporal cortex samples and showed that some of them appeared more abundant on one side compared to the other, such as the enzyme carbonyl reductase [28].

Brain tissue is composed of neurons, glia, and many other types of cells, and analysis of brain homogenates does not produce information regarding the specific protein profile of the various cellular components of the brain. This limitation is about to be overcome with the recent publications of proteomic studies using cultured cells. An example is the paper by Oguri et al., in which proteins from cultured hippocampal neurons of rat embryos were separated with a particularly high resolution using seven different narrow-range IPGs for isoelectric focusing (IEF) and two different polyacrylamide concentrations and gel sizes for the second dimension. This method allowed a gigantic, 70×67 cm large cyber-gel to be constructed, yielding 6677 spots to be separated [29], yet no identification was provided.

One essential prerequisite of clinical neuroscience prior to addressing the molecular mechanisms underlying neurological diseases consists of a clear understanding of how brain tissue physiologically develops, matures and changes with age, and whether gender may influence brain protein expression. Initial studies on brain development by Yoshida and Takahashi [30] and particularly by Geschwind and Hockfield [31] focused on early stages of neurogenesis and showed that fetal development of cortical tissue is indeed accompanied by considerable changes in the level of several unidentified, brain-specific proteins. Interestingly, despite minor differences, similar modifications were found in the embryonic spinal cord and in the cortex. These authors found a novel protein called TOAD-64, which was shown to be very early and only transiently expressed during neuronal differentiation and axon outgrowth [32], possibly acting as a major player in axon path finding. Neuron nuclear proteins were also studied by the same group during neurogenesis [33], from the stage of mitotic precursor cells to that of non-mitotic developing neurons, but prior to gliogenesis, again showing up- and down-regulation of several of them.

Aging-related processes have also been looked at by Fountoulakis et al. [34] who confirmed, using a conventional proteomics paradigm, that the expression of brain proteins is indeed quantitatively and qualitatively different between neonatal and

eight-month-old adult rat brain. For example, α-fetoprotein was detected only in the neonatal brain, and DRP-1, -3, -4, stathmin, T-complex protein 1 subunits, isoforms γ and ε of the 14-3-3 proteins, actin, tubulin, and vimentin were much more abundant than in adult brain. The reverse was found for SNAP-25, α- and β-synuclein, DRP-2, GFAP, and many enzymes, while dynamin-1, vesicular-fusion protein, pyruvate carboxylase, and 2-oxoglutarate dehydrogenase were present only in the adult brain. On the other hand, no consistent gender-related changes were detected, with the possible exception of one of the GFAP spots, which was found to be located differently in a significant proportion of male rats (pI 5.0, MW 70 kDa) as compared to females (pI 5.0, MW 55 kDa). Interestingly, when comparing both maps, it is obvious that this gender difference is unlike the GFAP-Duarte polymorphism reported by Comings [10], a potential cause of protein heterogeneity not examined at the time, which could be confidently ruled out a posteriori.

Recently, Taoka et al. [22] similarly studied changes of protein expression in the developing cerebellum of neonatal rat. They showed that protein levels change rapidly and dramatically during the first days of life and, while most of the proteins identified increased in abundance, a significant proportion of them appeared almost exclusively expressed at specific stages of the maturing cerebellum. Identification of some of these proteins showed that their known function was indeed relevant for neural maturation, including leucin-rich acidic nuclear protein, ubiquitin C-terminal hydrolase, FABP, vimentin, stathmin, and others. The results of Tsugita et al., who analyzed the change of protein expression during the adult mouse life, from week 10 to month 24, in five different brain areas, showed similar quantitative changes, suggesting that maturation and other age-related processes continue to occur during the whole lifespan [27].

Taken together, and while this work is still in progress, the proteomic studies of brain tissues have proved quite successful so far at constructing a rather comprehensive map of normal CNS proteins and at capturing some changes that occur during brain maturation and aging. This bulk of information is the necessary background to the study of neurological diseases in humans (see below) and animal models of these diseases, such as genetically modified animals or animals exposed to endogenous or exogenous toxins. Examples include the brain proteomic studies of tau [35] and glycogen synthase kinase-3 [36] transgenic mice as animal models of Alzheimer's disease and tauopathies, of *p53* knockout mice [37], or of rodents treated with the neurotoxic agents cytosine arabinoside, glutamate [38], or kainic acid [39, 40] as models of neuronal death as well as those involved in trauma, stroke, or neurodegeneration, or with the NMDA receptor antagonist MK-801 [41] as a model of schizophrenia.

17.2.2 Cerebrospinal Fluid

Analysis of CSF is by far the most convenient method to approach the biology of certain neurological diseases in living patients and, since the introduction of routine lumbar puncture by Quincke in 1891 has been used as a major diagnostic

tool for a wide range of infectious, inflammatory, immunological, and paraneoplastic conditions affecting both the CNS and the PNS. On the other hand, routine analyses of CSF, which commonly includes cell count, total protein content, and gross proteins and immunoglobulin separation using agarose gel electrophoresis and isoelectric focusing, the concentration of various compounds including glucose and chlorates, as well as selected viral serologies and specific autoantibodies, have proved considerably more limited when assessing other conditions, in particular neurodegenerative diseases, in which CSF is classically labeled as "normal". In fact, the recent use of some sophisticated CSF tests in certain normal-CSF neurological conditions, has helped to build the concept that CSF changes may sometimes be subtle and restricted to specific biological markers, the assessment of which is the only way to capture the biochemical abnormalities underlying the condition under consideration. Examples include the coupled levels of the microtubule-associated tau and the 1–14 fragment of the amyloid precursor protein in the diagnosis of Alzheimer's disease (AD) [42–50], the determination of the CSF concentrations of orexin-A (hypocretin) in narcolepsy [51] or of neopterin and biopterin in the GTP cyclohydrolase I gene-related form of dopa-responsive dystonia [52].

At variance with a traditional and perhaps somewhat limited concept, CSF cannot be viewed as a crude plasma filtrate formed from the passage of blood through the capillary of the choroid plexuses. It is a highly complex mixture encompassing several hundreds of compounds, the concentration of which is tightly regulated via specific mechanisms. With respect to its protein content, typically low with a plasma (70 $g\,l^{-1}$ protein) to CSF (0.35 $g\,l^{-1}$) ratio of about 200. Nearly 80% of it comes from the plasma through the blood-brain barrier (BBB). This high proportion merely reflects the fact that small subsets of plasma-derived proteins are the most abundant in CSF, including albumin, immunoglobulins, and haptoglobin, among others. While analysis of some of these plasma-derived proteins provides information regarding the function and integrity of the BBB, it is less useful at reflecting biological changes occurring within the CNS or the intrathecal compartment. Therefore, the remaining 20% of CSF proteins, which are not plasma-derived and present only in the CSF, constitutes a particular subset of compounds of considerable neurobiological interest. These CSF-specific or CSF-enriched proteins have been differently appreciated thus far, but 2-DE has proved a potent tool to study them in physiological and pathological conditions.

From a proteomics point of view, we propose a rather broad definition of CSF-specific proteins based on a set of five criteria that are not mutually exclusive. To fulfill the definition, CSF-specific proteins must: (1) be present in the CSF and absent in the plasma (e.g. β-trace, γ-trace, GFAP); or (2) be proportionally more abundant in the CSF than in the plasma or other biological fluids (e.g. transthyretin, apolipoproteins E and J); given the considerable difference of protein concentration in both samples, a similar amount of total protein have to be loaded onto the IPG strips to make the 2-DE measures of concentration (we routinely use the % volume) reliable; and/or (3) be qualitatively different in the CSF as compared to the plasma, suggesting that a given protein is simultaneously synthesized inside

and outside the CNS compartment by different sets of cells or that the final gene product results from different post-translational modifications (e.g. transferrin); or (4) be absent in both fluids in normal conditions but may appear, primarily in the CSF and not the plasma, in certain pathological conditions (e.g. 14-3-3 protein, tau); or (5) appear qualitatively different in the CSF in pathological conditions. So far, only a few, easily detectable CSF-specific proteins have been characterized [53, 54].

Some CSF-specific proteins were studied and characterized in detail using proteomic strategies, including transferrin by Gallo et al. [55], GFAP by Wiederkehr et al. [56], neuron-specific enolase by Wiederkehr et al. [57], transthyretin by Szilagyi et al. [58] and β-trace by Harrington et al. [59]. We recently used the concept of CSF-specific proteins for the detection of CSF in clear, aqueous rhinorrhea [60], a frequent rhinologic complaint which can be due to an excess of nasal mucus secretion or a more serious situation whereby CSF leaks through nasal cavities either spontaneously or as a complication of brain trauma, tumor, or surgery. The assessment of CSF rhinorrhea is notoriously difficult since nasal mucus may visually mimic CSF, but is of importance in view of the potentially life-threatening infectious complications a wrong diagnosis may lead to. We therefore established, using conventional 2-DE electrophoresis, five specific protein markers for CSF (transferrin, β-trace, transthyretin, and two unknown proteins, labeled U1 and U2, which were found to be consistently absent from plasma), plasma, and nasal mucus. Rhinorrhea samples were then analyzed with respect to the presence or absence of these markers, and results were compared with clinical data. In our first 18 patients, the method yielded both a high sensitivity and specificity. These results were recently confirmed in a larger population [61], validating the method for a routine use.

The CSF proteome has been extensively studied since the first modern reference map published by Goldman et al. and Merril et al. in the early 1980s [53, 61]. At the time, using an early 2-DE protocol and ultrasensitive silver-staining, these authors were able to separate over 300 spots from which 26 abundant proteins were tentatively identified, mostly by comparison with a previously published serum map by Anderson et al. Interestingly, they found six clusters of spots that were much more prominent in the CSF than the plasma, indeed corresponding to the definition of CSF-specific proteins mentioned above, but none was identified. Over the following two decades, many groups have similarly examined the CSF proteome and of particular interest are the works by Walsh et al. [54], Merril and Harrington [63], Bracco et al. [64], Jellum and Thorsrud [65], Wiederkehr et al. [56, 57, 66, 67], Endler et al. [68], Marshall and Williams [69], Wildenauer et al. [70], Yun et al. [71], Davidsson et al. [20, 72], Manabe et al. [73], Raymackers et al. [74], and Yuan et al. [75].

To date, the most comprehensive proteomic studies on proteins physiologically present in the CSF undertaken by a single laboratory and using the ultimate 2-DE technologies are the ones by Sickmann et al. [76, 77], who have generated nearly 500 spot identifications. However, the vast majority of them corresponded to albumin and its many fragments and isoforms, immunoglobulins, transferrin, and

other abundant plasma-derived proteins, while the number of brain-derived proteins remained relatively small. In fact, an overview of this literature leads to the conclusion that, despite considerable efforts, the number of different CSF proteins yet identified by proteomics probably does not exceed 70, which is far below the results obtained with other human samples including brain tissue and plasma. While this is essentially the result of the minute concentration of most proteins relevant for neuroscience, it is expected that new strategies aiming at improving sample preparation and protein fractionation will emerge in the future to overcome this limitation.

In our laboratory, we have progressively constructed our own CSF reference map which currently encompasses about 220 spots corresponding to over 50 proteins identified by peptide mass fingerprints using MALDI-TOF-MS and Q-TOF-MS and, for a few, by gel matching with published maps, in particular those provided by the ExPASy server (http://www.expasy.ch). This annotated map, obtained with conventional analytical 2-DE using non-linear, pH 3–10, 18-cm-long IPG strip for IEF and 18×20 cm homemade non-linear 9–16% T polyacrylamide gradient gels for SDS-PAGE, is shown in Figure 17.1. In accordance with previous studies, most identified proteins are plasma-derived while only 10 are CSF-specific, including β-trace (prostaglandin-D synthase), GFAP, γ-trace (cystatin C), β_2-microglobulin, transthyretin, transferrin, apolipoproteins D, E and J, and neuronal enolase.

Moreover, analysis of the non-eluted fraction of albumin-depleted CSF using affinity chromatography showed that nearly all areas of the gel are contaminated by many albumin fragments (data not shown), some of them intermingling with other polypeptides into the same spots (Fig. 17.1, spots 48 and 49), adding in some cases further uncertainty to the identification process.

Besides these global approaches of CSF protein content, some authors have started to focus on groups of CSF proteins related to particular neural structures or functions. For example, Borghini et al. [78] specifically analyzed and characterized the apolipoprotein content of lipoprotein particles in CSF and showed a considerable structural and size heterogeneity of these compounds, possibly mirroring the complexity of lipid transport processes going on in the CNS. Even more relevant for the purpose of identifying biomarkers for neurodegenerative diseases, Davidsson et al. [79] have studied the synaptic proteins SNAP-25, GAP-43, synaptotagmin, rab3a, and neurogranin in CSF using a specific procedure including TCA precipitation, liquid-phase IEF and western blotting. These authors were able to confirm that these compounds were indeed present in the CSF in trace amounts, opening the door to a quantitative CSF testing which may reflect altered synaptic function and integrity as it is known to occur in AD and schizophrenia.

The origin of CSF-specific proteins is probably as diverse as the number of cell types from which they are derived, as is their respective pathway to invade the CSF compartment. For example, the choroids plexuses are well-established sites for the synthesis of many polypeptides and proteins, including transferrin, transthyretin, insulin-like growth factor II, insulin-like growth factor-binding protein 2–6, endothelin-1, β-trace, among others [80]. Leptomeningeal cells have been

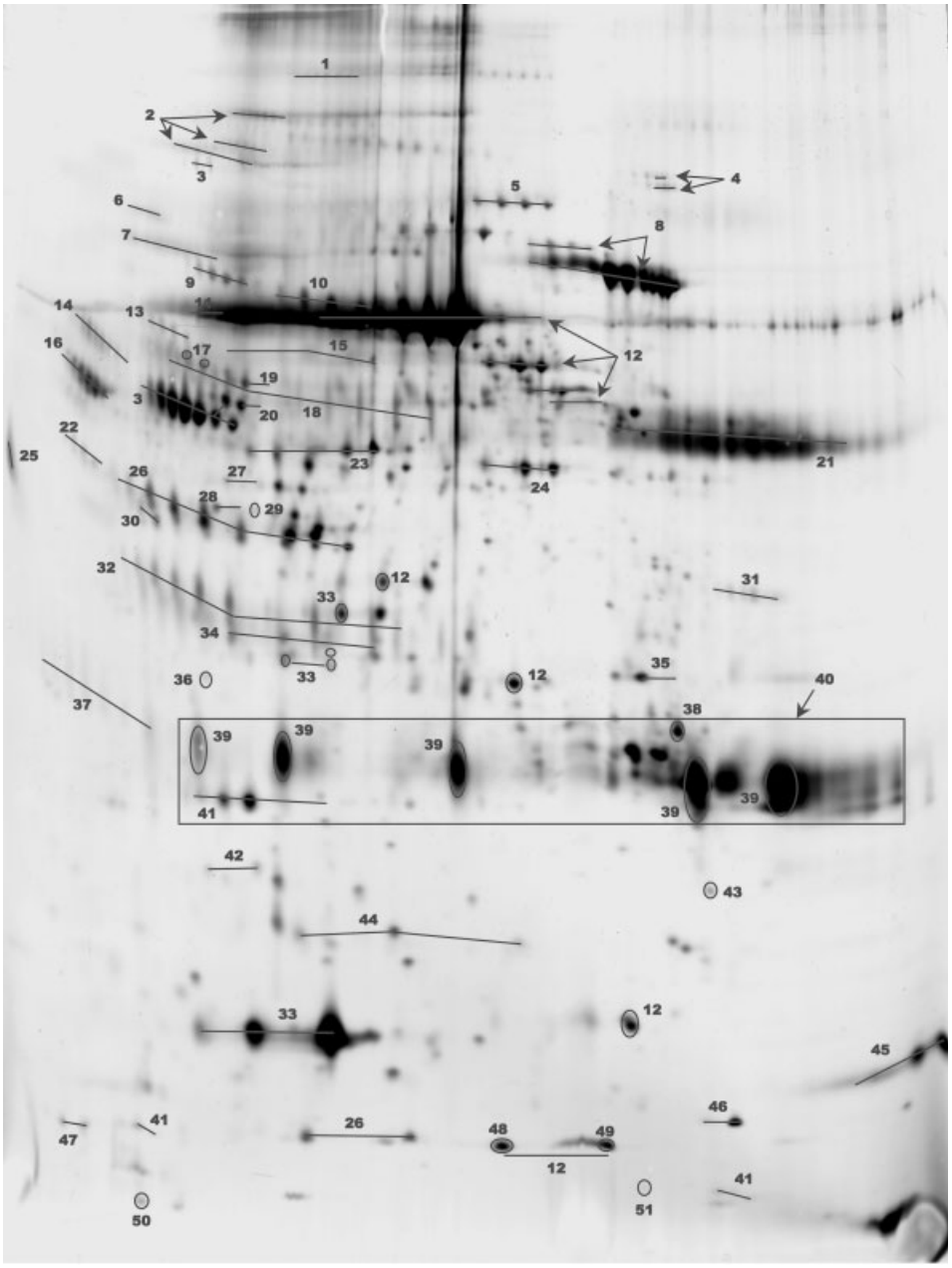

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51

Figure 17.1

demonstrated to secrete several major CSF proteins as well [81]. Very recently, using a proteomic approach, Lafon-Cazal et al. smartly confirmed that astrocytes are also a source of proteins which are actively secreted in the extracellular space and therefore can ultimately reach the CSF [82]. They first analyzed the profile of proteins released in the medium from cultured astrocytes mainly by vesicular transport or alternative mechanisms, identifying a total of 36 proteins including proteases and protease inhibitors, carrier, and antioxidant proteins. In a second step, they matched the mouse 2-DE gel maps from astrocyte-conditioned medium and CSF, showing that 23 astrocyte-secreted proteins were also present in the CSF, including major CSF constituents such as β_2-microglobulin, complement C3, apolipoprotein E, and cystatin C. The secretion capacity of astrocytes was also studied by Chang et al. [83] who demonstrated that these brain cells, when exposed to the proinflammatory cytokines IL-1β, TNFα, or basic fibroblast growth factor, dramatically increased their secretion of several proteins including plasminogen activation inhibitor type 1, ceruloplasmin, or complement C3, again pointing toward glia as a major source of proteins that can potentially be measured in the CSF.

17.3 Proteomics Studies in Neurological Diseases

Several proteomic strategies have been applied to the study of a variety of neurological diseases. We review here some common conditions upon which proteomics seems to have a significant clinical impact.

Figure 17.1 Annotated 2-DE reference map of CSF from a healthy individual. Each protein highlighted in red has been identified by MALDI-TOF-MS, Q-TOF-MS, or, occasionally, by gel matching with published maps of CSF or plasma. 1, α_2-macroglobulin; 2, ceruloplasmin; 3, α_1-antitrypsin; 4, plasminogen; 5, complement C3; 6, complement C1s; 7, prothrombin; 8, transferrin; 9, α_{1B}-glycoprotein; 10, hemopexin; 11, α_2-antiplasmin; 12, albumin; 13, UPA1; 14, α_1-antichymotrypsin; 15, immunoglobulin heavy chain α; 16, α_2-HS-glycoprotein; 17, GFAP; 18, angiotensinogen; 19, antithrombin-III; 20, vitamin D-binding protein; 21, immunoglobulin heavy chain γ; 22, leucine-rich α_2-glycoprotein; 23, fibrinogen γ-A chain; 24, immunoglobulin heavy chain μ, intermediate segment; 25, α_1-acid glycoprotein 1; 26, haptoglobin-1; 27, neuronal enolase; 28, apolipoprotein A-IV; 29, actin, cytoplasmic 1; 30, Zn-α_2-glycoprotein; 31, immunoglobulin heavy chain γ, intermediate segment; 32, apolipoprotein J; 33, transthyretin; 34, apolipoprotein E; 35, complement C4; 36, AMBP; 37, apolipoprotein D; 38, kallicrein 6; 39, prostaglandin D-synthase; 40, immunoglobulin light chains; 41, apolipoprotein A-I; 42, plasma retinol-binding protein; 43, phosphatidylethanolamine-binding protein; 44, haptoglobin-2; 45, cystatin C; 46 hemoglobin β-chain; 47, apolipoprotein C-II; 48, immunoglobulin heavy chain, variable region; 49, β_2-microglobulin; 50, apolipoprotein A-II; 51, kininogen.

17.3.1 Brain Tumor

While proteomics has been widely used in cancer research, only a limited number of studies were dedicated to brain tumors. Narayan et al. were the first to apply 2-DE to surgical biopsies of various types of human malignant brain tumors, including high- and low-grade gliomas, ependymomas and medulloblastomas [84]. They identified a number of proteins, such as actin, tubulin, GFAP, vimentin, glutamic oxaloacetic transaminase, neuron-specific enolase, and guanine nucleotide regulatory protein, and suggested that different protein profiles may parallel tumor histology. In addition, several spots were found to be present in tumor samples but not in normal cortex taken as control, and thus considered as tumor-associated proteins. Menzel and Unteregger [85] and Müller et al. [86] analyzed nuclear proteins obtained from different human tumor cell lines (glioblastoma, low-grade astrocytoma, non-glial tumor, and fibroblasts) by 2-DE and found that differential salt and urea extracts from these samples differ remarkably in their protein profiles. For example, different glioblastoma cell lines produced different protein patterns and, more importantly, exhibited numerous additional spots, which were not detectable in low-grade astrocytoma or fibroblasts cell lines, suggesting that some tumor-associated proteins are located within the nucleus of tumor cells and possibly bound to its DNA. Similarly, the group of Hanash et al. [87] has recently undertaken a large proteomics project aiming at establishing the prognosis of several types of brain tumor, in particular gliomas, based on integrated 2-DE gels obtained from a large number of patients. Preliminary data showed that 22 spots, including oncoprotein 18 and nucleoside diphosphate kinase A, were quantitatively different in low-grade compared to high-grade gliomas. Using a similar paradigm, Zhang et al. compared cultured fetal astrocytes and various glioblastoma cell lines, in particular one overexpressing a mutated form of the epidermal growth factor receptor (EGFR) gene, which results in a truncated receptor with aberrant functions and which is known to be frequently detected in primary glioblastomas [88]. Again, they found by 2-DE that 23 proteins were expressed only or mainly in astrocytes, while 29 were overexpressed or exclusively detected in glioblastoma cell lines. All were identified by MALDI-TOF-MS and LC-MS-MS. Ubiquitin, cystatin B and tissue transglutaminase (TTG) were 2- to 3-fold up-regulated in cell lines harboring the EGFR gene mutation as compared to the wild-type. Furthermore, comparing high-grade, low-grade glioma and non-malignant brain biopsy tissues, they demonstrated that Hsp27, major vault protein, cystatin B, and TTG were overexpressed in high-grade tumors, suggesting that these compounds may be used as potential diagnostic, prognostic, and therapeutic markers in the future.

The search for tumor-specific antigens, which may potentially lead to targeted anticancer immunotherapy, has recently proved fruitful in a number of tumors, including neuroblastomas, where a serum immune response has been found by Prasannan et al. and shown to be directed toward isoforms I and III of β-tubulin in a significant proportion of patients [89]. Moreover, the IgG1 and IgM anti-

bodies reaction was specific for the tumoral β-tubulin isoforms, since sera did not react with those present in normal brain tissue.

17.3.2 Multiple Sclerosis

Multiple sclerosis (MS) is an immune-mediated, inflammatory disease of the CNS of unknown origin characterized by the repeated occurrence of several demyelinating plaques in the white matter of the brain, brainstem, and spinal cord, the topography of which is the main determinant of the clinically apparent symptoms. The diagnosis of MS is based upon suggestive clinical features evolving over time along a typical pattern of relapses and remission periods associated with the presence of multiple and typical lesions seen on magnetic resonance imaging (MRI) scans, some of which are active during an attack of MS. The diagnosis is supported biologically by the presence of an intrathecal synthesis of immunoglobulins (Igs) mainly of the IgG type, taking the form of CSF-restricted oligoclonal bands (OBs), which are routinely detected by agarose gel electrophoresis (AGE) or IEF in up to 90–95% of the patients.

Initial proteomics studies in MS by Jellum and Thorsrud [65], Harrington et al. [90], Bracco et al. [64], Dams [91], Wiederkehr et al. [92], and Endler et al. [68] mainly focused on the oligoclonal pattern of the Ig light chains, which, on 2-DE gels, are particularly well separated at around MW 20–25 kDa and whose identification is easily confirmed using electroblotting and staining with anti-κ and anti-λ antibodies coupled to peroxidase. All groups reported similar findings. The oligoclonal pattern was characterized by the appearance of up to 35 additional spots, which are absent in both the corresponding serum from the same patient and the CSF of normal controls. These spots were slightly more abundant toward the basic half of the gels (cathodic shift) and showed some degree of spatial clustering, which is the presumed mechanism underlying the formation of bands in conventional, one-dimensional electrophoresis. The sensitivity of the 2-DE method to detect an oligoclonal Ig pattern appeared superior to AGE and IEF, based on the observation that some MS patients with negative IEF were found to have a definite oligoclonal pattern on 2-DE gels. It must be stressed that the definition of oligoclonal spots may be hampered by the presence, in the same gel area, of other, non-Ig proteins, the most prominent of which is the prostaglandin D synthase (β-trace protein, see Fig. 17.1). On the other hand, the proteomics approach did not allow increase in the specificity of the oligoclonal pattern, which is well known to occur in many non-MS inflammatory conditions affecting the CNS, although there has been some suggestions that the 2-DE oligoclonal pattern may somehow differ from one inflammatory disease to another.

These initial results have been largely confirmed by others and extended. The Ig content of CSF was further assessed and more specifically analyzed using 2-DE in studies by Walsh et al. [93, 94], Gallo et al. [95], and Wiederkehr et al. [96] in larger populations. Walsh et al. and Wiederkehr et al. isolated immune complexes from the CSF of patients with various neurological diseases using protein A-Seph-

arose affinity chromatography prior to 2-DE. The former group showed that the average number of Ig light-chain spots may vary considerably from one pathology to another, with infectious diseases, such as fungal, syphilitic, and tuberculosis infections, exhibiting up to 300 spots in the area. Moreover, these authors found, in MS, a particular and restricted pattern of CSF clonality not only for IgG but also for IgM and IgA light chains, suggestive of a finite clonal complexity, which did not change over time in sequential samples obtained after an 8-year interval from the same patients.

In addition to these consistent findings, Harrington et al. [90] reported that the expression of other, non-Ig spots may be altered, either increasing or decreasing, in MS, as compared with normal controls or other, non-MS, inflammatory and non-inflammatory neurological diseases. While most of these spots were from unidentified proteins, the modification of some of them was demonstrated to discriminate between the different disease groups, although there was considerable overlap between individual values. For example, one spot (102 in their nomenclature) corresponding to a component of the C3-activator protein was significantly decreased in MS and subacute sclerosing panencephalitis. Similarly, Bracco et al. [64] identified two additional non-Ig, unknown spots at MW 30 kDa and pH 7–7.5 which were found in 75% of the MS patients and not in the controls. Dams identified a train of spots at about MW 38 kDa present only in the CSF and not in the corresponding serum and found that this presumed CSF-specific protein tended to disappear in MS. Wiederkehr et al. were able to demonstrate the presence of an unknown protein (named Px, approximate MW 60 kDa) in some MS patients. This protein appeared only in the Ig fraction of CSF removed by protein A-Sepharose columns and was completely obscured by more abundant proteins in untreated CSF samples. Using a modified method, the same group also found a cluster of spots located above the acidic end of the Ig light chains, which was absent from plasma [96].

Harrington et al., in a 2-DE study on β-trace protein, showed that additional spots immunostained with anti-P5 serum appear in half of the MS CSF samples [59]. These spots, which were of higher molecular weight than the common β-trace spots, were absent in controls, suggesting that the CSF-specific β-trace protein may be qualitatively altered in MS. On the other hand, Wheeler et al. did not detect any protein alterations in the serum or plasma from MS patients, using an original, double-label 2-DE procedure [97], confirming the notion that immune-mediated or other changes of proteins are restricted to the CNS compartment in MS.

More recently, proteomics-based approaches have been used to address the pivotal, yet unsolved issue of defining the major antigen targets against which the CNS-limited humoral response is specifically directed in the form of OBs in MS. Using a combination of different electrophoretic procedures, including SDS-PAGE, 2-DE, non-equilibrium pH gradient electrophoresis (NephGE), and immunoblotting of purified brain and peripheral nerve myelin, Walsh and Murray showed that CSF and serum from MS patients demonstrated an intense humoral response directed against 2′,3′-cyclic nucleotid 3′ phosphodiesterase (CNP), a 46–

48 kDa membrane-associated protein of the oligodendrocyte, lymphocyte, and retina [98]. They demonstrated that only the CNPI, and not the CNPII isoform induced an antibody production that was predominantly of the IgM type, persistent over years, and specifically directed against CNS rather than PNS myelin. This anti-CNPI antibody response was present in 74% of the patients' sera, but in none of the healthy controls and in only a few patients with other neurological diseases. Altogether, these findings supported CNPI as a major target for the humoral response in MS.

In the study by Rand and Houck [99], where they used random phage library methodology to delineate the target antigens of OBs in MS, 2-DE was shown to be much more effective than one-dimensional electrophoresis to detect the absorption of individual OBs by antigenic epitopes. This methodological paper, however, did not provide further indication of the identity of these antigens.

To summarize the currently available data provided by proteomics studies when applied to MS, 2-DE has proved an efficient method to fragment the classical CSF OBs into many, highly patterned oligoclonal spots, whose detailed analysis carries a high sensitivity for the diagnosis of MS, probably superior to any available detection method. Why these results, which have already been consistently reported for two decades, have not propelled CSF proteomic analysis in the forefront of the biological routine diagnosis of MS is probably the result of methodological drawbacks, including results variability between laboratories, costs, and the fact that 2-DE in its present form is time consuming and cumbersome. On the other hand, new avenues of research in the field may overcome some of these limitations. For example, it has been recently demonstrated that OBs can be detected by IEF in tears of MS patients with a sensitivity of 72% and a specificity of 84% [100]. If, as expected, 2-DE exhibits similar, if not superior performances at detecting OBs from tears as conventional methods at detecting OBs from CSF, proteomics of tears may become an interesting diagnostic method for MS, making lumbar puncture potentially unnecessary. Furthermore, proteomic studies have already provided evidence that changes in other proteins than Igs may occur in the CSF of MS patients. Finally, as already demonstrated in recent studies, proteomic approaches may play a key role in the still ongoing search for major antigens underlying directly or indirectly the pathogenesis of plaque formation in MS.

17.3.3 Stroke and Cerebrovascular Diseases

For many years, a number of biological markers have been studied in the CSF and serum of patients with stroke, including CK-BB, lactate dehydrogenase, MBP, S-100β, NSE, GFAP, and tau. Most of them have proved quite useful, although imperfect, indicators of the extension of brain damage and accurate predictors of functional outcome. However, the diagnostic value of these markers has been hampered by their late appearance and delayed peak following brain insult, their poor sensitivity and specificity and the limited understanding of the mechanisms governing their release into the CSF and ultimately the blood. At variance with

the high clinical relevance of biomarkers for the diagnosis of myocardial infarct, for example, the routine use of biomarkers for stroke is still not recommended despite the limitations of clinical assessment (hence the notion of silent stroke) as well as available imaging techniques, notably brain computed tomography (CT) scan and MRI. The contribution of proteomics to the issue of biomarkers for stroke is extensively reviewed elsewhere in this book.

On the other hand, a few proteomics studies have reported protein changes in biological samples from certain uncommon cerebrovascular diseases which may constitute a useful aid to confirm their diagnostic or to understand the underlying pathophysiology. Cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy (CADASIL) is a rare hereditary condition associated with mutations in the *Notch-3* gene, responsible for the occurrence of strokes and a rapidly progressive vascular dementia in middle age, as a result of an occlusive vasculopathy of the small brain arteries due to granular osmiophilic material deposits. Unlu et al. [101] have studied CSF samples from three CADASIL patients using 2-DE and found that a small cluster of spots located above the albumin and transferrin spots appeared in the disease samples but in none of the six control samples. They identified these spots as comprising the complement factor B protein, which, interestingly, is synthesized by brain vessels.

Moyamoya disease is a rare cerebrovascular condition of unknown origin characterized by the development of an abundant network of neoformed vessels at the base of the brain resulting from a progressive stenosis and occlusion of the main trunks of the circle of Willis, causing strokes in children, adolescents, and young adults. Moyamoya disease is frequently associated with cerebral aneurysms, mainly in the posterior circulation, and, rarely, with arteriovenous malformation and cerebral arterial variations. More prevalent in Japan, it is genetically determined in some families and has been found in association with a variety of conditions including Down's syndrome. Some protein abnormalities have been reported in the CSF of such patients, in particular an increase of fibroblast growth factor. Hojo et al., using a conventional proteomic approach, studied CSF samples from three Moyamoya patients and detected a single small spot (MW 12 kDa, pI 5.35) located just below the transthyretin train of spots, which was absent in the samples from four controls. Unfortunately, no identification was carried out [102].

Mitochondrial encephalomyopathy, lactic acidosis, and stroke-like episodes (MELAS) is one of the well-established neurological conditions caused by a point mutation in the mitochondrial DNA. This phenotype is characterized by the association, in an adolescent or a young adult, of a myopathy, various signs of encephalopathic dysfunction and strokes which are typically not restricted to anatomically defined vascular territories and which are believed to be due to tissue anoxia and mitochondrial energy-depletion crises. Using 2-DE, Bentlage et al. [103] studied muscle complexes I, III, IV, and V of the mitochondrial respiratory chain in mitochondrial encephalomyopathies including MELAS, and found a reduced expression of these complexes that paralleled a decreased activity of the corresponding enzymes involved in oxidative phosphorylation. Thus, in this particular setting, 2-DE appeared suitable for an alternative diagnostic test, in particular for the

numerous patients carrying a suggestive phenotype but in whom no mutation could be found.

