Methods in Molecular Biology 753

Springer Protocols

Kris Gevaert
Joël Vandekerckhove *Editors*

Gel-Free Proteomics

Methods and Protocols

Humana Press

METHODS IN MOLECULAR BIOLOGY™

Series Editor
John M. Walker
School of Life Sciences
University of Hertfordshire
Hatfield, Hertfordshire, AL10 9AB, UK

For other titles published in this series, go to
www.springer.com/series/7651

Gel-Free Proteomics

Methods and Protocols

Edited by

Kris Gevaert

VIB and Ghent University, Ghent, Belgium

Joël Vandekerckhove

VIB and Ghent University, Ghent, Belgium

Humana Press

Editors
Kris Gevaert
VIB Department of Medical Protein Research and UGent Department of Biochemistry
VIB and Ghent University
Albert Baertsoenkaai 3
Ghent 9000, Belgium
kris.gevaert@vib-ugent.be

Joël Vandekerckhove
VIB Department of Medical Protein Research and UGent Department of Biochemistry
VIB and Ghent University
Albert Baertsoenkaai 3
Ghent 9000, Belgium
joel.vandekerckhove@vib-ugent.be

ISSN 1064-3745 e-ISSN 1940-6029
ISBN 978-1-61779-147-5 e-ISBN 978-1-61779-148-2
DOI 10.1007/978-1-61779-148-2
Springer New York Dordrecht Heidelberg London

Library of Congress Control Number: 2011927918

Printed on acid-free paper

Humana Press is part of Springer Science+Business Media (www.springer.com)

Preface

Proteomics by means of mass spectrometry has rapidly changed the way that we analyze proteomes. Electrospray ionization mass spectrometry coupled with nanoscale peptide liquid chromatography in particular is currently by far the most used proteomics technology. Mass spectrometry-driven, and thus gel-free proteomics, techniques all start by digesting a proteome or an isolated sub-proteome into peptides, these being readily analyzable by mass spectrometers. Whereas the generated peptide mixtures are very information rich, given that almost all extracted proteins will finally be represented by one or more peptides, such mixtures are also very complex as they hold tens of thousands of peptides present in highly different concentrations. Contemporary mass spectrometers are still not able to fully cope with such highly complex mixtures of analytes, and therefore several intelligent solutions have been proposed, many of which are described here. Especially cumbersome turned out to be the analysis of modified peptides, since these are present at much lower levels as non-modified peptides and are thus outcompeted for ionization and detection.

Gel-Free Proteomics: Methods and Protocols addresses contemporary methods for gel-free proteome research with a special focus on differential analysis and protein modifications.

Chapter 1 starts with a perspective overview of gel-free proteome analytical approaches, explaining their raisons d'être, potentials, and pitfalls. Given that proteomics is typically used to discover differentially expressed proteins or protein modifications, a section of this book deals with isotope labeling approaches for gel-free proteomics. Metabolic labeling of organisms is described in **Chapter 2**. Of note is that not all proteome samples can be labeled metabolically and hence various ways of introducing mass tags post-metabolically are described in **Chapters 3**, **4**, **5**, and **6**. Finally, this section of the book ends with a description of the PSAQ method in which fully isotopically labeled proteins are expressed in cell-free systems and then added to proteomes to be analyzed as internal standards for subsequent absolute quantification of proteins (**Chapter 7**).

Reproducible and highly effective sample preparation is of key importance for proteome research. Our field especially needs reproducible protocols for isolating organelles and membrane proteins, and these are described in **Chapters 8** and **9**, respectively. Further, a protocol for comprehensive proteome analysis by the so-called GeLCMS method – i.e., separation of proteins by SDS-PAGE, followed by LC-MS/MS analysis of in-gel digested proteins – is described in **Chapter 10**. A promising new tool for gel-free proteomics is the metalloendopeptidase Lys-N. In **Chapter 11**, a protocol is given that exploits this Lys-N protease to enrich for amino-terminal peptides and phosphorylated peptides, as well as more basic peptides for detailed analysis of proteomes. Further, a protocol that details the use of diagonal chromatography for the identification of newly synthesized proteins is given in **Chapter 12**.

A large section of this book is dedicated to the analysis of protein modifications. Protein phosphorylation is without a doubt the most extensively studied protein modification, with high numbers of approaches reported. An overview of phosphoproteomics approaches is given in **Chapter 13**, and one of the most used approaches – enrichment

of phosphorylated peptides on titanium dioxide beads – is presented in **Chapter 14**. Protein processing by endoproteases and aminopeptidases creates novel protein amino termini, and the introduction of several technologies by which amino-terminal peptides are specifically enriched or recognized upon tandem mass spectrometric analysis was recently observed in the field of protease degradomics. The positional proteomics strategy that enriches for amino-terminal peptides by affinity removal of the biotinylated internal peptides is described in **Chapter 15**. A somewhat opposite strategy during which protein amino termini are biotinylated and affinity enriched is given in **Chapter 16**. This procedure exploits the fact that blocking a protein alpha-amino terminus by, for instance, acetylation is a typical co-translational modification, implying that the alpha-amino termini formed following protease action are not blocked and can here be biotinylated and further enriched for analysis. **Chapters 17** and **18** describe procedures from the Overall lab by which protease specificities are characterized with high detail using peptide libraries (PICS, **Chapter 17**) and by which internal peptides are removed from protein digests using high molecular weight and soluble polymers (TAILS, **Chapter 18**).

Protein glycosylation is a prominent protein modification that is reported to affect more than one-third of all proteins. In this book the use of lectins to affinity isolate glycosylated proteins and peptides is described (**Chapter 19**), next to two methods to specifically isolate glycopeptides carrying sialic acid (**Chapters 20** and **21**). Further, a novel approach for enriching for O-glycosylated peptides is given (**Chapter 22**). Finally, this section of the book ends with a protocol to characterize ubiquitination (**Chapter 23**).

A key feature of mass spectrometry-driven proteomics is the enormous amount of data that are generated per experiment. Not surprisingly these data pose high demands on data storage, analysis, and interpretation. A critical viewpoint on the various challenges that bioinformaticians face when confronted with large-scale proteomics data is presented in **Chapter 24**. Further, quantification of proteomics data aiming at identifying regulated proteins was recently automated by the introduction of several software tools. A case study on the use of several of these tools is presented in **Chapter 25**.

All procedures needed to perform gel-free proteomics are described in *Gel-Free Proteomics: Methods and Protocols*. These range from sample preparation, isotope labeling for differential proteomics, enrichment technologies for modified proteins and peptides, and bioinformatics. As such we hope that this timely and critical overview of the promises of gel-free proteomics will be a guide for researchers who are both new to the field and already working on some aspect of proteomics.

Ghent, Belgium — ***Kris Gevaert***
Ghent, Belgium — ***Joël Vandekerckhove***

Contents

Preface . . . . . *v*

Contributors . . . . . *ix*

1. Mass Spectrometry-Driven Proteomics: An Introduction 1
Kenny Helsens, Lennart Martens, Joël Vandekerckhove, and Kris Gevaert

2. Metabolic Labeling of Model Organisms Using Heavy Nitrogen (^{15}N) 29
Joost W. Gouw, Bastiaan B.J. Tops, and Jeroen Krijgsveld

3. Trypsin-Catalyzed Oxygen-18 Labeling for Quantitative Proteomics 43
Wei-Jun Qian, Brianne O. Petritis, Carrie D. Nicora, and Richard D. Smith

4. ICPL Labeling Strategies for Proteome Research 55
Friedrich Lottspeich and Josef Kellermann

5. Quantitative Proteome Analysis Using Isobaric Peptide Termini Labeling (IPTL) 65
Magnus Ø. Arntzen, Christian J. Koehler, Achim Treumann, and Bernd Thiede

6. Complete Chemical Modification of Amine and Acid Functional Groups of Peptides and Small Proteins 77
Casey J. Krusemark, Brian L. Frey, Lloyd M. Smith, and Peter J. Belshaw

7. Production and Use of Stable Isotope-Labeled Proteins for Absolute Quantitative Proteomics 93
Dorothée Lebert, Alain Dupuis, Jérôme Garin, Christophe Bruley, and Virginie Brun

8. Organelle Proteomics 117
Sophie Duclos and Michel Desjardins

9. Membrane Protein Digestion – Comparison of LPI HexaLane with Traditional Techniques 129
Ping Sui, Tasso Miliotis, Max Davidson, Roger Karlsson, and Anders Karlsson

10. GeLCMS for In-Depth Protein Characterization and Advanced Analysis of Proteomes 143
Alicia Lundby and Jesper V. Olsen

11. Exploring New Proteome Space: Combining Lys-N Proteolytic Digestion and Strong Cation Exchange (SCX) Separation in Peptide-Centric MS-Driven Proteomics 157
Nadia Taouatas, Shabaz Mohammed, and Albert J.R. Heck

12. Quantitation of Newly Synthesized Proteins by Pulse Labeling with Azidohomoalanine 169
Gertjan Kramer, Piotr T. Kasper, Luitzen de Jong, and Chris G. de Koster

13. Analytical Strategies in Mass Spectrometry-Based Phosphoproteomics 183
Heidi Rosenqvist, Juanying Ye, and Ole N. Jensen

14. A Protocol on the Use of Titanium Dioxide Chromatography for Phosphoproteomics 215
Martijn W.H. Pinkse, Simone Lemeer, and Albert J.R. Heck

15. Positional Proteomics at the N-Terminus as a Means of Proteome Simplification . 229
Gemma R. Davidson, Stuart D. Armstrong, and Robert J. Beynon

16. N-Terminomics: A High-Content Screen for Protease Substrates and Their Cleavage Sites 243
John C. Timmer and Guy S. Salvesen

17. Protease Specificity Profiling by Tandem Mass Spectrometry Using Proteome-Derived Peptide Libraries 257
Oliver Schilling, Ulrich auf dem Keller, and Christopher M. Overall

18. Identification of Proteolytic Products and Natural Protein N-Termini by Terminal Amine Isotopic Labeling of Substrates (TAILS) 273
Alain Doucet, Oded Kleifeld, Jayachandran N. Kizhakkedathu, and Christopher M. Overall

19. Lectins as Tools to Select for Glycosylated Proteins 289
Els J.M. Van Damme

20. Strong Cation Exchange Chromatography for Analysis of Sialylated Glycopeptides 299
Katharina Lohrig, Albert Sickmann, and Urs Lewandrowski

21. Titanium Dioxide Enrichment of Sialic Acid-Containing Glycopeptides 309
Giuseppe Palmisano, Sara E. Lendal, and Martin R. Larsen

22. Chemical De-O-glycosylation of Glycoproteins for Applications in LC-Based Proteomics 323
Franz-Georg Hanisch

23. Ubiquitination and Degradation of Proteins 335
Yelena Kravtsova-Ivantsiv and Aaron Ciechanover

24. Bioinformatics Challenges in Mass Spectrometry-Driven Proteomics 359
Lennart Martens

25. A Case Study on the Comparison of Different Software Tools for Automated Quantification of Peptides 373
Niklaas Colaert, Joël Vandekerckhove, Lennart Martens, and Kris Gevaert

Subject Index 399

Contributors

ULRICH AUF DEM KELLER • *Institute of Cell Biology, Swiss Federal Institute of Technology, Zürich, Switzerland*

STUART D. ARMSTRONG • *Protein Function Group, Institute of Integrative Biology, University of Liverpool, Liverpool, UK*

MAGNUS Ø. ARNTZEN • *The Biotechnology Centre of Oslo, University of Oslo, Oslo, Norway; Institute of Immunology, Rikshospitalet HF, University of Oslo, Oslo, Norway; Department of Chemistry, Biotechnology and Food Science, Norwegian University of Life Sciences, Ås, Norway*

PETER J. BELSHAW • *Departments of Chemistry and Biochemistry, University of Wisconsin, Madison, WI, USA; Best Sensors Inc., Milton, ON, Canada*

ROBERT J. BEYNON • *Protein Function Group, Institute of Integrative Biology, University of Liverpool, Liverpool, UK*

CHRISTOPHE BRULEY • *CEA, DSV, iRTSV, Laboratoire d'Etude de la Dynamique des Protéomes, Grenoble, France; INSERM, Grenoble, France; Université Joseph Fourier, Grenoble, France*

VIRGINIE BRUN • *CEA, DSV, iRTSV, Laboratoire d'Etude de la Dynamique des Protéomes, Grenoble, France; INSERM, Grenoble, France; Université Joseph Fourier, Grenoble, France*

AARON CIECHANOVER • *The Rappaport Faculty of Medicine and Research Institute, Tumor and Vascular Biology Research Center, Technion-Israel Institute of Technology, Haifa, Israel*

NIKLAAS COLAERT • *Department of Medical Protein Research, VIB, Ghent University, Ghent, Belgium; Department of Biochemistry, Ghent University, Ghent, Belgium*

GEMMA R. DAVIDSON • *Protein Function Group, Institute of Integrative Biology, University of Liverpool, Liverpool, UK*

MAX DAVIDSON • *Nanoxis AB, Gothenburg, Sweden*

LUITZEN DE JONG • *Mass Spectrometry of Biomacromolecules of the Swammerdam Institute for Life Sciences, University of Amsterdam, Amsterdam, The Netherlands*

CHRIS G. DE KOSTER • *Mass Spectrometry of Biomacromolecules of the Swammerdam Institute for Life Sciences, University of Amsterdam, Amsterdam, The Netherlands*

MICHEL DESJARDINS • *Département de pathologie et biologie cellulaire, Département de microbiologie et immunologie, Université de Montréal, Montréal, QC, Canada*

ALAIN DOUCET • *Department of Biochemistry and Molecular Biology, Centre for Blood Research, University of British Columbia, Vancouver, BC, Canada; Department of Biochemistry, Microbiology and Immunology, Ottawa Institute of Systems Biology, University of Ottawa, Ottawa, ON, Canada*

SOPHIE DUCLOS • *Département de pathologie et biologie cellulaire, Université de Montréal, Montréal, QC, Canada*

ALAIN DUPUIS • *CEA, DSV, iRTSV, Laboratoire d'Etude de la Dynamique des Protéomes, Grenoble, France; INSERM, Grenoble, France; Université Joseph Fourier, Grenoble, France*

Brian L. Frey • *Department of Chemistry, University of Wisconsin, Madison, WI, USA*
Jérôme Garin • *CEA, DSV, iRTSV, Laboratoire d'Etude de la Dynamique des Protéomes, Grenoble, France; INSERM, Grenoble, France; Université Joseph Fourier, Grenoble, France*
Kris Gevaert • *VIB Department of Medical Protein Research and UGent Department of Biochemistry, VIB and Ghent University, Ghent, Belgium*
Joost W. Gouw • *Department of Biochemistry and Molecular Biology, Centre for High-Throughput Biology, University of British Columbia, Vancouver, BC, Canada*
Franz-Georg Hanisch • *Institute of Biochemistry II, Medical Faculty, and Center for Molecular Medicine Cologne, University of Cologne, Köln, Germany*
Albert J.R. Heck • *Biomolecular Mass Spectrometry and Proteomics Group, Utrecht Institute for Pharmaceutical Sciences, Bijvoet Center for Biomolecular Research, Utrecht University, Utrecht, The Netherlands*
Kenny Helsens • *Department of Medical Protein Research, VIB, Ghent University, Ghent, Belgium; Department of Biochemistry, Ghent University, Ghent, Belgium*
Ole N. Jensen • *Department of Biochemistry and Molecular Biology, University of Southern Denmark, Odense, Denmark*
Roger Karlsson • *Department of Chemistry, University of Gothenburg, Gothenburg, Sweden*
Anders Karlsson • *Nanoxis AB, Gothenburg, Sweden*
Piotr T. Kasper • *Mass Spectrometry of Biomacromolecules of the Swammerdam Institute for Life Sciences, University of Amsterdam, Amsterdam, The Netherlands*
Josef Kellermann • *Max Planck Institute of Biochemistry, Protein Analysis, Martinsried, Germany*
Jayachandran N. Kizhakkedathu • *Department of Pathology and Laboratory Medicine, Centre for Blood Research, University of British Columbia, Vancouver, BC, Canada*
Oded Kleifeld • *Department of Biochemistry and Molecular Biology, Centre for Blood Research, University of British Columbia, Vancouver, BC, Canada; Faculty of Biology, Israel Institute of Technology, Heifa, Israel*
Christian J. Koehler • *The Biotechnology Centre of Oslo, University of Oslo, Oslo, Norway*
Gertjan Kramer • *Mass Spectrometry of Biomacromolecules of the Swammerdam Institute for Life Sciences, University of Amsterdam, Amsterdam, The Netherlands*
Yelena Kravtsova-Ivantsiv • *The Rappaport Faculty of Medicine and Research Institute, Tumor and Vascular Biology Research Center, Technion-Israel Institute of Technology, Haifa, Israel*
Jeroen Krijgsveld • *EMBL, Genome Biology Unit, Heidelberg, Germany*
Casey J. Krusemark • *Department of Biochemistry, University of Wisconsin, Madison, WI, USA; Stanford University Biochemistry, Stanford, CA, USA*
Martin R. Larsen • *Department of Biochemistry and Molecular Biology, University of Southern Denmark, Odense, Denmark*
Dorothée Lebert • *CEA, DSV, iRTSV, Laboratoire d'Etude de la Dynamique des Protéomes, Grenoble, France; INSERM, Grenoble, France; Université Joseph Fourier, Grenoble, France*
Simone Lemeer • *Netherlands Proteomics Centre, Delft, The Netherlands; Technische Universitaet Muenchen, Freising, Germany*

SARA E. LENDAL • *Department of Biochemistry and Molecular Biology, University of Southern Denmark, Odense, Denmark*
URS LEWANDROWSKI • *Department of Bioanalytics, Leibniz-Institut für Analytische Wissenschaften – ISAS e.V., Dortmund, Germany*
KATHARINA LOHRIG • *Department of Bioanalytics, Leibniz-Institut für Analytische Wissenschaften – ISAS e.V., Dortmund, Germany*
FRIEDRICH LOTTSPEICH • *Max Planck Institute of Biochemistry, Protein Analysis, Martinsried, Germany*
ALICIA LUNDBY • *Faculty of Health Sciences, Novo Nordisk Foundation Center for Protein Research, University of Copenhagen, Copenhagen, Denmark; The Danish National Research Foundation, The Panum Institute, Centre for Cardiac Arrhythmia, University of Copenhagen, Copenhagen, Denmark*
LENNART MARTENS • *Department of Medical Protein Research, VIB, Ghent University, Ghent, Belgium; Department of Biochemistry, Ghent University, Ghent, Belgium*
TASSO MILIOTIS • *AstraZeneca R&D, Mölndal, Sweden*
SHABAZ MOHAMMED • *Biomolecular Mass Spectrometry and Proteomics Group, Utrecht Institute for Pharmaceutical Sciences, Bijvoet Center for Biomolecular Research, Utrecht University, Utrecht, The Netherlands*
CARRIE D. NICORA • *Biological Sciences Division, Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, WA, USA*
JESPER V. OLSEN • *Faculty of Health Sciences, Novo Nordisk Foundation Center for Protein Research, University of Copenhagen, Copenhagen, Denmark*
CHRISTOPHER M. OVERALL • *Department of Biochemistry and Molecular Biology, Department of Oral Biological and Medical Sciences, Centre for Blood Research, University of British Columbia, Vancouver, BC, Canada*
GIUSEPPE PALMISANO • *Department of Biochemistry and Molecular Biology, University of Southern Denmark, Odense, Denmark*
BRIANNE O. PETRITIS • *Biological Sciences Division, Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, WA, USA*
MARTIJN W.H. PINKSE • *Department of Biotechnology, Technical University of Delft, Delft, The Netherlands; Netherlands Proteomics Centre, Delft, The Netherlands*
WEI-JUN QIAN • *Biological Sciences Division, Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, WA, USA*
HEIDI ROSENQVIST • *Strathclyde Institute of Pharmacy and Biomedical Sciences, University of Strathclyde, Glasgow G4 0RE, Scotland, UK*
GUY S. SALVESEN • *Program in Apoptosis and Cell Death Research, Sanford-Burnham Medical Research Institute, La Jolla, CA*
OLIVER SCHILLING • *Institute for Molecular Medicine and Cell Research, University of Freiburg, Freiburg, Germany*
ALBERT SICKMANN • *Department of Bioanalytics, Leibniz-Institut für Analytische Wissenschaften – ISAS e.V., Dortmund, Germany; Medizinisches Proteom Center, Ruhr-Universität, Bochum, Germany*
LLOYD M. SMITH • *Department of Chemistry, University of Wisconsin, Madison, WI, USA*
RICHARD D. SMITH • *Biological Sciences Division, Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, WA, USA*

PING SUI • *AstraZeneca R&D, Mölndal, Sweden*
NADIA TAOUATAS • *Biomolecular Mass Spectrometry and Proteomics Group, Utrecht Institute for Pharmaceutical Sciences, Bijvoet Center for Biomolecular Research, Utrecht University, Utrecht, The Netherlands*
BERND THIEDE • *The Biotechnology Centre of Oslo, University of Oslo, Oslo, Norway*
JOHN C. TIMMER • *Department of Pharmacology, University of California San Diego, La Jolla, CA, USA*
BASTIAAN B.J. TOPS • *Biomolecular Mass Spectrometry and Proteomics Group, Bijvoet Center for Biomolecular Research, Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Utrecht, The Netherlands*
ACHIM TREUMANN • *NEPAF Proteome Analysis Facility, Newcastle, UK*
ELS J.M. VAN DAMME • *Department of Molecular Biotechnology, Ghent University, Ghent, Belgium*
JOËL VANDEKERCKHOVE • *VIB Department of Medical Protein Research and UGent Department of Biochemistry, VIB and Ghent University, Ghent, Belgium*
JUANYING YE • *Department of Biochemistry and Molecular Biology, University of Southern Denmark, Odense, Denmark*

Chapter 1

Mass Spectrometry-Driven Proteomics: An Introduction

Kenny Helsens, Lennart Martens, Joël Vandekerckhove, and Kris Gevaert

Abstract

Proteins are reckoned to be the key actors in a living organism. By studying proteins, one engages into deciphering a complex series of events occurring during a protein's life span. This starts at the creation of a protein, which is tightly controlled on both a transcriptional (Williams and Tyler, 2007, *Curr Opin Genet Dev* **17**, 88–93) and a translational level (Van Der Kelen et al., 2009, *Crit Rev Biochem Mol Biol* **44**, 143–168). During translation, a primary strand of amino acids undergoes a complex folding process in order to obtain a native three-dimensional protein structure (Gross et al., 2003, *Cell* **115**, 739–750). Proteins take on a plethora of functions, such as complex formation, receptor activity, and signal transduction, which ultimately adds up to a cellular phenotype. Consequently, protein analysis is of major interest in molecular biology and involves annotating their presence and localization, as well as their modification state and biochemical context. To accomplish this, many methods have been developed over the last decades, and their general principles and important recent advances in large-scale protein analysis or proteomics are discussed in this review.

Key words: Mass spectrometry, peptide-centric proteomics, proteomics bioinformatics, gel-free proteomics, protein modifications.

1. Introduction

Primary information about a protein is obtained through its amino acid sequence. Published in the 1950s, the Edman sequencing methodology presented a milestone for protein analysis by enabling amino acid sequencing for the first time (4). The method first requires the protein to be purified before applying a series of modification steps. First, the amino-terminal residue is modified into a cyclic phenylthiocarbamyl, which can subsequently be released from the protein under acidic conditions

K. Gevaert, J. Vandekerckhove (eds.), *Gel-Free Proteomics*, Methods in Molecular Biology 753,
DOI 10.1007/978-1-61779-148-2_1,

as a thiazolinone amino acid derivative. This modified amino acid is then converted to a phenylthiohydantoin (PTH) amino acid, which can then be identified by chromatographic separation as each PTH amino acid has a slightly different elution profile. By iterating this process, amino acid after amino acid is released, forming a sequence ladder that starts from the protein N-terminus. The main technical drawbacks of Edman sequencing are that N-terminally blocked proteins (e.g., by acetylation) are not compatible with the protocol, and that generally only up to 30 amino acids can be sequenced due to incomplete reactions. Moreover, since genome sequencing was still many years in the future, a sequence obtained by Edman sequencing could very often not yet be placed in an appropriate context. The evolution toward automated Edman sequencing in the 1970s dramatically increased the throughput of the method and thereby gave the means to sequence multiple proteins in a single study (5). Yet even automated Edman sequencing proved insufficient and insensitive when applied to a complete proteome consisting of thousands of proteins, spanning multiple orders of magnitude in abundance (6).

Prior to MS-driven proteomics, studying a proteome required proteins to be separated prior to sequence analysis, and different gel-based methods have therefore been described over the years to separate proteins by their physicochemical properties. In isoelectric focusing (IEF), proteins are separated by their isoelectric point, which corresponds to the pH where proteins carry a net charge of zero, and thereby no longer feel the force that the electrical field applied to the gel exerts on charge-carrying molecules (7). In SDS-PAGE, proteins are denatured by sodium dodecyl sulfate (SDS) and subsequently separated by their apparent molecular weight in a gel with a controllable pore size (gradient) (8). Different separation methods can also be combined in a multidimensional setup. The most commonly used setup is 2D-PAGE, in which proteins are separated by IEF in the first dimension and by SDS-PAGE in the second dimension (9, 10).

After separation, the proteins can be visualized on the gel by different means including Coomassie Brilliant blue, silver, or immunostaining methods (11). The resulting spots indicate proteins which can then be analyzed by Edman sequencing after electroblotting (12). However, in addition to the limitations of Edman sequencing described above, the method also requires a large amount of protein material, typically in the order of micrograms. This impaired sensitivity of the method allows only abundant proteins to fall within the scope of the method.

When the much more sensitive technique of mass spectrometry became applicable to proteins through the advent of new ionization methodologies, far less material was needed

for protein detection and analysis. Instead of reading amino acids by chromatography, however, mass spectrometry completely relies on accurate mass measurements of charged analytes like (poly)peptides and their fragments, which necessitates the use of an existing sequence database against which the recorded masses can be matched. After gel-based separation, a protein spot is subjected to in-gel digestion, the resulting peptides are typically extracted (e.g., in acetonitrile water), and their masses are finally determined by mass spectrometry. The list of peptide masses, resulting from a single gel spot, is then utilized as a fingerprint to identify the parent protein using a protein sequence database and a specialized software program called a search engine. This technique is known as peptide mass fingerprinting (PMF) and was the most commonly used proteome analysis method during the 1990s (13). The major drawbacks of this method are that the primary amino acid sequence cannot be directly determined by PMF, and similar to Edman sequencing, proteins must be purified prior to digestion.

A new standard of proteome analysis was introduced by peptide-centric proteomics, where the focus is transferred from the separation of proteins to the separation of peptides. Peptides can be separated by chromatography, which can be in-line coupled to tandem mass spectrometry (MS/MS). The resulting MS/MS spectra allow determination of peptide sequences, which can then be used to infer the parent proteins. As a consequence of this paradigm shift in favor of analyzing peptides, the probability for a protein to be identified increases since multiple peptides can be utilized to identify the parent protein. Moreover, peptides are less extreme in size and other physicochemical parameters than proteins; this in turn also dramatically increases the sensitivity of peptide-centric proteome analysis.

The technologies involved in peptide-centric proteomics are described in detail in the next section.

2. Technological Requirements

2.1. Peptide Separation

All peptide-centric proteomic methods need to consider the complexity of the peptide mixture. Let us, for instance, assume that half of the 20,334 proteins annotated in the human subset of UniProtKB/Swiss-Prot[1] are present at any one time in a cell. Subjecting these to a tryptic digest generates on average about

[1] UniProtKB/Swiss-Prot release 57.9.

50 peptides per protein, yielding roughly 500,000 different peptide molecules. If we then (conservatively) allow for a doubling of this complexity due to alternative splicing and again for different modification states, then a complex peptide mixture will easily contain a few million distinct peptides. Even using the latest generation of mass spectrometry, capable of analyzing up to 10 peptides per second (MS/MS mode), such a complex peptide mixture requires extensive separation prior to mass spectrometry analysis. This was achieved by applying both existing and new chromatographic methods.

In liquid chromatography (LC) methods (14), peptides dissolved in a mobile phase flow through a column containing a synthetic resin. Peptides interact with and bind to this stationary phase and can be gradually eluted by varying the mixture ratio (water/organic solvent) of the mobile phase to increasingly resemble the properties of the stationary phase. A number of stationary and mobile phases had already proven their value in separating peptides and were furthermore compatible with mass spectrometry and therefore ready to use in peptide-centric proteomics. One of the most common methods, reversed-phase high performance liquid chromatography (RP-HPLC), employs a highly hydrophobic stationary phase (typically, a column packed with C-18-coated beads) (15). Peptides bind to these beads through hydrophobic interactions, and by increasing the concentration of the organic solvent (typically acetonitrile) in the mobile phase, increasingly hydrophobic peptides are released and therefore eluted over time. Another widely used method is ion exchange chromatography (IEX), where the resin is coated with either negative (strong anion exchange, SAX) or positive (strong cation exchange, SCX) ionic groups, attracting molecules of opposite charge. An increasing concentration of counterions (with charges opposite to the resin) in the mobile phase or changes in the buffer solution's pH will then increasingly displace peptides of ever higher charge states from the resin, eluting them from the column.

2.2. Mass Spectrometry

Mass spectrometry was first conceptualized in 1906 by Thomson, who described mass separation as canal rays, and the importance of mass spectrometry has grown ever since, illustrated by the five Nobel Prizes that were awarded over the years to research performed in the field of mass spectrometry: Joseph John Johnson for his work on the conduction of electricity by gas in 1906, Francis William Aston for the discovery of isotopes in 1922, Ernest Orlando Lawrence for the development of the cyclotron in 1939, Wolfgang Paul for the development of the ion trap in 1989, and John Bennett Fenn and Koichi Tanaka for the development of soft ionization methods in 2002.

All mass spectrometers systematically employ an identical series of components. The first component is the ion source, which serves to charge the analytes that will be measured. These charged analytes then enter the second component, the mass analyzer, wherein their trajectories are directly affected by the force of an electrical or magnetic field leading to selection or separation of ions with different mass to charge (m/z) ratios. The third component, the detection device, accurately captures ions and reads out their specific m/z ratios. These three components are described in the following sections.

2.2.1. Ionization of Peptides

It lasted until the late 1980s before intact peptides could be ionized and measured by mass spectrometry. Earlier ionization methods (e.g., electron ionization, chemical ionization, fast atom bombardment) disrupted the molecular structures and thus failed measuring intact peptides and were further practically limited by low upper mass limits. The advent of two soft ionization methods enabled the ionization of intact peptides (plus other biomolecules), and their efforts were recognized by two Nobel Prize awardees in 2002.

In matrix-assisted laser desorption ionization (MALDI), simultaneously described by Tanaka (16) and Karas (17), the peptides are co-crystallized with an acidic organic matrix (e.g., alpha-cyano). Here, a solution of peptides with excess of such a matrix is spotted on a metallic plate and dried upon which crystallization occurs. These samples are then inserted into the ion source region of the mass spectrometer – which is held under very high vacuum – and pulsed laser light, typically a N_2 laser emitting light of 337 nm, is directed at the crystallized sample. Upon absorption of this light, the energy is converted into collision energy and heat, leading to desorption of matrix molecules and peptides. Somewhere along this process, peptides are ionized but the exact mechanism driving ionization is still debated (e.g., *see* (18)).

Electrospray ionization (ESI) was the second soft ionization method (19). Here, peptides eluting from a RP-HPLC column are sprayed through a tiny orifice needle upon which a voltage is applied. At the needle tip, the spray forms a cone, the "Taylor cone" (20), and the tip of this cone releases charged solvent droplets containing peptides. These charged droplets further move to the mass analyzer using an electrical field, and a stream of drying gas (e.g., heated nitrogen) is used to further evaporate solvent from the droplets until the Coulomb repulsion on alike charged ions overcomes the droplet's surface tension, known as the Rayleigh limit (21), and finally shatters the droplet into yet smaller droplets. This cycle of solvent evaporation and droplet fission continues until the charges are transferred onto the peptides which thereby become volatile.

The coupling with liquid chromatography methods is online in ESI and off-line in MALDI. As such, the peptides can be reanalyzed over time in MALDI, which can be useful for biomarker discovery. Then again, the continuous flow of analytes in ESI will measure the mass more accurately than repeated desorption of inhomogeneous peptide crystallization in MALDI eventually (22).

2.2.2. Mass Analysis of Ions

The ionized analytes formed in the ion source are transported to the mass analyzer, wherein their trajectories are controlled and analyzed in various ways, finally enabling accurate m/z measurements. Ion trajectories can be controlled by two general methods: either by applying a dynamic electrical field (e.g., quadrupole ion trap, linear ion trap, Orbitrap, quadrupole, time-of-flight) or by applying a magnetic field (e.g., Fourier transform ion cyclotron resonance).

The very popular quadrupole (Q) mass analyzer is an m/z filter by applying a radio frequency (RF) voltage between two pairs of rods (23). The stability of the ion's trajectory in the quadrupole is balanced by its m/z value and the applied RF field, such that by adjusting the parameters of the RF field, only ions with a specific m/z value are selected and reach the end of the quadrupole where they are detected or transferred to a second analyzer. Similar to the quadrupole is the quadrupole ion trap (QIT) as it also generates a 3D RF field, though here ions are first trapped and then sequentially ejected from the QIT (24). The linear ion trap (LIT) in turn is similar to the QIT but now ions are trapped and ejected in a 2D RF field, which results in higher ion injection efficiencies and ion storage capacities, thus increasing the overall sensitivity (25). The time-of-flight analyzer again uses an electrical field to accelerate ions in a vacuum tube. The kinetic energy acquired by the ions correlates with their mass, charge, and applied voltage, and measurement of the flight time finally allows calculation of their m/z value.

The Orbitrap mass analyzer was introduced about a decade ago and surpassed the accuracy of other mass analyzers by one to two orders of magnitude (26). An Orbitrap uses an inner and outer electrode, shaped to create an electrostatic field. When ions enter the Orbitrap, they are trapped in an orbit around the inner electrode and the frequency of their rotation is related to their m/z value. An image current of this rotating ion is then Fourier transformed into a frequency spectrum and in its turn converted into a mass spectrum. In the Fourier transform ion cyclotron resonance (FT-ICR) mass analyzer, the ions are subjected to an oscillating electrical field combined with the perpendicular force of a magnetic field, together bringing ions into rotation, and similar to the Orbitrap, a mass spectrum can be inferred by measuring the frequency of this rotation (27). Each of the mass analyzers

described above can be placed in tandem with another (compatible) mass analyzer (13), thus enabling tandem mass spectrometry. Now, the first mass analyzer measures intact ions (the precursor ions), and precursor ions with a given m/z value are selected and fragmented in the same (tandem-in-time) or a different (tandem-in-space) mass analyzer. The resulting fragments of the precursor ion are finally measured and lead to MS/MS spectra. Both the QIT and the LIT are able to perform tandem MS by trapping, measuring, selecting, and fragmenting ions within one analyzer. Their main drawbacks are their low mass accuracy and their inability of measuring small mass fragment ions such as immonium ions. Clearly, their main advantage is their ability to perform MS^n, during which fragment ions are isolated and further fragmented, and this was proven valuable for studying post-translationally modified peptides (28, 29). TOF–TOF setups offer both good mass accuracy (20–50 ppm) and ion sensitivity and are very often coupled to a MALDI ion source. The Q-TOF yields similar quality, but is preferentially coupled to an ESI source. Despite that the FT-ICR remains the most accurate mass analyzer to date; its requirement of a magnetic field makes the instrument far more cumbersome than the LTQ-Orbitrap. The LTQ-Orbitrap also provides highly accurate mass measurement of precursors and suffers from low fragment ion mass cutoff and fragment ion mass errors typical to ion traps. However, the recently introduced LTQ-Orbitrap Velos overcomes these problems by introducing a higher energy collisional dissociation cell (HCD) (30).

2.2.3. Detection

The third element required by a mass spectrometer is the detector and most often this is a microchannel plate detector (MCP) (31) consisting of an array of electron multipliers (the channels) (32). When a charged analyte exits the mass analyzer and collides into the metallic-coated surface (e.g., PbO) of an electron multiplier, an electron-emitting torrent is initiated through the channel. The resulting electron flow is measured and is proportional to the number of charged analytes that started the electron-emitting cascade.

2.3. Mass Spectrometry Methodologies

In peptide-centric proteomics, tandem MS is generally used to identify peptides (MS/MS analysis) or to specifically detect and quantify peptides by selected reaction monitoring (SRM). We will here describe the different methods that are mainly applied to generate fragment ions. The fragment ion nomenclature suggested by Roepstorff and Fohlman is used here (33) and illustrated in **Fig. 1.1**.

2.3.1. Peptide Fragmentation

Collision-induced dissociation (CID) is the most commonly used method for peptide fragmentation (34). In CID, precursor ions

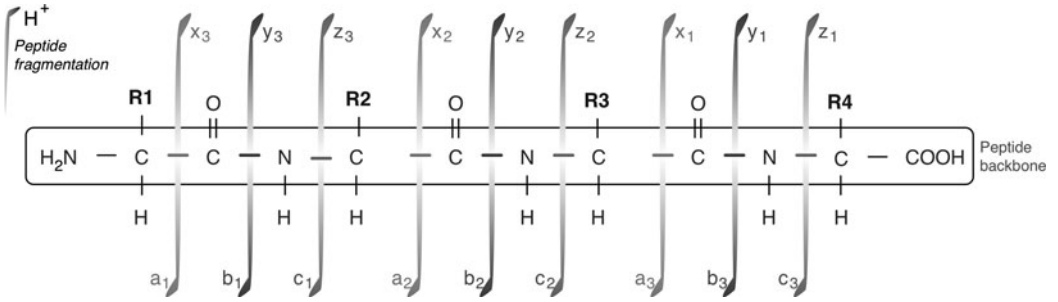


Fig. 1.1. This figure shows the different fragment ions generated upon peptide dissociation. Note that the figure follows the nomenclature suggested in (33).

collide with inert gas atoms (e.g., He and Ar) in a collision cell upon which mainly b- and y-fragment ions are created. CID also produces immonium ions specific for individual amino acids (35) and further readily dissociates labile peptide bonds (e.g., Xxx-Pro, Asp-Xxx) and unstable modified residues (e.g., *O*-phosphates and glycans) (36, 37).

Electron transfer dissociation (ETD) (38) and electron capture dissociation (ECD) (39) rely on an electron-based dissociation process and dominantly produce c- and z-ions along the peptide backbone in a sequence-independent manner, different from CID which prefers labile peptide bonds. ECD is limited to be used in the FT-ICR cells and not widely implemented due to the cost of this analyzer. ETD on the other hand is readily used in less expensive ion traps and thus more applied (e.g., for phosphoproteomics (40, 41)). Since ETD and ECD incorporate negatively charged electrons in the positively charged peptides, these peptides need to be highly charged (e.g., 3+, 4+) or else the signal of the fragment ions will be too weak resulting in less informative fragmentation spectra.

2.3.2. Selected Reaction Monitoring

Selected reaction monitoring was introduced three decades ago (42), but only recently got a fair amount of attention by the proteomics community (43, 44). Triple quadrupoles are best suited for SRM, in which the first quadrupole accurately filters a targeted precursor, the second quadrupole fragments this precursor ion, and the third quadrupole accurately filters for (a) specified fragment ion(s). Thus, a peptide ion is transferred from the first to the last quadrupole and a fragment ion is recorded, and such transitions are monitored through time. A few transitions per peptide (2–5) are often exceptionally specific and monitoring them surpasses other methods in terms of sensitivity. SRM is clearly predisposed for validation and quantification of previously identified peptides (45).

3. Proteomic Strategies

Two types of peptide-centric proteomic experiments are generally distinguished: qualitative proteomics aiming at comprehensively mapping the presence of all proteins in a sample and quantitative proteomics to quantify changes in protein abundance between samples. Representative methods were here selected and include those that had most impact on peptide-centric proteomics.

3.1. Qualitative Proteomics

3.1.1. Proteome Coverage

Inherent to proteomics is the absence of an amplification method (e.g., PCR in genomics) prior to identification of an analyte. Consequently, when a low-abundant or in fact any given protein is not identified, no conclusion can be drawn on whether it is absent from the sample or falls outside the detection limits of the mass spectrometer. Throughout the years, this has driven qualitative peptide-centric proteomics to produce ever-growing lists of identified peptides (from hundreds to thousands and recently ten thousands), continuously increasing proteome coverage.

Proteome coverage is mainly influenced by three factors (6). The first factor is the sensitivity of the mass spectrometer, which defines the lowest amount of analyte that can be detected. The latest generation instruments typically allow measurements in the order of femtomoles or even attomoles (30). The second factor is the dynamic range of the instrument, that is, the signal intensity range in which two distinct analytes can be detected, which typically spans two to three orders of magnitude. The third determining factor for proteome coverage is the duty cycle of the mass spectrometer, being the number of fragmentation spectra (with a fair amount of quality) the mass spectrometer can produce within a given time frame. This varies from 1 spectrum per second for slower instruments to 10 spectra per second for the last generation instruments (30). Combined with chromatographic resolution, this parameter also influences the number of peptides that will be identified. Now given the high complexity of proteome samples, it is clear that even with the fastest mass spectrometers not all peptides will be identified, a phenomenon known as random sampling (46).

To tackle this random sampling issue, both technological and methodological proteomic developments were made. Technological developments are mainly driven by vendors of mass spectrometers, who release better performing and more specialized instruments. Such steady developments enhance sensitivity, increase sequencing speed, and enlarge dynamic range. Methodological developments are mainly driven by academic researchers and include inventive strategies to reduce complexity of a peptide

mixture resulting from proteome digestion. In this respect, two main approaches can be distinguished: either the peptide mixture is extensively fractionated prior to LC-MS/MS analysis or only a targeted set of peptides related to the proteomic experiment is selected and analyzed. The former is metaphorically termed "shotgun proteomics," while the latter is often referred to as "targeted proteomics," but both diminish the problems associated with dynamic range and sequencing speed and thereby increase the probability for a peptide to be sequenced upon random sampling by mass spectrometers.

3.1.2. Shotgun Proteomics

MUltiDimensional Protein Identification Technology (MUDPIT) was the first method that employed extensive fractionation of a complex peptide mixture (47, 48). In their 2001 publication, the group of John Yates first separated a yeast tryptic proteome digest by SCX in 15 fractions. These were then individually analyzed by LC-MS/MS. By increasing the number of SCX fractions to 80, a further increase in proteome coverage was achieved (49).

Another popular method for shotgun proteomics is GeLC-MS/MS (6). Here, intact proteins are separated by SDS-PAGE, the gel is then cut into multiple slices, and proteins are in-gel digested and the resulting peptides are analyzed by LC-MS/MS. Compared to MUDPIT, GeLC-MS/MS requires an order of magnitude less protein material. Furthermore, abundant proteins will concentrate in distinct gel slices, such that peptides produced from these proteins are not smeared over LC-MS/MS runs thereby increasing the chance of identifying less abundant proteins. This combined protein–peptide separation was also exploited to detect protein processing events (50). Yet another method is peptide IEF-LC-MS/MS, in which a peptide mixture is fractionated by IEF prior to LC-MS/MS (51).

3.1.2.1. Drawbacks of Shotgun Proteomics

The strength of shotgun proteomics is that by random sampling a peptide mixture, an overview of the proteome composition is readily generated in which many proteins are identified by multiple peptides, which increases the reliability of such identifications. For instance, in the initial MUDPIT study (47), 5,530 distinct peptides were identified in 1,484 yeast proteins, yielding an average of 3.72 peptides per protein. A similar sample was studied with newer instruments using the GeLC-MS/MS method (6) and provided a fourfold of identified peptides (20,893) in 2,003 yeast proteins, thus yielding an average of 10.4 peptides per protein. While this increases confidence in protein identifications, the moderate increase in protein identifications illustrates that a high number of proteins remained subject to undersampling. A follow-up study was published in 2008 in which Orbitrap mass spectrometry resulted in 4,399 protein identifications (51), a number similar to previously published tandem

affinity purification and green fluorescent protein tagged benchmark proteomes (52, 53). However, a considerably more complex proteome such as that from human cells remains challenging for receiving full coverage with shotgun proteomics.

3.1.3. Targeted Proteomics

Targeted proteomics uses strategies to extract a selected set of peptides from a whole proteome digest and only analyzes these by LC-MS/MS. This selection is such that it is representative for the analyzed proteome or pertinent to the goal of the proteome study. Moreover, since a selection of peptides always yields a less dense peptide mixture, random sampling tends to be reduced.

3.1.3.1. Sample Complexity Reduction

On average, tryptic digestion of a protein generates 50 peptides (*see* **Section 2.1**), a large number given that typically only a few are required for protein inference (54). This in turn offers an opportunity to lower sample complexity by lowering the amount of peptides per protein.

Selection of cysteinyl peptides by isotope-coded affinity tag (ICAT) labeling (55) probably is the hallmark of targeted proteomics. The ICAT molecule binds covalently to the free thiol group of cysteine and carries a biotin group which enables enrichment of ICAT-modified cysteinyl peptides by avidin affinity chromatography. Moreover, isotopic variants of the ICAT molecule are introduced in the linker region, and these enable quantitative proteome studies (*see* **Section 3.2**).

Another effort to reduce sample complexity was introduced by the versatile COmbined FRActional DIagonal Chromatography (COFRADIC) methodology, capable of, among others, selecting cysteinyl and methionyl peptides (56, 57). COFRADIC's core is the separation of a complex peptide mixture by two identical and consecutive RP-HPLC separations. Between both separations, a modification reaction is performed that alters the physicochemical properties of a targeted group of peptides. As a result, altered peptides obtain a different elution profile during the second RP-HPLC separation by which they are distinguished from non-altered peptides. Clearly, by changing the actual modification reaction, different sets of peptides can be targeted and thus isolated. **Table 1.1** lists the various sorting protocols employed so far with the COFRADIC methodology.

The highest reduction in sample complexity comes from selecting only a single, though representative peptide per protein. This can be accomplished by selecting either its N- or C-terminal peptide and was in fact the motivation for developing N-terminal COFRADIC (58). The actual modification reaction uses 2,4,6-trinitrobenzene sulfonic acid that renders the non-N-terminal peptides more hydrophobic such that N-terminal peptides are readily isolated (*see* also (59)).

Table 1.1
Overview of the different COFRADIC technologies

Peptides targeted by COFRADIC	Sorting reaction	References
Methionyl peptides	H_2O_2	(56)
N-terminal peptides	TNBS	(58)
Cysteinyl peptides	TCEP reduction	(128)
Phosphorylated peptides	Phosphatases	(129)
N-glycosylated peptides	PNGaseF	(130)
ATP binding peptides	Alkaline hydrolysis	(131)
Sialylated peptides	Neuraminidase	(132)
3-Nitrotyrosine peptides	Dithionite reduction	(133)

3.1.3.2. Selection of Post-translationally Modified Peptides

Post-translational modifications (PTMs) are considered as a "cellular language" and therefore obtain plenty of interest as they are key to understanding cellular phenotypes (60). PTMs effect either amino acid side chains or peptide bonds (protein processing). Both are (in)directly detectable by MS; yet, since modified peptides are hard to distinguish in a complex mixture, various methods were developed to select or enrich particular modified peptides.

Proteolytic Cleavage

Several targeted proteomic methodologies are readily applied for studying protein processing events by selecting (neo-)N-terminal peptides (58, 61–64). The identified N-terminal peptides align either at protein N-termini or at internal regions, and if the latter also match criteria imposed by the applied selection strategy (e.g., modification state), the location of "neo-N-terminal peptides" points to the actual protein processing events.

Phosphorylation

Protein phosphorylation sites have been mapped in detail by targeting phosphorylated peptides in a complex peptide mixture. Various methods have been described to achieve this including immobilized metal ion chromatography (IMAC) (65, 66), titanium dioxide chromatography (TiO_2) (67), hydrophilic interaction chromatography (HILIC) (68), and SCX at low pH (69), all exploiting the unique physicochemical properties of the polar phosphate group to separate phosphorylated peptides from non-phosophorylated peptides prior to mass spectrometry analysis.

Glycosylation

Lectin affinity chromatography has been extensively employed to enrich, depending on the lectin(s) used, N-glycosylated or O-glycosylated peptides or proteins (70). A different approach is to chemically trap N-glycosylated peptides by hydrazide

chemistry, followed by PNGaseF-driven release (71). Yet another approach enriches O-GlcNAc modified peptides via a chemoenzymatic strategy (72).

Clearly, more protein modifications are studied by targeted proteomics, including lysine acetylation (73) and ubiquitination (74), but the plethora of possible protein modifications remain a formidable challenge in current proteomics. In that respect, MS-driven targeted proteomics will prove indispensable for generating maps of PTM prevalence. Yet, an additional step is required before PTMs can even be considered in the perspective of systems biology and this is measurement of their abundance in a variety of conditions (75). This is enabled by quantitative proteomics, which is the topic of the next section.

3.2. Quantitative Proteomics

While qualitative proteomics aims to generate a compositional map of proteins, quantitative proteomics extends this map with relative or absolute abundance information. Quantitative proteomics is performed on samples differing in cellular phenotype (e.g., benign versus malignant cancer), subjected to different stimuli (e.g., control versus growth factor), and followed over time (e.g., cell cycle checkpoints) or on many other cellular states in which a differential proteome composition might be expected by hypothesis-driven research (60).

Various methods have been applied to carry out quantitative proteomics and initially two groups can be distinguished. The first group of methods introduces mass tags that allow differentiation between peptides from distinct samples during MS analysis. The second group of label-free methods integrates aligned intensity profiles from LC-MS or MS/MS analyses to find differences between samples.

Quantification methods that use mass tags can further be distinguished based on the step in the protocol during which they introduce isotopic labels. Each protocol step introduces its own variation, thus the sooner the labels are introduced, the sooner the samples can be mixed together and less variation is introduced (*see* **Fig. 1.2**).

3.2.1. Metabolic Labeling

Metabolic labeling of living cells or organisms was first achieved by culturing cells in carbon-13- and/or nitrogen-15-containing nutrients, eventually rendering most carbon and/or nitrogen-containing molecules in a heavy form (76), but, for proteome analysis, metabolic labeling proved more efficient by using essential amino acids.

3.2.1.1. SILAC (Cell Culture to Organism)

Stable isotope labeling of amino acids in cell culture (SILAC) was introduced in 2002 by the group of Matthias Mann and has been widely adopted ever since (77, 78). In SILAC an isotopic label is introduced metabolically by growing cell cultures

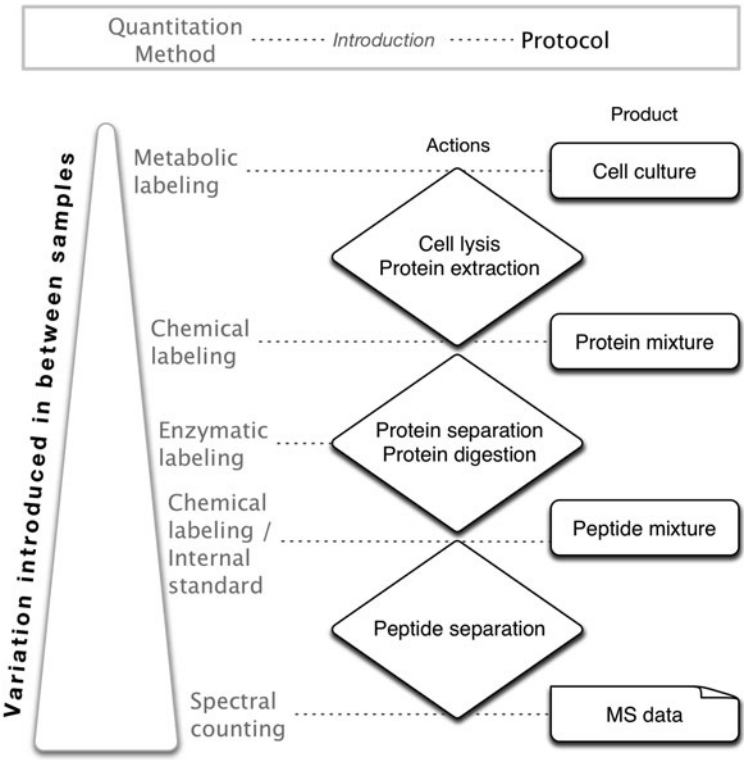


Fig. 1.2. General overview of the different quantification methodologies and indication of their introduction in a proteomic protocol. The gradient shape in the *left* illustrates how variation in between samples increases as the introduction of isotope labels occurs late in the overall protocol.

in natural medium or in SILAC medium in which one or more essential amino acids (e.g., Arg, Lys, and Met for mammalian cells) are only present in their "heavy form" (e.g., deuterated, carbon-13, and nitrogen-15). Following a number of cell population doublings, all proteins derived from SILAC-grown cell cultures are isotopically labeled. Normally, a control (light) peptide and one (or more) SILAC-labeled (heavy) peptide(s) elute identically in a chromatographic setup but segregate in MS spectra by the SILAC-introduced mass difference(s). The MS intensity profile of the differently labeled peptides is finally used for quantification.

The strength of SILAC-driven quantification lies in its early introduction in the experimental protocol as this is expected to minimize variation in ratio measurements, thus finally yielding more reproducible and accurate results. The main drawback is that an adequate protein turnover is required and that the method is limited to systems that use essential amino acids. Consequently, plants, bacteria, body fluids (e.g., urine, blood plasma), or other patient samples cannot be studied by SILAC. For some model

organisms, these limitations were overcome by breeding them on a metabolic labeled diet until the organism was completely labeled for subsequent quantitative experiments (79, 80).

3.2.2. Non-metabolic Labeling

The ICAT methodology enables both sample complexity reduction by selecting cysteinyl peptides (*see* **Section 3.1.3**) and protein quantification by introducing a mass tag via the ICAT label (81). Therefore, a control sample is post-metabolically labeled with one type of ICAT label, while a second sample is labeled with another type of ICAT label. After avidin purification of ICAT-tagged cysteinyl peptides, quantification of heavy and light peptides is done in MS spectra.

3.2.2.1. ICAT

3.2.2.2. Enzymatic Peptide Labeling with Oxygen-18

Oxygen-18 labeling is an enzymatic labeling strategy for relative quantification of peptides and proteins typically during or following tryptic digestion of proteomes (82). Tryptic digestion is conducted either in light ($H_2\ ^{16}O$) or in heavy ($H_2\ ^{18}O$) water and, when performed adequately, the resulting peptides are mass tagged by a 4 Da difference.

3.2.2.3. ITRAQ^{TM}

Isobaric tags for relative and absolute quantification (ITRAQ^{TM}) bind covalently to primary amines (the alpha-amine at a protein's N-terminus and the epsilon-amine at lysine side chains) (83). ITRAQ molecules are bivalent and built from a small reporter group and a balancer group together forming an isobaric combination (same nominal mass) and joined through a linker region that readily fragments in MS/MS mode. The isobaric nature of the tags implies that the mass of peptides tagged with different ITRAQ molecules (thus different samples) is identical during MS survey, yet the ITRAQ label renders reporter ions specific to each sample in MS/MS mode. As such, samples are compared using the intensities of the ITRAQ reporter ions in the peptide fragmentation spectra. Clearly, the main advantage of ITRAQ is that its multiplexed nature enables to compare up to eight samples in a single MS/MS spectrum, yet statistical variation in ratio measurement is expected to be enlarged due to late introduction in the overall protocol.

