

分子克隆

实验指南系列

蛋白质纯化实验指南

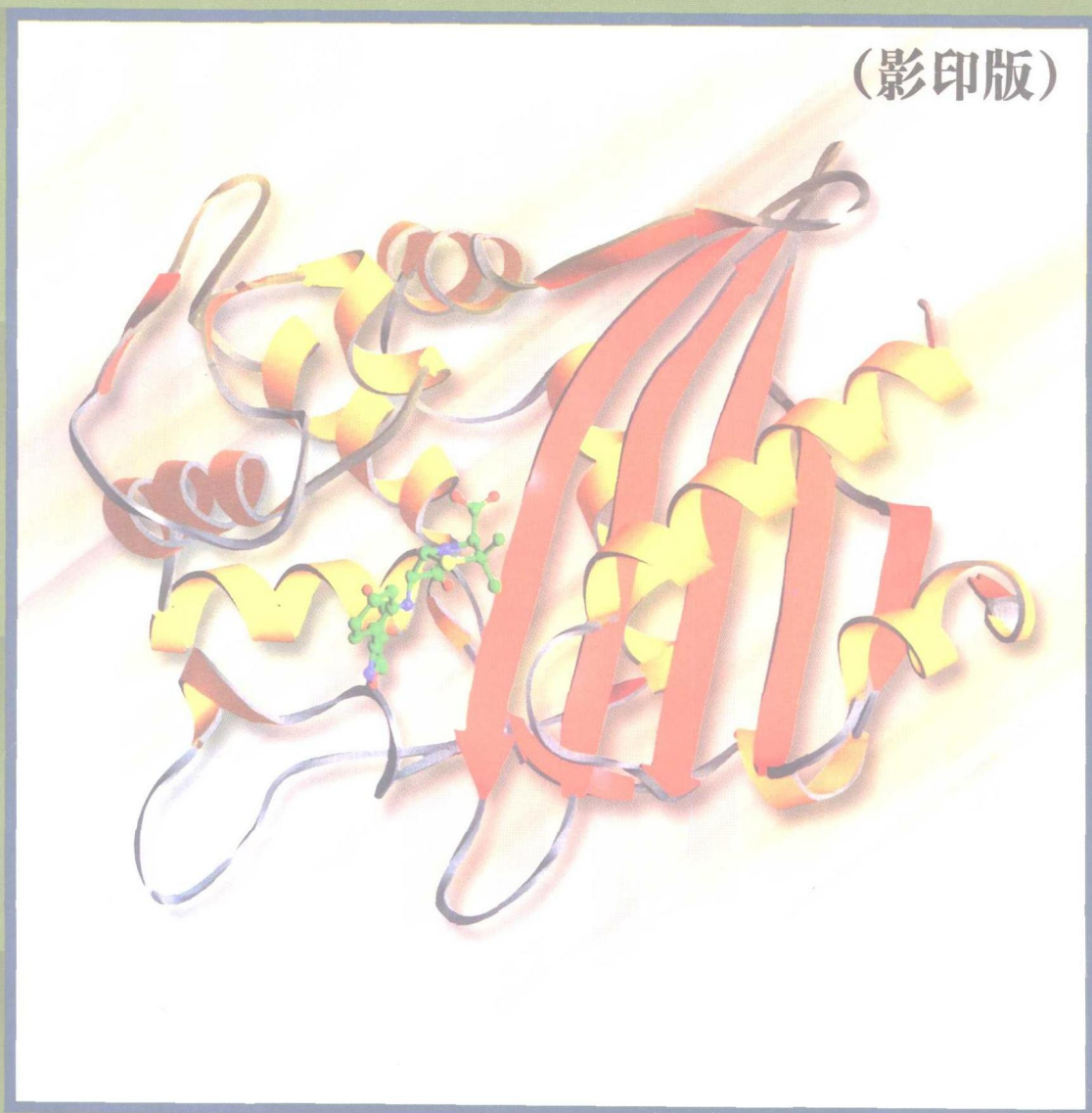
——用于蛋白质组学研究

Purifying Proteins for Proteomics

A LABORATORY MANUAL

[澳] R. J. 辛普森 主编

(影印版)



科学出版社

www.sciencep.com

内 容 简 介

蛋白质组学研究的第一步是要把蛋白质从复杂的大分子混合物中分离纯化出来。就此,本书提供了全面的背景资料、理论、实验方案和实验结果的优化方法,以及疑难解答,包括专门用于蛋白质组学研究的制备、色谱层析和电泳技术,用于结构蛋白质组学研究的蛋白质纯化分析技术等。本书是冷泉港实验室出版社最新推出的 *Purifying Proteins for Proteomics: A Laboratory Manual* 的英文影印版,是《蛋白质与蛋白质组学实验指南》的姐妹篇,秉承了冷泉港实验手册一贯的先进性、实用性和权威性,内容丰富,信息全面,方案设计完善,技术适用范围较广。

本书是遗传学、分子生物学者从基因组学、基因型研究转向蛋白质组和表型研究的必备工具书,适于生物化学、分子生物学、细胞生物学、遗传学、免疫学、蛋白质组学、功能基因组学等生命科学相关领域研究院所、高校相关院系、实验室的教师、研究生、科研人员,以及生物技术企业的研发者和决策者参考使用。

书名原文: *Purifying Proteins for Proteomics: A Laboratory Manual*
原著正文内彩图集中列于书末,请读者对应参考。

©2004 by Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York
All rights reserved.

AUTHORIZED EDITION FOR SALE IN P. R. CHINA ONLY

图书在版编目 (CIP) 数据

蛋白质纯化实验指南: 用于蛋白质组学研究/ (澳) 辛普森 (Simpson, R. J.) 主编. —影印本. —北京: 科学出版社, 2004.6
(分子克隆实验指南系列)

ISBN 7-03-013387-0

I. 用… II. 辛… III. 蛋白质组学-纯化-实验-方法-英文
IV. Q51-33

中国版本图书馆 CIP 数据核字 (2004) 第 038905 号

责任编辑: 莫结胜

责任印制: 安春生/封面设计: 王 浩

科 学 出 版 社 出 版

北京东黄城根北街16号

邮政编码: 100717

<http://www.sciencep.com>

中国科学院印刷厂印刷

科学出版社发行 各地新华书店经销

*

2004年6月第 一 版 开本: 890×1240 1/16

2004年6月第一次印刷 印张: 50 3/4 彩插 2

印数: 1—3 000 字数: 1 283 000

定价: 98.00 元

(如有印装质量问题, 我社负责调换〈路通〉)

Preface

WITH THE SUCCESSES OF THE MYRIAD GENOME SEQUENCING PROJECTS—more than 150 published completed projects, including the human genome—the attention of the international bioscientific community is now focusing more on the structure and function of the encoded proteins in the various genomes, i.e., trying to establish what genes really do! In response to the growing interest in this field, *Proteins and Proteomics: A Laboratory Manual* (Cold Spring Harbor Laboratory Press 2003) was published, providing researchers with a comprehensive introduction to proteomic methods, including those for isolating proteins, especially from one- and two-dimensional gels (2D gels), as well as classical “pull-down” procedures, using affinity tags, for isolating proteins and their interacting partners. Procedures for identifying proteins isolated by these means (and their posttranslational modifications, especially phosphorylation sites) using classical amino- and carboxy-terminal sequence analysis, as well as mass-spectrometry-based methods, were well covered.