17.3.4 Creutzfeldt-Jakob Disease

Sporadic Creutzfeldt-Jakob disease (CJD), one of the prion-related transmissible spongiform encephalopathies affecting humans, is a rare but devastating neurological condition leading to death in less than a year in most cases. Typical features include a very rapidly progressive dementia associated with other neurological signs, myoclonus, and a suggestive pseudoperiodic EEG pattern. From a proteomics perspective, CJD is the prototypical example of how a classical proteomic strategy can lead to the successful identification of a diagnostic biomarker relevant for clinical neuroscience. In their first article published in 1986, Harrington et al. demonstrated by 2-DE the abnormal appearance of two acidic, 25–30 kDa spots (130 and 131 in their nomenclature) in the CSF of all 21 CJD patients studied, which were not found in 100 healthy controls nor in a total of 420 patients suffering from a variety of other neurological diseases, with the exception of some cases of herpes simplex encephalitis [104]. Ten years later, the same group used amino acid sequencing to identify spot 130 as isoforms of the 14-3-3 protein, which was confirmed by a positive and specific reaction of these spots after immunoblotting using polyclonal anti-14-3-3 antibodies [105]. The "14-3-3 test" was then applied to CSF samples from 257 neurological patients, including 71 with CJD, and found positive in 68 CJD cases, 4 of 94 cases with dementia of various etiologies and 18 of 66 patients with other neurological diseases not involving dementia. Taken together, the assay demonstrated a sensitivity of 96% and a specificity of 88%. Following this pioneering work, many unrelated groups confirmed the validity of the 14-3-3 test [105–110], allowing the World Health Organization in 1998 to enthusiastically implement a positive test in the set of diagnostic criteria for sporadic CJD.

However, based on the observation that most of these data were from highly specialized referral centers for prion diseases, we recently revisited the true clinical usefulness of the 14-3-3 test when applied to unselected patients presenting with a dementing condition to be assessed [111]. Thus, 100 patients with various degenerative, vascular, secondary, and undetermined dementias were fully investigated using an extensive work-up including CSF analysis. Only two had sporadic CJD. The final diagnosis was then compared with the result of the 14-3-3 assay. We found that 12% of the non-CJD cases were 14-3-3 falsely positive, notably a significant proportion of those with Alzheimer's disease, frontotemporal dementia, and dementia with Lewy body, yielding an extremely low positive predictive value of 14.3%. According to these findings, we strongly questioned the allegedly robust validity parameters of the 14-3-3 tests. These results were further confirmed on over 200 patients with similar results (unpublished data). In fact, this controversial paper was quickly followed by others [112, 113], whose conclusions were in agreement with ours. Very recently, Geschwind et al. [114] examined the results of the 14-3-3 test in a cohort of 32 autopsy-proven sporadic CJD patients

and showed that 17 of them were 14-3-3 positive, granting the assay of a 53% sensitivity only, a figure close to what has been found in the new variant of CJD transmitted to humans by ingested cattle and in the familial form of the disease. Altogether, it seems that unfortunately the 14-3-3 test carries a limited sensitivity and specificity for the diagnosis of CJD and results of this test have to be interpreted both in the correct clinical context and with extreme caution.

On the other hand, because some 14-3-3 isoforms are enriched in human brain, notably neurons, the appearance of the 14-3-3 protein in the CSF can be viewed as a non-specific indicator of neuronal damage. As such, it can be detected in any neurological diseases as long as the magnitude of axonal or neuronal damage is sufficient and occurs at an acute or subacute pace. In keeping with this hypothesis, some preliminary studies have already suggested that the 14-3-3 test may be a reliable prognostic marker in early MS [115], in acute transverse myelitis [116] and in bacterial meningitis [117].

Other biomarkers are clearly needed for the pre-mortem diagnosis of CJD. Some 2-DE studies have been published aiming at profiling the prion proteins PrP(C) and PrP(Sc) by means of 2-DE and immunoblotting of CSF and brain samples [118–120] but, while useful for the molecular classification of prion diseases, it seems that the complexity of the methods used may make such analyses difficult for routine applications, at least at the moment. A few years ago, Hochstrasser et al. demonstrated a massive elevation of apolipoprotein E in the CSF of eight cows affected by bovine spongiform encephalopathy [121]. Choe et al. were able to identify a panel of seven differently expressed spots in the CSF of patients with sporadic CJD, variant CJD and degenerative dementia. Using heuristic clustering, they were able to separate quite convincingly the two forms of CJD, independent of the variations of each of the defined markers [122]. Surprisingly and at variance with the former study, apolipoprotein E was found less expressed in sporadic CJD than in controls. We too recently addressed the issue of biomarkers for CJD by examining CSF and plasma samples from affected patients [123]. We found that FABP increased significantly in CJD cases not only in the CSF but, more importantly, also in the plasma, providing for the first time a potential blood test for the diagnosis of CJD in living patients.

17.3.5
Alzheimer's Disease and Related Dementias

Because of their high prevalence and a major impact upon social, medical, and economic management in developed countries, AD and related degenerative conditions have been a hot topic over the years on which a considerable number of research and clinical studies have focused, most of them dedicated to the delineation of reliable diagnostic, prognostic, and therapeutic markers. Recently, proteomics studies have started to appear on other forms of degenerative dementia, including frontotemporal dementia [124] and dementia with Lewy body. The contribution of proteomics to this enormous work is substantial and is extensively reviewed in Chapter 19.

17.3.6 Parkinson's Disease

Parkinson's disease (PD) is the second most frequent neurodegenerative disease in the general population after AD. Essentially motor in nature, cardinal symptoms of the disease (tremor, rigidity, and akinesia) result from degenerative processes selectively involving the substantia nigra pars compacta, a tiny nucleus bilaterally located in the midbrain. Besides neuronal loss and mild gliosis, nigral dopaminergic neurons demonstrate, in PD, target-shaped intracytoplasmic inclusions, the Lewy bodies, which are mainly, yet not exclusively, composed of *α*-synuclein. This is a physiologic 140-amino-acid cytoplasmic protein found in presynaptic terminals and possibly involved in neurotransmitter release, from which some hydrophobic core fragments tend to aggregate into fibrils.

While quite straightforward in typical cases, the diagnosis of PD may occasionally be misled by other forms of degenerative parkinsonism, including multiple system atrophies, progressive supranuclear palsy and corticobasal degeneration, not-so-rare conditions which could clinically masquerade as PD, in particular at disease onset. So far, no reliable radiological, genetic, or biological tests are available to aid the diagnosis of sporadic PD.

In this setting, Harrington and Merril studied in 1984 CSF samples from 20 PD patients by conventional 2-DE and were able to detect in 75% of their cases, but in none of 91 healthy controls, the appearance of a single abnormal spot (129 in their nomenclature) located at MW 25 kDa, intermingled with Ig light-chains spots [90]. Quantitative changes of 20 additional spots were also noted, notably apolipoprotein A1 and orosomucoid. Unfortunately, these findings were never confirmed by this group nor replicated by others. In fact, we recently undertook a similar study using more recent 2-DE technologies in which 30 CSF samples from PD patients (21 men, 9 women, mean age 72 years) and 30 from age-matched healthy controls (16 men, 14 women, mean age 67 years) were compared. Among several differentially expressed spots, we observed two trains of spots (MW 57.2–57.5 kDa, pI 6.1–6.4 for the first, and MW 54.4–56.1 kDa, pI 6.2–6.6 for the second), which were significantly overexpressed in the PD population. Identification by Q-TOF-MS of several spots showed that these proteins were albumin fragments, a finding consistent with a non-specific alteration of the BBB (unpublished data). On the other hand, we were unable to confirm the presence of spot 129 in any of our PD samples.

To the best of our knowledge, no other proteomic study has been published so far for the purpose of finding biomarkers for PD. A potentially interesting avenue would be to analyze the pattern of synuclein isoforms in the CSF of PD patients, since the location of some forms of synuclein were recently defined on 2-DE gels of rat brain, at approximately MW 14–15 kDa and pI 5.0–5.4 [125].

17.3.7
Huntington's Disease

Huntington's disease (HD) is an autosomal dominantly inherited progressive neurodegenerative disease whose main clinical features include chorea and other movement disorders, neuropsychiatric, and cognitive disturbances. Its genetic defect consists of an unstable expanded CAG trinucleotide repeat in the 5′ coding region of the *huntingtin* gene and, therefore, no complimentary diagnostic biomarker appears yet necessary. However, very recently, Zabel et al. studied the proteome of the R6/2 mice model of HD which expresses exon 1 of the human HD gene, yielding highly expanded CAG repeats and producing a phenotype resembling human HD [126]. In addition, both exhibit intranuclear neuronal inclusions in the cerebral cortex and the striatum. Using very large 2-DE gels, these researchers showed that two distinct proteins, α_1-antitrypsin and αB-crystalline, were differently expressed during the course of the disease as compared with control mice followed over time. Specifically, both proteins appeared to have a progressive increase of expression from the age of 4 weeks to 12 weeks in control mice, while an early increase of α_1-antitrypsin was followed by a dramatic decrease in R6/2 mice at the 12-week time point corresponding to an advanced disease. For αB-crystalline, no modification was observed throughout the disease course, meaning that the age-dependent increase of expression of this protein seen in controls was somehow prevented in HD mice. Moreover, these authors assessed the expression of these proteins in human HD brain tissue and, interestingly, found a modified pattern of a group of spots, which turned out to be identifed as α_1-antitrypsin as well. Taken together, these data are consistent with an abnormal processing of α_1-antitrypsin, an acute-phase protein possibly secreted by astrocytes to counteract tissue damage, in the HD brain of both humans and the mouse model of HD.

17.3.8
Miscellaneous

Besides the quite common neurological diseases reviewed above, various protein abnormalities have also been found in brain trauma, amyotrophic lateral sclerosis, Down's syndrome, mental retardation, progressive supranuclear palsy (PSP), corticobasal degeneration (CBD), and in rare conditions, including Machado-Joseph disease, aceruloplasminemia, Norrie disease, and Hashimoto encephalopathy. Furthermore, proteomics approaches have been applied to a variety of psychiatric conditions, in particular schizophrenia, and promising findings have already been reported by some groups in CSF, brain tissue, and peripheral samples. Finally, neuropharmacology has also been the target of some preliminary proteomic studies, showing protein alterations after exposure to the antiepileptic vigabatrin or to the neuroleptic haloperidol.

17.4 Conclusions

The application of proteomics-based approaches to clinical neuroscience for about two decades has provided a tremendous bulk of information regarding the various nervous system-related proteomes, the molecular mechanisms underlying certain neural functions and dysfunctions, biomarkers for the diagnosis and perhaps the prognosis of neurological diseases and potential targets for future therapeutic interventions. However, neuroscience-oriented proteomics is still at an embryonic stage and several limitations related to access to relevant tissues or biological fluids, sample purification, and proteomic technologies, need to be solved. Moreover, the many, and perhaps somewhat inconsistent, if not contradictory, results obtained from the different laboratories engaged in this line of research would have to be critically and comprehensively evaluated, to allow the major neural proteome changes to be unravelled in normal and disease states. For example, given the complexity of molecular mechanisms operating dynamically and in parallel throughout the CNS, it appears quite clear that the concept of a unique protein abnormality taken as disease biomarker is swinging toward a more intricate framework of multiple and related protein modifications. Providing these issues are circumvented, there is little doubt that proteomics will rapidly emerge as one major player in basic and clinical neurosciences.

17.5 Acknowledgements

We are indebted to Proteome Sciences plc, which supported some studies described in this paper. This work would not have been possible without the precious help of Neftali Rodrigo, Mélanie Côte, Annie Philibert, Odile Carrette, Elisabeth Guillaume, Catherine Zimmermann-Ivol, Laure Allard, Véronique Rouge, and Denis Hochstrasser, to whom we are enormously grateful.

17.6 References

1 Rohlff C. *Electrophoresis* **2000**, *21*, 1227–1234.
2 Morrison R.S., Kinoshita Y., Johnson M.D. et al. *Mol. Cell. Proteomics* **2002**, *1*, 553–560.
3 Fountoulakis M. *Amino Acids* **2001**, *21*, 363–381.
4 Grant S.G., Blackstock W.P. *J. Neurosci.* **2001**, *21*, 8315–8318.
5 Husi H., Grant S.G. *Trends Neurosci.* **2001**, *24*, 259–266.
6 Fountoulakis M., Hardmeier R., Hoger H., Lubec G. *Exp. Neurol.* **2001**, *167*, 86–94.
7 Comings D.E. *Clin. Chem.* **1982**, *28*, 782–789.
8 Comings D.E., Carraway N.G., Pekkula-Flagan A. *Clin. Chem.* **1982**, *28*, 790–797.
9 Comings D.E. *Clin. Chem.* **1982**, *28*, 798–804.

10 Comings D.E. *Clin. Chem.* **1982**, *28*, 805–812.
11 Comings D.E., Pekkula-Flagan A. *Clin. Chem.* **1982**, *28*, 813–818.
12 Comings D.E. *Prog. Clin. Biol. Res.* **1982**, *79*, 181–186.
13 Eckerskorn C., Jungblut P., Mewes W., Klose J., Lottspeich F. *Electrophoresis* **1988**, *9*, 830–838.
14 Jungblut P., Zimny-Arndt U., Klose J. *Electrophoresis* **1989**, *10*, 464–472.
15 Jungblut P., Dzionara M., Klose J., Wittmann-Leibold B. *J. Protein Chem.* **1992**, *11*, 603–612.
16 Gauss C., Kalkum M., Lowe M., Lehrach H., Klose J. *Electrophoresis* **1999**, *20*, 575–600.
17 Fountoulakis M., Schuller E., Hardmeier R., Berndt P., Lubec G. *Electrophoresis* **1999**, *20*, 3572–3579.
18 Langen H., Berndt P., Roder D., Cairns N., Lubec G., Fountoulakis M. *Electrophoresis* **1999**, *20*, 907–916.
19 Karlsson K., Cairns N., Lubec G., Fountoulakis M. *Electrophoresis* **1999**, *20*, 2970–2976.
20 Davidsson P., Paulson L., Hesse C., Blennow K., Nilsson CL. *Proteomics* **2001**, *1*, 444–452.
21 Beranova-Giorgianni S., Pabst M.J., Russell T.M., Giorgianni F., Goldowitz D., Desiderio D.M. *Brain Res. Mol. Brain Res.* **2002**, *98*, 135–140.
22 Taoka M., Wakamiya A., Nakayama H., Isobe T. *Electrophoresis* **2000**, *21*, 1872–1879.
23 Friso G., Wikstrom L. *Electrophoresis* **1999**, *20*, 917–927.
24 Edgar P.F., Douglas J.E., Knight C., Cooper G.J., Faull R.L., Kydd R. *Hippocampus* **1999**, *9*, 644–650.
25 Gozal E., Gozal D., Pierce W.M. et al. *J. Neurochem.* **2002**, *83*, 331–345.
26 Beranova-Giorgianni S., Giorgianni F., Desiderio D.M. *Proteomics* **2002**, *2*, 534–542.
27 Tsugita A., Kawakami T., Uchida T. et al. *Electrophoresis* **2000**, *21*, 1853–1871.
28 Chen W., Ji J., Zhao R., Ru B. *Neurochem. Res.* **2002**, *27*, 871–881.
29 Oguri T., Takahata I., Katsuta K., Nomura E., Hidaka M., Inagaki N. *Proteomics* **2002**, *2*, 666–672.
30 Yoshida Y., Takahashi Y. *Neurochem. Res.* **1980**, *5*, 81–96.
31 Geschwind D.H., Hockfield S. *J. Neurosci.* **1989**, *9*, 4303–4317.
32 Minturn J.E., Fryer H.J., Geschwind D.H., Hockfield S. *J. Neurosci.* **1995**, *15*, 6757–6766.
33 Thormodsson F.R., Redmond L., Hockfield S. *J. Neurochem.* **1995**, *64*, 1919–1927.
34 Fountoulakis M., Hardmeier R., Schuller E., Lubec G. *Electrophoresis* **2000**, *21*, 673–678.
35 Tilleman K., Van den Haute C., Geerts H., van Leuven F., Esmans E.L., Moens L. *Proteomics* **2002**, *2*, 656–665.
36 Tilleman K., Stevens I., Spittaels K. et al. *Proteomics* **2002**, *2*, 94–104.
37 Araki N., Morimasa T., Sakai T. et al. *Electrophoresis* **2000**, *21*, 1880–1889.
38 Charriaut-Marlangue C., Dessi F., Ben-Ari Y. *Electrophoresis* **1996**, *17*, 1781–1786.
39 Krapfenbauer K., Berger M., Friedlein A., Lubec G., Fountoulakis M. *Eur. J. Biochem.* **2001**, *268*, 3532–3537.
40 Krapfenbauer K., Berger M., Lubec G., Fountoulakis M. *Electrophoresis* **2001**, *22*, 2086–2091.
41 Paulson L., Martin P., Persson A. et al. *J. Neurosci. Res.* **2003**, *71*, 526–533.
42 Andreasen N., Minthon L., Davidsson P. et al. *Arch. Neurol.* **2001**, *58*, 373–379.
43 Blennow K., Vanmechelen E., Hampel H. *Mol. Neurobiol.* **2001**, *24*, 87–97.
44 Galasko D., Chang L., Motter R. et al. *Arch. Neurol.* **1998**, *55*, 937–945.
45 Kanai M., Matsubara E., Isoe K. et al. *Ann. Neurol.* **1998**, *44*, 17–26.
46 Riemenschneider M., Lautenschlager N., Wagenpfeil S., Diehl J., Drzezga A., Kurz A. *Arch. Neurol.* **2002**, *59*, 1729–1734.
47 Shoji M., Matsubara E., Kanai M. et al. *J. Neurol. Sci.* **1998**, *158*, 134–140.
48 Sunderland T., Linker G., Mirza N. et al. *JAMA* **2003**, *289*, 2094–2103.
49 Tapiola T., Pirttila T., Mikkonen M. et al. *Neurosci. Lett.* **2000**, *280*, 119–122.
50 Vanmechelen E., Vanderstichele H., Hulstaert F. et al. *Mech. Ageing Dev.* **2001**, *122*, 2005–2011.

51 Nishino S., Ripley B., Overeem S. et al. *Ann. Neurol.* **2001**, *50*, 381–388.
52 Bandmann O., Wood N.W. *Neuropediatrics* **2002**, *33*, 1–5.
53 Goldman D., Merril C.R., Ebert M.H. *Clin. Chem.* **1980**, *26*, 1317–1322.
54 Walsh M.J., Limos L., Tourtellotte W.W. *J. Neurochem.* **1984**, *43*, 1277–1285.
55 Gallo P., Bracco F., Morara S., Battistin L., Tavolato B. *J. Neurol. Sci.* **1985**, *70*, 81–92.
56 Wiederkehr F., Ogilvie A., Vonderschmitt D.J. *J. Neurochem.* **1987**, *49*, 363–372.
57 Wiederkehr F., Bueler M.R., Wacker M., Vonderschmitt D.J. *Electrophoresis* **1989**, *10*, 480–488.
58 Szilagyi A.K., Ha N.T., Szelenyi J., Puskas E. *Electrophoresis* **1993**, *14*, 1079–1081.
59 Harrington M.G., Aebersold R., Martin B.M., Merril C.R., Hood L. *Appl. Theor. Electrophoresis* **1993**, *3*, 229–234.
60 Burkhard P.R., Rodrigo N., May D. et al. *Electrophoresis* **2001**, *22*, 1826–1833.
61 Ricchetti et al., *Head Neck* **2003**, in press.
62 Merril C.R., Goldman D., Sedman S.A., Ebert M.H. *Science* **1981**, *211*, 1437–1438.
63 Merril C.R., Harrington M.G. *Clin. Chem.* **1984**, *30*, 1938–1942.
64 Bracco F., Gallo P., Tavolato B., Battistin L. *Neurochem. Res.* **1985**, *10*, 1203–1219.
65 Jellum E., Thorsrud A.K. *Scand. J. Clin. Lab. Invest. Suppl.* **1986**, *184*, 71–76.
66 Wiederkehr F., Vonderschmitt D.J. *J. Suisse Med.* **1985**, *115*, 368–373.
67 Wiederkehr F., Ogilvie A., Vonderschmitt D.J. *Clin. Chem.* **1985**, *31*, 1537–1542.
68 Endler A.T., Young D.S., Yanagihara T., Currie R.M., Reid J. *J. Clin. Chem. Clin. Biochem.* **1987**, *25*, 61–70.
69 Marshall T., Williams K.M. *Electrophoresis* **1991**, *12*, 461–471.
70 Wildenauer D.B., Korschenhausen D., Hoechtlen W., Ackenheil M., Kehl M., Lottspeich F. *Electrophoresis* **1991**, *12*, 487–492.
71 Yun M., Wu W., Hood L., Harrington M. *Electrophoresis* **1992**, *13*, 1002–1013.
72 Davidsson P., Nilsson C.L. *Biochim. Biophys. Acta* **1999**, *1473*, 391–399.
73 Manabe T., Miyamoto H., Inoue K., Nakatsu M., Arai M. *Electrophoresis* **1999**, *20*, 3677–3683.
74 Raymackers J., Daniels A., De Brabandere V. et al. *Electrophoresis* **2000**, *21*, 2266–2283.
75 Yuan X., Russell T., Wood G., Desiderio D.M. *Electrophoresis* **2002**, *23*, 1185–1196.
76 Sickmann A., Dormeyer W., Wortelkamp S., Woitalla D., Kuhn W., Meyer H.E. *Electrophoresis* **2000**, *21*, 2721–2728.
77 Sickmann A., Dormeyer W., Wortelkamp S., Woitalla D., Kuhn W., Meyer H.E. *J. Chromatogr. B: Anal. Technol. Biomed. Life Sci.* **2002**, *771*, 167–196.
78 Borghini I., Barja F., Pometta D., James R.W. *Biochim. Biophys. Acta* **1995**, *1255*, 192–200.
79 Davidsson P., Puchades M., Blennow K. *Electrophoresis* **1999**, *20*, 431–437.
80 Chodobski A., Szmydynger-Chodobska J. *Microsc. Res. Techn.* **2001**, *52*, 65–82.
81 Ohe Y., Ishikawa K., Itoh Z., Tatemoto K. *J. Neurochem.* **1996**, *67*, 964–971.
82 Lafon-Cazal M., Adjali O., Galeotti N. et al. *J. Biol. Chem.* **2003**, *278*, 24438–24448.
83 Chang J.W., Young D.A., Coleman P.D., O'Banion M.K. *Neurochem. Int.* **2001**, *39*, 349–359.
84 Narayan R.K., Heydorn W.E., Creed G.J., Jacobowitz D.M. *Cancer Res.* **1986**, *46*, 4685–4694.
85 Menzel A., Unteregger G. *Electrophoresis* **1989**, *10*, 554–562.
86 Müller A., Henn W., Unteregger G. *Electrophoresis* **1991**, *12*, 515–523.
87 Hanash S.M., Bobek M.P., Rickman D.S. et al. *Proteomics* **2002**, *2*, 69–75.
88 Zhang R., Tremblay T.L., McDermid A., Thibault P., Stanimirovic D. *GLIA* **2003**, *42*, 194–208.
89 Prasannan L., Misek D.E., Hinderer R., Michon J., Geiger J.D., Hanash S.M. *Clin. Cancer Res.* **2000**, *6*, 3949–3956.
90 Harrington M.G., Merril C.R. *Clin. Chem.* **1984**, *30*, 1933–1937.

91 Dams E. *Acta Neurol. Belg.* **1986**, *86*, 233–242.

92 Wiederkehr F., Imfeld H., Vonderschmitt D.J. *J. Clin. Chem. Clin. Biochem.* **1986**, *24*, 1017–1021.

93 Walsh M.J., Tourtellotte W.W. *J. Exp. Med.* **1986**, *163*, 41–53.

94 Walsh M.J., Tourtellotte W.W., Roman J., Dreyer W. *Clin. Immunol. Immunopathol.* **1985**, *35*, 313–327.

95 Gallo P., Tavolato B., Bergenbrant S., Siden A. *J. Neurol. Sci.* **1989**, *94*, 241–253.

96 Wiederkehr F., Wacker M., Vonderschmitt D.J. *Electrophoresis* **1989**, *10*, 473–479.

97 Wheeler T.T., Jordan T.W., Ford H.C. *J. Neurol. Sci.* **1987**, *78*, 87–92.

98 Walsh M.J., Murray J.M. *J. Clin. Invest.* **1998**, *101*, 1923–1931.

99 Rand K.H., Houck H. *J. Neurosci. Meth.* **2000**, *101*, 131–139.

100 Devos D., Forzy G., de Seze J. et al. *J. Neurol.* **2001**, *248*, 672–675.

101 Unlu M., de Lange R.P., de Silva R., Kalaria R., St Clair D. *Neuroscience Lett.* **2000**, *282*, 149–152.

102 Hojo M., Hoshimaru M., Miyamoto S., Taki W., Kikuchi H., Hashimoto N. *Neurosurgery* **1999**, *45*, 170–173, discussion 173–174.

103 Bentlage H., de Coo R., ter Laak H. et al. *Eur. J. Biochem.* **1995**, *227*, 909–915.

104 Harrington M.G., Merril C.R., Asher D.M., Gajdusek D.C. *N. Engl. J. Med.* **1986**, *315*, 279–283.

105 Hsich G., Kenney K., Gibbs C.J., Lee K.H., Harrington M.G. *N. Engl. J. Med.* **1996**, *335*, 924–930.

106 Beaudry P., Cohen P., Brandel J.P. et al. *Dement. Geriatr. Cogn. Disord.* **1999**, *10*, 40–46.

107 Brandel J.P., Beaudry P., Delasnerie-Laupretre N., Laplanche J.L. *Rev. Neurol. (Paris)* **1999**, *155*, 148–151.

108 Zerr I., Bodemer M., Gefeller O. et al. *Ann. Neurol.* **1998**, *43*, 32–40.

109 Green A.J., Ramljak S., Muller W.E., Knight R.S., Schroder H.C. *Neurosci. Lett.* **2002**, *324*, 57–60.

110 Satoh J., Kurohara K., Yukitake M., Kuroda Y. *Eur. Neurol.* **1999**, *41*, 216–225.

111 Burkhard P.R., Sanchez J.C., Landis T., Hochstrasser D.F. *Neurology* **2001**, *56*, 1528–1533.

112 Mendez O.E., Shang J., Jungreis C.A., Kaufer D.I. *J. Neuroimag.* **2003**, *13*, 147–151.

113 Chapman T., McKeel D.W., Jr., Morris J.C. *Neurology* **2000**, *55*, 1396–1397.

114 Geschwind M., Martindale J., Miller D. et al. *Arch. Neurol.* **2003**, *60*, 813–816.

115 Martinez-Yelamos A., Saiz A., Sanchez-Valle R. et al. *Neurology* **2001**, *57*, 722–724.

116 Irani D., Kerr D. *Lancet* **2000**, *355*, 901.

117 Bonora S., Zanusso G., Raiteri R. et al. *Clin. Infect. Dis.* **2003**, *36*, 1492–1495.

118 Zanusso G., Righetti P.G., Ferrari S. et al. *Electrophoresis* **2002**, *23*, 347–355.

119 Pan T., Colucci M., Wong B.S. et al. *J. Biol. Chem.* **2001**, *276*, 37284–37288.

120 Castagna A., Campostrini N., Farinazzo A., Zanusso G., Monaco S., Righetti P.G. *Electrophoresis* **2002**, *23*, 339–346.

121 Hochstrasser D.F., Frutiger S., Wilkin M., Hughes G., Sanchez J. *FEBS Lett.* **1997**, *416*, 161–163.

122 Choe L., Green A., Knight R., Thompson E., Lee K. *Electrophoresis* **2002**, *23*, 2242–2246.

123 Guillaume E., Zimmermann C., Burkhard P.R., Hochstrasser D.F., Sanchez J.-C. *Proteomics* **2003**, *3*, 1495–1499.

124 Davidsson P., Sjogren M., Andrasen N. et al. *Brain Res. Mol. Brain Res.* **2002**, *109*, 128–133.

125 Zugaro L.M., Reid G.E., Ji H. et al. *Electrophoresis* **1998**, *19*, 867–876.

126 Zabel C., Chamrad D.C., Priller J. et al. *Mol. Cell. Proteomics* **2002**, *1*, 366–375.

18
Human Cerebrospinal Fluid

Pia Davidsson and Michael G. Harrington

18.1
Introduction

The impetus driving the study of cerebrospinal fluid (CSF) is its great potential to window biochemical events that participate in and reflect *in vivo* brain biochemistry. CSF is produced within the blood/brain barrier and drains mainly into the jugular venous system via the arachnoid granulations. Ready access to lumbar CSF in humans has led to most of our current knowledge of normal physiology and of relevant changes in disease states.

We have a long way to go! While CSF provides buoyancy for the brain and a few molecular trafficking processes have been identified, knowledge of its molecular (patho)physiology is rudimentary. Specifically, we do not know the detailed identity and concentration of each molecule in CSF at any age, gender, or chronobiological time point, for either healthy or diseased individuals. Nor do we know the complete synthesis, transport, metabolic, and catabolic pathways for any of the molecular components.

Technical advances – especially in computers, human genome sequence, mass spectrometers, and the chemistry and miniaturization of chromatographic and electrophoretic separations – provide hope. We are convinced that these advanced methods can enable major steps forward in understanding CSF and its source in health and disease.

In this chapter we address experimental design and sample preparation issues that are critical to better understand the CSF reflections of the brain. We then review how the new technologies have been applied to CSF proteins, a major molecular subset. We review all of the CSF proteins identified and quantified by high-resolution, two-dimensional electrophoresis (2-DE), liquid chromatography (LC), and mass spectrometry (MS). We review all proteins that have been implicated in protein studies of diseases of the central nervous system. Finally, we describe strategic and technical challenges for the immediate future of this active research field.

Biomedical Application of Proteomics
Edited by J.-C. Sanchez, G. L. Corthals, D. F. Hochstrasser

ISBN: 3-527-30807-5

18.2 Experimental Design

Clinical/experimental heterogeneity and the timing of collection both have a major influence on the success and interpretation of human CSF studies. Whether the purpose of the experiment is to identify all the proteins in a sample, measure their abundance, or detect differences between populations, reproducibility must be defined. This scientific rigor, though fundamental for any correct interpretation, has been ignored in our field too frequently and we make a plea to consider the following points in designing future strategies.

Clinical heterogeneity, obviously the hardest to manage, derives from the lack of recombinant inbred strains or clones of humans that are genetically restricted and environmentally controlled. The first solution is to define all study participants with the clinical "gold standard" that is relevant for their condition. This is most difficult for those conditions that are least easy to define, such as control groups. These merit the most detailed definition, yet are often only minimally assessed. The second solution is to have sufficient numbers in each group tested. The most relevant approach to determine the number of study participants is to identify "changes" using direct molecule-to-molecule comparisons of at least four per group, perform parametric or non-parametric tests of statistical significance between groups and, for borderline but non-significant changes, perform a power analysis to determine how many more samples are necessary to test for significance. All such candidate significant changes should be verified by testing a second population. As examples in the initial discovery phase of research, we currently use at least 5–10 individuals per clinical group and validate results on a similarly sized second population.

Experimental heterogeneity is compounded by technical variations across these involved and multistep procedures. The solution is to analyze replicates of each sample. The number of repeat experiments should be selected based on each experimental configuration and the variation assigned to each individual test sample. As examples, in our preliminary studies we currently perform 2-DE gel studies in triplicate, and the more variable solution digest LC/tandem MS experiments with seven or more replicates. For measurement of protein abundances, normalization is essential. While this is standard in genetic studies with reference to a constitutive or housekeeping gene, this practice is all too uncommon in proteome work.

Once study participants are selected using the aforementioned gold-standard approach, the timing for CSF collection is best chosen after two additional variables are defined and assessed: (1) time of day or month (especially for hormone-related proteins), and (2) recent behavior (e.g. relevant for sleep disturbance/sleep molecular components or exercise/endorphin peptides). We, for example, perform our research collections for all studies between 1 and 5 p.m. (except for specific time course studies), define immediate behavioral history by structured interview and select discrete days in a menstrual cycle, based on measured hormone levels, when relevant. There are limited studies that confirm a CSF chronobiology [1], but until more is known for each protein, we can assume that rhythmic, behavioral, and hormonal changes will contribute to measured variability.

18.2.1
Sample Collection and Preparation

CSF collection is most frequently and safely obtained by the standard lumbar puncture (LP) procedure in the L3/L4 or L4/L5 interspace. While either cisternal or ventricular collection offer alternative routes, the composition of proteins differs substantially along this neuraxis [2].

The gauge size of the LP needle is the only controllable variable that influences severe leakage of CSF (and its consequence, the post-LP headache syndrome) [3]. A 26-gauge is preferred to introduce drugs, but is too small for passive outflow of CSF for collection. We compromise with a 22-gauge needle. A traumatic tap (blood contamination) invalidates most detailed proteomic studies and should be discarded. The volume collected should be sufficient for anticipated studies. It has also been shown that during LP, a gradient of protein concentration occurs. Therefore, standardized procedures have to be adopted in CSF laboratories. We collected 20 mL in Pasadena and 12 mL in Molndal of CSF, immediately remove any cells and then aliquot with reference to the chronological order of outflow at collection to enable identification of any neuraxis gradient effect. Individual CSF proteins and peptides seem to be very stable when stored at low temperature (we use minus 80 °C), but protein complexes (such as lipoprotein particles) may be disrupted by such freezing and should be purified prior to any cryostorage.

Multiple choices of protein preparations are beyond the scope of this chapter, and enrichment methods are discussed in the section below, but three ubiquitous "prep" procedures merit mention here. Total protein is measured by any standard assay (normal range 0.2–0.5 g L^{-1}). Protein concentration is often required, most commonly by precipitation or ultrafiltration. For the former we use 9 parts ethanol:1 part water at –80 °C overnight; for the latter we use Viva Spin 500 with either 5000 or 3000 molecular weight cut-off membranes. Third, it is frequently necessary to desalt and/or remove small molecular weight components, for which we again use ultrafiltration.

18.3
Two-dimensional Gels of CSF with Protein Identification by Antibodies, Edman Degradation Chemistry, MALDI/MS or LC/MS of Individual "Spots"

We have summarized the identified human CSF proteins from all referenced proteome work [4–13] in Table 18.1. The common names and database accession numbers are provided. In addition to the many immunoglobulin kappa, lambda and heavy chain proteins, 52 of the unique proteins (starting with P) are from SwissProt, while the three digit-only ones are from the NCBI protein database. We encourage everyone to improve this resource (it will be at www.hmri.org, under "Programs" under "Molecular Neurology") and contact us for any errors or additions.

Table 18.1 CSF Protein names and database reference numbers.