3.2.2.4. Internal Standard Peptides

By adding a known amount of internal standard peptides into a peptide mixture, the further determined heavy-to-light peptide intensity ratio can be utilized to estimate the absolute amount of peptides present in the sample (45, 84). Therefore, heavy-labeled peptides (e.g., 13-carbon or 15-nitrogen) are typically used to be monitored by SRM (**Section 2.3.2**). One drawback is that only a limited number of internal standard peptides can be monitored simultaneously, yet Malmström et al. recently showed how only a few absolute protein quantifications are sufficient to extrapolate proteome-wide absolute quantification numbers with a moderate error rate (85).

3.2.3. Label-Free Methods

Quantitative proteomics can also be achieved without employing stable isotopes and such methods are accordingly referred to as label-free quantitative proteomics.

3.2.3.1. Spectral Counting

Spectral counting starts with the assumption that more abundant peptides are more likely to get selected for fragmentation than less abundant peptides (86). Thus, since a peptide can be expected to ionize equally well in different samples, the number of fragmentation events should be fairly alike too and can therefore be used as a relative quantification measure between multiple samples. The simplicity of the method is its main strength, and although it is by far the least accurate method to employ quantitative proteomics, it can still be useful for monitoring large quantitative proteomic perturbations of a system.

3.2.3.2. Intensity-Based Quantification

Intensity-based quantification considers a peptide as a feature with two coordinates, its retention time and its m/z value, and records an MS-derived intensity value for each feature (87). Separate LC-MS analyses from distinct samples can then be aligned and normalized by these coordinates, while the intensity values provide quantitative information for the aligned features across distinct samples. When an interesting (deviating) feature is observed, a MS/MS sequencing attempt is made to identify the origin of the feature, which is not necessarily successful. Since the samples are subjected to mass spectrometry more than once, offline LC-MALDI-MS(/MS) is the platform of choice for this type of quantification. The strength of this method is that it readily scales with the number of samples, thereby being the only quantitative method providing appropriate statistical analysis, which makes label-free intensity-based quantification a good candidate for biomarker discovery.

4. Working with Proteomic Data

From a reductionist viewpoint, a proteomic study has a typical structure of subsequent steps: breaking the cells, extracting the proteins, digesting the proteins into peptides, separating the peptides, and finally analyzing the peptides by mass spectrometry. The output of the mass spectrometer marks a drastic change, as it is purely numerical. For the remainder of the protocol, these data are then employed to reconstruct information about the studied sample, which is the domain of (bio-)informatics.

4.1. Data Processing

The raw MS and MS/MS mass spectrometry data are typically first processed by instrument-dependent or -independent signal processing software (e.g., Excalibur from Thermo Fisher,

MassLynx from Waters, Mascot Distiller from Matrix Science, OpenMS (88), MaxQuant (89)). The processing actions typically involve peak picking, smoothing, noise removal, mass calibration by internal standard, isotope correction, charge deconvolution, etc. (90). Some of these processing actions are optional and others are user customizable, yet since the impact of each of these processing actions on the raw data persists to the level of peptide and protein identification and quantitation, manipulation of these actions should be well understood by the user. Furthermore, it is critical that the applied processing steps are preserved in the data format. There is also an orthogonal aspect, requiring the raw (or processed) data from different instruments to be rendered in a common, readily consumable, and open standardized data format [RAW]. The mzML standard has been developed and released by the HUPO Proteomics Standards Initiative (PSI)[2] for these purposes (maintaining data processing history and providing a common, open representation of mass spectrometry data) and has been widely implemented in mass spectrometry software.

4.2. Peptide Identification

4.2.1. Sequence Assignment

In order to determine the identity of a peptide based on its MS/MS spectrum, the fragmentation data can be interpreted by various software methods, all exploiting the (partial) sequence information inherent in a fragmentation spectrum. Because the fragmentation process is at least partially sequence dependent, different fragment ions do not have equal chances of occurring when a given ion is fragmented. As a consequence, a fragmentation spectrum consists of a heterogeneous combination of high- and low-intensity ion signals, along with noise peaks. Certain fragment ions will furthermore be altogether absent from the spectrum. Still, such a complicated and incomplete fragmentation spectrum can capture enough information about the peptide sequence that an attempt can be made to identify the peptide.

The first method to identify peptides from MS/MS spectra, described by Mann and Wilm, was based on so-called peptide sequence tags in which a peptide sequence tag is formed by several consecutive ion signals in the fragmentation spectrum (91). The approach extracts a small stretch of sequence information directly from the spectrum and appends the remaining terminal mass at either end. In the next step, this sequence tag is searched against the peptides resulting from an in silico digested protein sequence database (92). Any peptide that contains the small sequence along with the correct flanking masses is then a candidate for identification.

[2] The HUPO Proteomics Standards Initiative (PSI) defines community standards for data representation in proteomics to facilitate data comparison, exchange, and verification. http://www.psidev.info/

A second, and currently most popular, method is implemented in database search algorithms, which also make use of sequence databases (93). These algorithms first constrain all possible peptides by checking their theoretical mass against a mass interval around the experimental precursor mass. The matching peptides are then in silico fragmented, and these computationally derived fragmentation spectra are then matched to the experimental fragmentation spectrum. The best ranked match is finally hypothesized as the peptide sequence that led to the experimental fragmentation spectrum. This best match is not necessarily correct, however, as the best match can still be a very poor match indeed. Many search algorithms therefore calculate some kind of score (typically probability based) that allows the distinction between reliable and spurious peptide hits. The first published search algorithm was SEQUEST (94), but several other commercial as well as free algorithms exist today, including Mascot (95), X!Tandem (96), Phenyx (97), and OMSSA (98).

A third method to assign a peptide sequence to a fragmentation spectrum is through de novo sequencing, in which no a priori information from a sequence database is used. In de novo sequencing, a peptide sequence is computationally derived purely from the information captured within a fragmentation spectrum (99, 100). The methodology performs rather well when fragmentation is fairly complete, such that a signal is found for most theoretically expected fragment ions in a series. But larger gaps have a detrimental impact on the performance of de novo sequencing algorithms, introducing substantial ambiguity in the possible sequence. De novo sequencing is therefore usually restricted to specialized applications, for instance when no protein sequence database is available for the organism under study.

4.2.2. Error Estimation

To provide adequate quality within a proteomic experiment, the peptide identifications are further evaluated mainly by estimating the rate of false-positive peptide identifications in results. This is mostly done by employing a target-decoy database searching strategy (101). This approach performs the spectrum to peptide matching process in duplicate. The first search employs the normal (or "target") sequence database that contains sequences relevant to the sample, while the second search relies on a decoy sequence database which contains only nonsense sequences that should not be present in the sample. This type of database is commonly created by reversing or shuffling the protein sequences from the target database. Each peptide identified in this decoy sequence database search can thus be considered a false-positive (or rather random) peptide identification. Assuming that this error rate in the decoy database is reflected in the target database, a false-positive rate can be estimated for the experimental results obtained from the target sequence database. Note that this

approach does not point out the actual false-positive peptide identifications themselves; it merely estimates the overall level of false-positive peptides identified in the obtained results.

The use and value of employing decoy sequence databases to verify whether the required quality is achieved within a results set has been a topic of debate ever since the onset of peptide-centric proteomics (*see* (93) for an excellent review).

Another database search algorithm parameter that can be evaluated is the false-negative rate within the obtained results. An example of high false-negative rate can be obtained by attempting to identify a set of fragmentation spectra generated from human liver tissue in a yeast protein sequence database. While some fragmentation spectra will be identified to peptides from highly conserved proteins, most fragmentation spectra remain unidentified simply because the appropriate sequence information is lacking. The amount of highly informative (or high-quality) spectra that remain unassigned in a given data set can be assessed by spectrum quality methods, which rate a fragmentation spectrum by its information content (e.g., number of ions, ion intensities, signal-to-noise ratio) (102, 103). When a large amount of highly informative fragmentation spectra remain unidentified, then the means for interpreting those fragmentation spectra should be scrutinized, as the results seemingly comprise a large amount of false negatives.

4.2.3. Quality Validation

After fragmentation spectra have been identified by a database search algorithm, the identifications can be subjected to validation methodologies. These approaches tend to employ information complementary to the database search algorithm to remove a maximum of false-positive peptide identifications while incurring a minimal loss in sensitivity (i.e., removing as few true positives as possible).

Historically, a peptide-centric proteomic analysis generated only a few hundreds of peptide identifications per analysis, thus manually validating these as a feasible endeavor for a mass spectrometry expert. However, as the number of peptide identifications grew to tens of thousands during the last decade, manual validation became unworkable and automated validation methods were introduced. These typically implement statistical methods (e.g., linear discriminant analysis (104), support vector machines (105), and hidden Markov models (106)) utilizing various parameters related to the quality of the peptide identification (e.g., precursor mass error, ion coverage, and max sequence tag length) to further improve the separation between correct and incorrect peptide identifications. Although these standard parameters perform well for standard analyses, they often struggle to cope with novel types of information relevant to new protocols. A possible solution is to build tools for the semi-automatic validation of

peptide identifications where the mass spectrometry expert interactively defines a set of applicable rules for a particular experiment, reflecting the expert's knowledge about the experiment (107). Once established, these rule sets can be used to automate the peptide identification evaluation, separating suspicious from trustworthy peptide identifications.

4.3. Protein Inference

After the peptides have been successfully identified from the acquired MS/MS spectra, they are mapped to their parent proteins. This mapping is again performed using the information in a protein sequence database and ideally results in a one-to-one mapping, in which a particular peptide sequence is uniquely mapped onto a single protein. For the specific case of the human complement of the UniProtKB/Swiss-Prot database, **Fig. 1.3** shows that more than 90% of the tryptic peptides can be mapped one-to-one. Other databases tend to fare much worse, with only 25%–45% of the tryptic peptides uniquely mapped, depending on the exact database used. The overall challenge of inferring proteins from a list of peptides is known as the protein inference problem, and the relevant issues are clearly and comprehensively described in (54). Even though no method to this problem can be absolutely conclusive (108), several approaches have been suggested to tackle this problem. The most optimistic approach includes all possible protein mappings, such that if peptide "A" maps to both protein "1" and protein "2," then both are included

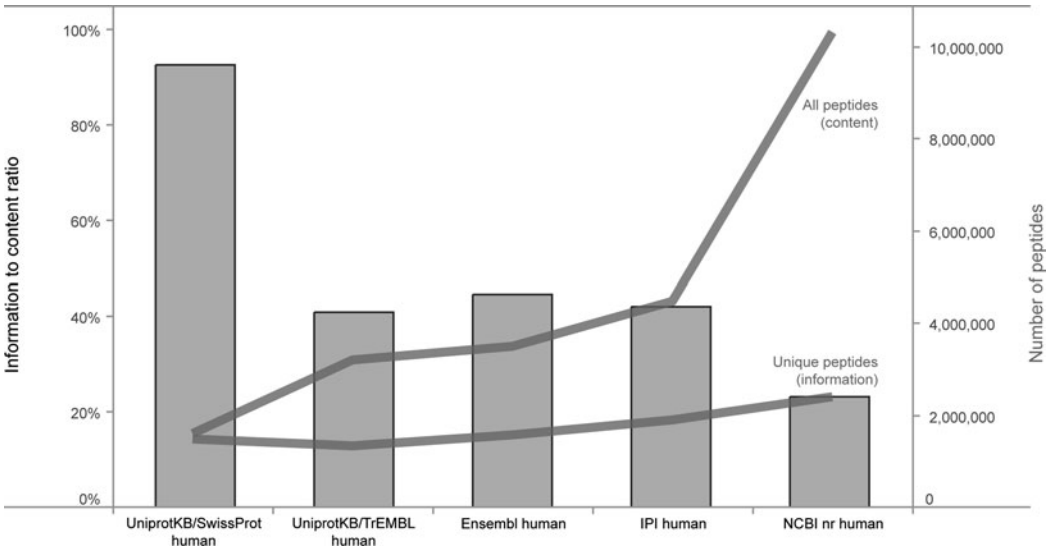


Fig. 1.3. Information versus content in popular protein sequence databases. This figure shows *lines* for the total number of peptides (content) and the number of unique peptides (information) generated by an in silico trypsin digestion of popular protein sequence databases (*right axis*). The information to content ratio is further shown by a bar chart (*left axis*), where the bar height correlates with the uniqueness of a peptide in each distinct sequence database.

in the protein results list. A more pessimistic approach assumes that only one of both proteins truly occurs within the sample and attempts to scrutinize the best option by utilizing additional information. The Protein Prophet software, for instance, first generates all possible peptide to protein mappings and then employs an iterative algorithm that gravitates toward those proteins that are identified by the largest number of (unique) peptides (109). A third approach relies on using the amount of annotation provided for a protein, postulating that highly annotated proteins are most likely to be detected than hypothetical protein entries for which no evidence has yet been encountered (108). Interestingly, quantitative information of peptides could potentially also be used to attribute peptides to proteins, as suggested by (54).

4.4. Functional Analyses

The resulting list of peptides and proteins is most often not yet meaningful in terms of the underlying biology (110). In order to facilitate the step from the experimental results to biological insight, several free software tools for functional analysis have been developed over the past few years.

The Cytoscape tool visualizes proteins and their interactions in dynamic networks that can be highly customized with biological annotations (111). By using publicly known protein–protein interaction data for example, regulation of protein complexes or signaling cascades is easily analyzed. Moreover, the BINGO plugin to Cytoscape enables gene ontology-driven analyses from parts of the network (112). Other tools such as DAVID (113) or PANDORA (114) also attempt to classify protein results lists into functional groups.

But beyond the annotation and contextualizing of the identified proteins, other analyses can be useful as well. A nice example is provided by tools that analyze sequence conservation in aligned nucleic or amino acid sequences, such as the sequence logos that were first described two decades ago, and have recently been refined for protein sequence analysis (115, 116).

4.5. Data Management

The data workflow described in the previous sections, from generating fragmentation spectra to identifying peptides and inferring proteins, has to be adequately managed at several levels. The first level of data management is typically established by a laboratory information management system (LIMS) implementation (117–121). These systems track and register data actions (e.g., storing newly generated fragmentation spectra, identifying fragmentation spectra into peptides) in order to create a queryable, historical log for the experimental results. Such a log enables monitoring of who performed which action at what time, essential for data provenance. Moreover, a convenient access point is thereby created for the retrieval of data and results. Furthermore, adequate data management greatly facilitates the implementation

and adoption of standardized data processing workflows, which in turn result in a net increase in productivity. Finally, organizing data from different experiments in an identical manner enables inspiring meta-analyses between distinct experiments.

The second level of data management is maintained by public data repositories such as PRIDE (122), NCBI Peptidome (123), PeptideAtlas (124), or GPMDB (125). Proteomic journal guidelines typically recommend or require storage of results in public data repositories to enable community-driven quality control for both peer reviewers and motivated readers. Furthermore, this also enables results aggregation of independent laboratories, which proves useful for genome annotating endeavors, meta-analyses discovering general proteomic result biases (126), and overall methodological evaluations (127).

Acknowledgment

K.H. is supported by a Ph.D. grant from the Institute for the Promotion of Innovation through Science and Technology in Flanders (IWT-Vlaanderen).

References

1. Williams, S.K. & Tyler, J.K. (2007) Transcriptional regulation by chromatin disassembly and reassembly. *Curr Opin Genet Dev* **17**, 88–93.
2. Van Der Kelen, K., Beyaert, R., Inzé, D. & De Veylder, L. (2009) Translational control of eukaryotic gene expression. *Crit Rev Biochem Mol Biol* **44**, 143–168.
3. Gross, J.D. et al. (2003) Ribosome loading onto the mRNA cap is driven by conformational coupling between eIF4G and eIF4E. *Cell* **115**, 739–750.
4. Edman, P. (1950) Method for determination of the amino acid sequence in peptides. *Acta Chem Scand* **4**, 283–293.
5. Niall, H. (1973) Automated Edman degradation: the protein sequenator. *Methods Enzymol* **27**, 942–1010.
6. de Godoy, L.M.F. et al. (2006) Status of complete proteome analysis by mass spectrometry: SILAC labeled yeast as a model system. *Genome Biol* **7**, R50.
7. Svensson, H. (1961) Isoelectric fractionation, analysis, and characterization of ampholytes in natural pH gradients. *Acta Chem Scand* **15**, 325.
8. Laemmli, U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680–685.
9. O'Farrell, P.H. (1975) High resolution two-dimensional electrophoresis of proteins. *J Biol Chem* **250**, 4007–4021.
10. Klose, J. (1975) Protein mapping by combined isoelectric focusing and electrophoresis of mouse tissues. *Hum Genet* **26**, 231–243.
11. Lauber, W.M. et al. (2001) Mass spectrometry compatibility of two-dimensional gel protein stains. *Electrophoresis* **22**, 906–918.
12. Vandekerckhove, J., Bauw, G., Puype, M., Van Damme, J. & Van Montagu, M. (1985) Protein-blotting on Polybrene-coated glass-fiber sheets. A basis for acid hydrolysis and gas-phase sequencing of picomole quantities of protein previously separated on sodium dodecyl sulfate/polyacrylamide gel. *Eur J Biochem* **152**, 9–19.
13. Pappin, D.J., Hojrup, P. & Bleasby, A.J. (1993) Rapid identification of proteins by

peptide-mass fingerprinting. *Curr Biol* **3**, 327–332.
14. Felinger, A. (2008) Molecular dynamic theories in chromatography. *J Chromatogr A* **1184**, 20–41.
15. Imoto, T. & Yamada, H. (1983) Peptide separation by reversed-phase high-performance liquid chromatography. *Mol Cell Biochem* **51**, 111–121.
16. Tanaka, K., Waki, H., Ido, Y., Akita, S. & Yoshida, Y. (1988) Protein and polymer analyses up to m/z 100,000 by laser ionization time-of-flight mass spectrometry. *Rapid Commun Mass Spectrom* **2**, 151–153.
17. Karas, M. & Hillenkamp, F. (1988) Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Anal Chem* **60**, 2299–2301.
18. Zenobi, R. & Knochenmuss, R. (1998) Ion formation in MALDI mass spectrometry. *Mass Spectrom Rev* **17**, 337–366.
19. Fenn, J.B., Mann, M., Meng, C.K., Wong, S.F. & Whitehouse, C.M. (1989) Electrospray ionization for mass spectrometry of large biomolecules. *Science* **246**, 64–71.
20. Taylor, G. (1964) Disintegration of water drops in an electric field. *Proc R Soc Lond A* **280**, 383–397.
21. Rayleigh, L. (1882) Further observations upon liquid jets. *Proc R Soc Lond* **34**, 130–145.
22. Rietschel, B. et al. (2009) The benefit of combining nLC-MALDI-Orbitrap MS data with nLC-MALDI-TOF/TOF data for proteomic analyses employing elastase. *J Proteome Res* **8**, 5317–5324.
23. Paul, W. & Steinwedel, H. (1953) Ein neues Massenspektrometer ohne Magnetfeld. *Zeitschrift Naturforschung Teil A* **8**, 448–448.
24. March, R. (2009) Quadrupole ion traps. *Mass Spectrom Rev* **28**, 961–989.
25. Makarov, A. et al. (2006) Performance evaluation of a hybrid linear ion trap/orbitrap mass spectrometer. *Anal Chem* **78**, 2113–2120.
26. Makarov, A. (2000) Electrostatic axially harmonic orbital trapping: a high-performance technique of mass analysis. *Anal Chem* **72**, 1156–1162.
27. Mclafferty, F.W. (1994) High-resolution tandem FT mass spectrometry above 10 kDa. *Acc Chem Res* **27**, 379–386.
28. Kameyama, A. (2006) Glycomics using mass spectrometry. *Trends Glycosci Glycotechnol* **18**, 323–341.
29. Villén, J., Beausoleil, S.A. & Gygi, S.P. (2008) Evaluation of the utility of neutral-loss-dependent MS3 strategies in large-scale phosphorylation analysis. *Proteomics* **8**, 4444.
30. Olsen, J. et al. (2009) A dual pressure linear ion trap – Orbitrap instrument with very high sequencing speed. *Mol Cell Proteomics* **8**, 2759–2769.
31. Wiza, J. (1979) Microchannel plate detectors. *Nucl Instrum Methods* **162**, 587–601.
32. Farnsworth, P. (1934) Electron multiplier. *US Patent* 1,969,399.
33. Roepstorff, P. & Fohlman, J. (1984) Letter to the editors. *Biol Mass Spectrom* **11**, 601.
34. Wells, J. & McLuckey, S. (2005) Collision-induced dissociation (CID) of peptides and proteins. *Methods Enzymol* **402**, 148–185.
35. Falick, A., Hines, W., Medzihradszky, K., Baldwin, M. & Gibson, B. (1993) Low-mass ions produced from peptides by high-energy collision-induced dissociation in tandem mass spectrometry. *J Am Soc Mass Spectrom* **4**, 882–893.
36. Huang, Y. et al. (2008) A data-mining scheme for identifying peptide structural motifs responsible for different MS/MS fragmentation intensity patterns. *J Proteome Res* **7**, 70–79.
37. DeGnore, J. & Qin, J. (1998) Fragmentation of phosphopeptides in an ion trap mass spectrometer. *J Am Soc Mass Spectrom* **9**, 1175–1188.
38. Syka, J., Coon, J. & Schroeder, M. (2004) Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc Natl Acad Sci* **101**, 9528–9533.
39. Zubarev, R., Kelleher, N. & McLafferty, F. (1998) Electron capture dissociation of multiply charged protein cations. A nonergodic process. *J Am Chem Soc* **120**, 3265–3266.
40. Molina, H., Horn, D.M., Tang, N., Mathivanan, S. & Pandey, A. (2007) Global proteomic profiling of phosphopeptides using electron transfer dissociation tandem mass spectrometry. *Proc Natl Acad Sci USA* **104**, 2199–2204.
41. Chi, A. et al. (2007) Analysis of phosphorylation sites on proteins from Saccharomyces cerevisiae by electron transfer dissociation (ETD) mass spectrometry. *Proc Natl Acad Sci USA* **104**, 2193–2198.
42. Kondrat, R.W., Mcclusky, G.A. & Cooks, R.G. (1978) Multiple reaction monitoring in mass spectrometry/mass spectrometry for direct analysis of complex mixtures. *Anal Chem* **50**, 2017–2021.
43. Stahl-Zeng, J. et al. (2007) High sensitivity detection of plasma proteins by multiple reaction monitoring of N-glycosites. *Mol Cell Proteomics* **6**, 1809–1817.

44. Picotti, P., Bodenmiller, B., Mueller, L.N., Domon, B. & Aebersold, R. (2009) Full dynamic range proteome analysis of S. cerevisiae by targeted proteomics. *Cell* **138**, 795–806.
45. Lange, V., Picotti, P., Domon, B. & Aebersold, R. (2008) Selected reaction monitoring for quantitative proteomics: a tutorial. *Mol Syst Biol* **4**, 222.
46. Liu, H., Sadygov, R.G. & Yates, J.R. (2004) A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Anal Chem* **76**, 4193–4201.
47. Washburn, M., Wolters, D. & Yates, J. (2001) Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat Biotechnol* **19**, 242–247.
48. Link, A. et al. (1999) Direct analysis of protein complexes using mass spectrometry. *Nat Biotechnol* **17**, 676–682.
49. Peng, J., Elias, J., Thoreen, C., Licklider, L. & Gygi, S. (2003) Evaluation of multidimensional chromatography coupled with tandem mass spectrometry (LC/LC-MS/MS) for large-scale protein analysis: the yeast proteome. *J Proteome Res* **2**, 43–50.
50. Dix, M.M., Simon, G.M. & Cravatt, B.F. (2008) Global mapping of the topography and magnitude of proteolytic events in apoptosis. *Cell* **134**, 679–691.
51. de Godoy, L.M.F. et al. (2008) Comprehensive mass-spectrometry-based proteome quantification of haploid versus diploid yeast. *Nature* **455**, 1251–1254.
52. Ghaemmaghami, S. et al. (2003) Global analysis of protein expression in yeast. *Nature* **425**, 737–741.
53. Huh, W.-K. et al. (2003) Global analysis of protein localization in budding yeast. *Nature* **425**, 686–691.
54. Nesvizhskii, A.I. & Aebersold, R. (2005) Interpretation of shotgun proteomic data: the protein inference problem. *Mol Cell Proteomics* **4**, 1419–1440.
55. Gygi, S.P. et al. (1999) Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat Biotechnol* **17**, 994–999.
56. Gevaert, K. et al. (2002) Chromatographic isolation of methionine-containing peptides for gel-free proteome analysis: identification of more than 800 Escherichia coli proteins. *Mol Cell Proteomics* **1**, 896–903.
57. Gevaert, K. et al. (2007) A la carte proteomics with an emphasis on gel-free techniques. *Proteomics* **7**, 2698–2718.
58. Gevaert, K. et al. (2003) Exploring proteomes and analyzing protein processing by mass spectrometric identification of sorted N-terminal peptides. *Nat Biotechnol* **21**, 566–569.
59. Staes, A. et al. (2008) Improved recovery of proteome-informative, protein N-terminal peptides by combined fractional diagonal chromatography (COFRADIC). *Proteomics* **8**, 1362–1370.
60. Jensen, O. (2006) Interpreting the protein language using proteomics. *Nat Rev Mol Cell Biol* **7**, 391–403.
61. Enoksson, M. et al. (2007) Identification of proteolytic cleavage sites by quantitative proteomics. *J Proteome Res* **6**, 2850–2858.
62. Timmer, J.C. et al. (2007) Profiling constitutive proteolytic events in vivo. *Biochem J* **407**, 41–48.
63. Doucet, A. & Overall, C.M. (2008) Protease proteomics: revealing protease in vivo functions using systems biology approaches. *Mol Aspects Med* **29**, 339–358.
64. Mahrus, S. et al. (2008) Global sequencing of proteolytic cleavage sites in apoptosis by specific labeling of protein N termini. *Cell* **134**, 866–876.
65. Andersson, L. & Porath, J. (1986) Isolation of phosphoproteins by immobilized metal (Fe-3+) affinity chromatography. *Anal Biochem* **154**, 250–254.
66. Bonenfant, D. et al. (2003) Quantitation of changes in protein phosphorylation: a simple method based on stable isotope labeling and mass spectrometry. *Proc Natl Acad Sci USA* **100**, 880–885.
67. Pinkse, M., Uitto, P., Hilhorst, M., Ooms, B. & Heck, A. (2004) Selective isolation at the femtomole level of phosphopeptides from proteolytic digests using 2D-nanoLC-ESI-MS/MS and titanium oxide precolumns. *Anal Chem* **76**, 3935–3943.
68. Mcnulty, D. & Annan, R. (2008) Hydrophilic interaction chromatography reduces the complexity of the phosphoproteome and improves global phosphopeptide isolation and detection. *Mol Cell Proteomics* **7**, 971.
69. Beausoleil, S. et al. (2004) Large-scale characterization of HeLa cell nuclear phosphoproteins. *Proc Natl Acad Sci USA* **101**, 12130–12135.
70. Geng, M., Zhang, X., Bina, M. & Regnier, F. (2001) Proteomics of glycoproteins based on affinity selection of glycopeptides from tryptic digests. *J Chromatogr B* **752**, 293–306.
71. Zhang, H., Yan, W. & Aebersold, R. (2004) Chemical probes and tandem mass spectrometry: a strategy for the quantitative analysis of proteomes and subproteomes. *Curr Opin Chem Biol* **8**, 66–75.

72. Khidekel, N. et al. (2007) Probing the dynamics of O-GlcNAc glycosylation in the brain using quantitative proteomics. *Nat Chem Biol* **3**, 339–348.
73. Choudhary, C. et al. (2009) Lysine acetylation targets protein complexes and co-regulates major cellular functions. *Science* **325**, 834–840.
74. Peng, J. et al. (2003) A proteomics approach to understanding protein ubiquitination. *Nat Biotechnol* **21**, 921–926.
75. Aebersold, R. & Mann, M. (2003) Mass spectrometry-based proteomics. *Nature* **422**, 198–207.
76. Lahm, H. & Langen, H. (2000) Mass spectrometry: a tool for the identification of proteins separated by gels. *Electrophoresis* **21**, 2105–2114.
77. Ong, S. et al. (2002) Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol Cell Proteomics* **1**, 376–386.
78. Mann, M. (2006) Functional and quantitative proteomics using SILAC. *Nat Rev Mol Cell Biol* **7**, 952–958.
79. Krueger, M. et al. (2008) SILAC mouse for quantitative proteomics uncovers kindlin-3 as an essential factor for red blood cell function. *Cell* **134**, 353–364.
80. Krijgsveld, J. et al. (2003) Metabolic labeling of C. elegans and D. melanogaster for quantitative proteomics. *Nat Biotechnol* **21**, 927–931.
81. Han, D.K., Eng, J., Zhou, H. & Aebersold, R. (2001) Quantitative profiling of differentiation-induced microsomal proteins using isotope-coded affinity tags and mass spectrometry. *Nat Biotechnol* **19**, 946–951.
82. Staes, A. et al. (2004) Global differential non-gel proteomics by quantitative and stable labeling of tryptic peptides with oxygen-18. *J Proteome Res* **3**, 786–791.
83. Ross, P. et al. (2004) Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. *Mol Cell Proteomics* **3**, 1154–1169.
84. Gerber, S., Rush, J., Stemman, O., Kirschner, M. & Gygi, S. (2003) Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. *Proc Natl Acad Sci USA* **100**, 6940–6945.
85. Malmström, J. et al. (2009) Proteome-wide cellular protein concentrations of the human pathogen Leptospira interrogans. *Nature* **460**, 762–765.
86. Choi, H., Fermin, D. & Nesvizhskii, A. (2008) Significance analysis of spectral count data in label-free shotgun proteomics. *Mol Cell Proteomics* **7**, 2373.
87. Wiener, M., Sachs, J., Deyanova, E. & Yates, N. (2004) Differential mass spectrometry: a label-free LC-MS method for finding significant differences in complex peptide and protein mixtures. *Anal Chem* **76**, 6085–6096.
88. Sturm, M. et al. (2008) OpenMS-An open-source software framework for mass spectrometry. *BMC Bioinformatics* **9**, 163.
89. Cox, J. & Mann, M. (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* **26**, 1367–1372.
90. Katajamaa, M. & Oresic, M. (2007) Data processing for mass spectrometry-based metabolomics. *J Chromatogr A* **1158**, 318–328.
91. Mann, M. & Wilm, M. (1994) Error-tolerant identification of peptides in sequence databases by peptide sequence tags. *Anal Chem* **66**, 4390–4399.
92. Reisinger, F. & Martens, L. (2009) Database on Demand – An online tool for the custom generation of FASTA-formatted sequence databases. *Proteomics* **9**, 4421–4424.
93. Nesvizhskii, A.I., Vitek, O. & Aebersold, R. (2007) Analysis and validation of proteomic data generated by tandem mass spectrometry. *Nat Methods* **4**, 787–797.
94. Yates, J.R., Eng, J.K., McCormack, A.L. & Schieltz, D. (1995) Method to correlate tandem mass spectra of modified peptides to amino acid sequences in the protein database. *Anal Chem* **67**, 1426–1436.
95. Perkins, D., Pappin, D., Creasy, D. & Cottrell, J. (1999) Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **20**, 3551–3567.
96. Fenyo, D. & Beavis, R. (2003) A method for assessing the statistical significance of mass spectrometry-based protein identifications using general scoring schemes. *Anal Chem* **75**, 768–774.
97. Colinge, J., Masselot, A., Giron, M., Dessingy, T. & Magnin, J. (2003) OLAV: towards high-throughput tandem mass spectrometry data identification. *Proteomics* **3**, 1454–1463.
98. Geer, L. et al. (2004) Open mass spectrometry search algorithm. *J Proteome Res* **3**, 958–964.
99. Frank, A. & Pevzner, P. (2005) PepNovo: de novo peptide sequencing via probabilistic network modeling. *Anal Chem* **77**, 964–973.
100. Johnson, R. & Taylor, J. (2002) Searching sequence databases via de novo peptide sequencing by tandem mass spectrometry. *Mol Biotechnol* **22**, 301–315.

101. Elias, J.E. & Gygi, S.P. (2007) Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat Methods* **4**, 207–214.
102. Nesvizhskii, A. et al. (2006) Dynamic spectrum quality assessment and iterative computational analysis of shotgun proteomic data – Toward more efficient identification of post-translational modifications, sequence polymorphisms, and novel peptides. *Mol Cell Proteomics* **5**, 652–670.
103. Flikka, K., Martens, L., Vandekerckhoe, J., Gevaert, K. & Eidhammer, I. (2006) Improving the reliability and throughput of mass spectrometry-based proteomics by spectrum quality filtering. *Proteomics* **6**, 2086–2094.
104. Keller, A., Nesvizhskii, A., Kolker, E. & Aebersold, R. (2002) Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal Chem* **74**, 5383–5392.
105. Kall, L., Canterbury, J.D., Weston, J., Noble, W.S. & MacCoss, M.J. (2007) Semi-supervised learning for peptide identification from shotgun proteomics datasets. *Nat Methods* **4**, 923–925.
106. Wan, Y., Yang, A. & Chen, T. (2006) PepHMM: a hidden Markov model based scoring function for mass spectrometry database search. *Anal Chem* **78**, 432–437.
107. Helsens, K., Timmerman, E., Vandekerckhove, J., Gevaert, K. & Martens, L. (2008) Peptizer: a tool for assessing false positive peptide identifications and manually validating selected results. *Mol Cell Proteomics* 7, 2363–2372.
108. Martens, L. & Hermjakob, H. (2007) Proteomics data validation: why all must provide data. *Mol Biosyst* **3**, 518–522.
109. Nesvizhskii, A., Keller, A., Kolker, E. & Aebersold, R. (2003) A statistical model for identifying proteins by tandem mass spectrometry. *Anal Chem* **75**, 4646–4658.
110. Mueller, M., Martens, L. & Apweiler, R. (2007) Annotating the human proteome: beyond establishing a parts list. *Biochim Biophys Acta* **1774**, 175–191.
111. Shannon, P. et al. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* **13**, 2498–2504.
112. Maere, S., Heymans, K. & Kuiper, M. (2005) BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics* **21**, 3448–3449.
113. Dennis, G. et al. (2003) DAVID: database for annotation, visualization, and integrated discovery. *Genome Biol* **4**, R60.
114. Kaplan, N., Vaaknin, A. & Linial, M. (2003) PANDORA: keyword-based analysis of protein sets by integration of annotation sources. *Nucleic Acids Res* **31**, 5617–5626.
115. Schneider, T.D. & Stephens, R.M. (1990) Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res* **18**, 6097–6100.
116. Colaert, N., Helsens, K., Martens, L., Vandekerckhove, J. & Gevaert, K. (2009) Improved visualization of protein consensus sequences by iceLogo. *Nat Methods* **6**, 786–787.
117. Matthiesen, R., Trelle, M., Hojrup, P., Bunkenborg, J. & Jensen, O. (2005) VEMS 3.0: algorithms and computational tools for tandem mass spectrometry based identification of post-translational modifications in proteins. *J Proteome Res* **4**, 2338–2347.
118. Rauch, A. et al. (2006) Computational Proteomics Analysis System (CPAS): an extensible, open-source analytic system for evaluating and publishing proteomic data and high throughput biological experiments. *J Proteome Res* **5**, 112–121.
119. Hakkinen, J., Vincic, G., Mansson, O., Warell, K. & Levander, F. (2009) The proteios software environment: an extensible multiuser platform for management and analysis of proteomics data. *J Proteome Res* **8**, 3037–3043.
120. Hartler, J. et al. (2007) MASPECTRAS: a platform for management and analysis of proteomics LC-MS/MS data. *BMC Bioinformatics* **8**, 197.
121. Helsens, K. et al. (2010) ms_lims, a simple yet powerful open source LIMS for mass spectrometry-driven proteomics. *Proteomics* **10**, 2560.
122. Martens, L. et al. (2005) PRIDE: the proteomics identifications database. *Proteomics* **5**, 3537–3545.
123. Slotta, D.J., Barrett, T. & Edgar, R. (2009) NCBI Peptidome: a new public repository for mass spectrometry peptide identifications. *Nat Biotechnol* **27**, 600–601.
124. Desiere, F. et al. (2006) The PeptideAtlas project. *Nucleic Acids Res* **34**, D655–D658.
125. Craig, R., Cortens, J.P. & Beavis, R.C. (2004) Open source system for analyzing, validating, and storing protein identification data. *J Proteome Res* **3**, 1234–1242.
126. Klie, S. et al. (2008) Analyzing large-scale proteomics projects with latent semantic indexing. *J Proteome Res* 7, 182–191.

127. Mueller, M. et al. (2008) Analysis of the experimental detection of central nervous system-related genes in human brain and cerebrospinal fluid datasets. *Proteomics* **8**, 1138–1148.
128. Gevaert K, Ghesquière B, et al (2004) Reversible labeling of cysteine-containing peptides allows their specific chromatographic isolation for non-gel proteome studies. Proteomics **4**, 897–908.
129. Gevaert K, Staes A, et al (2005) Global phosphoproteome analysis on human HepG2 hepatocytes using reversed-phase diagonal LC. Proteomics **5**, 3589–3599.
130. Ghesquière B, Van Damme J, et al (2006) Proteome-wide characterization of N-glycosylation events by diagonal chromatography. J Proteome Res **5**, 2438–2447.
131. Hanoulle X, Van Damme J, et al (2006) A new functional, chemical proteomics technology to identify purine nucleotide binding sites in complex proteomes. J Proteome Res **5**, 3438–3445.
132. Ghesquière B, Buyl L, et al (2007) A new approach for mapping sialylated N-glycosites in serum proteomes. J Proteome Res **6**, 4304–4312.
133. Ghesquière B, Colaert N, et al (2009) In vitro and in vivo protein-bound tyrosine nitration characterized by diagonal chromatography. Mol Cell Proteomics **8**, 2642–2652.

Chapter 2

Metabolic Labeling of Model Organisms Using Heavy Nitrogen (^{15}N)

Joost W. Gouw, Bastiaan B.J. Tops, and Jeroen Krijgsveld

Abstract

Quantitative proteomics aims to identify and quantify proteins in cells or organisms that have been obtained from different biological origin (e.g., "healthy vs. diseased"), that have received different treatments, or that have different genetic backgrounds. Protein expression levels can be quantified by labeling proteins with stable isotopes, followed by mass spectrometric analysis. Stable isotopes can be introduced in vitro by reacting proteins or peptides with isotope-coded reagents (e.g., iTRAQ, reductive methylation). A preferred way, however, is the metabolic incorporation of heavy isotopes into cells or organisms by providing the label, in the form of amino acids (such as in SILAC) or salts, in the growth media. The advantage of in vivo labeling is that it does not suffer from side reactions or incomplete labeling that might occur in chemical derivatization. In addition, metabolic labeling occurs at the earliest possible moment in the sample preparation process, thereby minimizing the error in quantitation. Labeling with the heavy stable isotope of nitrogen (i.e., ^{15}N) provides an efficient way for accurate protein quantitation. Where the application of SILAC is mostly restricted to cell culture, ^{15}N labeling can be used for micro-organisms as well as a number of higher (multicellular) organisms. The most prominent examples of the latter are *Caenorhabditis elegans* and *Drosophila* (fruit fly), two important model organisms for a range of regulatory processes underlying developmental biology. Here we describe in detail the labeling with ^{15}N atoms, with a particular focus on fruit flies and *C. elegans*. We also describe methods for the identification and quantitation of ^{15}N-labeled proteins by mass spectrometry and bioinformatic analysis.

Key words: Stable isotope labeling, proteomics, nitrogen, model organism, mass spectrometry, protein quantitation, *Drosophila*, *C. elegans*.

1. Introduction

Labeling of proteins with stable isotopes has provided a strong impetus to quantitative proteomics over the past few years (1). It entails the incorporation, either in vitro or in vivo, of stable

K. Gevaert, J. Vandekerckhove (eds.), *Gel-Free Proteomics*, Methods in Molecular Biology 753,
DOI 10.1007/978-1-61779-148-2_2,

isotope-labeled atoms like ^{2}H, ^{13}C, ^{15}N, or ^{18}O into proteins, resulting in a mass increase that can be detected by mass spectrometry. In mixtures of differentially labeled samples, spectral intensities of light and heavy peptides directly reflect their relative abundance, thus providing a measure for the relative expression levels of proteins. Proteins can be derivatized in vitro (e.g., by methods such as iTRAQ (2), ICAT (3), or reductive dimethylation (4)) or in vivo by the metabolic incorporation of the label during growth of the target organism. Isotope-tagged amino acids have been widely used for the labeling of cell cultures (SILAC; (3)) and (lower) organisms that can be cultured in defined media (5). Labeling with heavy nitrogen, replacing all naturally occurring ^{14}N by ^{15}N atoms, provides an alternative means with applications in (multicellular) organisms. While unicellular organisms (bacteria, yeast) can be readily cultured in ^{15}N-enriched media (6), we and others have demonstrated the labeling of *Caenorhabditis elegans* (7, 8), fruit flies (7, 9), plants (10), and mammals (11) with ^{15}N atoms. We have applied metabolic labeling of fruit flies to analyze the mother-to-zygote transition, uncovering both maternal- and zygote-specific proteins (12). In *C. elegans*, we have compared protein expression in male and female animals, disclosing expression of sex-specific proteins (13). The same approach has been used by others to analyze insulin signaling (8). Here we provide a step-wise protocol for metabolic labeling with heavy nitrogen of *Escherichia coli*, yeast, *C. elegans*, and *Drosophila*, along with procedures for protein analysis by mass spectrometry. Finally, methods are provided for bioinformatic interpretation of the data, including protein identification and quantitation.

2. Materials

Deionized water, 18 MΩ, needs to be used for all solutions and buffers.

2.1. Labeling of E. coli

1. *E. coli*, OP50 strain (streak plate or glycerol stock) (Caenorhabditis Genetics Center, University of Minnesota)
2. Media for labeling: Spectra 9-U medium (unlabeled) and Spectra 9-N medium (^{15}N-labeled) (Cambridge Isotope Laboratories)

2.2. Labeling of C. elegans

1. *C. elegans* (Caenorhabditis Genetics Center, University of Minnesota)

2. Pellets of labeled (^{14}N and ^{15}N) OP50 *E. coli* (from a freshly grown 400 ml culture)
3. 1 M $CaCl_2$
4. 1 M $MgSO_4$
5. 5 mg/ml cholesterol in ethanol
6. 1 M potassium phosphate buffer, pH 6.0: dissolve 108.3 g KH_2PO_4 and 35.6 g K_2HPO_4 in 1 l of water
7. Agarose
8. Media for labeling: Spectra 9-U and Spectra 9-N media
9. M9 buffer: 3 g KH_2PO_4, 6 g Na_2HPO_4, 5 g NaCl, 1 ml 1 M $MgSO_4$, add water to 1 l
10. Petri dishes (preferably 15 cm diameter)

2.3. Labeling of Yeast

1. Minimal medium: dissolve 1.7 g yeast nitrogen base without amino acids and ammonium sulfate (Difco), 20 g sucrose (nitrogen-free), and 5 g ^{15}N-labeled ammonium sulfate (Cambridge Isotope Laboratories) in 1 l of water (*see* **Note 1**).
2. *Saccharomyces cerevisiae* type II (Sigma).
3. Phosphate-buffered saline (PBS): dissolve 8 g NaCl, 200 mg KCl, 1.44 g Na_2HPO_4, and 240 mg KH_2PO_4 in 900 ml of water. Adjust the pH to 7.4 using HCl or NaOH and add water to 1 l.
4. Falcon tubes, 50 ml.
5. Sterile flasks of 50 ml and 2.5 l.

2.4. Labeling of Drosophila

1. Larva box that contains enough openings to allow for ventilation. However, all openings should be covered with fine gauze to prevent the escape of hatched flies (*see* **Note 2**).
2. Fly collection cage (cylindrical) with one side covered by fine gauze and the other side by a Petri dish (*see* **Note 3**).
3. Petri dishes that can be mounted onto the fly collection cage (i.e., the diameter should be similar to the fly collection cage, see above).
4. 10% (w/v) Tegosept: dissolve 5 g of *p*-hydroxybenzoic acid methyl ester in 50 ml of 95% ethanol.
5. Ampicillin (50 mg/ml): dissolve 0.5 g ampicillin in 10 ml of water.
6. Larva box mixture: dissolve 9 g sucrose (nitrogen-free) and 9 g of dry weight yeast in 70 ml of water (if fresh yeast culture is used instead of dry yeast, then the volume of water should be corrected accordingly). Add 37.5 μl of propionic

acid, 240 μl of phosphoric acid, 840 μl of Tegosept, and 60 μl of ampicillin to this sucrose/yeast mixture.

7. Sheet(s) of cotton wool.
8. Whatman #1 filter paper, about 3 cm in diameter.
9. Food and collection plates: dissolve 6.25 g sucrose (nitrogen-free) and 7.5 g pure agarose in 375 ml of water and autoclave this solution (this will also dissolve the agar). Cool down to 55°C and add 1.88 ml of Tegosept and 1.88 ml of ethyl acetate. Dispense 20 ml of this solution into the Petri dishes and let this solidify (15–30 min). Store the plates in plastic bags at 4°C.

2.5. Analyzing the Level of ^{15}N-Enrichment and Optimizing the Mixing Ratio of Labeled and Unlabeled Peptides

1. Polypropylene tubes (e.g., 1.5 ml from Eppendorf).
2. Micropestles (Eppendorf) or ultrasonic homogenizer (e.g., the LABSONIC® M from Sartorius Stedim).
3. 50 mM ammonium bicarbonate (it is not necessary to adjust the pH).
4. Lysis buffer: 8 M urea and 2 M thiourea in 50 mM ammonium bicarbonate supplemented with a Protease Inhibitor Cocktail tablet (Roche).
5. 200 mM dithiothreitol in 50 mM ammonium bicarbonate.
6. 200 mM iodoacetamide in 50 mM ammonium bicarbonate. This solution must be prepared fresh and should be kept in the dark.
7. Trypsin (sequencing-grade modified trypsin, Promega) in aliquots of 20 μg in 40 μl of 50 mM acetic acid per tube (0.5 μg/μl) should be kept at –20°C.

2.6. Analysis of Labeled Proteins by LC-MS/MS

1. High-performance liquid chromatography (HPLC)-grade acetonitrile and acetic acid.
2. HPLC buffer A: 0.1 M acetic acid in water or 0.1% (v/v) formic acid in water.
3. HPLC buffer B: 0.1 M acetic acid (or 0.1% (v/v) formic acid) in 8/2 (v/v) acetonitrile/water.
4. Nanoflow liquid chromatography system (e.g., Agilent 1200 system consisting of an optional vacuum degasser, binary (nano) pump, and (micro) autosampler).
5. Reversed-phase C_{18} analytical column (e.g., in-house packed; 50 μm × 20 cm fused-silica capillary (Polymicro) with ReproSil-Pur C18-AQ 3 μm particles, 200 Å pore size (Dr. Maisch GmbH)). Other columns (50–100 μm internal diameter, packed or monolithic) will work as well.
6. Reversed-phase C_{18} trap column (e.g., in-house packed; 100 μm × 2 cm fused-silica capillary (Polymicro) with Aqua C_{18} 5 μm particles, 200 Å pore size (Phenomenex)).

7. High-resolution electrospray hybrid mass spectrometer (e.g., Q-TOF, LTQ-Orbitrap, LTQ-FT).
8. The HPLC, columns, and mass spectrometer are connected essentially as described previously (14).

3. Methods

3.1. Labeling of E. coli

1. Inoculate a small volume (5 ml) of bacterial culture medium in a Falcon tube. For ^{14}N and ^{15}N labeling of bacteria, Spectra 9-U and Spectra 9-N media are used, respectively.
2. Culture bacteria for 8–15 h (overnight) at 37°C with shaking at 250 rpm.
3. Inoculate a large volume (400 ml) of the appropriate culture medium with 0.1–1 ml of the overnight culture.
4. Culture bacteria for 8–15 h (overnight) at 37°C with shaking at ~250 rpm (*see* **Note 4**).
5. Harvest bacteria by spinning at 2,400×*g* for 10 min.

3.2. Labeling of C. elegans

C. elegans animals are typically cultured on nematode growth medium (NGM) agar seeded with OP50 *E. coli*. NGM agar contains NaCl, bacterial nutrients (peptone), cholesterol, buffering salts (phosphate buffer, pH 6.0), and agar. To ensure optimal isotopic labeling of *C. elegans*, all nitrogen-containing ingredients of the NGM medium should be replaced either by the Spectra-9 medium or by nitrogen-free substitutes. Plates used to metabolically label *C. elegans* are therefore prepared with agarose (1% w/v) instead of agar and contain Spectra-9 medium (25% v/v) instead of peptone and NaCl. Preparation for 1 l:

1. Mix 10 g of agarose and 250 ml of Spectra-N medium in a flask or bottle. Add 725 ml of water. Autoclave for 15 min at 121°C (agarose cannot be autoclaved too long due to caramelization).
2. Cool flask in a water bath or stove at 55°C.
3. Add 1 ml of 1 M $CaCl_2$, 1 ml of 5 mg/ml cholesterol in ethanol, 1 ml of 1 M $MgSO_4$, and 25 ml of 1 M phosphate buffer.
4. Swirl to mix well.
5. Dispense the solution into Petri dishes (*see* **Note 5**).
6. Leave plates at room temperature for 1–2 days before use to allow excess moisture to evaporate.
7. Re-suspend equal amounts of pelleted ^{14}N- and ^{15}N-labeled *E. coli* in M9 buffer (typically 400 ml of bacterial culture is pelleted and re-suspended in 50 ml M9 buffer).

8. Put 2–3 ml of the bacterial suspension on the plates and spread using a glass rod.
9. Allow excess moisture to evaporate and the OP50 *E. coli* lawn to grow by leaving the plates overnight at room temperature.
10. At this point the plates can be used to culture *C. elegans* or be stored at 4°C for several weeks. To ensure optimal labeling of *C. elegans*, culture animals for at least four generations (preferably more) on the isotope plates before protein extracts are generated (*see* **Note 6**).
11. When sufficient numbers of animals are cultured, harvest the animals by rinsing the plates with M9 medium and spinning the animals at 400×*g* for 2 min. To remove excess bacteria, re-suspend pelleted *C. elegans* in M9 buffer in a 50 ml Falcon tube and allow to settle on ice (the bacteria will remain in suspension).

3.3. Labeling of Yeast

1. Re-suspend a couple of yeast grains in 5 ml of water.
2. Use this yeast to inoculate 5 ml minimal media in a 50 ml flask.
3. Grow overnight at 30°C with shaking at 230–270 rpm.
4. Inoculate 1 l of minimal media in a 2.5 l flask with 500 μl of the overnight culture.
5. Grow overnight at 30°C with shaking at 230–270 rpm.
6. Collect yeast in 500 ml GSA bottles and centrifuge at 2,400×*g* for 20 min at 4°C.
7. Decant the supernatant. The supernatant can be incubated overnight (at 30°C and shaking at 230–270 rpm) to collect additional yeast.
8. Re-suspend the remaining yeast (pellet) in 20 ml of PBS, transfer to a 50 ml Falcon tube, and centrifuge at 2,400×*g* for 20 min at 4°C. Remove the PBS and use this tube to pool subsequent yeast (e.g., from the incubated supernatant). If the supernatant is incubated, collect the yeast and pool this with the previously collected yeast.
9. Store the Falcon tube at 4°C for direct use or at –80°C for long-term storage (*see* **Note 7**).

3.4. Labeling of Drosophila

The following instructions assume the use of newly grown wild-type *Drosophila melanogaster* embryos (*see* **Note 8**).

1. Collect per larva box 20–30 mg of 0–12 h-old embryos and rinse them extensively with water, followed by a brief rinse with 70% ethanol (note that 100% ethanol will dehydrate the embryos). Store the embryos at room temperature until the larva box is ready.

2. Prepare the larva box by placing a layer of cotton on the bottom of the box and soak the cotton with the larva box mixture. The cotton should be completely soaked, but without leaving any pools of liquid in the box. Close the box until the embryos are ready.
3. Spray the Whatman filter with 70% ethanol, place the collected and washed embryos onto the filter, and distribute them evenly using a fine brush.
4. Place the Whatman filter with the embryos in the center of the larva box and close the box (*see* **Note 9**).
5. Prepare the Petri dishes required for the fly collection cage. Note that these dishes must be replaced when they are depleted of yeast or when (staged) embryos are to be collected.
6. When a significant amount of flies have hatched, transfer them to the fly collection cage. Prepare the collection cage by adding some labeled or unlabeled yeast to the Petri dish (*see* **Note 10**).
7. Transfer the hatched flies to the fly collection cage by holding the larva box upside down and filling the box with carbon dioxide gas. When holding the larva box upside down, be very careful that flies do not get wet due to condensed water. The CO_2 will anesthetize the flies making it easier to transfer them. Optionally, during this step, the volume of flies can be determined using a measuring cylinder. A total of 30 mg of embryos yield around 5 ml of flies.
8. Directly after transferring the flies to the collection cage, use the Petri dish to cover one side of the fly collection cage (the other side is covered by fine gauze).
9. Put the larva box to 25°C and 80% humidity to allow hatching of the remaining pupae. These flies can then be added to the previously collected flies using Step 8.

3.5. Analyzing the Level of ^{15}N-Enrichment

There are two separate, yet equally important, phenomena that contribute independently to the final percentage of the stable isotope in proteins: (1) the purity of the stable isotope that is obtained from the supplier (e.g., 99% ^{15}N) and (2) the degree of incorporation (efficiency) of that stable isotope into proteins. Although labeling with ^{15}N can be very efficient, even an incorporation efficiency as high as 98% causes some challenges for proper peak selection by the mass spectrometer, peptide identification, and quantitation (15). It is therefore recommended to use the highest purity heavy nitrogen (i.e., 99+% ^{15}N) that is available and to minimize other sources of unlabeled nitrogen throughout metabolic incorporation of the label. The final enrichment level should be determined to ensure complete incorporation

of the label, which can be determined by LC-MS as follows (*see* **Note 11**):

1. Transfer 1 mg of heavy-labeled sample (*E. coli*, yeast, flies, *C. elegans*) into an appropriately labeled tube. Add 100 μl of ice-cold lysis buffer and lyse the organisms on ice by sonication or by using micropestles.
2. Clear the mixture by centrifugation for 20 min at 20,000×*g* and 4°C.
3. Transfer the supernatant to a new tube and add dithiothreitol to a final concentration of 2 mM (1 μl of dithiothreitol stock solution per 100 μl of sample solution).
4. Incubate at 56°C for 25 min.
5. After cooling to room temperature, add iodoacetamide to a final concentration of 4 mM (2 μl of iodoacetamide stock solution per 100 μl of sample solution) and incubate for 30 min in the dark at room temperature.
6. Dilute the sample four times with ammonium bicarbonate (*see* **Note 12**), add 2 μl (1 μg) of trypsin, and incubate overnight at 37°C.
7. Purify the peptides by reversed-phase solid phase extraction (*see* **Note 13**).
8. Analyze around 1 μg of peptides by LC-MS/MS (*see* **Section 3.7**).
9. Search the data in a database search engine to identify the peptides (*see* **Note 14**).
10. Select several (i.e., 20) high-scoring, abundant peptides that differ in mass and hence chemical composition and determine their chemical formulae using their amino acid sequence, charge, and modifications if present (e.g., oxidized methionine, *see* **Note 15**).
11. Simulate (with software like IsoPro 3.0, *see* **Note 16**) isotope distributions using the peptide's chemical formula and decreasing ^{15}N-enrichments. Start with 100% ^{15}N and decrease this in a step-wise manner (e.g., 0.5% steps). Compare each distribution (i.e., the height of all of the isotopes) to the summed experimental isotope distribution of the peptide and the best "fit" is the peptide's actual ^{15}N-enrichment.
12. The actual ^{15}N-enrichment of each of the 20 peptides should be very similar. If this is not the case, this most likely indicates that labeling is not complete yet and that extended labeling (i.e., the next generation) is required. If, on the contrary, the actual enrichment of all the peptides is very similar, and if the average of these enrichments is close to the purchased purity of the label, full incorporation is reached (*see* **Note 17**).