Now that the field of proteomics is beginning to mature, it is clear that separation tools, other than the traditional one- and two-dimensional gels, are required for proteome research, especially if we aim to see beyond the tip of the iceberg where the vast majority of precious and rare proteins are still hidden, awaiting discovery. This is especially the case for low-abundance biomarkers of disease and certain classes of proteins that are essentially refractory to 2D gels, such as membrane proteins. For instance, for the majority of protein expression profiling experiments using 2D gels, only the most abundant proteins in a cell or tissue (i.e., the general “housekeeping” and structural proteins) are observed. This is not unexpected for the simple reason that the dynamic range of protein abundances in a cell is on the order of 10^5 to 10^6 . As an example, actin—one of the most abundant proteins in a cell—has a concentration of 10^8 molecules per cell, whereas some cellular receptors or transcription factors are present at only 10^2 to 10^3 molecules per cell. This problem can be exacerbated when studying biological specimens such as blood, where albumin is present at ~ 40 mg/ml and cytokines at low pg/ml, a dynamic range of protein abundances of $\sim 10^9$. Because the dynamic range for detection of proteins in 2D gels is $\sim 10^4$, clearly, some prefractionation techniques, preferably in combination with abundant protein depletion methods, must be invoked to move low-abundance proteins of interest onto the “radar screen.”

The aim of this book, *Purifying Proteins for Proteomics: A Laboratory Manual*, which is a companion to *Proteins and Proteomics: A Laboratory Manual*, is to provide the researcher with important purification strategies (Section I) and prefractionation approaches, both chromatographic (Section II) and electrophoretic (Section III), to circumvent the dynamic range impasse. Along with these prefractionation approaches, a variety of classical protein separation methods are covered. These are methods designed to facilitate the high-through-

put production of purified proteins for the purpose of high-resolution three-dimensional structure analysis, i.e., “structural genomics” (also referred to as structural proteomics) and protein microarray (protein chip) analysis. In addition, Section IV describes methods for determining the functional integrity of purified proteins (e.g., measurement of conformational stability, accurate molecular weight, aggregation state, and binding characteristics using surface plasmon resonance). Included in Section IV is a chapter devoted to the analysis of carbohydrate from glycoproteins; protocols are provided that detail the removal of glycans from glycoproteins and the preparation of monosaccharides for identification by gas chromatography coupled to mass spectrometry.

Purifying Proteins for Proteomics: A Laboratory Manual is aimed at those investigators who wish to isolate proteins and peptides for subsequent proteomic analysis. It is written for an audience ranging from early graduate students to experienced investigators studying problems in a broad range of disciplines, from clinical to basic sciences, and applied to a vast array of proteomic problems. I felt that a single author would be unable to do justice to the entire panoply of currently available methods. Accordingly, the book contains contributions from some of the experts in each of the specific fields that are covered.

In addition to the authors of the various contributions included in this book, I am greatly indebted to a number of people who have made significant contributions to bring this project to fruition. First, I thank the editorial and production staff at Cold Spring Harbor Laboratory Press for their dedication and tireless efforts in checking references, facts, and faulty constructions, and for keeping the book (and myself) on schedule. A special debt of gratitude goes to Inez Sialiano for coordinating the project, Dorothy Brown for editorial assistance, Susan Schaefer for page layout, Denise Weiss for her elegant design of the book, and most of all, Michael Zierler for his unstinting support as Project Manager in steering the book to completion. I also acknowledge the generous support of Jan Argentine, my Managing Editor, and John Inglis, the Director of Cold Spring Harbor Laboratory Press, for overseeing the project. I am indebted to a number of my colleagues in the “Parkville Strip” (especially, Tom Garret, Ray Norton, Geoff Howlett, Heung-Chin Cheng, Lindsay Sparrow, and Herbert Treutlein) for their critical reading of various chapters of the book. In addition, I wish to express my gratitude to Professor Joe Sambrook, senior author of the extraordinary manual *Molecular Cloning*, whose practical editorial guidance, interest, and encouragement throughout the writing of both of the Proteomics Laboratory Manuals have been invaluable. I must also thank Mary Whitham for her secretarial help in the early stages of this book and, more recently, my personal assistant, Simone Pakin, for her secretarial excellence and superb informatics skills. Finally, I would like to thank my partner, Donna Dorow, for her never-failing support during this endeavor.

Richard J. Simpson

Purifying Proteins for Proteomics Companion Web Site

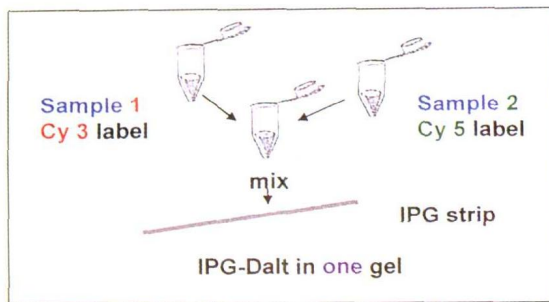
THIS VOLUME *PURIFYING PROTEINS FOR PROTEOMICS: A LABORATORY MANUAL* will share a Web Site (www.proteinsandproteomics.org) with its companion volume, *Proteins and Proteomics: A Laboratory Manual*. In addition to the on-line information already supplementing the latter volume, the Web Site will contain:

- All references from this volume linked to Medline.
- Expanded set of reference tables.
- Link to www.biosupplynet.com for suppliers mentioned in this volume.
- Further links to useful databases and Web Sites.

Additional information will be added after the book is published. To access the Web Site:

1. Open the home page of the site.
2. Follow the simple registration procedure that begins on that page (no unique access code is required).
3. Your e-mail address becomes your log-in information for subsequent visits to the site.