Name	*ID*	*Name*	*ID*
Actin	P02570	Haptoglobin	P00737
Albumin	P02768	Haptoglobin (cleaved)	P00738
Alpha-2 macroglobulin	P01023	Hemoglobin beta	P02023
Alpha-2HS glycoprotein	P02765	Hemopexin	P02790
Alpha-1 acid glycoprotein	P02763	HIC-3	Q96JB3
Alpha-1 antichymotrypsin	P01011	Ig alpha chain C	P01876
Alpha-1 antitrypsin (precurser)	P01009	IgG1	226787
Alpha-1B glycoprotein	P04217	IgG 4C	P01861
Alpha-1 microglobulin	P02760	Flavin containing	P31512
Alpha-2 microglobulin	P80188	monooxygenase 4	
Amyloid precurser-like protein	P05067	Neurogranin	Q92686
Angiotensinogen	P01019	Myosin beta heavy chain	P12883
Antithrombin III	P01008	Neuromodulin (GAP-43)	P17677
Apolipoprotein A1	P02647	Nitric oxide synthase ES-NOS	P29474
Apolipoprotein D	P05090	Plasminogen	P00747
Apolipoprotein E	EP02649	Prostaglandin D synthase	P41222
Apolipoprotein J (NA1)	P10909	Ras-related protein (Rab3a)	P20336
Apolipoprotein-AIV	P06727	Retinol binding protein	P02753
Beta-2 microglobulin precurser	P01884	Synaptosomal-associated	P13795
Calgranulin A	P05109	protein 25 (SNAP-25)	
Ceruloplasmin	P00450	Synaptotagmin 1	P21579
Chromogranin A	P10645	Retinol binding protein	P02753
Chromogranin B	P05060	Transferrin	P02787
Collagen type 1 alpha-1	1888409	Transthyretin	P02766
Collagen alpha-2 (VI) chain	87169	Ubiquitin	P02248
Complement C3	P01024	Vitamin D-binding protein	P02774
Complement factor B	P00751	precurser	
Cystatin C	P01034	Zn-alpha glycoprotein precurser	P41222
Fibrinogen beta chain	P02675	Ig kappa, lambda, and heavy	Many
Gelsolin	P06396	chains	
Glial fibrillary acid protein	P14136		

Two-dimensional (2-D) separation using isoelectric focusing (IEF) followed by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE), as originally described by O'Farrell and Klose [14, 15], and later adapted for serum [16], is a powerful method for separating thousands of proteins in biological fluids and an excellent visualization tool for protein expression. A map of human CSF proteins was first presented by Goldman et al. [17]. By comparing CSF samples from 20 unrelated individuals with the corresponding plasma, Goldman et al. found six clusters of proteins that were more prominent in CSF than in plasma. However, no information about their origin and identity was available. The identification of many plasma proteins in CSF was done by comparison with known 2-DE reference patterns of plasma [16, 18]. Since then, several groups have adapted this method [11, 19–35]. The identification of the proteins in CSF has been performed

by comigration of purified proteins, comparison with published 2-D maps of other body fluids, and immunoblotting using specific antibodies.

The combination of 2-DE with partial microsequencing by Edman degradation for protein identification has been used by just a few groups [8, 11, 29, 36]. However, other approaches for protein identification, such as matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF-MS) and electrospray ionization (ESI)-quadropole ion trap (Q-IT) are now the more widely used methods for identification. Fourier transform mass spectrometry (FT-MS), a hybrid quadropole-TOF (Q-TOF) and TOF-TOF instruments are also becoming available, promising increased detection sensitivity, resolution, and a larger dynamic range. These new hybrid MS detectors provide expanded detection capabilities for low-abundance components in CSF and will be instrumental in analyzing these complex and difficult to interpret samples. MALDI-MS is used for peptide mapping and MALDI-post source decay (PSD) analysis is used to determine the amino acid sequence of individual tryptic peptides. If N-terminal sequencing and MALDI-MS analysis fail to provide information, the method of choice is nanoliquid chromatography and tandem MS [37].

The combination of 2-DE with MALDI-TOF-MS and ESI-MS/MS have been demonstrated to be useful for the rapid identification of proteins in CSF. Today, 55 unique proteins, along with many immunoglobulin kappa, lambda and heavy chains, have been identified in human CSF using 2-DE and MS [4, 5, 7–11, 38–42]. The major components in human CSF are isoforms of serum albumin, transferrin, and different immunoglobulins. This result is in agreement with their known abundance in CSF, which contains >80% of these proteins. The large amounts of albumin and IgG also cause overlapping or superimposed protein spots, and therefore some low-abundance proteins may remain concealed on 2-DE gels. CSF contains several transport-binding proteins (transthyretin, hemoglobin β-chain, apolipoproteins), signal proteins (cystatin C, complement C4 β-chain, β_2-microglobulin, and antithrombin III), and immune-related proteins (Ig-chain, Ig-1 chain c-region).

The 2-DE gel approach has limitations in terms of dynamic range, sensitivity, and the high amount of salts, sugars and lipids versus the low levels of total protein in human CSF. Moreover, certain types of proteins are systematically excluded from the 2-DE map, many of which may have important functions. These include membrane/hydrophobic, very basic or acidic, very large (>150 kDa), small (<10 kDa), or low-abundance proteins. Up until now, only a few neuron-specific proteins have been identified by 2-DE and MS [11–13, 19, 35, 36]. Alternative approaches for the purification of these classes of CSF proteins are necessary.

Since 2-DE is unsuitable for peptides much smaller than 10 kDa, several different strategies are employed such as: ion-exchange, size-exclusion, affinity, and reverse phase chromatography followed by MS [43, 44]. Recently, a new concept for the analysis of peptides was developed, the Differential Peptide Display™ (DPD). In the DPD process, peptides were extracted and analyzed by a combination of reverse phase high-performance liquid chromatography (RP-HPLC) and MALDI-MS. The peptide pattern of a sample was plotted as multidimensional maps [43, 45, 46]. Another attractive and rapidly growing field is the protein chip technology

that captures proteins directly on a modified MALDI plate in which fractionation is based on differential protein affinities for varied surface chemistries (SELDI). This further reduces sample complexity in CSF [47].

18.3.1 Prefractionation of CSF Prior to 2-DE

It is frequently impossible to observe discrete minor changes in the pattern of total CSF proteins. Some of the abundant CSF proteins are "house-keeping" proteins, such as albumin, IgG, and transferrin. In order to detect low-abundance proteins in CSF, increased protein amounts were applied to the 2-D gel. This did not result in better detection, as disturbance of the 2-D proteins pattern was observed. One of the major drawbacks for studying proteins in CSF using 2-DE gel electrophoresis is its inherent limit in loading capacity despite its high resolution. The circumvention of these limitations led to alternative approaches to isolate and characterize CSF proteins with protein prefractionation prior to 2-DE. Various techniques have been described, but not many additional CSF proteins have been detected. Affinity chromatography with immobilized cibacron blue F3G-A has been used to deplete mainly albumin from CSF [48]. Another approach is to remove albumin and other abundant serum-derived proteins by immobilized anti-human serum proteins. This enhancement technique was first applied in 1982 [22]. However, affinity columns are not ideal because of non-specific binding due to different matrix effects and subsequent uncontrolled loss of proteins. Another way to fractionate CSF for subsequent 2-DE analysis is to use separation by size-exclusion chromatography [34].

Preparative IEF has been demonstrated to enrich low-abundance proteins from complex biological mixtures in sufficient quantities for detection of neuron-related proteins in CSF by immunoblotting [5, 49]. Tau-protein (a marker for neurodegeneration) and synaptic proteins (markers for synaptic function) have been detected in nanogram per liter concentrations in CSF. The mechanism by which preparative IEF has been so successful in identifying more CSF proteins appears to be more than would be expected simply from the enrichment of a narrow pI distribution of proteins. We speculate that there is removal of other "interfering" substances, presumably small molecules. In another strategy, liquid-phase IEF was used to prefractionate individual samples, then selected IEF fractions were analyzed on SYPRO Ruby stained 2-DE gels, with the final protein identification by MALDI-TOF-MS [7, 50]. Studies using other types of apparatus, like the Gradiflow [51] or the sol-IEF device [52], also showed the necessity of prefractionation when analyzing complex samples such as CSF or serum.

Although unpublished, an interesting enrichment strategy was reported by GenProt scientists recently (5th Sienna meeting "From Genome to Proteome" in Sienna, Italy, September 2002). They described pooling a large amount of CSF (over 2 L) in order to "dig deep" for low-abundance molecules that are common to a population group, using multiple LC methods followed by 2-DE or LC/MS. Their results will be of interest to compare with existing work, where the tendency is to

start with relatively small (<20 mL) amounts, and preserve the genotypic and phenotypic individuality of each sample.

18.3.2 Two-dimensional Liquid-phase Electrophoresis and Other LC-coupled MS Approaches

Despite the seemingly endless options, two promising protein digestion strategies should be considered. In one, intact proteins are first partially purified, followed by enzymatic digestion, and finally MS-based detection (with or without gas-phase fragmentation). The other "shotgun" method involves initial enzymatic digestion of all proteins, followed by varied LC separation techniques and direct introduction into to the MS (with or without gas-phase fragmentation).

Preparative two-dimensional liquid-phase electrophoresis (2-D LPE) and MALDI-TOF-MS was shown to be useful in the characterization of tryptic digests of low-abundance proteins in CSF [53]. 2-D LPE is based on the same IEF and gel electrophoresis principles as traditional 2-DE, except that the analytes remain in solution throughout the separation process. Analysis of intact proteins has also been performed using this strategy [54]. In another study, β-trace protein in CSF was isolated by HPLC and high-performance electrophoresis chromatography (HPEC) for MALDI analysis [55]. Another two-step electrophoretic procedure involving liquid phase IEF in conjunction with 1-DE gels and electroelution has been used for the purification and identification of proteins from human CSF [6, 56]. Several neuron-related proteins such as amyloid precursor-like protein, chromogranin A, chromogranin B, and glial fibrillary protein have been identified from human CSF using these approaches [6].

Alternatively, in the shotgun approach only a limited fractionation of a complex mixture of tryptic peptides is achieved by online LC, either one-dimensionally or two-dimensionally before MS analysis, and a significant amount of peptide "separation" is accomplished by the MS detector itself through a data-dependent acquisition algorithm. Employing this approach by coupling an optimized reverse phase micro-HPLC with nanospray interface to an ion trap mass spectrometer, we recently identified 299 different proteins, 200 of which had not been previously described by any reported 2-DE or LC/MS methods [57]. These molecules include ion channel components, blood/brain barrier components, cytoskeletal proteins, glycoproteins, a glycosyl transferase, oxidoreductases, vasoactive proteins, ion and electron transporters, ATP-binding proteins, lipid carrier/binding/synthesis and breakdown proteins, and circadian rhythm proteins. The known functions of these proteins provide working models of their involvement in the biochemical processes and clinical features of brain functions, and a rich framework to allow a better understanding of pathophysiology. More replication is needed to fully interpret the significance of these molecules. We have not included these proteins in Table 18.1, since these studies have not yet passed peer review. However, as soon as this is accomplished, we will add them to our website at www.hmri.org, under "Programs", under "Molecular Neurology".

Another promising work includes the optimization of capillary electrophoresis as well as reverse phase liquid chromatography coupled online to a Fourier transform ion cyclotron resonance mass spectrometry (FTICR-MS) of trypsin-digested CSF [58, 59]. Thus far, only a small number of unique proteins have been identified by this method, but its incorporation of an improved separation step that reduces ion suppression with the enhanced mass resolution of FT-MS offers great potential.

18.4 CSF Proteomic Applications in Central Nervous System Diseases

Several CSF proteins, discovered by 2-DE, have been proposed as diagnostic markers for neurodegenerative disorders; for example, the 14-3-3γ protein (originally p130 and p131) as a surrogate marker for Creutzfeldt-Jakob disease (CJD) [27, 60], and the middle isoform of α_2-haptoglobulin for Alzheimer's disease (AD) and schizophrenia [30]. In AD, many studies have found various assays for CSF tau and CSF β-amyloid to be very informative [61–63].

Since the 14-3-3γ protein assay is perhaps the first proteomic-derived test, some discussion of its application is worthwhile. Most CJD surveillance units incorporate this in their assessment of suspected cases of CJD in a dementia population. The original discovery of 14-3-3γ protein presented it as a surrogate marker to distinguish CJD in the context of clinical dementia, when we noted that a small number of false positive or negative testing would occur since it merely reflects neuronal damage from which 14-3-3γ leaks, apparently at a continuous rate of production mainly in the prion dementias. Occasional 14-3-3γ production will come from a number of other conditions, most of which can be clinically excluded. This seems to surprise some people [64], but if interpreted with appropriate caution, the overwhelming experience of several retrospective and prospective studies have validated the 14-3-3γ immunoassay for CJD [60, 65–71]. In this context, the American Academy of Neurology endorsed the use of this test in their guidelines for the work-up of dementia if CJD is suspected [72].

Only a few comparative proteomic analyses of CSF proteins have been employed for studies of pathophysiological mechanisms in neurodegenerative diseases. Using 2-DE gel and the silver staining technique, large differences in protein patterns in CSF between patients with brain disorders such as AD, schizophrenia, Parkinson's disease, and multiple sclerosis were shown [24, 25, 28, 73], however many of these proteins have not yet been identified. Fonteh and Harrington reported at the 2002 meeting in Siena an increase of specific isoforms of apolipoproteins A1, J, K and prostaglandin D synthase (PGDS) in the CSF of AD study participants [74]. Moreover, Harrington and colleagues have described significant changes in the abundance of 10 CSF proteins during the headache phase of migraine and more subtle changes in isoforms of two other proteins in the well-state of a migraineur population compared with controls who were not susceptible to headaches [75, 76]. In a proteomic study of CSF proteins in a group of AD

patients, the levels of six proteins and their isoforms, including pro-apolipoprotein, apolipoprotein E, β_2-microglobulin, retinol-binding protein, transthyretin, and ubiquitin were significantly altered. The most notable changes were seen among the apolipoproteins, especially pro-apolipoprotein [7], which is consistent with the previously reported results of reduced levels of apolipoproteins in AD [77].

Protein oxidation has long been discussed in the pathophysiology of AD and protein modifications to reactive oxygen species were demonstrated more frequently in AD, with α-enolase identified as one of the target proteins [20]. Aging is the highest risk factor for AD. Merril and Harrington demonstrated five proteins that were age-affected in CSF from normal controls [31], and a quantitative proteome analysis identified 35 proteins in relation to aging that may partly reflect some of the pathophysiological components involved in AD [78]. Normal aging itself is a complex event and the proteomic approach could be important in determining its molecular complexity.

One of the most common types of dementia is frontotemporal dementia (FTD), which is often misdiagnosed as AD. A recent study showed for the first time that several proteins involved in FTD pathology are not altered in the CSF of AD patients, and vice versa, thus establishing differences in pathophysiological mechanisms between FTD and AD, two of the most common neurodegenerative disorders [12].

18.5 Future Challenges

As mentioned, we are confident that application of the multiple new technologies will greatly improve our knowledge of the molecular details of CSF. The challenge is to employ the non-trivial experimental design concerns discussed above with judicial selection of the varied separation and detection tools. Understanding any CSF protein changes will require careful study of the source of each molecule and its dynamic fluctuations. Thus, time course experiments and analyses of blood, urine, and brain will be necessary. Multidisciplinary studies will always be the most informative. The coupling of psychological, neurological, psychiatric, physiological, neuroimaging, and molecular evaluations of the same person would be ideal, but obviously most difficult to realize. When future studies are impractical in humans, it will be necessary to employ animal models; at such a point, CSF extraction from mice and/or other convenient mammals will become helpful.

The most exciting advances and challenges are with the protein characterization methods. Instrumental improvements, in particular in the field of mass spectrometry, such as the development of high-capacity ion traps with faster scan speeds and higher sensitivity, will push the frontier in the detection of low-abundance molecules in CSF. The increased scan speed on linear ion trap MS detectors will provide greater opportunity for more tandem MS spectra that should lead to the reproducible identification of more peptides from complex mixtures in each ex-

periment. Hybrid detectors, such as the combination of a linear ion trap with an ICR detector (LTQ-FTMS) will give more accurate mass assignments and thus should provide better confidence in identifications. The challenge for most investigators will be that technical options are changing so rapidly (often with considerable cost per instrument) that no laboratory can utilize the full range of electrophoresis-, chromatography-, and mass spectrometry-based instruments.

Measures of protein abundance by MS methods are still in their infancy and have not been applied to CSF yet. Methods that quantify are essential to survey whether any proteins are up- or down-regulated. Relative quantification can be performed by integrating the area under the ion chromatograph and normalizing to a constitutive internal standard, similar to what has been achieved for integrating spot intensities by 2-D gel methods. Preliminary results of this relative quantification by LC/tandem MS are promising, with around 15–30% variation [79]. Absolute quantification can be achieved by using an isotopically labeled standard that is spiked in each sample and for which a standard curve can be generated. Our preliminary results with this approach on the TSQ Quantum^{TM} have great sensitivity as well as $<5\%$ variation (A. Fonteh, unpublished).

Protein chips are still mainly a dream. The best currently available is a multiple cytokine-based immunoassay format. It will be exciting if this field can realize the potential of generating massively multiplexed and quantitative specific protein assays.

2-DE gels are still very useful, yet they are a multistep process. Automation is necessary to improve the errors and variability of this procedure, yet there is only one "early" device that attempts to do this (Nextgen Sciences Ltd, UK).

Characterization of post-translational modifications (PTMs) is not currently feasible on the proteomic scale. The majority of the abundant CSF proteins are glycosylated and the literature cited above illustrates that there are PTM-specific changes of relevance to disease states. Thus it seems necessary to develop easier methods (if that is possible) to characterize the many PTM possibilities, and also PTM-specific immunoassays or other reagents to enable high-throughput assays or topological screening of PTM-specific isoforms. Recently, the combination of a rapid 2-D gel glycoprotein separation, glycoprotein visualization, in-gel digestion, and MS analysis in an FTIR instrument allowed identification of CSF glycoproteins and also revealed differences in glycosylation [80].

The most immediate and critical bottleneck is the data handling and data mining aspect of the new high-throughput approaches. For instance, in some of our experiments with nanobore LC/tandem MS, a single multidimensional LC/MS experiment can generate several thousand tandem MS spectra that represent potential peptide and protein identifications. Interpretation of the wealth of this information-rich fragmentation spectral data constitutes a substantial challenge and is virtually impossible without computer-guided automation. With replicates, it is not long before the manipulation of raw files, spreadsheets, and even storage/backup become limiting. Clearly, the new techniques to collect large proteomic data sets outstrip our ability to validate, interpret, and integrate such data for the purpose of creating biological knowledge. The instrument companies are

struggling to keep up with this non-traditional challenge and it remains to be seen whether the federally funded database organizations will solve some of these issues. Another bioinformatic issue is that better statistical determinations of what constitutes a unique protein identity from MS experiments are needed.

The challenges in understanding CSF proteins in their entirety are immense. We need to discover the identity and abundance of all of the CSF proteins, including their many post-translational modifications, as well as their age, gender, and chronobiological variations, and then study their alterations in disease states. However, the potential to make progress has never been as exciting. Maybe this is the beginning of the beginning for this field. We think so!

18.6 References

1 Kennedy, J.S., Gwirtsman, H.E., Schmidt, D.E. et al. *Life Sci.* **2002**, *71*, 1703–1715.
2 Merril, C.R., Goldman, D., Sedman, S.A., Ebert, M.H. *Science* **1981**, *211*, 1437–1438.
3 Evans, R.W., Armon, C., Frohman, E.M., Goodin, D.S. *Neurology* **2000**, *55*, 909–914.
4 Davidsson, P., Jahn, R., Bergquist, J., Ekman, R., Blennow, K. *Mol. Chem. Neuropathol.* **1996**, *27*, 195–210.
5 Davidsson, P., Puchades, M., Blennow, K. *Electrophoresis* **1999**, *20*, 431–437.
6 Davidsson, P., Paulson, L., Hesse, C., Blennow, K., Nilsson, C.L. *Proteomics* **2001**, *1*, 444–452.
7 Davidsson, P., Folkesson, S., Christiansson, M. et al. *Rapid Commun. Mass Spectrom.* **2002**, *16*, 2083–2088.
8 Raymackers, J., Daniels, A., De, B. et al. *Electrophoresis* **2000**, *21*, 2266–2283.
9 Sickmann, A., Dormeyer, W., Wortelkamp, S., Woitalla, D., Kuhn, W., Meyer, H.E. *Electrophoresis* **2000**, *21*, 2721–2728.
10 Sickmann, A., Dormeyer, W., Wortelkamp, S., Woitalla, D., Kuhn, W., Meyer, H.E. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* **2002**, *771*, 167–196.
11 Yun, M., Wu, W., Hood, L., Harrington, M. *Electrophoresis* **1992**, *13*, 1002–1013.
12 Davidsson, P., Westman-Brinkmalm, A., Nilsson, C.L. et al. *Neuroreport* **2002**, *13*, 611–615.
13 Davidsson, P., Sjogren, M., Andreasen, N. et al. *Mol. Brain Res.* **2002**, *109*, 128–133.
14 Klose, J., Spielmann, H. *Biochem. Genet.* **1975**, *13*, 707–720.
15 O'Farrell, P.H. *J. Biol. Chem.* **1975**, *250*, 4007–4021.
16 Anderson, L., Anderson, N.G. *Proc. Natl Acad. Sci. USA* **1977**, *74*, 5421–5425.
17 Goldman, D., Merril, C.R., Ebert, M.H. *Clin. Chem.* **1980**, *26*, 1317–1322.
18 Tracy, R.P., Currie, R.M., Young, D.S. *Clin. Chem.* **1982**, *28*, 890–899.
19 Bracco, F., Gallo, P., Tavolato, B., Battistin, L. *Neurochem. Res.* **1985**, *10*, 1203–1219.
20 Castegna, A., Thongboonkerd, V., Klein, J.B., Lynn, B., Markesbery, W.R., Butterfield, D.A. *J. Neurochem.* **2003**, *85*, 1394–1401.
21 Choe, L.H., Dutt, M.J., Relkin, N., Lee, K.H. *Electrophoresis* **2002**, *23*, 2247–2251.
22 Dermer, G.B., Silverman, L.M., Chapman, J.F. *Clin. Chem.* **1982**, *28*, 759–765.
23 Endler, A.T., Young, D.S., Yanagihara, T., Currie, R.M., Reid, J., *J. Clin. Chem. Clin. Biochem.* **1987**, *25*, 61–70.
24 Harrington, M.G., Merril, C.R. *Clin. Chem.* **1984**, *30*, 1933–1937.
25 Harrington, M.G., Merril, C.R., Torrey, E.F. *Clin. Chem.* **1985**, *31*, 722–726.
26 Harrington, M.G., Merril, C.R. *Psychopharmacol. Bull.* **1985**, *21*, 361–364.

27 Harrington, M.G., Merril, C.R., Asher, D.M., Gajdusek, D.C. *N. Engl. J. Med.* **1986**, *315*, 279–283.

28 Harrington, M.G., Merril, C.R. *J. Chromatogr.* **1988**, *429*, 345–358.

29 Hochstrasser, D.F., Frutiger, S., Wilkins, M.R., Hughes, G., Sanchez, J.C. *FEBS Lett.* **1997**, *416*, 161–163.

30 Johnson, G., Brane, D., Block, W., Van Kammen, D.P. et al. *Appl. Theor. Electrophor.* **1992**, *3*, 47–53.

31 Merril, C.R., Harrington, M.G. *Clin. Chem.* **1984**, *30*, 1938–1942.

32 Merril, C.R., Harrington, M.G. *Schizophr. Bull.* **1988**, *14*, 249–254.

33 Walsh, M.J., Tourtellotte, W.W., Roman, J., Dreyer, W. *Clin. Immunol. Immunopathol.* **1985**, *35*, 313–327.

34 Wiederkehr, F., Ogilvie, A., Vonderschmitt, D.J. *J. Neurochem.* **1987**, *49*, 363–372.

35 Wiederkehr, F., Bueler, M.R., Wacker, M., Vonderschmitt, D.J. *Electrophoresis* **1989**, *10*, 480–488.

36 Harrington, M.G., Aebersold, R., Martin, B.M., Merril, C.R., Hood, L. *Appl. Theor. Electrophor.* **1993**, *3*, 229–234.

37 Harrington, M.G., Biringer, R., Huhmer, A.F. et al. *Methods for the study of cerebrospinal fluid composition. Proceedings of the 5th Siena Conference on Functional Proteomics.* **2002**, p 240.

38 Burkhard, P.R., Sanchez, J.C., Landis, T., Hochstrasser, D.F. Neurology **2001**, *56*, 1528–1533.

39 Harrington, M.G., Swedberg, S., Biringer, R. et al. P*roteome Analysis of Cerebrospinal Fluid from Headache and Non-headache Cohorts. American Headache Society,* **2002a.**

40 Hoogland, C., Sanchez, J.C., Tonella, L. et al. *Nucleic Acids Res.* **2000**, *28*, 286–288.

41 Unlu, M., De Lange, R.P., De Silva, R., Kalaria, R., St Clair, D. *Neurosci. Lett.* **2000**, *282, 149–152.*

42 Yuan, X., Russell, T., Wood, G., Desiderio, D.M. *Electrophoresis* **2002**, *23*, 1185–1196.

43 Schulz-Knappe, P., Zucht, H.D., Jurgens, M., Hess, R., Schrader, M. *Comb. Chem. High Throughput. Screen.* **2001**, *4*, 207–217.

44 Stark, M., Danielsson, O., Griffiths, W.J., Jornvall, H., Johansson, J. *J. Chromatogr. B Biomed. Sci. Appl.* **2001**, *754*, 357–367.

45 Heine, G., Zucht, H.D., Schuhmann, M.U. et al. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* **2002**, *782*, 353–361.

46 Johnson, T., Bergquist, J., Ekman, R. et al. *Anal. Chem.* **2001**, *73*, 1670–1675.

47 Lewczuk, P., Esselmann, H., Meyer, M. et al. *Rapid Commun. Mass Spectrom.* **2003**, *17*, 1291–1296.

48 Davidsson, P., Ekman, R., Blennow, K. *J. Neural Transm.* **1997**, *104*, 711–720.

49 Sjogren, M., Davidsson, P., Tullberg, M. et al. *J. Neurol. Neurosurg. Psychiatry* **2001**, *70*, 624–630.

50 Westman-Brinkmalm, A., Davidsson, P. *Anal. Biochem.* **2002**, *301*, 161–167.

51 Corthals, G.L., Molloy, M.P., Herbert, B.R., Williams, K.L., Gooley, A.A. *Electrophoresis* **1997**, *18*, 317–323.

52 Zuo, X., Echan, L., Hembach, P. et al. *Electrophoresis* **2001**, *22*, 1603–1615.

53 Davidsson, P., Westman, A., Puchades, M., Nilsson, C.L., Blennow, K. *Anal. Chem.* **1999**, *71*, 642–647.

54 Puchades, M., Westman, A., Blennow, K., Davidsson, P. *Rapid Commun. Mass Spectrom.* **1999**, *13*, 2450–2455.

55 Hoffmann, A., Nimtz, M., Getzlaff, R., Conradt, H.S. *FEBS Lett.* **1995**, *359*, 164–168.

56 Davidsson, P., Nilsson, C.L. *Biochim. Biophys. Acta* **1999**, *1473*, 391–399.

57 Harrington, M.G., Fonteh, A.N., Brininger, R.G., Amato, H., Huhmer, A.H. *ICES* **2003**, *37*.

58 Bergquist, J., Palmblad, M., Wetterhall, M., Hakansson, P., Markides, K.E. *Mass Spectrom. Rev.* **2002**, *21*, 2–15.

59 Ramstrom, M., Palmblad, M., Markides, K.E., Hakansson, P., Bergquist, *J. Proteomics* **2003**, *3*, 184–190.

60 Hsich, G., Kenney, K., Gibbs, C.J., Lee, K.H., Harrington, M.G. *N. Engl. J. Med.* **1996**, *335*, 924–930.

61 Andreasen, N., Vanmechelen, E., Vanderstichele, H., Davidsson, P., Blennow, K. *Acta Neurol. Scand. Suppl.* **2003**, *179*, 47–51.

62 Andreasen, N., Hesse, C., Davidsson, P. et al. *Arch. Neurol.* **1999**, 56, 673–680.

63 Andreasen, N., Vanmechelen, E., Van de Voorde, A. et al. *J. Neurol. Neurosurg. Psychiatry* **1998**, *64*, 298–305.

64 Burkhard, P.R., Rodrigo, N., May, D. et al. Electrophoresis **2001**, *22*, 1826–1833.

65 Green, A.J. *Neuropathol. Appl. Neurobiol.* **2002**, *28*, 427–440.

66 Caboclo, L.O., Huang, N., Lepski, G.A. et al. *Arq. Neuropsiquiatr.* **2002**, *60*, 458–461.

67 Saiz, A., Nos, C., Yague, J., Dominguez, A., Graus, F., Munoz, P. *J. Neurol.* **2001**, *248*, 592–594.

68 Beaudry, P., Cohen, P., Brandel, J.P. et al. *Dement. Geriatr. Cogn Disord.* **1999**, *10*, 40–46.

69 Lee, K.H., Harrington, M.G. *Lancet* **1996**, *348*, 887.

70 Marzewski, D.J., Towfighi, J., Harrington, M.G., Merril, C.R., Brown, P. *Neurology* **1988**, *38*, 1131–1133.

71 Croxson, M., Brown, P., Synek, B. et al. *Neurology* **1988**, *38*, 1128–1130.

72 Knopman, D.S., Dekosky, S.T., Cummings, J.L. et al. *Neurology* **2001**, *56*, 1143–1153.

73 Harrington, M.G., Kennedy, P.G. *Postgrad. Med. J.* **1987**, *63*, 735–740.

74 Fonteh, A.N., Cowan, R., Harrington, M.G. *Proceedings of the 5th Siena Conference on Functional Proteomics* **2002**, p 91.

75 Harrington, M.G., Swedberg, S., Biringer, R. et al. *Proteome Analysis of Cerebrospinal Fluid from Headache and Non-headache Cohorts. American Headache Society,* **2002**.

76 Harrington, M.G., Stochaj, W., Cowan, R. *Cephalalgia* **2001**, *21*, 343.

77 Dietschy, J.M., Turley, S.D. *Curr. Opin. Lipidol.* **2001**, *12*, 105–112.

78 Tsuji, T., Shiozaki, A., Kohno, R., Yoshizato, K., Shimohama, S. *Neurochem. Res.* **2002**, *27*, 1245–1253.

79 Huhmer, A.H., Biringer, R., Amato, H., Harrington, M.G. *J. Biomol. Technol.* **2003**, *14*, 77.3.

80 Hakansson, K., Emmet, M.R., Marshall, A.G., Davidsson, P., Nilsson, C. *J. Proteome Res.* **2003**, in press.

19
Proteomic Applications for Molecular Assessment of Alzheimer's Disease

Odile Carrette, Pierre R. Burkhard, Denis F. Hochstrasser, and Jean-Charles Sanchez

19.1
Introduction

Dementia is characterized by a progressive loss of function in multiple cognitive domains. The most commonly used set of criteria for the diagnosis of dementia is provided by the DSM-IV (Diagnostic and Statistical Manual for Mental Disorders, American Psychiatric Association). Diagnostic features include memory impairment and at least one of the following: aphasia, apraxia, agnosia, and disturbances in executive functions. In addition, cognitive impairment must be severe enough to interfere with social and occupational functioning. Importantly, it must represent a decline from a previously higher level of functioning. Finally, the diagnosis of dementia should not be made when the cognitive deficits occur during the course of a delirium.

Dementia is a common disorder, involving about 10% of the general population over the age of 65, and possibly up to 50% of persons over the age of 85. More specifically, the prevalence of dementia doubles with every 5-year increase in age, and consequently constitutes a major health problem worldwide.

From the many causes of dementia, the most common include [1]:

- Degenerative dementia, such as Alzheimer's disease (AD), Lewy body dementia (LBD), frontotemporal dementia (FTD), Huntington's disease, and Binswanger disease.
- Vascular dementia (VD), such as multiple-infarct dementia.
- Secondary dementia such as normal pressure hydrocephalus, vitamin B12 deficiency, Lyme disease and other infectious conditions.

However, AD is by far the most frequent etiology of dementia, affecting between 50 and 70% of all people with dementia. Over four million people are currently suffering from AD in Europe.

In a recent article, Knopman et al. [2] reviewed the different criteria currently used for the diagnosis of dementia and their reliability. They compared 13 studies in which clinical assessment was performed using both the Diagnostic and Statistical Manual 3rd edition (DSM-IIIR) and the National Institute of Neurologic,

Biomedical Application of Proteomics
Edited by J.-C. Sanchez, G. L. Corthals, D. F. Hochstrasser

ISBN: 3-527-30807-5

Communicative Disorders and Stroke-AD and Related Disorders Association (NINCDS-ADRDA) definitions. This meta-analysis showed that the clinical diagnosis of probable AD could be established with an average sensitivity of 81% and an average specificity of 70%. Indeed, most of the clinical features of AD are shared by other forms of dementia, which can confuse the clinical diagnosis, particularly in the early stage of disease.

There is a need for complementary diagnostic tools of high sensitivity and specificity in addition to neuropsychological assessment and neuroimaging techniques in order to discriminate between normal aging, mild cognitive impairment, and the various dementia conditions at early stage. A consensus report of the working group on 'Molecular and Biochemical Markers of Alzheimer's disease' have suggested that an ideal biomarker for AD should have at least a sensitivity and a specificity of more than 80% [3].

Proteomics aims to detect proteins that are associated with diseases by the determination of their altered levels of expression between control and disease states. Proteome research may allow new protein markers to be found for diagnostic purposes and novel molecular targets for drug discovery. It is expected that proteomics will make a key contribution to the discovery of relevant protein markers of dementia with potential as diagnostic markers and novel therapeutic targets. Various technological approaches including two-dimensional gel electrophoresis, mass spectrometry, and protein chip array have been used to discover potential new biomarkers of dementia, especially AD. In the present chapter, we review the various proteomic studies that have been performed to date in the field of dementia research.

19.2 Two-dimensional Gel Electrophoresis Studies

Genomic sequence and mapping information from the Human Genome Project has facilitated the emergence of high-scale screening tools. Genomics uses DNA chips to determine which genes are turned on and turned off, and to uncover genes expressed in a particular disease. However, very few diseases are thought to have a simple genetic origin. Most physiological and pathological processes are related to quantitative variations in the amounts of genes products. Moreover, post-translation modifications of the expressed proteins can influence the protein function. Two-dimensional gel electrophoresis (2-DE) is an ideal tool for the resolution of complex protein mixtures. Hundreds to thousands of proteins can be visualized simultaneously. Proteins are separated according to their charge and their molecular weight, allowing resolution of multiple isoforms and variants of the same protein. When coupled to mass spectrometry, individual polypeptide components can be accurately identified. In addition, the quantitative differences between proteins in mixtures can be determined from 2-DE gel images and bioinformatic analysis. Altogether, these allow the direct detection of differentially expressed gene products and the discovery of new potential markers for diseases.