3.6. Optimizing the Mixing Ratio of Labeled and Unlabeled Proteins

A critical aspect in quantitative proteomics is the mixing of two or more differentially labeled samples. Preferably, proteins that do not change in abundance between conditions should be present in equal amounts (i.e., a 1:1 ratio). Therefore, one should aim for mixing samples in a 1:1 ratio based on protein content. When samples are mixed in suboptimal ratios, quantitation might be difficult for regulated proteins due to the limited dynamic range of mass spectrometers. There are several ways to mix differentially labeled samples which can be as simple as combining an equal number of cells or embryos. Other methods are based on absolute protein amounts estimated by protein assays or the intensity of separated proteins on a 1D gel using SDS-PAGE. These provide good approximations, but are not accurate enough to achieve an exact 1:1 ratio. To prepare a 1:1 mixture of unlabeled and ^{15}N-labeled protein the following procedure can be used:

1. Add 1 mg of unlabeled sample (*C. elegans* or *Drosophila*) to 3 mg of the corresponding labeled sample (tube 1).
2. In a second tube, add 1 mg of labeled sample (*C. elegans* or *Drosophila*) to 3 mg of the corresponding unlabeled sample (tube 2).
3. Add 100 μl of ice-cold lysis buffer to both samples and lyse them on ice by sonication or by using micropestles.
4. For the samples in both tubes, follow Steps 2–6 described in **Section 3.5**.
5. Analyze both samples by LC-MS by injecting 0.5 μl of each sample.
6. Search the data in a database search engine and quantify a number of abundant proteins (i.e., proteins with a high number of peptides) that are expected not to change in abundance using quantification software to determine the actual mix ratio. The ratio of these proteins should be similar and the average ratio represents the actual mix ratio in each of the tubes.
7. Based on these actual mixing ratios, the volumes can be calculated that need to be combined from tubes 1 and 2 to achieve a 1:1 ratio. Mix in a new tube calculated volumes of samples 1 and 2 using the following formula (*see* **Note 18**):

$$V_{\text{sample 1}} = V_{\text{total}} \frac{R_{\text{desired}} - R_{\text{sample 2}}}{R_{\text{sample 1}} - R_{\text{sample 2}}}$$

where $V_{\text{sample 1}}$ is the volume of sample 1 to be added to a new tube, V_{total} is the total volume of the new tube, R_{desired} is the desired mix ratio (usually 1), $R_{\text{sample 1}}$ and $R_{\text{sample 2}}$ are the actual mix ratios of samples 1 and 2, respectively. After calculating the volume of sample 1, the volume of sample 2

($V_{total} - V_{sample\ 1}$) should be added to get the desired mix ratio.

8. Purify the peptides by reversed-phase solid phase extraction (*see* **Note 13**).
9. This sample is ready for further proteomic analysis (LC-MS/MS) (*see* **Section 3.7**).

3.7. Analysis of Labeled Proteins by LC-MS/MS

Mixtures of labeled and unlabeled peptides are best analyzed by liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS) for protein identification and quantitation. This process involves the online separation of peptides by reversed-phase LC, electrospray ionization of peptides, and fragmentation of detected peptides in the mass spectrometer. This setup can be achieved on multiple platforms (i.e., various types of LC systems coupled to a range of mass spectrometers), and the exact protocol depends on available instrumentation (for reviews, *see* (16, 17)). Therefore, we will describe the workflow only in general terms, highlighting some aspects that should be kept in mind for optimal performance for a quantitative analysis.

1. A chromatographic system should be chosen that can deliver flow rates at 100–300 nl/min, either as a splitless nanoflow system or as a conventional system running at 100–300 μl/min and passive splitting to the desired flow rate.
2. Choose a reversed-phase column (either pre-packed or homemade) that efficiently captures both hydrophilic and hydrophobic peptides. The internal diameter should be in the range of 50–100 μm; length can vary between 10 and 40 cm depending on gradient length and flow rate, but 15–20 cm is a good starting point. Slope and length of the gradient should be optimized for the sample and column system. For complex samples, a typical gradient can be generated by raising the concentration of acetonitrile from 5% to approximately 40% over a 2 h period (*see* **Note 19**).
3. The amount of sample (peptide mixture) injected into the system should not exceed the capacity of the column. Overloading will cause peak broadening and (potentially) saturation of the detector in the mass spectrometer. Both will compromise proper quantitation.
4. Ideally, a high-resolution mass spectrometer should be used. Resolution should be sufficient to determine the charge state of the peptide and the mass of the mono-isotopic peak. Enhanced resolution also helps in distinguishing labeled from non-labeled peptides based on isotope pattern. High resolution (such as in TOF, Orbitrap, or FT instruments) is usually coupled to high mass accuracy, aiding in the identification process.

5. After processing raw data to peaklists (often by vendor-specific software), proteins can be identified by a range of database search algorithms. Most, but not all of them, have the option to identify ^{15}N-labeled proteins (*see* **Note 14**).
6. Protein quantitation is a critical process that is supported by an increasing number of software packages. Although standardization of mass spectral data formats is steadily progressing (e.g., mzXML, mzML), the use of most software packages for protein quantitation is often dictated (and limited) by the data format of the mass spectrometer that was used. There are a number of options for quantitation of ^{15}N-labeled proteins. MSQuant (18) and Census (19) support a range of data formats and are available free of charge. Mascot distiller (www.matrixscience.com) is a commercial package supporting nearly all data formats.

4. Notes

1. The minimal media should not contain any unlabeled nitrogen (in the form of amino acids or residual salts) and it is advised to use highly ^{15}N-enriched (>99%) ammonium sulfate. This will result in the highest possible enrichment in yeast and (eventually) in *Drosophila*, thereby enhancing the accuracy of quantitation by mass spectrometry.
2. The larva box used in these experiments had the dimensions of 20 × 15 × 10 cm and the amount of embryos for this box can vary between 10 and 500 mg.
3. When a cylindrical fly collection cage is used that fits on a Petri dish, flies can easily be fed by replacing the Petri dishes (with labeled or unlabeled yeast). In addition, (staged) embryos can also be collected simply by replacing the Petri dish at desired intervals. A fly cage with an approximate diameter of 9 cm and height of 10 cm can hold an optimal amount of 5 ml of flies.
4. *E. coli* tends to grow slightly slower in Spectra-9 N medium than in Spectra-9 U medium.
5. Since a rather large amount of *C. elegans* animals is needed for proteomic experiments, 15 cm ø Petri dishes are best suited.
6. Depending on the *C. elegans* strain(s) used, different culturing conditions might be required, but animals are usually cultured at 15–20°C. It will take approximately 12–15 days to culture four generations at 20°C. Since these

animals have a life span of 2–3 weeks, it is best to pick individual (young) animals when transferring animals to a new plate to get rid of the older partially metabolically labeled animals.

7. Yeast can be used for *Drosophila* labeling experiments after 12 months of storage at –80°C.
8. We have used the *D. melanogaster* strain Oregon-R, but this protocol should be adaptable to many other fly strains.
9. When the larva boxes are kept at 25°C and 80% humidity throughout larval and pupal developmental stages, flies hatch after approximately 9 days.
10. A small amount of labeled or unlabeled yeast is added to the Petri dish and serves as the food source for the flies. When the dish is depleted of yeast, it should be replaced with a fresh one. The amount of yeast spread on the plate depends on the amount of time the plate is left in the fly cage. For shorter periods of time (0–3 h), a medium strip of yeast (about 3 cm in diameter and 0.5 cm deep) is sufficient, but more yeast is required for longer periods (e.g., overnight).
11. Alternatively, if no LC-MS/MS system is available, the level of incorporation can be determined by MALDI-TOF mass spectrometry. To this end, labeled proteins should be separated by SDS-PAGE gel electrophoresis, preferably on a large (15 cm) gel. After Coomassie staining, a number of intense bands can be digested with trypsin, followed by MALD-TOF analysis and protein identification. The disadvantage of this approach is that it is very likely that even a single gel band will still contain multiple proteins, which might hamper the protein identification process.
12. The urea concentration needs to be 2 M or less to prevent inhibitory effects on trypsin activity.
13. Protein digests can be desalted using ZipTips (Millipore) or homemade tips. To create your own, pack a small plug of C_{18} material (3 M Empore C18 extraction disk) into a GELoader tip (Eppendorf) similar to what has been described previously (20). The tip (homemade or ZipTips) is first washed with 2 × 20 μl of acetonitrile, followed by 20 μl of 8/2 (v/v) acetonitrile/water with 0.1 M acetic acid and finally equilibrated with 2 × 20 μl of 0.1 M acetic acid. The sample is then added onto the material in steps of 20 μl, followed by washing with 20 μl of 0.1 M acetic acid. The peptides are eluted and collected in a 0.5 ml tube with 2 × 20 μl of 8/2 (v/v) acetonitrile/water with 0.1 M acetic acid and dried using a vacuum centrifuge.

14. There are several search algorithms that allow identification of ^{15}N-labeled proteins. In Mascot (http://www.matrixscience.com/), select "^{15}N metabolic" in the "quantitation" drop-down menu. In The GPM (http://human.thegpm.org), check the box "all ^{15}N amino acids." In Sequest, you need to define a new set of amino acids in which all ^{14}N masses are replaced by their ^{15}N-isotope.
15. MS-Isotope (http://prospector.ucsf.edu/prospector/cgi-bin/msform.cgi?form=msisotope) is a useful tool to calculate the chemical composition of peptides and their modifications. It also provides the isotope distribution of (unlabeled) peptides.
16. The software tool IsoPro 3.0 can be used to simulate isotope distributions based on varying abundances of natural occurring elements. IsoPro used to be available as freeware, but we noticed that the Website has been discontinued recently. Ask the authors (J.W.G or J.K.) for availability.
17. If the actual enrichment of the peptides is very similar and if the average of these enrichments is not close to the purchased purity of the label, this is most likely caused by incomplete labeling, and longer periods of labeling might still be required. Alternatively, other sources of unlabeled nitrogen can affect the purity of the label and introduce similar effects. Be absolutely certain that the chemicals used in the labeling experiments are nitrogen-free (e.g., use nitrogen-free sucrose instead of sugar to prepare the larva box mixture). Starting from a 99% pure source of ^{15}N, incorporation efficiencies of 97–99% should be achievable.
18. This method of mixing is based on the assumption that several abundant proteins do not change in abundance between both conditions. If there are differences expected between these proteins, another method to mix is recommended.
19. The easiest way to check the condition of your HPLC system and reversed-phase columns is by injecting a standard peptide mixture (e.g., tryptic digest of bovine serum albumin).

References

1. Ong, S. E., and Mann, M. (2005) Mass spectrometry-based proteomics turns quantitative, *Nat Chem Biol 1*, 252–262.
2. Ross, P. L., Huang, Y. N., Marchese, J. N., Williamson, B., Parker, K., Hattan, S., Khainovski, N., Pillai, S., Dey, S., Daniels, S., Purkayastha, S., Juhasz, P., Martin, S., Bartlet-Jones, M., He, F., Jacobson, A., and Pappin, D. J. (2004) Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents, *Mol Cell Proteomics 3*, 1154–1169.
3. Gygi, S. P., Rist, B., Gerber, S. A., Turecek, F., Gelb, M. H., and Aebersold, R. (1999)

Quantitative analysis of complex protein mixtures using isotope-coded affinity tags, *Nat Biotechnol 17*, 994–999.
4. Boersema, P. J., Raijmakers, R., Lemeer, S., Mohammed, S., and Heck, A. J. (2009) Multiplex peptide stable isotope dimethyl labeling for quantitative proteomics, *Nat Protoc 4*, 484–494.
5. Everley, P. A., Bakalarski, C. E., Elias, J. E., Waghorne, C. G., Beausoleil, S. A., Gerber, S. A., Faherty, B. K., Zetter, B. R., and Gygi, S. P. (2006) Enhanced analysis of metastatic prostate cancer using stable isotopes and high mass accuracy instrumentation, *J Proteome Res 5*, 1224–1231.
6. Oda, Y., Huang, K., Cross, F. R., Cowburn, D., and Chait, B. T. (1999) Accurate quantitation of protein expression and site-specific phosphorylation, *Proc Natl Acad Sci USA 96*, 6591–6596.
7. Krijgsveld, J., Ketting, R. F., Mahmoudi, T., Johansen, J., Artal-Sanz, M., Verrijzer, C. P., Plasterk, R. H., and Heck, A. J. (2003) Metabolic labeling of *C. elegans* and *D. melanogaster* for quantitative proteomics, *Nat Biotechnol 21*, 927–931.
8. Dong, M. Q., Venable, J. D., Au, N., Xu, T., Park, S. K., Cociorva, D., Johnson, J. R., Dillin, A., and Yates, J. R., 3rd. (2007) Quantitative mass spectrometry identifies insulin signaling targets in *C. elegans*, *Science 317*, 660–663.
9. Findlay, G. D., Yi, X., Maccoss, M. J., and Swanson, W. J. (2008) Proteomics reveals novel *Drosophila* seminal fluid proteins transferred at mating, *PLoS Biol 6*, e178.
10. Kim, J. K., Harada, K., Bamba, T., Fukusaki, E.-i., and Kobayashi, A. (2005) Stable isotope dilution-based accurate comparative quantification of nitrogen-containing metabolites in Arabidopsis thaliana T87 Cells using in vivo ^{15}N-isotope enrichment, *Biosci Biotechnol Biochem 69*, 1331–1340.
11. Wu, C. C., MacCoss, M. J., Howell, K. E., Matthews, D. E., and Yates, J. R., 3rd. (2004) Metabolic labeling of mammalian organisms with stable isotopes for quantitative proteomic analysis, *Anal Chem 76*, 4951–4959.
12. Gouw, J. W., Pinkse, M. W., Vos, H. R., Moshkin, Y. M., Verrijzer, C. P., Heck, A. J. R., and Krijgsveld, J. (2009) In vivo stable isotope labeling of fruit flies reveals post-transcriptional regulation in the maternal-to-zygotic transition, *Mol Cell Proteomics 8*, 1566–1578.
13. Krijgsveld, J., Tops, B. B., Gauci, S., and Heck, A. J. (2009) Worms from Venus and Mars: Proteomics profiling of sexual differences in *Caenorhabditis elegans* using in vivo 15 N isotope labeling, *J Proteome Res 9*, 341–351.
14. Meiring, H. D., van der Heeft, E., ten Hove, G. J., and de Jong, A. P. J. M. (2002) Nanoscale LC-MS$^{(n)}$: Technical design and applications to peptide and protein analysis, *J Sep Sci 25*, 557–568.
15. Gouw, J. W., Tops, B. B. J., Mortensen, P., Heck, A. J. R., and Krijgsveld, J. (2008) Optimizing identification and quantitation of ^{15}N-labeled proteins in comparative proteomics, *Anal Chem 80*, 7796–7803.
16. Motoyama, A., and Yates, J. R., 3rd. (2008) Multidimensional LC separations in shotgun proteomics, *Anal Chem 80*, 7187–7193.
17. Qian, W. J., Jacobs, J. M., Liu, T., Camp, D. G., 2nd, and Smith, R. D. (2006) Advances and challenges in liquid chromatography-mass spectrometry-based proteomics profiling for clinical applications, *Mol Cell Proteomics 5*, 1727–1744.
18. Mortensen, P., Gouw, J. W., Olsen, J. V., Ong, S. E., Rigbolt, K. T., Bunkenborg, J., Cox, J., Foster, L. J., Heck, A. J., Blagoev, B., Andersen, J. S., and Mann, M. (2010) MSQuant, an open source Platform for mass spectrometry-based quantitative proteomics, *J Proteome Res 9*, 393–403.
19. Park, S. K., Venable, J. D., Xu, T., and Yates, J. R., 3rd. (2008) A quantitative analysis software tool for mass spectrometry-based proteomics, *Nat Methods 5*, 319–322.
20. Rappsilber, J., Ishihama, Y., and Mann, M. (2003) Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics, *Anal Chem 75*, 663–670.

Chapter 3

Trypsin-Catalyzed Oxygen-18 Labeling for Quantitative Proteomics

Wei-Jun Qian, Brianne O. Petritis, Carrie D. Nicora, and Richard D. Smith

Abstract

Stable isotope labeling based on relative peptide/protein abundance measurements is commonly applied for quantitative proteomics. Recently, trypsin-catalyzed oxygen-18 labeling has grown in popularity due to its simplicity, cost-effectiveness, and its ability to universally label peptides with high sample recovery. In ^{18}O labeling, both C-terminal carboxyl group atoms of tryptic peptides can be enzymatically exchanged with ^{18}O, thus providing the labeled peptide with a 4 Da mass shift from the ^{16}O-labeled sample. Peptide ^{18}O labeling is ideally suited for generating a labeled "universal" reference sample used for obtaining accurate and reproducible quantitative measurements across large number of samples in quantitative discovery proteomics.

Key words: LC-MS, ^{18}O labeling, quantitative proteomics, stable isotope labeling, enzymatic labelling.

Abbreviations: ABC, ammonium bicarbonate; AMT, accurate mass and time; BCA, bicinchoninic acid; DTT, DL-dithiothreitol; ESI, electrospray ionization; FTICR, Fourier transform ion cyclotron resonance; IAA, iodoacetamide; I.D., inner diameter; LC, liquid chromatography; MS, mass spectrometry; NET, normalized elution time; SPE, solid phase extraction

1. Introduction

Stable isotope labeling is commonly applied in shotgun proteomics for relative peptide/protein abundance quantitation. The incorporation of stable isotopes into peptides results in a fixed mass shift, yet does not affect the chemical properties of the peptides (e.g., peptide LC elution times or ionization efficiency),

K. Gevaert, J. Vandekerckhove (eds.), *Gel-Free Proteomics*, Methods in Molecular Biology 753,
DOI 10.1007/978-1-61779-148-2_3,

thus allowing relative abundances of labeled peptides from different samples to be accurately quantified during a single analysis using liquid chromatography coupled with mass spectrometry (LC-MS) (1, 2). In trypsin-catalyzed ^{18}O labeling, a digested protein sample is placed into $H_2{}^{18}O$ where trypsin catalyzes the exchange of either one or both of the C-terminal carboxyl group ^{16}O atoms of the peptide into ^{18}O (*see* **Fig. 3.1**), thus providing the labeled peptide with a 2 or 4 Da mass shift, respectively, from the unlabeled (i.e., ^{16}O-labeled) sample (*see* **Fig. 3.2**) (3). Similarly, other serine proteases such as Lys-C and Glu-C can also catalyze C-terminal carboxyl group oxygen exchange for ^{18}O-labeling, although these enzymes target different amino acid C-termini than trypsin (4).

Trypsin-catalyzed ^{18}O labeling is a popular approach for relative quantitation because it is simple, cost-effective, and flexible in its ability to specifically label all peptides that end with Arg or Lys residue on peptide C-terminal. Moreover, $^{16}O/^{18}O$ labeling has

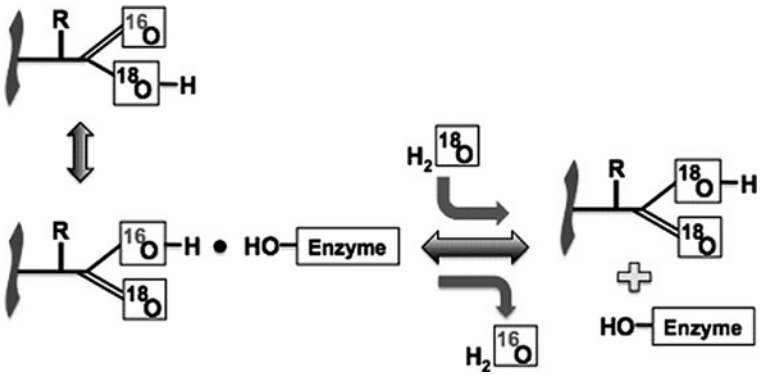


Fig. 3.1. Mechanism of enzyme-catalyzed oxygen exchange on the C-terminal carboxyl groups of peptides. Serine proteases, like trypsin, bind to the C-terminal carboxyl group and catalyze the exchange of both oxygen atoms with ^{18}O water.

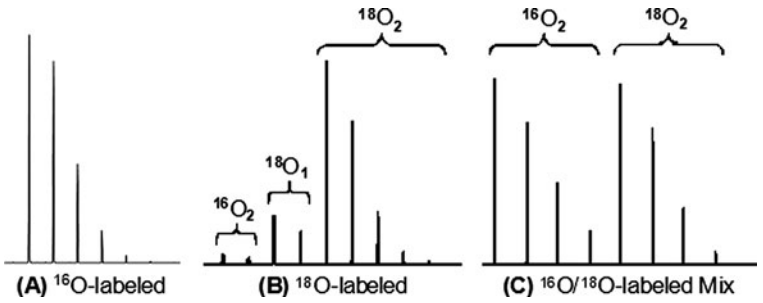


Fig. 3.2. Mass spectral patterns for $^{16}O/^{18}O$-labeled peptides. (**a**) An ^{16}O-labeled, or unlabeled, peptide. (**b**) An ^{18}O-labeled peptide. Since the ^{18}O water is not 100% pure, a percentage of the peptide will have ^{16}O incorporated into one or two C-terminal carboxyl groups. A well-labeled peptide will have minimal ^{16}O incorporation. (**c**) ^{16}O- and ^{18}O-labeled samples mixed in a 1:1 ratio (w:w).

no lower limit on the amount of peptides that can be labeled and the procedure results in nearly 100% sample recovery (5), which is important for processing small amounts of samples. However, there are two potential limitations of ^{18}O-labeling that need to be addressed in order to effectively apply such labeling for accurate quantitative measurements. The first limitation is the potentially incomplete labeling of different peptides, which has been recently addressed by de-coupling the enzymatic digestion and the ^{18}O-labeling steps. This strategy significantly improves the overall labeling efficiency where nearly all peptides can reach oxygen exchange equilibrium with optimized incubation time and enzyme concentration (5–8). The second limitation is the potential of oxygen "back-exchange" where ^{18}O is replaced with ^{16}O after labeling through a residual trypsin-catalyzed reaction when the sample is placed into a buffer containing ^{16}O water. Several approaches have been reported to minimize this back-exchange by inactivating trypsin by cysteine alkylation (8) or by using immobilized trypsin rather than solution-phase trypsin (9). More recently, we reported a simple procedure for effectively inactivating residue trypsin by boiling the labeled peptides for 10 min, followed by the addition of 5% (v/v) formic acid (5).

Stable ^{18}O-labeling is ideally suited for generating a labeled "universal" reference sample for large-scale quantitative discovery proteomics (10). As a pool across experimental samples, the "universal" reference ensures that all detected ^{16}O-labeled sample peptides will have their corresponding ^{18}O-labeled counterparts to serve as internal standards for quantitation. With an ^{18}O-labeled reference, it is unnecessary to label individual samples, thus eliminating additional sample handling and loss. The labeled reference that acts as comprehensive internal standards in every sample provides a means to obtain reproducible quantitative measurements for large-scale studies where LC-MS performance variation over time presents a significant challenge.

In this chapter, we present a detailed post-digestion trypsin-catalyzed ^{18}O-labeling protocol using solution-phase trypsin that provides effective labeling without the problem of oxygen back-exchange as well as a brief description of the quantitative data analysis strategy.

2. Materials

Unless otherwise noted, all reagents and equipment can be obtained from Sigma-Aldrich. Nanopure water is used for all experiments.

2.1. Protein Digestion

1. 50 mM ammonium bicarbonate (ABC) stock solution: dissolve 395 mg of ABC powder per 100 ml of water and adjust pH to 7.8 with HCl and/or NH_4OH. Store at 4°C for up to 6 months.
2. Urea in solid form.
3. 500 mM DL-dithiothreitol (DTT): prepare fresh DTT solution by dissolving 77 mg of DTT powder in 1 ml of water.
4. 400 mM iodoacetamide (IAA): prepare fresh IAA solution by dissolving 74 mg IAA powder in 1 ml of water. Note that iodoacetamide is light sensitive and should thus be kept in the dark!
5. 1 M calcium chloride: dissolve 147 mg of calcium chloride dihydrate powder in 1 ml of water.
6. 1 μg/μl trypsin: dissolve 20 μg of sequencing-grade porcine trypsin (Promega) in 20 μl of 50 mM ABC. Optional: incubate the re-dissolved trypsin at 37°C for 10 min before adding the trypsin to the sample.
7. 1.5 and 2.0 ml sterile, siliconized microcentrifuge tubes from Thermo Fisher Scientific, Inc.
8. 15 and/or 50 ml Falcon tubes (BD Biosciences).
9. Thermal mixer (Thermomixer R; Eppendorf) or oven that can be set to 37°C.
10. Bicinchoninic acid (BCA) protein assay kit (Thermo Scientific).
11. Absorbance reader (562 nm).

2.2. C18 Solid Phase Extraction (SPE) Cleanup Following Protein Digestion

1. 100% methanol.
2. 0.1% TFA (v/v) in water.
3. 5% (v/v) acetonitrile and 0.1% (v/v) TFA in water.
4. 80% (v/v) acetonitrile and 0.1% (v/v) TFA in water.
5. C18 SPE column of appropriate bed weight (i.e., a 1 ml column with a 100 mg bed weight can clean up to 5 mg of digested protein).
6. Vacuum manifold.
7. Speed-Vac® vacuum concentrator (Thermo Fisher Scientific, Inc.).

2.3. ^{18}O-Labeling

1. 97% $H_2{}^{18}O$ (*see* **Note 1**).
2. 1 M ammonium bicarbonate (ABC) buffer in ^{18}O water: dissolve 79.06 mg of ABC in ^{18}O water. This buffer can be stored at 4°C for up to 6 months in a tube that is wrapped in a parafilm and placed into a larger container with desiccant.

3. 50 mM ammonium bicarbonate (ABC) buffer in ^{18}O water: dilute the 1 M ABC buffer 20-fold with ^{18}O water (storage conditions, see above).
4. 1 M calcium chloride (*see* step 5 of **Section 2.1**).
5. 1 μg/μl trypsin: dissolve 1 volume of 20 μg of sequencing-grade porcine trypsin (Promega, Madison, WI) in 20 μl of 50 mM ABC dissolved in ^{18}O water.
6. 1 M formic acid in water.
7. Liquid nitrogen.

2.4. LC-MS Analysis

1. 65 cm long, 150 μm I.D. fused silica reversed-phase capillary column packed with 3 μm Jupiter C18-bonded particles.
2. Mobile phase A: 0.2% v/v acetic acid and 0.05% v/v trifluoroacetic acid in water.
3. Mobile phase B: 90% v/v acetonitrile and 0.1% v/v trifluoroacetic acid in water.

3. Methods

$^{16}O/^{18}O$-labeling can be applied for both pair-wise comparative quantitation and relative quantitation across many biological samples by employing an ^{18}O-labeled "universal" reference sample (10). Sample processing typically starts with cell lysis, protein digestion, followed by post-digestion $^{16}O/^{18}O$-labeling that utilizes solution-phase trypsin (*see* **Fig. 3.3**) that is later inactivated by boiling (5). Labeled samples with a mixture of $^{16}O/^{18}O$-labeled peptides can be analyzed by LC-MS and LC-MS/MS with a high-resolution mass spectrometer. For pair-wise labeling, both biological samples will be processed using identical protocols with the exception that ^{16}O-labeling uses ^{16}O water while ^{18}O-labeling uses ^{18}O water. For applying the ^{18}O-labeled "universal" reference strategy, only one pooled sample needs to be labeled with ^{18}O, while all biological samples can be processed without an additional labeling step; however, residual trypsin from protein digestion in each unlabeled sample must be de-activated by boiling the peptide sample for 10 min prior to mixing it with the ^{18}O-labeled reference to prevent ^{18}O back-exchange. LC-MS data of the labeled samples are analyzed by a software algorithm that performs de-isotoping, feature and pair finding, and relative quantitative information based on $^{16}O/^{18}O$-peptide pair abundance ratios (*see* **Fig. 3.4**).

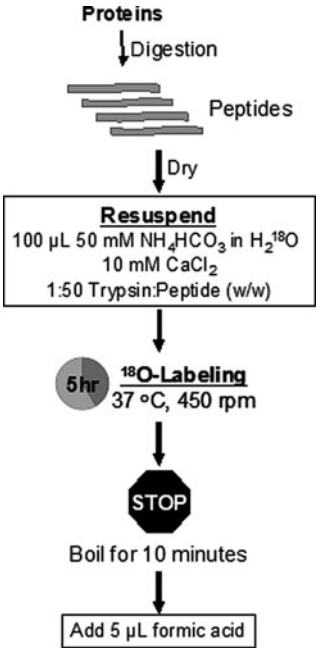


Fig. 3.3. A workflow for post-digestion ^{18}O labeling.

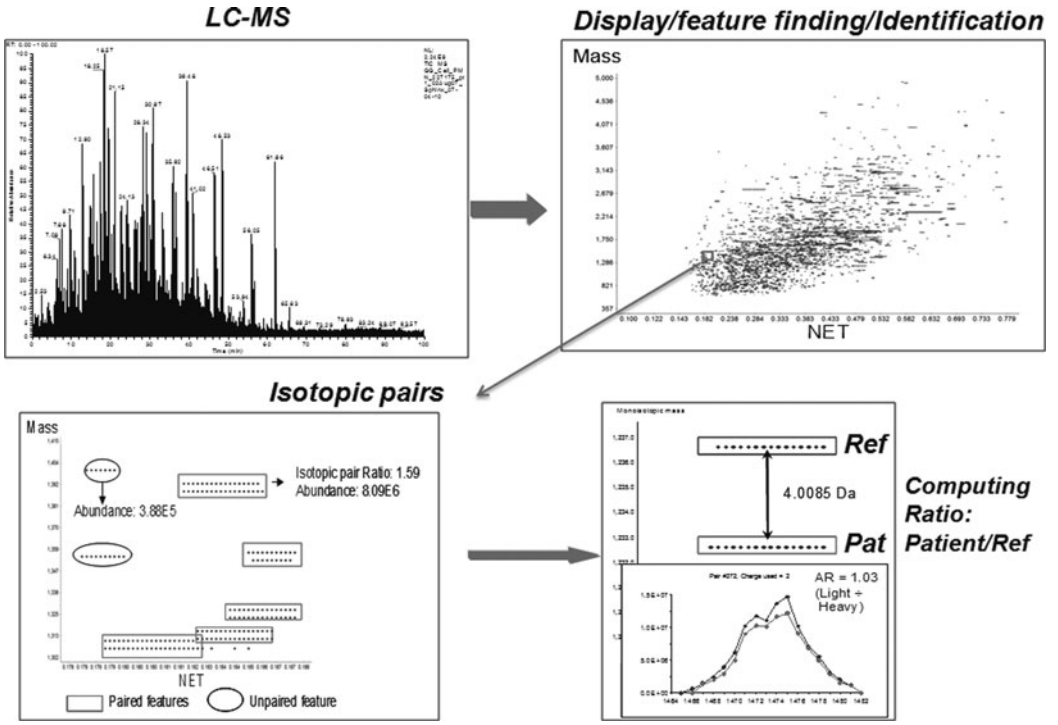


Fig. 3.4. An overview of LC-MS quantitative data analysis process. Each LC-MS data set was displayed in mass and normalized elution time (NET) dimensions for feature and pair finding and peptide identification. $^{16}O/^{18}O$ abundance ratio is calculated based on the light and heavy versions of peptide elution profiles for a given peptide.

3.1. Protein Digestion

1. Perform a BCA assay on the sample to determine protein concentration and total protein mass.
2. Dilute sample, if necessary, with 50 mM ABC so that the protein concentration does not exceed 10 mg/ml.
3. Add an appropriate amount of urea to obtain a final concentration of 8 M urea (i.e., 780 mg of urea per 1 ml of sample) to denature the proteins (*see* **Note 2**). Vortex briefly (i.e., ~1–3 s intervals) to mix thoroughly and spin down briefly to bring the sample and reagents to the bottom of the tube. Incubate at 37°C for 1 h (shaking at 450 rpm if possible), and following incubation, spin down briefly to remove condensation from the top of the lid.
4. To reduce the proteins, add DTT stock solution to the sample to obtain a final concentration of 10 mM DTT (i.e., 20 μl of 500 mM DTT per 1 ml of sample). Vortex briefly to mix thoroughly and spin the sample briefly down. Incubate at 37°C for 1 h (shaking at 450 rpm if possible) and spin down briefly to remove all condensation from the top of the lid.
5. To alkylate the proteins (optional), add IAA stock solution to the sample to obtain a final concentration of 40 mM IAA (i.e., 111 μl of 400 mM IAA per 1 ml of sample). Vortex, spin the sample down, and incubate for 37°C for 1 h in the dark. Following incubation, spin down briefly to remove condensation from the top of the lid.
6. For trypsin digestion, dilute the sample 10-fold with 50 mM ABC to reduce the salt concentration. Add calcium chloride to obtain a final concentration of 1 mM calcium chloride (i.e., 1 μl of 1 M calcium chloride per 1 ml of sample). Add sequencing-grade porcine trypsin to the sample in a 1:50 trypsin:protein (w:w) ratio. Vortex briefly, spin down the sample briefly, and incubate at 37°C for 3–6 h, with shaking at 450 rpm if possible. After incubation, spin down briefly to remove condensation from the top of the lid (*see* **Note 3**).

3.2. C18 SPE Cleanup of the Protein Digest

1. Place the C18 SPE cartridge on the vacuum manifold, turn on the vacuum.
2. Condition the column by adding 3 ml of 100% methanol and letting the methanol flow no faster than one drop per second. Make sure that the column does not go dry before adding the next solution.
3. Equilibrate the column by adding 2 ml of 0.1% trifluoroacetic acid (TFA), letting the solution flow no faster than one drop per second.

4. Apply the sample, decreasing the flow to as slow as possible (i.e., approximately one drop every 3 s or even longer).
5. Wash the column by adding 4 ml of 5% (v/v) acetonitrile and 0.1% (v/v) TFA in water, letting the solution flow no faster than one drop each second.
6. Elute the sample slowly (i.e., one drop every 3 s or longer), collecting into a sterile and siliconized 1.5 ml microcentrifuge tube, with 1 ml of 80% (v/v) acetonitrile, 0.1% (v/v) TFA in water.
7. Discard the C18 SPE cartridge.
8. Concentrate sample in a Speed-Vac® vacuum concentrator down to ~50–100 μl to remove most of the acetonitrile.
9. Perform a BCA assay to determine the peptide concentration (*see* **Note 4**).

3.3. Post-digestion ^{18}O-Labeling (see Note 5)

1. Completely dry the peptide sample for ^{18}O labeling in a Speed-Vac® vacuum concentrator (*see* **Note 6**).
2. Perform ^{18}O labeling (*see* **Note 7**) by adding 100 μl of 50 mM ABC buffer in ^{18}O water to the dried peptide sample. Mix the sample by brief sonication (i.e., ~10 s) and vortex briefly (i.e., 1–3 s) (*see* **Note 8**) followed by a brief spin down to collect the sample at the bottom of the tube. Add 1.0 μl of 1 M calcium chloride to the sample for a final concentration of 10 mM calcium chloride (*see* **Note 9**). Add solution-phase porcine trypsin in a 1:50 trypsin:peptide (w:w) ratio to the sample. Wrap the tube lid in parafilm, vortex briefly, and spin down briefly to collect the sample at the bottom of the tube. Incubate at 37°C for 5 h (shaking at 450 rpm (*see* **Note 10**)).
3. Stop the ^{18}O-labeling reaction by placing the tubes into a floating tube rack into the boiling water (*see* **Note 11**) and boil the sample for 10 min (*see* **Note 12**). After boiling, immediately snap-freeze the sample in liquid nitrogen, and once the sample has thawed, immediately add 5 μl of formic acid (*see* **Notes 13** and **14**). Vortex briefly to mix and spin the sample down briefly (*see* **Note 3**).
4. To prepare samples for MS analysis, perform a BCA assay to determine the peptide concentration (*see* **Note 15**). (a) Verify the labeling efficiency of the ^{18}O-labeled sample to ensure accurate quantitation by analyzing the labeled sample by LC-MS before mixing the ^{18}O- and ^{16}O-labeled, or unlabeled, samples together (*see* **Fig. 3.2b**) (*see* **Note 16**). (b) Once high labeling efficiency of the ^{18}O-labeled sample has been verified, boil all of the unlabeled samples for 10 min. (c) Then, mix the ^{16}O- and ^{18}O-labeled samples in a 1:1 (w:w) ratio for LC-MS analysis.

3.4. Capillary LC-MS Analyses

In our laboratory, peptide samples are analyzed using a custom-built capillary LC system (11) coupled online to an LTQ-Orbitrap ion trap mass spectrometer equipped with an ESI interface:

1. Degas mobile phases using a vacuum degasser.
2. Inject 10 μl of peptide sample onto the reversed-phase capillary column and apply the following solvent gradient: first 100% of solvent A for 20 min, followed by increasing the solvent composition to 70% of solvent B over 100 min, using a stainless steel mixing chamber (*see* **Note 17**). This same gradient is applied for both LC-MS/MS and LC-FTICR analyses.
3. The mass spectrometers are operated under standard conditions, details of which are available in references (10, 11).

3.5. Quantitative LC-MS Data Analysis

In our laboratory, LC-MS data analysis and processing steps are automated using an in-house developed software package that includes the informatics tools Decon2LS and VIPER (12, 13), which can be downloaded at ncrr.pnl.gov. The details of data analysis are available in references (7, 10). The process is briefly described as follows:

1. The initial analysis of raw LC-MS data involves a mass transformation or de-isotoping step using Decon2LS, an analysis tool based on the THRASH algorithm, which generates a text file report for each LC-MS data set including both the monoisotopic masses and the corresponding intensities for all detected species for each mass spectrum.
2. Following Decon2LS analysis, each data set is processed by the feature matching tool (VIPER) for peptide identification and quantification. Data sets are displayed in a two-dimensional mass and normalized LC elution time (NET) format. The feature matching process includes the "distinct feature" (i.e., a peak with unique mass and elution time) finding that searches for $^{16}O/^{18}O$ feature pairs and computes abundance ratios for feature pairs. An intensity report for all detected features is created, which normalizes LC elution times via alignment to a database that allows the peptides to be identified.
3. Feature identification is performed by matching the accurately measured masses and normalized elution time (NET) values of each detected feature to a pre-established accurate mass and time (AMT) tag database from prior LC-MS/MS analyses.
4. For each identified peptide, the relative $^{16}O/^{18}O$ abundance ratio is reported if the feature is paired. The details for computing the $^{16}O/^{18}O$ abundance ratios using VIPER have been previously described (7).

4. Notes

1. The purity of ^{18}O water is directly related to the labeling efficiency because oxygen exchange is dependent upon the concentration of the available isotopes. In other words, the higher the purity of the ^{18}O water, the higher the ^{18}O-labeling efficiency.
2. Each milligram of urea adds an additional ~0.8 μl to the sample volume.
3. At this point, the sample can be frozen.
4. Sample recovery after urea digestion and C18 SPE is ~50%, by weight (i.e., final peptide weight versus starting protein amount).
5. The procedure can also be applied toward ^{16}O labeling using $H_2{}^{16}O$ for comparative analyses. ^{16}O- and ^{18}O-labeled samples must be kept separate until the boiling step is done (step 4c of **Section 3.3**).
6. As much as 1 mg of digested protein in 100 μl of $H_2{}^{18}O$ can be successfully labeled.
7. Peptide samples should be cleaned using SPE or a method that can effectively remove detergents and contaminants such as SDS that may interfere with labeling.
8. For small sample amounts, sample mixing should be performed via sonication rather than vortexing to improve sample recovery.
9. 1 M of calcium chloride is dissolved in deionized ^{16}O water, but could be made in ^{18}O water. The small amount of ^{16}O from the calcium chloride solution has a minimal effect on labeling efficiency when compared to the amount of ^{18}O in the final labeling mixture.
10. Labeling for longer than 5 h will not hurt the labeling efficiency but it does not appear to improve it either.
11. Microcentrifuge tube lids can pop open during boiling if the ratio of the sample volume to tube volume is high enough. For this protocol, a 100 μl sample volume will not cause the lid of a 1.5 ml tube to pop open during the 10 min boil. However, 100 μl sample volume in a 0.6 ml tube will usually cause the lid to pop open during boiling, which may result in sample loss and/or contamination. To minimize this from happening, position the tube in the floating rack so that the sample is covered with boiling water, but not pushed to the maximum depth.

12. The boiling step to inactivate the residue trypsin is the most critical step in this protocol; otherwise, back-exchange will be a significant issue.
13. 5% formic acid (v/v) added to the sample was observed to be beneficial for achieving high labeling efficiency.
14. It is essential that the formic acid is added as soon after thawing as possible to ensure good labeling efficiency. Rub the tube back and forth between the hands to accelerate the thawing. Once partially thawed, add the formic acid. Then, vortex in short bursts (i.e., 1–3 s) to mix the formic acid with the sample until the sample is completely thawed.
15. Sample recovery after labeling is ~100% by weight since the labeling protocol does not introduce sample loss.
16. ^{18}O-Labeling efficiency can be determined by visually evaluating a handful of chromatograms representing ^{18}O-labeled tryptically digested peptides. Although labeling may vary slightly from peptide to peptide, a well-labeled sample will have a narrow range of high labeling efficiency (i.e., minimal incorporation of ^{16}O) across all tryptically digested peptides and will be very similar to that shown in **Fig. 3.2b**.
17. The ^{16}O/^{18}O-labeling strategy is compatible with any type of mobile phase.

Acknowledgments

The authors thank NIH National Center for Research Resources (RR018522) for support of portions of this research and the Environmental Molecular Sciences Laboratory (EMSL) at Pacific Northwest National Laboratory (PNNL), a national scientific user facility sponsored by the US Department of Energy's (DOE) Office of Biological and Environmental Research, for the development and use of the instrumentation applied in this research. PNNL is a multi-program national laboratory operated by Battelle Memorial Institute for the DOE under Contract DE-AC05-76RL01830.

References

1. Goshe, M. B., and Smith, R. D. (2003) Stable isotope-coded proteomic mass spectrometry, *Curr Opin Biotechnol 14*, 101–109.
2. Qian, W. J., Camp, D. G., and Smith, R. D. (2004) High throughput proteomics using Fourier Transform Ion Cyclotron Resonance (FTICR) mass spectrometry, *Expert Rev Proteomics 1*(1), 89–97.
3. Yao, X., Freas, A., Ramirez, J., Demirev, P. A., and Fenselau, C. (2001) Proteolytic 18O labeling for comparative proteomics: model

studies with two serotypes of adenovirus, *Anal Chem 73*, 2836–2842.
4. Fenselau, C., and Yao, X. (2009) 18O2-labeling in quantitative proteomic strategies: a status report, *J Proteome Res 8*, 2140–2143.
5. Petritis, B. O., Qian, W. J., Camp, D. G., 2nd, and Smith, R. D. (2009) A simple procedure for effective quenching of trypsin activity and prevention of 18O-labeling back-exchange, *J Proteome Res 8*, 2157–2163.
6. Heller, M., Mattou, H., Menzel, C., and Yao, X. (2003) Trypsin catalyzed 16O-to-18O exchange for comparative proteomics: tandem mass spectrometry comparison using MALDI-TOF, ESI-QTOF, and ESI-ion trap mass spectrometers, *J Am Soc Mass Spectrom 14*, 704–718.
7. Qian, W. J., Monroe, M. E., Liu, T., Jacobs, J. M., Anderson, G. A., Shen, Y., Moore, R. J., Anderson, D. J., Zhang, R., Calvano, S. E., Lowry, S. F., Xiao, W., Moldawer, L. L., Davis, R. W., Tompkins, R. G., Camp, D. G., 2nd, and Smith, R. D. (2005) Quantitative proteome analysis of human plasma following in vivo lipopolysaccharide administration using 16O/18O labeling and the accurate mass and time tag approach, *Mol Cell Proteomics 4*, 700–709.
8. Staes, A., Demol, H., Van Damme, J., Martens, L., Vandekerckhove, J., and Gevaert, K. (2004) Global differential non-gel proteomics by quantitative and stable labeling of tryptic peptides with oxygen-18, *J Proteome Res 3*, 786–791.
9. Sevinsky, J. R., Brown, K. J., Cargile, B. J., Bundy, J. L., and Stephenson, J. L., Jr. (2007) Minimizing back exchange in 18O/16O quantitative proteomics experiments by incorporation of immobilized trypsin into the initial digestion step, *Anal Chem 79*, 2158–2162.
10. Qian, W. J., Liu, T., Petyuk, V. A., Gritsenko, M. A., Petritis, B. O., Polpitiya, A. D., Kaushal, A., Xiao, W., Finnerty, C. C., Jeschke, M. G., Jaitly, N., Monroe, M. E., Moore, R. J., Moldawer, L. L., Davis, R. W., Tompkins, R. G., Herndon, D. N., Camp, D. G., and Smith, R. D. (2009) Large-scale multiplexed quantitative discovery proteomics enabled by the use of an (18)O-labeled "Universal" reference sample, *J Proteome Res 8*, 290–299.
11. Livesay, E. A., Tang, K., Taylor, B. K., Buschbach, M. A., Hopkins, D. F., LaMarche, B. L., Zhao, R., Shen, Y., Orton, D. J., Moore, R. J., Kelly, R. T., Udseth, H. R., and Smith, R. D. (2008) Fully automated four-column capillary LC-MS system for maximizing throughput in proteomic analyses, *Anal Chem 80*, 294–302.
12. Zimmer, J. S., Monroe, M. E., Qian, W. J., and Smith, R. D. (2006) Advances in proteomics data analysis and display using an accurate mass and time tag approach, *Mass Spectrom Rev 25*, 450–482.
13. Monroe, M. E., Tolic, N., Jaitly, N., Shaw, J. L., Adkins, J. N., and Smith, R. D. (2007) VIPER: an advanced software package to support high-throughput LC-MS peptide identification, *Bioinformatics (Oxford, England) 23*, 2021–2023.

Chapter 4

ICPL Labeling Strategies for Proteome Research

Friedrich Lottspeich and Josef Kellermann

Abstract

Stable isotope labeling in combination with mass spectrometry has emerged as a powerful tool to identify and quantify thousands of proteins within complex protein mixtures. Isotope-coded protein label (ICPL) is capable of high-throughput quantitative proteome profiling on a global scale. Since ICPL is based on stable isotope tagging at the free amino groups of intact proteins, it is applicable to any protein sample, including extracts from tissues or body fluids. After labeling of up to four different proteome states, the samples can be combined and the complexity reduced by any separation method currently employed in protein chemistry. After enzymatic cleavage of the protein fractions the ratios of peptides in the different proteome states can be calculated by simple MS-based mass spectrometric analyses. Only peptides that exhibit regulations in the different proteome states are further investigated for identification by tandem-mass spectrometry. The quantification of multiplexed ICPL experiments is greatly facilitated by the recently published ICPL*Quant* software, which was especially designed to cover the whole ICPL workflow. The method shows highly accurate and reproducible quantification of proteins, yields high sequence coverage, and is indispensable for the comprehensive detection of posttranslational modifications and protein isoforms.

Key words: ICPL, isotope-coded protein label, ICPL*Quant*, quantitative proteomics, post-metabolic labeling.

1. Introduction

The successful evolution of high-throughput mass spectrometry of peptides has strongly influenced the proteomic strategies. Tailored instrumentation and complex software packages were all optimized to facilitate rapid and sensitive peptide identification and peptide quantification. Therefore, it sounds logical that bottom-up (peptide-based) proteomics becomes the preferred strategy. Since for comparison of different proteome stages,

K. Gevaert, J. Vandekerckhove (eds.), *Gel-Free Proteomics*, Methods in Molecular Biology 753,
DOI 10.1007/978-1-61779-148-2_4,

reproducible sample preparation, cleavage, and peptide fractionation were difficult to achieve, stable isotope labeling using, e.g., iTRAQ (1) or similar isotopic reagents (2) was developed and was rated as the most accurate quantitative technique for general proteomic applications. However, due to the enzymatic cleavage of the complete proteome, the complexity of the sample is significantly increased and – even more severe – the information from which of the many possibly existing protein species a peptide is liberated is destroyed. A single gene will produce almost always an unpredictable multiplicity of different protein species by posttranslational modifications and processing events. Thus, protein isoforms, processed forms of proteins, and posttranslational modifications are hard to recognize, to annotate, and to quantify by bottom-up proteomics.

Consequently, in recent years, top-down (intact protein-based) proteomic strategies gain more attraction. However, efficient fractionation of proteins is much more demanding compared to peptide fractionation. More complex multistep workflows are inevitably connected with unpredictable protein loss. The yields of the individual proteins are strongly dependent on the actual composition of the protein mixture and the separation technique used. Therefore, especially when significant reduction of complexity on the protein level is mandatory, it is questionable if label free strategies can deliver valid quantitative results. Therefore, proteins of different proteomic states should be compared after labeling with different stable isotope reagents. Those reagents behave chemically identical but introduce a characteristic mass difference into all proteins of a certain proteomic state. From now on proteins of the different proteomic states are traceably marked, samples from different proteomic states can be combined (multiplexed), and analytical procedures to fractionate proteins, thereby reducing the complexity, can be performed. Isotopic markers may even be introduced already in living cells (e.g., SILAC) (3). Cells from different states, differentially labeled, can then be mixed already before cell lysis and subsequent steps of fractionation and purification do not affect the accuracy of relative quantification. SILAC became one of the most widely used strategies in quantitative proteomics. However, metabolic labeling is not always possible. With human tissues and body fluids, the earliest feasible time point to introduce isotopic labels is on the level of intact proteins. Several reagents (e.g., ICAT) were developed for protein labeling (4). However, due to several limitations none of them has become very popular.

Recently, the isotope-coded protein label, ICPL, was specially designed for protein labeling (5). A kit to compare four different proteomic samples is commercially available (SERVA, Bruker). The main advantages are as follows: complete reaction

yield, the possibility of multiplexing four samples, quantification on the mass spectrometry (MS) level, and the availability of a software, ICPL*Quant*, that was specially designed to support the ICPL workflow (6). The complete ICPL workflow is shown in **Fig. 4.1**. Experimentally, four protein mixtures obtained from four distinct cell states, tissues, or body fluids are individually reduced and alkylated to denature the proteins and to ensure easier access to free amino groups. Subsequently, proteins are labeled each with one of the four ICPL reagents ($ICPL_0$, $ICPL_4$, $ICPL_6$, $ICPL_{10}$). After combining all mixtures, any separation method can be adopted to reduce the complexity of the sample on the protein level and, after digestion, on the peptide level. Quantification and identification is done by high-throughput MS. Since peptides with identical amino acid sequence derived from the four differentially labeled protein samples differ in mass, they appear as quadruplets in the acquired MS spectra (**Fig. 4.2**). From the ratios of the ion intensities of these sister peptide pairs, the relative abundance of their parent proteins in the original samples can be determined. As an integral part of the ICPL strategy, the use of a reference sample is highly recommended. The reference sample is prepared by combining an equal aliquot of all four proteome samples to be analyzed. This combined sample is then split into four equal portions which are treated like the experimental samples, i.e., reduction, alkylation, ICPL labeling, protein

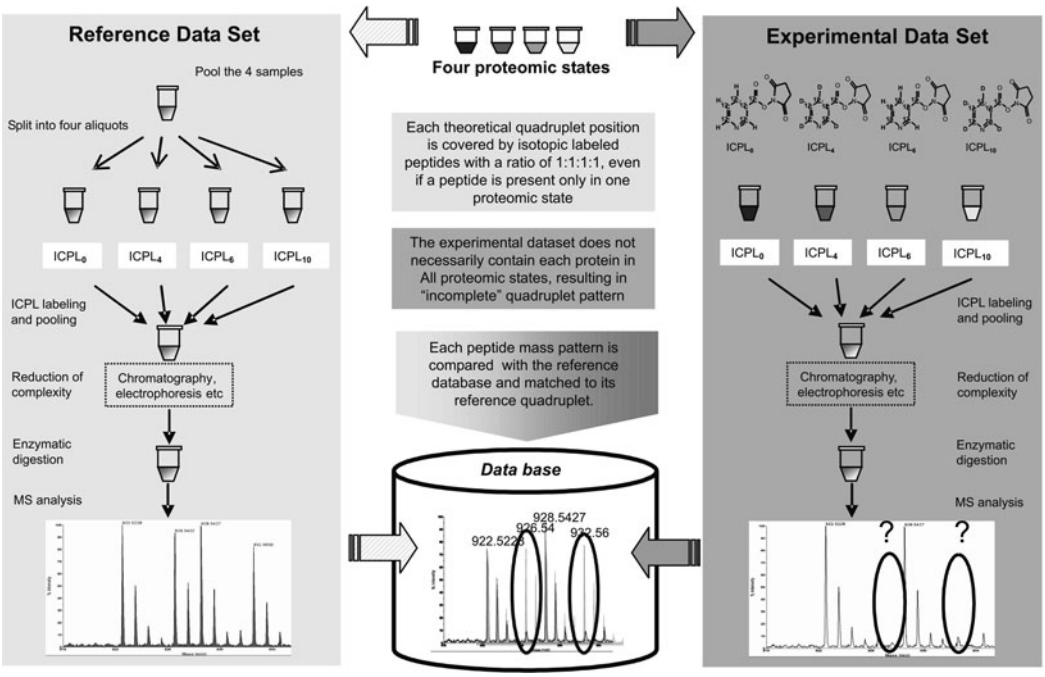


Fig. 4.1. ICPL workflow of four proteomic states with reference sample and experimental sample.

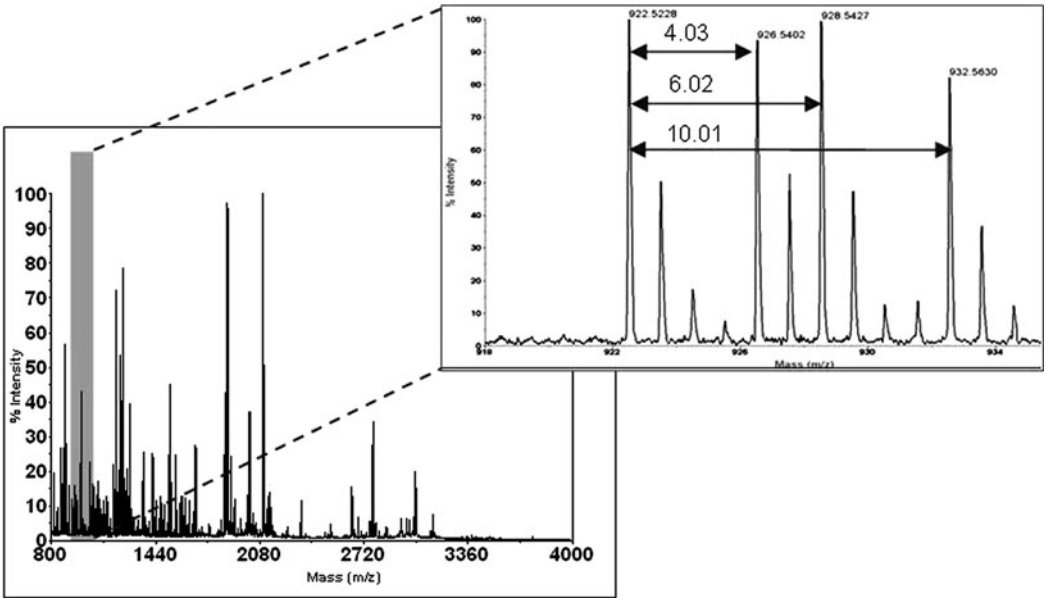


Fig. 4.2. Mass spectrum of an enzymatic digest of an ICPL-labeled protein fraction. The *insert* shows a zoom into the mass region of a quadruplet corresponding to an ICPL-derivatized peptide present in the same amount in all four proteomic states.

fractionation, and analysis. After enzymatic cleavage of the protein factions and mass spectrometry, each lysine-containing peptide will appear as an equally intense quadruplet. These quadruplets can be stored in a database and, by comparison with the data of the experimental samples, enable also the identification of incomplete quadruplets (i.e., existence of a peptide in only one or two out of the four experimental proteomic states). The MS data of the experimental sample and/or the reference sample should be submitted to ICPL*Quant*, a software especially designed for the ICPL workflow. This enables a rapid and highly automated quantification of the complex ICPL proteomic experiments.