The FAQ section of the site contains answers about the registration procedure. For additional assistance with registration, to inform us of other Web address changes, and for all other inquiries about the proteinsandproteomics.org Web Site, please e-mail support@proteinsandproteomics.org or call 1-800-843-4388 (in the continental U.S. and Canada) or 1-516-422-4100 (all other locations) between 8:00 A.M. and 5:00 P.M. Eastern U.S. time.



DIGE

of high pressure-stress
induced *Lactobacillus*
sanfranciscensis proteins

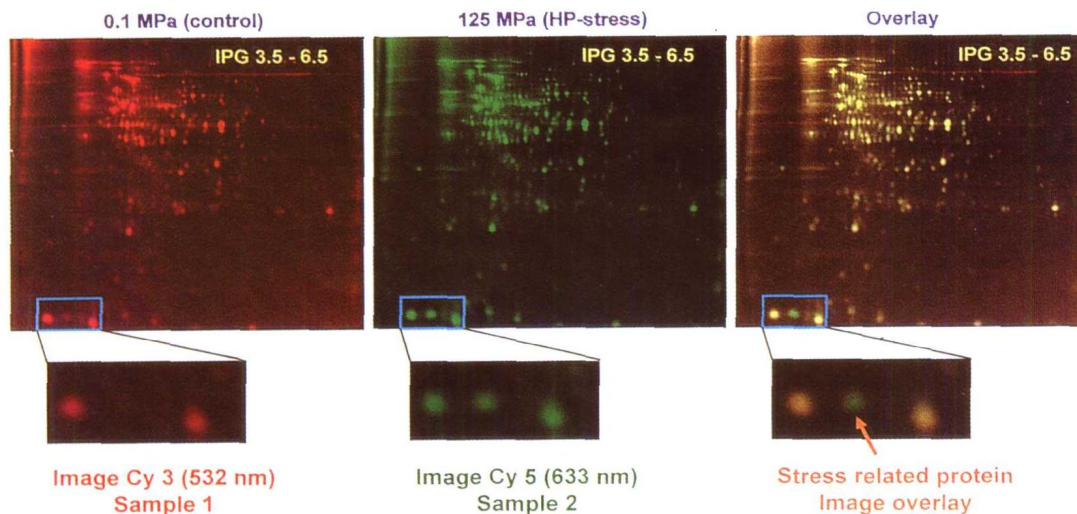


FIGURE 16.4. Difference gel electrophoresis of high-pressure inducible *Lactobacillus sanfranciscensis* proteins. Samples (Control [at 0.1 Mpa] and high-pressure stressed [at 125 MPa]) were labeled in vitro with two different fluorescent cyanine dyes (Cy3 and Cy5, respectively) differing in their excitation and emission wavelengths. The samples were mixed, and the mixture was separated on a single 2D gel. After consecutive excitation with both wavelengths, the resultant gel images were overlaid to visualize differences (e.g., up- or down-regulated proteins) between the samples.

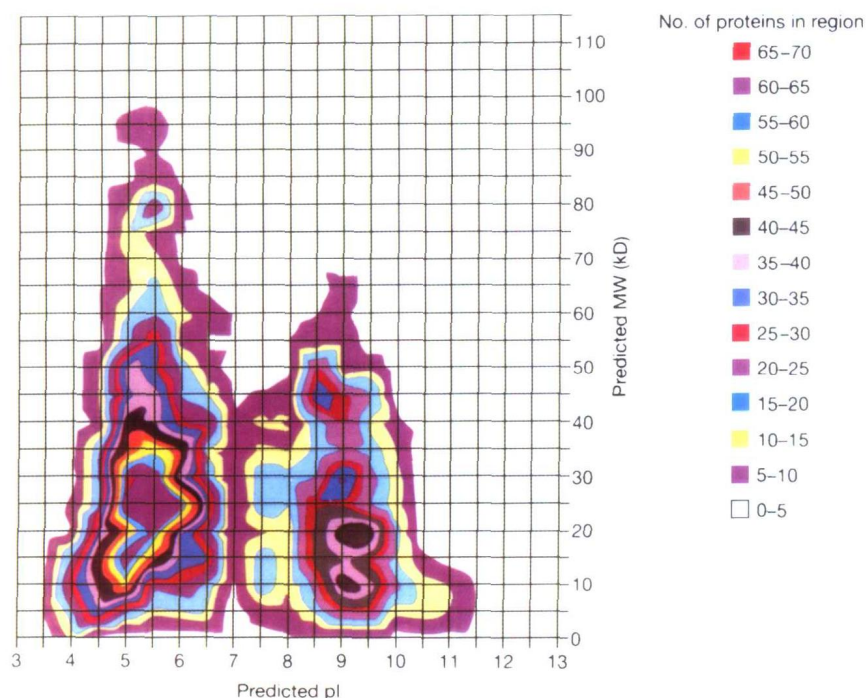


FIGURE 17.1. Theoretical *E. coli* proteome.

Syndrome	β -Linked Protein and Inclusion	
Familial encephalopathy with neuroserpin inclusion bodies, progressive myoclonic epilepsy, atypical myoclonic disease	Neuroserpin	Collins body
Familial Parkinson's disease, dementia with Lewy bodies, Lewy-body variant of Alzheimer's disease	α -Synuclein	Lewy body
Creutzfeldt-Jakob disease, bovine spongiform encephalopathy, variant Creutzfeldt-Jakob disease, Gerstmann-Sträussler-Scheinker disease, fatal familial insomnia, kuru	Prion protein	Variant Creutzfeldt-Jakob disease amyloid
Alzheimer's disease, Down's syndrome, familial Alzheimer's disease	β -Amyloid peptide	β -Amyloid plaque
Frontotemporal dementia (Pick's disease)	Tau protein	Pick body
Huntington's disease, dentatorubral-pallidoluysian atrophy (Haw-River syndrome), spinocerebellar ataxia types 1, 2, and 3 (Machado-Joseph disease)	Glutamine repeats (huntingtin)	Huntingtin inclusion

FIGURE 20.1. Classification of the dementias and neurodegenerative diseases on the basis of underlying protein abnormality. The typical endpoint formation of cellular inclusions or pericellular deposits of each protein illustrates the shared mechanism of disease resulting from the cross-bonding (β -linkage) aggregation of the conformationally unstable proteins. The glutamine-repeat disorders arise from a variety of proteins containing aberrant expansions of glutamine repeats; a prime example is the protein huntingtin in Huntington's disease. Only some forms of familial Parkinson's disease are due to α -synuclein (Lewy-body) aggregation (inclusion bodies, x400 to x600; amyloid deposits, x100; huntingtin inclusion, x800). (Reproduced, with permission, from Carrell and Lomas 2002b.)