19.2.1 General Screening Studies

Measurements of change in the expression of multiple proteins provide a powerful strategy for investigating complex pathophysiological processes such as the one observed in neurodegenerative diseases. As early as 1982, Comings et al. [4] have investigated the major proteins present in brain extracts of AD patients. They reported a relative decrease in the amount of three unidentified proteins in one out of the six AD patients studied, a genetic variant of glial fibrillary acidic protein (GFAP) in two patients, a higher quantity of tubulin, and variable patterns of spots in three protein groups. 2-DE analyses have also shown an increase in the myelin basic protein [5], heat shock proteins [6], heavy chain gamma globulins [7], and GFAP [7, 8]. In the study performed by Mattila and Frey [9] on AD brain extracts, four protein spots have been detected in AD as changed versus controls. One spot was undetectable, two spots were significantly weaker and one spot was stronger than those in controls. By immunostaining, they observed an increasing number of GFAP proteolytic fragments in AD cases.

Improved resolution and reproducibility of 2-DE analysis allowed Tsuji et al. [10, 11] to detect more protein changes in AD brain. Comparing the density of spots between AD patients and normal controls, they found that five protein spots were significantly increased, 28 spots were significantly decreased and nine spots were detected only in AD. Two increased spots were identified as GFAP.

A comparative analysis of hippocampal tissue from schizophrenic, AD patients, and controls was performed by Edgar et al. [12]. In comparison with the control hippocampal proteome, eight proteins in the schizophrenic hippocampal proteome were found to be decreased and eight increased in concentration, whereas in the AD hippocampal proteome, 35 proteins were decreased and 73 were increased in concentration ($P<0.05$). One protein, which was decreased in concentration in both diseases, was characterized as diazepam-binding inhibitor (DBI) by N-terminal sequence analysis. Later the same group performed a larger study on six brain regions from individuals with AD and compared them with the proteome from control subjects without dementia [13]. In severely affected brain regions, 76 proteins were differentially expressed in AD hippocampus compared with normal controls, 62 proteins were differentially expressed in temporal cortex, and 39 proteins were differentially expressed in enthorinal cortex. Thirty-seven of these proteins have been identified, which are involved in synaptic transmission, stress response, lipid transport, glycolysis, and associated with the pathogenesis of diabetes mellitus. Alterations in all these functions have previously been implicated in AD.

Finally, in CSF samples from AD patients compared with controls, Choe et al. [14] have also found nine potential markers that are not identified yet. However, according to these authors, a canonical correlation analysis using the panel of the nine markers allowed the segregation of disease cases from normal controls.

More recently, using mini 2-DE technology stained with SYPRO ruby, another group [15] has identified 10 proteins significantly altered in the CSF of AD pa-

tients. Three proteins were increased: transthyretin, retinol-binding protein and β_2-microglobulin. The seven proteins found decreased in AD were isoforms of proapolipoprotein and apolipoprotein E. Indeed, a genetic risk factor has been associated with AD, based on the variants of *apoE*, the gene that encodes apolipoprotein E, a constituent of low-density apolipoproteins. Three variants of the gene and the protein (referred to as ApoE ε2, ApoE ε3, and ApoE ε4) are found in humans and result from the change of a single amino acid in apolipoprotein E. Carrying one ApoE ε4 allele nearly doubles the lifetime risk of AD, whereas not carrying an ApoE ε4 allele reduces the risk by 40%.

In the brain of patients with AD, ApoE is deposited with β-amyloid protein in the form of senile plaques [16], probably because of its high avidity for the peptide. The three isoforms have different isoelectric points, so that they can be separated and characterized by 2-DE [17]. Moreover, 2-DE analysis of the heparin-Mn^{2+}-precipitable lipoprotein fraction in CSF showed that the ratio between the level of CSF ApoA-I and of CSF ApoE of controls was significantly higher than those of all AD patients ($P<0.01$) [18].

Plasma proteins from patients with AD were also resolved by 2-DE and compared with controls and patients with non-AD dementia [19]. This study showed that AD-related proteins, such as ApoE, tau, and presenilin 2 are detectable in small amounts in blood samples. However, in this study no quantitation was performed.

In these numerous screening studies performed with brain extracts or body fluids, several potential new markers were detected, although many of them remain to be identified. Besides, 2-DE has often been used to study specific group of proteins. In this chapter we report the findings of 2-DE gel studies oriented on the analysis of tau protein, β-amyloid peptides and oxidative stress proteins, which are known to be mainly involved in the molecular mechanisms underlying neurodegeneration.

19.2.2 Tau Protein

Tau protein is a microtubule-associated protein found in high concentration in the axon of neurons and glial cells of the central nervous system. It is a member of a protein family that promotes assembly and stabilization of microtubules. Tau protein exists as many heterogeneous isoforms derived from both differential splicing of tau-mRNA and post-translational modification of the protein. Each of the isoforms differs from the others in the size of the N-terminal inserts and the presence of 3 or 4 tandem repeat regions of 31–32 amino acids at the C-terminal end. In addition, tau isoforms can be phosphorylated at a number of different sites.

In several tauopathies, namely AD, Pick's disease, and FTD, tau is the principal component of insoluble proteinaceous deposits such as neurofibrillary tangles (NFTs). In particular, in AD, hyperphosphorylated tau is converted into paired helical filaments (PHF) forming the NFT.

A number of studies have evaluated measures of CSF tau as an ante-mortem marker for dementia. Tau is measured in the CSF using enzyme-linked immuno-

assay (ELISA), which has been developed using different antibodies. In most of these studies, total tau was measured, although assays for phosphorylated tau alone have been used more recently. Therefore, different sets of sensitivity and specificity have been reported from different sources. Sunderland et al. [20] recently performed a meta-analysis of 34 studies on CSF tau. They reported that all the studies showed a significant difference in CSF tau levels between AD patients and controls. When levels in normal individuals were compared with those in AD patients, the reported sensitivity of CSF tau for AD detection was between 60% and 95%, while the specificity varied from 56–90% [21–24].

Nevertheless, high CSF levels of tau are not specific for AD, and have also been found in vascular dementia [25], FTD [26], diffuse LBD [27], and Creutzfeldt-Jakob disease [28]. This lack of specificity may restrict the usefulness of CSF tau protein in distinguishing normal aging and depression from true dementia.

The use of proteomic techniques and the analysis of the tau proteins have shown that the tau profile is disease-specific. In AD, pathological tau proteins are composed of four main isoforms (tau 74, 69, 64, and 55 kDa), while in Pick's disease, mainly tau 64, 55, and a minor tau 69 kDa isoforms are found [29]. Moreover, the PHF tau proteins differ both qualitatively and quantitatively in their degree of phosphorylation when compared with native tau protein. For example, in AD, the 74 kDa isoforms corresponding to the longest human brain isoform in a hyperphosphorylated state was detected particularly in the younger and most severely affected patients [30]. Specific sets of tau could be used to distinguish between typical neuronal inclusions [31]. Janke et al. [32, 33] showed by 2-DE electrophoresis that the distribution of tau isoforms differs in white and gray matter. However, the authors did not observe significantly altered tau isoforms in AD compared with controls.

To increase the specificity of CSF tau measurement, an assay to detect hyperphosphorylated tau or specific isoforms should be developed. Some studies have shown promising results in distinguishing AD from non-AD dementia by measuring specific phosphorylated tau isoforms. This was first reported by Blennow et al. [34], who obtained a sensitivity of 88% and a specificity of 74%, when comparing AD versus depression, VD, Parkinson's disease (PD), frontal lobe dementia, and normal aging. Since then, other groups have replicated these results (review in [35]). In particular, the measurement of tau at phosphoserine 199 may help to distinguish AD from non-AD dementia with a sensitivity and a specificity of 85% [36].

19.2.3 Amyloid Precursor Protein

The amyloid precursor protein (APP) is a ubiquitously expressed transmembrane glycoprotein. Different pathways of processing APP can occur involving at least three types of proteolytic enzymes: α, β, and γ secretase. APP may be cleaved in the middle of the β-amyloid (Aβ) peptide sequence by α-secretase to generate non-amyloidogenic 100-kDa N-terminal fragment, called soluble APPα (sAPPα). The

larger sAPPα containing the entire sequence of the Aβ are neurotoxic, as demonstrated in cell or in animal models [37]. Alternatively, APP may be cleaved by the sequential actions of β and γ secretase to generate Aβ peptides, the size of which is governed by the exact site of γ secretase cleavage. There are two major Aβ peptides generated by this way: one which is 40 amino acids long, called Aβ 40, and one slightly longer with 42 amino acids, Aβ 42. Amyloid beta 40 (Aβ 40) and 42 (Aβ 42) are major components of senile plaques. Polymerization of β-amyloid and subsequent neuronal deposit leads to the degeneration of neurons involved in memory and cognition. Moreover, mutations in the *APP* gene also cause some forms of familial AD by releasing increased amounts of β-amyloid.

CSF Aβ 42 concentrations can be measured using a number of different ELISA that use either monoclonal or polyclonal antibodies. The CSF concentrations of Aβ peptides, particularly Aβ 42, are decreased in AD and it has been suggested that this may be due to the preferential deposition of Aβ peptides into senile plaques. The reduction of Aβ 42 peptide has been investigated as a potential diagnostic marker for AD in many studies [20, 38]. However, there is a considerable overlap with healthy controls [39]. Measurements of the ratio of Aβ 40 to Aβ 42 [40] or measures of Aβ 42 in combination with tau [41–43] results in higher sensitivity and specificity to diagnose probable AD than either test alone. Indeed, assays of the CSF Aβ ratio (Aβ 40/Aβ 42) showed a diagnostic sensitivity of 59% and a specificity of 88%, as compared with non-AD type dementia and controls [40], while combination of assays of the Aβ ratio and tau gave a sensitivity of 85% and a specificity of 81–91% [41].

The proteolytic fragments derived from APP in primary culture of rat hippocampal neurons have been analyzed by 2-DE [44]. Secretory neuronal APP is resolved into at least six spots with a molecular weight between 100 and 110 kDa and a pI ranging from 4.0 to 4.5. The wide isoelectric focusing range is mainly due to different amount of sialyl residues in secreted APP. No secreted APP could be detected in the medium of cultured neurons transfected with a clinical mutant associated with AD, APP670/671. This suggested that there is little α-secretase cleavage in neurons expressing the APP670/671. Sergeant et al. [45] have investigated and characterized the Aβ peptides involved in the first steps of the amyloid deposition by 2-DE. The aggregated species of Aβ were directly extracted from the brain tissue using formic acid and resolved on a 2-DE gel. Ten different spots detected on the immunoblot probed with specific anti-Aβ antibodies were also clearly stained by Coomassie blue and analyzed by mass spectrometry. Several spots corresponded to amino-truncated Aβ peptides and post-translationally modified variants of Aβ. The amino-truncated species accounted for 60% of all Aβ species found in these samples. However, it is possible that unpredictable secondary proteolytic processes may generate additional Aβ peptide fragments in the post-mortem tissues used in this study.

19.2.4 Oxidative Stress and Antioxidant Response

There is substantial evidence that oxidative damage that occurs in the aging brain of the normal individual is enhanced in AD brains. Recently, defects of mitochondrial electron transport chain enzymes were found repeatedly in various cells and tissues of patients with neurodegenerative diseases. Indeed, Kim et al. [46] observed reduced levels of complex III core protein 1 in AD (temporal cortex) and of complex V β-chain in Down's syndrome (DS) (frontal cortex). A consequence of defective mitochondrial energy production is the increased generation of free radicals (such as superoxide and hydroxyl radicals), which are normally produced as by-products of oxidative metabolism. Mitochondria are in turn damaged by reactive oxygen species.

Oxidative stress by mitochondrial damage has been implicated in the degeneration of both synapses and neurons in neurodegenerative diseases. Therefore, identification of oxidatively altered proteins in AD is important in understanding the relationship between protein oxidation, protein aggregation, and neurodegeneration. For this purpose, Korolainen et al. [47] have investigated the oxidation of proteins present in the AD brain. After derivatization with 2,4-dinitrophenylhydrazine (DNPH), the brain extracts were subjected to 2-DE followed by blotting onto PVDF membrane. The blots obtained were first stained with SYPRO ruby and then probed with an anti-dinitrophenyl antibody, revealing the oxidized proteins. This kind of western blot is also called oxyblot. The authors showed that six protein-bound carbonyls were decreased in AD versus controls, and one protein was increased in AD. However, no identification was achieved in this study.

Using the same technique, Castegna et al. [48, 49] identified three specifically oxidized proteins in AD brain: creatine kinase BB, glutamine synthase, and ubiquitin C-terminal hydrolase L-1. Three other proteins exhibited increase in specific oxidation in AD brain: dihydropyrimidinase-related protein 2, α-enolase, and the heat shock cognate 71. A similar study performed on plasma samples revealed that fibrinogen γ-chain precursor protein and α_1-antitrypsin precursor are more intensively oxidized in AD plasma samples than in controls [50]. These proteins exhibited a 2- to 6-fold higher oxidation index in plasma from AD subjects. Both proteins have been suggested to be involved in inflammation processes in AD.

Krapfenbauer et al. [51] investigated the expression of peroxiredoxins (Prxs), a family of six highly conserved antioxidant enzymes expressed in the frontal cortex and cerebellum of patients with AD, PD, and DS. Their results show that Prx II is significantly increased in DS, AD, and PD and Prx III is decreased in DS and PD. However, Prx VI is only increased in PD. They conclude that up-regulation of Prx VI could be used to discriminate PD from AD as well as DS.

In conclusion, identification, quantitation, and qualitative analysis of a large number of proteins involved in the pathogenesis of dementia have been achieved using 2-DE coupled with mass spectrometry. However, this method is laborious and cannot resolve proteins with extreme molecular weight, hydrophobicity, and isoelectric points. Therefore, others proteomic techniques, such as surface-en-

hanced laser desorption/ionization time-of-flight (SELDI-TOF) MS have been developed.

19.3 SELDI

Surface-enhanced laser desorption/ionization (SELDI) mass spectrometry is a novel proteomic approach for biomarker discovery [52]. It combines two powerful techniques: chromatography and mass spectrometry. Proteins are captured by adsorption, partition, electrostatic interaction, or affinity chromatography on a solid-phase proteinChip surface. The retained proteins are subsequently ionized and detected by time-of-flight (TOF) mass spectrometry. Although SELDI provides a unique preparation platform, it is similar to MALDI-MS in that a laser ionizes samples that have been co-crystallized with a matrix on a target surface. Unlike MALDI target surfaces, the SELDI patented ProteinChip arrays are designed to capture individual proteins or group of proteins with common biochemical properties such as pI or hydrophobicity. After adding a matrix solution, proteins can be ionized with a nitrogen laser and their molecular masses measured by TOF MS. SELDI can provide a rapid protein expression profile from a variety of complex biological samples such as plasma, urine, cell lysates, or CSF. This system is most effective at profiling low molecular weight proteins (i.e. <20 kDa), providing a complementary visualization technique to 2-DE. Moreover, SELDI is more sensitive and requires smaller amounts of sample than 2-DE. Finally, the software used to analyze the data generated by the SELDI provides tools to conduct cluster analysis to identify significant differences in relative protein abundances across many samples.

The currently available proteinChip array surfaces are based on either chromatographic-based chemistry (ion exchange, hydrophobic, immobilized metal affinity, etc.) that will bind large classes of proteins or biologically defined surfaces that can be used to investigate specific protein-interaction events. In the latter, the surfaces of proteinChip arrays are covalently modified with bait molecules, such as antibodies, to capture proteins that have an affinity to the bait. This technique was used to study all the variants of β-amyloid peptides produced in transfected cells using antibody specific for the Aβ peptides such as the 4G8, the 6E10, or the anti-NTA4 antibodies.

19.3.1 Amyloid Beta Peptide Analyzed by SELDI

Amyloid beta SELDI immunoassay was first developed to analyze the different Aβ peptides found in cell culture medium. Increased levels of Aβ fragments were found in transfected HEK293 cells incubated with cholesterol, indicating an increase of the activity of both β-secretase and γ-secretase [53, 54]. Beher et al. [55] have applied the same technique in combination with the use of specific inhibi-

tors of APP processing to study the mechanism of α- and β-secretase in cell culture models. Their results demonstrate that inhibition of γ-secretase blocks the generation of Aβ 40, Aβ 42 and all truncated species except Aβ 15 and Aβ 16, which are derived by α-secretase cleavage. This could indicate that a specific inhibitor of γ-secretase would have the potential to prevent the formation of amyloid deposits.

SELDI immunoassay of Aβ in brain tissues and sera has also been performed in a mouse model of Aβ peptide immunization [56]. As expected, the Aβ vaccination resulted in a decrease of Aβ deposits in the brain, an increase of soluble Aβ 40 in the CSF, and a decrease of Aβ levels in serum.

More recently, the pattern of Aβ peptides found in human CSF and in brain homogenates of AD patients have been investigated using the same method [57]. These data suggest the presence of a novel Aβ peptide, elongated at the C-terminus, corresponding to Aβ 1–45 or Aβ 2–46 according to its molecular mass. This peptide was observed only in AD patients. Moreover, the intensity of the signal corresponding to Aβ 42 was decreased in the CSF of AD patients when compared with non-demented controls, consistent with the results obtained by ELISA quantitation.

To the best of our knowledge, these few studies are the only ones reported in the literature using the SELDI technology to study neurodegenerative diseases. However, screening on proteinChip MS has already been used by several groups to detect potential novel biomarkers of prostate [58], bladder [59], or breast cancer [60] in serum, seminal plasma, nipple fluid, urine, or cell extracts. In our laboratory, SELDI has been used to profile the differences and similarities in proteins present in the CSF of patients diagnosed with AD, FTD, VD, and LBD compared with controls.

19.3.2
Screening of AD with SELDI on a Strong Anionic Exchange Surface [61]

A first pilot study was performed on nine AD samples compared with 10 controls using a strong anionic exchange (SAX2) proteinChip array. This surface specifically binds negatively charged proteins through interaction with aspartic acid and glutamic acid residues. Multiple protein changes were consistently and reliably found in the CSF of AD patients when compared with controls, including five differentially expressed polypeptides ($P<0.05$). The average SELDI mass associated with the five specific peaks was 4.82 kDa, 7.69 kDa, 11.78 kDa, 11.98 kDa, and 13.41 kDa.

In order to identify the proteins corresponding to these peaks, a fractionation of crude CSF on an SAX spin column was performed. The eluted fractions were analyzed by SELDI-TOF-MS on a SAX2 proteinChip array as shown in Figure 19.1. The differentially expressed peak of 13.4 kDa was eluted with buffer A (20 mM Tris-HCl and 5 mM NaCl pH 8.0) and B (20 mM sodium phosphate pH 7.0). The amount of proteins in these fractions was too low to see the peak by SELDI analysis. The differentially expressed peaks of 11.78 kDa and 11.98 kDa were found in the fractions eluted with buffer C (20 mM sodium phosphate pH 6.0) and D

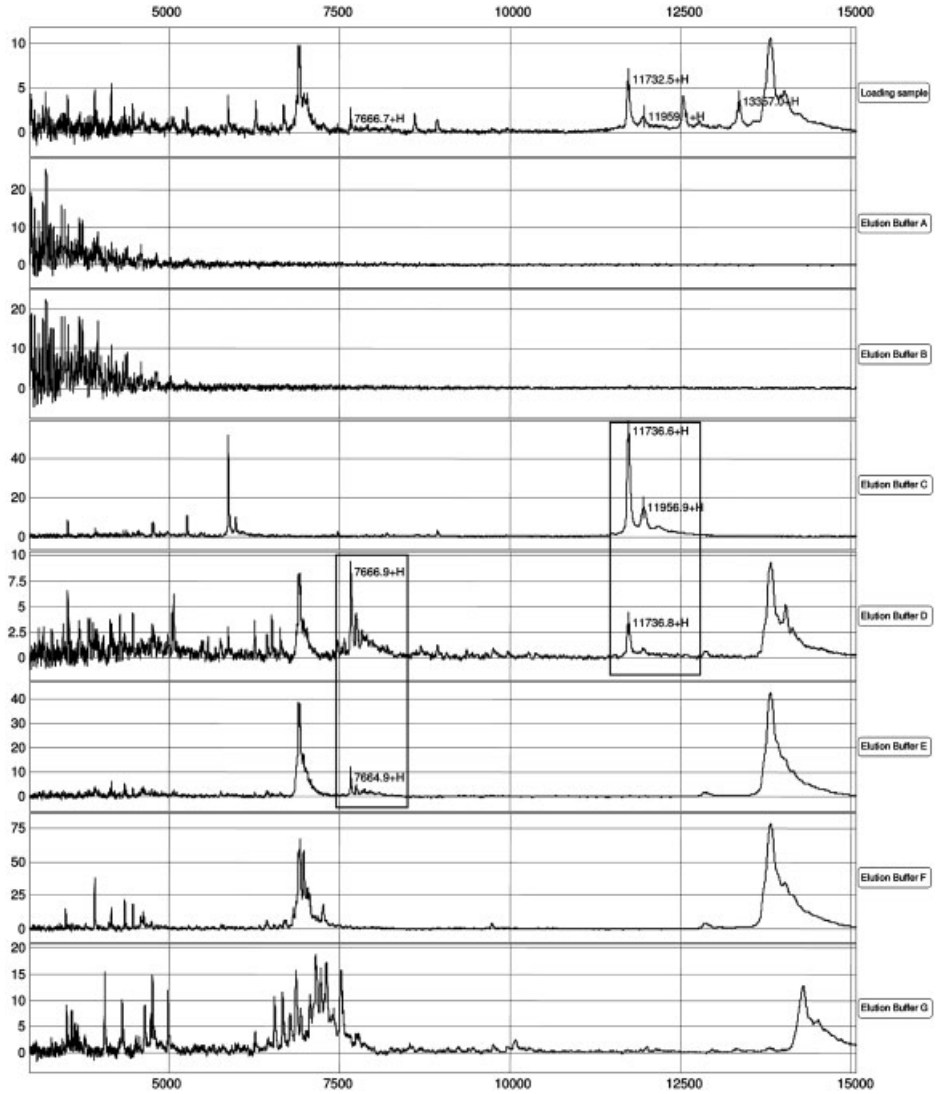


Figure 19.1 SELDI-TOF-MS protein profiles of CSF samples from healthy patients after SAX spin column fractionation analyzed on a SAX2 proteinChip array. The 7.69-kDa peak is eluted in fractions D and E. The 11.78-kDa and 11.98-kDa peaks were eluted in fractions C and D.

(20 mM sodium phosphate and citrate pH 5.0). The peak of 7.69 kDa was eluted with buffer D (20 mM sodium phosphate and citrate pH 5.0) and E (20 mM sodium phosphate and citrate pH 4.0). Each eluted fraction was loaded on a 16.5% Tris tricine SDS-PAGE. After staining with Coomassie blue, the bands of interest were cut out, digested with trypsin and analyzed by MS.

The 7.69 kDa peak tandem mass analysis did not match with any known human protein. Peptide sequences were the following XXAD(L/I)AGHG(Q/K)EV(L/I)(L/I)R and HGTVV(L/I)TA(L/I)GG(L/I)(L/I)K. The MS analysis of the 11.78 kDa and 11.98 kDa peaks identified both of them as β_2-microglobulin. The 13.4 kDa peak was identified by peptide mass fingerprint as cystatin C. Since, very small polypeptides (<6 kDa) tend to diffuse out of the polyacrylamide gels or are not stained with classical techniques, it is necessary to purify them in a liquid fraction before MS identification. Therefore, in order to identify the 4.82 kDa peak, the CSF was pre-purified with Centricon 30 and Centricon 3. This allowed the partial purification and enrichment of the 4.82 kDa peak mainly present in one fraction. The liquid fraction was reduced and alkylated, digested with trypsin and analyzed with liquid chromatography-quadrupole-TOF (LC-Q-TOF). One peptide of 2161 Da was sequenced as VGEEDEEAAEAEAEAEEAER. Mascot search revealed that this sequence corresponds to a protein present in TrEMBL with the accession number Q9UDW8, named WUGSC:H_DJ0747G18.3 protein with a score of 65. A BLAST search of this sequence revealed 100% identity with the neurosecretory protein VGF precursor (VGF8a protein) in rat; and 90% identity with the neurosecretory protein VGF precursor (VGF) in humans.

The relevance of cystatin C, β_2-microglobulin, and VGF peptide to AD was assessed by a literature review. Cystatin C (also known as γ-trace or post-β globulin) is a small cysteine proteinase inhibitor present in most human body fluids at physiologically relevant concentrations. It has been shown that cystatin C colocalizes with β-amyloid (Aβ) within the arteriolar walls in AD brains and cerebral amyloid angiopathy [62]. Moreover, hereditary cerebral hemorrhage with amyloidosis, Icelandic type (HCHWA-I), also called hereditary cystatin C amyloid angiopathy (HCCAA), is coupled to a decreased concentration of this major cysteine proteinase inhibitor in CSF and leads to its deposition in the brain as amyloid [63]. Finally, and in accordance with our findings, it has been recently reported that cystatin C is overexpressed in the neurons which are particularly susceptible to the AD degenerative processes, including the entorhinal cortex, hippocampus, and temporal cortex [64].

Beta-2-microglobulin constitutes the small constant component of the class I major histocompatibility complex (MHC) and its presence in biological fluids represents the balance between membrane protein turnover and elimination. This peptide is increased in some diseases characterized by an elevation of the immune response. A proteome analysis of CSF by 2-DE has recently shown a significant increase of β_2-microglobulin in AD patients, confirming our data [15]. Finally, in agreement with our data, Sugaya et al. [65] have also shown that the level of VGF mRNA is decreased in aged-impaired rats, suggesting down-regulation of the VGF peptides with aging.

In conclusion, this first screening on SAX2 chips provided us with proof that this new technology has potential to reveal new potential biomarkers of AD. We applied the same technique on another chip surface, a weak cationic exchanger (WCX2) to screen for potential biomarkers of AD compared with FTD, LBD, and VD.

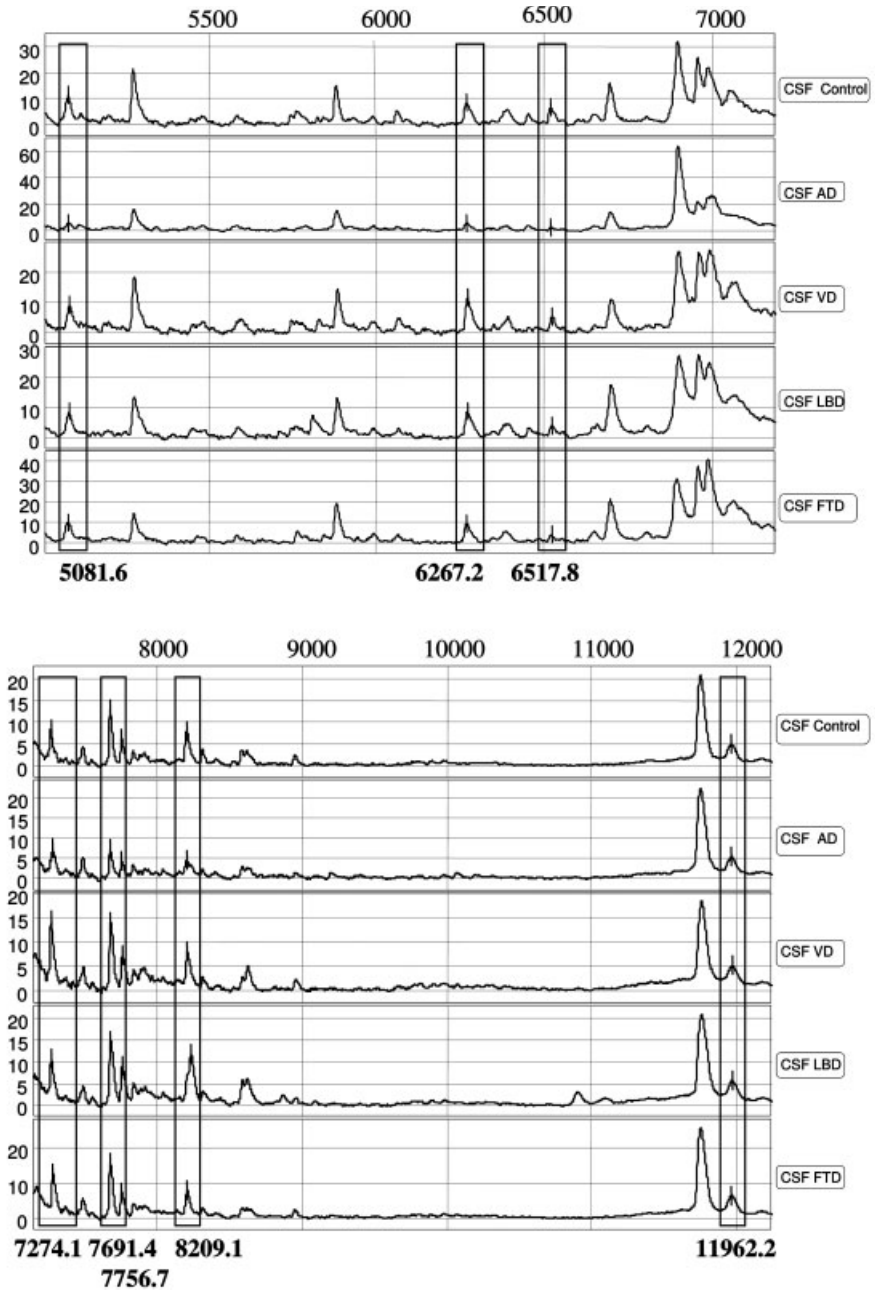


Figure 19.2 Representative SELDI-TOF-MS protein profile of CSF samples from one AD patient, one control, one VD, one LBD, and one FTD patient after capture on a WCX2 chip surface. The clusters of interest are boxed.

19.3.3
Screening of AD, FTD, LBD, and VD with SELDI on a Weak Cationic Exchange (WCX2) Surface

The WCX2 proteinChip array specifically binds positively charged proteins through interaction with the lysine, arginine, or histidine residues. Ten AD patients, 10 controls, 5 FTD, 5 LBD, and 5 VD patients were included in this study. The representative protein profiles obtained on WCX2 chips are shown in Figure 19.2 as mass spectra. The potentially interesting clusters found differentially expressed between groups are outlined in boxes. The average SELDI mass associated with the five differentially expressed proteins was 5.08 kDa, 6.27 kDa, 6.52 kDa, 7.27 kDa, 7.69 kDa, 7.75 kDa, 8.21 kDa, and 11.96 kDa. Figure 19.3 shows the average intensities of each cluster in the five groups studied.

Several peaks were found differentially expressed between AD patients and controls: 5.08 kDa, 6.27 kDa, 6.52 kDa, 7.27 kDa, and 8.21 kDa ($P<0.05$). Interestingly, those peaks are also decreased in the other demented individuals although the differences are not statistically significant when compared with the controls. The 6.52 kDa peak is significantly decreased in all the demented individuals analyzed. The 7.69 kDa and 7.75 kDa peaks are increased in the three demented groups, with a significant increased in the LBD compared with controls and AD patients. Finally, the 11.96 kDa peak is significantly increased in the LBD group compared with the AD patients. The three latter are potential markers that could allow to distinguished AD from the other types of dementia.

These protein profiles confirm the difficulty in finding biomarkers specific to each kind of dementia. These preliminary data should be confirmed on a larger population. Peaks of interest will need to be identified in order to validate their potential value as diagnostic markers in dementia. Finally, other existing protein chip surfaces such as the hydrophobic surface should be used to screen the CSF samples.

19.4
Conclusions

Many studies have demonstrated that clinical assessment, using well-established criteria, can yield relatively high diagnostic accuracy rates to distinguish demented individuals. This is especially true for patients who have reached the moderate to severe stages of dementia. However, existing diagnostic criteria are less reliable for patients in the early stages of the disease. As neuroprotective treatments become available, identifying AD patients in these stages will be increasingly important. In the past decades, considerable efforts have been made to assess biomarkers implicated in the pathogenesis of AD (such as tau and the amyloid-β peptides). However, a review of the clinical studies using these markers revealed poorly reliable results and difficulties in distinguishing probable AD patients from other demented individuals. At this time, it appears unlikely that one single biomarker could be used as an ante-mortem diagnosis to distinguish persons with AD from those with non-AD dementia.

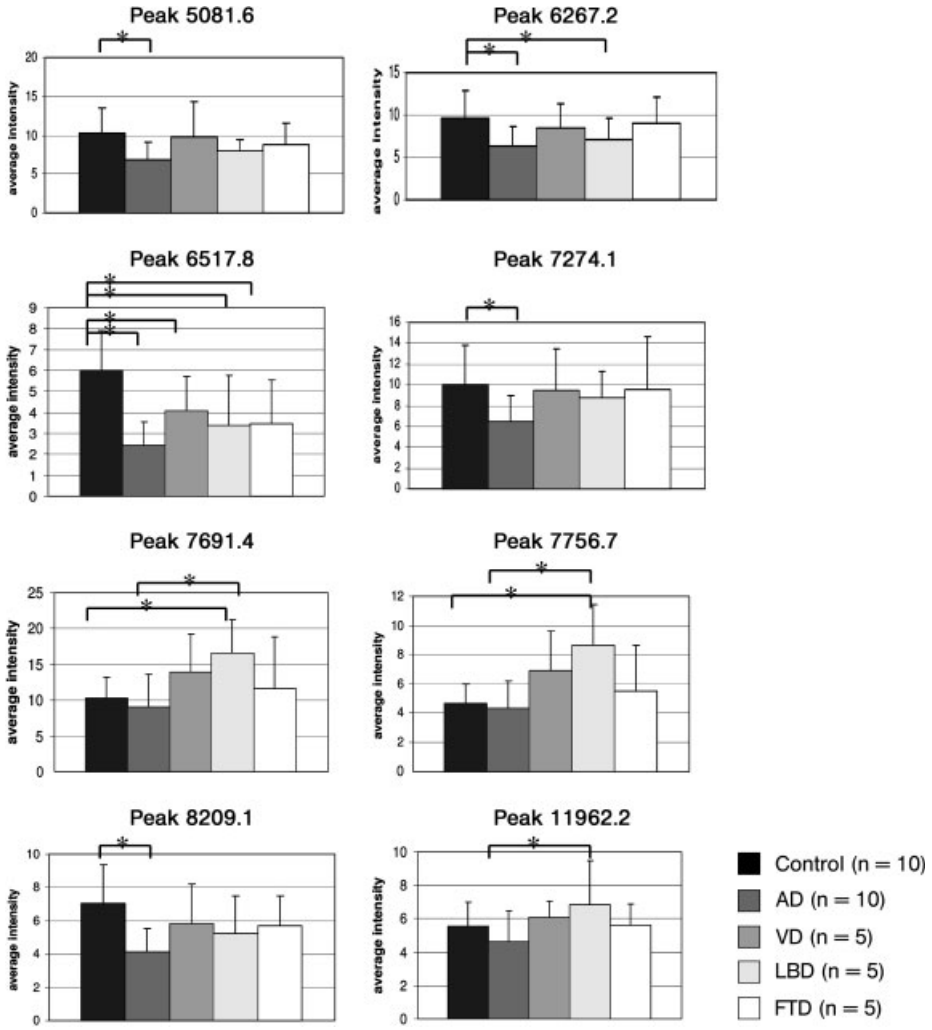


Figure 19.3 Average intensities of the clusters of interest outlined in Figure 19.2. *$P<0.05$ between two groups.

