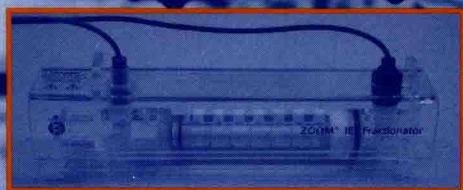
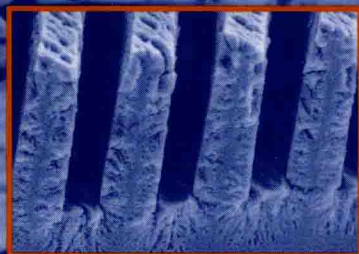
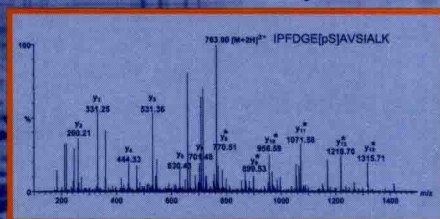
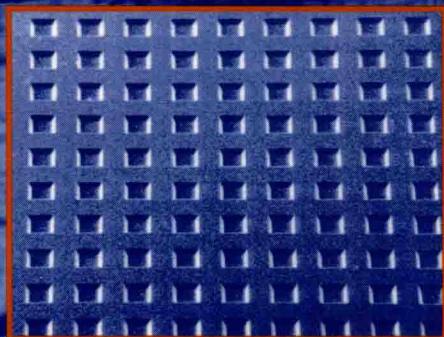
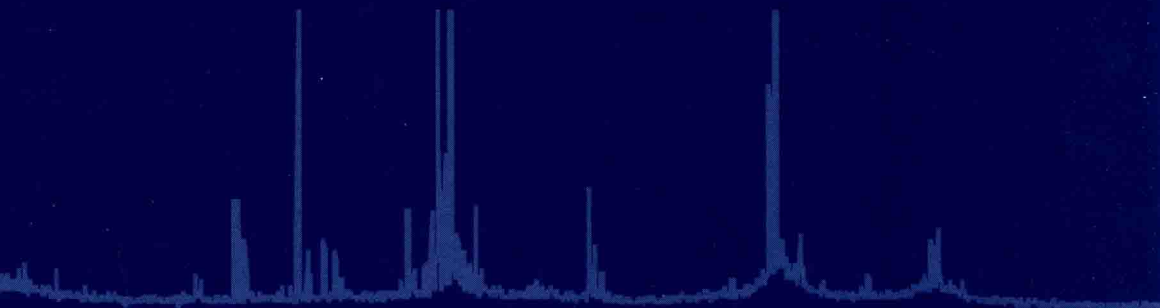




# PROTEOME ANALYSIS

## INTERPRETING THE GENOME



DAVID W. SPEICHER | EDITOR

# PROTEOME ANALYSIS INTERPRETING THE GENOME

Edited by

**David W. Speicher**

*The Wistar Institute  
3601 Spruce Street  
Philadelphia, PA 19104  
USA*

2004



ELSEVIER

Amsterdam – Boston – Heidelberg – London – New York – Oxford  
Paris – San Francisco – Singapore – Sydney – Tokyo

**ELSEVIER B.V.****Sara Burgerhartstraat 25****P.O. Box 211, 1000 AE Amsterdam****The Netherlands****ELSEVIER Inc.****525 B Street, Suite 1900****San Diego, CA 92101-4495****USA****ELSEVIER Ltd****The Boulevard, Langford Lane****Kidlington, Oxford OX5 1GB****UK****ELSEVIER Ltd****84 Theobalds Road****London WC1X 8RR****UK**

© 2004 Elsevier B.V. All rights reserved.

This work is protected under copyright by Elsevier B.V., and the following terms and conditions apply to its use:

**Photocopying**

Single photocopies of single chapters may be made for personal use as allowed by national copyright laws. Permission of the Publisher and payment of a fee is required for all other photocopying, including multiple or systematic copying, copying for advertising or promotional purposes, resale, and all forms of document delivery. Special rates are available for educational institutions that wish to make photocopies for non-profit educational classroom use.