2. Materials

2.1. Sample Preparation

1. Lysis buffer: 6 M guanidinium hydrochloride, 0.1 M HEPES, pH 8.5.
2. Protein concentrations of the samples are determined either by Bradford (BioRad Laboratories) or the ProteoQuant method (Serva Electrophoresis).

2.2. Reduction and Alkylation of Cysteine Residues (Carbamidomethylation)

1. Reduction solution: 0.2 M TCEP and 0.1 M HEPES, pH 8.5.
2. Alkylation solution: 0.4 M iodoacetamide and 0.1 M HEPES, pH 8.5. Prepare fresh every time.
3. Stop solution 1: 0.5 M *N*-acetyl-cysteine and 0.1 M HEPES, pH 8.5.

2.3. Isotope Labeling of the Protein Samples

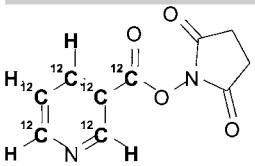
1. A 0.15 M solution of each derivative of the *N*-hydroxysuccinimide ester (Serva Electrophoresis) in dimethylsulfoxide (DMSO) is prepared:
   ^{12}C-*N*-hydroxysuccinimide ester ($ICPL_0$)
   $^{12}C^2H_4$-*N*-hydroxysuccinimide ester ($ICPL_4$)
   $^{13}C_6$-*N*-hdroxysuccinimide ester ($ICPL_6$)
   $^{13}C_6{}^2H_4$-*N*-hydroxysuccinimide ester ($ICPL_{10}$)
   (**Table 4.1**)
2. Stop solution 2: 1.5 M hydroxylamine-HCl, pH 8.5.
3. 2 N NaOH.
4. 2 N HCl.

2.4. Acetone Precipitation of the Labeled Proteins

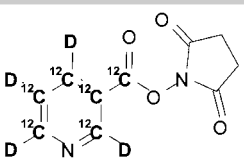
1. Acetone, cooled to –20°C.
2. 80% acetone, 20% water, cooled to –20°C.

Table 4.1
Structures of the different ICPL derivatives and masses of ICPL-labeled residues according to Unimod

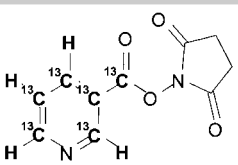
		Residues Lys		Protein Nterm	
		Monoisotopic	Average	Monoisotopic	Average
$ICPL_0$	ICPL:	233.116427	233.2664	106.029289	106.1020
$ICPL_4$	ICPL:2H(4)	237.141537	237.2911	110.054396	110.1267
$ICPL_6$	ICPL:13C(6)	239.136556	239.2223	112.049418	112.0579
$ICPL_{10}$	ICPL:13C(6)2H(4)	243.161663	243.2470	116.074525	116.0826



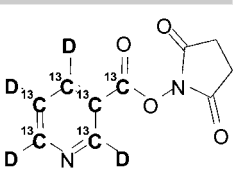
$ICPL_0$



$ICPL_4$



$ICPL_6$



$ICPL_{10}$

2.5. Enzymatic Digestion of the Labeled Proteins

1. Buffer for enzymatic cleavage: 4 M urea, 100 mM Tris-HCl, pH 8.0.
2. Trypsin (Roche Diagnostics).
3. Endoproteinase Glu-C (Roche Diagnostics).

2.6. ICPLQuant

The latest software version of ICPL*Quant* including a detailed tutorial can be downloaded from http://www.biochem.mpg.de/lottspeich/technologies/ICPLQuant/index.html

3. Methods

3.1. Experimental Sample Preparation for Four Proteomic Samples

Note that the experimental sample preparation is identical for all four samples:

1. Dissolve 100 μg of protein in 20 μl of lysis buffer (*see* **Note 1**) and vortex for 2 min.
2. Incubate sample with gentle agitation for 20 min at 25°C.
3. Vortex for 2 min and sonicate four times for 30 s each in an ultrasound bath (cool sample during pauses in ice water for 2 min).
4. Vortex for 1 min and incubate sample with gentle agitation for 15 min at 25°C.
5. Vortex for 2 min and spin sample for 30 min at 100,000×*g*.
6. Use the supernatant directly for a protein concentration assay and further analysis.
7. Adjust the protein concentration to 5 mg/ml with lysis buffer before proceeding with the labeling protocol (*see* **Note 1**).

3.2. Reduction and Alkylation of Cysteine Residues (Carbamidomethylation)

This carbamidomethylation protocol is identical for all four samples:

1. Check the pH of the sample buffer and if necessary adjust to 8.5 ± 0.1 by addition of 2 N HCl or 2 N NaOH. A micro pH electrode is highly recommended. Before every measurement, rinse the tip of the electrode with distilled water and dry very carefully with a dry, dust-free tissue to avoid sample dilution.
2. Add 0.5 μl of reduction solution to 20 μl of sample solution (equivalent to 100 μg of protein) and reduce proteins for 30 min at 60°C.
3. Cool sample to room temperature and spin down condensed solution from the lid.

4. Add 0.5 µl of the freshly prepared alkylation reagent to each sample, wrap samples quickly in aluminum foil for light protection, and leave samples for 30 min at 25°C.
5. Stop the alkylation reaction by adding 0.5 µl of stop solution 1 to each sample and incubate for 15 min at 25°C.

3.3. Isotope Labeling of the Protein Samples

1. Add 3 µl of the ^{12}C-Nic-reagent solution ($ICPL_0$) to the sample containing the first proteome state. Add 3 µl of the $^{12}C^2H_4$-Nic-reagent solution ($ICPL_4$) to the sample containing the second proteome state. Add 3 µl of the $^{13}C_6$-Nic-reagent solution ($ICPL_6$) to the sample containing the third proteome state. Add 3 µl of the $^{13}C_6{}^2H_4$-Nic-reagent solution ($ICPL_{10}$) to the sample containing the fourth proteome state.
2. Overlay all samples with argon (or equivalent) to exclude oxidation, vortex for 10 s, and sonicate for 1 min in ultrasound bath. Spin down samples.
3. Incubate samples for 2 h at 25°C.
4. Add 2 µl of stop solution 2 to each sample and shake for 20 min at 25°C to destroy excess reagent.
5. Combine all ICPL-labeled samples and vortex thoroughly.
6. Adjust the pH of the mixture to 11.9 ± 0.1 by adding 2 N NaOH (about 4 µl to 4 × 20 µl sample volume) to revert possible esterification products. After 20 min add the same amount of 2 N HCl to neutralize the sample (usually it is not necessary to check the pH).

3.4. Purification of the Labeled Proteins by Acetone Precipitation

1. Add an equal amount of distilled water to the sample.
2. Add fivefold excess (related to the total volume of the sample) of ice-cold acetone and leave sample at –20°C overnight.
3. Spin down precipitated proteins at 100,000×*g* for 30 min at 4°C.
4. Discard supernatant.
5. Overlay precipitated proteins with about 100–200 µl of ice-cold 80% acetone, shake carefully, and spin down again at 100,000×*g* for 5 min at 4°C.
6. Discard supernatant and let the remaining acetone evaporate at room temperature leaving the lid open.
7. The samples can be stored now at –80°C or can be directly dissolved in appropriate buffers for protein separation (1D, 2DE, free flow electrophoresis, or chromatography, *see* **Note 2**).

3.5. Enzymatic Digestion of the Labeled Proteins for Direct MS Analysis

The enzymatic digestion of the labeled sample is done according to common protocols. We recommend using trypsin, endoproteinase Glu-C, or a combination of both enzymes:

1. Dissolve the sample in 20 µl of tryptic cleavage buffer or endoproteinase Glu-C cleavage buffer (*see* **Note 3**). Dilute the urea concentrations recommended in the enzyme data sheets.
2. Add enzyme in a protein/enzyme ratio of 50:1 for trypsin or 30:1 for endoproteinase Glu-C. Incubate the sample for 4 h at 37°C.
3. After digestion, the samples can be directly analyzed by mass spectrometry, preferentially by LC-MALDI-TOF-/TOF or LC-ESI-MS-/MS.

3.6. Reference Sample Preparation for Four Proteomic Samples

For each of the four samples:

1. Each of the four samples is treated in the same way as described in steps 1–6 of **Section 3.1** for the experimental samples.
2. Combine the four samples.
3. Split this combined sample into four equal parts and treat these four samples as described for the experimental samples in **Section 3.2**.

3.7. Mass Spectrometry and Data Analysis by ICPL Quant

It is recommended to process the mass spectrometry data (MALDI or ESI) using the ICPL*Quant* software. The latest version of the software can be downloaded for free from http://www.biochem.mpg.de/lottspeich/technologies/ICPLQuant/index.html:

1. Follow the video tutorial with detailed explanation of the software.
2. After the modification of the lysine residues by ICPL, lysine is protected against proteolytic digestion. Trypsin therefore only cleaves C-terminal to arginine. For this reason database searches should be done using endoproteinase Arg C as enzyme entry (endoproteinase Lys C cannot be used at all!).
3. For Mascot searches, the ICPL-modified residues have to be added to the modification file (mod_file) on the local Mascot server (*see* **Note 4**).

4. Notes

1. The protocol is optimized for a protein concentration of 5 mg/ml. However, it works as well with protein concentrations of 2.5 mg/ml. As the recovery rate of the protein

precipitation step below depends strongly on the total protein concentration, losses are likely when working with lower protein concentrations. Therefore, it is extremely important to keep the concentrations of the reagents strictly as recommended. If you want to work with increased sample volumes of 40 μl (for example, to facilitate the pH measurement), you rather have to double the sample amount and also have to double volumes of the reagents given in this protocol!

2. The further protocol depends on the complexity of the used sample. Complex proteome samples should be separated to a convenient complexity by any protein fractionation method (1DE, 2DE, free flow electrophoresis, chromatography, or any combination of these techniques) before MS analysis. In combination with 2D electrophoresis, the ICPL technology provides some improvements over current 2DE/MS technologies. The multiplexing of several proteome states allows for the simultaneous separation of differentially labeled protein samples in the same gel. Therefore, problems of electrophoretic variations between gels are avoided and protein quantification is more accurate and confident. The modification of the basic amino groups with Nic-NHS changes the migration behavior of labeled basic proteins during 2DE toward the more acidic side, making extreme basic proteins more accessible for analysis. Therefore pH ranges for isoelectric focussing from 3 to 6 should be used (optimal range: 3,5 to 4,5).
3. The protein mixture also can be cleaved directly by trypsin. For a better solubility, we recommend to dissolve the acetone precipitate using a buffer containing 4 M urea. The solution has to be diluted to an urea concentration tolerated by trypsin as well as by endoproteinase Glu-C (check the enzyme data sheet).
4. Working with Mascot version 2.1 or below, all four modifications have to be selected as variable modifications. Working with Mascot version 2.2, carbamidomethylation of cysteines has to be defined as a fixed modification. Mass differences of labeled peptides are as follows: 4.0251 Da (light/medium), 6.0201 Da (light/heavy), and 10.0452 Da (light/extra heavy) per labeled amino group. Masses of ICPL-labeled residues according to Unimod are given in **Table 4.1**.

References

1. Ross, P. L., Huang, Y. N., Marchese, J. N., Williamson, B., Parker, K., Hattan, S., Khainovski, N., Pillai, S., Dey, S., Daniels, S., Purkayastha, S., Juhasz, P., Martin, S., Bartlet-Jones, M., He, F., Jacobson, A., and Pappin, D. J. (2004) Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging

reagents. *Molecular & Cellular Proteomics* 3(12), 1154–1169
2. Kellermann, J. Sample Preparation for Differential Proteome Analysis: Labeling Technologies for Mass Spectrometry. In: Proteomics Sample Preparation. von Hagen, J. (ed.), pp. 105–118 (Wiley-VCH, Weinheim, 2008)
3. Ong, S. E., Blagoev, B., Kratchmarova, I., Kristensen, D. B., Steen, H., Pandey, A., and Mann, M. (2002) Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Molecular & Cellular Proteomics* 1(5), 376–386
4. Gygi, S. P., Rist, B., Gerber, S. A., Turecek, F., Gelb, M. H., and Aebersold, R. (1999) Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nature Biotechnology* 17, 994–999
5. Schmidt, A., Kellermann, J., and Lottspeich, F. (2005) A novel strategy for quantitative proteomics using isotope-coded protein labels. *Proteomics* 5, 4–15
6. Brunner, A., Keidel, E., Dosch, D., Kellermann, J., and Lottspeich F. (2010) ICPL*Quant* – a software for non-isobaric isotopic labeling proteomics. *Proteomics* 10, 315–323

Chapter 5

Quantitative Proteome Analysis Using Isobaric Peptide Termini Labeling (IPTL)

Magnus Ø. Arntzen, Christian J. Koehler, Achim Treumann, and Bernd Thiede

Abstract

The quantitative comparison of proteome level changes across biological samples has become an essential feature in proteomics that remains challenging. We have recently introduced isobaric peptide termini labeling (IPTL), a novel strategy for isobaric quantification based on the derivatization of peptide termini with complementary isotopically labeled reagents. Unlike non-isobaric quantification methods, sample complexity at the MS level is not increased, providing improved sensitivity and protein coverage. The distinguishing feature of IPTL when comparing it to more established isobaric labeling methods (iTRAQ and TMT) is the presence of quantification signatures in all sequence-determining ions in MS/MS spectra, not only in the low mass reporter ion region. This makes IPTL a quantification method that is accessible to mass spectrometers with limited capabilities in the low mass range. Also, the presence of several quantification points in each MS/MS spectrum increases the robustness of the quantification procedure.

Key words: Chemical labeling, isobaric labeling, IsobariQ, IPTL, iTRAQ, mass spectrometry, quantitative proteomics, TMT.

1. Introduction

In addition to biosynthetic labeling strategies such as SILAC, chemical labeling at the peptide level with isobaric reagents has contributed to the growth of mass spectrometry-based comparative proteomic studies (1). Early isobaric labeling reagents such as TMT (2) or iTRAQ (3) are composed of three different segments. The first segment can be of different molecular weights and generates intense fragment ions in peptide MS/MS spectra. The second segment also has different molecular weights, but it is

K. Gevaert, J. Vandekerckhove (eds.), *Gel-Free Proteomics*, Methods in Molecular Biology 753,
DOI 10.1007/978-1-61779-148-2_5,

designed to balance out the first segment so as to ensure that the total reagent has the same molecular weight for all different labels. The third segment supplies a reactive group to ensure quantitative derivatization of peptides. These reagents can be used to analyze up to eight samples simultaneously (4).

Recently, we have introduced IPTL as a new approach to isobaric relative protein identification by derivatizing both peptide termini with complementary isotopically labeled reagents (5). The mixed isotopic labeling results in isobaric precursor masses and provides several quantification data points per peptide in MS/MS spectra, providing the option of statistical treatment of the result with increased confidence in quantification accuracy. Furthermore, MS/MS spectra after IPTL derivatization are suitable for relative quantification using ion trap mass spectrometers as the low mass cutoff in ion trap mass spectra does no longer interfere with quantification. In addition, the presence of pairs of b-ions and y-ions in CID-MS/MS spectra with reverse quantification ratios increases the confidence of database hits and/or aids in the assignment of ions when de novo sequencing has been carried out.

2. Materials

2.1. Protein Digestion with Endoproteinase Lys-C

1. Endoproteinase Lys-C, sequencing grade (e.g., Roche Applied Science, Sigma-Aldrich).
2. 1 pmol/μl solution of a standard protein (e.g., bovine serum albumin (BSA)).
3. 25 mM Tris-HCl, pH 8.5, 1 mM EDTA.
4. Pipette tips with C-18 microcolumns (e.g., Millipore, Proxeon, Varian).
5. HPLC grade (or better) water.
6. HPLC grade (or better) acetonitrile.

2.2. Derivatization of Lysine Residues with 2-Methoxy-4,5-Dihydro-1H-imidazole (MDHI)

1. 2-Methoxy-4,5-dihydro-1H-imidazole (MDHI) and the tetradeuterated form 2-methoxy-4,5-dihydro-1H-imidazol-4,4,5,5-d4 (MDHI-d4) of the reagent (C/D/N Isotopes).
2. Pipette tips with C-18 microcolumns (e.g., Millipore, Proxeon, Varian).

2.3. Derivatization of Alpha-N-Termini with Succinic Anhydride (SA)

1. Succinic anhydride (SA) and the tetradeuterated form succinic anhydride-d4 (SA-d4) (e.g., Cambridge Isotopes, C/D/N Isotopes, Sigma-Aldrich).
2. 200 mM disodium hydrogen phosphate (Na_2HPO_4).

3. Ammonium hydroxide (NH_4OH, 28–30% (v/v) solution).
4. Pipette tips with C-18 microcolumns (e.g., Millipore, Proxeon, Varian).

2.4. MALDI-MS Analysis

1. α-Cyano-4-hydroxycinnamic acid (20 mg/ml) in 0.3% aqueous trifluoroacetic acid/acetonitrile (1/1, v/v).

2.5. Nano-LC-ESI-MS Analysis

1. Solvent A: 0.1% (v/v) formic acid in water.
2. Solvent B: 0.1% (v/v) formic acid, 90% (v/v) acetonitrile in water.
3. Reversed phase C18 pre- and analytical columns.

2.6. Data Analysis

1. PC.
2. Protein search engine (e.g., Mascot).
3. Quantification software (IsobariQ).

3. Methods

IPTL is performed by crosswise peptide termini labeling to produce isobaric peptides. Thus, two different states of a protein sample can be compared and distinguished after mass spectrometry data acquisition. An outline of the IPTL approach is shown in **Fig. 5.1**. First, the proteins are digested with endoproteinase Lys-C to generate peptides with lysines at the C-terminal end. The free lysines are subsequently modified with MDHI-d4 (state A) and MDHI (state B), respectively. The second chemical modification is performed to modify peptide alpha-N-termini with succinic anhydride (SA). Here, the peptides from state A are labeled with SA whereas the peptides from state B are modified with the SA-d4. The doubly labeled peptides with single lysines from both combined states result in isobaric masses with identical physicochemical properties. Moreover, isobaric peptides co-elute during reverse phase LC-separation and are selected for MS/MS acquisition at the same time. Therefore, corresponding peptides of the two states result in single peaks in MS mode. The relative quantitative abundance of the peptides derived from the two different states can be detected by the ion intensities of peptide fragment ions in the MS/MS spectrum, occurring in pairs with 4 Da mass shifts. Here, the b-ions of the peak pairs with lower masses (d0) are derived from state A proteins and the higher masses (d4) from state B proteins. For y-ions, the lower mass ions (d0) in peak pairs are derived from state B proteins and the higher mass ions in peak pairs from state A proteins (**Fig. 5.1**). An example of a CID-MS/MS spectrum is shown in **Fig. 5.2**.

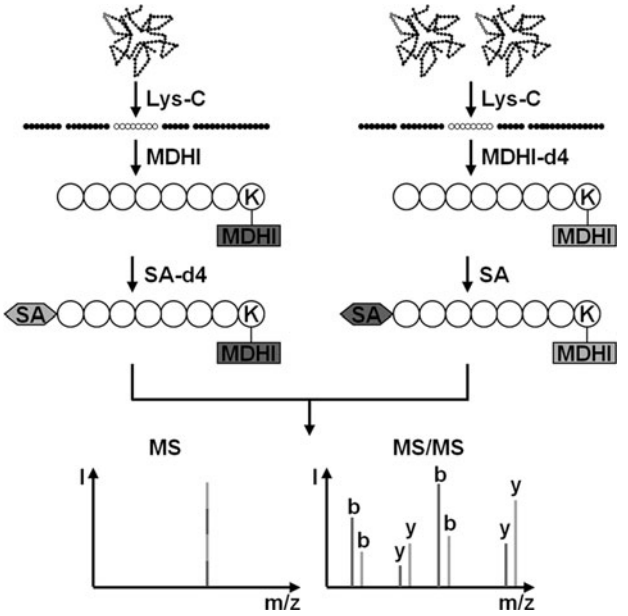


Fig. 5.1. Flowchart of the IPTL approach.

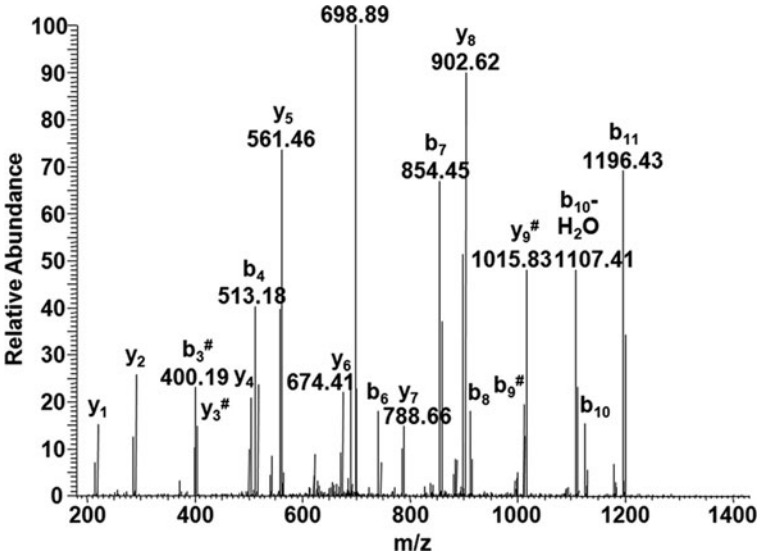


Fig. 5.2. CID-MS/MS spectrum of a parent ion at *m/z* 707.90 (2+), showing the identified sequence SA-ADLINNLGTIAK-MDHI-d4 of heat shock protein 90-beta b (Q58FF8). The b-series was identified as the light version of an IPTL ion pair, whereas the y-series was identified as the heavy version of an IPTL ion pair. The ions labeled with # could not be used for ratio calculation due to overlapping ion pairs between the fragment series. In total, 14 ion pairs were used for calculating the median peptide ratio of 1.73 (SD 0.34).

3.1. Protein Digestion with Endoproteinase Lys-C

1. Reconstitute 5 μg Lys-C in 50 μl water.
2. Add 950 μl of 25 mM Tris-HCl, pH 8.5, 1 mM EDTA to generate a stock Lys-C solution.
3. Add an appropriate amount of the Lys-C stock solution to all samples, which can be either a protein sample in solution or protein sample in a gel piece excised from one- or two-dimensional SDS-PAGE gels. Include three tubes containing each 10 μl (10 pmol) of the standard protein solution (*see* **Notes 1** and **2**).
4. Mix by vortexing.
5. Incubate for 16 h at 37°C under continuous shaking (e.g., at 1,000 rpm in a thermomixer).
6. Purify the digestion products using C-18 microcolumns using the instructions provided by the manufacturer of these microcolumns (*see* **Note 3**).
7. Evaporate the solvent from all sample tubes by vacuum drying.
8. Check the efficiency of the Lys-C digest by analyzing one of the three aliquots of the digest of the control protein using either MALDI-MS or ESI-MS analysis.

3.2. Derivatization of Lysine Residues with MDHI

1. Make stock solutions of MDHI and MDHI-d4 by dissolving 10 mg of these compounds in 250 μl of water (final concentration: 800 mM) (*see* **Note 4**).
2. Add 20 μl of the MDHI or MDHI-d4 solution, respectively, to each tube containing the dried peptides (*see* step 7 of **Section 3.1**). Add MDHI or MDHI-d4, respectively, to the tubes containing the standard protein digests.
3. Mix by vortexing.
4. Incubate for 3 h at 55°C under continuous shaking (e.g., at 1,000 rpm in a thermomixer).
5. Purify the MDHI-derivatized peptides using C-18 microcolumns (*see* **Note 3**).
6. Remove 10% aliquots (1 pmol each) from the two tubes containing the MDHI and MDHI-d4-derivatized control protein digests.
7. Evaporate the solvent from all tubes by vacuum drying.
8. Check the yield of derivatization of the control protein digest with MALDI-MS or ESI-MS analysis using the aliquots taken in step 6 of **Section 3.2** (*see* **Note 5** and **Fig. 5.3**).

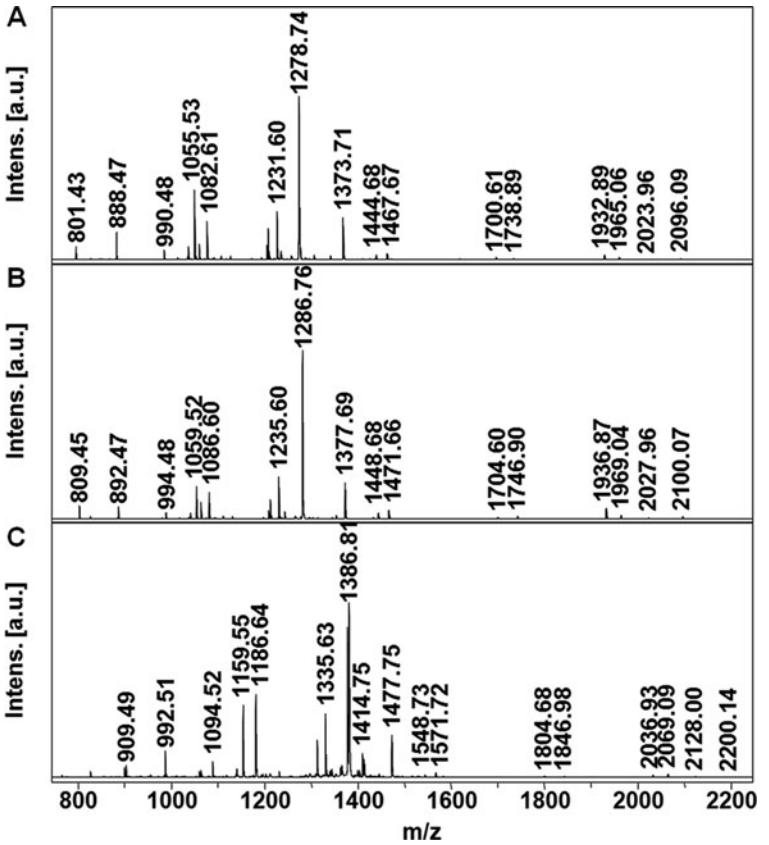


Fig. 5.3. MALDI-MS spectrum of BSA digested with Lys-C and modified with MDHI (**a**), with MDHI-d4 (**b**), and with MDHI and SA-d4 (**c**).

3.3. Derivatization of Peptide Alpha-N-Termini Using Succinic Anhydride

1. Dissolve 5 mg succinic anhydride (SA) and succinic anhydride (SA-d4), respectively, in 500 µl of 200 mM Na_2HPO_4 (*see* **Note 4**).
2. Add 9 µl of 10% (v/v) NH_4OH to the SA and SA-d4 solutions (final pH 6.5) (*see* **Note 6**).
3. Add 20 µl of SA solution to samples that are derivatized with MDHI-d4 (*see* step 5 of **Section 3.2**) and mix thoroughly.
4. Add 20 µl of SA-d4 solution to the samples that are derivatized with MDHI (*see* step 5 of **Section 3.2**) and mix thoroughly.
5. Incubate for 1 h at 37°C under continuous shaking (e.g., at 1,000 rpm in a thermomixer).
6. Purify the doubly derivatized peptides using C-18 microcolumns (*see* step 6 of **Section 3.1** and **Note 3**).
7. Remove 10% aliquots (1 pmol each) from the tubes containing the MDHI/SA-d4 and MDHI-d4/SA-derivatized control protein digests.

8. Evaporate solvent from all samples by vacuum drying.
9. Check the yield of derivatization of the fully derivatized BSA peptides with MALDI-MS or ESI-MS using 100 fmol of the aliquots taken in step 7 of **Section 3.3** (*see* **Note 5** and **Fig. 5.3**).
10. Combine samples (SA/MDHI-d4 and SA-d4/MDHI).
11. Analyze by nano-LC-ESI-MS(/MS).

3.4. MALDI-TOF/TOF-MS

1. Mix 0.5 μl of the α-cyano-4-hydroxycinnamic acid stock solution with an equal volume of sample.
2. Record peptide mass fingerprints.
3. Check for the completeness of the derivatization reactions by comparing the peptide mass fingerprints of the Lys-C digest with those of the MDHI-modified Lys-C digest, the MDHI-d4-modified Lys-C digest, and the doubly modified digests (MDHI-SA-d4-modified Lys-C digest and MDHI-d4-SA-modified Lys-C digest) (*see* **Note 5** and **Fig. 5.3**).

3.5. Nano-LC-ESI Mass Spectrometry

1. Dissolve the dried peptides in 10 μl of 1% formic acid and 5% acetonitrile.
2. Inject 5 μl into the nano-LC system.
3. Separate peptides by capillary RP-HPLC by applying a linear gradient from 5% of solvent B to 45% of solvent B over 45 min (*see* **Note 7**).
4. Record MS and MS/MS spectra (*see* **Notes 8** and **9**).

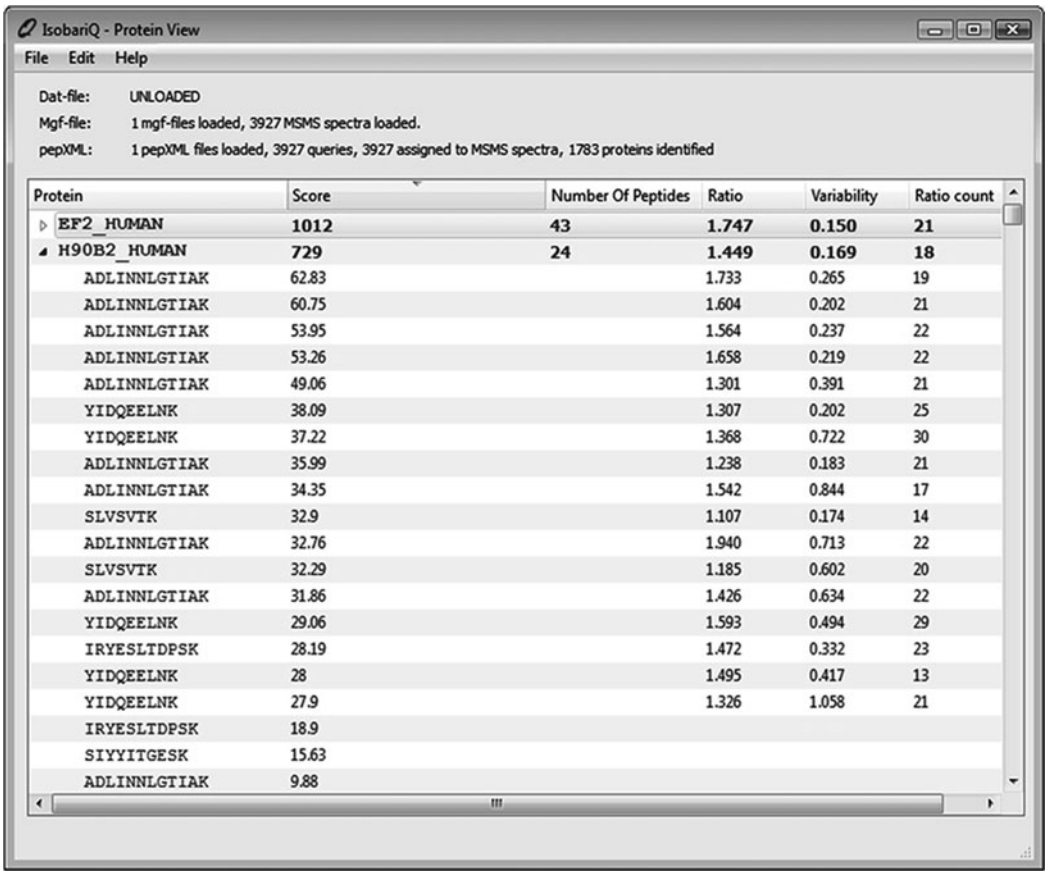
3.6. Protein Identification and Score-Based Quantification Using Mascot

1. Process raw data to mgf files (*see* **Note 10**).
2. Perform a database search with Mascot using the following search parameters: Lys-C as enzyme with no missed cleavage sites, N-terminal protein acetylation and methionine oxidation as variable modifications, fixed modifications set to either SA/MDHI-d4 or SA-d4/MDHI, respectively, and automatic decoy database to determine the false discovery rate. The mass accuracies are dependent on the used instrument.
3. Quantify (here, approximate relative quantification) peptides by comparing the ratios of the Mascot protein scores using SA/MDHI-d4 and SA-d4/MDHI as fixed modifications.

3.7. Protein Identification Using Mascot and Quantification Using IsobariQ

1. Process raw data to mgf files (*see* **Note 10**).
2. Perform a database search with Mascot using the following search parameters: Lys-C as enzyme with no missed cleavage sites, N-terminal protein acetylation, methionine oxidation, SA, SA-d4, MDHI, MDHI-d4 as variable modifications. Set no fixed modifications and apply the automatic decoy database to determine the false discovery rate.

3. Copy the Mascot result file (dat-file) to a local folder.
4. Launch the IsobariQ application and from the file menu select "Open Mascot dat-file." In the dialog select the correct dat-file and click "open." IsobariQ processes the Mascot dat-file and displays the Mascot results in a table (**Fig. 5.4**) where every protein can be clicked to display its assigned peptides and their individual scores. IsobariQ is not limited to Mascot as search engine because it can also read separate mgf files with corresponding identifications in a pepXML file (*see* **Note 11**).
5. Double click on a protein to load this protein with its assigned peptides and MS/MS spectra into the quantification module QuaIPTL (**Fig. 5.5**). The first MS/MS spectrum identifying this protein is shown in the bottom panel and all Mascot hits to this MS/MS spectrum are shown in



Protein	Score	Number Of Peptides	Ratio	Variability	Ratio count
EF2_HUMAN	1012	43	1.747	0.150	21
H90B2_HUMAN	729	24	1.449	0.169	18
ADLINNLGTIAK	62.83		1.733	0.265	19
ADLINNLGTIAK	60.75		1.604	0.202	21
ADLINNLGTIAK	53.95		1.564	0.237	22
ADLINNLGTIAK	53.26		1.658	0.219	22
ADLINNLGTIAK	49.06		1.301	0.391	21
YIDQEELNK	38.09		1.307	0.202	25
YIDQEELNK	37.22		1.368	0.722	30
ADLINNLGTIAK	35.99		1.238	0.183	21
ADLINNLGTIAK	34.35		1.542	0.844	17
SLVSVTK	32.9		1.107	0.174	14
ADLINNLGTIAK	32.76		1.940	0.713	22
SLVSVTK	32.29		1.185	0.602	20
ADLINNLGTIAK	31.86		1.426	0.634	22
YIDQEELNK	29.06		1.593	0.494	29
IRYESLTDPSK	28.19		1.472	0.332	23
YIDQEELNK	28		1.495	0.417	13
YIDQEELNK	27.9		1.326	1.058	21
IRYESLTDPSK	18.9				
SIYYITGESK	15.63				
ADLINNLGTIAK	9.88				

Fig. 5.4. IsobariQ – protein view. This is the main window of IsobariQ where all identified proteins and their respective peptides are shown. When a peptide is quantified in the QualPLT module (**Fig. 5.5**), the ratio and variability of every peptide is transferred back to this protein view where the overall protein ratio and variability is calculated. This protein list can be exported to a spreadsheet application for further analysis.

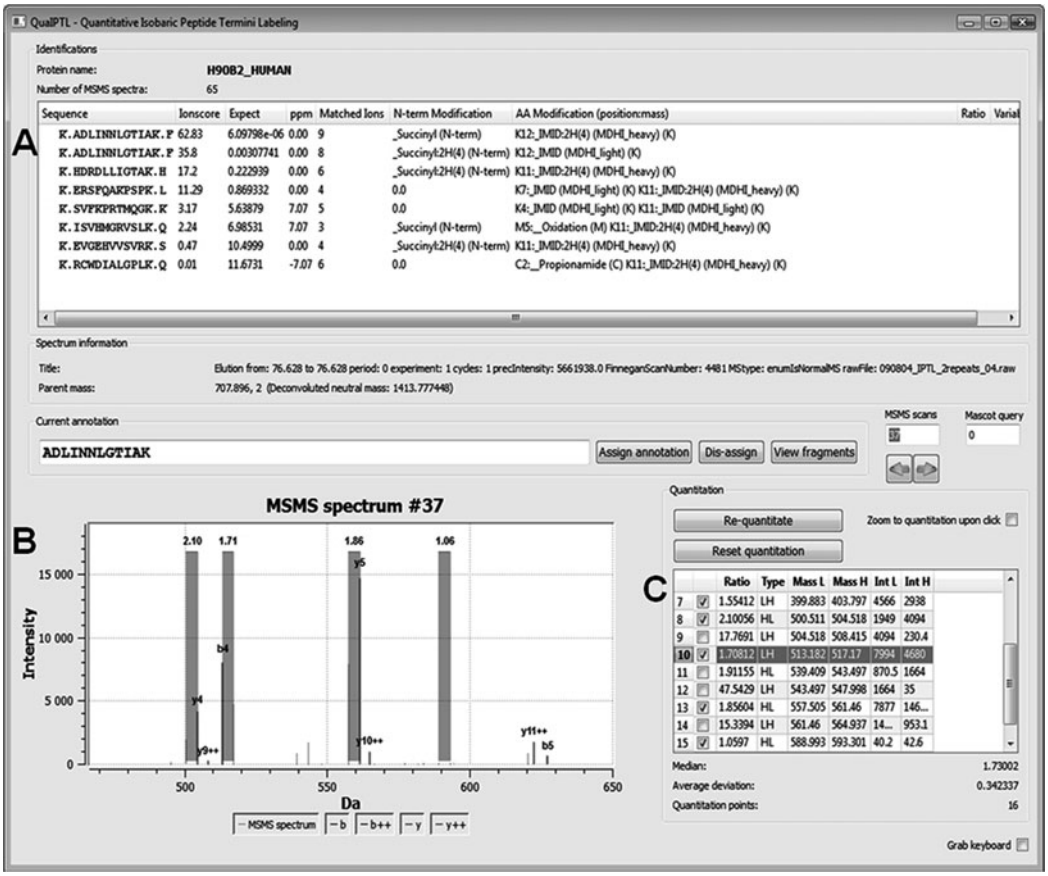


Fig. 5.5. IsobariQ – the QuaIPTL module. When a protein has been double-clicked in the protein view (**Fig. 5.4**) all MS/MS spectra assigned to this protein are displayed here. All Mascot hits assigned to a given MS/MS spectrum are displayed in the *top panel* (**a**) and the annotated MS/MS spectrum is shown in the *bottom panel* (**b**). When quantified, all the quantification events are listed in the quantification table (**c**) where the user can select which ratios to include or exclude for this particular MS/MS spectrum.

the top panel with detailed information about sequence, ions score, modifications, and ppm error (**Fig. 5.5a**). This is the same information as given by Mascot when hovering over a query number with the mouse in the web-view. The spectrum updates its annotations accordingly by clicking on different sequence annotations (**Fig. 5.5b**). For IPTL-labeled peptides, two identifications for the same peptide should be found: one with the modifications SA/MDHI-d4 and the other with SA-d4/MDHI. As a control, clicking on these two should change the b- and y-ion series annotation from light to heavy and vice versa.

6. Click on the "Quantitate"-button to inform QuaIPTL to detect all ion pairs which have been assigned to a sequence fragment and to calculate their individual ratios. The results

are shown in the quantification table (**Fig. 5.5c**), in which a user can select which ratios to include or exclude in the final quantification of this peptide.

7. Once a peptide is quantified, the ratio and the ratio variance for this peptide are transferred back to the protein table where also the overall protein ratio and its variance are calculated.
8. For further analysis, the protein list and all of its quantification information can be exported to a spreadsheet application via a tab-separated values (tsv) file. From the file menu choose "Save As" and give the file a unique name ending with.tsv. In the spreadsheet application the protein list can be opened and processed for post-quantification analysis like ratio normalization and significance determination.

4. Notes

1. For in-gel digestions, the gel pieces must be prepared according to the standard protocols including shrinking and expanding to destain the gel pieces and equilibration to digestion buffer (6). All solutions are best aliquoted into 1.5 ml tubes to reduce the potential for keratin contamination.
2. For in-gel digestion, the gel pieces must be covered by the buffer. Twenty samples can be digested using 1 ml Lys-C-containing solution if 50 μl is sufficient per sample. The Lys-C solution should be aliquoted before adding Tris-HCl buffer. If fewer samples have to be processed, then the remaining solution can be used for a week if stored at 4°C. According to the manufacturer, a solution of Lys-C in water may be used for 1–2 days at maximum, if stored at +2 to +8°C. For in-solution digestion, the recommended amount of Lys-C is 1/100 to 1/20 of the protein by weight. In the case of proteins which are hard to solubilize, urea (up to 4 M), SDS (up to 0.1%), or guanidine hydrochloride (up to 0.1 M) can be added to the digestion buffer prior to solubilization of the protein.
3. Different companies offer pipette tips with C-18 microcolumns with slightly different protocols, e.g., ZipTip μ-C18 (Millipore, Billerica, MS, USA), StageTips (Proxeon, Odense, Denmark), and OMIX pipette tips C18 (Varian Inc., Palo Alto, CA, USA).

4. The solutions of MDHI, MDHI-d4, SA, and SA-d4 must be prepared fresh shortly before use.
5. Peptide mass fingerprinting of a control protein (e.g., BSA) should be performed to check if the Lys-C digest and the chemical reactions have been complete. Start with enough control protein (e.g., 10 pmol) to ensure that the reactions have been complete even with a high amount of sample. Theoretical peptide masses of protein digests can be calculated, e.g., at http://au.expasy.org/tools/peptide-mass.html. After derivatization of the Lys-C digest, mass differences of +68 Da (MDHI), +72 Da (MDHI-d4), and +172 Da (MDHI + SA-d4 and MDHI-d4 + SA) must be detected in comparison to the Lys-C digest of the control protein. Furthermore, the detected masses of the precedent spectra must disappear. It is recommended to repeat a derivatization if the yield of the reaction was below 97%. Peptide mass fingerprinting can easily be performed using MALDI-MS. If a MALDI mass spectrometer is not available, the control sample can also be analyzed using ESI-MS or LC-ESI-MS.
6. Ammonium hydroxide can be purchased as a 28–30% solution. Ammonium hydroxide must be rapidly pipetted to avoid evaporation. Work under a fume hood and comply with the health and safety regulations that apply in your country for handling this hazardous compound.
7. In average, longer peptides are produced using Lys-C in comparison to trypsin digestion of proteins. Therefore, a steeper gradient should be used (e.g., 45% of solvent B at the end of the separating gradient).
8. Peptide fragmentation by collision-induced dissociation (CID) can be used. At least two MS/MS scans of each precursor mass should be recorded, because slight differences of the ion pairs in MS/MS from scan to scan have been observed.
9. The mass differences within the MS/MS spectra according to IPTL labeling are dependent on the charge of the fragments: $m/z = 4$ for $z = 1$, $m/z = 2$ for $z = 2$, $m/z = 1.33$ for $z = 3$.
10. mgf files can be generated using different software routines depending on the instrument used and the type of raw data. For Thermo instruments we recommend using DTASuperCharge available free of charge from http://msquant.alwaysdata.net/msq/download/
11. IsobariQ (7) is freely available for academic users at www.biotek.uio/research/thiede_group/software/

Acknowledgments

IPTL was supported by the National Program for Research in Functional Genomics in Norway (FUGE, project no. 183418/S10) of the Norwegian Research Council, MLS^{UiO} and FUGE-Øst to BT.

References

1. Ong, S. E., and Mann, M. (2005) Mass spectrometry-based proteomics turns quantitative. *Nat. Chem. Biol.* **1**, 252–262.
2. Thompson, A., Schafer, J., Kuhn, K., Kienle, S., Schwarz, J., Schmidt, G., et al. (2003) Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Anal. Chem.* **75**, 1895–1904.
3. Ross, P. L., Huang, Y. N., Marchese, J. N., Williamson, B., Parker, K., Hattan, S., et al. (2004) Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. *Mol. Cell Proteomics* **3**, 1154–1169.
4. Choe, L., D'Ascenzo, M., Relkin, D., Ross, P., Williamson, B., Guertin, S. et al. (2007) 8-plex quantitation of changes in cerebrospinal fluid protein expression in subjects undergoing intravenous immunoglobulin treatment of Alzheimer's disease. *Proteomics* **7**, 3651–3660.
5. Koehler, C. J., Strozynski, M., Kozielski, F., Treumann, A., and Thiede, B. (2009) Isobaric peptide termini labeling for MS/MS-based quantitative proteomics. *J. Proteome. Res.* **8**, 4333–4341.
6. Webster, J., and Oxley, D. (2005) Peptide mass fingerprinting: protein identification using MALDI-TOF mass spectrometry. *Methods Mol. Biol.* **310**, 227–240.
7. Arntzen, M.Ø., Koehler, C.J., Barsnes, F.S., Berven, F.S., Treumann, A., and Thiede, B. (2011) IsobariQ: software for isobaric quantitative proteomics using IPTL, iTRAQ, and TMT. *J. Proteome Res.* **10**, 913–920.

Chapter 6

Complete Chemical Modification of Amine and Acid Functional Groups of Peptides and Small Proteins

Casey J. Krusemark, Brian L. Frey, Lloyd M. Smith, and Peter J. Belshaw

Abstract

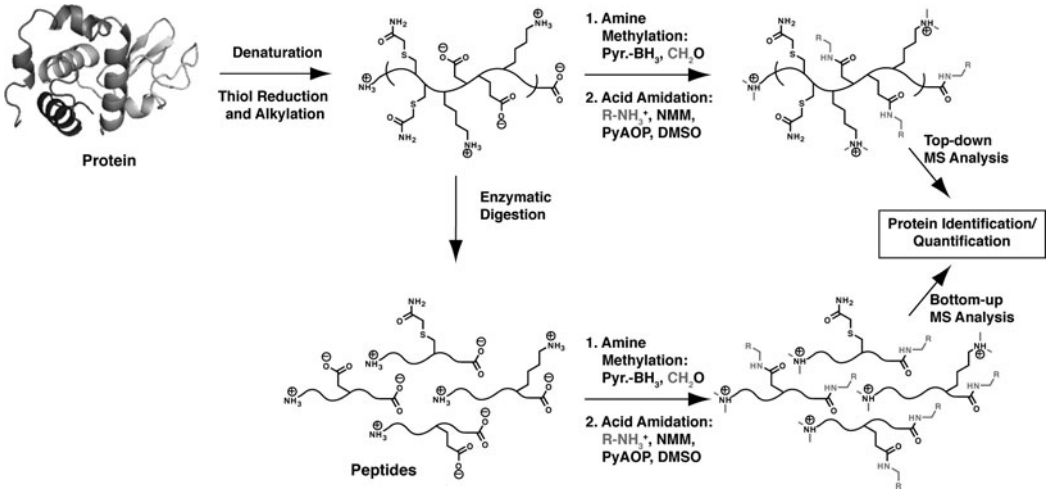
The chemical modification of protein thiols by reduction and alkylation is common in the preparation of proteomic samples for analysis by mass spectrometry (MS). Modification at other functional groups has received less attention in MS-based proteomics. Amine modification (Lys, N-termini) by reductive dimethylation or by acylation (e.g., iTRAQ labeling) has recently gained some popularity in peptide-based approaches (bottom-up MS). Modification at acidic groups (Asp, Glu, C-termini) has been explored very minimally. Here, we describe a sequential labeling strategy that enables complete modification of thiols, amines, and acids on peptides or small intact proteins. This method includes (1) the reduction and alkylation of thiols, (2) the reductive dimethylation of amines, and (3) the amidation of acids with any of several amines. This chemical modification scheme offers several options both for the incorporation of stable isotopes for relative quantification and for improving peptides or proteins as MS analytes.

Key words: Mass spectrometry (MS), stable isotope labeling, acylation, dimethylation, amidation, proteomics, protein derivatization, peptide derivatization.

1. Introduction

The chemical modification of peptides or proteins prior to MS analysis is generally carried out for any of three purposes: (1) to attach an affinity handle to purify a subset of the proteome, (2) to incorporate stable isotope labels for relative quantification of two or more samples, and/or (3) to modify the physiochemical properties of peptides and proteins to improve them as MS analytes or to facilitate upstream sample preparation (e.g., chromatography). In this chapter, we present a detailed protocol for the complete

K. Gevaert, J. Vandekerckhove (eds.), *Gel-Free Proteomics*, Methods in Molecular Biology 753,
DOI 10.1007/978-1-61779-148-2_6,



Scheme 6.1. Protein and peptide modification strategy.

modification of amines and acids on peptides or small intact proteins (**Scheme 6.1**). Given that these are very common functional groups, this approach presents useful possibilities for incorporating labels for quantification and for broadly modifying the properties of proteins or peptides in proteomic samples (1). We have shown that modifying proteins in this manner can significantly affect charging during electrospray ionization mass spectrometry (ESI-MS) and even reveal insights into the ESI mechanism (2, 3). We are currently exploiting this modification strategy for peptides, including tryptic digests of yeast lysate samples, to increase ionization efficiency and charge states during ESI.

The primary consideration when choosing chemical reactions for the modification of proteins and peptides in MS analysis is the purity of the transformation. Reactions should proceed to completion and modify only targeted functional groups. Incomplete modifications and side products produce additional complexity that can hinder analysis considerably. This is more problematic for the modification of intact proteins, as side reactions and missed modifications are quickly compounded. For example, the modification of a protein with 10 targeted sites leads to over 1,000 possible products ($2^n - 1$) when the transformation is not complete at each site (4). Additionally, a transformation with an efficiency of 95% at each site would produce only 60% (0.95^{10}) of the desired product. Other considerations in choosing modification reactions may include generality, cost, ease of use, and the retention of desired properties for MS analysis. Ideally, reactions should enable facile modification with any of several functionalities at low cost without complicating MS analysis.

We have developed conditions for the modification of intact proteins in part to facilitate "top-down" proteomics, where

proteins are analyzed intact, as opposed to the "bottom-up" approach, where proteins are digested into peptides. There are several advantages to the analysis of intact proteins. The increased complexity generated by proteolysis is avoided. Measurement of the intact mass of a protein can contribute significantly to its identification. Also, post-translational modifications can be more accurately located and assigned to distinct protein forms. The "top-down" approach in MS, however, has seen limited implementation partially due to the physiochemical properties of many intact proteins, which can present problems of insolubility, limited charging in MS, and poor separation in chromatography (5, 6). Chemical modification may be able to address these issues. Furthermore, the development of methods for modification of intact proteins can enable quantitative proteomics on whole proteins through labeling with stable isotopes. Labeling with a mixture of two isotopic versions of the same reagent has the additional application that the number of functional groups in a protein can be determined by the mass difference between the heavy and light forms (**Fig. 6.1**). This information, together with the intact mass, can be used to uniquely identify a protein from a protein database or genetic sequences, which can obviate the need for MS/MS sequencing (1). This approach for protein identification has been applied with metabolic isotopic labeling in cell culture (7, 8).

We have previously demonstrated the generality of global amine and acid labeling on four small proteins (≤17 kDa) (1). Further work may be required for the modification of larger proteins, as the possibilities of side products and missed modifications will increase with increasing protein size. Also, large proteins may

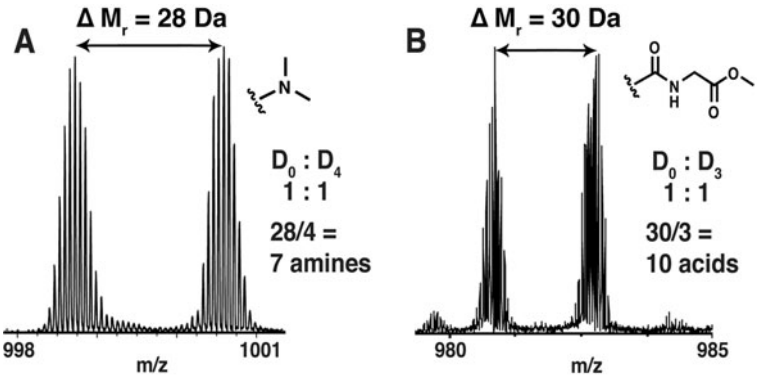


Fig. 6.1. Relative quantification of protein samples and determination of the number of protein functional groups. (**a**) ESI-FTMS spectrum of 15+ ions of thiol-alkylated lysozyme-dimethylated with either d_0 or d_2 formaldehyde and mixed in a 1:1 ratio. (**b**) ESI-FTMS spectrum of 16+ ion of thiol-alkylated, dimethylated lysozyme amidated with either d_0 or d_3 glycine methyl ester and mixed in a 1:1 ratio. Spacing of peaks indicates the numbers of functional groups present in the protein (from Ref. (1), reprinted with permission from American Chemical Society, copyright 2008).

present difficulties for complete modification due to the increased likelihood of higher order structure and incomplete denaturation. However, in the previous work, we did find that the smallest protein tested, ubiquitin (8.6 kDa), was the most difficult to modify at the acidic functional groups. This may have been due to the relatively high density of acidic groups or otherwise related to the particular sequence of this protein. We anticipate that the presented conditions would be directly applicable for modification of a low-mass proteomic mixture (9).

We have found modifying peptides to be generally easier than modifying intact proteins due to their smaller size and more limited secondary and tertiary structures. Thus, the reaction protocols developed herein for proteins also worked well for peptide labeling. However, we have further optimized the reaction and cleanup conditions specifically for peptide modification to produce a robust protocol for bottom-up proteomic applications.

1.1. Amine Modification

Chemical modification of amines in peptides has seen several applications in MS-based proteomics. Generally, these have been limited to guanidination, acylation, and reductive alkylation reactions. Guanidination of peptides with *O*-methylisourea has been used to increase signal in MALDI-MS and also to incorporate stable isotopes for quantitative bottom-up proteomics (10). However, we experienced difficulty in achieving complete guanidination on intact proteins.

Acylation of amines with active esters has seen several proteomic applications involving either modification of peptides or intact proteins. Two notable examples include iTRAQ® labeling of peptide amines (11) and ICPL™ for intact proteins (12), both of which employ stable isotopes to enhance relative quantification of protein abundance. Acylation of amines is an attractive modification reaction in that several activated esters are commercially available (typically *N*-hydroxysuccinimide (NHS) esters) or are easily prepared from carboxylic acids. While acylation of amines proceeds efficiently with activated esters, our experience and additional studies have shown that selective amine acylation is not possible and that hydroxyl groups of tyrosine, serine, and threonine are acylated as well (1, 12–14). The O-acylations can be selectively removed, but harsh conditions are required, such as alkaline hydrolysis, lysis with high concentrations of hydroxylamine, or high temperatures. These conditions are less than desirable for proteomic analysis due to side reactions or loss of post-translational modifications (14). Presumably, the iTRAQ reagent avoids this complication due to the electron-withdrawing nature of *N*-methylpiperazine acetic acid esters, which likely hydrolyze readily at neutral pH. An additional drawback of amine modification via acylation is the loss of the amine basic group that is important for charging in positive mode MS, the typical mode

for proteomic analysis. This can be avoided by including a non-nucleophilic basic group, such as a tertiary or quaternary amine, in the activated ester. The preparation of such an active ester is not trivial, and including a basic site in the active ester reduces the generality of this approach.

