

INTRODUCTION TO PROTEOMICS

Introduction to Proteomics

Tools for the New Biology

By

DANIEL C. LIEBLER, PhD

*College of Pharmacy
The University of Arizona
Tucson, AZ*

Foreword by

JOHN R. YATES, III, PhD

*Department of Cell Biology
The Scripps Research Institute
La Jolla, CA*



**Springer Science+Business
Media, LLC**

© 2002 Springer Science+Business Media New York
Originally published by Humana Press Inc. in 2002
Softcover reprint of the hardcover 1st edition 2002

humanapress.com

For additional copies, pricing for bulk purchases, and/or information about other Humana titles, contact Humana at the above address or at any of the following numbers: Tel.: 973-256-1699; Fax: 973-256-8341, E-mail: humana@humanapr.com; or visit our Web site at: www.humanapr.com

All rights reserved.

No part of this book may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, microfilming, recording, or otherwise without written permission from the Publisher. The content and opinions expressed in this book are the sole work of the authors and editors, who have warranted due diligence in the creation and issuance of their work. The publisher, editors, and authors are not responsible for errors or omissions or for any consequences arising from the information or opinions presented in this book and make no warranty, express or implied, with respect to its contents.

Cover design by Patricia Cleary.

Production Editor: Kim Hoather-Potter.

This publication is printed on acid-free paper. 

ANSI Z39.48-1984 (American National Standards Institute) Permanence of Paper for Printed Library Materials.

Photocopy Authorization Policy:

Authorization to photocopy items for internal or personal use, or the internal or personal use of specific clients, is granted by Springer Science+Business Media, LLC, provided that the base fee of US \$10.00 per copy, plus US \$00.25 per page, is paid directly to the Copyright Clearance Center at 222 Rosewood Drive, Danvers, MA 01923. For those organizations that have been granted a photocopy license from the CCC, a separate system of payment has been arranged and is acceptable to Springer Science+Business Media, LLC. The fee code for users of the Transactional Reporting Service is: [0-89603-991-9/02 \$10.00 + \$00.25].

10 9 8 7 6 5

Library of Congress Cataloging-in-Publication Data

Liebler, Daniel C.

Introduction to proteomics: tools for the new biology/Daniel C. Liebler.

p. cm.

Includes bibliographical references and index.

ISBN 978-0-89603-992-6 ISBN 978-1-59259-130-5 (eBook)

DOI 10.1007/978-1-59259-130-5

1. Proteins—Research—Methodology. I. Title.

QP551.L467 2002

572'.6'072—dc21

2001051465

Foreword

Mass spectrometry has evolved tremendously since Professor Klaus Biemann first analyzed amino acids in a mass spectrometer in 1958. The clear challenge in Biemann's first experiment was how to introduce nonpolar molecules into the mass spectrometer to create ions. In the years since 1958, several new ionization techniques and sample introduction methods appeared and stimulated much progress in the analysis of biomolecules. As these new ionization techniques, such as chemical ionization, field desorption, field ionization, plasma desorption, and finally fast atom bombardment (FAB) emerged, new methods for peptide and protein characterizations also developed. Mass spectrometry technology leapt forward in 1987 with the introduction of matrix-assisted laser desorption ionization (MALDI) and the application of electrospray ionization (ESI) to biomolecules. Both ionization methods led to dramatic improvements in the analysis of peptides and proteins. A key mass spectrometry technique that benefited from the new ionization methods was tandem mass spectrometry.

In the early 1980s Professor Donald Hunt began developing and applying tandem mass spectrometry to the sequence analysis of peptides and proteins. FAB, a soft ionization technique, created intact protonated molecules and allowed the refinement of approaches for peptide sequencing. FAB was a major breakthrough for peptide sequencing, because peptides could now be readily ionized without derivatization to increase volatility. By incorporating FAB with tandem mass spectrometry, a rapid peptide sequencing methodology was developed. Most approaches used off-line HPLC separations when complicated peptide mixtures were encountered. Many proteins were sequenced by this approach and many important methods were developed. Unfortunately, on-line coupling of separation methods with FAB was never able to create a robust, easy-to-use method. This problem wasn't resolved until electrospray ionization facilitated the direct coupling of separation techniques to the mass spectrometer. All aspects of peptide and protein analyses were improved by increases in the sensitivity of analysis, easier sample handling, and automation.

These developments in mass spectrometry dovetailed very nicely into the worldwide efforts to sequence the human genome. The genome sequencing efforts encompassed not only the human genome, but also genomes of many model organisms and have resulted in the generation of a large amount of sequence information. In 1993 several groups discovered that mass spectrometry data could be used to search databases to identify the protein under study. In 1994 methods to search sequence databases using tandem mass spectrometry data were developed allowing one to "look up the answer in the back of the book." If the "book" was an organism whose genome was sequenced, then the answer was most assuredly in the back. The complex issues of post-translational modifications and amino acid sequence variations can also be addressed by knowing the sequences of proteins from a genome sequence.