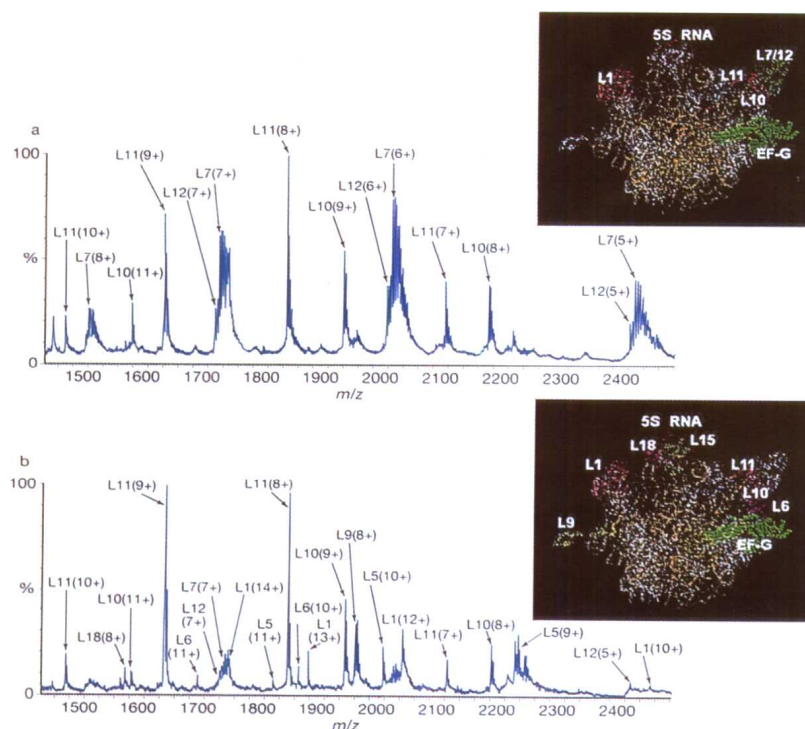


FIGURE 21.15. Mass spectra of the m/z region 1450–2500 of the ribosome EF-G complex in the presence of fusidic acid (a) and thiostrepton (b). The two spectra are markedly different. The low- m/z region of the mass spectrum recorded in the presence of fusidic acid is similar to that observed for ribosomes in the absence of EF-G under these solution and MS conditions. In contrast, the complex inhibited by thiostrepton demonstrates the absence of L7/L12 and the presence of additional proteins L5, L6, and L18. The structures of the 50S ribosomal subunit and EF-G are shown in the insets to each figure and the proteins colored in the two structures are those released from the two complexes. (Reproduced, with permission, from Hanson et al. 2003.)

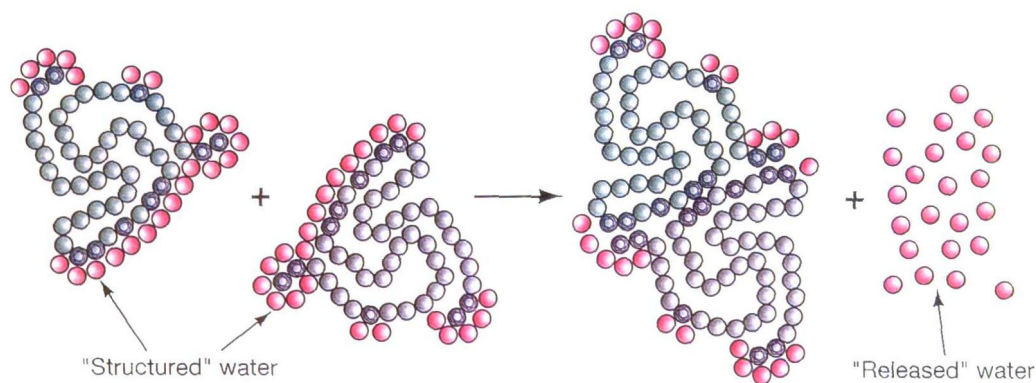


FIGURE A1.5. Solubilization of nonpolar solutes in water. Nonpolar solutes are surrounded by *ordered* shells of water molecules. The association of two nonpolar solute molecules allows some water molecules to return to a *less ordered* and “thermodynamically favorable” state, thereby facilitating their solubility. (Gray or green balls) Hydrophilic amino acids; (blue balls) hydrophobic amino acids; (pink balls) water molecules.

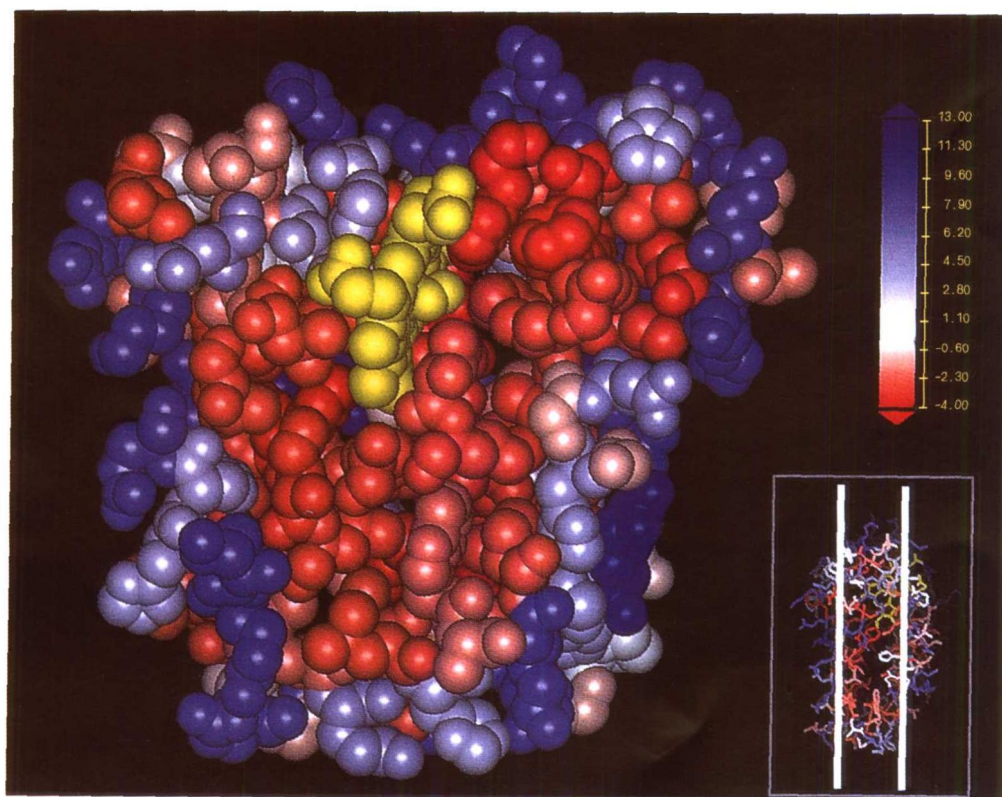
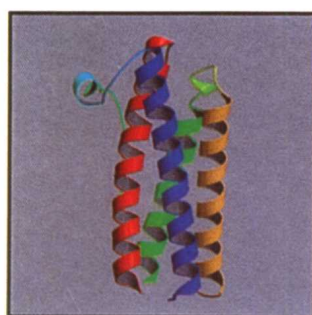
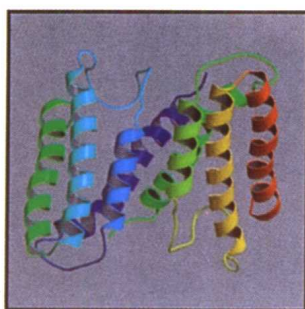