The reductive alkylation of peptide amines with formaldehyde has been gaining popularity for proteomic applications, primarily as a means to incorporate stable isotopes for quantitative proteomics (15). Also, dimethylation significantly enhances the signal of a_1 ions in collisionally activated dissociation (CAD) (16). A drawback of the approach has been side products and incomplete modification for reactions using the common reductant sodium cyanoborohydride (16, 17). By modifying the procedures of Means and Feeney (18), we have developed conditions using pyridine-borane as the reductant that enable complete dimethylation of amines to yield modified peptides and whole proteins with excellent purity (**Fig. 6.2**) (1). In addition to high purity, the benefits of this reaction include the ease of modification, the preservation of the amine group for MS charging, and the inexpensive availability of stable isotope forms of formaldehyde. The reaction is limited in generality, since other aldehydes react much less efficiently than formaldehyde. Recently, however, complete reductive alkylation of an intact protein was demonstrated with the dialdehyde, glutaraldehyde (19).

1.2. Acid Modification

The modification of peptide carboxylic acid groups has seen more limited application in MS proteomics. Esterification of acidic groups using anhydrous, acidic methanol or ethanol has been employed to incorporate stable isotopes, to prevent preferential cleavage at acidic sites in CAD, and also to eliminate copurification of carboxylic acid-containing peptides with phosphopeptides during immobilized metal affinity chromatography (IMAC) (20–22). The esterification reaction is limited to short chain alcohols that will solubilize peptides and is not applicable to proteins, which are often poorly soluble in alcohols (23). Amidation of carboxylic acids by reaction with amines has been used classically for the determination of protein acids by hydrolytic amino acid analysis (24), but it has not seen application in MS-based proteomics. A study by Tanaka and coworkers on the amidation of peptide acids with ammonia and a water-soluble carbodiimide found incomplete transformations and the dehydration of peptides due to intramolecular cyclization reactions (25). Another study by Xu et al. reported excellent reaction efficiency for carbodiimide coupling of a particular amine, 1-(2-pyrimidyl) piperazine, to peptide acid groups for a few model peptides (26).

We have developed conditions for the complete modification of peptide and protein carboxylic acids in high purity (1). Our approach involves the capping of both thiol and amine

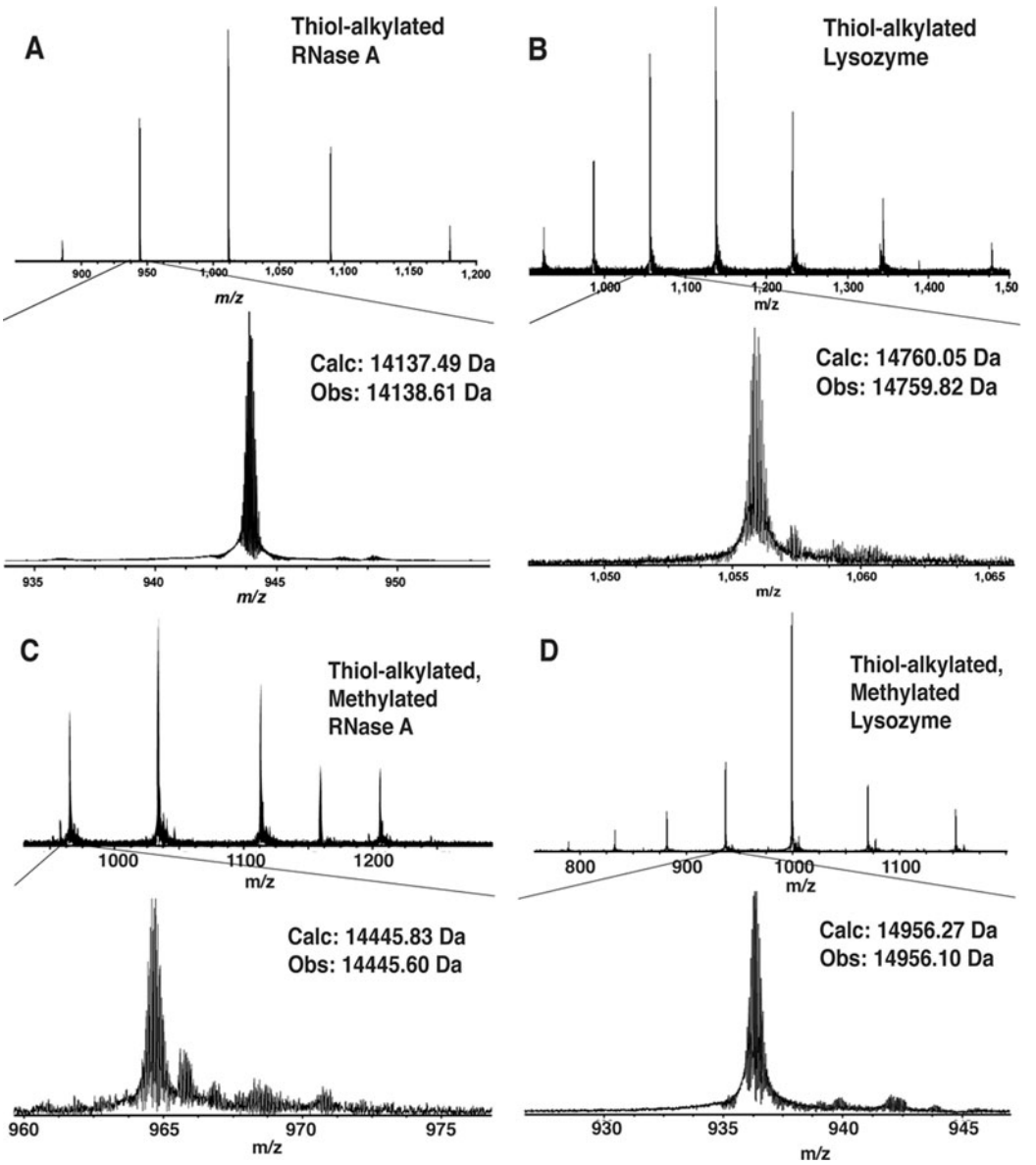


Fig. 6.2. ESI-FTMS spectra of (**a**) thiol-alkylated RNase A, (**b**) thiol-alkylated lysozyme, (**c**) thiol-alkylated, amine-dimethylated RNase A, and (**d**) thiol-alkylated, amine-dimethylated lysozyme. The expanded spectra show a cluster of peaks due to the natural isotopic distribution of the completely modified protein product. If the reaction did not go to completion, peaks would be observed at lower *m/z* values, but no such peaks are present (from Ref. (1), reprinted with permission from American Chemical Society, copyright 2008).

nucleophilic groups on the protein prior to the acid coupling reaction in order to prevent intramolecular cyclization (**Scheme 6.1**). The spectra in **Fig. 6.3** demonstrate the high purity of modified protein obtained from this method. The modification of carboxylic acids by amidation has several advantages as a proteomic labeling reaction. The main advantage is its generality. We

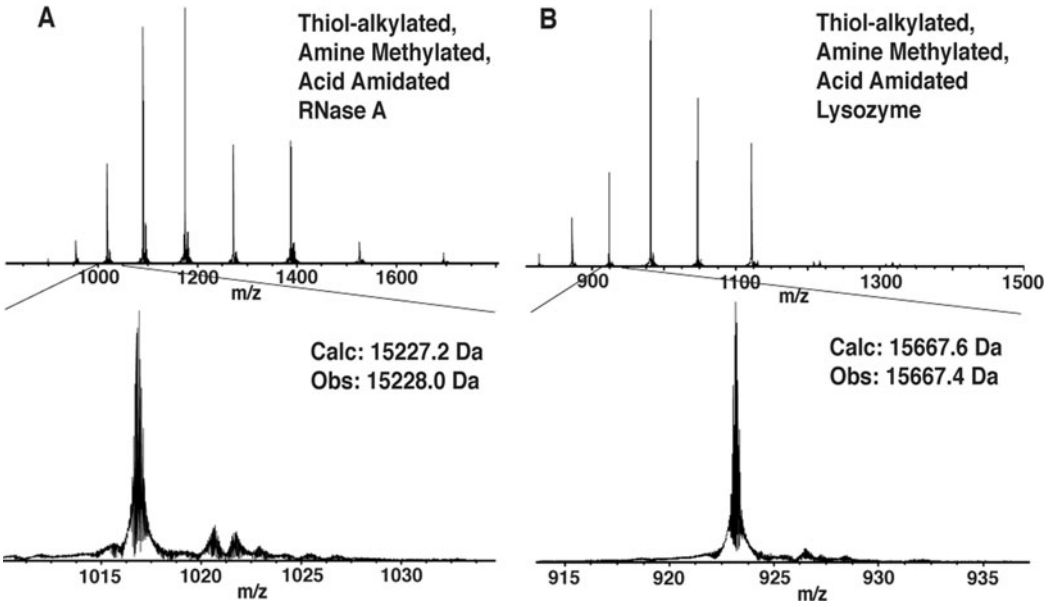


Fig. 6.3. ESI-FTMS spectra of (**a**) thiol-alkylated, amine-dimethylated, acid-amidated (with glycine methyl ester) RNase A and (**b**) thiol-alkylated, amine-dimethylated, acid-amidated (with glycine methyl ester) lysozyme. As in **Fig. 6.2**, the near absence of peaks at lower *m/z* values indicates that the reactions have gone to completion (from Ref. (1), reprinted with permission from American Chemical Society, copyright 2008).

have demonstrated efficient coupling with a variety of primary amines (**Figs. 6.4** and **6.5**). Primary amines are widely commercially available. Additionally, carboxylic acids do not contribute to charging in positive mode MS (2), as so their modification does not hinder MS detection. As with esterification, amidation can limit preferential cleavage at acidic sites in CAD (25). Importantly, we found that amidation does not complicate MS/MS sequencing by CAD, electron transfer dissociation (ETD), or infrared multiphoton dissociation (IRMPD) (1).

2. Materials

2.1. Thiol Modification and Digestion

1. Protein alkylation buffer: 6 M guanidine-HCl and 60 mM NH_4HCO_3. Prepare freshly before use.
2. 500 mM stock solution of dithiothreitol (DTT) in water.
3. Iodoacetamide (IAA).
4. Sequencing-grade modified trypsin (Promega).
5. Ultra-Clean^{TM} SPE C18 4 ml cartridge (Grace/Alltech).

2.2. Amine Modification

1. Amine modification buffer: 6 M guanidine-HCl, 300 mM triethanolamine buffer, pH 7.5. Prepare freshly before use.

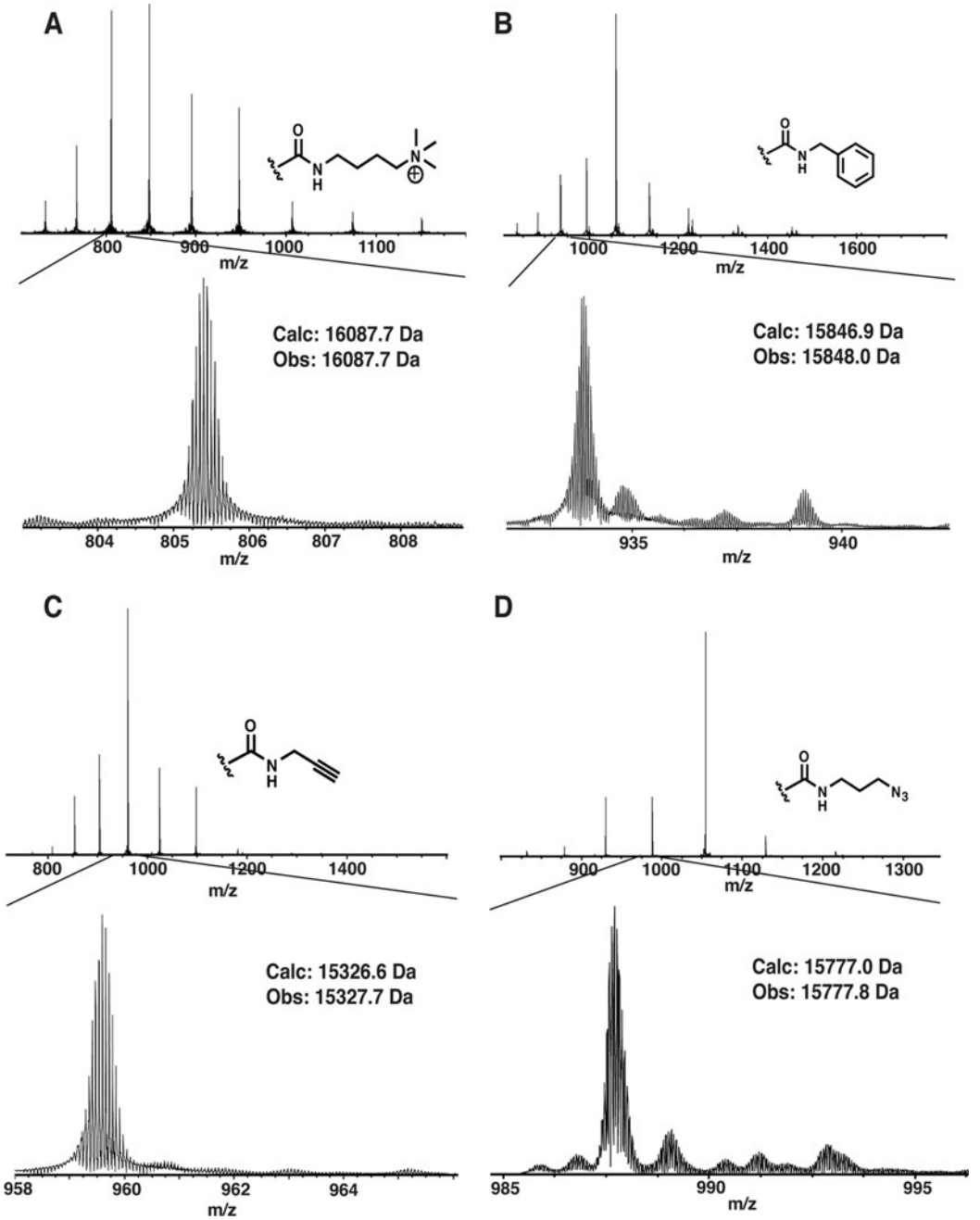


Fig. 6.4. ESI-FTMS spectra of thiol-alkylated, amine-dimethylated, acid-amidated lysozyme. Amidation was performed with (**a**) (4-aminobutyl)trimethylamine, (**b**) benzylamine, (**c**) propargylamine, or (**d**) 3-azidopropylamine. (from Ref. (1), reprinted with permission from American Chemical Society, copyright 2008).

2. 8 M pyridine-BH_3 in pyridine.
3. Methanol.
4. d_0 formaldehyde (formalin, 37% in water).
5. d_2 formaldehyde (Cambridge Isotope Laboratories).

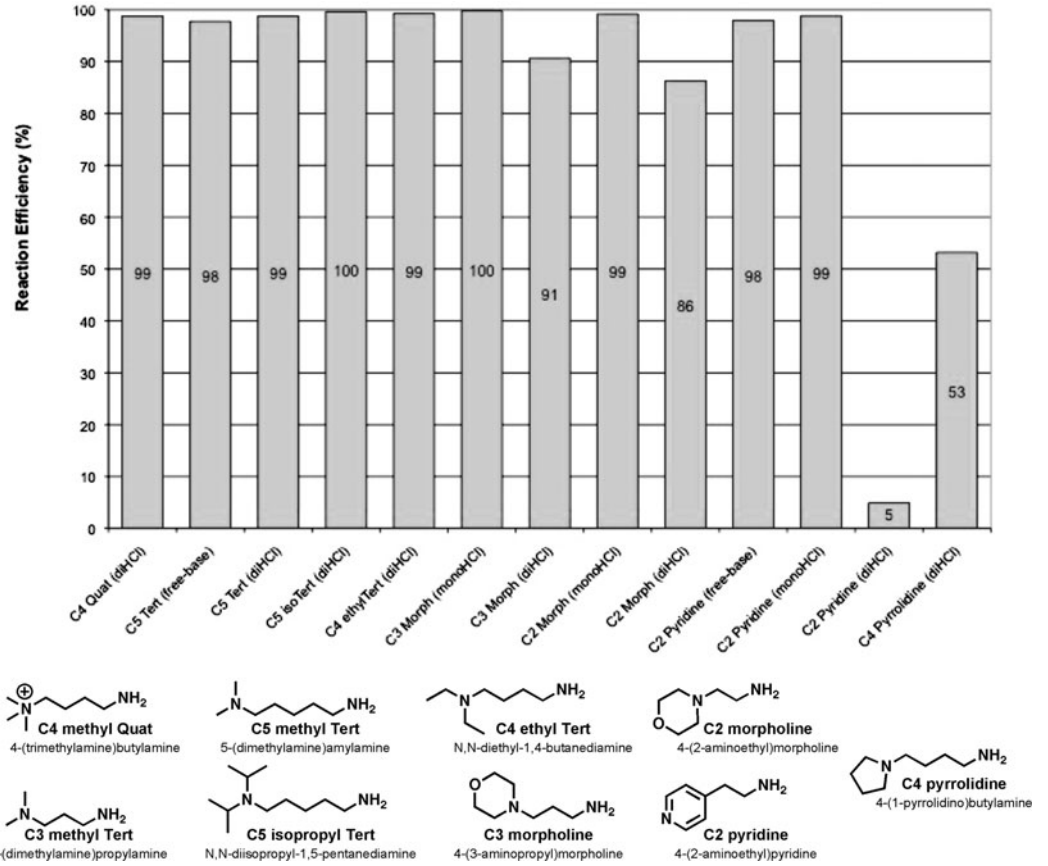


Fig. 6.5. The reaction efficiency for various amines used to amidate the peptide neurotensin (pyroELYENKPRRPYIL). Efficiencies were calculated from the peak intensities of the doubly amidated product, the singly amidated product, and the starting material (dimethylated peptide) obtained by direct injection MS. The results are given as a percentage of the total number of acid sites, which means that the singly amidated product was counted as one site reacted and one site not reacted. Different forms of the amine were used for modification as indicated (free base, mono-HCl, and/or di-HCl). Greater than 99% reaction efficiency was obtained for all amines in at least one of the forms, except for the case of pyrrolidine, which gave poor results for the di-HCl salt and the other forms were not tested.

2.3. Acid Amidation

1. DMSO (*see* **Note 1**).
2. (7-Azabenzotriazol-1yloxy)tripyrrolidinophosphonium hexafluorophosphate (PyAOP, Applied Biosystems)
3. *N*-Methylmorpholine (NMM)
4. Desired amine HCl salts such as those listed in **Section 2.3.1**
5. Aqueous solution of 0.1% formic acid
6. Aqueous solution of 0.1% trifluoroacetic acid (TFA)
7. Chloroform
8. 5K MWCO Amicon Ultra centrifugal filter device (Millipore)
9. C18 Ultramicrospin columns (The Nest Group)

2.3.1. Amine HCl Salts

1. Glycine methyl ester HCl (Aldrich)
2. d_3 glycine methyl ester HCl, prepared from d_3 methanol as previously described (27)
3. Benzylamine HCl
4. Propargylamine (Aldrich)
5. 3-Azidopropylamine, prepared as previously described (28)
6. (4-Aminobutyl)trimethylammonium dichloride (C4 Quat), prepared as described in (1)
7. 5-(Dimethylamino)amylamine (C5 Tert, Matrix Scientific)
8. *N*,*N*-diisopropyl-1,5-pentanediamine (C5 isoTert, Alfa Aesar)
9. *N*,*N*-diethyl-1,4-butanediamine (C4 ethylTert, Alfa Aesar)
10. 4-(3-Aminopropyl)morpholine (C3 Morph, Alfa Aesar)
11. 4-(2-Aminoethyl)morpholine (C2 Morph, Aldrich)
12. 4-(2-Aminoethyl)pyridine (C2 Pyridine, Alfa Aesar)
13. 4-(1-Pyrrolidino)butylamine (C4 Pyrrolidine, Alfa Aesar)
14. 1 M HCl in ether (Alfa Aesar)
15. Acetone

2.4. HPLC Purification

1. Mobile phase A: 0.5% TFA in water
2. Mobile phase B: 0.5% TFA in acetonitrile

3. Methods

3.1. Protein Modification

3.1.1. Thiol Alkylation

1. Dissolve protein sample in protein alkylation buffer to 5 mg/ml concentration.
2. Add 500 mM DTT stock to 20 mM concentration.
3. Heat at 37°C for 1 h to reduce disulfide bonds.
4. Dilute solution to 2.5 mg protein/ml with protein alkylation buffer.
5. Add solid IAA to 25 mM concentration.
6. Incubate at room temperature for 1.5 h in the dark, e.g., in foil-covered tube.
7. Purify protein by RP-HPLC on a semi-preparative C18 reversed phase column using a gradient of 15–95% B over 25 min (*see* **Note 2**).
8. Remove acetonitrile from collected fractions by rotary evaporation.

9. Freeze protein solution and lyophilize to obtain pure labeled protein.

3.1.2. Amine Methylation

1. Dissolve thiol-alkylated protein or protein lacking thiols in amine modification buffer to 1.25 mg/ml. Be certain that the proteins are well dissolved and heat gently at 50°C if needed.
2. Dilute sample to 1 mg/ml with methanol.
3. Add pyridine-BH_3 complex to 30 mM.
4. Add formaldehyde (d_0 or d_2) to 20 mM.
5. Sonicate reactions for 1 min in water bath sonicator to incorporate pyridine-BH_3.
6. Incubate reaction at room temperature for 2 h.
7. Purify by reverse phase HPLC and lyophilize, as described in steps 7–9 of **Section 3.1.1**.

3.1.3. Acid Amidation

1. Dissolve thiol-alkylated, amine-dimethylated protein in DMSO to 1.5 mg/ml concentration.
2. Heat solution (50°C) and sonicate until dissolved.
3. Add glycine methyl ester HCl (d_0 or d_3) to 1 M and *N*-methylmorpholine (NMM) to 750 mM. Heat solution and sonicate if necessary to dissolve (*see* **Notes 3, 4,** and **5**).
4. Add solid PyAOP to 40 mM.
5. Allow reactions to proceed at room temperature for 2 h.
6. Quench with an equal volume of water.
7. Purify by reverse phase HPLC, as described in steps 7–9 of **Section 3.1.1** (*see* **Note 6**).
8. If reaction components (typically tripyrrolidinophosphine oxide) co-elute with modified protein on RP-HPLC, remove small molecular weight species with a molecular weight cut-off filter device prior to HPLC. Dilute the reaction mixture 10- to 20-fold in protein alkylation buffer and centrifuge over the MWCO filter to concentrate. Repeat dilution and centrifugation until contaminants are removed.

3.1.4. MS Analysis

1. Dissolve purified, lyophilized protein in 49.5/49.5/1% H_2O/methanol/formic acid.
2. Analyze by MS method of choice, e.g., direct electrospray or LC-MS.

3.2. Peptide Modification

3.2.1. Thiol Alkylation

Peptides or proteins containing cysteine residues are reduced and alkylated with DTT and IAA according to any standard procedure, such as the one outlined in **Section 3.1.1**. Proteins are then digested enzymatically to produce peptides. Subsequent

purification with a C18 stationary phase is necessary to remove the excess DTT and IAA before proceeding with the methylation and amidation reactions.

3.2.2. Amine Methylation

1. Dissolve thiol-alkylated peptides in 50:50 methanol:H_2O containing 40 mM formaldehyde, 60 mM pyridine-BH_3 complex, and 50 mM NMM to a peptide concentration of 2 mg/ml.
2. Incubate reaction at room temperature for 1–2 h.
3. Evaporate to dryness in a vacuum centrifuge, which removes most of the excess reagents.

3.2.3. Acid Amidation

1. Dissolve thiol-alkylated, amine-dimethylated peptides in DMSO to 1.5 mg/ml concentration.
2. In a separate tube, prepare the following reagent mixture (note that the amounts are for 100 μg of peptides; scale accordingly). Dissolve 30 μmol of the mono- or di-HCl salt of the desired amine in 5 μl of water and 2.6 μl of NMM. If the amine does not completely dissolve, heat the solution or add an additional 1–2 μl of water (*see* **Note 7**).
3. Transfer the peptide/DMSO solution into the amine reagent mixture and mix thoroughly. The resulting amine and NMM concentrations are 450 and 350 mM, respectively.
4. Add solid PyAOP to 60 mM.
5. Allow reaction to proceed at room temperature for 2 h.
6. Quench with 0.1% aqueous formic acid to dilute DMSO to 5% (v/v) (*see* **Note 8**).
7. Extract the quenched reaction with chloroform (double the sample volume). Shake vigorously for 2 min, with periodic venting. Transfer most of the aqueous layer to a new tube with another aliquot of chloroform and shake again. Transfer most of the aqueous layer to a new tube.
8. Evaporate to dryness in a vacuum centrifuge.
9. Dissolve peptides in 0.1% aqueous TFA to 0.3 mg/ml and purify by C18 solid-phase extraction. Elute with 75% acetonitrile and 0.1% formic acid.
10. Evaporate to dryness or dilute into a suitable (LC)-MS solvent.

3.3. Preparation of Amine HCl Salts

1. Add a stoichiometric amount of 1 M HCl in ether to the amine-free base.
2. Remove solvent by rotary evaporation.
3. Dissolve in acetone.
4. Remove solvent by rotary evaporation.

4. Notes

1. We have found that "wet" DMSO works as well or better than dry DMSO. Simply use a bottle of standard, reagent-grade DMSO. No special precautions were taken to limit water absorption from air: the solvent was not distilled, did not contain a water adsorbent or a septum cap, and was not blanketed with inert gas after each use.
2. While we found quick HPLC gradients to be the most convenient way to purify proteins between reaction steps, we also explored acetone or ethanol precipitation. This approach provided clean products, but there were significant losses of protein and often carryover of guanidinium salts. There are several, potentially higher throughput options for the purification of the high molecular weight proteins from the small molecules used for modification. These include dialysis/molecular weight cutoff filters, reverse phase spin columns, and more quantitative precipitation with sodium deoxycholate/trichloroacetic acid (29).
3. We found these weakly basic, buffered conditions to be critical to achieve high purity transformations. Using amine-free bases or a stronger base, such as triethylamine, in the place of NMM gave poor purity products.
4. If using other amines, the ratio of amine HCl salt to NMM base should be determined empirically to optimize efficiency of labeling. The typical ratio of amine HCl to NMM with more basic amines is 1 M amine HCl to 500 mM NMM, which was used for benzyl amine and 3-azidopropylamine (**Fig. 6.4b, d**). A ratio of 500 mM amine HCl and 250 mM NMM was used for propargylamine labeling (**Fig. 6.3c**).
5. If using a highly polar amine that is poorly soluble in DMSO, a minimum amount of water should be added to solubilize. In the case of (4-aminobutyl)trimethylammonium dichloride (**Fig. 6.4a**), 500 mM of this di-HCl salt and 350 mM NMM were used with up to 10% H_2O added.
6. Proteins modified with glycine methyl ester should be quickly reduced in volume and frozen to minimize ester hydrolysis while in acidic HPLC buffers.
7. The amidation works best when employing hydrochloride salts of the amines. If attaching a diamine, then the di-HCl salt should be used except when one of the amine groups has a p*K*a < 8. In that case, we have found that the mono-HCl salts give better yields with the NMM buffer conditions used here.

8. Acidification of the solution keeps peptides protonated, which helps retain them in the aqueous phase during chloroform extraction.

Acknowledgments

This work was supported by the National Heart Lung Blood Institute (NHLBI) Proteomics Program N01-HV-28182 and NIH grant P01GM081629. We wish to thank Samuel Sondalle for preparing and analyzing the modified neurotensin samples.

References

1. Krusemark, C. J., Ferguson, J. T., Wenger, C. D., Kelleher, N. L., Belshaw, P. J. (2008) Global amine and acid functional group modification of proteins. *Anal Chem* 80, 713–720.
2. Krusemark, C. J., Frey, B. L., Belshaw, P. J., Smith, L. M. (2009) Modifying the charge state distribution of proteins in electrospray ionization mass spectrometry by chemical derivatization. *J Am Soc Mass Spectrom* 20, 1617–1625.
3. Frey, B. L., Krusemark, C. J., Belshaw, P. J., Smith, L. M. (2008) Ion-ion reactions with fixed-charge modified proteins produce ion in a single, very high charge state. *Int J Mass Spectrom* 276, 136–143.
4. Zhao, J. Y., Waldron, K. C., Miller, J., Zhang, J. Z., Harke, H., Dovicki, N. J. (1992) Attachment of a single fluorescent label to peptides for determination by capillary zone electrophoresis. *J Chromatogr* 608, 239–242.
5. Chait, B. T., Kent, S. B. H. (1992) Weighing naked proteins: practical, high-accuracy mass measurement of peptides and proteins. *Science* 257, 1885–1894.
6. Chait, B. T. (2006) Mass spectrometry: bottom-up or top-down. *Science* 314, 65–66.
7. Veenstra, T. D., Martinovic, S., Anderson, G. A., Pasa-Tolic, L., Smith, R. D. (2000) Proteome analysis using selective incorporation of isotopically labeled amino acids. *J Am Soc Mass Spectrom* 11, 78–82.
8. Du, Y., Parks, B. A., Sohn, S., Kwast, K. E., Kelleher, N. L. (2006) Top-down approaches for measuring expression ratios of intact yeast proteins using Fourier transform mass spectrometry. *Anal Chem* 78, 686–694.
9. Lee, J. E., Kellie, J. F., Tipton, J. D., Catherman, A. D., Thomas, H. M., Ahlf, D. R., Durbin, K. R., Vellaichamy, A., Ntai, I., Marshall, A. G., Kelleher, N. L. (2009) A robust two-dimensional separation for top-down tandem mass spectrometry of the low-mass proteome. *J Am Soc Mass Spectrom* 20, 2183–2191.
10. Btancia, F. L., Montgomery, H., Tanaka, K., Kumashiro, S. (2004) Guanidino labeling derivatization strategy for global characterization of peptide mixtures by liquid chromatography matrix-assisted laser desorption/ionization mass spectrometry. *Anal Chem* 76, 2748–2755.
11. Ross, P. L., Huang, Y. M., Marchese, J. N., Williamson, B., Parker, K., Hattan, S., Khainovski, N., Pillai, S., Dey, S., Daniels, S., Purkayastha, S., Juhasz, P., Martin, S., Bartlet-Jones, M., He, F., Jacobson, A., Pappin, D. J. (2004) Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. *Mol Cell Proteomics* 3, 1154–1169.
12. Schmidt, A., Kellermann, J., Lottspeich, F. (2005) A novel strategy for quantitative proteomics using isotope-coded protein labels. *Proteomics* 5, 4–15.
13. Yang, J., Gitlin, I., Krishnamurthy, V. M., Vazquez, J. A., Costello, C. E., Whitesides, G. M. (2003) Synthesis of monodisperse polymers from proteins. *J Am Chem Soc* 125, 12392–12393.
14. Abello, N., Kerstjens, H. A. M., Postma, D. S., Bischoff, R. (2007) Selective acylation of primary amines in peptides and proteins. *J Proteome Res* 6, 4770–4776.
15. Boersema, P. J., Raijmakers, R., Lemeer, S., Mohammed, S., Heck, A. J. R. (2009) Multiplex peptide stable isotope dimethyl labeling

for quantitative proteomics. *Nat Protoc* 4, 484–494.
16. Fu, Q., Li, L. (2005) De novo sequencing of neuropeptides using reductive isotopic methylation and investigation of ESI QTOF MS/MS fragmentation pattern of neuropeptides with N-terminal dimethylation. *Anal Chem* 77, 7783–7795.
17. Gidley, M. J., Sanders, J. K. (1982) Reductive methylation of proteins with sodium cyanoborohydride. Identification, suppression and possible uses of N-cyanomethyl byproducts. *Biochem J* 224, 331–334.
18. Means G. E., Feeney R. E. (1995) Reductive alkylation of proteins. *Anal Biochem* 224, 1–16.
19. Russo, A., Chandramouli, N., Zhang, L., Deng, H. (2008) Reductive glutaraldehydation of amine groups for identification of protein N-termini. *J Proteome Res* 7, 4178–4182.
20. Goodlett, D. R., Keller, A., Watts, J. D., Newitt, R., Yi, E. C., Purvine, S., Eng, J. K., Von Haller, P., Aebersold, R., Kolker, E. (2001) Differential stable isotope labeling of peptides for quantitation and de novo sequence derivation. *Rapid Comm Mass Spectrom* 15, 1214–1221.
21. Ma, M., Kutz-Naber, K. K., Li, L. (2007) Methyl esterification assisted MALDI FTMS characterization of the orcokinin neuropeptide family. *Anal Chem* 79, 673–681.
22. Fixarro, S. B., McCleland, M. L., Stukenberg, P. T., Burke, D. J., Ross, M. M., Shabanowitz, J., Hunt, D. F., White, F. M. (2002) Phosphoproteome analysis by mass spectrometry and its application to *Saccharomyces cerevisiae*. *Nat Biotech* 20, 301–305.
23. Houen, G., Svaerke, C., Barkholt, V. (1999) The solubilities of denatured proteins in different organic solvents. *Acta Chem Scand* 53, 1122–1126.
24. Hoare, D. G., Koshland, D. E., Jr (1967) A method for the quantitative modification and estimation of carboxylic acid groups in proteins. *J Biol Chem* 242, 2447–2453.
25. Sekiya, S., Wada, Y., Tanaka, K. (2004) Improvement of the MS/MS fragmentation ion coverage of acidic residue-containing peptides by amidation with ^{15}N-substituted amine. *Anal Chem* 76, 5894–5902.
26. Xu, Y., Zhang, L., Lu, H., Yang, P. (2008) Mass spectrometry analysis of phosphopeptides after peptide carboxy group derivatization. *Anal Chem* 80, 8324–8328.
27. Jacobsen, J. R., Cochran, A. G., Stephans, J. C., King, D. S., Schultz, P. G. (1995) Mechanistic studies of anti-body-catalyzed pyrimidine dimmer photocleavage. *J Am Chem Soc* 117, 5453–5461.
28. Lewis, W. G., Magallon, F. G., Fokin, V. V., Finn, M. G. (2004) Discovery and characterization of catalysts for azide-alkyne cycloaddition by fluorescence quenching. *J Am Chem Soc* 126, 9152–9153.
29. Arnold, U., Ulbrich-Hofmann, R. (1999) Quantitative protein precipitation from guanidine hydrochloride-containing solutions by sodium deoxycholate/trichloroacetic acid. *Anal Biochem* 271, 197–199.

Chapter 7

Production and Use of Stable Isotope-Labeled Proteins for Absolute Quantitative Proteomics

Dorothée Lebert, Alain Dupuis, Jérôme Garin, Christophe Bruley, and Virginie Brun

Abstract

In the field of analytical chemistry, stable isotope dilution assays are extensively used in combination with liquid chromatography-mass spectrometry (LC-MS) to provide confident quantification results. Over the last decade, the principle of isotope dilution has been adopted by the proteomic community in order to accurately quantify proteins in biological samples. In these experiments, a protein's concentration is deduced from the ratio between the MS signal of a tryptic peptide and that of a stable isotope-labeled analog, which serves as an internal standard. The first isotope dilution standards introduced in proteomics were chemically synthesized peptides incorporating a stable isotope-tagged amino acid. These isotopically labeled peptide standards, which are currently widely used, are generally added to samples after protein isolation and digestion. Thus, if protein enrichment is necessary, they do not allow correction for protein losses that may occur during sample pre-fractionation, nor do they allow the tryptic digestion yield to be taken into account. To reduce these limitations we have developed the PSAQ (Protein Standard Absolute Quantification) strategy using full-length stable isotope-labeled proteins as quantification standards. These standards and the target proteins share identical biochemical properties. This allows standards to be spiked into samples at an early stage of the analytical process. Thanks to this possibility, the PSAQ method provides highly accurate quantification results, including for samples requiring extensive biochemical pre-fractionation. In this chapter, we describe the production of full-length stable isotope-labeled proteins (PSAQ standards) using cell-free expression devices. The purification and quality control of protein standards, crucial for good-quality and accurate measurements, are also detailed. Finally, application of the PSAQ method to a typical protein quantification assay is presented.

Key words: Mass spectrometry, proteomics, absolute quantification, isotope dilution, stable isotope-labeled protein, PSAQ, cell-free expression.

1. Introduction

The field of absolute quantitative proteomics covers a wide range of analytical approaches aiming to determine protein concentrations in biological samples (1–4). Quantitative proteomic data

K. Gevaert, J. Vandekerckhove (eds.), *Gel-Free Proteomics*, Methods in Molecular Biology 753,
DOI 10.1007/978-1-61779-148-2_7, © Springer Science+Business Media, LLC 2011

are required for specific applications including biomarker research (5–7) and systems biology, where patterns of protein expression levels have to be quantitatively characterized before data integration and mathematical modeling (8, 9).

Although elemental analysis by inductively coupled plasma-mass spectrometry (ICP-MS) is gaining interest for absolute protein quantification (10, 11), most quantitative experiments are performed using LC-MS on peptides released by trypsin proteolysis of protein samples (12). These peptide digests are highly complex; therefore, many strategies have been developed to improve the detection of moderate to low-abundance proteins in biological samples. These developments involve most of the steps from sample preparation to analysis. MS analytical performance has been significantly enhanced by the introduction of targeted analyses enabling the selective monitoring of defined peptide sets. One of these applications, known as SIM (selected ion monitoring), allowed Hanke et al. to quantify 150 amol of protein in a cell lysate without the need for any fractionation prior to LC-MS (13). The SRM (selected reaction monitoring) mode of analysis, and particularly time-scheduled SRM, has rendered the monitoring of several hundreds of proteolytic peptides in a single analysis possible (14, 15). Recently, Picotti et al. harnessed the specific advantages of SRM to detect and quantify 100 proteins over the full dynamic range of the yeast proteome and reached protein abundances down to 50 protein copies/cell without sample enrichment (16). However, in samples which display a very wide dynamic range of protein concentrations, such as plasma (17), SRM encounters sensitivity problems (5). To help treat this type of sample, various pre-fractionation strategies significantly improving the overall sensitivity of the MS-based quantification process have been developed. These strategies include the following approaches: (i) immunoenrichment of target proteins (18) or proteolytic peptides (19); (ii) depletion of abundant proteins (20, 21); (iii) isolation of subproteomes, such as N-glycosylated peptides (22) or N-terminal peptides (23); and (iv) off-gel isoelectrofocusing (24).

Recently, novel strategies attempting to evaluate protein abundance at a proteome-wide scale have emerged (25). Nevertheless, most absolute quantitative proteomic analyses focus on protein sets presenting a particular biological interest. These experiments frequently use internal standards and are generally combined with targeted MS assays such as SRM. Classically, one or several standards are selected for each protein to be quantified. Several types of standards are available and can be classed according to their sequence length (peptide/polypeptide/protein) and the specific modification which allows them to be distinguished from their endogenous counterparts during mass spectrometry analysis. Various modifications have been used to generate internal standards: (i) chemical derivatization (26, 27); (ii) slight

sequence modifications or use of a protein analogue from another species (28); and (iii) stable isotope labeling (1). Absolute quantitative proteomics mainly uses stable isotope-labeled internal standard-based approaches. Standards are generally tagged with ^{13}C and/or ^{15}N on a particular amino acid residue. This labeling results in identical behavior of the standard and the target endogenous tryptic peptides during chromatography, ionization, and fragmentation in the mass spectrometer. However, the isotope-labeled standard peptide ("heavy" peptide) and the corresponding endogenous tryptic peptide ("light" peptide) can be unambiguously distinguished during MS analysis. Quantification is performed by comparing the relative intensities of the "light" and "heavy" peptide signals. Three types of standards can be used in stable isotope dilution strategies; these have been fully reviewed in (1) and will be briefly described here. The first, most commonly used technique for MS-based protein absolute quantification is based on AQUA (absolute quantitation) standards. These are chemically synthesized stable isotope-labeled peptides which are added to samples in defined quantities before LC-MS analysis (29). The AQUA method is very straightforward to perform, easily multiplexed, and enables quantitative analysis of protein phosphorylation. Conversely, it is the least accurate protein quantification method when protein enrichment is necessary or in cases where tryptic digestion is incomplete. Indeed, as AQUA peptides are added just prior to LC-MS analysis, quantification accuracy is severely compromised in experimental setups where any protein is lost during the process of biochemical pre-fractionation or when partial protein digestion is an issue (13, 30, 31). A second option for absolute quantification is based on the use of QconCAT standards. These are artificial concatemers of stable isotope-labeled peptides which can be spiked into samples before trypsin digestion. QconCAT standards are especially suited to the quantification of numerous proteins or protein complexes in a single multiplexed assay (32). Like the AQUA approach, this quantification strategy lacks accuracy when sample pre-fractionation is performed (30). Finally, the PSAQ (protein standard absolute quantification) method makes use of full-length isotope-labeled proteins as quantification standards. PSAQ standards can be added at very early stages of the analytical process and display specific advantages over peptide or polypeptide standards. A PSAQ approach is particularly recommended for the accurate quantification of low-abundance proteins which require significant sample pre-fractionation (30). Indeed, of the three isotope dilution strategies commonly used, PSAQ is the only one allowing accurate quantification of this set of proteins. In addition, because they correspond to full-length proteins, PSAQ standards offer maximal protein coverage for quantification. This allows many tryptic peptides to be monitored and renders this approach especially valuable for isoform discrimination and quantification (33, 34).

Finally, PSAQ standards are also applicable for the quantification of proteins displaying post-translational modifications; Ciccimaro et al. have recently demonstrated their use for the efficient characterization and quantification of protein phosphorylation (35).

This chapter presents a detailed description of the protocol for synthesizing, labeling, and purifying a full-length stable isotope-labeled protein standard (PSAQ standard). We describe the production of these standards using a cell-free (in vitro) expression system. Steps to control the quality of the protein standard, including verification of isotope labeling and accurate calibration, are also described. These steps are crucial for good-quality experiments allowing accurate quantification. Finally, an example of absolute protein quantification using a full-length isotope-labeled protein as reference is presented.

2. Materials

2.1. Gene Cloning: Incorporating the Target Gene into an Expression Vector

1. RTS pIVEX His-tag, 2nd Generation Vector Set (RiNA GmbH) (*see* **Note 1**).
2. Specific primers for cloning (Eurogentech).
3. DNA template encoding the target protein, e.g., cDNA library or DNA plasmid.
4. DNA polymerase with proofreading activity such as iProof™ DNA polymerase (Bio-Rad).
5. PCR reagents: dNTP, polymerase buffer.
6. 0.2 or 0.5 ml thin-wall microtubes for PCR reaction.
7. Thermocycler for DNA amplification.
8. Agarose (SeaKem® GTG® Agarose, Cambrex) and low-melting agarose (SeaPlaque® Agarose, Cambrex).
9. Tris-borate EDTA (TBE) buffer 10X (Gibco, Invitrogen).
10. DNA electrophoresis apparatus.
11. Restriction enzymes (New England Biolabs).
12. PCR/linear DNA extraction and purification kit (Nucleospin Extract II, Macherey-Nagel).
13. Rapid DNA Ligation Kit (Roche).
14. XL1-Blue competent cells (Stratagene).
15. Heating block.
16. SOC medium (Invitrogen).
17. Petri dishes.
18. 37°C incubator for plated bacterial culture.
19. LB medium.

20. Ampicillin.
21. Shaking incubator for suspension bacterial culture (37°C).
22. Plasmid DNA Mini-prep kit (Nucleospin® Plasmid, Macherey-Nagel).
23. T7 promoter primer [5′-primer: 5′-TAATACGACTCAC TATAGGG-3′] and T7 terminator primer [3′-primer: 5′-GCTAGTTATTGCTCAGCGG-3′] (Eurogentech).
24. Vector NTI Advance™ Software.
25. Sterile glycerol.
26. Cryotubes.
27. Clean latex or nitrile gloves.

2.2. Cell-Free Expression of Stable Isotope-Labeled Protein

1. RNaseKiller (5 PRIME).
2. LB medium.
3. Ampicillin.
4. Shaking incubator.
5. Plasmid DNA Midi-prep kit (Nucleobond® Xtra Midi Plus, Macherey-Nagel).
6. Isopropanol (*see* **Note 2**).
7. Molecular Biology grade, RNase-free H_2O (Eppendorf).
8. Ethanol/RNase-free water, 70/30, v/v (*see* **Note 2**).
9. Spectrophotometer (260 nm).
10. RTS Amino Acid Sampler (RiNA GmbH).
11. Isotopically labeled amino acids (Cambridge Isotope Laboratories, Inc.) (*see* **Notes 3** and **4**).
12. RTS 500 Proteomaster *Escherichia coli* HY kit (RiNA GmbH).
13. Proteomaster Instrument (Roche) or MTP Thermoblock (Eppendorf) equipped with an RTS 500 Adapter (RiNA GmbH).
14. Clean latex or nitrile gloves.

2.3. Stable Isotope-Labeled Protein Dialysis and Analysis of Solubility

1. Dialysis cassettes, 1–3 ml, 3,500–10,000 Da.
2. Dialysis buffer: 20 mM Tris-HCl pH 7.5, 0.5 M NaCl, 20 mM imidazole.
3. Microtube centrifuge.
4. Acetone.
5. Equipment and reagents for SDS-PAGE analysis.
6. 1.5 ml microtubes
7. Clean latex or nitrile gloves.

2.4. Purification of Stable Isotope-Labeled Protein

1. Ni Sepharose resin (Ni Sepharose™ 6 Fast Flow, GE Healthcare).
2. Equilibration buffer and elution buffers for purification of soluble stable isotope-labeled protein:
   - Equilibration buffer: 20 mM Tris-HCl pH 7.5, 0.5 M NaCl, 20 mM imidazole.
   - A series of elution buffers: 20 mM Tris-HCl pH 7.5, 0.5 M NaCl, gradient from 25 to 500 mM imidazole.
3. Equilibration buffer and elution buffers for purification of insoluble stable isotope-labeled protein:
   - Equilibration buffer: 6 M guanidine, 20 mM Tris-HCl pH 7.5, 0.5 M NaCl, 20 mM imidazole.
   - A series of elution buffers: 6 M guanidine, 20 mM Tris-HCl pH 7.5, 0.5 M NaCl, gradient from 25 to 500 mM imidazole.
4. Disposable empty polypropylene columns (Poly-prep chromatography columns, Bio-Rad).
5. Equipment and reagents for SDS-PAGE analysis.
6. Renaturation buffer: 4 M guanidine, 20 mM Tris-HCl pH 7.5, 0.5 M NaCl.
7. Ammonium bicarbonate.
8. pH-meter.
9. Dialysis cassettes, 1–3 ml, 3,500–10,000 Da.
10. Ultrafiltration devices: Vivaspin 2, Hydrosart, 5 kDa (Sartorius Stedim).
11. Bicinchoninic acid protein assay kit.
12. Spectrophotometer (540 nm).
13. Low-adsorption tubes (SafeSeal Microcentrifuge tubes, low binding polymer technology, 0.5 ml, Sorenson Bioscience Inc.).
14. CHAPS.
15. Lyophilizer.
16. 1.5 ml microtubes.
17. Clean latex or nitrile gloves.

3. Methods

Stable isotope-labeled proteins can be produced using various types of expression systems including cell-free systems (13, 30), bacterial expression systems (13, 33), or expression in mammalian

cells (36). Here, we describe the production of a stable isotope-labeled protein (PSAQ standard) using cell-free extracts, an expression strategy also known as in vitro translation. Cell-free expression is especially suited to the incorporation of an isotopic label into proteins as follows: (i) the endogenous pools of unlabeled amino acids, which could lead to reduced isotope incorporation yields, are depleted as part of the cell-free extract preparation process; (ii) the metabolism of isotope-labeled amino acids is significantly reduced; this helps to avoid isotope dispersion through metabolic conversion of labeled amino acids. Due to these specific features very high isotope incorporation levels (>95%) can be achieved.

Various cell-free expression systems are commercially available (*see* **Note 5**). In this chapter, we describe the synthesis of a stable isotope-labeled protein using the RTS 500 Proteomaster *E. coli* HY kit (RiNA GmbH) which we have found to provide a good success rate for protein expression (18 proteins tested for expression) (**Fig. 7.1**). However, depending on the target protein, other lysates, such as wheat germ lysates, may provide better expression rates. As a general rule, prokaryotic proteins should be preferentially expressed using *E. coli* lysates. For human and other eukaryotic proteins, we advise a first expression attempt using *E. coli*

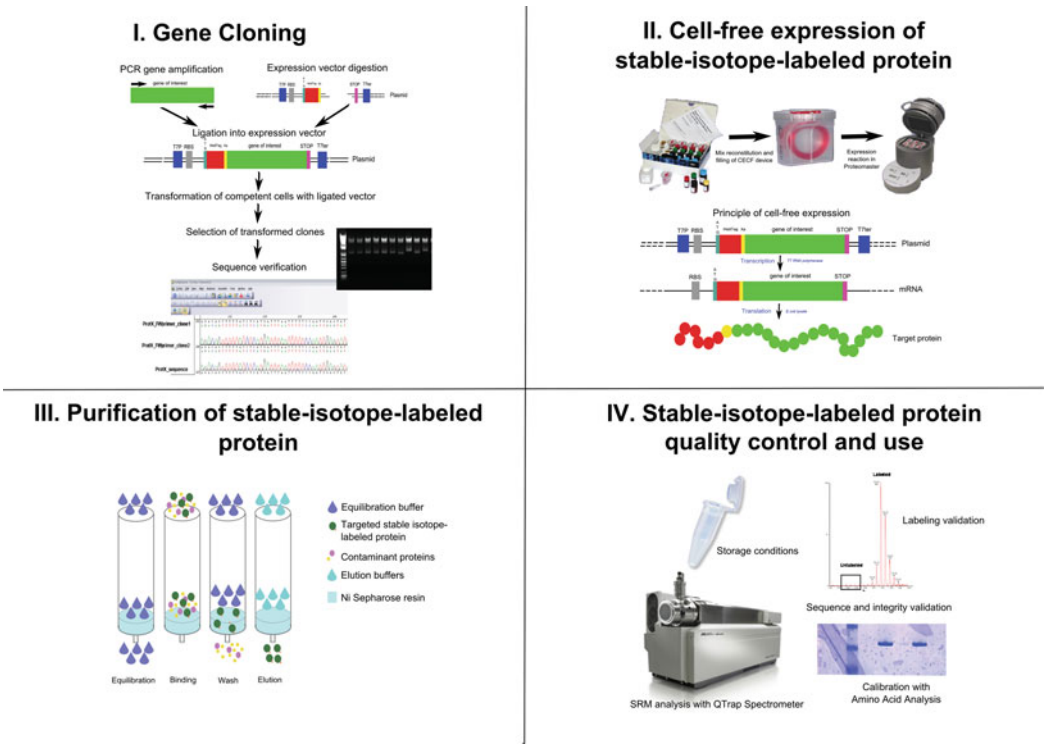


Fig. 7.1. Production and use of a stable isotope-labeled protein standard for absolute quantitative proteomic analysis.

lysates. If expression using this lysate is unsuccessful, a second attempt can be made using wheat germ lysates in which the translational machinery is more suited to the expression of eukaryotic genes.

3.1. Gene Cloning: Incorporating the Target Gene into an Expression Vector

RTS pIVEX vectors have been optimized for high-level protein expression using RTS cell-free systems. The protocol described below uses the 2.4d pIVEX plasmid. This plasmid has been specifically developed for protein expression using a combination of T7 RNA polymerase and RTS *E. coli* lysates. The 2.4d pIVEX vector allows fusion of a hexahistidine purification tag at the N-terminal extremity of the expressed protein (*see* **Notes 6** and **7**). If necessary, this N-terminal hexahistidine tag can be removed by factor Xa proteolysis. Due to the incorporation of this factor Xa cleavage site between the target protein sequence and the hexahistidine tag (*see* **Note 8**), the 2.4d pIVEX vector allows a non-tagged protein to be obtained **(Fig. 7.2)**.

3.1.1. Primer Design for PCR Gene Amplification

Design forward and reverse primers of about 20 bases following the 2.4d pIVEX plasmid booklet instructions. Both 5′-primer ends must incorporate a restriction enzyme cleavage site, and these two sites must be for distinct restriction enzymes (*see* **Note 9**).

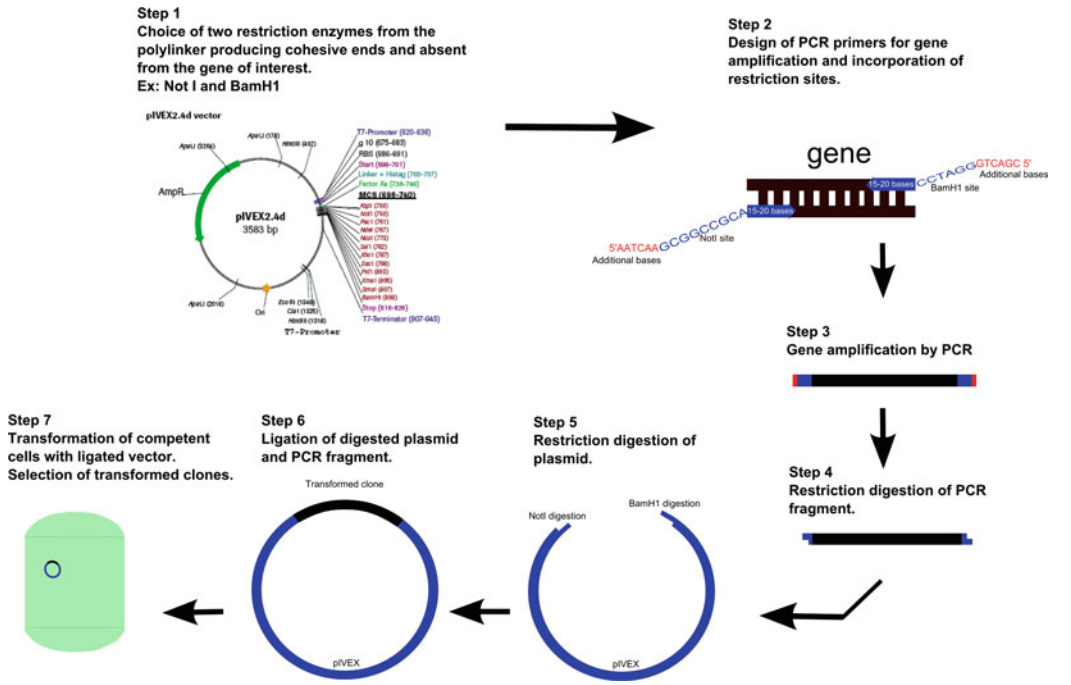


Fig. 7.2. Cloning a target gene into the pIVEX2.4d vector.

1. Design primers with comparable melting temperatures (Tm), typically 50–60°C.
2. Perform the PCR reaction with the DNA template carrying the gene encoding the target protein. The PCR experiment should be designed taking the DNA polymerase supplier's instructions as a guide. **Tables 7.1** and **7.2** describe a typical PCR reaction using the commercial Iproof™ High-Fidelity polymerase.