A proteomic approach allows parallel analysis of hundreds of proteins in one single experiment. Therefore current advances in the proteomics field have been applied to screen for new biomarkers specific for AD. In this chapter, we have summarized the reported studies using 2-DE and protein chip array technology that have been performed to date. Many proteins have been shown to be differentially expressed between the disease patients and the controls. Currently each individual protein has to be identified and differential expression confirmed by classical immunological tests. However, the variety of these potential markers confirms the need to implement new approaches in the routine analysis laboratory. Indeed,

automated proteomics profiling techniques combined with elaborate bioinformatic tools will be well suited to discriminate disease from control by considering several markers in a single analysis step.

19.5 Acknowledgements

We are grateful to Isabelle Demalte and Alexander Scherl for providing technical support. Much of the work reported in this chapter could only be achieved thanks to the outstanding collaboration with Peter Gautschi and Oezkan Yalkinoglu. The authors acknowledge the support of Bayer AG.

19.6 References

1 B.A. Vicioso. *Am. J. Med. Sci.* **2002**, *324*, 84–95.
2 D.S. Knopman, S.T. DeKosky, J.L. Cummings et al. *Neurology* **2001**, *56*, 1143–1153.
3 Anon. *Neurobiol. Aging* **1998**, *19*, 109–116.
4 D.E. Comings. *Clin. Chem.* **1982**, *28*, 798–804.
5 D.J. Selkoe, B.A. Brown, F.J. Salazar, C.A. Marotta. *Ann. Neurol.* **1981**, *10*, 429–436.
6 N. Perez, J. Sugar, S. Charya et al. *Brain Res. Mol. Brain Res.* **1991**, *11*, 249–254.
7 D.E. Comings. *Clin. Chem.* **1982**, *28*, 805–812.
8 S.S. Panter, J.D. McSwigan, J.R. Sheppard, C.R. Emory, W.H. Frey II. *Neurochem. Res.* **1985**, *10*, 1567–1576.
9 K.M. Mattila, H. Frey. *Electrophoresis* **1994**, *15*, 721–725.
10 T. Tsuji, S. Shimohama, S. Kamiya, T. Sazuka, O. Ohara. *J. Neurol. Sci.* **1999**, *166*, 100–106.
11 T. Tsuji, A. Shiozaki, R. Kohno, K. Yoshizato, S. Shimohama. *Neurochem. Res.* **2002**, *27*, 1245–1253.
12 P.F. Edgar, S.J. Schonberger, B. Dean, R.L. Faull, R. Kydd, G.J. Cooper. *Mol. Psychiatry* **1999**, *4*, 173–178.
13 S.J. Schonberger, P.F. Edgar, R. Kydd, R.L. Faull, G.J. Cooper. *Proteomics* **2001**, *1*, 1519–1528.
14 L.H. Choe, M.J. Dutt, N. Relkin, K.H. Lee. *Electrophoresis* **2002**, *23*, 2247–2251.
15 P. Davidsson, A. Westman-Brinkmalm, C.L. Nilsson et al. *Neuroreport* **2002**, *13*, 611–615.
16 T. Wisniewski, B. Frangione. *Neurosci. Lett.* **1992**, *135*, 235–238.
17 C. Hesse, C.L. Nilsson, K. Blennow, P. Davidsson. *Electrophoresis* **2001**, *22*, 1834–1837.
18 R. Fukuyama, T. Mizuno, S. Mori, K. Yanagisawa, K. Nakajima, S. Fushiki. *Eur. Neurol.* **2000**, *43*, 161–169.
19 I. Ueno, T. Sakai, M. Yamaoka, R. Yoshida, A. Tsugita. *Electrophoresis* **2000**, *21*, 1832–1845.
20 T. Sunderland, G. Linker, N. Mirza et al. *JAMA* **2003**, *289*, 2094–2103.
21 M. Vandermeeren, M. Mercken, E. Vanmechelen et al. *J. Neurochem.* **1993**, *61*, 1828–1834.
22 M. Jensen, H. Basun, L. Lannfelt. *Neurosci. Lett.* **1995**, *186*, 189–191.
23 N. Andreasen, L. Minthon, A. Clarberg et al. *Neurology* **1999**, *53*, 1488–1494.
24 M. Shoji, E. Matsubara, T. Murakami et al. *Neurobiol. Aging* **2002**, *23*, 363–370.
25 K. Nagga, J. Gottfries, K. Blennow, J. Marcusson. *Dement. Geriatr. Cogn. Disord.* **2002**, *14*, 183–190.

26 A. J. Green, R. J. Harvey, E. J. Thompson, M. N. Rossor. *Neurosci. Lett.* **1999**, *259*, 133–135.

27 H. Arai, Y. Morikawa, M. Higuchi et al. *Biochem. Biophys. Res. Commun.* **1997**, *236*, 262–264.

28 M. Otto, J. Wiltfang, H. Tumani et al. *Neurosci. Lett.* **1997**, *225*, 210–212.

29 A. Delacourte, L. Buee. *Int. Rev. Cytol.* **1997**, *171*, 167–224.

30 N. Sergeant, J. P. David, M. Goedert et al. *J. Neurochem.* **1997**, *69*, 834–844.

31 N. Sergeant, J. P. David, D. Lefranc, P. Vermersch, A. Wattez, A. Delacourte. *FEBS Lett.* **1997**, *412*, 578–582.

32 C. Janke, M. Holzer, J. Klose, T. Arendt. *FEBS Lett.* **1996**, *379*, 222–226.

33 C. Janke, M. Beck, M. Holzer, V. Bigl, T. Arendt. *Brain Res. Brain Res. Protoc.* **2000**, *5*, 231–242.

34 K. Blennow, A. Wallin, H. Agren, C. Spenger, J. Siegfried, E. Vanmechelen. *Mol. Chem. Neuropathol.* **1995**, *26*, 231–245.

35 A. Mitchell, N. Brindle. *Int. J. Geriatr. Psychiatry* **2003**, *18*, 407–411.

36 N. Itoh, H. Arai, K. Urakami et al. *Ann. Neurol.* **2001**, *50*, 150–156.

37 Y. H. Suh. *J. Neurochem.* **1997**, *68*, 1781–1791.

38 N. Andreasen, C. Hesse, P. Davidsson et al. *Arch. Neurol.* **1999**, *56*, 673–680.

39 P. C. Southwick, S. K. Yamagata, C. L. Echols et al. *J. Neurochem.* **1996**, *66*, 259–265.

40 M. Shoji, M. Kanai. *J. Alzheimers Dis.* **2001**, *3*, 313–321.

41 F. Hulstaert, K. Blennow, A. Ivanoiu et al. *Neurology* **1999**, *52*, 1555–1562.

42 M. Shoji, E. Matsubara, M. Kanai et al. *J. Neurol. Sci.* **1998**, *158*, 134–140.

43 E. Kapaki, G. P. Paraskevas, I. Zalonis, C. Zournas. *Eur. J. Neurol.* **2003**, *10*, 119–128.

44 M. Simons, P. J. Tienari, C. G. Dotti, K. Beyreuther. *FEBS Lett.* **1995**, *368*, 363–366.

45 N. Sergeant, S. Bombois, A. Ghestem et al. *J. Neurochem.* **2003**, *85*, 1581–1591.

46 S. H. Kim, R. Vlkolinsky, N. Cairns, G. Lubec. *Cell. Mol. Life Sci.* **2000**, *57*, 1810–1816.

47 M. A. Korolainen, G. Goldsteins, I. Alafuzoff, J. Koistinaho, T. Pirttila. *Electrophoresis* **2002**, *23*, 3428–3433.

48 A. Castegna, M. Aksenov, M. Aksenova et al. *Free Radic. Biol. Med.* **2002**, *33*, 562–571.

49 A. Castegna, M. Aksenov, V. Thongboonkerd et al. *J. Neurochem.* **2002**, *82*, 1524–1532.

50 J. Choi, C. A. Malakowsky, J. M. Talent, C. C. Conrad, R. W. Gracy. *Biochem. Biophys. Res. Commun.* **2002**, *293*, 1566–1570.

51 K. Krapfenbauer, E. Engidawork, N. Cairns, M. Fountoulakis, G. Lubec. *Brain Res.* **2003**, *967*, 152–160.

52 M. Merchant, S. R. Weinberger. *Electrophoresis* **2000**, *21*, 1164–1177.

53 E. R. Frears, D. J. Stephens, C. E. Walters, H. Davies, B. M. Austen. *Neuroreport* **1999**, *10*, 1699–1705.

54 B. M. Austen, E. R. Frears, H. Davies. *J. Pept. Sci.* **2000**, *6*, 459–469.

55 D. Beher, J. D. Wrigley, A. P. Owens, M. S. Shearman. *J. Neurochem.* **2002**, *82*, 563–575.

56 A. K. Vehmas, D. R. Borchelt, D. L. Price et al. *DNA Cell Biol.* **2001**, *20*, 713–721.

57 P. Lewczuk, H. Esselmann, M. Meyer et al. *Rapid Commun. Mass Spectrom.* **2003**, *17*, 1291–1296.

58 R. B. Rubin, M. Merchant. *Am. Clin. Lab.* **2000**, *19*, 28–29.

59 A. Vlahou, P. F. Schellhammer, S. Mendrinos et al. *Am. J. Pathol.* **2001**, *158*, 1491–1502.

60 J. D. Wulfkuhle, K. C. McLean, C. P. Paweletz et al. *Proteomics* **2001**, *1*, 1205–1215.

61 O. Carrette, I. Demalte, A. Scherl, O. Yalkinoglu, G. Corthals, P. Burkhard, D. F. Hochstrasser, J.-C. Sanchez. *Proteomics* **2003**, *3*, 1486–1494.

62 E. Levy, M. Sastre, A. Kumar et al. *J. Neuropathol. Exp. Neurol.* **2001**, *60*, 94–104.

63 A. Grubb, O. Jensson, G. Gudmundsson, A. Arnason, H. Lofberg, J. Malm. *N. Engl. J. Med.* **1984**, *311*, 1547–1549.

64 A. Deng, M. C. Irizarry, R. M. Nitsch, J. H. Growdon, G. W. Rebeck. *Am. J. Pathol.* **2001**, *159*, 1061–1068.

65 K. Sugaya, R. Greene, D. Personett et al. *Neurobiol. Aging* **1998**, *19*, 351–361.

Part VII
Mass Spectrometry and Bioinformatics

Biomedical Application of Proteomics
Edited by J.-C. Sanchez, G. L. Corthals, D. F. Hochstrasser

ISBN: 3-527-30807-5

20
MALDI-MS Imaging in Biomedical Research

Markus Stoeckli and Terry B. Farmer

20.1
Introduction

Molecular imaging techniques play an increasingly important role in the search for new and better treatments for diseases. Techniques such as magnetic resonance imaging (MRI) or X-ray have been traditionally applied in biomedical research. However, there are a number of technologies and improvements emerging from the need to gain deeper and more specific insight into the biological sample under investigation. The required information tends to go from structural imaging towards the analysis of the distributions at the molecular level. Molecular imaging can contribute to biomedical research in a number of different ways, including target-finding activities, following the modulation of the target upon treatment, and then imaging the drug and metabolite distributions (Fig. 20.1). Recent developments in MRI, near-infrared fluorescence (NIRF) imaging, bioluminescence imaging, and positron

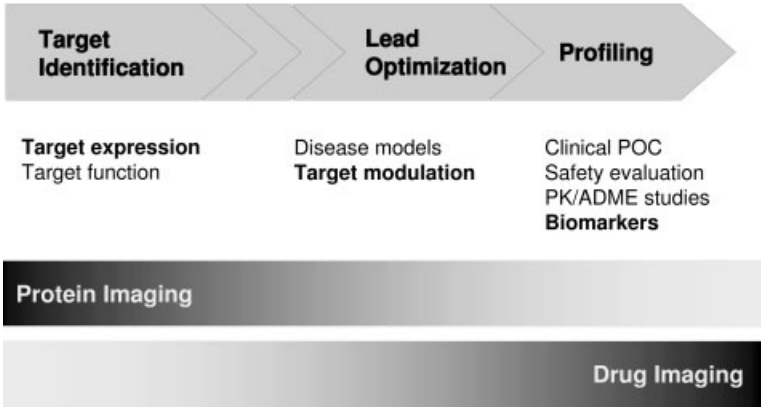


Figure 20.1 Molecular imaging in the drug discovery process. In early phases the focus is on localization of targets. As hits become available, the modulation of the target is monitored as well as the distribution of the drug and its metabolites in the tissue.

Biomedical Application of Proteomics
Edited by J.-C. Sanchez, G. L. Corthals, D. F. Hochstrasser

ISBN: 3-527-30807-5

Table 20.1 Comparison between different molecular imaging modalities. While MRI, NIRF, PET, and bioluminescence are *in vivo* techniques, they require the use of a specific reporter molecule for each analyte.

	MRI	***NIRF***	***PET***	***Biolum***	***Hist***	***MMSI***
In vivo	Y	Y	Y	Y	N	N
Sensitivity	µmol	nmol	pmol	fmol	amol	amol
Resolution	50 µm	1 mm	1 mm	1 mm	1 µm	50 µm
Time	min	s	mon	s	s	min
Typical application	Cells, proteins	Proteins, cells	Metabolites	Genes, cells	Proteins	Proteins, metabolites
Label	Y	Y	Y	Y	N	N
Dimensions	3-D	2(3)-D	3-D	2(3)-D	2-D	2-D
Cost	$$$	$	$$$$	$	$	$$

MRI, magnetic resonance imaging; NIRF, near-infrared fluorescence; PET, positron emission tomography; Biolum, bioluminescence; Hist, histology; MMSI, MALDI mass spectrometric imaging.

emission tomography (PET) all provide exciting new possibilities to analyze changes at the molecular level in model organisms and in humans.

While these methods allow *in vivo* imaging, they depend on the use of reporter molecules which specifically interact with the molecules of interest and thereby reveal their presence. This constraint limits applications in biomedical research in two ways. First, it is expensive and time consuming to develop a reporter molecule for every analyte and specificity is often not constrained to a single analyte. Furthermore, these techniques do not allow imaging of unknown molecules. For these reasons the mass spectrometric imaging technique presented in this chapter, matrix-assisted laser desorption ionization mass spectrometry (MALDI-MS), is gaining importance and is able to partially fill the need for a priori molecular imaging. MALDI-MS also has sensitivity comparable to immunochemical methods. A rough comparison between different imaging techniques showing their relative advantages is given in Table 20.1.

The scope of this chapter does not include the field of secondary ion mass spectrometric (SIMS) imaging and laser desorption ionization (LDI) imaging. These techniques are commonly applied for imaging of small organic molecules and metal ions, but the low accessible mass range limitation does not allow the imaging of peptides and proteins.

20.1.1 Background

Matrix-assisted laser desorption ionization/time-of-flight mass spectrometry (MALDI-TOF-MS) is a sensitive technique for the analysis of peptides and proteins, allowing detection in the picomole to femtomole range. This sensitivity

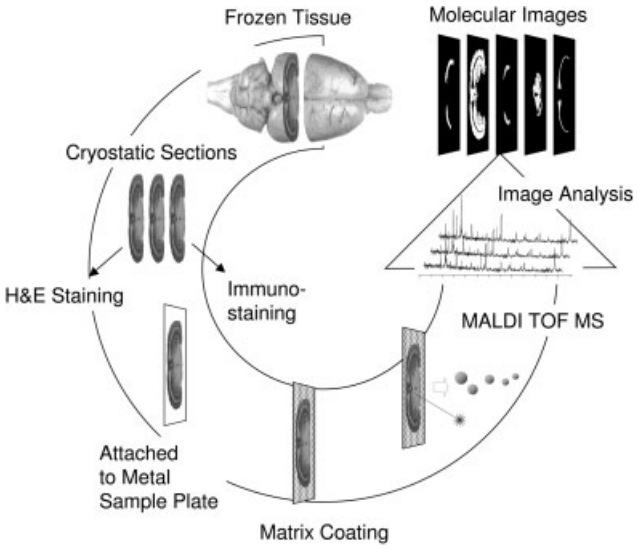


Figure 20.2 MALDI-MS imaging process. A tissue section is placed on a metal sample support, covered with a matrix solution and introduced into a mass spectrometer. The sample is rastered over with a pulsed UV laser while acquiring mass spectra from every location. Multiple molecular images are calculated out of this data set.

brings the direct analysis of biomolecules in tissue into the realm of possibility. MALDI-MS uses a matrix that absorbs the energy from the impinging laser beam at the wavelength of the laser used, and coupled with a TOF analyzer allows the desorption and analysis of biomolecules larger than 100 000 Da.

In 1987 Tanaka et al. first reported experiments that led to the development of the MALDI process. They used a fine powder of cobalt metal in glycerol as the matrix for the observation of ions with an m/z of 34 000 [1]. Tanaka and colleagues also reported the observation of lysozyme oligomers (up to seven lysozyme molecules for an observed mass of ~100 kDa) using a nitrogen laser with a glycerol/metal powder matrix. Later, Hillenkamp and Karas used nicotinic acid as the matrix for a frequency quadrupled Nd-YAG laser to yield "soft" desorption of mellitin (2843 Da) [2]. The use of an organic matrix, that absorbed the majority of the laser power leading to explosive desorption of the matrix and analyte into the gas phase without degradation of the analyte, demonstrated the utility of laser desorption mass spectrometry for the analysis of large biomolecules. Soon afterward MALDI-MS results of serum albumin (67 000 Da) were reported using nicotinic acid as the matrix with a quadrupled Nd-YAG laser [3].

MALDI-TOF-MS has now become an indispensable tool in biomedical research. While it was originally used to analyze body fluids and tissue extracts, MALDI-MS goes one step further by spatially analyzing biological tissue sections [4]. In MALDI-MS imaging, tissue sections are placed on a sample support, covered with ma-

trix and transferred to the mass spectrometer (Fig. 20.2). A pulsed laser is used to raster over the sample while acquiring mass spectra of every image location. From one array of acquired mass spectra, images representing the biological distribution of hundreds of substances can be generated.

20.2 Methods

20.2.1 Tissue Preparation Technique

Sample preparation is one of the most critical factors in MALDI-MS and significant research effort has been undertaken to optimize preparation procedures. Biological tissue represents an extremely complex and challenging environment for direct analysis by MALDI-MS. The multiple components present in a tissue section (e.g. proteins, lipids, oligonucleotides, carbohydrates, small organic molecules, and salts) interfere with each other in the MALDI process and prevent optimal detection. A high salt concentration, for example, strongly limits the ionization efficiency of other molecules of interest produced in the MALDI process. There are multiple approaches to overcoming this problem and these are described below.

20.2.1.1 **Blotting**

One of the possible approaches is to extract the proteins out of the tissue, while conserving their relative spatial distribution. This has been achieved using blotting techniques on different substrates. An example of this approach is to use a surface prepared by attaching C18 beads typically used in reverse phase high-performance liquid chromatography (RP-HPLC) to a metal plate and blotting freshly cut tissue sections onto it [4]. The surface is then washed with water before matrix application. This method is successfully used to analyze the spatial distribution of neuropeptides in rat pituitary tissue and to assign them to specific brain regions.

The chemical properties of the blotting surface determine which analytes are preferentially detected. Typically, hydrophobic surfaces are applied, because they allow an aqueous washing step after the binding of the peptides. Conductive polyethylene is one of the surfaces that has been tested for tissues blots [5]. It combines the binding ability of a hydrophobic polymer with the conductivity of the embedded carbon particles. The polymer film (Goodfellow, Berwyn, PA, USA) is attached to the metal sample support with conductive tape so the surface is maintained at the same electric potential, which is crucial for achieving a high-quality mass measurement.

More complex surfaces can be designed to either extract specific proteins or a class of proteins from the tissue sample. In the past we have tested several differ-

ent surfaces including glass, metal oxides on glass, Teflon, gold-sputtered plates, PVDF, PET, and antibodies bound to surfaces. All of them proved to be able to bind peptides and proteins. A more systematic study will follow that may enable the classification of the surfaces and their use. Nevertheless, while the concept of blotting is straightforward, the practical process itself has some complications. The blotting surface is first covered with a solvent, typically methanol. The semi-dry tissue is brought in contact with the surface and removed after a few seconds. The solvent layer and the liquid from the tissue cause diffusion of peptides and proteins. This diffusion phenomenon is necessary for the transfer to the blotting surface, but leads to loss of spatial resolution in the imaging experiment.

20.2.1.2 **Frozen Sections**

The direct use of tissue sections provides advantages over blotting which are crucial for the imaging process. Since the laser is rastered directly over the tissue, one can safely assume that the ions produced are from the underlying tissue. Smearing effects that can occur in blotting techniques are eliminated. By using frozen tissue for sectioning, the proteins will remain in place during the cutting process and only negligible mechanical smearing occurs by the moving blade. This method is also compatible with immunostaining techniques so that direct comparison between molecular images and stained slides is possible. No suitable method of analysis by MALDI-MS imaging for paraffin-embedded tissues has yet been found.

For sectioning, tissues are frozen by immersion in liquid nitrogen or in an isopropanol solution cooled with dry ice. The latter has the advantage of minimizing the formation of ice crystals inside the tissue, which causes cracking effects. Special care has to be taken not to contaminate the outer surface of the tissue with any foreign material. When attaching a large piece of tissue, OCT (Tissue-Tek, Zoeterwoude, The Netherlands) may be applied to the back of the tissue but great care must be taken to avoid contamination of the region where the tissue is cut. OCT, a polymer mixture which is commonly used to attach frozen sections to the sample holder, results in a large number of intense signals in the MALDI-MS spectrum, which overlap with signals from the compounds in tissues. OCT is therefore not suitable for embedding tissues to be analyzed by MALDI-MS imaging. For smaller tissues, attachment with water is an alternative approach.

The frozen sections are picked up with a special support which can be interfaced with the MALDI instrument. In the simplest case it is a blank stainless steel plate, which is also used for analysis of discrete samples. While brain samples adhere well to stainless steel, other tissues like kidney or skin detach easily and may contaminate the instrument. For these tissues, glass or silicon wafers used in the semiconductor industries are an alternative. Silicon offers the benefit of conductivity, allowing good mass spectral quality (see section on “Blotting”). Depending on the size of the plate and the MS instrument, multiple sections may be placed on the same support.

Some final details regarding analysis:

- The plates are stored at –80°C prior to analysis.
- Rinsing and fixation techniques may improve the sample quality and need to be developed for every sample individually.
- Mild fixation with ethanol or acetone give better results for brain sections.

20.2.1.3 **MALDI Matrix Deposition**

Matrix deposition on the tissue section is a very important step of the sample preparation and much effort has been invested in the development of an optimized method. It is generally accepted that the analytes need to co-crystallize with the matrix in order to yield high analyte signal in the mass spectrum. In fact, in practical experiments, where dry matrix crystals were deposited directly onto the tissue surface, no protein signals could be detected. The most direct way to achieve protein co-crystallization is to apply the matrix solution as a small drop with a pipette onto a tissue section. The mass spectrum from such a preparation presents a high number of signals (Fig. 20.3) in a mass range up to 30 kDa. This preparation is not well suited for imaging mass spectrometry, since the proteins tend to diffuse from their original position during the matrix application and crystallization process and are finally detected at a different location in the tissue section. Nevertheless, this matrix application method is often used if peptide/protein profiling of a distinct tissue region is required rather than imaging.

It is possible to minimize analyte diffusion using smaller matrix droplets, comparable in size to the image raster. This can be achieved either using a spray tech-

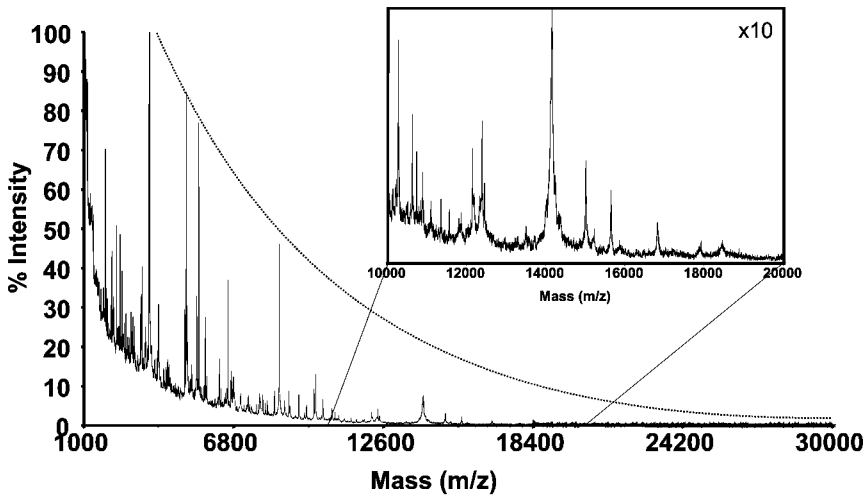


Figure 20.3 Typical mass spectrum acquired from a frozen tissue section, after drop-deposition of sinapinic acid. A large number of signals are observed in the low mass regime, with sensitivity of the technique decreasing towards higher masses.

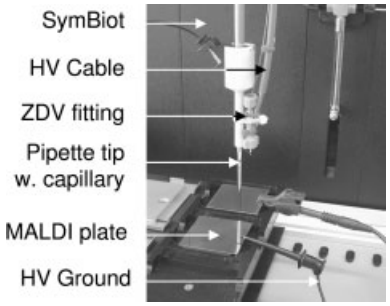


Figure 20.4 Electrospray coating of matrix results in a fine and homogeneous layer of matrix. A robotic device is used to move a spray tip held at high voltage over the sample plate.

nique or through droplet deposition with picoliter pipettes. Two methods of spraying have been tested on biological samples. Electrospray coating is based on the same principle as used in LC-MS electrospray sources. The matrix is dissolved in a conductive buffer (e.g. 10 mg ml^{-1} matrix in 50:50 acetonitrile/1% trifluoroacetic acid) and pumped to a small-diameter capillary using a syringe pump at a typical rate of 2 µL min^{-1}. The end of the capillary is connected to a metal fitting by which a voltage of 3.5 kV is applied. At the other end of the fitting, a short length of fused silica capillary serves as a spray tip and is moved over the grounded sample plate at a constant speed and at a distance of 8 mm (Fig. 20.4). This set-up produces a homogeneous and reproducible layer of matrix with only fine droplets reaching the sample surface. Due to the desolvation effect of the electrospray process, the droplets reaching the target do not contain enough liquid to efficiently dissolve proteins from the tissue. Recently published results that used this method showed poor sensitivity, but excellent spatial resolution.

Pneumatic sprays deliver bigger droplets than electrospray and allows a faster spraying rate to be used. A sprayer similar to the one used to develop TLC (thin-layer chromatography) plates is filled with matrix solution and the sample plate is sprayed in multiple steps [6]. In every step a thin film of matrix is deposited onto the surface and allowed to dry before the next coating. Up to 10 layers are sprayed before the sample is analyzed in the mass spectrometer.

Airbrushing the matrix is a similar approach to deliver fine droplets which dry fast on the surface because of the presence of a stream of air. A dual-action airbrush is used which allows the experimenter to independently control the flux of matrix solution and air. The size of the droplets is controlled by different needle diameters.

Pneumatic techniques currently have the disadvantage of not being extremely reproducible since it is not easy to control the exact amount of matrix sprayed per sample area. Further developments in this area may overcome this problem.

Covering the complete area to be sampled with a large matrix volume is currently the approach which yields the strongest ion signals (Fig. 20.5). In this procedure, the sample plate is cooled to 4 °C and covered with a solution of 20 mg ml^{-1} matrix in 50:50 acetonitrile/0.1% trifluoroacetic acid and applying sufficient solution on the tissue to a final concentration of 20 µL cm^{-2}. The preparation is then brought to room temperature to allow evaporation of the matrix solvent. Using sinapinic acid as the matrix yields crystals with a size up to 200 µm.

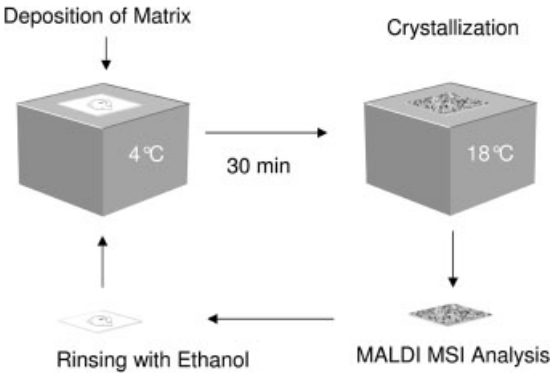


Figure 20.5 Matrix crystallization procedure. The sample is cooled to 4 °C and the matrix is added as a bulk volume. The slow temperature gradient determines the formation of large crystals. The process is repeated after imaging and rinsing with ethanol.

Due to large volumes of solvent, migration of proteins is often observed and needs to be considered when evaluating the molecular images. Several images can be acquired from the same section, by rinsing the matrix layer on the tissue with ethanol after imaging and recoating it with new matrix solution.

Complete immersion of the sample in a matrix solution is another approach for production of a homogeneous coating. The plate with the attached tissue section is completely immersed in matrix solution and allowed to air dry. This is a modification of the technique described above with the difference that the proteins that are eluted out of the tissue will be washed out from the surface and fewer artifacts will occur. Due to the smaller amount of matrix solution remaining on the surface after evaporation of the solvent, the crystals layer is thinner and results in modest signal intensity.

Once the sample is covered with matrix, it may be stored under vacuum for several days without significant deterioration of the MALDI MS spectra quality.

20.2.1.4 **Instrumentation**

Imaging by MALDI-MS differs from conventional single sample analysis in that a series of mass spectra are acquired from different locations in an ordered pattern covering the area of interest. An imaging instrument must therefore facilitate movement of the sample plate under the laser beam or of the laser beam relative to the sample. High-performance instruments typically have a very narrow optimal ionization region (< 0.5 mm) and therefore they are not suitable for movement of the laser over larger areas. In modern instruments, motor-driven stages are used for precise positioning of sample plates under the stationary laser.

Spatial resolution is an important consideration in imaging particularly if the analyst is examining cellular processing. Two aspects of MALDI imaging are in-

volved in spatial resolution: the size of the cells that are being examined and the diameter of the laser beam when it strikes the sample surface.

Cell size can vary from 10 nm to 500 μm in diameter depending on the species, tissue cell type, and its location. *Lymnaea stagnalis* contain neurons with a diameter of 50–500 μm and these neurons have been used for single-cell MALDI analysis. *Aplysia californica* bag neurons, 50 μm in diameter, and single dense core vesicles in the exocrine atrial gland, 1–2 μm diameter, have been analyzed and peptide processing in these cells has been delineated using MALDI-TOF-MS [7]. MALDI-MS can also be used to map groups of cells as demonstrated by the imaging of insulin to localize a pancreatic islet (~250 μm in diameter) [4].

Initial work in MALDI-TOF-MS was performed with a focused laser spot size that was 200–300 μm in diameter upon reaching the target. As long as the object was to desorb all analyte and matrix from the target, regardless of its spatial resolution, this was acceptable. In contrast, work done with microprobe or LAMMAS was performed by rastering an ion beam across the surface with an ion beam diameter of 1–10 μm, enabling work to be done on individual cells. In order to examine signals from individual cells by MALDI-MS a smaller laser spot size was required. Work was begun in the Caprioli lab to decrease the laser spot size for MALDI-MS by placing a mask with a 3-mm hole in front of the UV laser exit aperture. This mask, in conjunction with the focusing system, gave a laser spot size on the target of approximately 25 μm. Mask holes smaller than 3 mm were also investigated to shrink the laser spot on target, but the results showed the laser fluence was too low to desorb enough ions into the gas phase to be detected [4]. Currently, an extra iris at the laser beam exit and before the focusing system is incorporated in commercial instruments to produce a spot of approximately 30 μm on the target.

Due to the interest in individual cell processes, decreasing the size of the laser spot on target while retaining enough laser power to desorb the analytes became a priority. Spengler's lab had been using a microscope focusing system in their microprobe work to generate a spot system of approximately 1 μm. They developed a system to focus lasers and used this to develop an instrument with a focused laser beam that can be used in bimolecular applications [8].

When acquiring thousands of spectra in imaging mode, the overall speed of the system becomes a critical consideration. The most important factors are the repetition rate of the laser, the speed of the control electronics, and the data acquisition system. With the availability of lasers with repetition rates in the hundreds of Hertz, acquisition times can be greatly reduced and are typically around an hour for an image of 100×100 points [9].

20.2.1.5 **Software**

MALDI-MS imaging software comprises two aspects: data acquisition and image processing. Acquisition software must be specifically written for each instrument hardware. Integrated versions have been developed for the Applied Biosystems Voyager [9, 10] systems and Sciex QStar [11] systems. On all instruments with

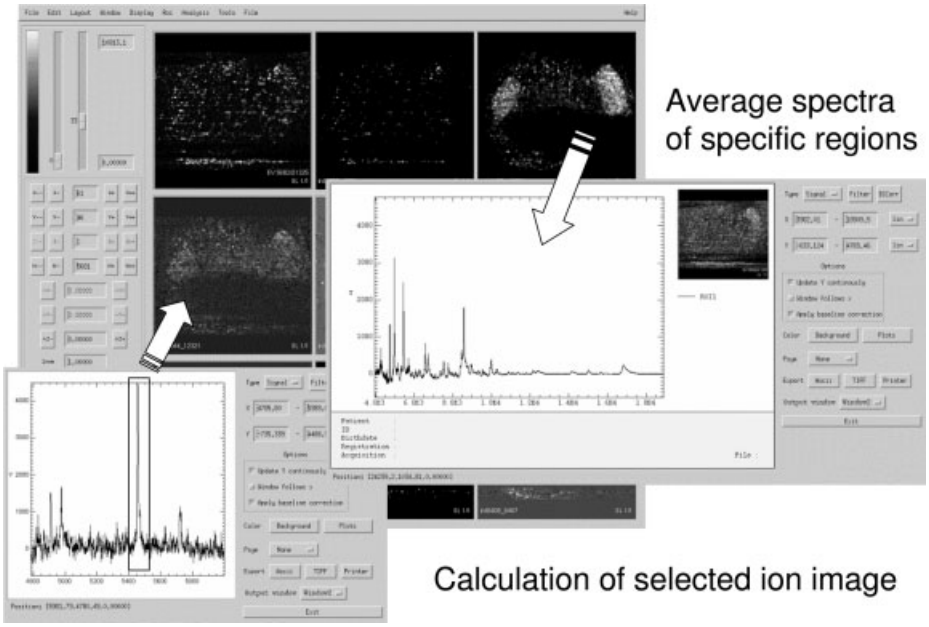


Figure 20.6 During image acquisition the mass spectra of each image point are acquired. Image analysis software is then used to (a) display the intensity distribution of a specific molecular weight and (b) show the mass spectra for areas of interest.

macro language capability, acquisition routines for sequential acquisition of spectra can easily be implemented [12–14].