Permissions may be sought directly from Elsevier's Rights Department in Oxford, UK: phone (+44) 1865 843830, fax (+44) 1865 853333, e-mail: [permissions@elsevier.com](mailto:permissions@elsevier.com). Requests may also be completed on-line via the Elsevier homepage (<http://www.elsevier.com/locate/permissions>).

In the USA, users may clear permissions and make payments through the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, USA; phone: (+1) (978) 7508400, fax: (+1) (978) 7504744, and in the UK through the Copyright Licensing Agency Rapid Clearance Service (CLARCS), 90 Tottenham Court Road, London W1P 0LP, UK; phone: (+44) 20 7631 5555; fax: (+44) 20 7631 5500. Other countries may have a local reprographic rights agency for payments.

**Derivative Works**

Tables of contents may be reproduced for internal circulation, but permission of the Publisher is required for external resale or distribution of such material. Permission of the Publisher is required for all other derivative works, including compilations and translations.

**Electronic Storage or Usage**

Permission of the Publisher is required to store or use electronically any material contained in this work, including any chapter or part of a chapter.

Except as outlined above, no part of this work may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without prior written permission of the Publisher.

Address permissions requests to: Elsevier's Rights Department, at the fax and e-mail addresses noted above.

**Notice**

No responsibility is assumed by the Publisher for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions or ideas contained in the material herein. Because of rapid advances in the medical sciences, in particular, independent verification of diagnoses and drug dosages should be made.

First edition 2004

**Library of Congress Cataloging in Publication Data**

A catalog record is available from the Library of Congress.

**British Library Cataloguing in Publication Data**

A catalogue record is available from the British Library.

ISBN: 0 444 51024 9

© The paper used in this publication meets the requirements of ANSI/NISO Z39.48-1992 (Permanence of Paper).  
Printed in Hungary.

**PROTEOME ANALYSIS  
INTERPRETING THE GENOME**

# Preface

The new discipline of proteomic can trace its birth to the mid-1990's when the terms "proteome" and "proteomics" were first introduced by Wilkins and Williams. Although some investigators initially considered "proteomics" to simply be a new name for conventional protein biochemistry, it rapidly became clear that this was a misperception. Proteomics represents an exciting new way to pursue biological and biomedical science at an unprecedented pace. Proteomics takes a broad, comprehensive, systematic approach to understanding biology that is generally unbiased and not dependent upon existing knowledge. It is primarily discovery-based rather than "hypothesis limited". That is, conventional one protein-at-a-time reductionist research is based upon and limited by existing knowledge. A reductionist scientist usually chooses to study a protein because prior research indicates that it is, or may be, important in a biological or disease process of interest. However, there could easily be 10 or more proteins that are more critical to the biological or disease process being studied, which have not yet been identified as associated with this process. Proteomics provides the means for identifying some or all of these more critical protein targets. Hence proteomics should have a dramatic impact on biological, biomedical and pharmaceutical research for the foreseeable future.

In comparison to genomics, proteomics is a far more complex and challenging problem. Genomes, particularly mammalian genomes, are certainly quite complex and, for example, simply assigning all open reading frames (ORFs) and splice variants of the human genome remains a daunting, incomplete task. However, even these complex genomes are finite and largely static over the lifetime of an organism. In contrast, proteomes are constantly changing in response to external stimuli and during development. The proteomes of higher eukaryotes are especially complex because alternative gene splicing and posttranslational modifications produce up to 20 to 100 times more unique protein components with distinct biological activities than the number of ORFs in the genome. The simplest form of a complex organism, a single cell type in culture, will express different subsets of this vast array of biologically distinct protein components under different conditions, and the moment the cells are exposed to a new condition, a different proteome will result. Hence, there are an essentially infinite number of proteomes that even such a relatively simple biological system can express. Despite this nearly infinite complexity, current proteomic tools provide the capacity to discover many novel individual proteins and groups of proteins (biosignatures) that are associated with normal biological processes as well as polygenetic disorders such as cancers and cardiovascular disease. These targets

and biosignatures have substantial potential for development of novel therapies and diagnostics. Furthermore, proteomics together with modern genetic methods such as nucleic acid arrays and siRNA approaches are starting to generate complex datasets that should begin to provide a systems biology view of cells, organisms and complex diseases.