3.1.2. PCR Product Restriction Digestion

Control completion of the PCR reaction by DNA electrophoresis on an aliquot of the reaction mixture using a 1% agarose gel and 1X TBE buffer.

1. Use the pre-selected restriction enzymes to digest the PCR fragment. This digestion can be performed in two successive steps or with the two enzymes simultaneously. Digestion buffers and conditions should be selected according to the manufacturer's instructions. In the case of a simultaneous digestion it is important to use an appropriate

Table 7.1
Typical PCR reaction mix using Iproof™ high-fidelity Taq polymerase

Reagent	Volume for a 20 µl reaction	Final concentration
Iproof™ HF buffer (5X)	4 µl	1X
dNTP (dATP, dCTP, dTTP, dGTP) (25 mM each)	1.6 µl	0.5 mM each
Forward primer (10 µM)	1 µl	0.5 mM
Reverse primer (10 µM)	1 µl	0.5 mM
Molecular biology grade H_2O	11.2 µl	
DNA template (1 ng/µl)	1 µl	
Iproof™ polymerase	0.2 µl	

Table 7.2
Sample PCR program

Cycle step	Temperature (°C)	Time	Number of cycles
Initial denaturation	98	60 s	1
Denaturation	98	15 s	30
Annealing	45–72	20 s	30
Extension	72	30 s (30 s/kb)	30
Final extension	72	5 min	1

buffer, compatible with both enzymes. Information regarding enzyme and reaction buffer compatibility is available in the New England BioLabs® Inc., catalog.

2. Purify the digested PCR fragment using a DNA extraction and purification kit and analyze by agarose gel electrophoresis.

3.1.3. Restriction Digestion of the 2.4d pIVEX Vector

1. Digest the 2.4d pIVEX vector with the restriction enzymes used in **Section 3.1.1** for PCR fragment digestion. As above, appropriate digestion buffers and conditions should be applied.
2. Check plasmid digestion by loading a small aliquot of the digestion mixture (1/10 of total volume) onto a 1% agarose gel. Carry out DNA electrophoresis in 1X TBE buffer.
3. If digestion is complete, load the remaining sample onto a 1% low-melting agarose gel to separate the linearized vector from any remaining undigested vector or other, smaller DNA fragments that may be present.
4. Excise a band surrounding the linearized 2.4d pIVEX from the gel.
5. Extract the linearized vector using a DNA extraction and purification kit.

3.1.4. PCR Product Ligation into Linearized 2.4d pIVEX Vector

Ligate the digested and purified PCR fragment into the linearized 2.4d pIVEX vector using the Rapid DNA Ligation Kit according to the procedure described below. A control ligation reaction should also be performed:

1. Dissolve the purified 2.4d pIVEX vector and PCR fragment in 1X DNA dilution buffer (included in the Rapid DNA Ligation Kit) in a final volume of 10 μl (*see* **Note 10**).
2. Add 10 μl of vortexed T4 DNA ligation buffer to the reaction vial. Mix thoroughly.
3. Add 1 μl of T4 DNA ligase to the reaction mix. Mix thoroughly.
4. Incubate for 5–10 min at 25°C.
5. Proceed directly to *E. coli* transformation.

3.1.5. E. coli Transformation with 2.4d pIVEX Vector

For bacterial transformation with the ligation product we use XL1-Blue Competent Cells (*see* **Note 11**). For a detailed transformation protocol for your competent cells, refer to the supplier's instructions. An outline of the main experimental steps for bacterial transformation is presented below. A control transformation reaction should also be performed.

1. Thaw the cells on ice. When thawed, gently mix and transfer 100 µl of cells to a pre-chilled 1.5 ml sterile microtube.
2. Add 1–20 µl of the ligation reaction to the aliquot of cells and swirl the tube gently (use of more than 1/10 of the volume of the ligation reaction mix for the transformation assay is not recommended, *see* **Note 12**).
3. Incubate the tube on ice for 30 min.
4. Heat-pulse the tube at 42°C for 45 s. The duration of this pulse is critical.
5. Cool the tube on ice for 2 min, then add 0.9 ml of pre-heated (42°C) SOC medium and incubate the tube at 37°C for 1 h with gentle shaking.
6. Spread 150–200 µl of the transformation reaction on an agar plate supplemented with ampicillin (100 µg/ml). Allow bacteria to grow overnight at 37°C.

3.1.6. Selection of Transformed Clones

1. Pick a few isolated clones from the agar plates. For each clone, start a 3 ml culture in ampicillin (100 µg/ml)-supplemented LB medium. Grow the cultures overnight at 37°C in a shaking incubator.
2. Extract and purify pIVEX 2.4d plasmid DNA from each mini-culture using a plasmid DNA Mini-prep kit (*see* **Notes 13** and **14**).
3. Select two transformed clones for each target gene and verify the insert sequence and orientation by DNA sequencing (*see* **Note 15**).
4. Store 2.4d pIVEX-transformed clones in sterile cryotubes. For this, mix 300 µl of a sterile 50% glycerol/water (v/v) solution with 700 µl of a 1 ml ampicillin (100 µg/ml)-supplemented LB culture of the selected clone.
5. Freeze clearly labeled cryotubes at –80°C for future use.

3.2. Cell-Free Expression of Stable Isotope-Labeled Protein

Throughout the following experiments, extreme care should be taken at all handling steps to prevent RNase contamination. Wear clean latex or nitrile gloves and use only RNase-free solutions. All work surfaces should also be cleaned using RNaseKiller solution.

3.2.1. Preparation of DNA Template for Cell-Free Protein Expression

1. Inoculate 100 ml of ampicillin-supplemented LB medium with XL1 blue cells transformed with the pIVEX2.4d vector into which the gene of interest has been inserted.
2. Grow the bacterial cultures overnight at 37°C in a shaking incubator.

3. Extract and purify the pIVEX2.4d plasmid using a plasmid DNA Midi-prep kit according to the manufacturer's instructions.
4. Elute the purified plasmid DNA with RNase-free Molecular Biology Grade water. Store at –20°C for future use.
5. 10–15 μg of purified insert-containing pIVEX2.4d vector is required for each RTS 500 reaction.

3.2.2. Preparation of Amino Acid Mix for Stable Isotope Labeling

The RTS Amino Acid Sampler contains stock solutions of the 20 amino acids necessary for cell-free translation. PSAQ standards are generally labeled on arginine and lysine residues (*see* **Note 3**). Consequently, when making the amino acid mix, the unlabeled arginine and lysine solutions must be replaced by stable isotope-labeled arginine and lysine solutions.

1. Prepare 168 mM stock solutions of labeled arginine and lysine by re-suspending each lyophilized amino acid in an appropriate volume of RTS re-suspension buffer. These solutions should be stored at –20°C in 150 μl aliquots.
2. To prepare the "Amino Acid mix" (*see* **Note 16**), pre-heat each amino acid solution at 37°C for 30 min with vigorous shaking. Mix 150 μl of each of the 18 unlabeled amino acid solutions from the RTS Amino Acid Sampler (*see* **Note 17**) and add 150 μl of labeled arginine and labeled lysine. Finally, add 330 μl of 40 mM DTT (*see* **Note 18**).

3.2.3. Stable Isotope-Labeled Protein Expression Using the Cell-Free Reaction Device

The following steps are based on the RTS *E. coli* Proteomaster HY kit instruction manual.

1. Reconstitute the RTS *E. coli* Proteomaster HY kit solutions: the "*E. coli* lysate" with 0.525 ml of RTS reconstitution buffer, the "reaction mix" with 0.25 ml RTS reconstitution buffer, and the "feeding mix" with 8.1 ml of RTS reconstitution buffer.
2. Prepare the "feeding solution" by adding 3 ml of the "amino acid mix" into the bottle containing the "feeding mix."
3. Prepare the "reaction solution," by adding 225 μl of "reaction mix," 300 μl of "amino acid mix," and 10–15 μg of DNA template (suspended in no more than 30 μl) to the "*E. coli* lysate" bottle.
4. Fill the RTS 500 CECF reaction device (**Fig. 7.3**): fill the reaction chamber with 1 ml of "reaction solution" and feeding chamber with 10 ml of "feeding solution."
5. Place the RTS 500 CECF reaction device in the Proteomaster instrument (or the appropriately adapted thermomixer).
6. Allow protein expression to proceed for 24 h at 900 rpm and 30°C (*see* **Notes 19**, **20,** and **21**).

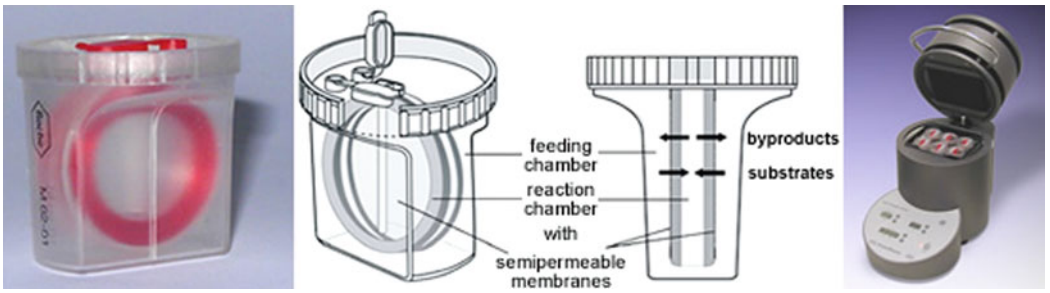


Fig. 7.3. The RTS 500 CECF reaction device.

3.3. Stable Isotope-Labeled Protein Dialysis and Verification of Solubility

3.3.1. Preparation of Dialysis Buffer and Dialysis

1. Prepare at least 2 l of pre-chilled dialysis buffer.
2. Fill a 1 ml dialysis cassette (MWCO 3,500 Da) with the contents of the RTS 500 reaction chamber.
3. Dialyze against 2 l of equilibration buffer overnight at 4°C.

3.3.2. Separation of Soluble and Insoluble Fractions

1. Transfer the dialyzed solution to a microtube and spin at 11,000×*g*, 4°C, for 10 min to separate the soluble and insoluble fractions. Pipette the supernatant carefully and transfer it to a clean microtube. Keep both fractions on ice.
2. Using a disposable pipette tip, take a small fraction of the insoluble fraction and dissolve it in SDS-PAGE loading buffer.
3. Precipitate soluble proteins by adding 100 µl of acetone to a 10 µl aliquot of the soluble fraction. Incubate at –20°C for 1 h (*see* **Note 22**). Spin for 10 min at 14,000×*g* (4°C) and then remove the acetone supernatant. Dry the pellet in air and re-suspend it in SDS-PAGE loading buffer.
4. Run these samples on an SDS-PAGE gel to check protein expression in both fractions.
5. If the target protein is expressed in the soluble fraction, follow the purification protocol described in **Sections 3.4.1** and **3.4.2**.
6. If the target protein is found in the insoluble fraction, purification must be carried out under denaturing conditions by adding 6 M of guanidium HCl to the purification buffers. In this case, the purification protocol is detailed in **Sections 3.4.1** and **3.4.3**.

3.4. Purification of Stable Isotope-Labeled Proteins

Stable isotope-labeled proteins produced with a hexahistidine tag can easily be purified using a protocol based on immobilized metal affinity chromatography (IMAC).

3.4.1. Preparation of IMAC Resin

1. Wash 100 μl of Ni Sepharose 6 Fast Flow resin slurry with 500 μl of equilibration buffer.
2. Spin using a bench-top centrifuge and discard the supernatant.
3. Proceed immediately to purification step (do not allow the resin to dry out).

3.4.2. Purification of Soluble Stable Isotope-Labeled Proteins

1. Re-suspend the washed resin in the soluble protein extract and incubate for 1 h at 4°C on a rotating wheel at 10 rpm.
2. Transfer the solution with the resin into an empty Poly-prep chromatography column and eliminate the liquid by gravity flow. Keep an aliquot of the flow-through to serve as a binding control.
3. Wash the IMAC resin first with 2–5 ml of equilibration buffer and then with 2–5 ml of equilibration buffer. Keep an aliquot of the wash solution as a control (store on ice).
4. Elute IMAC-bound protein by washing the resin with 500 μl of each elution buffer, starting with the buffer containing the lowest imidazole concentration and working through to the highest (*see* **Note 23**). Keep all eluted fractions on ice.
5. Perform SDS-PAGE analysis of all individual fractions.
6. Pool the fractions containing the target protein, change the buffer to 25 mM ammonium bicarbonate by dialysis, and proceed to protein concentration (**Section 3.4.5**).

3.4.3. Purification of Insoluble Stable Isotope-Labeled Proteins

When the stable isotope-labeled protein is expressed in the insoluble fraction, a chaotropic agent must be added to buffers to allow protein solubilization and purification.

1. Solubilize the insoluble proteins in 0.5–1 ml of equilibration buffer. Store on ice for 1 h, centrifuge at 11,000×*g* for 5 min, and recover the supernatant.
2. Re-suspend the washed IMAC resin in the re-dissolved protein extract and incubate for at least 1 h at 4°C under rotation.
3. Transfer the contents of the tube to an empty Poly-prep chromatography column and allow liquid to flow through.
4. Recover the flow-through and keep an aliquot as a control of protein binding.
5. Wash the IMAC resin with 2–5 ml of equilibration buffer (allow the wash buffer to run off the column by gravity flow), wash a second time with 2–5 ml equilibration buffer, and keep an aliquot of the wash solution as a control. Store on ice.

6. Elute IMAC-bound proteins by washing the resin with 500 µl of each elution buffer, starting with the buffer containing the lowest imidazole concentration and working up to the highest. Store all eluted fractions on ice.
7. Perform SDS-PAGE analysis of all recovered fractions (*see* **Note 24**).
8. Pool the fractions containing the target protein and proceed to protein renaturation (**Section 3.4.4**) before concentrating the protein solution (**Section 3.4.5**).

3.4.4. Renaturation of Stable Isotope-Labeled Protein

1. Prepare chilled renaturation buffer.
2. Pool the eluted fractions containing the purified isotope-labeled protein and fill a dialysis cassette.
3. Dialyze for 4 h at 4°C.
4. Allow the protein to refold by gradually decreasing the guanidine concentration of the dialysis buffer: 4, 2, 1, 0.5, 0.25, 0.1, 0.05, 0 M changing dialysis buffer every 4 h.
5. Finally, dialyze the protein solution against 25 mM ammonium bicarbonate, pH 7.5.
6. Before storage, load a fraction on an SDS-PAGE gel to check protein purity.

3.4.5. Concentrating Stable Isotope-Labeled Protein Solutions

1. Concentrate the solution containing the stable isotope-labeled proteins using Vivaspin 2 Hydrosart ultrafiltration devices according to the manufacturer's protocol.
2. Following concentration, proceed to protein storage (**Section 3.4.8**).

3.4.6. Evaluation of Protein Quantities for Stable Isotope-Labeled Standards

To estimate the amount of purified stable isotope-labeled protein, perform a colorimetric protein concentration assay, such as BCA (*see* **Note 25**). It can also be useful to measure UV absorption at λ_{280}. Use the predicted molar extinction coefficient to calculate the concentration of the protein (http://www.expasy.org/tools/protparam.html).

3.4.7. Storage of Stable Isotope-Labeled Protein Standards

For optimal storage, the use of special low-adsorption plastic tubes is recommended to avoid protein loss through adsorption. Biochemical properties of each protein (oligomerization, solubility, and hydrophobicity) are highly variable and there is no "standard" storage buffer. Therefore, before long-term storage at –80°C, we recommend testing several storage conditions for each stable isotope-labeled protein. Three main parameters should be tested:

1. pH of the storage buffer. This can be acidic or basic (*see* **Note 26**), but best results are commonly found with buffers

of pH distinct from the specific p*I* of the stable isotope-labeled protein.

2. Lyophilization, which generally avoids protein degradation, is recommended. However, some proteins may be better stored in solution.
3. Non-denaturing detergents such as CHAPS (0.5%) can also be added.

3.5. Quality Control: Verification of Label Incorporation, Accurate Quantification

Stable isotope-labeled proteins (PSAQ standards) have to be carefully checked before use. In addition to verification of protein sequence and labeling efficiency, other features may have to be controlled. Indeed, depending on the sample pre-fractionation to be performed on the biological sample, some "critical" biochemical properties involved in the partitioning process must be identical between the PSAQ standard and the protein to be quantified. Control of these biochemical characteristics is imperative.

1. The protein sequence can be verified either directly by mass spectrometry analysis (MS/MS mode) after trypsin digestion of the purified stable isotope-labeled protein (*see* **Note 27**) or indirectly by in silico translation of the oligonucleotide sequence of the pIVEX vector containing the gene of interest.
2. Stable isotope incorporation is estimated after trypsin digestion of the purified stable isotope-labeled protein using mass spectrometry analysis, either in MS mode or SRM mode. This experiment verifies that the isotope-labeled protein does not generate unlabeled tryptic peptides which could adversely affect quantification results. The minimal level of labeling required for a stable isotope-labeled protein standard is 95%. All PSAQ standards that we have produced using the previously described protocol displayed over 98% isotope incorporation.
3. Protein folding may be checked by circular dichroism analysis.
4. Non-denaturing gel electrophoresis is suitable to control for protein oligomerization.
5. Finally, accurate quantification of the stable isotope-labeled protein should be carried out by amino acid analysis (*see* **Note 28**). A single amino acid analysis generally requires between 50 and 100 μg of purified protein.

3.6. Use of Stable Isotope-Labeled Proteins for Absolute Quantification

To illustrate the use of stable isotope-labeled proteins for absolute quantification, we describe below the development of a typical "PSAQ" protein quantification assay. Each protein to be quantified should be synthesized in a stable isotope-labeled form (PSAQ standard), highly purified and accurately quantified, as described above.

3.6.1. Selection of Proteotypic Peptides

For each target protein, signature peptides released by trypsin digestion (so-called proteotypic peptides) must be selected. In addition to representing a unique amino acid sequence (*see* **Note 29),** these peptides must also be easily detectable by MS. To select proteotypic peptides, we typically digest the PSAQ standards or unlabeled target proteins (when these are commercially available) and perform LC-MS analysis on samples at very low concentrations. When the final analysis is to be carried out in SRM mode, we also select the best SRM transitions to be monitored for each proteotypic peptide.

3.6.2. Determining the Analytical Performances of the Quantification Assay

Having selected the proteotypic peptides, a recovery experiment is designed and carried out for each target protein (*see* **Note 30**). This experiment consists in spiking increasing amounts of each target protein into the matrix to be analyzed. Using the corresponding PSAQ standards and LC-MS analysis, the amounts of protein added are then estimated. To design a recovery experiment, several considerations must be taken into account:

1. The concentration range for each target protein tested should be selected according to the biological question addressed. For example, to develop an absolute quantitative assay for a disease protein biomarker, the concentration range must vary between the physiological level and the highest pathological level.
2. Ideally, the spiked matrix should be "analyte free," but very often this is not possible; in these cases, unknown quantities of endogenous target proteins are present in the matrix (*see* **Note 31**).
3. After spiking the unlabeled target protein into the matrix, the corresponding stable isotope-labeled protein (PSAQ standard) must be added in defined amounts to serve as internal reference. We generally spike a constant concentration of PSAQ standard. Increased assay sensitivity can be achieved by spiking the PSAQ standard at relatively high concentrations. However, we advise that the ratio between the highest target protein concentration and the PSAQ standard concentration not be greater than 5, since otherwise quantification accuracy may be affected.
4. After spiking target proteins and the corresponding PSAQ standards into the test sample, proceed to biochemical prefractionation, trypsin digestion, and LC-MS analysis.
5. From LC-MS data and in reference to internal standard MS signals, it is possible to re-estimate the concentrations of the spiked unlabeled protein. From these results, for each proteotypic peptide monitored, a titration curve can be obtained (**Fig. 7.4**). This curve indicates the linearity of quantification in the range of concentrations tested, the

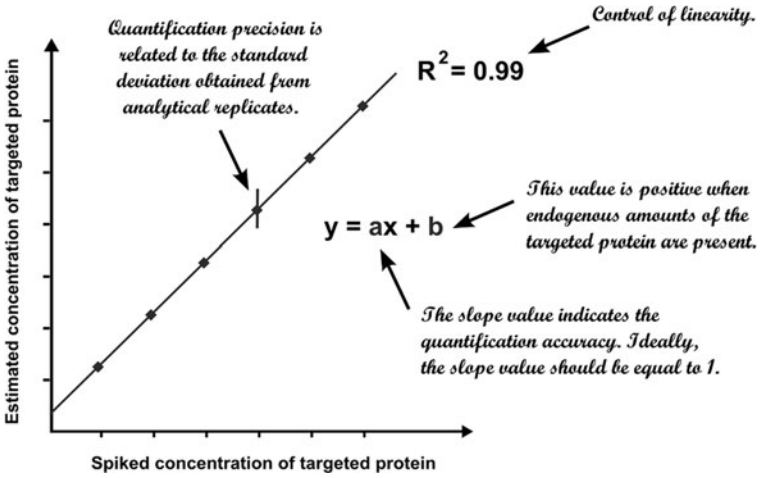


Fig. 7.4. Recovery experiment and titration curve.

quantification accuracy, and the quantification precision if each concentration point was replicated.

6. The limit of detection (LOD) and limit of quantification (LOQ) can be estimated using signal to noise ratios; these are generally given as 3/1 and 10/1, respectively.

3.6.3. Analysis of Biological Samples

1. Collect or prepare the samples to be analyzed (biological fluids, cell lysates, etc.)
2. As soon as possible after collection, spike the PSAQ standard(s) in defined amounts into the samples.
3. Allow the endogenous protein target(s) and the PSAQ standard(s) to equilibrate before starting biochemical prefractionation (1 h at 4°C is generally sufficient).
4. Proceed to biochemical treatment, trypsin digestion, and LC-MS analysis.
5. From LC-MS data, and relative to the PSAQ concentration, the concentration of the target proteins in the samples can be deduced.

4. Notes

1. Two different expression vectors are supplied with each RTS 500 expression kit: pIVEX2.3d allows the target protein to be fused with a hexahistidine purification tag at its

C-terminus, whereas pIVEX2.4d provides an N-terminal tag in combination with a factor Xa protease cleavage site.

2. To prevent RNase contamination, isopropanol and ethanol solutions should exclusively be used for this application.
3. When trypsin protein proteolysis is performed before LC-MS analysis, we recommend using L-lysine:2 HCl (U-13C6, 97-99%; U-15N2, 97-99%) and L-arginine:HCl (U-13C6, 97–99%; U-15N4, 97–99%) to label PSAQ standards. This particular label combination leads to "singly labeled" tryptic peptides with a constant mass increment, greatly simplifying LC-MS data processing.
4. Other stable isotope labeling strategies can be used, depending on proteolysis specificity. For example, if EndoLysC is used for proteolytic cleavage, it is appropriate to label only lysine residues.
5. Qiagen, Promega, and CellFree Sciences also provide alternative cell-free expression systems.
6. The pIVEX 2.3d vector provides a hexahistidine tag located at the C-terminal end of the expressed protein. In contrast to the N-terminal tag, the C-terminal tag cannot be removed.
7. The pIVEX 2.3d and 2.4d vectors allow fusion of the expressed protein with a hexahistidine purification tag. Other pIVEX vectors are available providing maltose-binding protein tag; hemagglutinin tag; or glutathione *S*-transferase tag. We generally choose a hexahistidine tag as it allows simple and efficient protein purification using immobilized metal affinity chromatography (IMAC). In addition, this small tag rarely affects target protein folding. Larger tags, such as maltose-binding protein or glutathione *S*-transferase, are generally used to increase the apparent solubility of the expressed protein.
8. Removal of the hexahistidine tag is generally not required to obtain similar behavior between the stable isotope-labeled protein and the endogenous protein target during sample pre-fractionation. However, it can be crucial in certain cases, typically when immunocapture is performed with an antibody directed against the N-terminal extremity of the target protein.
9. Cleavage at restriction sites will be improved if 5–6 additional bases are added. Other key elements to be taken into consideration for the selection of restriction enzymes are as follows: (i) the 5′- and 3′-restriction sites must be absent from the DNA sequence of the gene to be cloned; (ii) the inserted restriction sites should, however, be present

in the polylinker of the pIVEX plasmid; and (iii) enzymes that produce cohesive ends are preferable as they favor the correct orientation of the insert upon ligation with the plasmid.

10. For a standard assay, mix around 50 ng of linearized pIVEX vector with 150 ng of insert DNA.
11. Other competent cell strains can be used such as MH1.
12. The standard DNA quantity to use for optimal transformation is generally around 1 ng of plasmid DNA per 100 μl of bacterial cells.
13. It is also possible to perform PCR directly on the bacterial clone without plasmid DNA extraction and purification.
14. Whether 2.4d pIVEX vector transformation was successful or not can be checked. This is done either by restriction mapping using a restriction enzyme such as *Hin*dIII or by PCR. If you wish to do a PCR check, use primers annealing to the T7 promoter and T7 terminator sequences at either extremity of the insert (the sequences for these primers can be found in **Section 2.1**). An ordinary (non-proofreading) Taq polymerase can be used for this PCR. Design the PCR experiment taking the enzyme supplier's recommendations into consideration.
15. Use the T7 promoter and T7 terminator primers described in **Section 2.1** to sequence both strands.
16. The amino acid mix is only stable for 4 h at 4°C and should always be freshly prepared.
17. To avoid precipitation, Leu, Trp, and Tyr solutions should be added last.
18. Avoid using a DTT solution that has previously been defrosted.
19. Reaction temperature can be optimized: temperatures lower than 30°C can slow protein production, but generally result in improved protein solubility.
20. Additives such as chaperones or detergents can be added to the reaction mix to help produce soluble proteins. See the manufacturer's instructions for further details.
21. It is not unusual to observe a white precipitate in the cup at the end of the cell-free expression reaction: this may be precipitated salts or the target protein in an insoluble form.
22. Acetone precipitation is necessary to eliminate colloids (present in the reaction mix) which would otherwise interfere with protein migration on the SDS-PAGE gel.
23. Most hexahistidine-tagged proteins elute at an imidazole concentration between 50 and 200 mM.

24. Guanidine tends to precipitate in the presence of SDS. To facilitate loading onto the gel, fractions can be pre-heated at 95°C.
25. Other colorimetric assays, such as Bradford assay, can be used.
26. We generally use HCl or 25 mM ammonium bicarbonate for pH adjustment.
27. For MS-based sequence analysis, the band corresponding to the target protein should be excised from the gel used to control protein purification on the IMAC column (*see* **Section 3.4.2** or **3.4.3**). Trypsin digestion can then be performed in-gel and tryptic peptides extracted for LC-MS analysis.
28. A triple-A protein quantification service is provided by MScan, http://www.m-scan.com/home.
29. Amino acid sequence uniqueness can be checked using the BLAST application, found at http://www.uniprot.org/help/sequence-searches.
30. The target proteins can be added separately (monoplex assay) or simultaneously (multiplex assay).
31. If detectable amounts of endogenous target protein are present in the matrix, the titration curve will not have an intercept at the origin.

References

1. Brun, V., Masselon, C., Garin, J. and Dupuis, A. (2009) Isotope dilution strategies for absolute quantitative proteomics. *J Proteomics.* **72**, 740–749.
2. Bantscheff, M., Schirle, M., Sweetman, G., Rick, J. and Kuster, B. (2007) Quantitative mass spectrometry in proteomics: a critical review. *Anal Bioanal Chem.* **389**, 1017–1031.
3. Ong, S. E. and Mann, M. (2005) Mass spectrometry-based proteomics turns quantitative. *Nat Chem Biol.* **1**, 252–262.
4. Wilm, M. (2009) Quantitative proteomics in biological research. *Proteomics.* **9**, 4590–4605.
5. Huttenhain, R., Malmstrom, J., Picotti, P. and Aebersold, R. (2009) Perspectives of targeted mass spectrometry for protein biomarker verification. *Curr Opin Chem Biol.* **13**, 518–525.
6. Rifai, N., Gillette, M. A. and Carr, S. A. (2006) Protein biomarker discovery and validation: the long and uncertain path to clinical utility. *Nat Biotechnol.* **24**, 971–983.
7. Ye, X., Blonder, J. and Veenstra, T. D. (2009) Targeted proteomics for validation of biomarkers in clinical samples. *Brief Funct Genomic Proteomic.* **8**, 126–135.
8. Malmstrom, J., Lee, H. and Aebersold, R. (2007) Advances in proteomic workflows for systems biology. *Curr Opin Biotechnol.* **18**, 378–384.
9. Gstaiger, M. and Aebersold, R. (2009) Applying mass spectrometry-based proteomics to genetics, genomics and network biology. *Nat Rev Genet.* **10**, 617–627.
10. Rappel, C. and Schaumloffel, D. (2008) The role of sulfur and sulfur isotope dilution analysis in quantitative protein analysis. *Anal Bioanal Chem.* **390**, 605–615.
11. Bettmer, J., Montes Bayon, M., Encinar, J. R., Fernandez Sanchez, M. L., Fernandez de la Campa Mdel, R. and Sanz Medel, A. (2009) The emerging role of ICP-MS in proteomic analysis. *J Proteomics.* **72**, 989–1005.
12. Wolters, D. A., Washburn, M. P. and Yates, J. R., 3rd (2001) An automated multidimensional protein identification technology

for shotgun proteomics. *Anal Chem.* **73**, 5683–5690.
13. Hanke, S., Besir, H., Oesterhelt, D. and Mann, M. (2008) Absolute SILAC for accurate quantitation of proteins in complex mixtures down to the attomole level. *J Proteome Res.* **7**, 1118–1130.
14. Lange, V., Picotti, P., Domon, B. and Aebersold, R. (2008) Selected reaction monitoring for quantitative proteomics: a tutorial. *Mol Syst Biol.* **4**, 222.
15. Stahl-Zeng, J., Lange, V., Ossola, R., Eckhardt, K., Krek, W., Aebersold, R., et al. (2007) High sensitivity detection of plasma proteins by multiple reaction monitoring of N-glycosites. *Mol Cell Proteomics.* **6**, 1809–1817.
16. Picotti, P., Bodenmiller, B., Mueller, L. N., Domon, B. and Aebersold, R. (2009) Full dynamic range proteome analysis of S. cerevisiae by targeted proteomics. *Cell.* **138**, 795–806.
17. Hortin, G. L. and Sviridov, D. (2010) The dynamic range problem in the analysis of the plasma proteome. *J Proteomics.* **73**, 629–636.
18. Ackermann, B. L. and Berna, M. J. (2007) Coupling immunoaffinity techniques with MS for quantitative analysis of low-abundance protein biomarkers. *Expert Rev Proteomics.* **4**, 175–186.
19. Anderson, N. L., Anderson, N. G., Haines, L. R., Hardie, D. B., Olafson, R. W. and Pearson, T. W. (2004) Mass spectrometric quantitation of peptides and proteins using Stable Isotope Standards and Capture by Anti-Peptide Antibodies (SISCAPA). *J Proteome Res.* **3**, 235–244.
20. Roche, S., Tiers, L., Provansal, M., Seveno, M., Piva, M. T., Jouin, P., et al. (2009) Depletion of one, six, twelve or twenty major blood proteins before proteomic analysis: the more the better? *J Proteomics.* **72**, 945–951.
21. Liu, T., Qian, W. J., Mottaz, H. M., Gritsenko, M. A., Norbeck, A. D., Moore, R. J., et al. (2006) Evaluation of multiprotein immunoaffinity subtraction for plasma proteomics and candidate biomarker discovery using mass spectrometry. *Mol Cell Proteomics.* **5**, 2167–2174.
22. Domon, B. (2009) Glycosylation as means of reducing sample complexity to enable quantitative proteomics. *Proteomics.* **9**, 1488–1491.
23. Van Damme, P., Van Damme, J., Demol, H., Staes, A., Vandekerckhove, J. and Gevaert, K. (2009) A review of COFRADIC techniques targeting protein N-terminal acetylation. *BMC Proc.* **3(Suppl 6)**, S6.
24. Lange, V., Malmstrom, J. A., Didion, J., King, N. L., Johansson, B. P., Schafer, J., et al. (2008) Targeted quantitative analysis of Streptococcus pyogenes virulence factors by multiple reaction monitoring. *Mol Cell Proteomics.* **7**, 1489–1500.
25. Malmstrom, J., Beck, M., Schmidt, A., Lange, V., Deutsch, E. W. and Aebersold, R. (2009) Proteome-wide cellular protein concentrations of the human pathogen Leptospira interrogans. *Nature.* **460**, 762–765.
26. Becher, F., Pruvost, A., Clement, G., Tabet, J. C. and Ezan, E. (2006) Quantification of small therapeutic proteins in plasma by liquid chromatography-tandem mass spectrometry: application to an elastase inhibitor EPI-hNE4. *Anal Chem.* **78**, 2306–2313.
27. Ji, C., Sadagopan, N., Zhang, Y. and Lepsy, C. (2009) A Universal strategy for development of a method for absolute quantification of therapeutic monoclonal antibodies in biological matrices using differential dimethyl labeling coupled with ultra performance liquid chromatography-tandem mass spectrometry. *Anal Chem.* 81, 9321–9328.
28. Dubois, M., Fenaille, F., Clement, G., Lechmann, M., Tabet, J. C., Ezan, E., et al. (2008) Immunopurification and mass spectrometric quantification of the active form of a chimeric therapeutic antibody in human serum. *Anal Chem.* **80**, 1737–1745.
29. Gerber, S. A., Rush, J., Stemman, O., Kirschner, M. W. and Gygi, S. P. (2003) Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. *Proc Natl Acad Sci USA.* **100**, 6940–6945.
30. Brun, V., Dupuis, A., Adrait, A., Marcellin, M., Thomas, D., Court, M., et al. (2007) Isotope-labeled protein standards: toward absolute quantitative proteomics. *Mol Cell Proteomics.* **6**, 2139–2149.
31. Arsene, C. G., Ohlendorf, R., Burkitt, W., Pritchard, C., Henrion, A., O'Connor, G., et al. (2008) Protein quantification by isotope dilution mass spectrometry of proteolytic fragments: cleavage rate and accuracy. *Anal Chem.* **80**, 4154–4160.
32. Beynon, R. J., Doherty, M. K., Pratt, J. M. and Gaskell, S. J. (2005) Multiplexed absolute quantification in proteomics using artificial QCAT proteins of concatenated signature peptides. *Nat Methods.* **2**, 587–589.
33. Janecki, D. J., Bemis, K. G., Tegeler, T. J., Sanghani, P. C., Zhai, L., Hurley, T. D., et al. (2007) A multiple reaction monitoring method for absolute quantification of the human liver alcohol dehydrogenase ADH1C1 isoenzyme. *Anal Biochem.* **369**, 18–26.
34. Izrael-Tomasevic, A., Phu, L., Phung, Q. T., Lill, J. R. and Arnott, D. (2009) Targeting

interferon alpha subtypes in serum: a comparison of analytical approaches to the detection and quantitation of proteins in complex biological matrices. *J Proteome Res.* **8**, 3132–3140.

35. Ciccimaro, E., Hanks, S. K., Yu, K. H. and Blair, I. A. (2009) Absolute quantification of phosphorylation on the kinase activation loop of cellular focal adhesion kinase by stable isotope dilution liquid chromatography/mass spectrometry. *Anal Chem.* **81**, 3304–3313.
36. Heudi, O., Barteau, S., Zimmer, D., Schmidt, J., Bill, K., Lehmann, N., et al. (2008) Towards absolute quantification of therapeutic monoclonal antibody in serum by LC-MS/MS using isotope-labeled antibody standard and protein cleavage isotope dilution mass spectrometry. *Anal Chem.* **80**, 4200–4207.

Chapter 8

Organelle Proteomics

Sophie Duclos and Michel Desjardins

Abstract

Proteomics has significantly contributed to improve our understanding of cell structures and functions in the last decade. The possibility to identify large sets of proteins from minute amount of material, linked with the isolation of cellular organelles using various cell fractionation methods, has provided unique insights into the molecular mechanisms governing cell functions in health and disease. The success of this approach relies on the isolation of highly enriched cell fractions enabling the separation of organelles with minimal contamination by other cellular structures.

Key words: Phagocytosis, endocytosis, plasma membrane, cell fractionation, centrifugation.

1. Introduction

Mass spectrometry (MS) and its use for the identification of proteins in biological samples has significantly contributed to the understanding of cell structures and functions in the last few years. The approach to understand cell functions through the identification of the various molecular machines that constitute cellular organelles has been systematically applied for decades. The constant refinement of the methods and instruments to identify a large number of proteins from tiny quantities of starting material, together with the development of powerful bioinformatics tools, has redefined this approach. In the last decade, proteomics has contributed to improve our understanding of various cell organelles including the Golgi complex, the nuclear pore, mitochondria, and various endocytic organelles, to name just a few (for a review *see* (1, 2)). In the field of immunology, the characterization of phagosomes, the cellular organelle where microorganisms

K. Gevaert, J. Vandekerckhove (eds.), *Gel-Free Proteomics*, Methods in Molecular Biology 753,
DOI 10.1007/978-1-61779-148-2_8,

are killed and degraded following their internalization by phagocytosis, has contributed to a better understanding of molecular mechanisms linked to both innate and adaptive immunity. An important conclusion of the work accomplished in the last decade in the characterization of cellular organelles by proteomics is the fact that highly enriched fractions of these structures are needed. Various methods can be used to separate organelles after cell fractionation. Three distinct approaches can be highlighted. Differential centrifugation can be used to separate organelles such as lysosomes, mitochondria, the Golgi apparatus, and the nucleus based on their intrinsic density. Immunoisolation methods based on the use of specific antibodies were originally developed to isolate COP-1-enriched transport vesicles (3). This method has been adapted and found to be particularly efficient to obtain highly enriched plasma membrane preparations from human tissues (4). Finally, the use of flotation gradients to isolate organelles of the endocytic and phagocytic pathways has proven to be an invaluable tool to study the dynamic modulation of endosomes and phagosomes in a variety of cells and physiological conditions. This approach has, for example, provided unprecedented details regarding the molecular mechanisms by which interferon-gamma (IFN-γ) modulates phagosome functions (5, 6).

2. Materials

2.1. Cell Culture

1. Dulbecco's Modified Eagle's Medium (DMEM), high glucose, pH 7.4, supplemented with 10% heat-inactivated fetal bovine serum (FBS), 10 mM HEPES, 1% L-glutamine, 100 U/ml penicillin, and 100 μg/ml streptomycin.
2. Disposable cell lifters.

2.2. Phagocytosis Assay and Isolation of Phagosomes

1. IFN-γ (PBL Biomedical Laboratories).
2. DMEM without serum (*see* **Note 1**).
3. 0.8 μm blue-dyed polystyrene microspheres (Estapor) (*see* **Note 2**).
4. Rubber cell scraper.
5. 1X Dulbecco's phosphate-buffered saline (PBS) (*see* **Note 3**), pH 7.
6. MS-grade sucrose solutions: 8.55% sucrose (HB); 10% sucrose; 25% sucrose; 35% sucrose; 62% sucrose, each in 300 mM imidazole, pH 7.4 (*see* **Notes 4** and **5**).

7. Complete protease inhibitor cocktail tablets (Complete, Roche Diagnostics), from which to prepare 50X aliquots in sterile water, and store at –20°C.
8. Dura-Grind Stainless Steel Dounce Tissue Grinder, 7 ml (Fisher Scientific) (*see* **Note 6**).
9. 0.5% filtered trypan blue solution in PBS.
10. SW 41 Ti rotor and corresponding ultracentrifuge tubes.

2.3. Magnetic Immunoisolation of the Plasma Membrane

1. Rat anti-CD9 (BD).
2. Rat anti-CD11b (RDI Division of Fitzgerald Industries Intl).
3. Rat anti-CD29 (BD).
4. Rat anti-CD44 (Cedarlane).
5. Rat anti-CD45 (BD).
6. Goat anti-rat IgG microbeads.
7. Benzonase (Nuclease), 250 U/μl (Sigma).
8. Complete protease inhibitor cocktail tablets, EDTA-free (Complete EDTA-free, Roche Diagnostics).
9. Homogenization buffer (HB): 250 mM sucrose, 10 mM Tris–HCl, 5 mM $MgCl_2$, 1 mM $CaCl_2$, pH 7.4.
10. 1X Dulbecco's phosphate-buffered saline (PBS), pH 7.4.
11. Labquake tube shaker/rotator (Fisher Scientific).
12. LS separation column (Miltenyi Biotec, Auburn, CA).
13. Dura-Grind Stainless Steel Dounce Tissue Grinder (Fisher Scientific).
14. 0.5% filtered trypan blue solution in PBS.
15. MidiMACS separation unit (magnet) (Miltenyi Biotec).
16. MACS MultiStand (Miltenyi Biotec).
17. Tris/sucrose buffer (TS): 10 mM Tris–HCl, 250 mM sucrose, pH 7.4.
18. PVP/milk solution: 0.2% (w/v) polyvinylpyrrolidone-40T (PVP-40T), 0.05% (w/v) instant skimmed milk powder in PBS.
19. PVP/KI solution: 0.2% (w/v) PVP-40T and 0.6 M potassium iodide (KI) in PBS.
20. HB1 solution: HB, 100 U/ml Benzonase, 1X Complete EDTA-free (*see* **Note 7**).
21. HB2 solution: HB, 200 U/ml Benzonase, 1X Complete EDTA-free.
22. HB1/KI solution: HB, 100 U/mm Benzonase, 1X Complete EDTA-free, 1.2 M KI.

3. Methods

The use of small inert particles to load endosomes or phagosomes allows for the isolation of endocytic/phagocytic organelles in different conditions. Two types of particles are mostly utilized. Polystyrene particles are useful for the isolation of endocytic/phagocytic organelles because they display a low buoyant density that allows the flotation of endosomes or phagosomes on step gradients, at an interface where no other cell constituent can be found (6). They can be purchased in sizes ranging from 0.1 to 10 μm, in various colors (including a range of fluorescent spectra), and with surfaces enabling the attachment of a wide variety of opsonins including ligands that selectively bind cellular receptors, thus permitting the study of circumscribed pathways. Typically, beads around 1.0 μm and larger are used for the study of phagocytosis, while smaller beads can be used to isolate endosomes (0.1 μm). We observed that phagosomes usually contain a single particle, while the internalization of smaller beads by endocytosis results in the formation of endosomes displaying a large number of particles (unpublished observations). The second type of particles used to isolate endocytic/phagocytic organelles is magnetic microspheres that can be isolated with the help of a magnet after cell lysis. The following protocol, describing the isolation of phagosomes from IFN-γ-treated cells using polystyrene particles, has been optimized for RAW 264.7 macrophages but can easily be adapted to any cell type.

3.1. Isolation of IFN-γ-Treated Phagosomes from RAW 264.7 Murine Macrophage Cell Line on a Discontinuous Sucrose Gradient

1. Grow RAW 264.7 macrophages in 100 mm Petri dishes at 37°C in 5% CO_2 in DMEM so that they reach confluency on the day of the experiment. A passage is made by gently scraping a confluent dish 36 h before the experiment. For cytokine-treated cells, add 100 units/ml IFN-γ in the medium 24 h prior to the isolation of phagosomes. After centrifugation (500×*g*, 5 min), cells are split 1:8 in new DMEM pre-warmed to 37°C. For a typical experiment, 84 dishes of cells are prepared, of which 42 are control and 42 are IFN-γ-treated cells (*see* **Note 8**).
2. On the day of the experiment, briefly sonicate microspheres in a sonicator bath and then dilute 1:50 in pre-warmed DMEM without serum. Prepare 4 ml of DMEM with microspheres for each dish and vortex vigorously.
3. At room temperature, quickly aspirate the medium from the dishes, seven dishes at a time, and replace with 4 ml of DMEM-containing microspheres. Always pour the medium along the side of the dish so that the cells do not

detach. Start the timer for the internalization (1 h) as soon as the dishes are back in the incubator. Proceed like this for all the dishes.

4. After the pulse, wash cells three times for 5 min in cold PBS (*see* **Note 9**) to stop the internalization and remove the non-internalized microspheres.
5. After the last wash, replace the PBS by 5 ml of pre-warmed DMEM without serum and place in the incubator for the chase (1 h).
6. During the chase, cool the centrifuge to 4°C, place the stainless steel grinder on ice, thaw the protease inhibitors, and prepare the sucrose solutions (*see* **Note 10**).
7. Place the dishes on ice and slowly rock them so that the cells are always covered with liquid.
8. Gently detach the cells using the rubber cell scraper, always keeping the scraper in the liquid, and transfer the contents of seven dishes into one 50 ml tube, one dish at a time with a plastic bulb pipette.
9. Complete each tube to 50 ml with cold PBS and centrifuge at 500×*g* for 5 min.
10. Discard the supernatant, re-suspend in 30 ml of cold PBS, and centrifuge at 500×*g* for 5 min. Repeat this step.
11. Discard the supernatant, re-suspend in 10 ml of cold PBS, and transfer the contents of each 50 ml tube into a 15 ml tube. Centrifuge at 500×*g* for 5 min.
12. Discard the supernatant and add 1 ml of cold HB (8.55% sucrose solution) containing protease inhibitors to each 15 ml tube. Re-suspend the pellet by tapping the tube on the side of the ice bucket, so that the pellet is completely solubilized.
13. Centrifuge at 500×*g* for 5 min. Aspirate the supernatant and add another 1 ml of cold HB to each 15 ml tube.
14. On ice, gently re-suspend the samples and transfer in the cold stainless steel grinder, 3 ml at a time. Break the cells by moving the pestle up and down until 90% of the cells are broken without major breakage of the nucleus as monitored by light microscopy, every five strokes (*see* **Note 11**). Repeat with all the samples, cleaning the grinder between control and IFN-γ-treated samples. Keep an aliquot corresponding to 10% of the preparation for further Western blot analysis (total cell lysate (TCL) control).
15. Transfer back the homogenates into the 15 ml tubes (1 ml/tube) and centrifuge at 900×*g* for 5 min in order to prepare a post-nuclear supernatant (PNS).

16. During centrifugation, prepare the required number of SW 41 ultracentrifuge tubes by adding 1 ml of 62% sucrose solution in each tube. Also add 1 ml of 62% sucrose solution to the same number of 2 ml tubes.
17. Carefully recover the top supernatant (clear part, around 1 ml) with a p1000 Pipetman. This fraction contains the phagosomes.
18. Deposit this fraction on the previously prepared 1 ml of 62% sucrose solution in the 2 ml tubes. Mix very well and lay on top of the 62% sucrose solution already in the SW 41 ultracentrifuge tubes, using a plastic bulb pipette.
19. Sequentially add 2 ml of each sucrose solution (35, 25, 10%) very carefully on top of each other to form the discontinuous gradient. Take a plastic bulb pipette to do this, and use the graduations on the 15 ml tubes to calculate the 2 ml. Do this for each tube, and equilibrate tubes by adding 10% sucrose solution.
20. Ultracentrifuge at 71,000×*g* in the SW 41 Ti rotor for 1 h at 4°C.
21. Phagosomes are found within the blue band at the 10% sucrose and 25% sucrose interface, as shown in **Fig. 8.1**. With a 14.5 cm Pasteur pipette, or with a Sarstedt transfer pipette (1 ml, fine tip), carefully aspirate the blue band by making circular motions at the surface of the

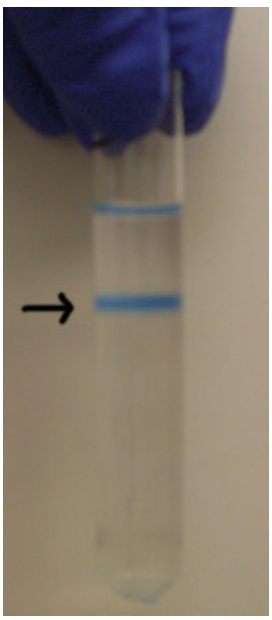


Fig. 8.1. Isolated phagosomes formed by the internalization of 0.8 μm blue-dyed Estapor® polystyrene microspheres after ultracentrifugation on a discontinuous sucrose gradient. *Arrow* indicates the band where phagosomes are collected, at the 10–25% interface of the gradient.

band, avoiding disturbing the gradient. Put in a new ultracentrifuge tube, pooling two gradients of phagosomes per tube.

22. Fill the tubes with cold PBS (about 10 ml), put a Parafilm on top, and mix by inversion. Equilibrate with PBS and ultracentrifuge at 28,000×*g* in the SW 41 Ti rotor for 15 min at 4°C. The pellet will contain the phagosomes. An example of a preparation of isolated phagosomes as seen under an electron microscope is shown in **Fig. 8.2**.
23. Remove the supernatant by inverting the tubes and aspirate drops of PBS on the sides of the tubes.
24. Re-suspend the pellet in 200 µl of PBS, keep an aliquot corresponding to 10% of the preparation (20 µl) for protein quantitation and Western blot analysis, and transfer the remaining into a 1.5 ml tube. An example of a Western blot analysis to assess purity of the preparation is shown in **Fig. 8.3**.
25. Centrifuge 5 min at 21,000×*g* in a tabletop microcentrifuge. Remove PBS and either store these dry pellets at –80°C or re-suspend in a minimum volume of buffer suitable for MS analysis (for example, 5 mM ammonium bicarbonate buffer, pH 7.8, at a volume of 2 µl per Petri dish used in the experiment).

3.2. Isolation of Plasma Membrane from BMA3.1A7 (BMA) Murine Macrophage Cell Line

The successful immunomagnetic isolation of plasma membranes relies on the specificity and affinity of the cell surface antibodies used in the protocol. The use of several specific antibodies ensures

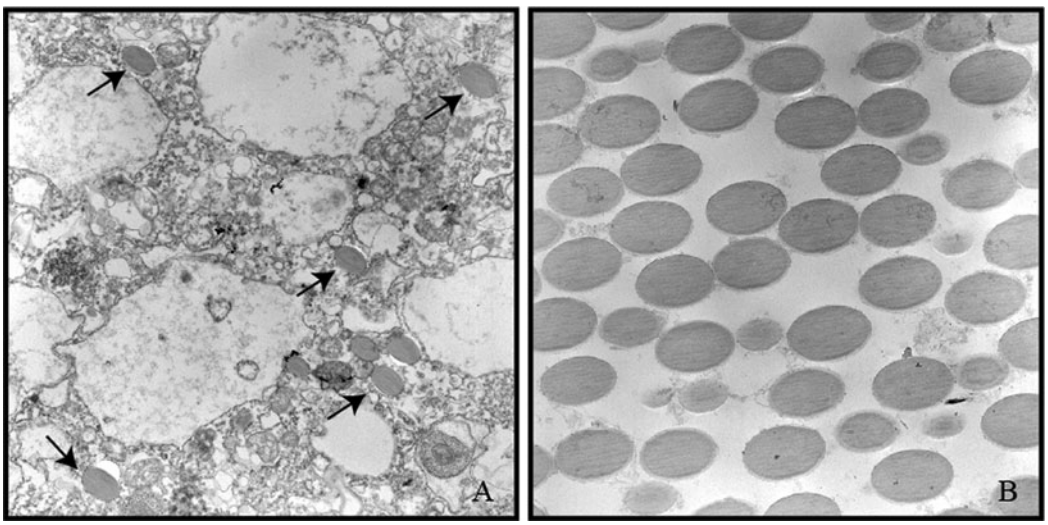


Fig. 8.2. Electron micrographs of a PNS (**a**) (7000X) compared to isolated phagosomes (**b**) (12,000X), showing the high level of enrichment and the absence of major apparent contaminants in phagosome preparation. *Arrowheads* point at phagosomes in the PNS sample.

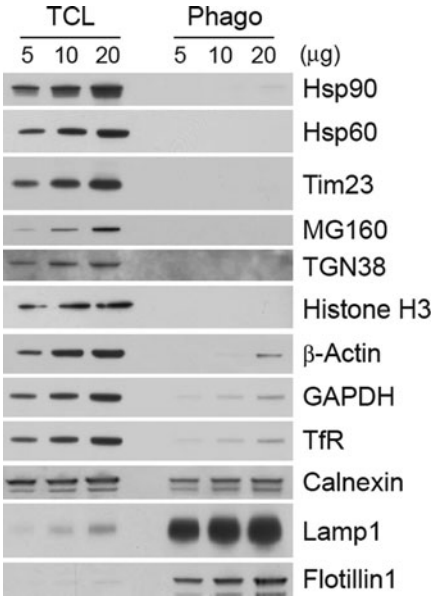


Fig. 8.3. Assessment of organelle contamination of 0.8 μm polystyrene microsphere-containing phagosomes isolated from macrophages. The methodology was validated using mammalian J774 macrophages and Western blotting. In this system the presence of contaminating organelles could be detected using the following antibodies: histone H3 (nucleus), Hsp90 (cytoplasmic/nucleus), Hsp60 (mitochondria), Tim23 (mitochondria), MG160 (Golgi), TGN38 (Golgi), β-actin (cytoskeletal), GAPDH (cytoplasma), transferrin receptor (TfR) (plasma membrane/recycling endosomes), calnexin (ER), Lamp1 (lysosome), and flotillin 1 (late endosome/lysosome). Total cell lysates and phagosomes were run for comparison (reproduced from (7)).

a higher yield of the plasma membrane preparation. A thorough screening of candidate antibodies using FACS is therefore needed in order to choose the perfect antibody cocktail for a specific cell type.

1. Grow BMA macrophages in 100 mm Petri dishes at 37°C in 5% CO_2, in DMEM, so that they reach confluency on the day of the experiment. A passage is made by gently detaching cells (gently going up and down with the pipette (*see* **Note 12**) from a confluent dish 48 h before the experiment). After centrifugation (500×*g*, 3 min), cells are split 1:8 in new DMEM pre-warmed to 37°C. For a typical experiment, prepare 10 dishes of BMA cells (*see* **Note 13**).
2. On the day of the experiment, thaw Complete EDTA-free and keep on ice.
3. Cool the centrifuge to 4°C and place 150 ml of PBS on ice.
4. Pre-warm 150 ml of DMEM at room temperature and wash the cells twice in this medium (*see* **Note 14**).

5. Detach macrophages from each Petri dish and transfer to the appropriate number of 50 ml tubes (five dishes per 50 ml tube).
6. Centrifuge cells at 500×*g* for 5 min at 4°C.
7. Re-suspend in a volume of 2 ml/dish of ice-cold PBS and pool pellets into a 50 ml tube.
8. Count cells by light microscopy using a hemocytometer (10 μl of cell suspension in 90 μl of 0.5% trypan blue).
9. Adjust the concentration to 5×10^6 cells/ml with PBS and then add 1 μg of each antibody (CD9, CD11b, CD29, CD44, and CD45 antibodies) per 10^6 cells.
10. Incubate at 4°C for 30 min in Labquake tube shaker/rotator (8 rpm).
11. Wash cells three times in 10 ml of cold PBS by centrifuging at 500×*g* for 5 min at 4°C.
12. Gently re-suspend the pellet in cold PBS (30 ml of PBS for 150×10^6 cells) and add goat anti-rat IgG MicroBeads to obtain a final concentration of 5×10^6 cells/ml and 12.5 μl of MicroBeads/10^6 cells.
13. Incubate at 4°C for 30 min in Labquake tube shaker/rotator.
14. Wash cells three times in 10 ml of cold PBS by centrifuging at 500×*g* for 5 min at 4°C.
15. Prepare HB1 and HB2 solutions.
16. Gently re-suspend pellet in HB1 (500 μl/dish) and perform a cell count.
17. Add HB1 to the cell suspension to obtain a final cell density of 20×10^6 cells/ml (*see* **Note 15**).
18. On ice, transfer the sample to the pre-cooled stainless steel grinder and add an equal volume of HB2 solution to obtain a final concentration of 10×10^6 cells/ml (*see* **Note 16**).
19. Break the cells by moving the pestle up and down until 90% of the cells are broken without major breakage of the nucleus as monitored by light microscopy.
20. Keep an aliquot corresponding to 10% of the preparation for further Western blot analysis (total cell lysate (TCL) control).
21. Centrifuge homogenate at 900×*g* for 10 min at 4°C.
22. Recover the PNS (post-nuclear supernatant).
23. Prepare the HB1/2KI solution (*see* **Note 17**).
24. Add an equal volume of HB1/2KI solution to the PNS and incubate for 15 min at 4°C in the Labquake tube shaker/rotator.