The file format of mass spectrometric image data is not standardized and different approaches are described in the literature. The Analyse™ 7.5 file format (Mayo Foundation, Rochester, MN, USA) was chosen by our labs because it is an open format which can easily be implemented in any application. A number of clinical image analysis packages also support this file type. This capability is particularly useful if the MS data are to be compared with data from other imaging methods.

Image analysis tools range from commonly used software like SigmaPlot™ (SPSS Inc., Chicago, IL, USA) to more sophisticated tools [10, 15]. BioMap [9] was originally designed for MRI data and then extended to support other imaging techniques including MS. The data can be analyzed in two different ways (Fig. 20.6). A mass range or peak is defined and the corresponding MS image is calculated or the mass spectrum corresponding to a specific image region is displayed. Both modes are used in conjunction in practical use. Typically, an image is calculated, the region of higher signal abundance is localized, and the mass spectra of this region are analyzed for coexpressed analytes which are then again imaged. Beside these basic functions, routines for baseline correction of spectra, spatial filtering, averaging of spectra, and others greatly enhance the information obtainable from a data set.

20.3 Applications

MALDI-MS is a relatively young field and applications are starting to be published at an increasing rate. In a number of publications the analysis of single cells using MALDI as a profiling tool is shown [16]. This technique was demonstrated in the analysis of cells from a freshwater snail *Lymnaea stagnalis* by removing a single brain cell, pipetting out the contents of the cell and mixing with matrix to obtain mass spectral information. Removing a single neuron from the snail, preventing cell rupture, and then adding matrix with subsequent rupturing of the cell was used to generate information on the peptides present in the neuron [17]. Peptide profiles for several more cell types were obtained in addition to peptide sequencing in order to delineate the peptide processing that occurs in the cells of *Lymnaea stagnalis* [18–20]. This work led to the analysis of single neurons of *Aplysia californica* using the matrix solution as a salt-removal agent to facilitate MALDI-MS analysis and reveal prohormone processing [21]. It was also shown that MALDI-MS not only is able to detect signals from different neuropeptides, but also allowed their identification by PSD analysis [22]. In the above studies the analysed neurons were large enough to be isolated from the surrounding cells.

For smaller cells, laser-capture microdissection (LCM) is used to extract single cells or a group of cells for analysis with MALDI-MS [23]. The film with the attached cells is directly attached to a MALDI sample plate and microdroplets of matrix are deposited on the surface. The advantage of this approach lies in the accuracy and specificity of the information gained. By isolating the cell and depositing matrix afterwards, it is assured that the signals obtained by the MS measurement originate from only these cells. In contrast, the matrix deposition methods described above all show diffusion of proteins from their original location to some degree. It is mostly due to not having a suitable matrix deposition procedure available that a number of publications limit studies to profiling of different regions in tissue sections.

The ability to analyze the entire tissue sample and to localize peptides and proteins has enormous potential for biological and biomedical research. The use of MALDI-MS to map proteins in biological samples was first demonstrated by imaging human buccal mucosa cells [4]. Tissue sample preparation was also examined in an effort to acquire MALDI-TOF-MS images. Rat pituitary tissue was blotted onto a surface consisting of C18 beads attached to double-sided tape on a metal target plate and matrix was electrosprayed onto the blotted beads. The resulting analysis allowed the mapping of peptides and proteins localized to specific regions of the pituitary. Direct tissue analysis was also performed on rat pancreas and the image of a rat pancreatic islet was obtained by extracting the signal for insulin. Chaurand and Caprioli [24] applied MALDI-MS to mouse cauda epididymis section, where they detected an increased concentration of most proteins in the lumen of the epididymal tubule. MALDI-MS images obtained from mouse brain sections [6] reveal the localization of several peptides which show a distinct distribution and reduced diffusion of the peptides.

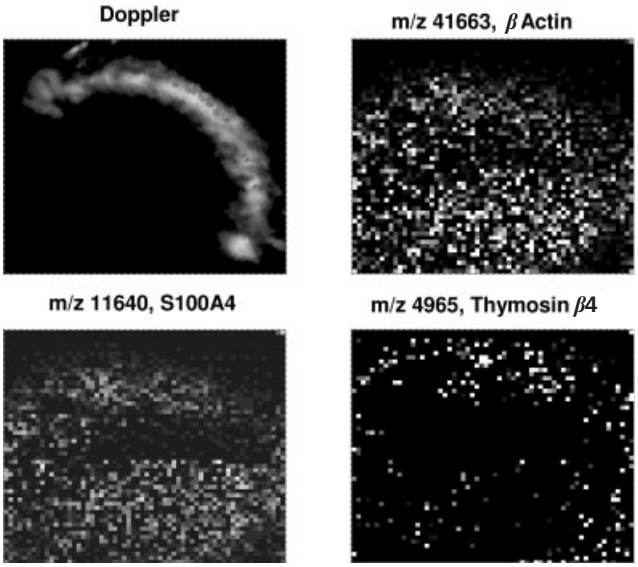


Figure 20.7 MS imaging of a human glioma xenograft. Thymosin β4 was detected in the growing area of the tumor, while S-100A4 was present in the ischemic area.

MS imaging of a human glioma xenograft grown in a nude mouse revealed several peptides which had a distribution attributed to the tumor growth [25]. One of these peptides, thymosin β4, is known to be present in higher levels in fast-growing tissues. The top left panel in Figure 20.7 contains a Doppler image highlighting the area of tumor growths. The MS image of thymosin β4 (lower right) shows a similar distribution, with higher levels present at the growing rim of the tumor. The identification of the peptide was based on the molecular weight of MALDI-MS and was confirmed by analyzing a brain extract.

20.3.1 Beta-amyloid Imaging

The spatial mass spectrometric analysis of β-amyloid peptides in mouse brain section is a typical application of MS imaging in biomedical research and is discussed in detail as an example for highly specific imaging.

The formation of plaques in the brain is directly linked to the progression of Alzheimer's disease (AD). The plaques contain β-amyloid (Aβ) peptides which are cleavage products of the amyloid precursor protein (APP) by different secretases. The types and concentration levels of the different peptides are of interest for the study of the disease progression. Brains of APP23 transgenic mice develop the pathogenic features of AD and a high plaque load is observed, beginning at eight months. The Aβ peptides have molecular weights up to 4500 Da, which is an ideal range for MALDI-MS in terms of sensitivity and mass resolution.

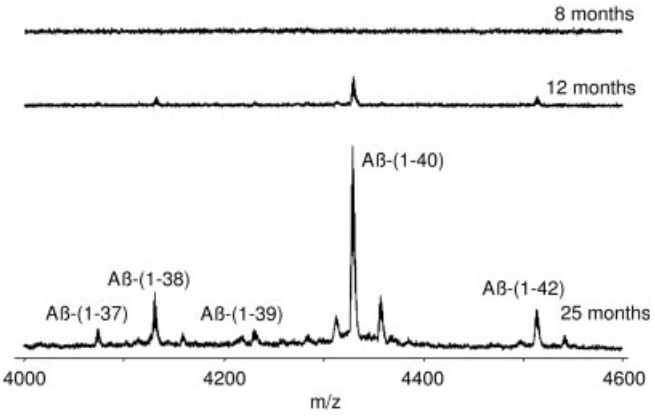


Figure 20.8 Mass spectra of APP23 mouse brain extracts. At 25 months high levels of Aβ peptides are detected.

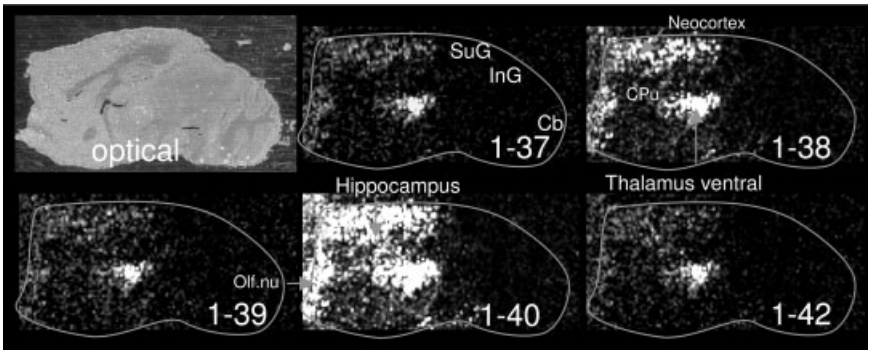


Figure 20.9 Molecular images acquired from two sagittal APP23 mouse brain sections. The section at lateral 0.4 mm position shows high concentration of the peptides in the thalamus and cortex.

For image analysis, brain sections from 24-month-old animals were prepared using the drop deposition procedure, with sinapinic acid, as described in the "MALDI Matrix Deposition" section. The result is a homogeneous preparation with crystals up to 200 µm in size. Figure 20.8 shows a mass spectrum of a brain region with high plaque load. The peaks are labeled with the corresponding Aβ species assigned by the measured molecular weight. The height of the peaks does not directly correlate to the relative amount of the corresponding species. In fact the different structures lead to differences in ionization efficiency which need to be taken into account.

Figure 20.9 shows molecular images of the different Aβ peptides in a sagittal mouse brain sections. The MS image was acquired with a resolution of 100 µm,

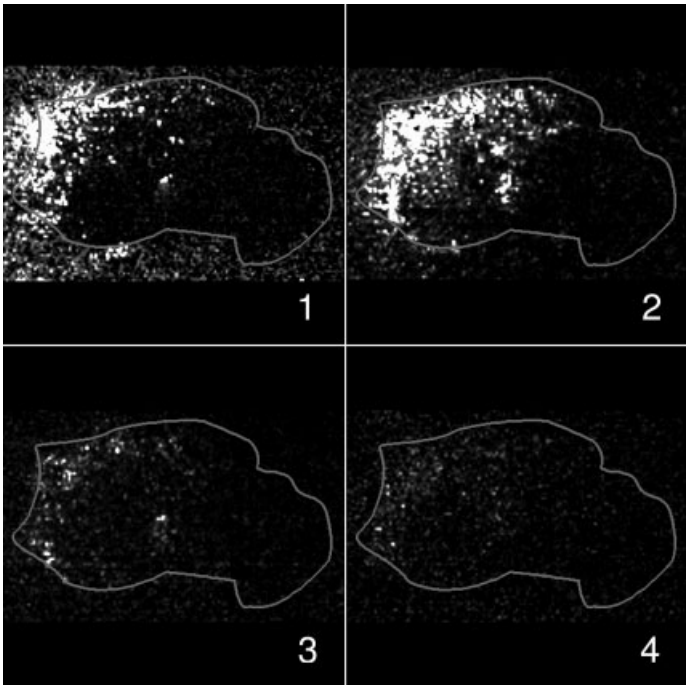


Figure 20.10 A mouse brain tissue section was subject to a repetitive MALDI-MS imaging analysis (1–4). Between each measurement, the section was washed with ethanol. The MS images correspond to the mass of Aβ 1-40.

resulting in an array of 120×70 image points. In the left-hand corner, an optical image of an adjacent section is shown for comparison. The MS image of Aβ 1-40 shows by far the highest intensity. Areas of high concentration of the peptide are the neocortex as well as the thalamus. These are also the regions where a high plaque load was found using Thioflavin S stain on adjacent sections. Compared with conventional stains, MS imaging has two major advantages: it delivers very specific information based on the molecular weight of the analyte and it allows the simultaneous measurement of hundreds of analytes. In the Aβ example of Figure 20.9, the different Aβ peptides show unique distributions and concentrations.

The nature of the sample preparation techniques (as described above) can lead to variations in the final results. This issue is typically addressed by measuring multiple adjacent sections or by remeasuring the same sample. Figure 20.10 shows four distributions of Aβ 1-40 acquired from the same brain section, but with rinsing and matrix-recoating steps between measurements. In the first run, some peptide signals are measured outside the tissue section. This is attributed to the migration of peptides and matrix crystals during the coating process. Rinsing with ethanol and recoating delivers a more consistent image of the Aβ distribu-

tion. Also, crystallization of the matrix is more homogeneous as the matrix crystallizes on the same spots where the laser hit the sample on the previous run. Even on the third and fourth measurements of the series, the distribution of the Aβ peptides could be reproduced.

20.4 Conclusions

MALDI-MS imaging is an emerging technology with enormous potential for applications in biomedical research. The ability to image analytes without the need for tagging reagents is of great value for target finding. Images of hundreds of analytes can be acquired within minutes with the currently established protocols and instrumentation.

The sensitivity of MALDI-MS is a critical factor for biological applications. It is dependent on many factors, most of which are not controlled by the analyst. Tissues are complex systems that contain salts, lipids, etc. that can interfere with the acquisition of a signal for the desired analyte. The analyte in a tissue may be detectable if the protein or peptide is concentrated in a very small area rather than dispersed over the whole tissue area. Spotting experiments with a protein (insulin) on brain tissue sections were used to estimate the required protein concentration for successful MALDI-MS. An area concentration of 12 fmol mm^{-2} still resulted in a signal in the MALDI mass spectrum with a signal-to-noise ratio of 5. With the laser spot diameter of 50 μm, this results in an absolute detection level of 25 amol per image point. Current applications of MALDI-MS demonstrate that this sensitivity is in a range where it can be used to detect distributions of relevant biomolecules.

Currently there is no commercial instrument manufacturer offering MALDI-MS and this limits the number of users to a small group of research laboratories. The first MALDI-MS imaging instruments are expected to become commercially available by the year 2004 for MALDI-TOF and MALDI-Q-TOF systems.

There are known limitations to MS imaging which influence the way it is applied to tissue imaging. For one, the instrumentally limited low sensitivity in the high mass range limits MS imaging to the analysis of peptides, small proteins, and highly abundant high mass proteins. Second, the spatial resolution of the images is limited by the crystallization procedure as well as instrumental parameters. Typical matrix-coating procedures result in large crystals and migration of the proteins. This problem can be overcome by using spray deposition or small drop deposition. The laser beam diameter rastering over the surface determines the maximal resolution. While current MALDI-MS instrumentation uses a 50-μm beam, this can be improved down to a few micrometers. Fewer ions are produced per laser shot resulting in lower sensitivity which must be compensated by improving the detector.

Despite the current limitations discussed above, MALDI-MS imaging fits well in the current portfolio of alternative molecular imaging techniques and it allows

us a unique view of the complex processes in biological tissues. With the full potential remaining to be explored, MALDI-MS imaging is a promising tool for further developments in biomedical research.

20.5 References

1 Tanaka, K., Ido, Y., Akita, S., Yoshida, Y., Yoshida, T. *Second Japan China Joint Symposium on Mass Spectrometry*, **1987**.

2 Karas, M., Bachmann, D., Bahr, U., Hillenkamp, F. *Int. J. Mass Spectrom. Ion Processes* **1987**, *78*, 53–68.

3 Karas, M., Hillenkamp, F. *Anal. Chem.* **1988**, *60*, 2299–2301.

4 Caprioli, R.M., Farmer, T.B., Gile, J. *Anal. Chem.* **1997**, *69*, 4751–4760.

5 Chaurand, P., Stoeckli, M., Caprioli, R.M. *Anal. Chem.* **1999**, *71*, 5263–5270.

6 Chaurand, P., Schwartz, S.A., Caprioli, R.M. *Curr. Opin. Chemical Biol.* **2002**, *6*, 676–681.

7 Rubakhin, S.S., Garden, R.W., Fuller, R.R., Sweedler, J.V. *Nature Biotechnol.* **2000**, *18*, 172–175.

8 Spengler, B., Hubert, M. *J. Am. Soc. Mass Spectrom.* **2002**, *13*, 735–748.

9 Stoeckli, M., Staab, D., Staufenbiel, M., Wiederhold, K.H., Signor, L. *Anal. Biochem.* **2002**, *311*, 33–39.

10 Stoeckli, M., Farmer, T.B., Caprioli, R.M. *J. Am. Soc. Mass Spectrom.* **1999**, *10*, 67–71.

11 Reyzer, M.L., Korfmacher, W.A., Ng, K., Hsieh, Y., Caprioli, R.M. *J. Mass Spectrom.* **2003**, *38*, 1081–1092.

12 Binz, P.A., Mueller, M., Walther, D. et al. *Anal. Chem.* **1999**, *71*, 4981–4988.

13 Bienvenut, W.V., Sanchez, J.C., Karmime, A. et al. *Anal. Chem.* **1999**, *71*, 4800–4807.

14 Garden, R.W., Sweedler, J.V. *Anal. Chem.* **2000**, *72*, 30–36.

15 Muller, M., Gras, R., Appel, R.D., Bienvenut, W.V., Hochstrasser, D.F. *J. Am. Soc. Mass Spectrom.* **2002**, *13*, 221–231.

16 Li, L., Garden, R.W., Sweedler, J.V. *Trends Biotechnol.* **2000**, *18*, 151–160.

17 van Veelen, P.A., Jimenez, C.R., Li, K.W., Wildering, W.C., Geraerts, W.P.M., Tjaden, U.R., van der Greef, J. *Org. Mass Spectrom.* **1993**, *28*, 1542–1546.

18 Li, K.W., Hoek, R.M., Smith, F. et al. *J. Biol. Chem.* **1994**, *269*, 30288–30292.

19 Jimenez, C.R., van Veelen, P.A., Li, K.W. et al. *J. Neurochem.* **1994**, *62*, 404–407.

20 Li, K.W., Jimenez, C.R., van Veelen, P.A., Geraerts, W.P.M. *Endocrinology* **1994**, *134*, 1812–1819.

21 Garden, R.W., Moroz, L.L., Moroz, T.P., Shippy, S.A., Sweedler, J.V. *J. Mass Spectrom.* **1996**, *31*, 1126–1130.

22 Li, L., Garden, R.W., Romanova, E.V., Sweedler, J.V. *Anal. Chem.* **1999**, *71*, 5451–5458.

23 Bhattacharya, S.H., Gal, A.A., Murray, K.K. *J. Proteome Res.* **2003**, *2*, 95–98.

24 Chaurand, P., Caprioli, R.M. *Electrophoresis* **2002**, *23*, 3125–3135.

25 Stoeckli, M., Chaurand, P., Hallahan, D.E., Caprioli, R.M. *Nature Med. (New York)* **2001**, *7*, 493–496.

21
Protein Variations: Resources and Tools

Yum Lina Yip, Maria Livia Famiglietti, Elisabeth Gasteiger, and Amos Bairoch

21.1
Introduction

The mapping of the human genome has provided a large volume of data, which, rather than representing the conclusion of years of work, forms the basis for future projects such as the characterization of all human genes and the study of the role of genetic diversity in determining health or disease status [1]. Very few traits have a single genetic origin. Most depend on the combination of various genetic factors, together with environmental influences. For complex diseases it is a very arduous task to decipher the genetic background contributing to the manifestation of the disease. A major challenge in medical genetics is, indeed, to understand the relationship between genetic and phenotypic variation. A first step to reach this goal is the identification and characterization of all variants of human genes.

Since most genomic diversity is due to single nucleotide polymorphisms (SNPs), which represent about 90% of human DNA sequence variation [2], a particular effort is currently being put into their detection and mapping. High-throughput genotyping techniques [3, 4] are generating huge amount of data [5–13] whose collection and integration with other medical, biological, and biochemical information is crucial in unravelling disease mechanisms.

The need to develop "a considered, integrated, and systematic approach to the problem of mutation documentation" was recognized as early as 1994, and resulted in the HUGO Mutation Database Initiative [14]. This initiative evolved into the Human Genome Variation Society (http://www.hgvs.org/), with the major aim of providing up-to-date documentation of genetic variations via a system of cross-linked specialized and central databases whose content and structure follow uniform guidelines and recommendations [15, 16]. Presently, a variety of databases [17] and tools of search and information retrieval are available on the Web. In this chapter, we illustrate the main topics concerning the annotation process and present the main online resources documenting human variation, including genomic and protein databases, and related search tools. Finally, we describe different analysis tools for *in silico* identification and characterization of protein variants. Variation at the protein level can be due to a plethora of different modifications. In

Biomedical Application of Proteomics
Edited by J.-C. Sanchez, G. L. Corthals, D. F. Hochstrasser

ISBN: 3-527-30807-5

this chapter, the term "protein variants" indicates protein sequences containing amino acid substitutions, due to non-synonymous coding SNPs.

21.2 Medical Protein Annotation

The expression "medical protein annotation" is used for the process of collecting, reviewing, and integrating information on disease-associated genes/proteins, susceptible of inclusion in a database for public use. It can cover different fields. Hence, disease phenotypes, clinical information, therapeutic information, genetic and structural data, might be more or less relevant depending on the main focus of the database and the needs of the particular scientific community to which the database is addressed. Nevertheless, independent of the content of a database, the general major challenge in the annotation process and, at the same time, obligation toward the scientific community is to make sure that the information is correct, accurate, clear, and not misleading.

In the annotation of sequence variants, two issues are particularly relevant to the problem of data quality in mutation databases, namely nomenclature and the effect of sequence variations. A consistent and widely accepted nomenclature is indispensable for unambiguous and accurate reporting of sequence variants. Recommendations have been published for the description of simple and complex sequence changes [18, 19]. However, due to the complexity of the subject, discussions are necessary in order to improve variant designation rules. Official recommendations are available online and scientists are invited to submit remarks and nomenclature suggestions for complicated cases that are not yet covered (http://www.genomic.unimelb.edu.au/mdi/mutnomen/).

Correct variants nomenclature in publications and database submissions is an essential basis for the quality control that each scientific database annotator should undertake before including variation data in a database. In particular, errors due to wrong assignment of base position, incorrectly deduced amino acid changes, erroneous use of the one-letter amino acid code, and typographical mistakes can be recognized if the name of the variant follows the basic nomenclature rules [18, 19]: (1) variants are best described at the DNA level, using either a genomic or a cDNA reference sequence with the accession and version number from a primary sequence database; (2) the description should be preceded by a letter indicating the type of reference sequence used ("g" for genomic, "c" for cDNA, "m" for mitochondrial,); (3) the nucleotide numbering should designate the A of the ATG of the initiator methionine as nucleotide +1; (4) it should be clearly indicated at which level sequence variation has been analyzed: "r" or "p" should precede variant descriptions if RNA or protein sequences have been analyzed. For example, given the cDNA sequence with accession number AF151980 and version AF151980.2, the designation *c.226C>A* indicates the substitution of cytosine at nucleotide position 226 by an adenosine, corresponding to the deduced amino acid change *p.Arg76Ser (p.R76S)* [20].

A curation tool for checking consistency between described nucleotide changes and corresponding deduced amino acid changes [21] is available online (http://www.ebi.ac.uk/cgi-bin/mutations/check.cgi).

A second major difficulty in the annotation process is to describe the effect of a sequence variation. Establishing a disease association, especially for missense variants, may not be an easy task, as recently outlined [22, 23]. A few criteria can help in designating a missense variant as disease-causing:

- The extent of DNA analysis: the entire gene should be analyzed to rule out the possibility of missing the actual relevant mutation.
- Segregation of the disease phenotype with the mutation within pedigrees is a good indication of pathogenicity.
- The type of missense change: substitutions of a residue by amino acids of different physical character are more likely to influence protein structure, particularly if they affect residues conserved among species.
- The frequency of the variant, since a mutation occurring at a frequency less than 1% in the population is rare and may be disease-causing.
- Functional analysis of the variant is the definitive test to confirm its phenotypic effect.

The retinal-specific ATP-binding cassette transporter gives a good example of the complexity of this issue. The corresponding gene (*ABCA4* or *ABCR*) is mutated in retinal disorders [24–27]. However, while the association with Stargardt disease [28, 29], cone-rod dystrophy [25, 26, 30, 31], and retinitis pigmentosa [30] is widely accepted, the role of *ABCR* mutations in age-related macular dystrophy is still unclear due to studies that either attest [32, 33] or confute [34] association with the disease. The annotation process in such cases should consist of the objective and complete reporting of the conflicting data, providing database users with references from which the information has been obtained.

21.3 Databases

Many databases that describe sequence variations, in the form of mutations and polymorphisms, and their relationship to phenotype or disease are currently freely available from the Internet. It is expected that medical research will benefit greatly from these resources. To make the best use of the data, it is essential for a user to know what is available and how to approach the overwhelming data space. In this section, we aim to provide a brief description of the main databases documenting human variation and discuss their utility and limitations. Some techniques of search will also be suggested.

21.3.1 Central Databases

In general, there are two main types of mutation database: (1) central databases, which contain information on variation across the whole genome [35]; and (2) specialized databases, which concentrate on variations within a single gene or in genes associated with a single phenotype.

If the central databases always contain a large amount of information, they can differ from each other in several aspects. First, the source of data and how they are accepted in the database can vary. Some databases are intensively curated and may contain only peer-reviewed, published information. Others accept submitted data directly from researchers and do little validation. There are also some databases that only collect data from many sources and present them with a common format at a single site. The quality of the databases and the degree of validity of the presented data is therefore variable. Secondly, the type of information provided can differ. For example, some databases may put their focus on disease phenotype and clinical relevance, while others record the biochemical consequences of the variations.

In the following, we provide examples of some major central databases in which human mutation information can be documented.

21.3.1.1 Online Mendelian Inheritance in Man (OMIM)

OMIM (http://www.ncbi.nlm.nih.gov/omim/) is a knowledgebase of human genes and genetic disorders [36, 37]. Originally, it was based on the textbook version of MIM [38]. OMIM entries are created for each unique genetic disorder or gene for which sufficient information exists. Each OMIM entry has a full-text summary of a genetically determined phenotype and/or gene and has numerous links to PubMed references, genetic databases (DNA and protein sequence), or other central and specialized databases. It also provides a highly detailed map-viewer and information on inheritance patterns. OMIM can be searched by title, OMIM number, allelic variants, text, references, clinical synopsis, gene map disorder, and contributors.

In terms of information on allelic variants, OMIM does not contain all known variations in a gene, but concentrates on the disease-causing mutations and some most common polymorphisms. It should also be noted that the variants are germline mutations and are distinct from somatic mutations, as seen in non-familial cases of cancer, which are generally not included. Allelic variants can be investigated for each entry under the "Allelic variants" section of the "Table of contents" of OMIM. This section contains a list of variants, which includes a brief description of the mutation in terms of amino acid change. It also describes the discovery and clinical details of a particular mutation.

Besides the richness of textual information provided in each entry, one of the major advantages of the OMIM database is that it can serve as a gateway to a wealth of genetic databases through its relevant links. However, the current textual nature of OMIM makes automatic parsing of information difficult. Furthermore, OMIM is mainly directed to geneticists and is therefore not specifically con-

cerned with sequence-oriented data. This results in a low correspondence between allelic variants and sequence information and makes it difficult to faithfully map allelic variants onto a sequence.

21.3.1.2 The Human Gene Mutation Database (HGMD)

The Human Gene Mutation Database [39, 40] (http://archive.uwcm.ac.uk/uwcm/mg/hgmd0.html) is a collection of published germline mutations associated with inherited diseases [41]. HGMD comprises single base-pair substitutions in coding, regulatory, and splicing regions of human nuclear genes, as well as deletions, duplications, insertions, repeat expansions, combined micro insertions and deletions (indels), and complex rearrangements. HGMD usually only includes mutations with obvious phenotypic consequences (disease-causing mutations). However, some polymorphisms with a statistically significant association to disease have been included on the basis that they significantly reduce the expression of a given gene or they alter the functional activity of its protein product. HGMD can be searched either by disease, gene name, or gene symbol.

All HGMD entries have a unique accession number. Each entry provides a reference to the first published report of a mutation, the number of entries by mutation type, associated disease, the gene name, symbol, and chromosomal location. HGMD also provides links to OMIM, GDB, GenAtlas, the HUGO nomenclature, and specialized databases.

21.3.1.3 The SNP Databases

Apart from OMIM and HGMD, where the database's focus is on disease-associated variation, high-throughput SNP data can be documented via the "SNP databases". In general, these databases serve as repositories and do not provide, at least for the moment, additional information about associated phenotype or biological significance. Two major databases of this type are dbSNP and the Human Genome Variation database (HGVbase).

DbSNP (http://www.ncbi.nlm.nih.gov/SNP/) is a repository for single-base nucleotide substitution, short deletion and insertion polymorphisms, and microsatellite repeats from all species [42, 43]. It should be noted that in dbSNP, the use of the term "polymorphism" does not imply minimum allele frequencies (1% of the population) for SNPs. DbSNP contains about 3 million human SNPs as well as 0.5 million from other organisms. The Web-based interface allows flexible searches by gene name and by cross-reference to other databases, such as OMIM or structure databases.

HGVbase [44] (http://hgvbase.cgb.ki.se/) was constructed to provide a non-redundant collection of all known, both functionally important and silent human DNA variants. Data are derived from different sources such as publications, batch submissions, or information from other Web-based databases. They are integrated in HGVbase on the basis of extensive curation. Records include neutral polymorphisms as well as disease-related mutations. They are grouped in proven and

suspected variants. Generally, variants detected only by computational methods are classified as "suspected". The tag "proven" is used for records that have been experimentally confirmed. New data can be submitted directly to the database by researchers using a Web page submission form. The data are then indexed and presented using the EBI SRS system (http://srs.ebi.ac.uk/) so that search can be performed by SNP identifier (ID), gene name, gene symbol, variation type and so forth. Each HGVbase entry is presented in the context of its surrounding sequence and mapped to the genome sequence. Population allele frequencies, when available, are also presented. In addition, the record provides information about the submitter and publications, genotyping assays and functional prediction, and links to GenBank and other databases.

SNP-related information can also be retrieved from the SNP consortium website [45], from the JSNP repository of Japanese SNP data [46], from the Seattle SNPs website (http://pga.mbt.washington.edu/), or from the ALFRED [47, 48], which is designed to store and disseminate frequencies of alleles at human autosomal polymorphic sites and multiple defined population samples. Some non-mutation-oriented databases, such as the Genome Database (GDB) [49] or the Swiss-Prot Protein Knowledgebase can also provide valuable information on variants.

21.3.1.4 **Advantages and Drawbacks of Central Databases**

The major advantage of central databases lies in the fact that they usually offer a consistent and well-maintained interface to access a large amount of variation data across the whole genome. As these databases often receive direct funding, the database design can reach a greater sophistication and consequently, flexible search options, together with good data visualization tools, can be made available. As a common base of information about all variations, a centralized database may also allow statistical analysis to be conducted for a spectrum of variations [35].

However, while striving to provide a large amount of data, the central databases fail to have specialized knowledge of any particular locus. The curators are less likely to have close ties to a particular research community and, therefore, data produced in small laboratories or more specific information may be missed out from these databases.

21.3.2 **Specialized Databases**

For many years now, efforts have been made to create specialized databases. These efforts are of tremendous value to the biological research community, since they help to organize the massive amounts of data into specific, tractable research areas. Many more types of data, such as molecular structure and function, macromolecular interactions, geographic location of the allele, can be included in these specific databases. Two subcategories of specialized databases can be distinguished, based on their focus: gene-oriented, locus-specific databases (LSDBs) (Table 21.1) [50], and disease-oriented databases (Table 21.2).

Table 21.1 Examples of locus-specific databases.

Database name	*Website*	*Data source*	*Content*	*Links* [a)]
Androgen receptor gene mutations database [54]	http://www.mcgill.ca/androgendb/	Published articles; curated	Mutations in androgen receptor gene Phenotype description Functional effects References Information on androgen receptor interacting proteins	PubMed Swiss-Prot EMBL Indexed by SRS
BTKbase [55]	http://bioinf.uta.fi/BTKbase 758 entries Last update: September 2002	Direct electronic submissions by authors	Mutations and polymorphisms in Bruton's tyrosine kinase gene Phenotype description; population data Technical data Textual description of protein domains Information on mutations effect References Special feature: mutation checker	OMIM EMBL Swiss-Prot Swiss-Change PubMed Indexed by SRS
Cystic fibrosis mutation database	http://www.genet.sickkids.on.ca/cftr/	Submissions by authors before publication	Mutations and polymorphisms in the cystic fibrosis transmembrane conductance regulator Phenotype description; population data References	OMIM GDB Indexed by SRS
G6PD mutations [62]	http://www.rubic.rdg.ac.uk/g6pd/ 193 entries Last update: December 2001	Published articles; favism website; electronic submission; manual validation	Mutations in G6PD gene Phenotype classification 3D structural data References	Indexed by SRS

Table 21.1 (cont.)

Database name	*Website*	*Data source*	*Content*	*Links* [a)]
Haemophilia B mutation database	http://www.kcl.ac.uk/ip/petergreen/haemBdatabase.html 2511 patient entries Version 12, 2003	Direct submissions to coordinators	Mutations and polymorphisms in factor IX gene Phenotype description Heterogeneous comments References	Indexed by SRS
Human PAX6 allelic variant databases [56]	http://pax6.hgu.mrc.ac.uk/ 228 record Last update: January 2003	Published articles and electronic data submissions; automatic curation (MuStar software)	Mutations and polymorphisms in the paired box protein Pax-6 Phenotype description Information on protein domains Information on mutations effect References	OMIM GDB GeneCards HGMD EMBL Swiss-Prot PubMed Indexed by SRS
Human type I and III collagen mutation database [57, 58]	http://www.le.ac.uk/genetics/collagen/ Last update: March 2003 No statistics available online	Published articles; curated	Mutations and polymorphisms in *COL1A1*, *COL1A2*, and *COL3A1* Phenotype description References	OMIM PubMed
KinMutBase [59]	http://www.uta.fi/imt/bioinfo/KinMutBase/ 320 entries Last update: December 2001	Direct submissions to curators	Mutations in protein kinases genes Phenotype description References	RefSeq Swiss-Prot

Table 21.1 (cont.)