This volume is a compilation of the current status of proteomics written by an international panel of experts in the field. The major components of proteomics from basic discovery using a range of alternative analytical methods, to discovery validation, and use of proteome methods for clinical applications are discussed. State-of-the-art protein profiling methods described include high resolution two-dimensional gels, two-dimensional differential in-gel electrophoresis, LC-MS and LC-MS/MS using accurate mass tags, and protein identifications of proteins from gels. These chapters reflect the rapid evolution of key technologies, including “old” methods such as 2-D gels, over this discipline’s first decade. However, further refinement of existing methods and continued introduction of new approaches are still clearly needed. Further advances are critical because no existing single analytical method or combination of methods can profile and quantitatively compare the majority of proteins in complex samples such as cells, tissues or biological fluids from higher eukaryotes. To address these limitations, most proteomics experts have turned to additional methods of sample separation. Some of the most promising approaches to obtain more global analysis of complex proteomes are to prefractionate proteomes into a series of well resolved fractions or to reproducibly isolate a single sub-proteome for detailed study. These approaches are discussed in a chapter describing electrophoretic prefractionation methods for global proteome analysis, and several chapters characterizing selected sub-proteomes based on specific posttranslational modifications including the phospho-proteome, the glyco-proteome, and nitrated proteins.

These conventional proteome analysis chapters are complemented by discussion of emerging technologies and approaches including affinity-based biosensor proteomics as well as the use of protein microarrays, microfluidics and nanotechnology. Strategies for improving throughput by automation are also discussed. Other chapters address the application of current proteome techniques to clinical problems and the availability of protein expression library resources for proteome studies.

Finally, I would like to express sincere appreciation to all the authors for their invaluable scientific contributions in a field where the rapid pace makes overcommitment and lack of time the norm. I would also like to acknowledge the invaluable administrative assistance of Ms. Emilie Gross with all aspects of editing and preparation of this volume for publication.

David W. Speicher

# List of Contributors

Natalie G. Ahn

*Department of Chemistry and Biochemistry, University of Colorado, Boulder, CO 80309, USA*

and

*Howard Hughes Medical Institute, University of Colorado, Boulder, CO 80309, USA*

David G. Camp II

*Macromolecular Structure and Dynamics, Environmental Molecular Sciences Laboratory, Battelle, Pacific Northwest National Laboratory, P.O. Box 999/MSIN K8-98, Richland, WA 99352, USA*

Bruno Catimel

*The Ludwig Institute for Cancer Research, Melbourne Tumour Biology Branch, P.O. Box 2008, The Royal Melbourne Hospital, Parkville, Vic. 3050, Australia*

Angelika Görg

*Technische Universität München, Fachgebiet Proteomik, D-85350, Freising-Weihenstephan, Germany*

Patrick R. Griffin

*Department of Molecular Profiling Proteomics, Merck Research Laboratories, Rahway, NJ 07065, USA*

Sam M. Hanash

*Department of Pediatrics, University of Michigan, 1150 W. Medical Center Drive, MSRB1, Room A520, Ann Arbor, MI 48109, USA*

William A. Hanlon

*Department of Molecular Profiling Proteomics, Merck Research Laboratories, Rahway, NJ 07065, USA*

Peter James

*Protein Technology, Wallenberg Laboratory II, Lund University, P.O. Box 7031, SE-220 07 Lund, Sweden*