25. Add Complete EDTA-free to a final concentration of 1X to the PVP/milk and the PVP/KI solutions.
26. In a cold room, place a LS column on the MidiMACS magnet and equilibrate with 3 ml of PVP/milk.
27. Add the sample to the column and let it completely drip through.
28. Wash the column three times with 8 ml of PVP/KI and once with 8 ml of TS buffer.
29. Take the column off the magnet and place it over a 2 ml tube.
30. Add 1.5 ml of TS buffer to the column and use the plunger to force the magnetic fraction containing the isolated plasma membrane through the column.
31. Keep 10% of the eluate for protein quantification and Western blot analysis of purity.
32. Centrifuge the remaining eluate for 30 min at 21,000 × *g* at 4°C.
33. Discard supernatant and either freeze the yellow pellet for future MS analysis or immediately re-suspend the pellet in MS-compatible buffer.

4. Notes

1. Polystyrene microspheres rapidly get opsonized with serum proteins in FBS-containing DMEM. These serum proteins highly interfere with the detection of phagosomal proteins by the mass spectrometer. This is why most of our preparations destined to MS analysis are performed in DMEM without serum.
2. The same assay can be used to isolate endosomes, using this time 0.1 μm polystyrene microspheres (Estapor).
3. If analyzing phosphoproteins, use Tris-buffered saline (TBS) instead of PBS and add phosphatase inhibitors together with protease inhibitors (Complete).
4. It is very important to work in dust-free conditions, with gloves, MS-grade solutions, and with particular care to all the lids and bottle caps that have to be closed immediately after usage.
5. All the sucrose solutions are weight/weight solutions in water. For example, to make the 8.55% solution, we would weigh 8.55 g of sucrose, add bidistilled water up to 99 g,

and add 1 ml of 300 mM imidazole pH 7.4 to buffer the solution. All the solutions are adjusted to pH 7.4. The 62% solution has to be slightly heated to dissolve and cooled down before adjusting pH. All the solutions except for the 62% sucrose are filtered on a 0.45 μm filter. We routinely prepare a large quantity that we aliquot in sterile bottles kept at –20°C.

6. If you do not have a stainless steel grinder, you can break the cells using a 1 cc tuberculin syringe with 22 G $1^1/_2$ needle (because the number of strokes with a syringe can vary, it is advisable to monitor at the microscope to avoid breakage of nucleus).
7. Complete, Complete EDTA-free, and Benzonase are always added fresh to the solutions.
8. Six to eight confluent dishes are needed to make one ultracentrifugation gradient, so 84 dishes give a total of 12 gradients. The SW 41 rotor contains six tubes, so the use of two ultracentrifuges is recommended, as ideally the control and cytokine-treated cells are done at the same time.
9. It is better to use cold PBS in order to stop the internalization process and have a more synchronized population of phagosomes, but the PBS has to be poured very carefully on the side of the dishes to prevent the cells from detaching.
10. To prepare the sucrose solutions, calculate 2 ml of each solution for one gradient and put in 15 ml tubes. For 12 gradients, you will have 2 tubes containing 12 ml for each sucrose solution. Add protease inhibitors to the HB and the 62% sucrose to a final concentration of 1X.
11. To monitor cell breakage, put a drop of the cell homogenate on a glass slide and add a drop of trypan blue to clearly distinguish the nuclei.
12. BMA macrophages attach very loosely to Petri dishes. When working with more adherent cells, the use of a cell scraper is recommended.
13. About 100×10^6 cells are needed to get 100 μg of isolated plasma membrane proteins with this protocol. A confluent dish of BMA roughly contains around 15×10^6 cells.
14. This step and all the following ones have to be performed in dust-free conditions, ideally under a clean bench when preparing the samples for MS analysis.
15. At this stage, labeled cells can be stored in the freezer until proceeding further. When ready to carry on with the experiment, thaw labeled cells in a water bath at RT.

Once thawed, immediately remove the tube from water and transfer on ice.

16. Adjust your sample preparation so that the total volume in the grinder is no less than 3 ml and no more than 5 ml at a time.
17. Potassium iodide is added in order to dissociate cytoskeletal proteins from the plasma membranes (4).

Acknowledgments

This work was supported by the Canadian Institute for Health Research and the Canadian Research Chairs.

References

1. Brunet, S., Thibault, P., Gagnon, E., Kearney, P., Bergeron, J.J., and Desjardins, M. (2003) Organelle proteomics: looking at less to see more. *Trends Cell Biol* **13**, 629–638.
2. Au, C.E., Bell, A.W., Gilchrist, A., Hiding, J., Nilsson, T., and Bergeron, J.J. (2007) Organellar proteomics to create the cell map. *Curr Opin Cell Biol* **19**, 376–385.
3. Lanoix, J., Ouwendijk, J., Lin, C.C., Stark, A., Love, H.D., Ostermann, J., and Nilsson, T. (1999) GTP hydrolysis by arf-1 mediates sorting and concentration of Golgi resident enzymes into functional COP I vesicles. *EMBO J* **18**, 4935–4948.
4. Sircar, K., Gaboury, L., Ouadi, L., Mecteau, M., Scarlata, E., Saad, F., Aprikian, A., Tanguay, S., Lapointe, S., Lussier, C., Miletti, T., and Lanoix, J. (2006) Isolation of human prostatic epithelial plasma membranes for proteomics using mirror image tissue banking of radical prostatectomy specimens. *Clin Cancer Res* **12**, 4178–4184.
5. Jutras, I., Houde, M., Currier, N., Boulais, J., Duclos, S., LaBoissière, S., Bonneil, E., Kearney, P., Thibault, P., Paramithiotis, E., Hugo, P., and Desjardins, M. (2008) Modulation of the phagosome proteome by interferon-gamma. *Mol Cell Proteomics* **7**, 697–715.
6. Desjardins, M., Huber, L.A., Parton, R.G., and Griffiths, G. (1994) Biogenesis of phagolysosomes proceeds through a sequential series of interactions with the endocytic apparatus. *J Cell Biol* **124**, 677–688.
7. Stuart, L.M., Boulais, J., Charriere, G.M., Hennessy, E.J., Brunet, S., Jutras, I., Goyette, G., Rondeau, C., Letarte, S., Huang, H., Ye, P., Morales, F., Kocks, C., Bader, J.S., Desjardins, M., and Ezekowitz, R.A. (2007) A systems biology analysis of the Drosophila phagosome. *Nature* **445**, 95–101.

Chapter 9

Membrane Protein Digestion – Comparison of LPI HexaLane with Traditional Techniques

Ping Sui, Tasso Miliotis, Max Davidson, Roger Karlsson, and Anders Karlsson

Abstract

Membrane protein profiling and characterization is of immense importance for the understanding of vital processes taking place across cellular membranes. Traditional techniques used for soluble proteins, such as 2D gel electrophoresis, are sometimes not entirely applicable to membrane protein targets, due to their low abundance and hydrophobic character. New tools have been developed that will accelerate research on membrane protein targets. Lipid-based protein immobilization (LPI) is the core technology in a new approach that enables immobilization and digestion of native membrane proteins inside a flow cell format. The presented method is described in the context of comparing the method to traditional approaches where the sample amount that is digested and analyzed is the same.

Key words: Membrane protein, immobilization, LPI, profiling, in-gel, in-solution, sequential digestion.

1. Introduction

The proteome is the complete set of proteins in a certain organism, cell, organelle, or body fluid at a given time under precisely defined conditions. Of all the proteins encoded in the human genome, 20–30% are estimated to belong to the class of membrane proteins (1); however, as much as 70% of all the drug targets are membrane proteins such as receptors and ion channels (2). Despite constituting ~30% of the total genome, membrane proteins are also generally under-represented in proteome expression studies (3–6). This under-representation of membrane

K. Gevaert, J. Vandekerckhove (eds.), *Gel-Free Proteomics*, Methods in Molecular Biology 753,
DOI 10.1007/978-1-61779-148-2_9,

proteins is primarily attributed to the heterogeneous, hydrophobic, and low abundance nature of these proteins.

To overcome these limitations, various strategies have been applied for the enrichment, solubilization, separation, and digestion steps involved in membrane protein studies. The common approaches for analyzing membrane proteins normally proceed via protein solubilization and denaturation, as their hydrophobic domains resist exposure to aqueous solvents, causing aggregation, adsorption, and precipitation that may lead to sample loss. Thus, reagents must be carefully chosen to maintain membrane protein solubility and facilitate digestion without interfering with downstream LC separation and MS detection.

Shotgun proteomics generally involves digesting a complex protein sample with subsequent separation of the resulting peptides using one or several dimensions of peptide chromatography. Again, the complexity of the sample is huge since a sample may contain thousands of proteins and once digested into peptides, the sample complexity rises even more. Numerous protocols to improve the solubilization and the digestion of membrane proteins have been reported, including the use of organic solvents (7–12), organic acids (13–15), detergents (8, 12), chaotropic reagents (13, 16–19), chemical cleavage reactions (13, 19, 20), and nonselective proteases (21, 22). However, the use of such detergents and organic solvents for membrane protein solubilization must not interfere with subsequent protein separation, enzymatic digestion, and MS analysis. Recently, "MS-compatible" detergents have been introduced that include the acid-labile surfactants RapiGest (ionic, Waters) (23) and PPS (zwitterionic, Protein Discovery) (24).

The lipid-based protein immobilization technology (LPI), developed by Nanoxis AB, enables immobilization of native membrane proteins. The technology has been developed as a flow cell platform to allow easy and quick exchange of the solution environment around the membrane proteins without loss or dilution of the sample and provides effective and precise preparation of samples for downstream mass spectrometry analysis. The LPI^{TM} technology uses a proprietary surface that binds intact proteins embedded in proteoliposomes. Membrane proteins are produced directly from a wide variety of cells and tissues by a tailored membrane preparation protocol. The LPI approach differs from the traditional methods in one very important aspect. Instead of trying to solubilize the membrane proteins, the LPI approach retains the membrane proteins in their native lipid bilayer and no detergents are used in the sample preparation step. This makes the LPI approach more generally applicable to all types of membrane proteins compared to traditional assays and gives the researcher a unique and beneficial opportunity to probe the solvent-exposed parts of the membrane proteins,

e.g., by proteolytic digestion. When digestion is performed on bilayer-embedded native membrane proteins, structural information regarding topography or surface exposure of soluble domains of the membrane proteins can be probed with the help of different proteases and other chemical cleavage agents. In the case of proteolytic digestion, reaction time is easy to control and limited digestion strategies (minute scale) are applicable. Moreover, immobilization enables the implementation of multiple digestions of the sample in sequence, simply by changing the protease solution in the flow cell. This will produce several different and complementary peptide fractions from the same sample and can help to increase the sequence coverage of the soluble parts of the membrane proteins. The flow cell accepts membrane proteoliposomes derived from mammalian cell lines, bacteria, and yeast, as well as purified membrane fractions, making it a versatile and generic tool for membrane protein studies.

Membrane samples can be prepared in many different ways, and described below is a general method for use with mammalian cells applicable to the LPI technology. The method produces membranes predominantly from the plasma membrane, ER, and Golgi. First, cells grown from tissue culture or prepared from tissue specimens are washed and pelleted. Cell pellets are re-suspended in hypotonic lysis buffer in order to swell the cells (25). It is important to perform the preparation at low temperature in order to reduce protein degradation and minimize DNA gelling. Cells are disrupted using a homogenizer and rapidly subjected to low-speed centrifugation in order to pellet out cell debris and nuclei from the homogenate. The supernatant is examined by microscopy to assure proper disruption and removal of nuclei. Mitochondria are then pelleted out using medium speed centrifugation. The supernatant contains ER, Golgi, and plasma membranes in a solution of cytosol. Further purification of the membranes requires one or more ultracentrifugation steps to remove soluble proteins. During this stage treatment of the membranes with high-salt and high-pH solutions can be used to strip off proteins loosely attached to the membranes, thereby increasing the proportion of integral and lipid-anchored membrane proteins in the sample (26). As a final step the membrane fragments are transformed into small proteoliposomes using either extrusion or sonication. Tip sonication, used in this study, provides a high-frequency vibration that breaks apart large membrane fragments into smaller ones which spontaneously seal into small proteoliposomes. Sonication-assisted formation of proteoliposomes leads to a mixture of membrane orientation, i.e., both the intra- and extracellular domains of the integral membrane proteins (IMPs) are exposed on the outside of the proteoliposomes.

For the presented method(s) a comparative study was made between the LPI approach and the traditional approaches of

in-solution digestion and in-gel digestion to analyze and identify membrane proteins. Specifically, the study exemplifies results where the same amount of sample was digested. To compare the outcome of the different approaches, the digested peptides of each sample were analyzed with a chip-LC-MS/MS system. The details of the instruments, settings, as well as the bioinformatics part used in this study are found under instruments and software.

2. Materials

2.1. Sample Preparation

1. Phosphate-buffered saline (PBS), pH 7.4
2. NOVEX Stainer A and B
3. NuPAGE 4–12% Bis-Tris gel
4. NuPAGE CDS sample Buffer 4x
5. NuPAGE MES SDS running buffer 20x
6. NuPAGE reducing agent
7. NuPAGE antioxidant
8. Xcell *SureLock*™ Mini-cell
9. LPI reagent Kit ionic buffer (10 mM Tris, 300 mM NaCl, pH 8)
10. LPI reagent Kit basic buffer (20 mM ammonium bicarbonate, pH 8)
11. Precision plus protein™ all blue standards
12. Bio-Rad reagent A and reagent B
13. Protease inhibitor cocktail
14. 2,2,2-trifluoroethanol (TFE)
15. PPS Silent Surfactant
16. Sequencing-grade modified trypsin, stock solution 20 μg/ml in ammonium bicarbonate
17. Lysis buffer: 10 mM $NaHCO_3$, protease inhibitor cocktail
18. Mobile phase A: 97% water, 3% acetonitrile, 0.1% formic acid
19. Mobile phase B: 95% acetonitrile, 5% water, 0.1% formic acid
20. PPS stock solution: add 500 μl of 50 mM ammonium bicarbonate to 1 mg of PPS (0.2% (w/v))
21. Solution A: 40 ml Milli-Q water, 50 ml methanol and 10 ml acetic acid

22. Solution B: 55 ml Milli-Q water, 20 ml methanol, 20 ml NOVEX Stainer A
23. Wash solution: 35% acetonitrile in 25 mM ammonium bicarbonate
24. Extraction solution: 1% formic acid in 5% acetonitrile

2.2. Instruments and Software

Membrane preparation: The homogenizer used for cell lysis was purchased from Wheaton Industries Inc. (Millville, NJ, USA). Tip sonication was done with a Vibra Cell (model 501) from Sonics & Materials, Inc., equipped with a 2 mm tip.

3. Methods

3.1. Sample Preparation

Keep tubes, solutions, and the Dounce homogenizer on ice during the whole procedure. The following protocol has been tested for Chinese hamster ovary cells using a 7 ml glass Dounce homogenizer.

1. Grow CHO K1 cells to confluency in Ham's F12 medium with glutamine supplemented with 10% fetal bovine serum.
2. Cells are harvested using Accutase and can be stored at –80°C in Ham's F12 supplemented with 20% fetal bovine serum and 10% DMSO or used directly.
3. If the cells are frozen, thaw them, and maintain the cell sample on ice.
4. Pellet the cells by centrifugation at 500×*g* for 3 min at 4°C.
5. Remove and discard the supernatant.
6. Add PBS buffer and re-suspend the cells using a wide-bore pipette.
7. Re-pellet the cells by centrifugation at 500×*g* for 3 min at 4°C.
8. Repeat Steps 5–7 two more times. 80 million cells should roughly give 200 µl of cell pellet.
9. Add 10 volumes of ice-cold lysis buffer to the cell pellet and re-suspend the cells gently using a wide-bore pipette.
10. Place the sample on ice and wait 10 min for the cells to swell.
11. Transfer the sample to the Dounce homogenizer and lyse the cells using 20 strokes with a tight-fitting pestle (*see* **Note 1**).
12. Quickly transfer the homogenate to a tube and centrifuge at 500×*g* for 3 min at 4°C to pellet cell debris and nuclei.

13. Carefully collect the supernatant, transfer to a new tube, and centrifuge at 10,000×*g* for 10 min at 4°C to pellet mitochondria (*see* **Note 2**).
14. Transfer the supernatant to a glass vial. Add Na_2CO_3 stock solution so that the final concentration of Na_2CO_3 is 100 mM. This carbonate wash of membranes removes peripheral membrane proteins and is an optional step. If not needed, skip to Step 15 (*see* **Note 3**).
15. Use bath sonication in an ice-cold bath for 20 min to disrupt and stir the membranes. Transfer the solution to a centrifuge tube.
16. Pellet the membranes by ultracentrifugation at 120,000×*g* for 1.5 h at 4°C.
17. Carefully remove and discard the supernatant (containing soluble protein).
18. Check if there is a pellet at the bottom of the tube.
19. Add a volume of solution of choice and re-suspend the pellet. The membrane pellet can be difficult to disperse after carbonate treatment, in which case use brief tip sonication.
20. If desired, rinse the membranes by adding one volume of deionized water in Step 17 and repeat Steps 14–17 (*see* **Note 4**).
21. Use tip sonication to break the membrane fragments into small proteoliposomes. Dilute the membrane preparation to about 1 mg/ml with buffers compatible with downstream processing and analysis.
22. Tip-sonicate the solution using short pulses (typically 1 s) with the same rest time in between pulses for 2 min on an ice bath (*see* **Note 5**). Set the amplitude to medium setting and check the preparation using a light microscope. Repeat if needed.
23. Determine the protein concentration (*see* **Note 6**).

3.2. Protein Digestion

1. Dilute a 7.5 μg protein aliquot of the membrane preparation to 40 μl with 10 mM Tris–HCl, 300 mM NaCl, pH 8, in a 1.5 ml Eppendorf tube.

3.2.1. TFE In-Solution Digestion

2. Add 20 μl of 100 mM Ambic to the sample followed by 60 μl of TFE, which is used as the denaturation agent.
3. Perform reduction of disulfide bridges at 55°C for 45 min by adding DTT to a final concentration of 10 mM. Subsequently perform alkylation in the dark for 45 min after adding IAA to a final concentration of 33 mM. Add an extra 6 μl of 200 mM DTT and incubate for 45 min in the dark to destroy excess of IAA.

4. Trypsin will be denatured or less active if the TFE level exceeds 5% (v/v). Hence, dilute the TFE concentration in the sample by adding 786 µl of Milli-Q water and 262 µl of 100 mM Ambic. Digest the sample by trypsin with a final enzyme to protein concentration ratio of 1/25 (w/w) and incubate overnight at 37°C.
5. Inactivate trypsin by adding 12 µl of formic acid and vacuum dry the tryptic peptides using a SpeedVac. Re-dissolve the dried peptides with 15 µl of mobile phase A and store at –80°C until analysis.

3.2.2. PPS In-Solution Digestion

1. Dilute a 7.5 µg protein aliquot of the membrane preparation to 40 µl with 10 mM Tris–HCl, 300 mM NaCl, pH 8, in a 1.5 ml Eppendorf tube followed by the addition of 40 µl of PPS stock solution.
2. Perform reduction of disulfide bridges at 55°C for 30 min by adding DTT to a final concentration of 10 mM. Subsequently perform alkylation in the dark for 45 min by adding IAA to a final concentration of 33 mM. Add an extra 6 µl of 200 mM DTT and incubate for 45 min in the dark to destroy excess IAA.
3. Digest the sample by trypsin with a final enzyme to protein concentration ratio of 1/25 (w/w) and incubate overnight at 37°C. Stop digestion by adding 2 µl HCl and incubate for 1 h.
4. Store the sample at –80°C until analysis.

3.2.3. In-Gel Digestion

1. Mix a protein aliquot (7.5 µg) of the membrane preparation in an Eppendorf tube with NuPAGE LDS sample Buffer 4x, NuPAGE reducing agent, and Milli-Q water in proportions according to the manufacturer's recommendations.
2. Heat the samples to 70°C for 10 min.
3. Load the denatured samples (15 µl) onto the wells of the gel (4–12% Bis-Tris gel and NuPAGE MES SDS running buffer). Use Precision plus protein™ all blue standards as molecular weight standard.
4. Add NuPAGE antioxidant to the running buffer in the upper buffer chamber of the Xcell SureLock™ Mini-cell to keep the proteins in a reduced state.
5. Run the gels at 200 V for 35 min.
6. Fix the separated proteins with solution A for 10 min and stain with solution B for 10 min, then add 5 ml of NOVEX Stainer B and shake for a minimum of 3 h.
7. De-stain the gel with 200 ml of Milli-Q water and shake for at least 7 h.

8. Cut the gel lane into 12 bands and mark them 1–12. Cut each band into three pieces and keep in the corresponding Eppendorf tube.
9. Wash the gel pieces with 150 μl wash solution and wash three times by incubating for 5 min, mix and incubate for another 5 min, then remove the liquid.
10. Vacuum dry the gel pieces in a SpeedVac.
11. Reduce the dry gel pieces with 40 μl of 10 mM DTT at 56°C for 45 min.
12. Remove the liquid and alkylate the samples by adding 40 μl of 55 mM IAA and incubate at room temperature in the dark for 30 min. Discard the supernatant.
13. Wash the gel pieces with 150 μl of 25 mM Ambic and 70% of acetonitrile in 25 mM Ambic for 5 min. Remove the supernatant and vacuum dry the gel pieces in a SpeedVac.
14. Perform in-gel digestion overnight at 37°C by adding 15 μl of freshly made trypsin solution (10 ng/μl).
15. After digestion, discard the liquid present at the tube lid and extract peptides by adding 40 μl of extraction solution for 1 h. Such extracted peptides are now ready for LC-MS/MS analysis.

3.2.4. HexaLane FlowCell Digestion

1. Load a 7.5 μg protein aliquot of the membrane preparation into each lane of the HexaLane FlowCell.
2. For protein digestion, using a 200 μl pipette, inject 50 μl of proteoliposome solution into the single lanes and leave the LPI^{TM} HexaLane FlowCell to incubate at 37°C for 1 h (*see* **Note 7**).
3. Inject 400 μl of 10 mM Tris–HCl, 300 mM NaCl, pH 8, followed by 400 μl of 20 mM Ambic, pH 8, to rinse the flow cell (these wash buffers must be pre-warmed to 37°C). Remove excess buffer from the out port continuously.
4. Inject 100 μl of freshly made trypsin with a final enzyme to protein concentration ratio of around 1/25 (w/w) in Ambic. Incubate 2 h at 37°C. Remove excess fluid from the out port.
5. Inject 200 μl 20 mM Ambic, pH 8, to elute peptides. Transfer the peptide sample from the out port to a tube.
6. Add 2 μl of formic acid to the eluted peptides to inactivate trypsin.
7. Vacuum dry the eluted peptides using a SpeedVac, redissolve the dried peptides in 15 μl of mobile phase A, and store at –80°C until analysis.

For multistep protocols *see* **Note 8**.

3.3. Comparison of Different Digestion Methods

3.3.1. Comparative Results

The HexaLane FlowCell digestion (single step, 2-step), in-solution digestion (TFE and PPS), and in-gel digestion were compared on the basis of the number of identified transmembrane proteins. The 2-step HexaLane digestion was performed by digesting the same sample in two steps, thus resulting in two different tryptic fractions. The number of transmembrane proteins was roughly equal in all of the approaches (*see* **Fig. 9.1**). When comparing the in-solution approaches, the PPS protocol was better than the TFE method.

The overlap between the identified proteins in the different methods was also assessed (**Fig. 9.2**). Overall, the four different approaches appear to identify the same proteins, but each method appears to contribute with some unique identifications. When comparing HexaLane 2-step, in-solution PPS, and in-gel digestion, HexaLane 2-step (17.4%) and in-gel (13%) digestion make up the biggest part in the diagram. In the comparison of HexaLane single step, in-solution PPS, and in-gel digestion, TMP identification is more or less equal when using these three methods. Moreover, the operation time of these three methods is 4, 24, and 48 h, respectively. The results acquired by HexaLane digestion method and in-gel digestion are comparable; however, the HexaLane digestion method only requires 1/12 or 1/2 of the time taken by the in-gel protocol.

3.3.2. LC-MS Instrument Operation for Peptide Analysis

Sample (~1 μl) was injected onto an LC/MS system consisting of a 1200 Series liquid chromatograph, HPLC-Chip Cube MS interface, and a 6520 ESI-Q-TOF mass spectrometer (all Agilent Technologies). Chromatography was performed on an HPLC-Chip (Agilent Technologies) that incorporated a 40 nl enrichment column and a 150 mm × 75 μm analytical column packed with ZORBAX 300SB-C18, 5 μm particles. The tryptic peptides were loaded onto the enrichment column with 97% solvent A (water with 0.1% formic acid and 2.5% acetonitrile) and 3% B (acetonitrile with 0.1% formic acid and 5% water) at 4 μl/min. They were then eluted with a gradient from 3% B to 40% B in 60 min, at a flow rate of 300 nl/min. Washing of the column at 95% B was performed for 5 min, subsequently followed by equilibration (3% B) for 20 min. The total run time of the separation including column reconditioning was 85 min. Q-TOF MS/MS condition: drying gas: 5 l/min (350°C), fragmentor: 175 V, skimmer: 60 V, capillary voltage: 1,800 V, acquisition rate and time: 4 spectra/s (threshold 200 Abs, 0.01% rel.) and 250 ms/spectrum, MS scan range and rate: 296–2,500 at 4 Hz; MS/MS, acquisition rate and time: 3 spectra/s (threshold 5 Abs, 0.01% rel.) and 333.3 ms/spectrum, MS/MS scan range and rate: 50–2,500 at 3 Hz, collision energy: slope 3.3 V,

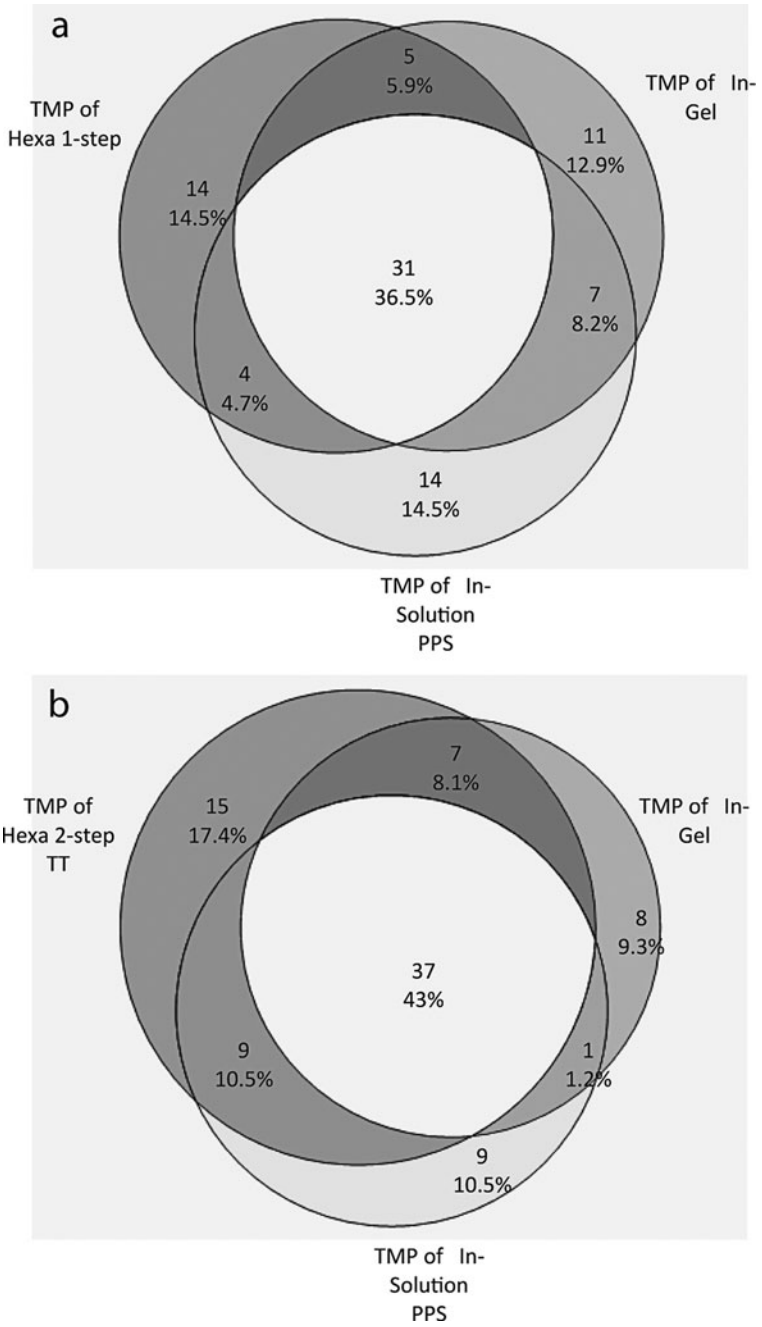


Fig. 9.1. Overlap of transmembrane protein (TMP) identifications between different methodologies. (**a**) Comparison between HexaLane 1-step digestion, in-gel digestion and in-solution digestion regarding the overlap of TMP identifications. (**b**) Comparison between HexaLane 2-step digestion, in-gel digestion and in-solution digestion regarding the overlap of TMP identifications.

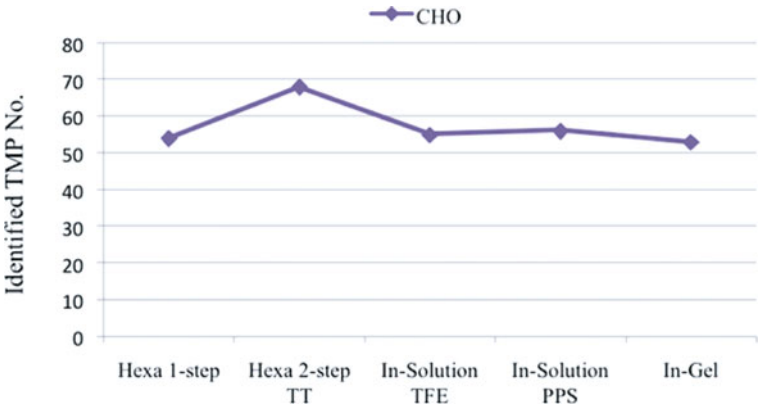


Fig. 9.2. Number of transmembrane protein (TMP) identifications depending on the method used.

offset 2.5 V, auto MS/MS: 5 precursor, active exclusion on with two repeat and with release after 0.1 min. Preferred charge state: 2, 3, >3, unknown. Internal reference mass correction was enabled.

3.3.3. Data Analysis Procedure

The raw data files were exported to an MGF format by MassHunter (Version B.02.00) to create peak lists on the basis of the recorded fragmentation spectra. Peptides and proteins were identified by Mascot V2.2.0 (Matrix Science, London, UK) against Swiss-Prot database release of May 11, 2009, with a precursor mass tolerance of 10 ppm, a fragment ion mass tolerance of ±0.3 Da, and strict trypsin specificity allowing for up to one missed cleavage. Carbamidomethylation of cysteine was set as a fixed modification, and methionine oxidation was allowed as a variable modification. Peptides were rejected if the Mascot score was less than the 95% confidence limit based on the "identity" score of each peptide, and a minimum of two different tryptic peptides were required for a positive protein identification. Mascot estimated the false-positive rates (FPR) by searching against a randomized decoy database. Mapping of the transmembrane (TM) domains for the identified proteins was conducted using the TM hidden Markov model (TMHMM) algorithm available at http://www.cbs.dtu.dk/services/TMHMM-2.0/, to which the FASTA files from the positively identified proteins were submitted. Information on the subcellular location of identified proteins was obtained from gene ontology (GO) component terms using GOSlim (http://www.geneontology.org). Sequence coverage was visualized with the TOPO2 TM protein graphics program (http://www.sacs.ucsf.edu/TOPO2/).

4. Notes

1. It is convenient to monitor both swelling and lysis of cells using a light microscope. The lysis buffer composition, duration of swelling, and number of Dounce strokes may need to be adjusted according to the cell type. The major portion of cells should be lysed and the nuclei should still be intact. Excessive Dounce treatment may lead to lysis of nuclei with a subsequent "gelation" of the sample, which should be minimized.
2. Centrifugation at 10,000×*g* for 10 min will pellet unbroken cells, intact nuclei, and mitochondria. Membranes remaining in the supernatant will mostly be from ER, Golgi, and plasma membrane. This fraction is also called the PNS (post-nuclear supernatant).
3. The high pH caused by Na_2CO_3 (pH > 11) will disrupt bonds between the membrane and the proteins associated with it by electrostatic interaction. Transmembrane proteins and lipid-anchored membrane protein will remain in the membrane.
4. There will be some loss of membrane sample during the membrane washing steps; so make sure you have enough starting material. Remember to check the pellet size in **Section 3.1**, Step 16. If you cannot see a pellet at that stage you have probably not enough material to start with.
5. Tip sonication should be performed at low temperature to prevent overheating of the sample. This can be done by immersing the sonication vessel in an ice bath. Sonication is preferably done in a glass or stainless steel vessel, since plastic vessels may absorb the vibrations. Use a probe sonicator with a tip compatible with the volume of your membrane preparation according to the manufacturer's specifications.
6. The protein concentration of the membrane preparation was in this study determined using the DC protein assay. The standard curve was made using samples with known concentrations of BSA: 0.25, 0.5, 1.0, and 1.5 mg/ml. BSA was diluted with ionic buffer (10 mM Tris–HCl, 300 mM NaCl, pH 8). The assay was performed in duplicates of each sample following the manufacturer's protocol. The absorbance was measured at 750 nm. The samples were stored at 4°C for further analysis.
7. Insert the pipette tip into the port on one side of the flow cell. Press lightly to obtain a good seal. About 50 μl will fill an empty lane and yield an excess of approximately 10 μl. The injection should be done in one movement with

duration of 5–10 s. Apply pressure gently to inject the liquid. Maintain pressure and observe the liquid's surface inside the tip until it has stopped to ensure that the entire volume is injected. Do not inject air. To prevent evaporation during incubation, make sure to use the protection stickers on top of the ports. If you are planning to use less than six lanes on the flow cell, remove individual stickers along the perforations.

8. For multistep protocols, using the same or other proteases, the protocol can be iterated several times. In this case, an extra wash step is required to be performed before the additional digestion step(s). This is done by injecting 800 µl Ambic and re-iterating Steps 4–6 of the protocol.

References

1. Ulmschneider, M. B., Sansom, M. S., and Di Nola, A. (2005) Properties of integral membrane protein structures: derivation of an implicit membrane potential. *Proteins* 59:252–265.
2. Terstappen, G. C., and Reggiani, A. (2001) In silico research in drug discovery. *Trends Pharmacol Sci* 22(1):23–26.
3. Macher, B. A., Yen, T. Y. (2007) Protein at membrane surfaces-a review of approaches. *Mol Biosyst* 3:705–713.
4. Santoni, V., Molloy, M., and Rabilloud, T. (2000) Membrane proteins and proteomics: Un amour impossible? *Electrophoresis* 21:1054–1070.
5. Stevens, T. J., and Arkin, I. T. (2000) Do more complex organisms have a greater proportion of membrane proteins in their genomes? *Proteins* 39:417–420.
6. Wallin, E., and Heijine, G. (1998) Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms. *Protein Sci* 7:1029–1038.
7. Blonder, J., Conrads, T. P., Yu, L. R., Terunuma, A., Janini, G. M., Issaq, H. J., Vogel, J. C., and Veenstra, T. D. (2004) A detergent– and cyanogen bromide- free method for integral membrane proteomics: application to Halobacterium purple membranes and the human epidermal membrane proteome. *Proteomics* 4(1):31–45.
8. Chen, E. I., Cociorva, D., Norris, J. L., and Yates, J. R. (2007) Optimization of mass spectrometry-compatible surfactants for shotgun proteomics. *J Proteome Res* 6(7):2529–2538.
9. Mitra, S. K., Gantt, J. A., Ruby, J. F., Clouse, S. D., and Goshe, M. B. (2007) Membrane proteomic analysis of Arabidopsis thaliana using alternative solubilization techniques. *J Proteome Res* 6(5):1933–1950.
10. Russell, W. K., Park, Z. Y., and Russel, D. H. (2001) Proteolysis in mixed organic-aqueous solvent systems: applications for peptide mass mapping using mass spectrometry. *Anal Chem* 73(11):2682–2685.
11. Soskic, V., and Godovac-Zimmermann, J. (2001) Improvement of an in-gel tryptic digestion method for matrix-assisted laser desorption/ionization-time of flight mass spectrometry peptide mapping by sue of volatile solubilizing agents. *Proteomics* 1(11):1364–1367.
12. Zhang, N., Chen, R., Young, N., Wishart, D., Winter, P., Weiner, J. H., and Li, L. (2007) Comparison of SDS- and methanol-assisted protein solubilization and digestion methods for Escherichia coli membrane proteome analysis by 2-D LC-MS/MS. *Proteomics* 7(4):484–493.
13. Washburn, M. P., Wolters, D., and Yates, J. R. (2001) Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat Biotechnol* 19(3):242–247.
14. Martosella, J., Zolotarjova, N., Liu, H., Moyer, S. C., Perkins, P. D., and Boyes, B. E. (2006) High recovery HPLC separation of lipid rafts for membrane proteome analysis. *J Proteome Res* **5**(6):1301–1312.
15. Da Cruz, S., Xenarios, I., Langridge, J., Vilbois, F., Parone, P. A., and Marinou, J. C. (2003) Proteomic analysis of the mouse liver mitochondrial inner membrane. *J Biol Chem* 278(42):41566–41571.
16. Lasonder, E., Ishihama, Y., Andersen, J. S., Vermunt, A. M., Pain, A., Sauerwein, R. W., Eling, W. M., Hall, N., Waters, A. P., Stun-

nenberg, H. G., and Mann, M. (2002) Analysis of the Plasmodium falciparum proteome by high-accuracy mass spectrometry. *Nature* 419(6909):537–542.
17. Mhatre, R., Woodard, J. and Zeng, C. (1999) Strategies for locating disulfide bonds in a monoclonal antibody via mass spectrometry. *Rapid Commun Mass Spectrom* 13(24):2503–2510.
18. Zhang, W., Marzilli, L. A., Rouse, J. C., and Czupryn, M. J. (2002) Complete disulfide bond assignment of a recombinant immunoglobulin G4 monoclonal antibody. *Anal Biochem* 311(1):1–9.
19. Quach, T. T., Richards, D. P., Zheng, J., Keller, B. O., and Li, L. (2003) Development and applications of in-gel CNBr/tryptic digestion combined with mass spectrometry for the analysis of membrane proteins. *J Proteome Res* 2(5):543–552.
20. Ball, L. E., Oatis, J. E. Jr., Dharmasiri, K., Busman, M., Wang, J., Cowden, L. B., Galijatovic, A., Chen, N., Crouch, R. K., and Knapp, D. R. (1998) Mass spectrometric analysis of integral membrane proteins: application to complete mapping of bacteriorhodopsins and rhodopsin. *Protein Sci* 7(3):758–764.
21. Wu, C. C., and Yates, J. R. (2003) The application of mass spectrometry to membrane proteomics. *Nat Biotechnol* 21(3):262–267.
22. Rodriguez-Ortega, M. J., Norais, N., Bensi, G., Liberatori, S., Capo, S., Mora, M., Scarselli, M., Doro, F., Ferrari, G., Garaguso, I., Maggi, T., Neumann, A., Covre, A., Telford, J. L., and Grandi, G. (2006) Characterization and identification of vaccine candidate proteins through analysis of the group A Streptococcus surface proteome. *Nat Biotechnol* 24(2): 191–197.
23. Yu. Y. Q., Gilar, M., and Gebler, J. C. (2004) A complete peptide mapping of membrane proteins: a novel surfactant aiding the enzymatic digestion of bacteriorhodopsin. *Rapid Commun Mass Spectrom* 18:711–715.
24. Norris, J. L., Porter, N. A., and Caprioli, R. M. (2003) Mass spectrometry of intracellular and membrane proteins using cleavable detergents. *Anal Chem* 75(23): 6642–6647.
25. Lochner, N., Pittner, F., Wirth, M., and Gabor, F. (2003) Preparation, characterization and application of artificial Caco-2 cell surfaces in the silver nanoparticle enhanced fluorescence technique. *J Control Release* 89:249–259.
26. Speer, A. E. and Wu, C. C. (2007) Proteomics of integral membrane proteins – theory and application. *Chem Rev* 107:3687–3714.

Chapter 10

GeLCMS for In-Depth Protein Characterization and Advanced Analysis of Proteomes

Alicia Lundby and Jesper V. Olsen

Abstract

In recent years the array of mass spectrometry (MS) applications to address questions in molecular and cellular biology has greatly expanded and continues to grow. Modern mass spectrometers allow for identification, characterization, as well as quantification of protein compositions and their modifications in complex biological samples. Prior to MS analysis any biological sample needs to be properly prepared for the experiment. Here we present a protocol that combines pre-separation of proteins by 1D gel electrophoresis followed by analysis of in situ digested protein products by tandem mass spectrometry (MS/MS). All steps of the sample preparation are explained in detail, and the procedure is compatible with downstream analysis on any mass spectrometer available. With minor adjustments the protocol can be used with 2D gels as well. The protocol provided can be applied to analyze specific proteins of particular interest as well as entire proteomes. If SILAC-labeled protein samples are mixed prior to gel separation, the protein content of the sample can furthermore be accurately quantified.

Key words: In-gel digest, mass spectrometry, proteomics, LTQ-Orbitrap, Velos, protein identification, post-translational modifications, LC-MS/MS, GeLCMS.

1. Introduction

Mass spectrometry is a powerful tool for the analysis of complex protein samples (1). The technique can be applied to investigate a multitude of biological questions covering a diverse array of research fields. Advances made in recent years have manifested its applicability in biological research, with MS-based studies characterizing complete cellular proteomes (2), identifying novel protein–protein interactions in an unbiased manner (3), and providing large-scale mapping of changes to post-translational

K. Gevaert, J. Vandekerckhove (eds.), *Gel-Free Proteomics*, Methods in Molecular Biology 753,
DOI 10.1007/978-1-61779-148-2_10,

modifications such as phosphorylation (4) or lysine acetylation sites (5) in response to a given stimulus. The range of potential applications of mass spectrometry-based proteomics to address questions in molecular and cellular biology seems to be ever expanding.

Prior to MS analysis the biological sample of interest needs to be appropriately prepared for the experiment. This usually involves fractionation to reduce sample complexity and enzymatic digestion of proteins into peptides by a specific protease such as trypsin (6). There are several different methods available for protein and peptide fractionation, which traditionally involves gel electrophoresis, ion-exchange, or reversed-phase chromatography. Here we present a protocol for gel-enhanced liquid chromatography mass spectrometry (GeLCMS) that combines preseparation of proteins by gel electrophoresis followed by analysis of the peptides resulting from an in-gel digest step by online high-performance liquid chromatography-tandem mass spectrometry (HPLC-MS/MS) (**Fig. 10.1**). All steps of sample preparation are explained in detail, and the procedure is compatible with downstream analysis on any electrospray mass spectrometer available. The provided protocol can be applied to analyze specific proteins of particular interest as well as entire proteomes. If stable isotope labeling by amino acids in cell culture (SILAC) (7) encoded protein samples are mixed prior to gel separation, the protein content of the sample can be both identified and also accurately quantified (8). SILAC-based approaches enable quantification of relative changes in protein expression levels as well as protein modification changes in response to a certain stimulus, as for instance application of growth factors (9) or drugs (10).

The power of gel electrophoresis-based separation of proteins in combination with mass spectrometric identification of proteins was first shown in a classic study by Shevchenko, Mann, and coworkers (11) where they presented an in-gel digestion protocol that enabled MS identification of proteins from Coomassie and silver-stained gels. The main advantages of identifying proteins from polyacrylamide gels are that SDS-PAGE is the highest resolution separation method for most proteins, and that gels in general are good containers for handling, concentrating, and storing proteins down to the femtomolar range (12). In addition, polyacrylamide gels provide efficient filtering of low molecular weight impurities from the samples, such as detergents and buffer components. The gel visualization step in the protocol is furthermore very informative, as it provides a measure to estimate relative protein abundance in the sample, which is not directly provided by the downstream MS measurements. However, if efforts are made to, for instance, characterize complete proteomes, it should be kept in mind that the estimated recovery of in-gel digested peptides is lower than that of in-solution digest (13). In general, it is

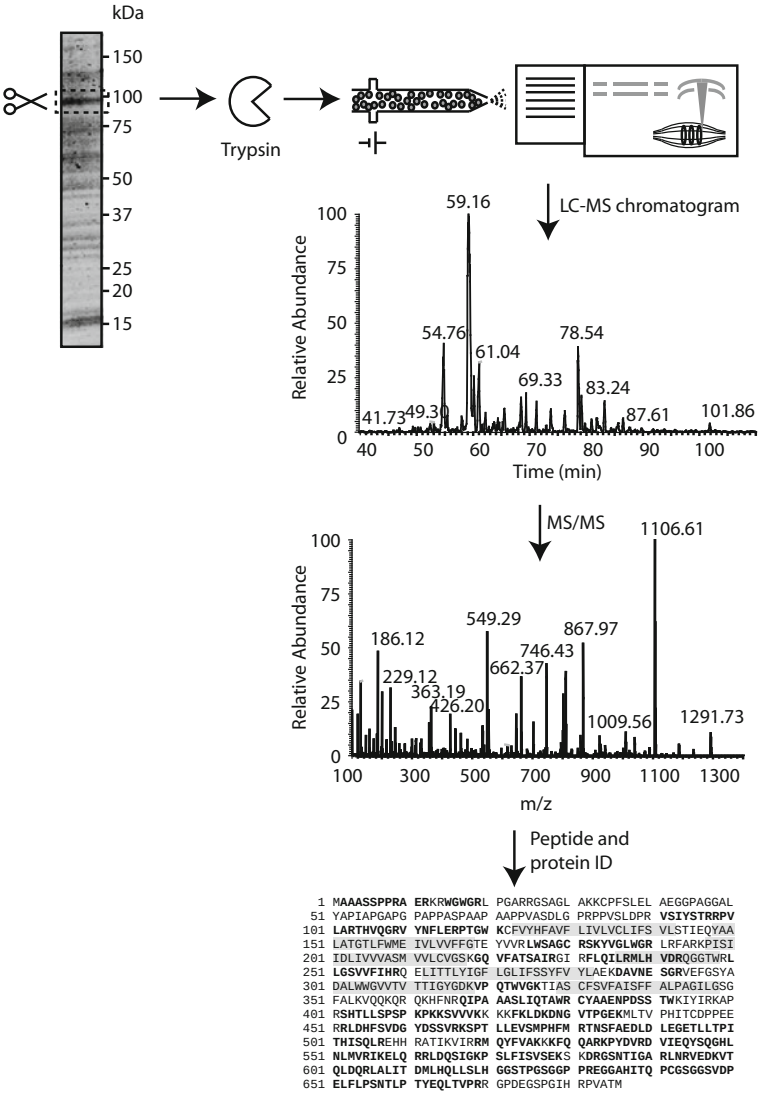


Fig. 10.1. Outline of workflow for protein identification by in-gel digest and LC-MS/MS. A GFP immunoprecipitation product from an MDCK cell line stably expressing a human Kv7.1 GFP fusion protein was separated by SDS-PAGE and the gel was stained with colloidal Coomassie and proteins were fixed in the gel. The indicated band was excised and in-gel digested by trypsin. Extracted peptides were subsequently analyzed by LC-MS/MS on a Thermo Scientific LTQ-Orbitrap Velos instrument providing high-resolution MS and MS/MS spectra. A total of 62 peptides identified the protein of interest, leading to total protein sequence coverage of 60.1%. Putative transmembrane segments are indicated as *boxed regions* and identified peptides are highlighted in *bold*.

beneficial to separate complex protein mixtures into several fractions to reduce sample complexity and increase the dynamic range of peptide identification. Splitting the proteins to be analyzed in several gel slices, which are individually processed, increases the number of proteins and post-translational modifications that

can be identified. A considerable disadvantage of the in-gel protein digestion procedure compared to that of in-solution digest is the risk of contaminating the samples with, for instance, skin and wool keratins, when handling or excising protein bands from the gel. In standard laboratories keratin contaminations are difficult to avoid, and it is important to consider peptide identifications from keratins as well as from enzymatic autolysis of the digestion enzyme used as a source of "false positives" when analyzing the data.

In brief, the workflow of the procedure is as follows. Once a biological sample of interest is collected, the proteins of the sample are separated according to their molecular weight by SDS-PAGE. Following the gel separation step the proteins are fixed in the gel and then visualized, for example, by colloidal Coomassie staining. Next, either selected protein bands are excised or entire gel lanes are separated into gel slices, followed by in-gel digestion by site-specific proteases such as trypsin or endoproteinase Lys-C. The resulting peptides are extracted from the gel plugs by acidic buffers, followed by a desalting and concentration step before LC-MS/MS analysis. At this point the peptide mixture is separated by reversed-phase C_{18} HPLC in microbore columns by a linear gradient of increasing acetonitrile in acidified water and the column effluent is directly electrosprayed in the MS source. In the mass spectrometer, the peptide mixture is analyzed by full-scan mass measurements and the most abundant peptides are fragmented in turn by collision-induced dissociation (CID). The peptide sequences and their modifications can be identified from the resulting MS/MS spectra by in silico matching these spectra against a protein sequence database using peptide search algorithms such as Mascot (14) and Sequest (15). Quantification and statistical evaluation of the resulting peptide data sets are then performed in dedicated software suites such as MaxQuant (16, 17) or MSQuant (18).

2. Materials

All solvents in the protocol are prepared with ultrapure water of 18.2 MΩ cm resistivity (PURELAB Ultra, ELGA LabWater).

2.1. Commonly Used Buffers

1. Buffer ABC: 25 mM ammonium bicarbonate (NH_4HCO_3) in H_2O.
2. Buffer A: 0.5% (v/v) acetic acid in H_2O.
3. Buffer B: 80% (v/v) acetonitrile and 0.5% (v/v) acetic acid in H_2O.

4. Buffer A*: 2% (v/v) acetonitrile and 0.1% (v/v) trifluoroacetic acid (TFA) in H_2O (*see* **Note 1**).

2.2. SDS-PAGE

1. Running buffer (5X): 125 mM Tris, 960 mM glycine, 0.5% (w/v) SDS in H_2O (*see* **Note 2**).
2. Prestained molecular weight markers such as the kaleidoscope marker (Bio-Rad).
3. Sample buffer: 62 mM Tris–HCl, pH 6.8, 10% glycerol, 2% SDS, 5% β-mercaptoethanol, and 0.00125% bromophenol blue in H_2O (*see* **Note 3**).
4. Precast NuPAGE Novex Bis-Tris gels (Invitrogen), XCell SureLock gel runners (Invitrogen), and PowerPac HC power supply (Bio-Rad Laboratories) or similar gel systems available.

2.3. Visualization of Protein Bands

1. Novex Colloidal Blue Staining Kit (Invitrogen).
2. Fixing solution: 40 ml H_2O, 50 ml methanol, and 10 ml acetic acid.
3. Staining solution: 55 ml H_2O, 20 ml methanol, and 20 ml Stainer A (from Novex Colloidal Blue Staining Kit).
4. Washing solution I: 50% (v/v) 25 mM buffer ABC and 50% (v/v) acetonitrile.

2.4. In-Gel Protein Digest

1. Washing solution II: 30% (v/v) acetonitrile and 3% (v/v) trifluoroacetic acid in H_2O.
2. Alkylation buffer (CAA): 550 mM chloroacetamide in buffer ABC (*see* **Note 4**).
3. Reduction buffer (DTT): 1 M dithiothreitol in buffer ABC (*see* **Note 5**).
4. Trypsin (modified sequencing grade, Promega): resuspend trypsin in buffer ABC at 12.5 ng/μl (*see* **Note 6**).

2.5. Desalting and Concentrating Peptides on Stage Tips

1. C_{18} discs (3 M).
2. 96-well elution plates (Thermo Scientific) and elution plate covers.

3. Methods

Prior to following the protocol described below, the biological sample to be analyzed should be prepared such that it is ready to be run on an SDS-PAGE gel (*see* **Note 7**). All steps take place at room temperature unless stated otherwise.

3.1. SDS-PAGE

1. Boil the sample in sample buffer for 10 min at 95°C.
2. Prepare 700 ml running buffer (dilute in H_2O).
3. Remove the plastic cover and the white plastic tape at the bottom from the gel.
4. Put the gel in the running chamber with loading wells backward and lock the gel in position.
5. Add 200 ml of running buffer to the inner chamber and 500 ml of the running buffer to the outer chamber.
6. Carefully remove the comb from the gel and use a pipette to wash the wells with running buffer.
7. Load 25–35 μl of sample in each well. Remember to include at least one well for the prestained molecular weight markers.
8. Completely assemble the gel unit, connect to a power supply, and run the gel at 200 V (*see* **Note 8**).
9. Run the gel for approximately 55 min or until the dye front reaches the bottom of the gel (*see* **Note 9**).

3.2. Visualization of Protein Bands

Various staining methods can successfully be applied, including Coomassie Brilliant Blue R-250 or G-250 or silver staining. Here we describe visualization of protein bands by colloidal Coomassie staining (*see* **Note 10**).

1. Prepare 100 ml fixing solution in a clean plastic box.
2. Crack the gel free from the hard plastic cover, place it in fixing solution, and let it shake on a rocking table for 10 min.
3. Remove the fixing solution and replace it with staining solution. Let the gel shake in staining solution on a rocking table for 5 min, then add 5 ml of Stainer B.
4. Leave the gel for staining on a rocking table for 3–17 h.
5. Destain the gel in H_2O for at least 7 h, exchanging the water multiple times.
6. Excise protein bands of interest or cut the entire gel lane into slices with a scalpel and place gel pieces in 1.5 ml Eppendorf tubes containing 500 μl of washing solution I (*see* **Notes 11** and **12**).

3.3. In-Gel Protein Digest

1. Destain gel pieces: wash samples three times for 20 min in 500 μl washing solution I on an Eppendorf Thermomixer at 1,200 rpm to remove all protein-bound dye (*see* **Note 13**).
2. Dehydrate gel pieces by incubating them in 500 μl acetonitrile on an Eppendorf Thermomixer at 1,200 rpm for

10 min or until the gel pieces are shrunk and turned white opaque (*see* **Note 14**).