Database name	*Website*	*Data source*	*Content*	*Links* [a]
NRMD [63]	http://cmbipc60.cmbi.kun.nl:8080/cgi-bin2/nrmd/nrmd.py 1095 entries Last update: November 2002	Published articles; Swiss-Prot and other Web-based resources; limited curation	Mutations in nuclear receptors Phenotype description Position in structural alignment	Swiss-Prot NucleaRDB OMIM VDR, PNR, GRR pages PubMed
PAHdb [60, 61]	http://www.pahdb.mcgill.ca/	Published articles and personal communications; manual curation	Mutations and polymorphisms in the phenylalanine-4-hydroxylase gene Phenotype description; clinical information Population data; haplotype data Mouse models; *in vitro* expression analysis results Molecular modeling data References	OMIM GeneBank GeneCards Swiss-Prot Indexed by SRS
PHEXdb	http://www.phexdb.mcgill.ca/		Mutations in *PHEX* gene Phenotype description Information on mutation effects References	OMIM
TGRAP	http://tGRAP.uit.no Entries: 10500 Last update: April 2001	Published articles; curation	Mutations in G protein-coupled receptors References	Swiss-Prot OMIM PubMed
VWF database	http://www.shef.ac.uk/vwf/ Continual update No statistics available	Electronic submissions	Mutations and polymorphisms in von Willebrand factor Phenotype description; functional studies References	Indexed by SRS

a) Only clickable links from an entry to a corresponding relevant entry of another database are described here. Links that are provided by the website to other websites are not included.

Table 21.2 Examples of disease-oriented databases.

Database name	*Website*	*Data source*	*Content*	*Links* [a)]
Alzheimer disease mutation database	http://molgen-www.uia.ac.be/ADMutations/	Published articles and direct submissions	Mutations in genes involved in Alzheimer's disease Phenotype descriptions Family details References	OMIM GDB GeneBank Swiss-Prot PubMed
eMelanoBase [64]	http://www.wmi.usyd.edu.au:8080/melanoma.html	Submissions by E-mail; curated	Germline variants in genes involved in familial melanoma susceptibility Population and geopolitical data References	NCBI nucleotide NCBI protein OMIM Ensembl PubMed
Gene connection for the heart	http://pc4.fsm.it:81/cardmoc/ Regular update	Database maintained and regularly updated under the supervision of editors in a European-wide collaborative effort	Linkage data and mutations underlying inherited arrhythmias Phenotype descriptions *In vitro* expression data References	
HbVar [65]	http://globin.cse.psu.edu/globin/hbvar/m enu.html 1165 entries (May 2003) Continual update	Books on hemoglobin variants and thalassemia; curators for additional entries	Hemoglobin variants and thalassemia causing mutations Population data; clinical information Stability data Detection methods	PubMed Indexed by SRS

Table 21.2 (cont.)

Database name	*Website*	*Data source*	*Content*	*Links* [a]
INFEVERS [66]	http://fmf.igh.cnrs.fr/infevers/	Electronic submissions; curated	Mutations and polymorphisms in genes responsible for hereditary inflammatory disorders Population data Phenotype description Detection methods References	PubMed
Leiden Muscular Dystrophy Pages	http://www.dmd.nl/	Published articles and direct submissions	Mutations and polymorphisms in genes involved in muscular dystrophies Gene and protein descriptions Phenotypes description Information on molecular diagnosis techniques References	OMIM LocusLink GDB HGMD PubMed Indexed by SRS
NCL mutations	http://www.ucl.ac.uk/ncl/ Last update: October 2002	Published articles and personal communications	Mutations and polymorphisms in genes involved in neuronal ceroid lipofuscinoses Phenotype description Information on mutation effects References	Indexed by SRS

a) Only clickable links from an entry to a corresponding relevant entry of another database are described here. Links that are provided by the website to other websites are not included.

21.3.2.1 An Example of a Locus-specific Database: the IARC TP53 Database

The IARC TP53 mutation dataset [51] (http://www.iarc.fr/p53/index.html) is the largest data set available on the variations of any human gene [52]. The database includes all somatic and germline *p53* mutations, as well as polymorphisms, that have been reported in the published literature since 1989. The IARC TP53 database is extensively annotated, in particular with respect to the description of tumor pathologies. It is interesting to note that the "central unit" of the database schema is the cancer patient whose tumor contains a mutation. Therefore, clinical information, as well as information on individual exposure and risk factors is provided for each individual whenever possible. For germline mutations, the data also include the description of the family, information on each family member, information on tumor samples, and reference to the publication in which the family is described.

The database can be searched via a Web-based application. The user can identify and select specific sets of data based on queries. The results, such as mutation patterns, codon distribution and tumor spectrum, can be visualized in the form of either graphs or tables. The entire data set, or sets of data selected according to the user's queries, can be downloaded as tab-limited text. In addition, the website offers a comprehensive user guide, a slide shown on p53 mutations as well as many links to other websites or database entries on TP53.

Apart from the IARC TP53 database, other p53 databases are also available. For example, the UMD-p53 database (http://p53.curie.fr/index.html) [53]. Although it has been developed independently from the IARC TP53 database, the UMD-p53 database contains essentially similar entries derived from the published literature. The two databases differ in their format, their annotations (more extensive in the IARC database), and the design of their respective Web-search applications.

21.3.2.2 An Example of a Disease-oriented Specialized Database: Retina International's Scientific Newsletter – Mutation Database

In contrast to "gene-oriented" specialized databases that focus on gathering mutant information for a single locus, the "disease-oriented" databases contain all variants implicated in a particular disease and these variants can be from different loci.

The Retina International's Scientific Newsletter (http://www.retina-international.com/sci-news/database.htm) provides an up-to-date synopsis of genes and mutations underlying retinal disorders. The data are organized in three different databases. In the mutation database, the mutations are sorted by gene and are presented following the official mutation nomenclature recommendations [18, 19]. The protein database offers an overview of all proteins directly or indirectly involved in the vision process and in the maintenance of the integrity of retinal cells and tissues. Information derived from the literature and links to other databases (GeneBank, OMIM, LocusLink, Swiss-Prot, MEDLINE) are also provided. The disease database is a collection of various retinal phenotypes and provides information on chromosomal location and the markers linked to a given locus. Finally, an animal model database presents a description of animal models for inherited retinal disorders.

21.3.2.3 Other Locus-specific Databases

There are numerous other LSDBs available on the Internet. Some examples [50, 54–66] are listed in Tables 21.1 and 21.2. It is possible to obtain a nearly exhaustive list on the following website: http://www.expasy.org/alinks.html#Human_mut. As the LSDBs were developed independently of each other, they have very different content and structure depending on gene and disease characteristics. In a recent study, a total of 94 independent websites that describe mutations associated with human disease were evaluated according to 80 content criteria defined by the authors [50]. It was found that 54% of the LSDBs studied were easy to use, whereas 11% were hard to follow. The majority of the databases (73%) were displayed through HTML and a certain redundancy can be found among the databases. Not surprisingly, there was extreme heterogeneity between databases and the type of criteria met by them. Clearly, this study provided a strong case for having a certain uniformity of data among the databases in order to make the content as useful as possible.

21.3.2.4 Advantages and Drawbacks of Specialized Databases

Specialized databases have several advantages. First, they often present the best state of knowledge in a particular area, and information for each variant is provided in greater depth. This is because specialized databases are usually set up, maintained, and curated by expert(s) with a particular interest in the gene(s) or field in question. They are consequently very well adapted and responsive to the needs of the corresponding research field. Secondly, they allow a large range of information to be retrieved for a particular gene. This information would otherwise have to be collected from multiple sites.

However, because of their smaller size and usually a lack of direct funding, specialized databases may become stagnant and/or disappear.

21.3.3 The Swiss-Prot Protein Knowledgebase and Information on Disease and Sequence Variations

In this section, a detailed description of the Swiss-Prot Protein Knowledgebase for the use of proteomic medical research will be provided. Although Swiss-Prot is not a disease or mutation-centered database, it stores a wealth of information of interest to the medical community.

Swiss-Prot [67] was created in 1986 at the Department of Medical Biochemistry of the University of Geneva. It is now an equal partnership between the European Molecular Biology Laboratory (EMBL) and the Swiss Institute of Bioinformatics (SIB). The database contains data originating from a wide variety of biological organisms. Currently (Release 41.15 of 03.07.2003), there are a total of about 130 000 annotated entries from more than 8200 different species. A particular effort, however, is put on the annotation of all known human protein sequences and their mammalian orthologues (The Human Proteomics Initiative) (http://

NiceProt View of Swiss-Prot: P06400

Printer-friendly view | Quick BlastP search

General information about the entry	
Entry name	**RB_HUMAN**
Primary accession number	**P06400**
Secondary accession number	P78499
Entered in Swiss-Prot in	Release 06, January 1988
Sequence was last modified in	Release 13, January 1990
Annotations were last modified in	Release 42, September 2003
Name and origin of the protein	
Protein name	**Retinoblastoma-associated protein**
Synonyms	**PP110** **P105-RB** **RB**
Gene name	**RB1**
From	Homo sapiens (Human) [TaxID: 9606]
Taxonomy	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

References

[1] SEQUENCE FROM NUCLEIC ACID.
MEDLINE=88014238; PubMed=3657987; [NCBI, ExPASy, EBI, Israel, Japan]
Lee W.-H., Shew J.-Y., Hong F.D., Sery T.W., Donoso L.A., Young L.-J., Bookstein R., Lee E.Y.-H.P.;
"The retinoblastoma susceptibility gene encodes a nuclear phosphoprotein associated with DNA binding activity.";
Nature 329:642-645(1987).

[2] REVISIONS.
MEDLINE=87149066; PubMed=3823889; [NCBI, ExPASy, EBI, Israel, Japan]
Lee W.-H., Bookstein R., Hong F.D., Young L.-J., Shew J.-Y., Lee E.Y.-H.P.;
"Human retinoblastoma susceptibility gene: cloning, identification, and sequence.";
Science 235:1394-1399(1987).

[3] .....

[25] VARIANT RB GLU-310.
MEDLINE=99138683; PubMed=9973307; [NCBI, ExPASy, EBI, Israel, Japan]
Klutz M., Horsthemke B., Lohmann D.R.;
"RB1 gene mutations in peripheral blood DNA of patients with isolated unilateral retinoblastoma.";
Am. J. Hum. Genet. 64:667-668(1999).

[26] VARIANTS RB GLY-500 AND GLU-616.
MEDLINE=21415586; PubMed=11524739; [NCBI, ExPASy, EBI, Israel, Japan]
Yu Y.S., Kim I.-J., Ku J.-L., Park J.-G.;
"Identification of four novel RB1 germline mutations in Korean retinoblastoma patients.";
Hum. Mutat. 18:252-252(2001).

Figure 21.1 Excerpt from Swiss-Prot entry P06400 (Nice-Prot format). Various reference blocks, cross-references, variants in the feature table, and part of the sequence have been deleted for the sake of simplicity.

www.expasy.org/sprot/hpi/) [68]. In release 41.15 (03.07.2003), there were 9503 annotated human sequences. These entries were associated with 24448 literature references; 23378 experimental or predicted post-translational modifications (PTMs), 3222 splice variants and 15770 variants. The majority (about 60%) of the variants annotated in Swiss-Prot are linked with disease states.

Comments

- ***FUNCTION***: Probably acts as a regulator of other genes. Forms a complex with adenovirus E1A and with SV40 large T antigen. Acts as a tumor suppressor. May bind and modulate functionally certain cellular proteins with which T and E1A compete for pocket binding. Potent inhibitor of E2F-mediated trans-activation. Recruits and targets histone methyltransferase SUV39H1 leading to epigenetic transcriptional repression.
- ***SUBUNIT***: Interacts preferentially with transcription factor E2F1. Interacts with DNMT1. The not phosphorylated form interacts with SUV39H1.
- ***SUBCELLULAR LOCATION***: Nuclear.
- ***PTM***: Phosphorylated from S to M phase of the cell cycle and is dephosphorylated in G1. T, but not E1A, binds only to the unphosphorylated form.
- ***DISEASE***: Defects in RB1 are the cause of childhood cancer retinoblastoma (RB)[MIM:180200]. RB is a congenital malignant tumor that arises from the nuclear layers of the retina. It occurs in about 1:20'000 live births and represents about 2% of childhood malignancies. It is bilateral in about 30% of cases. Although most RB appear sporadically, about 20% are transmitted as an autosomal dominant trait with incomplete penetrance. The diagnosis is usually made before the age of 2 years when strabismus or a gray to yellow reflex from pupil ("cat eye") is investigated.
- ***DISEASE***: Defects in RB1 are a cause of osteogenic sarcoma [MIM:259500].
- ***SIMILARITY***: BELONGS TO THE RETINOBLASTOMA PROTEIN (RB) FAMILY.
- ***DATABASE***: NAME=RB1base; NOTE=RB1 mutation db; WWW="http://www.d-lohmann.de/Rb/mutations.html".
- ***DATABASE***: NAME=Atlas Genet. Cytogenet. Oncol. Haematol.; WWW="http://www.infobiogen.fr/services/chromcancer/Genes/RB1ID90.html".

Copyright

Cross-references

EMBL	L41870; AAB59465.1; -. [EMBL / GenBank / DDBJ] [CoDingSequence] M27866; AAA53484.1; -.[EMBL / GenBank / DDBJ] [CoDingSequence] .....
PIR	JS0276; RBHU.
PDB	1AD6; 26-AUG-98. [ExPASy / RCSB] 1GUX; 13-JAN-99. [ExPASy / RCSB] 1GH6; 15-NOV-01.[ExPASy / RCSB] 1H25; 01-FEB-03. [ExPASy / RCSB] Detailed list of linked structures.
TRANSFAC	T00722; -.
Genew	HGNC:9884; RB1.
CleanEx	HGNC:9884; RB1.
MIM	180200 [NCBI / EBI]. 109800 [NCBI / EBI]. 259500 [NCBI / EBI].
GeneCards	RB1.
GeneLynx	RB1; Homo sapiens.
GO	GO:0005717;Cellular component: chromatin *(traceable author statement)*. GO:0005634;Cellular component: nuclear chromosome *(traceable author statement)*. GO:0003700;Molecular function: transcription factor activity *(traceable author statement)*. GO:0000075;Biological process: cell cycle checkpoint *(traceable author statement)*. GO:0000122;Biological process: negative regulation of transcription from Pol II promoter *(traceable author statement)*.
.....	.....

Figure 21.1 (continued)

A detailed description of the structure of a Swiss-Prot entry (Fig. 21.1), and of the type of information annotated in the different line types, is available from the Swiss-Prot user manual: http://www.expasy.org/sprot/userman.html. Information on diseases and the characteristics of the variants can be found in various sections of Swiss-Prot, as discussed below.

Keywords

Transcription regulation; **DNA-binding**; **Nuclear protein**; **Phosphorylation**; **Anti-oncogene**; **Disease mutation**; **3D-structure**.

Features

Feature table viewer Feature aligner

Key	From	To	Length	Description	FTId
DOMAIN	10	18	9	POLY-ALA.	
DOMAIN	20	29	10	POLY-PRO.	
DOMAIN	373	771	399	POCKET (BINDS T AND E1A).	
DOMAIN	373	579	207	DOMAIN A.	
DOMAIN	580	639	60	SPACER.	
DOMAIN	640	771	132	DOMAIN B.	
MOD_RES	249	249		PHOSPHORYLATION (BY CDC2).	
MOD_RES	252	252		PHOSPHORYLATION (BY CDC2).	
MOD_RES	373	373		PHOSPHORYLATION (BY CDC2).	
MOD_RES	807	807		PHOSPHORYLATION (BY CDC2).	
MOD_RES	811	811		PHOSPHORYLATION (BY CDC2).	
VARIANT	72	72		E -> Q (IN RB).	VAR_005572
VARIANT	137	137		E -> D (IN RB; UNILATERAL FORM).	VAR_005573
VARIANT	185	185		I -> T (IN RB).	VAR_005574
VARIANT	310	310		G -> E (IN RB; COULD BE A POLYMORPHISM).	VAR_010045
VARIANT	358	358		R -> G (IN RB).	VAR_010046
VARIANT	358	358		R -> Q (IN RB).	VAR_005575
VARIANT	404	404		F -> K (IN RB).	VAR_010047
VARIANT	447	447		K -> Q (IN RB).	VAR_010048
VARIANT	457	457		M -> R (IN RB).	VAR_005576
VARIANT	480	480		MISSING (IN RB; MILD FORM).	VAR_005577
VARIANT	500	500		R -> G (IN RB).	VAR_011580
VARIANT	530	530		K -> R (IN RB).	VAR_010049

.....

Sequence information

Length: **928 AA** Molecular weight: **106159 Da** CRC64: **C8E746111E19CC32** [This is a checksum on the sequence]

```
        10         20         30         40         50         60
         |          |          |          |          |          |
MPPKTPRKTA ATAAAAAAEP PAPPPPPPPE EDPEQDSGPE DLPLVRLEFE ETEEPDFTAL

        70         80         90        100        110        120
         |          |          |          |          |          |
CQKLKIPDHV RERAWLTWEK VSSVDGVLGG YIQKKKELWG ICIFIAAVDL DEMSFTFTEL

       130        140        150        160        170        180
         |          |          |          |          |          |
QKNIEISVHK FFNLLKEIDT STKVDNAMSR LLKKYDVLFA LFSKLERTCE LIYLTQPSSS

       190        200        210        220        230        240
         |          |          |          |          |          |
ISTEINSALV LKVSWITFLL AKGEVLQMED DLVISFQLML CVLDYFIKLS PPMLLKEPYK

       250        260        270        280        290        300
         |          |          |          |          |          |
TAVIPINGSP RTPRRGQNRS ARIAKQLEND TRIIEVLCKE HECNIDEVKN VYFKNFIPFM

       310
         |
NSLGLVTSNG ..........
```

P06400 in FASTA format

Figure 21.1 (continued)

21.3.3.1 Gene Names

In Swiss-Prot, the name of the gene coding for a specific protein is indicated in a line-type called "GN" (Gene Name). Swiss-Prot uses the gene symbol proposed by the HUGO Gene Nomenclature Committee (HGNC) [69] whenever possible in order to avoid confusion in gene nomenclature, and to facilitate information retrieval from public databases. It also lists, if they exist, symbols that were previously used in literature or database reports.

21.3.3.2 Description of Diseases

For proteins involved in disease, the description of the disease can be found in the comment lines ("CC") prefixed by the token "DISEASE". Disease descriptions can be very concise, but links to corresponding OMIM entries are provided, allowing the user to retrieve more detailed information. Diseases associated with chromosomal translocations are also described, and, if known, information on the translocation break point is indicated.

Another comment token, "POLYMORPHISM" contains data on alleles and polymorphisms not associated with a specific disease.

21.3.3.3 Proteins as Therapeutic Drugs

Swiss-Prot stores information relevant to the use of specific proteins as therapeutic drugs (pharmaceuticals). For such proteins, the brand name(s), the name(s) of the company(ies) that develops and/or sells the drug(s) and a brief description of the therapeutic use(s) of the protein are provided. This is implemented through a comment topic called "PHARMACEUTICAL". It should also be noted that the generic name of a protein drug is recorded in the description lines ("DE") of the relevant entry. While some generic names correspond to the main scientific name (example: insulin, somatotropin), for some others the generic name is one of the synonyms of the official protein name.

21.3.3.4 Data on Variants

One of the major aims of Swiss-Prot is to provide the users with non-redundant data. Therefore, in contrast to other often automatically created databases or repositories which maintain one sequence record per unique sequence and can consequently have up to hundreds of records corresponding to the same gene, Swiss-Prot chooses to display the most common isoform of a protein as its "master" sequence. All variations related to this "master" sequence are annotated in the feature tables using diverse keys according to the origin of the difference (alternative splicing, polymorphism, conflictual sequencing results).

Genetic variation data are stored in the feature table ("FT" lines) of the relevant entries using the "VARIANT" key. As Swiss-Prot is a "proteocentric" resource, only data on point mutations, small deletions, or insertions are collected. Frame-

shift or nonsense mutations are excluded, as they generally have a deleterious effect on the structure and function of a protein. Each variant annotated in Swiss-Prot is manually checked, and information on disease association and functional effects is provided according to literature reports. If a variant occurs in different disease phenotypes, it is registered only once. However, information on the associated disease phenotypes can be found in the corresponding FT descriptions. Each variant is also given a unique identifier (FTId) to allow a direct link to the relevant entries in disease mutation databases, as well as to provide these databases with a method to implement reciprocal links. While Swiss-Prot is focusing its effort in the annotation of experimentally proven and published variants, sequence differences revealed by sequence alignments are also stored in the database, namely in the feature table using the "CONFLICT" key. Although it cannot be excluded that these data might reflect sequencing errors, they might well correspond to not yet identified variants and can be considered as the Swiss-Prot counterpart of *in silico* detected polymorphisms stored in SNPs repositories such as dbSNP and HGVbase.

Recently, a specific variant web page has been implemented through the ExPASy server [70] to group and provide additional information on each variant (Fig. 21.2) (Yip et al., unpublished data). This page can be accessed via a link provided through the FTId. Apart from general information such as the amino acid change, position and effect of the variant, and its association with diseases, the page provides new structural information of the variant. In particular, when available, 3-D models generated by an automatic homology modeling method can be obtained so that the mutation can be visualized directly on the protein structure (Yip et al., unpublished data). Results of a sequence alignment can also be retrieved to evaluate residue conservation among different species. In addition, the variant web page gives a list of specific literature references, and direct links to relevant dbSNP, HGVbase and OMIM entries.

When compared with the information on missense variants listed in other mutation databases, it should be noted that Swiss-Prot does not simply catalogue amino acid changes predicted from nucleotide variations, but it stores, when available, information on direct protein sequencing and characterization including post-translational modifications (PTMs). This is important, as the real effect of missense variants on proteins, such as PTM and structural phenotype [71], cannot be deduced from simple translation of single nucleotide substitutions at the DNA level. In this regard, Swiss-Prot offers links to proteomic databases, for example Swiss_2DPAGE (http://www.expasy.org/ch2d/) [72] and Siena_2DPAGE (http://www.bio-mol.unisi.it/2d/2d.html), and can be regarded as an integration platform between genomics and proteomics.

Finally, it should be noted that while the concept of redundancy minimization and the organization of variant data in feature tables have many advantages for human readers, it may represent a certain danger to computer programs which only look at the "master" sequence and disregard the annotation in the feature tables. Similarity or motif search programs, such as BLAST [73] or ScanProsite [74], can miss alternatively spliced isoforms, or disease state isoforms with a high de-

Swiss-Prot variant: VAR_011580 in P06400

General information	
SwissProt ID (AC)	**RB_HUMAN (P06400)**
Gene symbol(s)	**Official:** RB1
Chromosomal location	13q14.2
Genew	HGNC: 9884
Protein name	Retinoblastoma-associated protein
Length of the protein	928
Information on the variant	
FTId	**VAR_011580**
Amino acid position of the variant	500
Residue change	**From Arg (R) to Gly (G), R500G**
Status (Disease, polymorphism or unclassified)	Disease
Disease	**Childhood cancer retinoblastoma (RB)** Defects in RB1 are the cause of childhood cancer retinoblastoma (RB) [MIM:180200]. RB is a congenital malignant tumor that arises from the nuclear layers of the retina. It occurs in about 1:20'000 live births and represents about 2% of childhood malignancies. It is bilateral in about 30% of cases. Although most RB appear sporadically, about 20% are transmitted as an autosomal dominant trait with incomplete penetrance. The diagnosis is usually made before the age of 2 years when strabismus or a gray to yellow reflex from pupil ("cat eye") is investigated
Comment	None
Structural information on the variant	
Location on the sequence	480 NIFHMSLLACALEVVMATYS R STSQNLDSGTDLSFPWILNV 520 ↓ G
Protein features in neighborhood	**Key From To Length Description** DOMAIN 373 771 399 POCKET (BINDS T AND E1A) DOMAIN 373 579 207 DOMAIN A
Residue conservation	Alignment from Blast search
Physico-chemical property	Change from large size and basic (R) to glycine (G)
3D homology model	**Model Visualization** | **Template Structure** ExPASy/ AstexViewer | 1AD6 ***Disclaimer:*** The result of any modelling procedure is non-experimental and must be considered with care. This is especially true since there is no human intervention during model building process.
References for the variant	
[1] VARIANTS RB GLY-500 AND GLU-616. MEDLINE=21415586; PubMed=11524739; [NCBI, ExPASy, EBI, Israel, Japan] Yu Y.S., Kim I.-J., Ku J.-L., Park J.-G. "Identification of four novel RB1 germline mutations in Korean retinoblastoma patients." Hum. Mutat. 18:252-252(2001)	
Cross-references for the variant	
OMIM	180200 [NCBI / EBI].
dbSNP	Not available
Specialized database	NAME=RB1base; NOTE=RB1 mutation db; WWW="http://www.d-lohmann.de/Rb/mutations.html" NAME=Atlas Genet. Cytogenet. Oncol. Haematol.; WWW="http://www.infobiogen.fr/services/chromcancer/Genes/RB1ID90.html"

Figure 21.2 Example of a web page for a Swiss-Prot variant (VAR_011580 in Swiss-Prot entry P06400). The structures of the homology models can be obtained via a hyperlink in the "3D homology model" section.

gree of sequence variation. In order to remedy this problem, the program var_splic.pl [75] was written to reconstruct alternative sequences as annotated in the feature tables of Swiss-Prot and its computer-annotated supplement TrEMBL. Var_splic.pl was conceived in particular for alternative splice isoform sequences, but it can also be used to derive alternative sequences resulting from polymorphisms or disease variants. The FASTA-formatted sequences obtained in this manner can be downloaded from the ExPASy and EBI FTP servers. The program is available free of charge from ftp://ftp.ebi.ac.uk/pub/software/swissprot/. When searching Swiss-Prot or TrEMBL, most sequence search tools on the ExPASy and EBI servers take into account these isoform sequences.

21.3.3.5 **Cross-references**

Swiss-Prot is directly cross-referenced [76] to about 50 other databases, such as the EMBL/GenBank/DDBJ international nucleotide sequence database [77], the PDB structural database [78, 79], OMIM, MEDLINE/PubMed, the human gene nomenclature database Genew [80], which stores all HGNC-approved gene symbols, the Gene Ontology database (GO) [81], and many others. GO is a database of controlled terms for the description of the molecular function, biological process, and cellular component of gene products. Most of the cross-references from Swiss-Prot to other databases are stored in a line type called "DR". However, a direct link from the relevant protein entries to specific mutation databases can be found within a comment topic called "DATABASE" (Fig. 21.1). In addition to these explicit cross-references hard coded in the raw Swiss-Prot entry, human sequence entries also have implicit links to databases such as GeneCards [82] and GeneLynx [83]. Implicit links are not annotated in the "DR" lines of the raw Swiss-Prot entry, but are added automatically when viewing the entry on the ExPASy server. GeneCards and GeneLynx are meta-databases of human genes, which offer concise information about the function of the gene products as well as their involvement in diseases. Both are created by automatically extracting relevant information from diverse internet resources, including Swiss-Prot, MIM, GenAtlas, and GDB.

On the ExPASy web server, all explicit and implicit cross-references are implemented as hypertext links that allow users to seamlessly browse from one database to another.

21.3.3.6 **Medical-oriented Keywords**

In order to facilitate the retrieval of data on medically relevant proteins, some specific medical-oriented keywords stored in the "KW" lines have been created in Swiss-Prot. These are: "Disease mutation", which is used for sequences in which there is at least one known disease-linked mutation; "Polymorphism", which is used in each entry containing protein sequence variants for which no disease association has been reported (at the level of the protein sequence), and "Chromosomal translocation", which is used to indicate proteins whose genes are known to

be involved in chromosomal translocations. The keyword "Pharmaceutical" was created to indicate that a protein is used as a therapeutic drug. Other specific keywords such as "Albinism", "Charcot-Marie-Tooth disease", "Deafness", "Diabetes", "Hemophilia", "Phenylketonuria", "Retinitis pigmentosa", and "Xeroderma pigmentosum" were designed for genetic diseases linked with more than a single gene/protein. This list is constantly expanding, and new disease-related keywords are added while relevant Swiss-Prot entries are being annotated. Keywords such as "Allergen", "Anti-oncogene", and "Proto-oncogene" are also available.

In summary, the Swiss-Prot database offers a high-quality annotation and a high level of integration with other biomolecular databases [76]. Swiss-Prot is distributed by anonymous FTP (see http://www.expasy.org/sprot/download.html) and can be accessed through the ExPASy (http://www.expasy.org/sprot/) and EBI (http://www.ebi.ac.uk/swissprot/) servers.

21.3.4 Techniques of Search

With this plethora of databases available through the Internet and each offering specific features [17], the investigator's needs will clearly determine which database(s) is most suitable for use. For example, a clinician may want to find clinical information about patients with mutations. In this case, a database such as OMIM can provide the desired information. However, when a researcher studying a single gene or a disease wants to find more biochemical or experimental data about the mutations, well-curated and maintained LSDBs would be preferable for this purpose. In general, it is usually good to start by searching one or more central databases to obtain some general information about mutations in a gene. Then, by using the data or particular links provided in these databases, one can easily move to the LSDBs for more complete and up-to-date data [35].

It is also recommended that researchers should not limit their studies to the use of one particular database. A recent study assessing the utility of seven international public or private SNPs databases (dbSNP, HGVbase, JSNP, GeneSNP, HOWDY, CGAP-GAI, LEELAB) showed that there was very little overlap among the databases [84]. Therefore, researchers who do not explore the entire information space provided by various databases take the risk of missing out information. Moreover, special care should be taken when using databases that solely rely on *in silico* detection methods of the variants, as they may contain false positives.

A comprehensive interrogation of multiple databases is clearly necessary to take full advantage of the available online resources. Several valuable systems are available that can serve as central hubs for querying across different databases.

At the European Bioinformatics Institute (EBI), resources related to mutation data are indexed under the Sequence Retrieval System (SRS) [85], which is a powerful program for parsing, indexing, querying, viewing, and linking independent textual databases. The SRS system at EBI indexes more than 40 databases containing information on human nuclear gene mutations. These include many LSDBs, OMIM, variation data from the EMBL and Swiss-Prot sequence databases

(EMBLCHANGE and SWISSCHANGE, respectively), and SNP information from dbSNP. In addition, key fields from many LSDBs are combined into a central resource HUMUT, which contains a subset of validated information common to all its component databases with limited redundancy. HUMUT is therefore a federated human mutation database that uses SRS as a tool for analysis. The data in HUMUT are not simply stored in a database, but are organized in a standardized format that allows easy computational analysis of the data. In addition, this database includes cross-links between databases [21, 86].

A similar system is also provided by the NCBI. Entrez is an integrated search and retrieval system for the major databases supported by the NCBI, including PubMed, GenBank, GenPept Sequences, dbSNP, LocusLink, and many others. It enables text searching of databases and provides links to related information from other NCBI-maintained resources [86].

21.3.5 Challenges for Databases

With the completion of the human genome map, the research environment has changed dramatically during the last few years. Sequence variations and their biological consequences were once considered as a field of peripheral interest. They are now widely recognized to be one of the major focuses of the post-genomic era. Consequently, mutation databases should respond to the demands coming from the scientific community.

The first challenge databases are confronted with is to keep up with new data. This is particularly true for central databases that constantly have to deal with a large amount of heterogeneous information. Electronic submission forms may provide an option to encourage authors to directly deposit data. The development of programs for setting up, managing, and curating databases is also becoming a constant priority. These programs should aim to facilitate data retrieval and storage without diminishing data quality [87, 88]. Some databases also strive to develop automatic or semi-automatic annotation tools. In the case of medical annotation of the entries, an automatic approach is less trivial. In fact, for a complete and correct annotation, a large number of documents have to be retrieved and read, and every entry checked manually. As a result, the implementation of advanced data-mining tools may become more and more frequent [89]. Alternatively, interfaces can be built to allow external experts to edit, or participate in the curation process.

The second challenge is to allow the information to be fully interconnected as a whole so that biological knowledge can be derived. Such integration can possibly be achieved on two levels. On the first level, ideally, the various data in different databases should be comprehensively cross-linked so that when starting with an entry from one database, information on the same entry provided by other databases can be accessed easily. On the second level, it should be made possible to query across different databases. Indexing and designing sophisticated data structure, which connects optimally the data objects, could achieve this. Complex

queries for discovering indirect relationships among data could then be carried out through an efficient search and retrieval system. The EBI SRS system is an example of the successful implementation of such a system. In the near future, these two levels of data integration will continue to present a formidable intellectual, technical, and logistical challenge.

A third challenge for databases is related to the information content. At the present time, most LSDBs have been created with the intention of listing gene mutations and directly relating the genotypes to their phenotype expression. However, as noted by Gottlieb et al., variable expressivity due to somatic mutations or mosaicism can cause phenotypic differences [90]. This suggests that it may be necessary to adapt the databases to incorporate additional data on genetic heterogeneity, and variable expressivity. For example, tissue specificity and the timing of the mutation may have important consequences on the phenotype and should be recorded in the databases. At the same time, a clear distinction between germline and somatic mutations should be made for the entries. Apart from adapting the content of the data, it is essential for the different LSDBs to share a minimum of core elements and strive to provide certain uniformity in data structure. From the survey conducted by Claustres et al., it is encouraging to see that the guidelines provided by the HUGO Mutation Database Initiative are being increasingly accepted and used in the design of new databases. For example, 42% of LSDBs use the HUGO-MDI nomenclature, and 29% of LSDBs follow the complete HUGO-MDI guidelines [50].

The last, but not the least important challenge is to assure the quality of the data. Therefore, even in cases where a degree of automation is implemented for information assembly and annotation, it is necessary to have quality control over the input into the databases. Continuous update and validation of the current database content has to be ensured. Clearly, it goes without saying that the quality of the data you get from a database is only as good as the quality of the input data. Quality assurance is thus essential for the retrieval of most updated biologically relevant information from the databases.

It can be envisioned that in the not too far future, central databases and LSDBs will continue to complement each other. Central databases would store a large amount of mutations, each with a minimal set of information. Links to LSDBs, where the major curation job would remain, would provide access to more detailed information for a particular field. Meta-analysis of mutation data across many genes would be available either in central databases or in specially designed retrieval systems.

21.4
Analysis Tools in the Context of Protein Variants

Currently, most information about variants stored in databases stems from high-throughput genomic projects. Changes in amino acid residues are thus often deduced from changes in the nucleotide sequence. However, advances in the field of

proteomics now allow new variants to be identified or deduced nucleotide changes to be confirmed with online proteomic tools. At the same time, there are more analysis and prediction programs being made available through the Internet, which provide researchers with the opportunity to conduct *in silico* analysis of protein variants. The development of these computational tools is essential because they can be employed on a scale that is consistent with the large number of variants being identified.