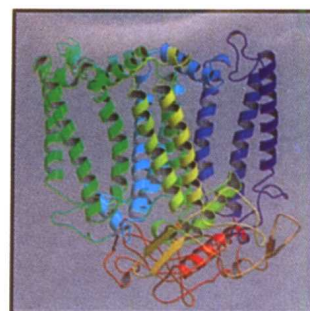
FIGURE A1.6. Distribution of hydrophobic and hydrophilic amino acids observed in myoglobin. Shown is a slice of myoglobin (PDB code 1A6M) through its center, allowing a view of the distribution of amino acids in the interior and on the surface of the protein (*insert*: the slice taken is the area between the white lines). (Red) Highly hydrophilic residues; (blue) highly hydrophobic residues. Residues with intermediate character are color-coded according to the Engelman-Steitz-Goldman hydrophobicity scale (Engelman et al. 1986) shown in the right panel. (Yellow) The prosthetic haem group observed in myoglobin spans a distance almost from the center of the protein to its surface. The hydrophobic core of the protein (composed of mainly red amino acids) is clearly visible, as is the hydrophilic (blue) surface layer. (This figure was kindly provided by Herbert Treutlein.)



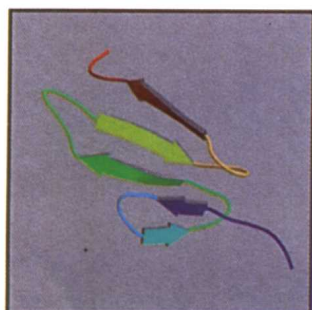
1rhg



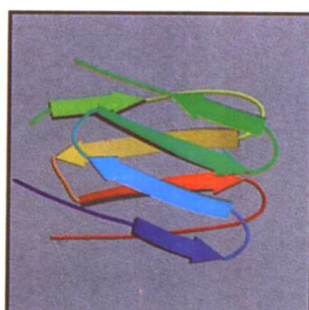
2ccy



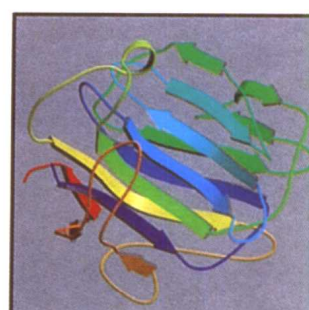
1pcr



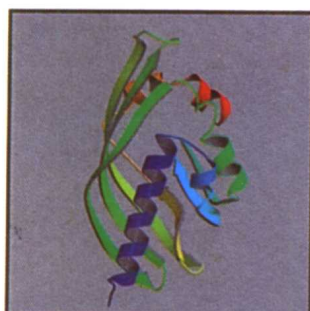
1tpm



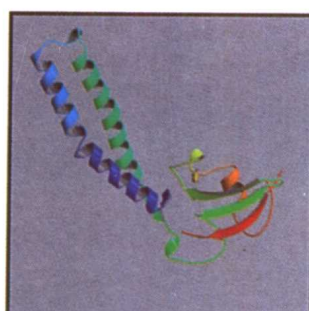
1btd



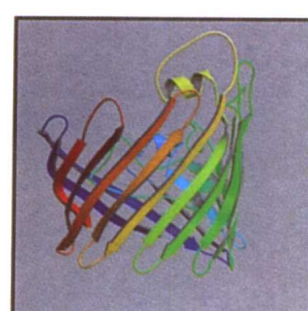
1cpn



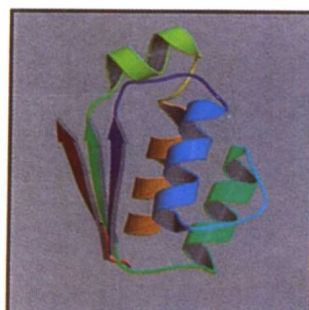
1std



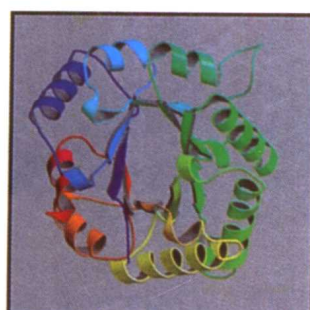
1grj



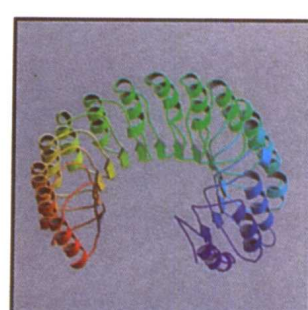
2por



1bta



1hti



2bnh

FIGURE A1.7. (See facing page for legend.)

Contents

Preface

Purifying Proteins for Proteomics Companion Web Site

Section I: Preparing Proteins for Proteomic Analysis

- | | | |
|----------|--|----|
| 1 | Role of Separation Science in Proteomics | 1 |
| | <i>Richard J. Simpson</i> | |
| 2 | Strategies for Purifying Proteins | 17 |
| | <i>Richard J. Simpson</i> | |

Section II: Chromatographic Methods for Preparing Proteins for Proteomics

- | | | |
|----------|---|-----|
| 3 | Introduction to Chromatographic Methods for Protein and Peptide Purification | 41 |
| | <i>Richard J. Simpson</i> | |
| 4 | Multidimensional Chromatography of Intact Proteins | 75 |
| | <i>Alex Apffel</i> | |
| 5 | High-throughput Screening of Soluble Recombinant Proteins | 101 |
| | <i>Ting-Fang Wang and Andrew H.-J. Wang</i> | |
| 6 | Ion-exchange Chromatography | 121 |
| | <i>Bengt Westerlund</i> | |
| 7 | Size-exclusion Chromatography | 147 |
| | <i>Helena Hedlund</i> | |
| 8 | Role of Reversed-phase High-performance Liquid Chromatography in Protein Isolation and Purification | 179 |
| | <i>C. David Carr and Robert L. Moritz</i> | |
| 9 | Hydrophobic Interaction Chromatography | 209 |
| | <i>Richard J. Simpson and Paul A. O'Farrell</i> | |