Judith Jebanathirajah

*Department of Biochemistry and Molecular Biology, University of Southern Denmark, Camusvej 55, DK-5230 Odense M, Denmark*

Ole Nørregaard Jensen

*Department of Biochemistry and Molecular Biology, University of Southern Denmark, DK-5230 Odense M, Denmark*

Joshua Labaer

*Institute of Proteomics, Harvard Medical School, Department of Biological Chemistry and Molecular Pharmacology, 250 Longwood Avenue, Boston, MA 02115, USA*

Thomas Laurell

*Department of Electrical Measurements, Lund Institute of Technology, Lund University, P.O. Box 118, SE-221 00 Lund, Sweden*

KiBeom Lee

*The Wistar Institute, 3601 Spruce Street, Philadelphia, PA 19104, USA*

Georgy Marko-Varga

*Department of Analytical Chemistry, Lund University, P.O. Box 124, SE-221 00 Lund, Sweden*

Gerald Marsischky

*Institute of Proteomics, Harvard Medical School, Department of Biological Chemistry and Molecular Pharmacology, 250 Longwood Avenue, Boston, MA 02115, USA*

Edouard Nice

*The CRC for Cellular Growth Factors, Melbourne, Australia*  
and

*The Ludwig Institute for Cancer Research, Melbourne Tumour Biology Branch, P.O. Box 2008, The Royal Melbourne Hospital, Parkville, Vic. 3050, Australia*

Johan Nilsson

*Department of Electrical Measurements, Lund Institute of Technology, Lund University, P.O. Box 118, SE-221 00 Lund, Sweden*



Katheryn A. Resing

*Department of Chemistry and Biochemistry, University of Colorado, Boulder,  
CO 80309, USA*

Peter Roepstorff

*Department of Biochemistry and Molecular Biology, University of Southern  
Denmark, Camusvej 55, DK-5230 Odense M, Denmark*

Richard D. Smith

*Macromolecular Structure and Dynamics, Environmental Molecular Sciences  
Laboratory, Battelle, Pacific Northwest National Laboratory, P.O. Box 999/MSIN  
K8-98, Richland, WA 99352, USA*

David W. Speicher

*The Wistar Institute, 3601 Spruce Street, Philadelphia, PA 19104, USA*

Walter Weiss

*Technische Universität München, Fachgebiet Proteomik, D-85350, Freising-  
Weihenstephan, Germany*

Xun Zuo

*The Wistar Institute, 3601 Spruce Street, Philadelphia, PA 19104, USA*

# Contents

Preface	v
List of Contributors	xv

## Chapter 1

<b>Overview of Proteome Analysis</b>	<b>1</b>
--------------------------------------	----------

*David W. Speicher*

1. Introduction	1
2. Scope of the Proteomics Problem	2
3. Global versus Targeted Proteomics	4
4. Top Down Protein Profiling Methods	5
4.1 Two-Dimensional Gels	6
4.2 Two-Dimensional Differential Gel Electrophoresis (2D DIGE)	7
4.3 Non-2D Gel Separation Methods	8
4.4 Protein and Antibody Arrays	9
5. Bottom Up Protein Profiling Methods	10
5.1 Multi-Dimensional Protein Identification Technology	10
5.2 Isotope-Coded Affinity Tags	11
5.3 Accurate Mass Tag Based Protein Profiling	12
6. Automation, Miniaturization, and Future Prospects	13
7. Summary	14
Acknowledgements	15
References	15

## Chapter 2

<b>Protein Profile Comparisons of Microorganisms, Cells and Tissues using 2D Gels</b>	<b>19</b>
---	-----------