Interest in and use of mass spectrometry in the biological sciences has grown rapidly during the 1990s and threatens to become as ubiquitous and important as SDS-PAGE in the new millennium. Biologists will come to rely on mass spectrometry to determine the outcomes of their experiments. Given the need for biologists to use mass spectrometry technology to analyze their experiments, how does a biologist learn about the art of mass spectrometry and the methods of proteomics? This book, *Introduction to Proteomics: Tools for the New Biology* by Professor Daniel Liebler, presents a tutorial on mass spectrometry and its use in proteomics. The basics of mass spectrometers and ionization techniques are described, which is important to ascertain what type of mass spectrometer is most appropriate for a particular study. The ability to use mass spectrometry data to search databases is an important advance for the nonspecialist, because it no longer requires the development of the skills to interpret mass spectra. A basic understanding of the fundamentals of the search algorithms and their limitations is described in the book. Finally, applications of mass spectrometry to proteomics are described. This book provides an excellent introduction and overview of proteomics for the graduate student or for any biologist interested in understanding the basics of this rapidly evolving area.

John R. Yates, III
Scripps Research Institute
La Jolla, CA

Preface

This book is an introduction to the new field of proteomics. It is intended to describe how proteins and proteomes can be analyzed and studied. Despite widespread, growing interest in proteomics, an understanding of proteomics tools and technologies is only slowly penetrating the research community at large. This book addresses the need to introduce biologists to new tools and approaches, and is for both students of biology and experienced, practicing biologists. Anyone who has taken a graduate level biochemistry course should be able to take from this book a reasonable understanding of what proteomics is all about and how it is practiced. The experienced biologist should encounter much here that is familiar, but refocused to facilitate studies of the proteome.

The achievement of long-sought milestones in genome sequencing, analytical instrumentation, computing power, and user-friendly software tools has irrevocably changed the practice of biology. After years of studying the individual components of living systems, we can now study the systems themselves in comprehensive scope and in exquisite molecular detail. We therefore face the tasks of effectively employing new technologies, of dealing with mountains of data, and, most important, of adjusting our thinking to understand complex systems as opposed to their individual components.

Introduction to Proteomics: Tools for the New Biology had its origins in a short course on peptide sequencing by mass spectrometry, which was taught by Dr. Donald F. Hunt at the 1998 Association of Biomedical Resource Facilities meeting in Durham, North Carolina. At that time, my colleague Dr. Tom McClure and I were establishing a new proteomics facility in the Center for Toxicology and the Arizona Cancer Center at the University of Arizona. Tom attended the Hunt course and, upon his return, taught the material to a handful of us. We subsequently put together a four-day workshop on mass spectrometry and proteomics, which we taught to 50 participants at the University of Arizona in August, 1999. The participants included graduate students, laboratory staff, and faculty. The enthusiastic response to this workshop reflected the need for some accessible means of introducing scientists to the new

techniques of proteomics and their potential applications in research. That experience provided the impetus for this book.

This is a book for beginners. My goal here is to familiarize the inexperienced reader with the important tools and applications of proteomics. Thus the description of certain instrumentation and applications is not highly rigorous. This book is not intended to be a laboratory manual or a compilation of the latest techniques. There are several excellent volumes available that provide more detailed descriptions of protein analytical techniques, mass spectrometry instrumentation and techniques, and applications of these technologies. The evolution of methods and applications in this area is now so rapid that no book really could be truly up-to-date. What is exciting about my experience in introducing proteomics to colleagues has been the creativity with which they then apply these tools. Ultimately, the exciting potential of proteomics rests with those who can put new technologies to work to address important questions.

I have divided the book into three parts. Part I introduces the subject of proteomics, describes its place in the new biology, and examines the nature of proteomes. Part II introduces the tools of proteomics research and explains how they work. Part III explains how these tools are integrated to solve different types of problems in biology.

I would like to thank Jeanne Burr, Laura Tiscareno, Julie Jones, Dan Mason, Beau Hansen, Hamid Badghisi, Linda Manza, Richard Vaillancourt, Tom McClure, Arpad Somogyi, and George Tsaprailis, who provided valuable suggestions, read and commented on several drafts of book chapters and provided sample data for some of the illustrations. I thank Elizabeth Hedger for excellent secretarial assistance. Finally, I thank my wife Karen and my son Andrew for their patience with me every time I went off with my laptop to write.

Daniel C. Liebler, PhD

Contents

Foreword by J. R. Yates, III	v
Preface	vii
I. Proteomics and the Proteome	1
1. <i>Proteomics and the New Biology</i>	3
2. <i>The Proteome</i>	15
II. Tools of Proteomics	25
3. <i>Overview of Analytical Proteomics</i>	27
4. <i>Analytical Protein and Peptide Separations</i>	31
5. <i>Protein Digestion Techniques</i>	49
6. <i>Mass Spectrometers for Protein and Peptide Analysis</i>	55
7. <i>Protein Identification by Peptide Mass Fingerprinting</i>	77
8. <i>Peptide Sequence Analysis by Tandem</i> <i>Mass Spectrometry</i>	89
9. <i>Protein Identification with Tandem Mass</i> <i>Spectrometry Data</i>	99
10. <i>SALSA: An Algorithm for Mining Specific Features</i> <i>of Tandem MS Data</i>	109
III. Applications of Proteomics	123
11. <i>Mining Proteomes</i>	125
12. <i>Protein Expression Profiling</i>	137
13. <i>Identifying Protein–Protein Interactions</i> <i>and Protein Complexes</i>	151
14. <i>Mapping Protein Modifications</i>	167
15. <i>New Directions in Proteomics</i>	185
Index	195