10	Affinity and Immunoaffinity Chromatography	221
	<i>Keith Brocklehurst, Albert J. Courey, Sheraz Gul, Sue-Hwa Lin, and Robert L. Moritz</i>	
11	Dye-Ligand Affinity Chromatography in Proteomic Research	275
	<i>Nikos E. Labrou</i>	
12	Immobilized Metal-ion Affinity Chromatography	301
	<i>Lars Haneskog</i>	
13	Protein Chromatography on Hydroxyapatite	337
	<i>Richard J. Simpson</i>	
14	Chromatofocusing	355
	<i>Jamil Mohammad</i>	

Section III: Electrophoretic Methods for Preparing Proteins for Proteomics

15	Introduction to Strategies Based on Electrophoresis for Separating Proteins	381
	<i>Richard J. Simpson</i>	
16	Separation of Proteins Using Two-dimensional Gel Electrophoresis	391
	<i>Angelika Görg, Oliver Drews, and Walter Weiss</i>	
17	Sample Preparation for High-resolution Two-dimensional Electrophoresis by Isoelectric Fractionation in an MCE	431
	<i>Ben R. Herbert, Pier Giorgio Righetti, John McCarthy, Jasmine Grinyer, Annalisa Castagna, Matthew Laver, Matthew Durack, Gerard Rummery, Rebecca Harcourt, and Keith L. Williams</i>	
18	An Electrophoresis Method for Purifying Proteins: The Gradiflow BF400 Instrument	443
	<i>Vicki L. Locke and Theresa M. Thomas</i>	
19	Protein Purification Using Free-flow Electrophoresis	463
	<i>Peter J.A. Weber, Gerhard Weber, and Christoph Eckerskorn</i>	

Section IV: Analysis of Purified Proteins prior to Structural Proteomics

20	Introduction to Methods for Characterizing Purified Proteins	479
	<i>Richard J. Simpson</i>	
21	Characterization of Proteins by Mass Spectrometry	489
	<i>Gavin E. Reid</i>	
22	Analytical Ultracentrifugation: Protein Size, Conformation, and Interactions	517
	<i>Christopher A. MacRaild and Geoffrey J. Howlett</i>	

23	Measuring the Conformational Stability of a Protein	535
	<i>Gerald R. Grimsley, Beatrice M.P. Huyghues-Despointes, C. Nick Pace, and J. Martin Scholtz</i>	
24	Combining Surface Plasmon Resonance with Mass Spectrometry: Identifying Binding Partners and Characterizing Interactions	567
	<i>Jos Buijs and Tohru Natsume</i>	
25	Analysis of Carbohydrate from Glycoproteins	579
	<i>David Oxley, Graeme Currie, and Antony Bacic</i>	

Appendices

1.	An Overview of Protein Structure	637
	<i>Richard J. Simpson</i>	
2.	Measuring the Concentration of Proteins	659
	<i>Richard J. Simpson</i>	
3.	Concentrating Solutions of Protein	701
	<i>Richard J. Simpson</i>	
4.	Stabilization of Proteins for Storage	733
	<i>Richard J. Simpson</i>	
5.	One-dimensional Polyacrylamide Gel Electrophoresis	751
6.	Troubleshooting Guides	765
	HPLC	
	2D Gels	
7.	Cautions	777
	Index	785

Role of Separation Science in Proteomics

Richard J. Simpson

*Joint ProteomicS Laboratory (JPSL) of the Ludwig Institute for Cancer Research
and the Walter and Eliza Hall Institute of Medical Research, Melbourne, Australia*

PRODUCTION OF PURIFIED PROTEINS: A BOTTLENECK FOR PROTEIN MICROARRAYS AND STRUCTURAL PROTEOMICS, 3

MASS SPECTROMETRY-BASED PROTEIN IDENTIFICATION, 4

A KEY TECHNICAL CHALLENGE CONFRONTING PROTEOMICS: REDUCING SAMPLE COMPLEXITY, 4

MS-based Approaches for Simplifying Complex Peptide Mixtures, 6

Dynamic Range of Protein Expression: Is There a Need for Subfractionation?, 7

OVERVIEW OF THE BOOK, 11

REFERENCES, 14

RECOMMENDED READING, 15

WWW RESOURCES, 15

PROTEOMICS—THE SYSTEMATIC ANALYSIS OF PROTEINS—is a branch of “post-genome” discovery science that aims to unravel the biological complexity encoded by the genome at the protein level. There are two main facets of proteomics:

- **Expression proteomics** aims to catalog the proteome, i.e., the full complement of all the proteins expressed by the genome in any given cell or tissue at a given time.
- **Targeted proteomics** attempts to determine the cellular functions of genes directly at the protein level (Figure 1.1).

The former approach may be regarded as an exercise in obtaining a protein atlas (or database) for a given genome that should be viewed as a static entity. Targeted proteomics, in contrast, aims to detect the dynamic expression of all proteins in a cell, including all protein isoforms and their posttranslational modifications (PTMs) (for a review of posttranslational modifications, see Mann and Jensen 2003), protein-protein interactions and their cellular localization (Golemis 2002; Phizicky et al. 2003; Simpson 2003a), organelle composition, and the structural description of proteins both individually and in their higher-order complexes (Sali et al. 2003; Simpson 2003a). For an overview of the impact that proteomics has made to the biological sciences in the past decade, and future challenges for the technique, see Patterson and Aebersold (2003).

Proteomics:	The physical mapping of cellular proteins and their interacting partners.
<hr/>	
Expression proteomics:	Study of global changes in protein expression in disease.
Targeted proteomics:	Systematic study of protein-protein interactions through the isolation of protein complexes.
<hr/>	
Objectives:	
	<ul style="list-style-type: none"> • Expression proteomics: Define the proteome, i.e., <i>create an atlas</i> of proteins in a given cell/tissue at a given time. • Targeted proteomics: Use proteomics as a <i>biological tool or assay</i> to define physiological function of a gene.

FIGURE 1.1. The two major facets of proteomics: Expression proteomics and targeted proteomics.

