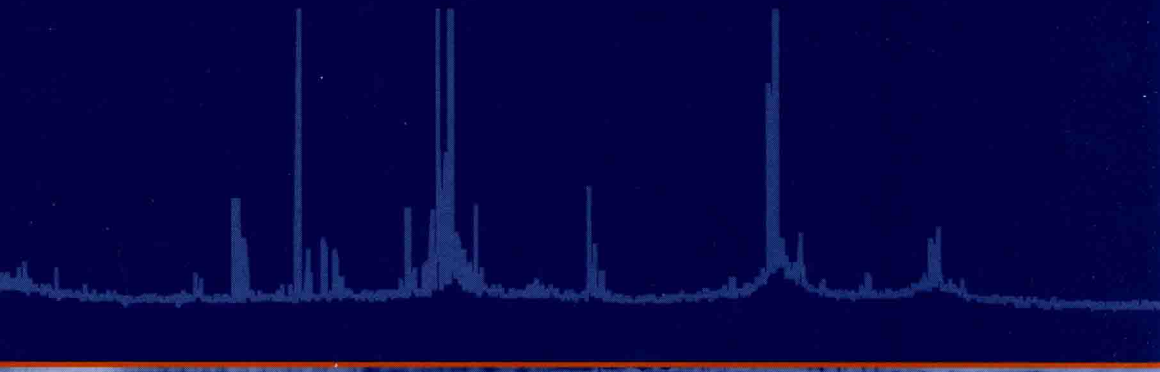
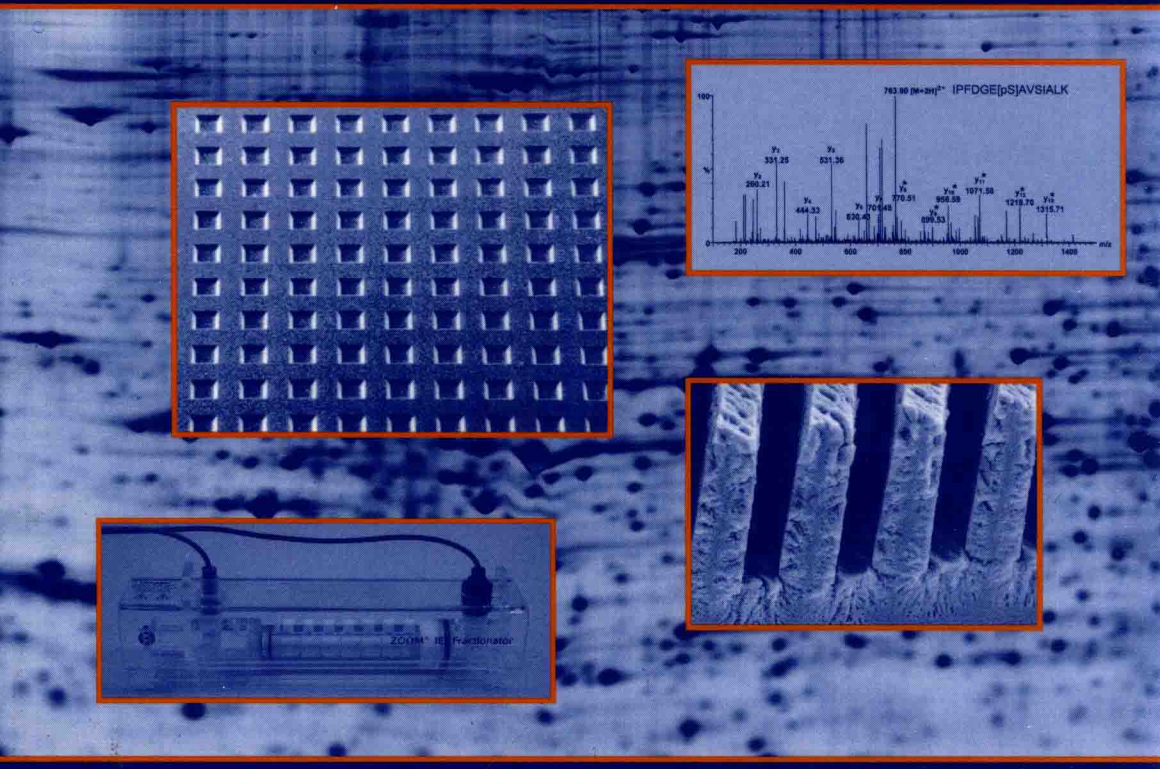
PROTEOME ANALYSIS

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

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, MSRBI, 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

Proteome Analysis. Interpreting the Genome.
D.W. Speicher (editor)

Chapter 1

Overview of proteome analysis

DAVID W. SPEICHER*

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

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 (D.W. Speicher).