21.4.1 Proteomic Tools for Protein Identification and the Characterization of Variants

21.4.1.1 Protein Identification Tools

Peptide mass fingerprinting provides a common method to identify proteins [91–93]. The experimental data are a list of peptide mass values obtained from an enzymatic digestion of a protein. The protein is identified by comparing these experimental masses with theoretical masses obtained by an *in silico* digestion of all proteins in a database. This is most effective for the simple (and often theoretical) scenario where the sample is not post-translationally modified, and does not contain any residues that differ from the "master" sequence, e.g. sequence variants. Both types of sequence alterations considerably influence the mass of a peptide, and an identification method based on an identity check of theoretical and experimental masses is very likely to miss out peptides modified in this manner. In these cases, more sophisticated identification and characterization tools are required. In the following enumeration of such tools, we concentrate on programs that are freely available to the scientific community via a Web interface, or can be downloaded free of charge. Others are only mentioned briefly.

Mascot Mascot (http://www.matrixscience.com/cgi/index.pl?page=/search?form?select.html) is a powerful search engine, which uses mass spectrometry data to identify proteins from primary sequence databases [94]. In particular, the Mascot peptide mass fingerprinting tool has a comprehensive database menu, which includes Swiss-Prot (although not TrEMBL). If Swiss-Prot is selected, the tool searches not only the master Swiss-Prot sequences, but also all splice isoforms, annotated variants and conflicts, and combinations of these. These alternative entries can be recognized in the result by their pseudo-accession numbers, which are of the form P12345-01-02-03. The isoform described by this number would possess the splicing variations belonging to the first annotated alternatively spliced isoform (01), the variant features belonging to the second alternative variant form (02), and the conflicts belonging to the third alternative conflict form (03).

SEQUEST The SEQUEST package (http://fields.scripps.edu/sequest/), distributed by Thermo Finnigan, correlates uninterpreted tandem mass spectra of peptides with amino acid sequences from protein and nucleotide databases [95]. A mode to identify amino acid substitutions resulting from SNPs, applicable to both HPLC and microspray tandem mass spectrometry has been investigated, but is not available to the general public.

Additional algorithms have been published that address the improvement of identification of mutated proteins using mass spectrometry: SALSA [96] is a pattern-recognition algorithm developed to screen large numbers of peptide MS-MS spectra for fragmentation characteristics indicative of specific peptide modifications. MS-Convolution and MS-Alignment [97] use the spectral alignment approach. However, to the best of our knowledge, no tools implementing these algorithms are available to the scientific community.

21.4.1.2 **Peptide Characterization Tools**

Besides identification of proteins with variants, several online tools allow detailed peptide characterization and thus help to determine if a variant is present.

FindMod The FindMod tool (http://www.expasy.org/tools/findmod.html) was conceived to identify potential amino acid substitutions (as well as post-translational modifications) in peptides from mass spectrometry experiments [98]. For a protein identified as a potential candidate, theoretical and experimental masses are compared, and their mass differences are then matched with the mass differences induced by all of the possible substitutions of any of the amino acids in the peptide by any other residue. If a mass match is found, the corresponding amino acid substitution is suggested to the user. If the substitution further corresponds to a sequence variant or conflict annotated for this peptide in the underlying Swiss-Prot entry, this is highlighted in color. This allows distinguishing between merely potential substitutions and those already observed and documented in the literature, or it might help to confirm a variation that was previously annotated as a conflict. As additional interpretation help, the BLOSUM62 score [99] is provided for each suggested potential amino acid substitution.

PFMUTS (ProteinFragmentMUTationS) PFMUTS (http://www.mcs.vuw.ac.nz/~aleksand/pfmuts.html) is based on the same idea. PFMUTS only accepts raw sequence input, and the analysis cannot, therefore, be combined with the wealth of related information already known and annotated in Swiss-Prot. However, the program allows prediction not only of single, but also potential double mutations of a peptide fragment, i.e. it is possible to find peptides in which two amino acids are replaced by two other amino acids.

Mascot In the context of peptide characterization, Mascot (see also above) allows a subset of the matching candidate proteins to be selected for submission to an "error tolerant search". This can be done using the MS-MS search mode that tries to increase the number of matching peaks by predicting potential unspecific cleavage, post-translational modifications, and primary sequence variations. Only the amino acid substitutions that result from single base substitutions are taken into account, which is an attempt to reduce the number of suggested substitutions of an otherwise potentially very long list.

As can be seen from the above list, the number of available identification tools that allow the discovery of new SNPs, based on peptide mass fingerprinting alone, is still limited. However, if one candidate protein is available, several characterization programs are available that allow the score for this candidate protein to be refined by investigating potential amino acid substitutions. This is probably due to the huge combinatorial overhead that results from the large number of possible substitutions: If every amino acid of every peptide of every protein in a database is potentially substituted, there is a high risk of obtaining a large number of unrelated, false-positive hits. Identification algorithms have to be extremely robust and well designed in order to deal with this problem in a sophisticated manner.

21.4.2 Tools for Analyzing and/or Predicting the Effects of Protein Variants

Many protein variants cause disease in humans, and these disease-causing mutations often tend to occur in structurally and functionally important sites. However, without in-depth biochemical or functional studies, it is often difficult to distinguish causative mutations from polymorphisms with little or no clinical significance. Techniques that could predict the functional consequences of a protein variant would therefore be invaluable for the clinical management of the related disease, as well as for medical research. Currently, there are two major approaches for the analysis of the impact of protein variations on the structure/function of proteins.

21.4.2.1 Sequence-based Analysis or Prediction Tools

The first strategy involves sequence-based analysis. In this type of analysis, protein sequence comparison among species is often used to infer preliminary information about protein function. The basic assumption is that in evolution, sequences coding for the most critical structural or functional amino acids tend to be the most highly conserved due to selection pressure against functionally altered proteins. Consequently, evolutionarily conserved amino acids could serve as a surrogate to identify functionally critical amino acids. There are several online analysis tools that use this assumption to predict the effect of variants. Most of these tools use empirical weighting schemes to measure amino acid similarity and score the substitution based on how often each amino acid replaces other amino acids in alignments of evolutionarily related proteins.

SIFT (Sorting Intolerant From Tolerant) SIFT [100] (http://blocks.fhcrc.org/~pauline/SIFT.html) uses sequence homology and position-specific information to predict whether a substitution affects protein function or not [101]. Given a protein sequence, this tool searches for related proteins and performs a multiple sequence alignment (MSA) of these proteins with the query. Then, based on the amino acid appearing at each position in the alignment, the program classifies

the substitutions as tolerated or deleterious. A position in the protein query that is conserved in the alignment will be scored by SIFT as intolerant to most changes; whereas a position that is poorly conserved will be scored by SIFT as tolerant.

Whilst SIFT can choose the sequences for alignment automatically, it was found that better prediction results were obtained when a list of homologous sequences was provided. In addition, orthologous sequences will be more appropriate than paralogous sequences with distinct biochemical functions. The latter may confound prediction.

ConSeq ConSeq (http://conseq.bioinfo.tau.ac.il/) is also a Web server that allows structurally and functionally important residues in proteins to be identified based on MSA. The program accepts either a unique sequence of a protein or a domain, or an MSA provided by the user as input. In cases where a single sequence is provided, the server automatically carries out a PSI-BLAST search [73] for close homologous sequences in the Swiss-Prot database, and produces an MSA using the CLUSTALW program [102]. A phylogenetic tree consistent with the MSA (either calculated or provided by users) is then built. The conservation score for each residue is calculated using the Rate4Site algorithm [103], which is based on the maximum likelihood principle [104]. In addition, the server uses the MSA to predict the relative solvent accessibility of each residue, whether it is in the protein core or exposed to the solvent. It then uses these two parameters to predict if a residue is biologically important. For example, slowly evolving and exposed residues are marked with an "f" (predicted to be functional), whereas slowly evolving and buried residues are marked with an "s" to indicate that they may have an important structural role. The ConSeq server provides a color scheme to visualize the conservation scores, the relative accessibility prediction for each residue, and indicates those residues with a predicted functional and structurally role.

Both SIFT and ConSeq base their predictions on sequence data alone and thus do not require protein structural information. They are thus well suited for the analysis of proteins with homologous sequences but of unknown three-dimensional (3-D) structure. However, sequence-based methods may not be sensitive enough to help analyze the effect of point mutations in more depth.

21.4.2.2 **Structure-based Analysis or Prediction Tools**

The second strategy is structure-based. In fact, knowledge of protein structure has long been used as one of the main approaches to rationalize the effects of disease-causing mutations [105, 106]. Amino acid variants may have an impact on the folding, interaction sites, solubility or stability of the protein, and thus directly affect protein structure and function. These effects cannot be directly analyzed by simple alignment programs, but they can be estimated from physico-chemical considerations. Therefore, several online servers use additional structural data to study or predict the effects of mutations.

ConSurf (Conservation Surface-mapping) The ConSurf server (http://consurf.tau.ac.il/) can be considered as a complement to the ConSeq server described above [107]. It has been developed by the same group. ConSurf enables the identification of functionally important regions on the surface of a protein or domain, based on the phylogenetic relations between its close sequence homologues [108]. Therefore, although ConSurf uses structural information, the algorithm used for the analysis is still mainly sequence-based.

ConSurf accepts a PDB file of a protein structure as input. It extracts the sequence from the PDB file and automatically carries out a search for homologous sequences of proteins of known structure. A phylogenetic tree consistent with the computed MSA is then built using the maximum parsimony method [109], and the conservation scores calculated [103]. Similar to ConSeq, ConSurf also accepts user-provided MSA to build the phylogenetic tree. The conservation scores are color-coded and "projected" onto the molecular surface of the protein. Users can use the online Protein Explorer engine (http://molvis.sdsc.edu/protexpl/frntdoor.htm) to visualize patches of highly conserved residues that are often of important biological function.

MutaProt MutaProt (http://bioinfo.weizmann.ac.il/cgi-bin/MutaProt/mutations0.cgi) is a tool designed to analyze pairs of PDB files that differ in one or two amino acids and to examine the microenvironment surrounding the mutated residue(s) [110]. The server has a collection of PDB files (resolution of 3.5 Å or better) concerning structural distinction between proteins with point substitutions in amino acid. Users can query these data in three ways: by specifying a PDB ID, keywords, or by designating a particular pair of amino acids. Searches can also be restricted to a particular level of residue solvent accessibility. Links are provided to the detailed structural analysis of the mutation site. This includes the relative accessibility of the mutated residues, the displacement between equivalent atoms of the contacting residues, and a list of contacts involving the mutated site and its surrounding residues. In addition, users can visualize the superimposed regions around the mutation site(s) via an interactive 3-D presentation.

The MutaProt database was last updated on November 2002 and currently contains 17668 PDB file-pairs, including 5654 single mutations, and 12014 double mutations.

PolyPhen (Polymorphism Phenotyping) PolyPhen (http://tux.embl-heidelberg.de/ramensky/) is an automatic tool for the prediction of the possible impact of an amino acid substitution on the structure and function of a human protein [111]. The prediction is based on a set of straightforward empirical rules that take into account the protein annotation of the substitution site, the profile analysis of homologous sequences and the 3-D protein structure [112].

For a given amino acid substitution in a human protein, PolyPhen first checks if the amino acid replacement occurs at a site that is annotated in the Swiss-Prot/TrEMBL database feature table as DISULFID, BINDING, ACT_SITE, LIPID, METAL, SITE, or MOD_RES, or at a site located in a TRAMSMEM, SIGNAL, or

PROPEP region. The server also takes into account predictions made with computational programs, such as the TMHMM algorithm for transmembrane regions (http://www.cbs.dtu.dk/services/TMHMM/) and the SignalP program (http://www.cbs.dtu.dk/services/SignalP/) for signal peptide regions of the protein sequences. As a second step, PolyPhen identifies homologues of the input sequences via a BLAST search of the NRDB database (a database of non-identical protein sequences extracted from EMBL CDS translations + PDB + Swiss-Prot + PIR) and uses aligned sequences with 30–94% sequence identity to the input sequence to construct a profile matrix by using the position-specific independent counts (PSIC) method [113]. Finally, a number of structural parameters, such as secondary structure, solvent accessible surface area, ϕ–ψ dihedral angles, as well as contacts with "critical sites", ligands, and other polypeptide chains are determined. This is done by mapping the amino acid substitution to known protein 3-D structure or, if the 3-D structure is not available, to a homologous protein structure of at least 50% sequence identity to the query.

PolyPhen accepts a protein sequence, or the Swiss-Prot/TrEMBL ID or AC together with sequence position and two amino acid variants characterizing the polymorphism as input. The program then sorts the mutations into possibly damaging (supposed to affect protein function), probably damaging, benign (most likely lacking any phenotypic effect), or unknown. Not all the parameters calculated are used for the decision rule, if some of them do not help to increase sensibility without significant loss of specificity of predictions.

Apart from these analysis or prediction tools for variants, several recent studies have performed in-depth analysis of different structural biophysical parameters to estimate the likely impact of an amino acid substitution on the structure and function of a protein. These studies either seek to differentiate between neutral and functional polymorphisms [114, 115], or attempt to identify important structural parameters for the disease-causing effects of particular mutations [116–118]. Most of these studies performed mapping of a set of protein variants from public databases onto existent 3-D structures of the corresponding proteins and analyzed their structural features. The availability of 3-D structures is thus critical to determine the data set used. To date, the number of experimental protein structures (19151, PDB holding 10 June 2003) is still limited when compared with the number of available sequences (more than 1 million Swiss-Prot Release 41.15 and TrEMBL release 24.1 of 04.07.2003). In this context, the use of predictive methods, such as comparative or homology modeling, could represent an alternative solution to the problem.

21.4.2.3 **The Swiss-Prot Variant Page and Comparative Modeling**

The Swiss-Prot variant page (see Fig. 21.2) provides 3-D comparative models of protein variants (Yip et al., unpublished data). These homology models were constructed using the automated comparative protein modeling machine PromodII underlying the SWISS-MODEL server (http://swissmodel.expasy.org/) [119]. Com-

parative modeling is based on the observation that families of proteins with members sharing similar sequences often fold into similar 3-D structures. This conservation allows, in principle, the prediction of structure in a family of proteins when only the structure of a single family member is known. Clearly, the reliability and the accuracy of the models are of great importance in such an approach and these two factors largely depend on how closely related the target sequence and the template structure are. Therefore, for the Swiss-Prot variant page, proteins are chosen to be modeled only when crystal template structure(s) of better than 2.5 Å resolutions and of at least 70% sequence identity can be found. This threshold was selected on the basis of several facts: (1) the four successive Critical Assessment of Protein Structure Prediction (CASP) meetings have demonstrated that deviation of models from experimental structures are typically <1 Å for Cα atoms at >60% sequence identity [120]; (2) models which were built using templates with a high degree of sequence identity (>70%) have been proven useful during drug design projects, thus indicating their high reliability and accuracy [121, 122].

Currently, more than 3500 variants of the 15 000-plus human protein variants recorded in the Swiss-Prot database have corresponding 3-D models. Of these 3500 variants, about 1600 are related to disease. The model data, together with information on Swiss-Prot sequences/variants, and experimentally determined protein structures from the PDB are stored in the regularly updated ModSNP database (Yip et al., unpublished data). The models can be accessed through the Swiss-Prot variant page and can be further analyzed using the platform-independent molecular graphic suite Astax Viewer.

21.4.2.4 **Remarks**

The analysis or prediction tools, either sequence- or structure-based, all have their own pros and cons. Sequence-based tools that use sequence homology rather than protein structure can potentially analyze a larger number of variants than tools based on protein structure alone. However, in instances where fewer than 5–10 homologues are available, it is recommended that structural information should be included if available [123]. In terms of performance, it is currently found that sequence- and structure-based tools have similar prediction abilities [100], and that the MSA-based profile scores appear to contribute most to the prediction when both structural and sequence information is used [111]. Nevertheless, as it is at present still not clear which structural parameters give the strongest predictive power and because few detailed large-scale structural analyses of protein variants have been performed due to the deficit in protein structures, it could be possible that structure-based analysis has not yet reached its full potential. The availability of high-quality comparative models will surely help to provide a larger data set for analysis.

21.5 Conclusions

Following the completion of the Human Genome Project, the increased availability of public databases and online analysis tools is surely influencing the way medical research is done. *In silico* approaches may be used to identify candidate genes for genetic disorders and to set priorities in their analysis [124, 125]. For example, a wise use of the analysis and prediction tools of the effect of protein variants can provide a fast and useful guide for the design and analysis of mutagenesis studies. This can reduce the number of functional assays required and give a higher proportion of affected phenotypes. Careful sequence and structural analysis of proteins involved in similar disorders may also uncover novel features and suggest new pathological pathways [126]. In this respect, a comprehensive and correct annotation is of crucial importance as any *in silico* approach is highly dependent on the quality of the data to which it is applied. The integration of mutation data, functional data, expression profiles, protein sequence and structural information, protein/protein interaction data, and phenotypic descriptions in databases will play a major role in the elucidation of the chain of events leading from a molecular defect to a pathology.

21.6 References

1 Collins, F.S., Patrinos, A., Jordan, E., Chakravarti, A., Gesteland, R., Walters, L. *Science* **1998**, *282*, 682–689.
2 Collins, F.S., Brooks, L.D., Chakravarti, A. *Genome Res.* **1998**, *8*, 1229–1231.
3 Gut, I.G. *Hum. Mutat.* **2001**, *17*, 475–492.
4 Tsuchihashi, Z., Dracopoli, N.C. *Pharmacog. J.* **2002**, *2*, 103–110.
5 Cargill, M., Altshuler, D., Ireland, J. et al. *Nature Genet.* **1999**, *22*, 231–238.
6 Cambien, F., Poirier, O., Nicaud, V. et al. *Am. J. Hum. Genet.* **1999**, *65*, 183–191.
7 Halushka, M.K., Fan, J.B., Bentley, K. et al. *Nature Genet.* **1999**, *22*, 239–247.
8 Altshuler, D., Pollara, V.J., Cowles, C.R. et al. *Nature* **2000**, *407*, 513–516.
9 Clifford, R., Edmonson, M., Hu, Y., Nguyen, C., Scherpbier, T., Buetow, K.H. *Genome Res.* **2000**, *10*, 1259–1265.
10 Martin, E.R., Lai, E.H., Gilbert, J.R. et al. *Am. J. Hum. Genet.* **2000**, *67*, 383–394.
11 Johnson, G.C., Esposito, L., Barratt, B.J. et al. *Nature Genet.* **2001**, *29*, 233–237.
12 Sachidanandam, R., Weissman, D., Schmidt, S.C. et al. *Nature* **2001**, *409*, 928–933.
13 Iida, A., Saito, S., Sekine, A. et al. *J. Hum. Genet.* **2002**, *47*, 285–310.
14 Cotton, R.G., McKusick, V., Scriver, C.R. *Science* **1998**, *279*, 10–11.
15 Scriver, C.R., Nowacki, P.M., Lehvaslaiho, H. *Hum. Mutat.* **1999**, *13*, 344–350.
16 Scriver, C.R., Nowacki, P.M., Lehvaslaiho, H. *Hum. Mutat.* **2000**, *15*, 13–15.
17 Baxevanis, A.D. *Nucleic Acids Res.* **2003**, *31*, 1–12.
18 Antonarakis, S.E. *Hum. Mutat.* **1998**, *11*, 1–3.
19 den Dunnen, J.T., Antonarakis, S.E. *Hum. Mutat.* **2000**, *15*, 7–12.
20 Paznekas, W.A., Boyadjiev, S.A., Shapiro, R.E. et al. *Am. J. Hum. Genet.* **2003**, *72*, 408–418.

21 Lehvaslaiho, H., Stupka, E., Ashburner, M. *Hum. Mutat.* **2000**, *15*, 52–56.
22 Cotton, R.G., Scriver, C.R. *Hum. Mutat.* **1998**, *12*, 1–3.
23 Cotton, R.G., Horaitis, O. *Hum. Mutat.* **2000**, *15*, 16–21.
24 Rozet, J.M., Gerber, S., Souied, E. et al. *Eur. J. Hum. Genet.* **1998**, *6*, 291–295.
25 Paloma, E., Martinez-Mir, A., Vilageliu, L., Gonzalez-Duarte, R., Balcells, S. *Hum. Mutat.* **2001**, *17*, 504–510.
26 Briggs, C.E., Rucinski, D., Rosenfeld, P.J., Hirose, T., Berson, E.L., Dryja, T.P. *Invest. Ophthalmol. Vis. Sci.* **2001**, *42*, 2229–2236.
27 Webster, A.R., Heon, E., Lotery, A.J. et al. *Invest. Ophthalmol. Vis. Sci.* **2001**, *42*, 1179–1189.
28 Gerber, S., Rozet, J.M., van de Pol, T.J. et al. *Genomics* **1998**, *48*, 139–142.
29 Nasonkin, I., Illing, M., Koehler, M.R., Schmid, M., Molday, R.S., Weber, B.H. *Hum. Genet.* **1998**, *102*, 21–26.
30 Cremers, F.P., van de Pol, D.J., van Driel, M. et al. *Hum. Mol. Genet.* **1998**, *7*, 355–362.
31 Maugeri, A., Klevering, B.J., Rohrschneider, K. et al. *Am. J. Hum. Genet.* **2000**, *67*, 960–966.
32 Allikmets, R., Shroyer, N.F., Singh, N. et al. *Science* **1997**, *277*, 1805–1807.
33 Allikmets, R. *Am. J. Hum. Genet.* **2000**, *67*, 487–491.
34 Guymer, R.H., Heon, E., Lotery, A.J. et al. *Arch. Ophthalmol.* **2001**, *119*, 745–751.
35 Porter, C.J., Talbot, C.C., Cuticchia, A.J. *Hum. Mutat.* **2000**, *15*, 36–44.
36 Hamosh, A., Scott, A.F., Amberger, J., Valle, D., McKusick, V.A. *Hum. Mutat.* **2000**, *15*, 57–61.
37 Hamosh, A., Scott, A.F., Amberger, J., Bocchini, C., Valle, D., McKusick, V.A. *Nucleic Acids Res.* **2002**, *30*, 52–55.
38 McKusick, V.A. *Mendelian Inheritance in Man. A Catalog of Human Genes and Genetic Disorders*, 12th edn. Johns Hopkins University Press, Baltimore, **1997**.
39 Krawczak, M., Ball, E.V., Fenton, I. et al. *Hum. Mutat.* **2000**, *15*, 45–51.
40 Stenson, P.D., Ball, E.V., Mort, M. et al. *Hum. Mutat.* **2003**, *21*, 577–581.
41 Cooper, D.N., Ball, E.V., Krawczak, M. *Nucleic Acids Res.* **1998**, *26*, 285–287.
42 Sherry, S.T., Ward, M., Sirotkin, K. *Genome Res.* **1999**, *9*, 677–679.
43 Sherry, S.T., Ward, M.H., Kholodov, M. et al. *Nucleic Acids Res.* **2001**, *29*, 308–311.
44 Fredman, D., Siegfried, M., Yuan, Y.P., Bork, P., Lehvaslaiho, H., Brookes, A.J. *Nucleic Acids Res.* **2002**, *30*, 387–391.
45 Thorisson, G.A., Stein, L.D. *Nucleic Acids Res.* **2003**, *31*, 124–127.
46 Hirakawa, M., Tanaka, T., Hashimoto, Y., Kuroda, M., Takagi, T., Nakamura, Y. *Nucleic Acids Res.* **2002**, *30*, 158–162.
47 Osier, M.V., Cheung, K.H., Kidd, J.R., Pakstis, A.J., Miller, P.L., Kidd, K.K. *Am. J. Phys. Anthropol.* **2002**, *119*, 77–83.
48 Rajeevan, H., Osier, M.V., Cheung, K.H. et al. *Nucleic Acids Res.* **2003**, *31*, 270–271.
49 Cuticchia, A.J. *Hum. Mutat.* **2000**, *15*, 62–67.
50 Claustres, M., Horaitis, O., Vanevski, M., Cotton, R.G. *Genome Res.* **2002**, *12*, 680–688.
51 Olivier, M., Eeles, R., Hollstein, M., Khan, M.A., Harris, C.C., Hainaut, P. *Hum. Mutat.* **2002**, *19*, 607–614.
52 Hernandez-Boussard, T., Rodriguez-Tome, P., Montesano, R., Hainaut, P. *Hum. Mutat.* **1999**, *14*, 1–8.
53 Beroud, C., Soussi, T. *Hum. Mutat.* **2003**, *21*, 176–181.
54 Gottlieb, B., Lehvaslaiho, H., Beitel, L.K., Lumbroso, R., Pinsky, L., Trifiro, M. *Nucleic Acids Res.* **1998**, *26*, 234–238.
55 Vihinen, M., Kwan, S.P., Lester, T. et al. *Hum. Mutat.* **1999**, *13*, 280–285.
56 Brown, A., McKie, M., van H., V., Prosser, J. *Nucleic Acids Res.* **1998**, *26*, 259–264.
57 Dalgleish, R. *Nucleic Acids Res.* **1997**, *25*, 181–187.
58 Dalgleish, R. *Nucleic Acids Res.* **1998**, *26*, 253–255.
59 Stenberg, K.A., Riikonen, P.T., Vihinen, M. *Nucleic Acids Res.* **2000**, *28*, 369–371.
60 Scriver, C.R., Waters, P.J., Sarkissian, C. et al. *Hum. Mutat.* **2000**, *15*, 99–104.
61 Scriver, C.R., Hurtubise, M., Konecki, D. et al. *Hum. Mutat.* **2003**, *21*, 333–344.

62 Kwok, C.J., Martin, A.C., Au, S.W., Lam, V.M. *Hum. Mutat.* **2002**, *19*, 217–224.

63 Van Durme, J.J., Bettler, E., Folkertsma, S., Horn, F., Vriend, G. *Nucleic Acids Res.* **2003**, *31*, 331–333.

64 Fung, D.C., Holland, E.A., Becker, T.M., Hayward, N.K., Bressac-de Paillerets, B., Mann, G.J. *Hum. Mutat.* **2003**, *21*, 2–7.

65 Hardison, R.C., Chui, D.H., Giardine, B. et al. *Hum. Mutat.* **2002**, *19*, 225–233.

66 Sarrauste de Menthiere, C., Terriere, S., Pugnere, D., Ruiz, M., Demaille, J., Touitou, I. *Nucleic Acids Res.* **2003**, *31*, 282–285.

67 Boeckmann, B., Bairoch, A., Apweiler, R. et al. *Nucleic Acids Res.* **2003**, *31*, 365–370.

68 O'Donovan, C., Apweiler, R., Bairoch, A. *Trends Biotechnol.* **2001**, *19*, 178–181.

69 Povey, S., Lovering, R., Bruford, E., Wright, M., Lush, M., Wain, H. *Hum. Genet.* **2001**, *109*, 678–680.

70 Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R.D., Bairoch, A. *Nucleic Acids Res.* **2003**, *31*, 3784–3788.

71 Jacoby, E.S., Kicman, A.T., Iles, R.K. *J. Mol. Endocrinol.* **2003**, *30*, 239–252.

72 Hoogland, C., Sanchez, J.C., Tonella, L. et al. *Nucleic Acids Res.* **2000**, *28*, 286–288.

73 Altschul, S.F., Madden, T.L., Schaffer, A.A. et al. *Nucleic Acids Res.* **1997**, *25*, 3389–3402.

74 Gattiker, A., Gasteiger, E., Bairoch, A. *Appl. Bioinform.* **2002**, *1*, 107–108.

75 Kersey, P., Hermjakob, H., Apweiler, R. *Bioinformatics* **2000**, *16*, 1048–1049.

76 Gasteiger, E., Jung, E., Bairoch, A. *Curr. Issues Mol. Biol.* **2001**, *3*, 47–55.

77 Stoesser, G., Baker, W., van den, B.A. et al. *Nucleic Acids Res.* **2002**, *30*, 21–26.

78 Bhat, T.N., Bourne, P., Feng, Z. et al. *Nucleic Acids Res.* **2001**, *29*, 214–218.

79 Berman, H.M., Westbrook, J., Feng, Z. et al. *Nucleic Acids Res.* **2000**, *28*, 235–242.

80 Wain, H.M., Lush, M., Ducluzeau, F., Povey, S. Genew: the human gene nomenclature database. *Nucleic Acids Res.* **2002**, *30*, 169–171.

81 Ashburner, M., Ball, C.A., Blake, J.A. et al. *Nature Genet.* **2000**, *25*, 25–29.

82 Rebhan, M., Chalifa-Caspi, V., Prilusky, J., Lancet, D. *Bioinformatics* **1998**, *14*, 656–664.

83 Lenhard, B., Hayes, W.S., Wasserman, W.W. *Genome Res.* **2001**, *11*, 2151–2157.

84 Marsh, S., Kwok, P., McLeod, H.L. *Hum. Mutat.* **2002**, *20*, 174–179.

85 Etzold, T., Ulyanov, A., Argos, P. *Methods Enzymol.* **1996**, *266*, 114–128.

86 Wheeler, D.L., Church, D.M., Federhen, S. et al. *Nucleic Acids Res.* **2003**, *31*, 28–33.

87 Beroud, C., Collod-Beroud, G., Boileau, C., Soussi, T., Junien, C. *Hum. Mutat.* **2000**, *15*, 86–94.

88 Brown, A.F., McKie, M.A. *Hum. Mutat.* **2000**, *15*, 76–85.

89 Dobrokhotov, P.B., Goutte, C., Veuthey, A.L., Gaussier, E. *Bioinformatics* **2003**, *19 Suppl. 1*, i91–i94.

90 Gottlieb, B., Beitel, L.K., Trifiro, M.A. *Hum. Mutat.* **2001**, *17*, 382–388.

91 Henzel, W.J., Billeci, T.M., Stults, J.T., Wong, S.C., Grimley, C., Watanabe, C. *Proc. Natl Acad. Sci. USA* **1993**, *90*, 5011–5015.

92 Mann, M., Hojrup, P., Roepstorff, P. *Biol. Mass Spectrom.* **1993**, *22*, 338–345.

93 Yates, J.R., III, Speicher, S., Griffin, P.R., Hunkapiller, T. *Anal. Biochem.* **1993**, *214*, 397–408.

94 Perkins, D.N., Pappin, D.J., Creasy, D.M., Cottrell, J.S. *Electrophoresis* **1999**, *20*, 3551–3567.

95 Gatlin, C.L., Eng, J.K., Cross, S.T., Detter, J.C., Yates, J.R., III. *Anal. Chem.* **2000**, *72*, 757–763.

96 Liebler, D.C., Hansen, B.T., Davey, S.W., Tiscareno, L., Mason, D.E. *Anal. Chem.* **2002**, *74*, 203–210.

97 Pevzner, P.A., Mulyukov, Z., Dancik, V., Tang, C.L. *Genome Res.* **2001**, *11*, 290–299.

98 Wilkins, M.R., Gasteiger, E., Gooley, A.A. et al. *J. Mol. Biol.* **1999**, *289*, 645–657.

99 Henikoff, S., Henikoff, J.G. *Proc. Natl Acad. Sci. USA* **1992**, *89*, 10915–10919.

100 Ng, P.C., Henikoff, S. *Genome Res.* **2002**, *12*, 436–446.

101 Ng, P.C., Henikoff, S. *Genome Res.* **2001**, *11*, 863–874.

102 Thompson, J.D., Higgins, D.G., Gibson, T.J. *Nucleic Acids Res.* **1994**, *22*, 4673–4680.

103 Pupko, T., Bell, R.E., Mayrose, I., Glaser, F., Ben Tal, N. *Bioinformatics* **2002**, *18 Suppl. 1*, S71–S77.

104 Felsenstein, J. *J. Mol. Evol.* **1981**, *17*, 368–376.

105 Erlandsen, H., Stevens, R.C. *Mol. Genet. Metab.* **1999**, *68*, 103–125.

106 Sigal, A., Rotter, V. *Cancer Res.* **2000**, *60*, 6788–6793.

107 Glaser, F., Pupko, T., Paz, I. et al. *Bioinformatics* **2003**, *19*, 163–164.

108 Armon, A., Graur, D., Ben Tal, N. *J. Mol. Biol.* **2001**, *307*, 447–463.

109 Felsenstein, J. *Methods Enzymol.* **1996**, *266*, 418–427.

110 Eyal, E., Najmanovich, R., Sobolev, V., Edelman, M. *Bioinformatics* **2001**, *17*, 381–382.

111 Ramensky, V., Bork, P., Sunyaev, S. *Nucleic Acids Res.* **2002**, *30*, 3894–3900.

112 Sunyaev, S., Ramensky, V., Koch, I., Lathe, W., III, Kondrashov, A.S., Bork, P. *Hum. Mol. Genet.* **2001**, *10*, 591–597.

113 Sunyaev, S.R., Eisenhaber, F., Rodchenkov, I.V., Eisenhaber, B., Tumanyan, V.G., Kuznetsov, E.N. *Protein Eng.* **1999**, *12*, 387–394.

114 Wang, Z., Moult, J. *Hum. Mutat.* **2001**, *17*, 263–270.

115 Chasman, D., Adams, R.M. *J. Mol. Biol.* **2001**, *307*, 683–706.

116 Terp, B.N., Cooper, D.N., Christensen, I.T. et al. *Hum. Mutat.* **2002**, *20*, 98–109.

117 Ferrer-Costa, C., Orozco, M., de la Cruz, X. *J. Mol. Biol.* **2002**, *315*, 771–786.

118 Stitziel, N.O., Tseng, Y.Y., Pervouchine, D., Goddeau, D., Kasif, S., Liang, J. *J. Mol. Biol.* **2003**, *327*, 1021–1030.

119 Guex, N., Peitsch, M.C. *Electrophoresis* **1997**, *18*, 2714–2723.

120 Venclovas, C., Zemla, A., Fidelis, K., Moult, J. *Proteins* **2001**, *Suppl 5*, 163–170.

121 Marti-Renom, M.A., Stuart, A.C., Fiser, A., Sanchez, R., Melo, F., Sali, A. *Annu. Rev. Biophys. Biomol. Struct.* **2000**, *29*, 291–325.

122 Baker, D., Sali, A. *Science* **2001**, *294*, 93–96.

123 Saunders, C.T., Baker, D. *J. Mol. Biol.* **2002**, *322*, 891–901.

124 Perez-Iratxeta, C., Bork, P., Andrade, M.A. *Nature Genet.* **2002**, *31*, 316–319.

125 van Driel, M.A., Cuelenaere, K., Kemmeren, P.P., Leunissen, J.A., Brunner, H.G. *Eur. J. Hum. Genet.* **2003**, *11*, 57–63.

126 Scheel, H., Tomiuk, S., Hofmann, K. *Hum. Mol. Genet.* **2002**, *11*, 1757–1762.

Subject Index

Biomedical Application of Proteomics
Edited by J.-C. Sanchez, G. L. Corthals, D. F. Hochstrasser

ISBN: 3-527-30807-5