Proteomics is a multifaceted, rapidly evolving, and open-ended endeavor. Today, proteomics embraces several technical disciplines (also called **platform technologies**), all of which are of equal importance for a successful outcome and which face significant technical challenges. The major platform technologies include:

- **Sample preparation.** One-dimensional (1D) or two-dimensional (2D) slab gel electrophoresis, liquid-based isoelectric focusing (IEF), chromatography, affinity capture methods, and various multidimensional combinations of these methods are being employed in the service of proteomics (for an overview of sample preparation methodologies, see Simpson 2003a, and Sections II and III of this volume). Indeed, success in proteomics very much depends on careful study design and the availability of high-quality biological samples. Fundamental issues, such as biological variability, preanalytical factors, and analytical reproducibility—issues that beset genomics and gene expression (microarray) studies—are also of paramount importance in proteomics (for a review of biomedical study design and sample quality, see Boguski and McIntosh 2003).
- **Mass spectrometry (MS).** This has become the method of choice for the analysis of complex protein mixtures, largely as a result of the discovery and development of soft ionization methods for proteins, which were recognized with the 2002 Nobel Prize in chemistry. For a review of general MS-based proteomics, see Aebersold and Goodlett (2001) and Aebersold and Mann (2003), and for a “hands-on” description of current MS-based proteomics methods, see the proteomics laboratory manual of Simpson (2003a).
- **Informatics.** With the advent of rapid, high-throughput MS-based methods for analyzing complex protein mixtures, our ability to generate data now clearly outstrips our ability to analyze it. In fact, data analysis is considered by many researchers to be “the Achilles heel of proteomics” (Patterson 2003). (For MS-derived data, sample documentation, and annotation of gene function, see Ashburner et al. [2000]; Camon et al. [2003]; see also <http://www.geneontology.org>.) This dilemma has prompted a systematic approach to modeling, capturing, and disseminating proteomics experimental data, using agreed-upon early-stage

documentation, XML-based definitions, and controlled vocabularies that allow different tools to exchange primary data sets (Taylor et al. 2003; Tyers and Mann 2003).

Other powerful proteomic approaches for defining the biological complexity of proteins include the following:

- Cell imaging by light and electron microscopy, confocal microscopy using fluorescently tagged proteins, and FRET (fluorescence resonance energy transfer) technology for defining intracellular localization of proteins and their binding partners (Phizicky et al. 2003; Simpson 2003a).
- Surface plasmon resonance (SPR) studies (Chapter 24), array- and chip-based proteomics, and genetic readout experiments (e.g., yeast two-hybrid assay) for defining protein-protein interactions (Phizicky et al. 2003).
- Protein microarray technologies, which have recently emerged (Zhu and Synder 2001), offer the potential for high-throughput genomic-scale analysis of protein expression, interaction, and function (for a current state-of-the-art assessment of protein arrays, see Cutler 2003).

PRODUCTION OF PURIFIED PROTEINS: A BOTTLENECK FOR PROTEIN MICROARRAYS AND STRUCTURAL PROTEOMICS

Today, one of the bottlenecks to several proteomics approaches, especially protein microarrays (Houseman et al. 2002; Mitchell 2002; Templin et al. 2002) and structural proteomics (Chance et al. 2002), is the production of large arrays of highly purified, correctly folded proteins. This has been largely due to the limited availability of validated genome-wide complementary DNA and to the production of correctly folded recombinant proteins. The FlexGene consortium between academic institutes and industry, described by LaBaer (Braun et al. 2002), has been established to develop a comprehensive cDNA collection in recombination-based cloning formats for the biomedical community (see <http://www.hip.harvard.edu>). A comparison of commonly used tagging methods for the rapid purification of recombinant proteins for the purpose of array-based proteomics and structural proteomics purposes is described in Chapter 5.

CAVEATS IN PROTEOMICS

"Aside from data analysis, the accuracy of quantitative data is an issue in proteomics, whether it be based on two-dimensional gels or other means, but this needs to be considered in light of the experimental aim. In some instances, knowing that expression varies by 20% could be critical; in others, 200% changes may represent the minimum change considered of interest. In addition, proteomics measurements are essentially all relative, because global measures of absolute abundance are currently not feasible, although they are being discussed (Patterson and Aebersold 2003). ...proteomics is substrate-limited and, together with the somewhat limited dynamic range of protein concentrations interrogated, only a portion of the proteome can be examined at one time. This is an important caveat that was lost on many investigators over the past few years in the hype of what proteomics could produce." (Patterson 2003)

MASS-SPECTROMETRY-BASED PROTEIN IDENTIFICATION

MS is now firmly entrenched as the method of choice for identifying proteins and their post-translational modifications. In general, MS-based methods identify proteins, and to some extent their posttranslational modifications, not by analyzing the intact proteins directly, but by identifying their constituent peptides (Figure 1.2). Thus, the task of identifying a protein by MS is reduced to one of identifying peptides. First, proteins are proteolytically cleaved (typically, with trypsin) into smaller peptides. Resultant peptides are then separated and analyzed in a mass spectrometer. MS data acquired using this approach are processed through a series of computer algorithms that determine the identity of a protein on the basis of the masses of the constituent peptides (peptide mass fingerprinting), correlation of collision-induced fragmentation (CID) spectra of peptides obtained by tandem mass spectrometry (MS/MS), and/or combinations of these approaches (Simpson 2003a).

These approaches introduce several advantages. First, small peptides are more easily eluted from acrylamide gels than are proteins. (During the early days of proteomics, and until relatively recently, two-dimensional polyacrylamide gel electrophoresis [2D-PAGE] in combination with MS has been the mainstay of protein identification.) Second, MS has greater sensitivity for small molecules, and, finally, small peptides yield better CID fragmentation data than large intact proteins. It should be emphasized that only a small subset of peptides, not the entire sequence of a protein, is typically determined using this MS approach. Sequence coverage typically is in the 30–80% range. Indeed, using MS/MS, a protein in a complex mixture in many cases can be identified after the identification of only a few peptides. Note also that protein identification by MS-based peptide sequencing relies on obtaining experimentally derived data and the use of these data to search peptide data deposited in sequence databases. (Of course, this approach does not work if the genome sequence is not in the databases. In these situations, the peptide sequence can be determined from MS/MS spectra by manual “de novo” sequence approaches followed by conventional informatics approaches [Verhagen et al. 2000; see also Simpson 2003a].)

Promising MS methods are currently being developed that allow fragmentation of intact proteins in the mass spectrometer, followed by identification of the protein using MS/MS of the generated protein fragments. This approach is referred to as “top-down sequencing” as compared to the present “bottom-up” sequencing (Simpson 2003b).