*Angelika Görg and Walter Weiss*

1. Introduction	20
2. Challenges of Protein Separation Methods for Proteome Analysis	21
2.1 Highly Alkaline Proteins	22
2.2 Zoom Gels, Non-Linear IPGs and Extended Separation Distances for Higher Resolution and Improved Detection of Low Copy Number Proteins	25
2.3 Protein Enrichment and Sample Pre-Fractionation Procedures	28
2.4 Low and High Molecular Mass Proteins	30
2.5 Very Hydrophobic Proteins	32
2.6 Protein Detection and Quantitation	34

2.7 Automated Procedures	37
3. Current Technology of 2D Electrophoresis with IPGs (IPG-Dalt)	38
3.1 Protein Profile Comparisons of Microorganisms, Cells and Tissues using 2D Gels	38
3.2 Sample Preparation	39
3.3 Two-Dimensional Electrophoresis with IPGs (IPG-Dalt)	44
3.4 Protein Visualization	58
3.5 Computer-Aided Image Analysis	61
4. 2D PAGE Databases	64
5. Summary	64
Acknowledgements	65
References	65

### Chapter 3

#### **Protein Profiling using Two-Dimensional Gel Electrophoresis with Multiple Fluorescent Tags 75**

*William A. Hanlon and Patrick R. Griffin*

1. Introduction	76
2. Mechanics of DIGE Technology	77
2.1 Sample Preparation	77
2.2 Cyanine Dye Labeling	77
2.3 Cyanine Image Visualization	78
2.4 Gel Image Analysis	78
3. Characterization of DIGE Technology	80
3.1 Cyanine Dye Labeling Bias	80
3.2 Assessment of Variability	82
3.3 Limit of Detection of DIGE	83
3.4 Importance of a Pooled Standard	86
3.5 Comparison to Other 2D Gel Methods	87
4. MS Identification of Cyanine Labeled Proteins	88
4.1 Excision of Cyanine Labeled Proteins	88
4.2 Protein Identification by LC-MS/MS	89
5. Summary	89
Acknowledgements	90
References	90

### Chapter 4

#### **Electrophoretic Prefractionation for Comprehensive Analysis of Proteomes 93**

*Xun Zuo, KiBeom Lee and David W. Speicher*

1. Introduction	94
2. Electrophoretic Prefractionation Methods	96
2.1 Rotofor	96

2.2	Free Flow Electrophoresis	97
2.3	IsoPrime and Related Multicompartment Electrolyzers	98
2.4	Microscale Solution Isoelectrofocusing Combined with Narrow pH Range 2D PAGE	99
3.	Strategies for Analysis of Large Soluble Proteins and Insoluble Proteins	103
3.1	Detection of Insoluble Proteins	103
3.2	Detection of Large Soluble Proteins	104
4.	Downstream Proteome Analysis After Sample Fractionation	105
4.1	Narrow pH Range 2D PAGE	105
4.2	1D PAGE	108
4.3	2D DIGE	109
4.4	LC–MS/MS and LC/LC–MS/MS Methods	110
5.	Summary	114
	Acknowledgments	115
	References	115

## Chapter 5

### Modification Specific Proteomics Applied to Protein Glycosylation and Nitration

119

*Judith Jebanathirajah and Peter Roepstorff*

1.	Introduction	119
2.	Glycosylation	121
2.1	Why Study Protein Glycosylation?	121
2.2	Strategies for Studying Glycosylation	123
3.	Tyrosine Nitration	133
4.	Summary	134
	References	134

## Chapter 6

### Phosphoproteomics: Mass Spectrometry Based Techniques for Systematic Phosphoprotein Analysis

139

*Ole Nørregaard Jensen*

1.	Introduction	140
2.	Modification-Specific Proteomics and Phosphoproteomics	141
2.1	Detection and Visualization of Phosphoproteins	142
2.2	Enrichment of Phosphoproteins and Phosphopeptides	145
3.	Detection and Sequencing of Phosphopeptides by Mass Spectrometry	148
3.1	Phosphoprotein Analysis by MALDI Mass Spectrometry	149
3.2	Phosphoprotein Analysis by MALDI Tandem Mass Spectrometry	150
3.3	Phosphoprotein Analysis by ESI Tandem Mass Spectrometry	150
3.4	Integrated Strategies for Phosphoprotein and Phosphoproteome Analysis by Mass Spectrometry	152

3.5 Phosphoprotein Characterization by Edman Degradation and Mass Spectrometry	153
4. Cellular Dynamics: Integrated Methods for Quantitative Phosphoproteome Analysis	154
5. Bioinformatics Tools for Phosphoprotein Sequence Analysis and Mass Spectrometry Data Interpretation	155
6. Summary	156
Acknowledgements	157
References	157

## Chapter 7

### **Protein Identification by In-Gel Digestion and Mass Spectrometry 163**

*Katheryn A. Resing and Natalie G. Ahn*

1. Introduction	164
2. In-Gel Digestion	165
2.1 Gel Electrophoresis and Staining	165
2.2 Gel Excision	166
2.3 Digestion	166
2.4 Peptide Extraction	168
2.5 Assessing Peptide Recovery	169
3. Mass Spectrometry	170
3.1 Sample Preparation and Data Collection	170
3.2 Peptide Mass Fingerprinting	171
3.3 Protein Identification using MS/MS Spectra	172
3.4 Database Searching Algorithms	176
4. Troubleshooting	176
4.1 Using Standards	176
4.2 What if there are no Peptides in the MS Spectra?	176
4.3 Common Contaminants	177
4.4 Evaluating Database 'Hits'	178
5. Summary	178
Acknowledgements	179
References	179

## Chapter 8

### **The Use of Accurate Mass and Time Tags Based Upon High-Throughput Fourier Transform Ion Cyclotron Resonance Mass Spectrometry for Global Proteomic Characterization 183**

*David G. Camp II and Richard D. Smith*

1. Introduction	184
2. Overall Experimental and Data Processing Approach	186
3. Sample Processing	188

3.1	Fractionation of Complex Peptide Mixtures	188
3.2	Preparation of Membrane Proteins	188
3.3	Stable-Isotope Labeling Methods	189
4.	High-Resolution Separations	191
5.	Chromatographic Separations Coupled to FTICR	194
6.	Generation of Accurate Mass and Time Tags and Their Utilization	195
7.	The Dynamic Range of Proteome Coverage	196
7.1	Dynamic Range Expansion by DREAMS FTICR MS	199
8.	Demonstration of Global Proteomic Characterization	205
9.	Overcoming Challenges to Proteome-wide Measurements	211
9.1	The Membrane Subproteome	211
9.2	The Phosphoproteome	212
10.	Summary	215
	Acknowledgements	218
	References	219

## Chapter 9

### **Clinical Applications of Proteomics** **225**

*Sam M. Hanash*

1.	Introduction	225
2.	Correlative Studies Using Proteomics and Transcriptomics	226
3.	Disease Marker Identification Using Proteomics	228
4.	Disease Tissue Analysis Using Proteomics	231
5.	Protein Microarrays as a Novel Technology for Disease Investigations	234
6.	Summary	237
	References	237

## Chapter 10

### **Affinity-based Biosensors, Microarrays and Proteomics** **243**

*Edouard Nice and Bruno Catimel*

1.	Introduction	244
2.	Biosensor Technology	246
2.1	Instrumentation	246
2.2	Sensitivity	250
2.3	Sensor Surfaces	250
2.4	Surface Immobilisation	253
3.	Biosensor Applications	253
3.1	Biosensor-based Ligand Searching	253
3.2	Preparative Biosensor Ligand Fishing and Proteomics	254
3.3	Cuvette-based Biosensors as Microaffinity Purification Platforms	258
4.	Protein Chip Mass Spectrometry Using SELDI	260

4.1 SELDI Technology	260
4.2 SELDI Applications	263
5. Protein Chips and Microarrays	266
5.1 Protein Profiling Arrays	267
5.2 Protein Function Arrays	271
6. Summary	275
References	275