A KEY TECHNICAL CHALLENGE CONFRONTING PROTEOMICS: REDUCING SAMPLE COMPLEXITY

Many of the proteomics methods mentioned thus far pose technical challenges owing to the high degree of complexity of cellular proteomes and the large dynamic-range considerations of cellular proteins ($\sim 10^5$ – 10^6 -fold for protein abundance alone). In fact, high-throughput proteomic analysis of complex protein (or peptide) mixtures remains an apparently insuperable technical challenge, at least for the foreseeable future. In recent years, there have been two main strategies for overcoming this problem: MS-based strategies for dealing with complex peptide mixtures and sample prefractionation prior to MS-based analysis. Given the enormity of the problem in understanding the complexity of cellular and tissue proteomes, a combination of these two approaches, undoubtedly, will be required.

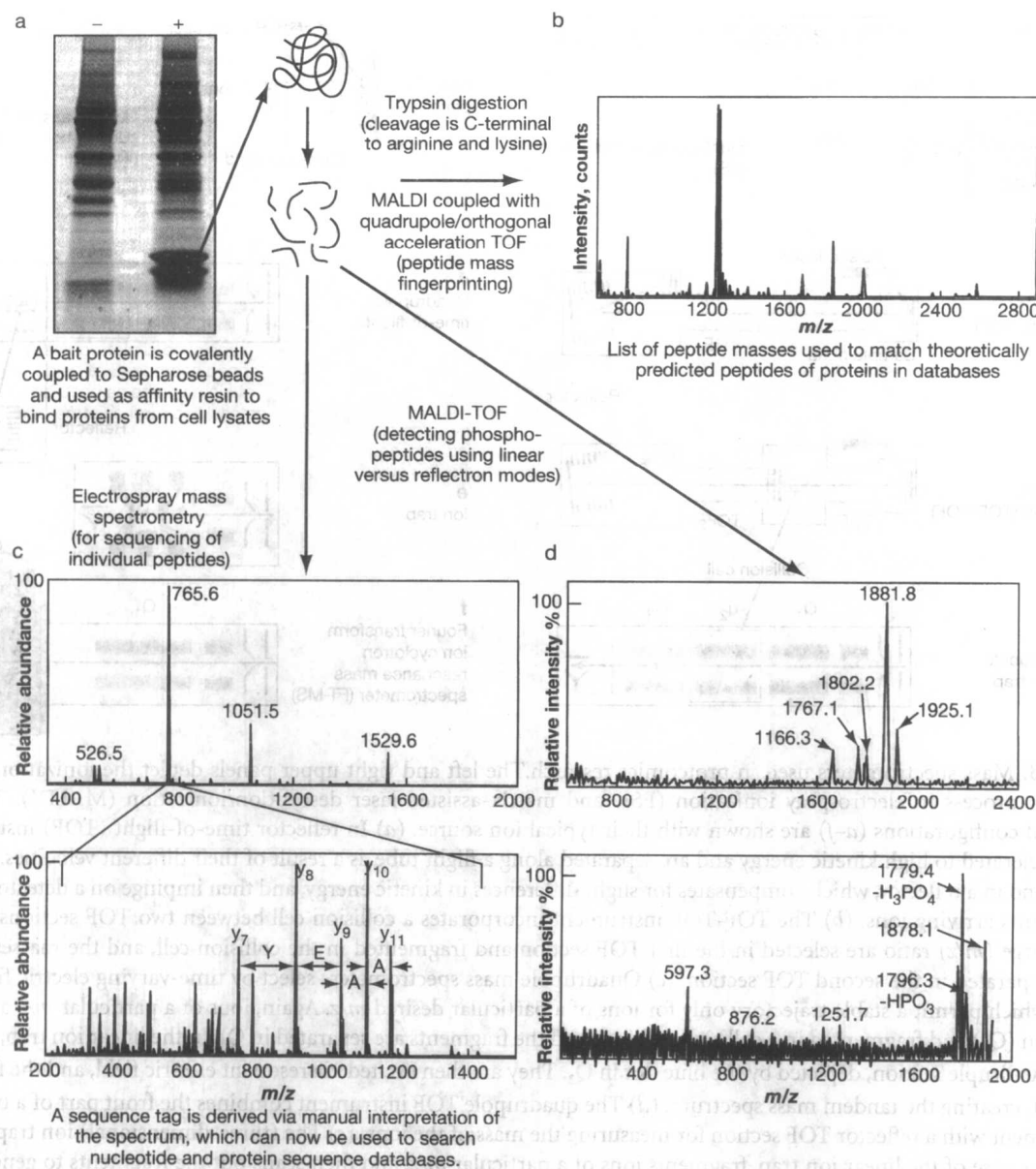


FIGURE 1.2. A strategy for mass spectrometric identification of proteins and posttranslational modifications. (a) Affinity capture using a bait protein to isolate protein-binding partners from a cell lysate. A GST fusion protein containing the SOCS-1 SOCS box sequence (Zhang et al. 1999) is used to illustrate the principle of affinity capture. SDS-PAGE analysis of affinity column eluates from GST-control (–), and GST-SOCS-1-SOCS-box glutathione (+) are shown. After 1D gel electrophoresis, the gel was stained with Coomassie and the protein bands of interest excised and subjected to trypsin digestion (see Simpson 2003). (b) Analysis of an aliquot of the tryptic peptide mixture using a MALDI quadrupole/orthogonal acceleration TOF mass spectrometer (see Chapter 8 in Simpson 2003). The resultant spectrum represents a peptide mass fingerprint (PMF) of a protein. The peptide masses can be entered into an algorithm, which matches them against theoretically predicted peptides of proteins in publicly available databases (see Chapter 8 in Simpson 2003). (c) Analysis of another aliquot of the peptide mixture using an electrospray ion-trap mass spectrometer, which is coupled on-line to a capillary RP-HPLC. (Top panel) Peptide masses at a given time. A typical experiment entails isolating the most intense ion in the spectrum (i.e., m/z 765.6), and performing collision-induced dissociation to generate sequence ions. (Bottom panel) From the resultant MS/MS spectrum, amino acid sequence information can be derived via manual interpretation or by using an algorithm that correlates the experimental spectrum with those in a database. (d) For phosphopeptide analysis, peptide-containing fractions from a capillary RP-HPLC separation are subjected to MALDI-TOF analysis operated in “linear mode” (top panel) and also in “reflectron mode” (bottom panel). It can be seen that the major peptide ion detected in “linear mode” (m/z 1881.8) is metastable, resulting in a major ion loss of 98 daltons ($-H_3PO_4$) and a lesser ion loss of 80 daltons ($-HPO_3$) that are observable in “reflectron mode.” Ion losses of 98 and 80 daltons, or multiples thereof, indicate that a parent ion is phosphorylated (see Chapter 9 in Simpson 2003) (Zugaro et al. 1998).