## **Chapter 11**

### **Protein Expression Library Resources for Proteome Studies 287**

*Joshua LaBaer and Gerald Marsischky*

1. Introduction	287
1.1 Public Full-Length cDNA Clone Projects	288
1.2 Cloning Formats	288
1.3 Site-specific Recombination-based Cloning Systems	290
2. Clone Collections	290
2.1 Human	290
2.2 Mouse	294
2.3 <i>Caenorhabditis Elegans</i>	295
2.4 <i>Drosophila Melanogaster</i>	296
2.5 <i>Saccharomyces Cerevisiae</i>	297
2.6 <i>Pseudomonas Aeruginosa</i>	299
2.7 <i>Arabidopsis Thaliana</i>	300
3. Summary	300
Acknowledgements	300
References	301

## **Chapter 12**

### **Automation of Proteome Analysis 305**

*Peter James*

1. Introduction	306
2. Experimental Design	307
2.1 Experimental Approaches	307
2.2 Experimental Design Factors	308
3. Sample Preparation	308
3.1 Pre-Fractionation	310
3.2 Protein Extraction	310
4. 2D PAGE	311
4.1 Resolution	311
4.2 Gel Stability	312
5. Image Analysis	312
6. Robotics for Cutting, Digestion and Spotting	313

7. Protein Fingerprinting by MALDI MS	314
7.1 Automated In-Gel Digestion and Data Acquisition	318
7.2 Data Extraction and Database Searching	318
7.3 Confidence and Coverage Levels	319
8. Peptide Fingerprinting by MS/MS	319
8.1 Algorithms	320
8.2 New MALDI MS-based Workflows	320
8.3 New ESI-based Workflows	321
9. The Crucial Elements: LIMS and Data Mining	321
9.1 Pre-Packed Solutions	321
9.2 Data Mining	323
10. Summary	323
Acknowledgements	324
References	324

## Chapter 13

### **Micro- and Nanotechnology for Proteomics 327**

*G. Marko-Varga, J. Nilsson and T. Laurell*

1. Introduction	327
2. Benefits of Miniaturisation	329
3. Miniaturisation in Proteomics	330
4. Fabrication of Microstructures	332
5. Microstructures for Proteomics	335
5.1 Protein Digestion On-Chip	335
5.2 Microchip Solid Phase Enrichment	338
5.3 Microdispensing to Interface MALDI	341
5.4 Nanovial MALDI Target Arrays	347
6. Summary	364
Acknowledgements	365
References	365

<i>Index</i>	371
--------------	-----



Chapter 1

Overview of proteome analysis

DAVID W. SPEICHER\*

*The Wistar Institute, 3601 Spruce Street, Philadelphia, PA 19104, USA*

1. Introduction	1
2. Scope of the proteomics problem	2
3. Global versus targeted proteomics	4
4. Top down protein profiling methods	5
4.1. Two-dimensional gels	6
4.2. Two-dimensional differential gel electrophoresis (2D DIGE)	7
4.3. Non-2D gel separation methods	8
4.4. Protein and antibody arrays	9
5. Bottom up protein profiling methods	10
5.1. Multi-dimensional protein identification technology	10
5.2. Isotope-coded affinity tags	11
5.3. Accurate mass tag based protein profiling	12
6. Automation, miniaturization, and future prospects	13
7. Summary	14
Acknowledgements	15
References	15

1. Introduction

Proteomics emerged as an exciting new discipline in the mid-1990s and has continued to grow rapidly since its inception. This new approach to biology and biomedical research was enabled by the emerging availability of complete sequences of genomes and the development of high sensitivity mass spectrometry techniques and instruments for analyzing proteins and peptides. The term

\* Tel.: +1-215-898-3972; fax: +1-215-898-0664. E-mail: [speicher@wistar.upenn.edu](mailto:speicher@wistar.upenn.edu) (D.W. Speicher).