3. Reduce disulfide bonds by re-swelling the gel pieces in 50 µl of freshly made 10 mM DTT in buffer ABC on an Eppendorf Thermomixer at 1,200 rpm for 45 min (*see* **Note 15**).
4. Alkylate cysteines to prevent reformation of disulfide bonds by incubating the gel pieces in same volume as used in the previous step of freshly made 55 mM CAA in buffer ABC on an Eppendorf Thermomixer at 1,200 rpm for 30 min. Cover the Thermomixer with tin foil to allow the reaction to proceed in dark (*see* **Notes 16** and **17**).
5. Wash the samples twice with 500 µl of buffer ABC on an Eppendorf Thermomixer at 1,200 rpm for 10 min.
6. Dehydrate the gel pieces in 500 µl of acetonitrile on an Eppendorf Thermomixer at 1,200 rpm for 10 min or until the gel pieces turn white opaque.
7. Remove all acetonitrile and add 10–50 µl of the trypsin solution to each gel piece. Digest for 6 h or overnight on an Eppendorf Thermomixer at 500 rpm at 37°C or alternatively at room temperature (*see* **Note 18**).
8. Label a new set of 1.5 ml Eppendorf tubes for collection of digested peptides and add 10% of trifluoroacetic acid in H_2O to each sample (the volume added should result in a final trifluoroacetic acid concentration of 3%). Centrifuge samples for 2 min at $0.5\times g$ and transfer each supernatant to a new 1.5 ml Eppendorf tube (*see* **Note 19**).
9. Add 20–50 µl of washing solution II to the gel pieces and mix for 30 min at 800 rpm on an Eppendorf Thermomixer. Centrifuge samples for 2 min at $0.5\times g$ and transfer supernatants to the same 1.5 ml Eppendorf tube as in Step 8 (*see* **Notes 20** and **21**).
10. Add 20–50 µl of buffer B to the gel pieces and mix for 30 min at 800 rpm on an Eppendorf Thermomixer. Centrifuge samples for 2 min at $0.5\times g$ and transfer supernatants to the same 1.5 ml Eppendorf tube as in Step 8.
11. Add 20–50 µl of acetonitrile to the gel pieces and mix for 30 min at 800 rpm on an Eppendorf Thermomixer. Centrifuge samples for 2 min at $0.5\times g$ and transfer supernatants to the same 1.5 ml Eppendorf tube as in Step 8 (*see* **Note 22**).
12. Remove organic solvents from the collected samples by evaporation using a vacuum centrifuge until all acetonitrile has evaporated (*see* **Note 23**).

3.4. Desalting and Concentrating Peptides on Stage Tips

1. Prepare stage tips by adding two C_{18} discs to a 200 µl pipette tip. One pipette tip is used per sample (*see* **Note 24**).
2. Place each stage tip in a custom-made holder inside a 2 ml Eppendorf tube, add 20 µl of methanol to the end of the pipette tip, and wash the filter by centrifugation at 0.5×*g* for 2 min.
3. Wash stage tips with 20 µl of buffer B using the same procedure as in Step 2.
4. Re-equilibrate the stage tips by adding two times 20 µl of buffer A using the same procedure as in Step 2.
5. Load the extracted peptide sample to the stage tip (*see* **Note 25**) and centrifuge at 0.5×*g* until the solution has spun through the filter.
6. Wash the stage tips once with 50 µl of buffer A (*see* **Note 26**).
7. Elute peptides into a 96-well microtiter plate (*see* **Note 27**) by adding two times 10 µl of buffer B to the back of the pipette and slowly forcing the solution through the filter and into the chosen well of the plate by applying back pressure with a 10 ml syringe (*see* **Note 28**).
8. Vacuum dry the plate until all organic material has evaporated (*see* **Note 29**).
9. Add 5 µl of buffer A* to each sample-containing well.

The samples are now ready to be analyzed by LC-MS/MS.

4. Notes

1. All buffers should be prepared and stored in glassware. Make 500 ml stock solutions of buffers A and ABC and 100 ml stock solutions of buffers B and A*. All buffers, except buffer B, can be stored at room temperature for several months. Buffer B should be made on a weekly basis due to evaporation.
2. Mix all ingredients except SDS and store at room temperature. Add SDS to a final concentration of 0.1% when the 1X buffer is made, just prior to use.
3. Mix all ingredients except β-mercaptoethanol and store 500 µl aliquots at –20°C for long-term storage. Once β-mercaptoethanol has been added the buffer can be kept at 4°C for 1 month. Let the buffer adjust to room temperature before use, as it is difficult to pipette otherwise.

4. Make 200 μl aliquots and store at –20°C. Thaw one vial just before use, do not re-freeze.
5. Make 50 μl aliquots and store at –20°C. Thaw one vial just before use, do not re-freeze.
6. Lyophilized trypsin is stored at –80°C. Trypsin should be resuspended in buffer ABC just prior to use. If aliquots are made, these can be stored at –20°C for 1 month. Cycles of thawing and freezing should be avoided. Buffer ABC is used as it provides an optimal pH range for trypsin.
7. In general, the presence of detergents in the sample diminishes the quality of the LC-MS/MS spectra that can be obtained, as detergents ionize well and bind to the LC column and interfere with the number of peptides that can be identified. Thus, it is preferential to use as low detergent concentrations as possible, even with the polyacrylamide gel as low molecular weight filter. The salt concentrations used for cell lysis and protein extraction are not as critical as they are for in-solution digest, as the salts are also filtered away by the gel matrix. If necessary, high concentrations of urea can be used in preceding steps to denature proteins without interfering with the LC-MS/MS data, as urea will effectively be filtered by the gel.
8. Check that the current is 100–125 mA at the beginning of the run.
9. If desired, larger quantities of sample can be loaded by stopping the gel when samples are concentrated into a single band (after about 2 min) and adding another 35 μl of sample to each well.
10. If protein bands are visualized by colloidal Coomassie or silver staining methods, the proteins need to be fixed in the gel matrix prior to visualization. Proteins stained with Coomassie Brilliant Blue are automatically fixated. Any staining method can be used. However, reagents that introduce covalent modifications to proteins should be avoided. Thus, gels should not be treated with any cross-linking reagent or strong oxidizers.
11. If the gel pieces are kept in 500 μl of H_2O, they can be stored at 4°C for several weeks before sample processing.
12. At all stages preceding protein digestion, particular care should be taken to reduce contamination of the samples with proteins such as keratin as much as possible. For protein band excision it is recommended to use a new disposable scalpel every time, change gloves frequently, and rinse glassware with ethanol prior to use. To excise the bands of interest as precisely as possible, it is recommended to

put the gel on a glass plate and place this plate on a light box. The gel should not be allowed to dry out while it is on the light box, so keep it wet. The efficiency of the downstream protein digest is improved the smaller the gel pieces are. Aim for pieces of about 2 × 1 mm. Cutting the pieces smaller than this might hamper pipetting in subsequent steps.

13. The duration of the washing steps depends on the intensity of the dye used for the protein stain. Washing is sufficient when the gel pieces are transparent with no traces of dye remaining.
14. In general, gel pieces are dehydrated to aid the uptake of the reagent added in the following step.
15. The volume added should be large enough to cover the gel slice, so for larger gel pieces, larger volumes will have to be added. If the SDS-PAGE sample buffer contains 1 mM of DTT or more, then this step of the protocol can be omitted.
16. It is important to remove as much of the DTT-containing solution as possible before CAA is added, as CAA will alkylate any remaining DTT before alkylating proteins in the gel. This step can be omitted from the protocol if the SDS-PAGE sample buffer contained 5 mM CAA. It is often advantageous to perform the reduction and alkylation step prior to loading proteins on the gel, as this significantly reduces gel processing time and minimizes cysteine oxidation.
17. In many protein digestion protocols iodoacetamide (IAA) is used as the alkylating agent to block cysteine residues. However, iodoacetamide can make a covalent adduct to lysine residues by attachment of acetamide molecules that mimic a lysine diglycine tag. To avoid introduction of such in vitro artifacts, it is important to use chloroacetamide instead of iodoacetamide (19).
18. The volume of added trypsin solution should be large enough to cover the gel pieces. At the same time the volumes added should be kept as small as possible, as an excess of trypsin will cause trypsin autolysis, and peptides generated from this digest will interfere with the MS analysis of the peptides of interest. To keep gel pieces wet overnight, two volumes of buffer ABC can be added. However, it is beneficial to keep the total volume low to keep the concentration of trypsin high to ensure efficient protein cleavage. The procedure outlined can be applied to prepare gel pieces for digest by any protease. Beware that the efficiency of in-gel digestion decreases with increasing

molecular weight of the protease. If using a different protease than trypsin, it should also be kept in mind that much greater enzyme concentrations might be used in this protocol, as for instance compared to in-solution digestion, since the enzymatic digestion relies on enzyme diffusion into the gel matrix.

19. If the start volume is large, add trifluoroacetic acid. The important thing is that the final concentration is about 3%. A common volume to add is 20 μl.
20. The volume added should be as small as possible, but large enough to cover the gel pieces.
21. All extracted peptides from each gel piece are pooled in one Eppendorf tube.
22. Although the extracted gel pieces are not used further in this protocol, it is a good idea to save them until MS analysis has been completed. In case digestion fails, the procedure can be repeated using the same gel pieces and either the same or another digestion enzyme.
23. It is important to remove as much of acetonitrile as possible, as any acetonitrile remaining in the sample will reduce binding of peptides to the column in the downstream liquid chromatography. One option is to calculate the amount of acetonitrile in each collected sample and vacuum dry till a greater volume than this has evaporated. Beware that the samples do not dry out, as they will be difficult to redissolve. It is common to vacuum dry till 10–20 μl of sample remains.
24. "Stage" tip is an abbreviation for "stop and go extraction" tip (20). The principle behind the stage tips is that C_{18} filters are hydrophobic and bind peptides. The peptides can be eluted from the filter by applying a hydrophobic solvent such as acetonitrile or methanol. For those who do not wish to prepare their own stage tips, they are also commercially available at Proxeon Biosystems.
25. Make sure the filter is still wet from the re-equilibration in buffer A when the sample is added. If not, re-equilibrate once more.
26. At this point the stage tips can be stored at 4°C for months before proceeding with the next step if necessary. The function of stage tips is thus dual: they are used to purify the peptide mixture and they can be used as a peptide storage device.
27. Buffer B is very hydrophobic, and flow-through of buffer B therefore leads to elution of the peptides from the C_{18} discs.

28. Make sure not to introduce air bubbles. If air bubbles are present, remove them by gently tapping the plate against the table.
29. If the sample has been eluted with 20 μl of buffer B, this means vacuum drying until a volume of less than 4 μl remains.

Acknowledgments

The authors thank all members of the Department for Proteomics at the Novo Nordisk Foundation Center for Protein Research for helpful scientific discussions and in particular Brian Tate Weinert for feedback on the chapter. This work was in part supported by the Danish National Research Foundation Centre for Cardiac Arrhythmia. The NNF Center for Protein Research is supported by a generous donation from the Novo Nordisk Foundation.

References

1. Aebersold, R., and Mann, M. (2003) Mass spectrometry-based proteomics, Nature **422**, 198–207.
2. de Godoy, L. M., Olsen, J. V., Cox, J., Nielsen, M. L., Hubner, N. C., Frohlich, F., Walther, T. C., and Mann, M. (2008) Comprehensive mass-spectrometry-based proteome quantification of haploid versus diploid yeast, Nature **455**, 1251–1254.
3. Berkefeld, H., Sailer, C. A., Bildl, W., Rohde, V., Thumfart, J. O., Eble, S., Klugbauer, N., Reisinger, E., Bischofberger, J., Oliver, D., Knaus, H. G., Schulte, U., and Fakler, B. (2006) BKCa-Cav channel complexes mediate rapid and localized Ca^{2+}-activated K+ signaling, Science **314**, 615–620.
4. Olsen, J. V., Blagoev, B., Gnad, F., Macek, B., Kumar, C., Mortensen, P., and Mann, M. (2006) Global, in vivo, and site-specific phosphorylation dynamics in signaling networks, Cell **127**, 635–648.
5. Choudhary, C., Kumar, C., Gnad, F., Nielsen, M. L., Rehman, M., Walther, T. C., Olsen, J. V., and Mann, M. (2009) Lysine acetylation targets protein complexes and co-regulates major cellular functions, Science **325**, 834–840.
6. Olsen, J. V., Ong, S. E., and Mann, M. (2004) Trypsin cleaves exclusively C-terminal to arginine and lysine residues, Mol Cell Proteomics **3**, 608–614.
7. Ong, S. E., Blagoev, B., Kratchmarova, I., Kristensen, D. B., Steen, H., Pandey, A., and Mann, M. (2002) Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics, Mol Cell Proteomics **1**, 376–386.
8. Blagoev, B., Kratchmarova, I., Ong, S. E., Nielsen, M., Foster, L. J., and Mann, M. (2003) A proteomics strategy to elucidate functional protein–protein interactions applied to EGF signaling, Nat Biotechnol **21**, 315–318.
9. Blagoev, B., Ong, S. E., Kratchmarova, I., and Mann, M. (2004) Temporal analysis of phosphotyrosine-dependent signaling networks by quantitative proteomics, Nat Biotechnol **22**, 1139–1145.
10. Bose, R., Molina, H., Patterson, A. S., Bitok, J. K., Periaswamy, B., Bader, J. S., Pandey, A., and Cole, P. A. (2006) Phosphoproteomic analysis of Her2/neu signaling and inhibition, Proc Natl Acad Sci USA **103**, 9773–9778.
11. Shevchenko, A., Wilm, M., Vorm, O., and Mann, M. (1996) Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels, Anal Chem **68**, 850–858.

12. Wilm, M., Shevchenko, A., Houthaeve, T., Breit, S., Schweigerer, L., Fotsis, T., and Mann, M. (1996) Femtomole sequencing of proteins from polyacrylamide gels by nano-electrospray mass spectrometry, Nature **379**, 466–469.
13. Wisniewski, J. R., Zougman, A., Nagaraj, N., and Mann, M. (2009) Universal sample preparation method for proteome analysis, Nat Methods **6**, 359–362.
14. Perkins, D. N., Pappin, D. J., Creasy, D. M., and Cottrell, J. S. (1999) Probability-based protein identification by searching sequence databases using mass spectrometry data, Electrophoresis **20**, 3551–3567.
15. Yates, J. R., 3rd, Eng, J. K., McCormack, A. L., and Schieltz, D. (1995) Method to correlate tandem mass spectra of modified peptides to amino acid sequences in the protein database, Anal Chem **67**, 1426–1436.
16. Cox, J., and Mann, M. (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification, Nat Biotechnol **26**, 1367–1372.
17. Cox, J., Matic, I., Hilger, M., Nagaraj, N., Selbach, M., Olsen, J. V., and Mann, M. (2009) A practical guide to the MaxQuant computational platform for SILAC-based quantitative proteomics, Nat protocols **4**, 698–705.
18. Mortensen, P., Gouw, J. W., Olsen, J. V., Ong, S. E., Rigbolt, K. T., Bunkenborg, J., Cox, J., Foster, L. J., Heck, A. J., Blagoev, B., Andersen, J. S., and Mann, M. (2010) MSQuant, an open source platform for mass spectrometry-based quantitative proteomics, J Proteome Res **9**(1), 393–403.
19. Nielsen, M. L., Vermeulen, M., Bonaldi, T., Cox, J., Moroder, L., and Mann, M. (2008) Iodoacetamide-induced artifact mimics ubiquitination in mass spectrometry, Nat Methods **5**, 459–460.
20. Rappsilber, J., Mann, M., and Ishihama, Y. (2007) Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using stage tips, Nat Protoc **2**, 1896–1906.

Chapter 11

Exploring New Proteome Space: Combining Lys-N Proteolytic Digestion and Strong Cation Exchange (SCX) Separation in Peptide-Centric MS-Driven Proteomics

Nadia Taouatas, Shabaz Mohammed, and Albert J.R. Heck

Abstract

The current advances in mass spectrometry technology have led to the possibility of analyzing more complex biological samples such as entire proteomes. Here, we describe a new and powerful methodology that combines the use of the metalloendopeptidase Lys-N and strong cation exchange with mass spectrometric analysis. The approach described here allows one to separate peptides with different functional groups. The peptides we are able to isolate are N-terminal peptides, phosphorylated peptides with a single lysine, peptides with a single basic residue (lysine), and peptides with multiply basic residues. When this separation strategy is combined with tandem mass spectrometry that involves both collision-induced dissociation and electron transfer dissociation, one can achieve an optimal targeted strategy for proteome analysis.

Key words: Strong cation exchange (SCX), Lys-N, mass spectrometry, tandem mass spectrometry, electron transfer dissociation.

1. Introduction

The complexity of the proteome is staggering: thousands of proteins with many of them having multiple splice variants and isoforms. Although attempts have been made to separate full proteomes at the protein level, either by 2D gels or chromatography, recently a more peptide-centric approach, combining proteolytic cleavage of the lysate with LC-MS/MS at the peptide level, has gained importance. The advent of powerful peptide separation strategies that can be combined and hyphenated and directly coupled to mass spectrometry has allowed the possibility

K. Gevaert, J. Vandekerckhove (eds.), *Gel-Free Proteomics*, Methods in Molecular Biology 753,
DOI 10.1007/978-1-61779-148-2_11,

to screen proteomes post-proteolysis (1). The challenge, at the peptide level, is the need to separate hundreds of thousands of peptides. Several multidimensional LC methods have been introduced to reduce sample complexity prior to MS analysis in order to decrease under-sampling by the mass spectrometer and to increase the level of proteome characterization (2–4). One of the earliest approaches that was introduced, notably still one of the most powerful for such a task, is MudPIT (multidimensional protein identification technology). The technique couples 2D-LC to MS/MS using a bi- or triphasic microcapillary column packed with strong cation exchange (SCX) beads and reversed-phase (RP) beads (3, 5). Currently, this is one of the most widely used approaches for separation of peptides of a complex sample. It has a broad range of application areas not only for membrane protein analysis (6) and phosphoproteome analysis (7–9) but also, for instance, for the targeted isolation of N-acetylated protein N-termini (10).

The most typical protease used for peptide separation strategies is trypsin, an enzyme that cleaves adjacent C-terminal to lysine and arginine. A common reason for such a choice is the preferred option for mass spectrometric sequencing, collision-induced dissociation (CID) (11). Tryptic peptides often generate MS/MS spectra which are predictable and thus easy to interpret. Although trypsin is the most frequently used protease in the field of proteomics, it is occasionally necessary to use other proteases or combinations of different enzymes for a more comprehensive sequence analysis of proteins (12–14). The use of an alternative enzyme, or multiple enzymes, has recently become more popular particularly when the relatively new peptide fragmentation method electron transfer dissociation (ETD) is used (15). Although CID is still the most commonly used fragmentation approach for analysis, it is not a complete solution. Certain types of peptides, such as those that contain labile side groups, for example, a phosphate group (16), or possess multiple basic residues (17, 18), are relatively poorly dealt with by using CID. Electron transfer dissociation (ETD) has been shown to be able to accommodate some of these areas.

Our group and others recently evaluated the specific characteristics of ETD for peptide sequencing by manipulating average peptide properties through the use of various different enzymes (12–14). It was found that switching the location of the basic residues from the C-terminal to the N-terminal side of the peptide has significant benefits for peptide sequencing by ETD (19). We found in particular that the use of a relatively unexplored enzyme, termed Lys-N, in combination with ETD, provided a clear advantage for spectral interpretation due to peptide MS/MS spectra consisting almost exclusively of *c*-type fragment ions. Such spectra, providing simple sequence ladders, may also prove to be a

valuable strategy for de novo sequencing and for the analysis of peptide post-translational modifications (19).

Lys-N is a zinc metalloendopeptidase that can be purified from the fruiting bodies of the edible mushroom *Grifola frondosa*. Lys-N is an enzyme with interesting properties as it exhibits high thermostability, high tolerance toward detergents, and a proteolytic activity in a broad pH range and can also be used for in-gel digestion. Most importantly, it was shown to have a very definite specificity for cleaving acyl-lysine bonds (20, 21).

Based on prior knowledge of peptide separation with low-pH SCX, we hypothesized that Lys-N proteolytic peptides containing certain functional groups could be largely separated and fractionated (22, 23). Here, we describe in detail the methods developed for these applications. The four categories we are able to distinguish and separate to near completion are (I) acetylated N-terminal peptides, (II) singly phosphorylated peptides containing a single basic (Lys) residue, (III) peptides containing a single basic (Lys) residue, and (IV) peptides containing more than one basic residue. Interestingly, ETD provides a facile method for site localization of phosphorylated peptides in category II and facilitates a database-independent method for sequencing of "normal" single lysine-containing peptides (category III).

2. Materials

2.1. In-Solution Digestion

1. Lysis buffer: 50 mM ammonium bicarbonate, 8 M urea, 5 mM sodium phosphate, 1 mM potassium fluoride, and 1 mM sodium orthovanadate (*see* **Note 1**). One tablet of protease inhibitors, without EDTA (Complete Mini, Roche) (*see* **Note 2**), per 10 ml of buffer. This buffer should be prepared fresh and protease inhibitors added shortly before use.

   Wash buffer: 50 mM ammonium bicarbonate, 25 mM sodium phosphate, 1 mM potassium fluoride, and 1 mM sodium orthovanadate. This buffer should be prepared fresh.
2. Reducing agent: prepare a fresh stock of 45 mM dithiothreitol (DTT) in 50 mM ammonium bicarbonate, pH 8.2.
3. Alkylation agent: prepare a fresh stock of 100 mM iodoacetamide (IAA) in 50 mM ammonium bicarbonate, pH 8.2.
4. Lys-N (Seikagaku Corporation or U-Protein Express): prepare a stock solution of 0.225 μg/μl in 50 mM ammonium bicarbonate. The enzyme can be stored at –80°C for several months (*see* **Note 3**).

2.2. Peptide Desalting and Strong Cation Exchange

1. Desalting Solvent A: 0.05% formic acid, pH 2.7.
2. Desalting Solvent B: 80% acetonitrile, 0.05% formic acid, pH 2.7.
3. SCX buffer A: 5 mM potassium dihydrogen phosphate, 30% acetonitrile, and 0.05% formic acid, pH 2.7.
4. SCX buffer B: 350 mM potassium chloride, 5 mM potassium dihydrogen phosphate, 30% acetonitrile, and 0.05% formic acid, pH 2.7.
5. Agilent 1100 HPLC system (with a MWD UV detector and a binary pump with two A and B lines) with two in-line coupled C18 Opti-Lynx (Optimized Technologies) guard columns (*see* **Note 4**) and a polysulfoethyl A SCX column (200 mm × 2.1 mm i.d., 5 μm, 200 Å) (PolyLC).

2.3. RP-HPLC and Mass Spectrometry

1. Trap column: 20 mm length × 100 μm internal diameter, Aqua C_{18} (Phenomenex).
2. Analytical column: 200 mm length × 50 μm internal diameter, ReproSil-Pur C18-AQ, 3 μm, 120 Å (Dr. Maisch).
3. HPLC solvent A: 0.1 M acetic acid.
4. HPLC solvent B: 0.1 M acetic acid, 80% acetonitrile.
5. LTQ XL mass spectrometer with an ETD source at the back (Thermo Fisher) operating in positive ionization, data-dependent mode, automatically switching between MS and MS/MS and between CID and ETD fragmentation. Full-scan MS spectra (from m/z 350 to 1,500) are acquired in the ion trap after accumulation to target value of 100,000. The two most intense ions are selected for collision-induced fragmentation in the linear ion trap at normalized collision energy of 35% after accumulating to a target value of 30,000. The two most intense ions are likewise selected for ETD fragmentation. ETD is performed with supplemental activation (24), fluoranthene as reagent anion, and an ion/ion reaction in the ion trap for 100 ms after accumulation to a target value of 100,000. An overview of the whole procedure is shown in **Fig. 11.1** (*see* **Note 5**).

2.4. Computational Analysis

1. Bioworks Browser software, version 3.1.1.
2. Xcalibur Software (Thermo Fisher). XDK component must be installed (*see* **Note 6**).
3. Mascot v2.2 search engine (Matrix Science).

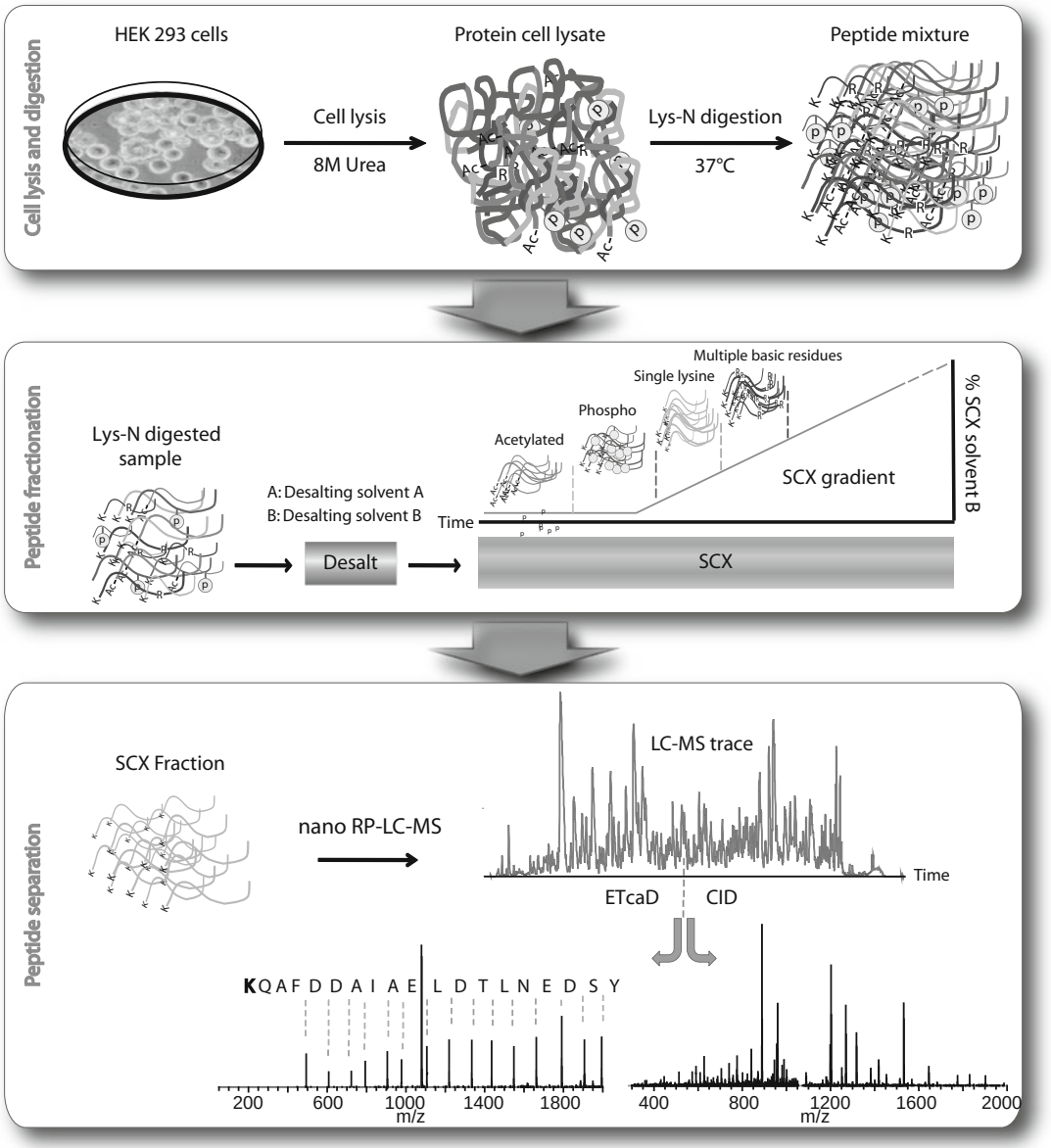


Fig. 11.1. Overview of the strategy. HEK293 cells are lysed using 8 M urea. Subsequently, the proteins are overnight digested with Lys-N at 37°C. The Lys-N-generated peptide mixture is desalted on a trap column prior to separation and fractionation by strong cation exchange (SCX). During the SCX gradient peptides with different functional groups will be separated and fractions containing these peptides are collected. Each fraction obtained is subjected to RP-LC mass spectrometry analysis using both CID and ETcaD to fragment the peptides.

3. Methods

3.1. In-Solution Digestion

1. Harvest cells by centrifugation at 14,000×*g* for 5 min at 4°C. Wash the cell pellet with 1 ml of wash buffer, centrifuge again, and remove the supernatant.
2. Lyse cell pellets by adding 3 ml of lysis buffer and keep on ice for at least 30 min (*see* **Note 7**).
3. Centrifuge at 20,000×*g* for 10 min at 4°C to spin down cell debris. Collect the soluble fraction (supernatant) into a new vial.
4. Determine the protein concentration by Bradford, Lowry, or BCA assays. Pipette the volume equivalent to 1 mg of protein into a new vial (the following steps are optimized for such an amount of starting material).
5. Reduce cysteine bonds by adding 62.5 µl of the 45 mM DTT stock solution per 200 µl of protein lysate (at 5 µg/µl) and incubate at 50°C for 15 min (*see* **Note 8**).
6. Block cysteine side chains by adding 62.5 µl of the 110 mM IAA stock solution per 200 µl of protein lysate (at 5 µg/µl) and incubate for 15 min in the dark at room temperature.
7. Add 11.8 µg of Lys-N to the protein solution (final ratio of 1:85, enzyme:substrate) and incubate the mixture overnight at 37°C (*see* **Note 9**).
8. Vacuum dry the digested cell lysate and re-dissolve in 0.05% formic acid (*see* **Note 10**). At this point samples are ready for desalting prior to SCX fractionation.

3.2. Peptide Desalting and Strong Cation Exchange

1. Equilibrate the SCX system by washing with SCX buffer B followed by washing with SCX buffer A (five times column volume or until a stable UV signal is reached).
2. Wash the trap columns with desalting solvent B followed by desalting solvent A for 10 min each at 200 µl/min. Switch the flow rate back to 100 µl/min for loading (*see* **Note 11**).
3. Load samples onto the C18 guard columns for 10 min at 100 µl/min using desalting solvent A.
4. Elute peptides onto the SCX column for 20 min at 100 µl/min using desalting solvent B.
5. Elute peptides from the SCX column using the following SCX gradient: 0% SCX buffer B for 10 min, 0–85% SCX buffer B for 35 min, 85–100% SCX buffer B for 6 min, and 100% SCX buffer B for 4 min. The gradient is performed at a flow rate of 200 µl/min. Collect approximately 49 SCX fractions (1 min/fraction) that contain the different subgroups of peptides.

6. Following SCX separation, equilibrate the SCX system for 10 min in SCX buffer A prior to the subsequent analysis.
7. Vacuum dry the fractions (lyophilizing is preferred) and resuspend the samples in 50 μl 10% of formic acid. At this point samples are ready for RP-HPLC and mass spectrometry.

3.3. RP-HPLC and Mass Spectrometry

1. Inject samples onto the RP-based nano-LC system.
2. Trap peptides are trapped on the C18 trap column at a flow rate of 5 μl/min for 10 min in 100% HPLC solvent A (*see* **Note 12**).
3. Apply a 75 min linear gradient from 0 to 40% HPLC solvent B during which the peptides elute from the trap column onto the analytical column for separation at a flow rate of 100 nl/min (*see* **Note 13**).
4. Subsequent MS analysis of peptides present in the low-pH SCX fractions containing Lys-N-generated peptides results in fractionation profiles in which peptides from different functional categories are well separated. The four categories that are separated to near completion are (I) acetylated N-terminal peptides, (II) singly phosphorylated peptides containing a single basic (Lys) residue, (III) peptides containing a single basic (Lys) residue, and (IV) peptides containing more than one basic residue (**Fig. 11.2**).

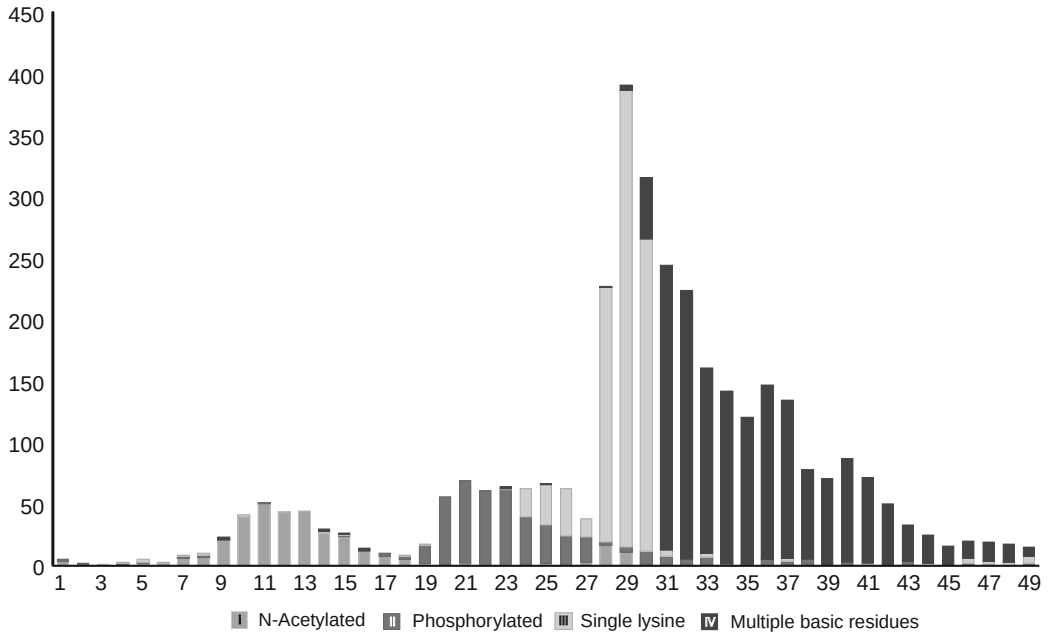


Fig. 11.2. SCX fractionates different classes of Lys-N-generated peptides. The graph plotting the total number of unique peptides versus SCX fraction (analysis performed by ETcaD). The color annotations (with group numbers) of the different peptide subgroups are described at the *bottom*. Distinct separation of peptide subgroups is observed (groups I–IV).

3.4. Computational Analysis

1. Convert flow-through and elution "raw" MS files from the SCX fractions into peak lists using Bioworks Browser software, version 3.1.1 (*see* **Note 6**).
2. For database searching, submit all spectra contained within the peak list to an in-house licensed Mascot 2.2.0 search

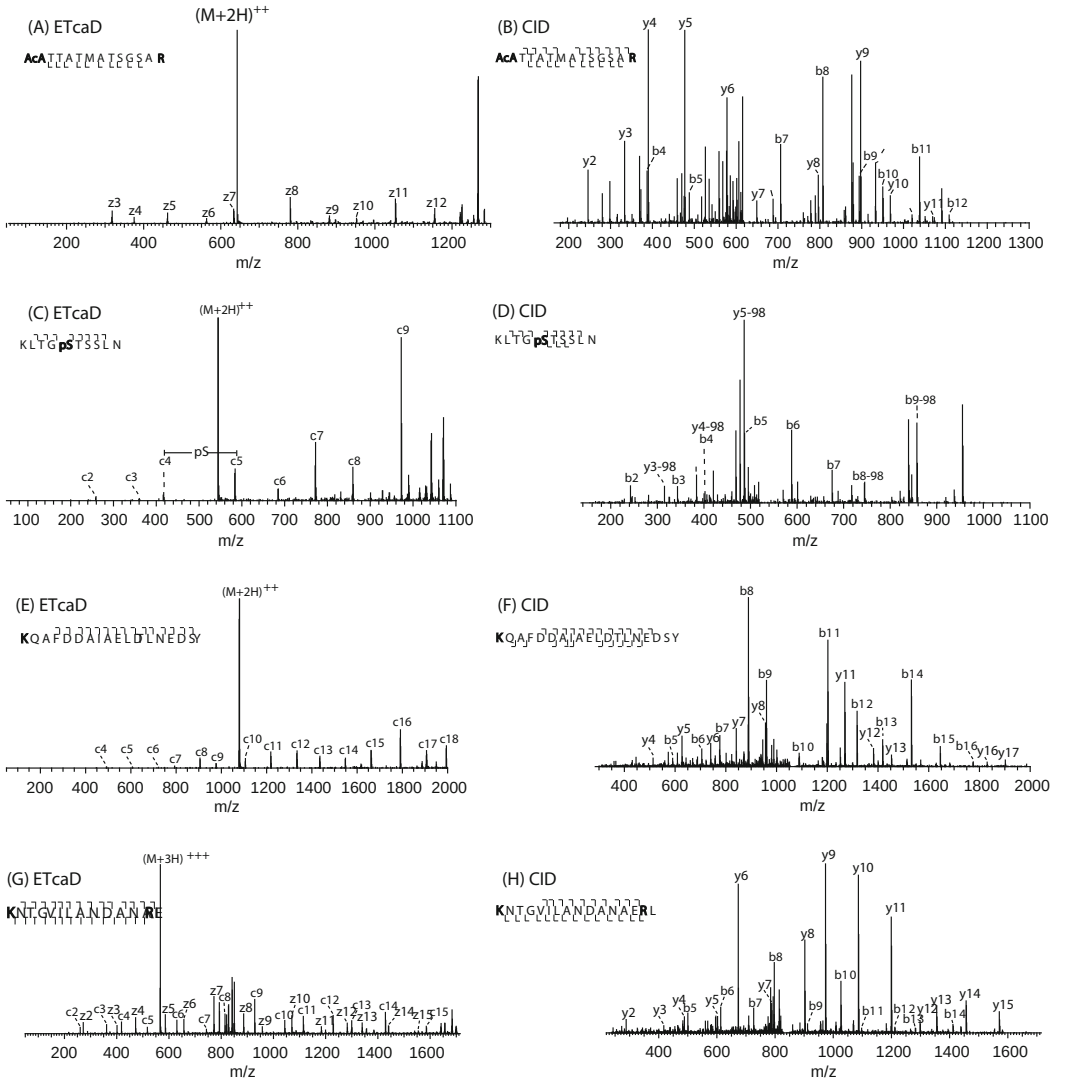


Fig. 11.3. Examples of Lys-N peptide fragmentation spectra using both CID and ETcaD. ETcaD and CID spectra of N-terminally acetylated doubly charged ions, phosphorylated doubly charged ions, and doubly and triple charged ions originated from a Lys-N digest of a HEK293 cell lysate. (**a**) ETcaD spectrum of the N-terminal acetylated peptide **AcA**TTATMATSGSAR. (**b**) CID spectrum of the N-terminal acetylated peptide **AcA**TTATMATSGSAR. (**c**) ETcaD spectrum of serine phosphorylated KLTG**pS**TSSLN. From the sequence of *c*-ions it is easy to determine the phosphorylation site. (**d**) CID MS/MS spectrum of serine phosphorylated KLTG**pS**TSSLN. (**e**) ETcaD spectrum of doubly charged peptide **K**QAFDDAIAELDTLNEDSY. (**f**) CID spectrum of doubly charged peptide **K**QAFDDAIAELDTLNEDSY. (**g**) ETcaD spectrum of triple charged peptide **K**NTGVILANDANAE**R**L. (**h**) CID spectrum of doubly charged **K**NTGVILANDANAE**R**L.

engine and search against the appropriate protein sequence database. Set carbamidomethyl cysteine as a fixed modification; oxidation of methionine, protein N-acetylation and phosphorylation of serine, threonine, and tyrosine are set as variable modifications. Set the peptide mass tolerance to 0.5 Da and fragment mass tolerance to 0.6 Da (*see* **Note 14**). **Figure 11.3** shows examples of MS/MS spectra of peptides obtained in the different subgroups.

4. Notes

1. Alternatively, sodium phosphate, potassium fluoride, and sodium orthovanadate (common phosphatase inhibitors) can be replaced by one tablet of phosphatase inhibitors (PhosStop, Roche) per 10 ml of buffer.
2. EDTA is a strong chelating agent binding to metal ions. As Lys-N is a metalloendopeptidase the enzymatic activity might be compromised in the presence of EDTA and should therefore be avoided.
3. The metalloendopeptidase Lys-N can also be purchased at U-Protein Express BV, Utrecht.
4. Desalting of peptides can be performed off-line prior to SCX fractionation and does not have to be performed online as described here.
5. Analysis of Lys-N-digested peptides can be performed on other instruments with an ETD source. The number of peptides identified in **Fig. 11.2** will be higher if an LTQ-Orbitrap is being used (23). If an LTQ-Orbitrap is used, full-scan MS spectra (from m/z 350 to 1,500) are acquired in the Orbitrap at a resolution of 60,000 at m/z 400 after accumulation to target value of 500,000. The five most intense ions at a threshold of 500 are selected for collision-induced fragmentation in the linear ion trap at normalized collision energy of 35% after accumulating to a target value of 30,000.
6. Proteome discoverer must be installed if samples have been analyzed on an LTQ-Orbitrap ETD instrument for processing "raw" MS files (Thermo Fisher).
7. For human embryonic kidney (HEK) 293T cells this is sufficient to ensure cell disruption; if other cell types are used a more stringent method might be needed to ensure cell lysis.
8. Do not use temperatures higher than 60°C as urea can decompose into isocyanic acid leading to artificial

carbamoylation of free amino groups (protein N-termini and side chains of lysine residues) (25).

9. A higher protease/protein ratio can be used (1/100) without influencing the digestion efficiency. Here 1/85 was used based on previous work (26).
10. The sample should be re-dissolved according to the capacity of the sample loop on the SCX system. Multiply injection step can also be performed.
11. It is important that desalting solvents A and B are not going through the SCX column when washing the trap columns.
12. The later fractions may require a more extensive desalting online (extra desalting step for 10–20 min before the analysis method). If no trap column is being used all fractions must be desalted off-line prior to LC-MS analysis.
13. The length of the gradient depends on the sample complexity and sample amount. For this work a 75 min gradient was sufficient. Where maximal proteome coverage is required we generally employ gradients of 3 h per SCX fraction.
14. Precursor tolerance and fragment ion tolerance are instrument dependent.

Acknowledgments

The work was supported by the Netherlands Proteomics Center.

References

1. Gevaert K., Van Damme P., Ghesquiere B., Impens F., Martens L., Helsens K., et al. (2007) A la carte proteomics with an emphasis on gel-free techniques. *Proteomics* **7**, 2698–718.
2. Link A. J., Eng J., Schieltz D. M., Carmack E., Mize G. J., Morris D. R., et al. (1999) Direct analysis of protein complexes using mass spectrometry. *Nat Biotechnol* **17**, 676–82.
3. Washburn M. P., Wolters D., Yates J. R., 3rd (2001) Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat Biotechnol* **19**, 242–7.
4. Boersema P. J., Divecha N., Heck A. J., Mohammed S. (2007) Evaluation and optimization of ZIC-HILIC-RP as an alternative MudPIT strategy. *J Proteome Res* **6**, 937–46.
5. Motoyama A., Xu T., Ruse C. I., Wohlschlegel J. A., Yates J. R., 3rd (2007) Anion and cation mixed-bed ion exchange for enhanced multidimensional separations of peptides and phosphopeptides. *Anal Chem* **79**, 3623–34.
6. Wu C. C., MacCoss M. J., Howell K. E., Yates J. R., 3rd (2003) A method for the comprehensive proteomic analysis of membrane proteins. *Nat Biotechnol* **21**, 532–8.
7. Villen J., Gygi S. P. (2008) The SCX/IMAC enrichment approach for global phosphorylation analysis by mass spectrometry. *Nat Protoc* **3**, 1630–8.
8. Gruhler A., Olsen J. V., Mohammed S., Mortensen P., Faergeman N. J., Mann M., et al. (2005) Quantitative phosphoproteomics applied to the yeast pheromone

signaling pathway. *Mol Cell Proteomics* **4**, 310–27.
9. Gnad F., de Godoy L. M., Cox J., Neuhauser N., Ren S., Olsen J. V., et al. (2009) High-accuracy identification and bioinformatic analysis of in vivo protein phosphorylation sites in yeast. *Proteomics* **9**, 4642–52.
10. Dormeyer W., Mohammed S., Breukelen B., Krijgsveld J., Heck A. J. (2007) Targeted analysis of protein termini. *J Proteome Res* **6**, 4634–45.
11. Steen H., Mann M. (2004) The ABC's (and XYZ's) of peptide sequencing. *Nat Rev Mol Cell Biol* **5**, 699–711.
12. Molina H., Horn D. M., Tang N., Mathivanan S., Pandey A. (2007) Global proteomic profiling of phosphopeptides using electron transfer dissociation tandem mass spectrometry. *Proc Natl Acad Sci USA* **104**, 2199–204.
13. Good D. M., Wirtala M., McAlister G. C., Coon J. J. (2007) Performance characteristics of electron transfer dissociation mass spectrometry. *Mol Cell Proteomics* **6**, 1942–51.
14. Mohammed S., Lorenzen K., Kerkhoven R., van Breukelen B., Vannini A., Cramer P., et al. (2008) Multiplexed proteomics mapping of yeast RNA polymerase II and III allows near-complete sequence coverage and reveals several novel phosphorylation sites. *Anal Chem* **80**, 3584–92.
15. Coon J. J. (2009) Collisions or electrons? Protein sequence analysis in the 21st century. *Anal Chem* **81**, 3208–15.
16. Leitner A., Foettinger A., Lindner W. (2007) Improving fragmentation of poorly fragmenting peptides and phosphopeptides during collision-induced dissociation by malondialdehyde modification of arginine residues. *J Mass Spectrom* **42**, 950–9.
17. Bridgewater J. D., Srikanth R., Lim J., Vachet R. W. (2007) The effect of histidine oxidation on the dissociation patterns of peptide ions. *J Am Soc Mass Spectrom* **18**, 553–62.
18. Coon J. J., Ueberheide B., Syka J. E., Dryhurst D. D., Ausio J., Shabanowitz J., et al. (2005) Protein identification using sequential ion/ion reactions and tandem mass spectrometry. *Proc Natl Acad Sci USA* **102**, 9463–8.
19. Taouatas N., Drugan M. M., Heck A. J., Mohammed S. (2008) Straightforward ladder sequencing of peptides using a Lys-N metalloendopeptidase. *Nat Methods* **5**, 405–7.
20. Nonaka T., Ishikawa H., Tsumuraya Y., Hashimoto Y., Dohmae N. (1995) Characterization of a thermostable lysine-specific metalloendopeptidase from the fruiting bodies of a basidiomycete, Grifola frondosa. *J Biochem* **118**, 1014–20.
21. Nonaka T., Dohmae N., Hashimoto Y., Takio K. (1997) Amino acid sequences of metalloendopeptidases specific for acyl-lysine bonds from Grifola frondosa and Pleurotus ostreatus fruiting bodies. *J Biol Chem* **272**, 30032–9.
22. Taouatas N., Altelaar A. F., Drugan M. M., Helbig A. O., Mohammed S., Heck A. J. (2009) Strong cation exchange-based fractionation of Lys-N-generated peptides facilitates the targeted analysis of post-translational modifications. *Mol Cell Proteomics* **8**, 190–200.
23. Gauci S., Helbig A. O., Slijper M., Krijgsveld J., Heck A. J., Mohammed S. (2009) Lys-N and trypsin cover complementary parts of the phosphoproteome in a refined SCX-based approach. *Anal Chem* **81**, 4493–501.
24. Swaney D. L., McAlister G. C., Wirtala M., Schwartz J. C., Syka J. E., Coon J. J. (2007) Supplemental activation method for high-efficiency electron-transfer dissociation of doubly protonated peptide precursors. *Anal Chem* **79**, 477–85.
25. Righetti P. G. (2006) Real and imaginary artefacts in proteome analysis via two-dimensional maps. *J Chromatogr B Analyt Technol Biomed Life Sci* **841**, 14–22.
26. Rao K. C., Carruth R. T., Miyagi M. (2005) Proteolytic 18O labeling by peptidyl-Lys metalloendopeptidase for comparative proteomics. *J Proteome Res* **4**, 507–14.

Chapter 12

Quantitation of Newly Synthesized Proteins by Pulse Labeling with Azidohomoalanine

Gertjan Kramer, Piotr T. Kasper, Luitzen de Jong, and Chris G. de Koster

Abstract

Measuring protein synthesis and degradation rates on a proteomic scale is an important step toward modeling the kinetics in complicated cellular response networks. A gel-free method, able to quantify changes in the formation of new proteins on a 15 min timescale, compatible with mass spectrometry is described. The methionine analogue, azidohomoalanine (azhal), is used to label newly formed proteins during a short pulse-labeling period following an environmental switch in *Escherichia coli*. Following digestion a selective reaction against azhal-containing peptides is applied to enrich these peptides by diagonal chromatography. This technique enables quantitation of hundreds of newly synthesized proteins and provides insight into immediate changes in newly synthesized proteins on a proteomic scale after an environmental perturbation.

Key words: Pulse-chase labeling, translational regulation, azhal, quantitative proteomics, *E. coli*, BONCAT, COFRADIC.

1. Introduction

Changes in protein levels in cells during adaptation from one environmental condition to another may be regulated both by transcription and by translation. However, little is known about the contribution of translational regulation. This requires information about genome-wide changes in translation rates and mRNA levels upon environmental perturbations. Determination of protein synthesis and degradation rates on a proteomic scale as such is an important step toward understanding the kinetics in complicated cellular response networks on the level of both transcription and

K. Gevaert, J. Vandekerckhove (eds.), *Gel-Free Proteomics*, Methods in Molecular Biology 753,
DOI 10.1007/978-1-61779-148-2_12,

translation. Determination of protein synthesis and degradation rates has mainly involved pulse labeling with radiolabels combined with separation of proteins by two-dimensional gel electrophoresis (1, 2). This approach has drawbacks, such as difficulties to detect very acidic, basic, or hydrophobic proteins (e.g., membrane proteins), and the possible occurrence of more than one protein in a gel spot. The use of amino acids labeled with stable isotopes rather than radioisotopes is a method that is applicable to a mass spectrometry-based proteome-wide approach (3–7). However, it requires extensive labeling times as the unlabeled bulk of the protein content of the cell will also be detected. As a consequence this approach only provides limited temporal resolution.

An alternative to the labels described above is the use of the methionine analogue azidohomoalanine (azhal) to pulse label newly synthesized proteins. Pulse labeling of proteins with azhal has been applied in *Escherichia coli* (8) and mammalian cells (9). To separate the small amount of labeled material from the bulk of the unlabeled material, the azide group of azhal can be used to enrich labeled material by different selective chemical reactions employed in various enrichment schemes (8–10). This provides the opportunity to use short labeling times and increases temporal resolution.

Here a protocol to label *E. coli* with azhal immediately following an environmental perturbation (temperature switch) is described. The enrichment approach uses a chemically induced retention time shift between two reversed-phase separations with identical chromatographic conditions and enables the relative quantitation of hundreds of proteins newly synthesized during a 15 min pulse-labeling period with azhal (8).

2. Materials

2.1. Cell Culture and Lysis

1. LB medium: 10 g Bacto Tryptone, 5 g yeast extract, and 10 g NaCl in 1 l of ddH_2O, autoclaved for sterility.
2. LB plates: add 1% (w/v) of Bacto Agar to LB medium, autoclave, and pour plates in sterile plastic culture dishes.
3. Water bath shaker set at 37°C and a water bath shaker set at 44°C.
4. Minimal medium, M9 complete: 6.8 μM $CaCl_2$, 1.0 mM $MgSO_4$, 59.3 μM thiamine HCl, 57 nM Na_2SeO_3, 5 μM $CuCl_2$, 10 μM $CoCl_2$, 5.2 μM H_3BO_3, 99.9 μM $FeCl_3$, 50.5 μM $MnCl_2$, 25.3 μM ZnO, 0.08 μM Na_4MoO_4, 111 mM glucose and 60 mg/l for each of the 19 natural amino acids and 40 mg/l for tyrosine in ddH_2O.

5. Minimal medium, M9 minimal: as described above but without methionine.
6. Azidohomoalanine (azhal) (*see* **Note 1**).
7. Sonicator, Branson-250, with a 3 mm tapered microtip (Branson).
8. Benchtop centrifuge with swing out rotor.
9. Lysis solution: 8 M urea, 50 mM HEPES, pH 8.0, in Millipore water (*see* **Note 2**).
10. Dialysis tubing 5 kD cutoff.

2.2. Digestion and iTRAQ Labeling

1. Trypsin Gold, mass spectrometry grade (Promega): 1 μg/μl in 50 mM acetate (*see* **Note 3**).
2. iTRAQ 4-plex kit (Applied Biosystems).
3. ICAT strong cation-exchange cartridge (SCX cartridge) and cartridge holder (Applied Biosystems).
4. Hamilton syringe (500 μl) with a blunt-end needle.
5. SCX loading buffer: 20% (v/v) acetonitrile and 0.1% (v/v) formic acid in water.
6. SCX elution buffer: 20% (v/v) acetonitrile in 2 M ammonium formate buffer, pH 6.8.

2.3. Diagonal Chromatography and Mass Spectrometry

1. 20 mM tris(2-carboxyethyl)phosphine in water (prepare freshly).
2. 1 M HEPES, pH 8.0.
3. 50 mM NaN_3 and 150 mM iodoacetamide (IAA) in water (prepare freshly and keep in dark).
4. Human [Glu1]fibrinopeptide B.
5. SMART system (Pharmacia) equipped with a 200 μl sample loop and a fraction collector (or a comparable HPLC system with fraction collector), using a Jupiter Proteo C12 column (2 mm internal diameter, 150 mm length, Phenomenex). Solvent A is 0.1% TFA in water and solvent B is 0.1% TFA in acetonitrile.
6. 1100 series LC system (Agilent) fitted with a nanoscale reversed-phase HPLC setup (solvent A′ 0.1% formic acid in water and solvent B′ 0.08% formic acid in acetonitrile) involving Deans' switching as described by Meiring et al. (11), with a 2 cm 100 μm inner diameter C18 trapping column (Nanoseparations) and a 63 cm 50 μm inner diameter C18 reversed-phase analytical column (Nanoseparations) coupled to a QSTAR XL mass spectrometer via a liquid junction with nebulizer using an uncoated fused silica emitter (New Objective) operating around 4.7 kV (Applied Biosystems) or a comparable LC-MS setup.

2.4. Data Analysis and Statistics

1. MASCOT search engine version 2.1 or higher with in-house license (Matrix Science).
2. Microsoft Excel.
3. Quant: iTRAQ quantitation software (12) (http://www.protein-ms.de).
4. ASAP utilities (http://www.asap-utilities.com/).

3. Methods

To ensure complete labeling of newly synthesized proteins with azhal, a methionine-auxotrophic *E. coli* strain (MTD123) is used and cells are washed prior to labeling, so no residual methionine remains. This is necessary because the k_{cat}/K_M of methionyl-tRNA synthetase for azhal is 390 times lower than for methionine (13). In principle any auxotrophic organism that grows at a normal rate for at least several minutes in the presence of azhal instead of methionine and incorporates it into proteins can be used. The example here uses *E. coli* cells, which are labeled for 15 min following a change in growth temperature.

Apart from efficient labeling of newly synthesized proteins with azhal, the other important principle for sensitive detection of newly synthesized proteins is the selective enrichment of labeled material from the bulk of unlabeled material. The enrichment is based on the principle of COmbined FRActional DIagonal Chromatography (COFRADIC), first used to selectively enrich methionine-containing peptides by a retention time shift induced by H_2O_2 between two reversed-phase chromatographic runs (14). To isolate azhal-containing peptides, a retention time shift is induced by a selective reaction with tris(2-carboxyethyl)phosphine (TCEP) yielding four different reaction products that have a different retention time than the parent peptide (**Fig. 12.1**) (8, 15). Previously this reaction was used to enrich cross-linked peptides with an artificially introduced linker containing an azide moiety (16). By collection of fractions containing reaction products that shift their retention time upon reaction with TCEP, the small amount of azhal-containing peptides is highly enriched from the unlabeled background. This enrichment facilitates the use of short pulse-labeling times, while subsequent iTRAQ labeling provides relative quantitation of azhal peptides between samples. Using this approach around 300 newly synthesized proteins can typically be quantified after a 15 min pulse-labeling period.

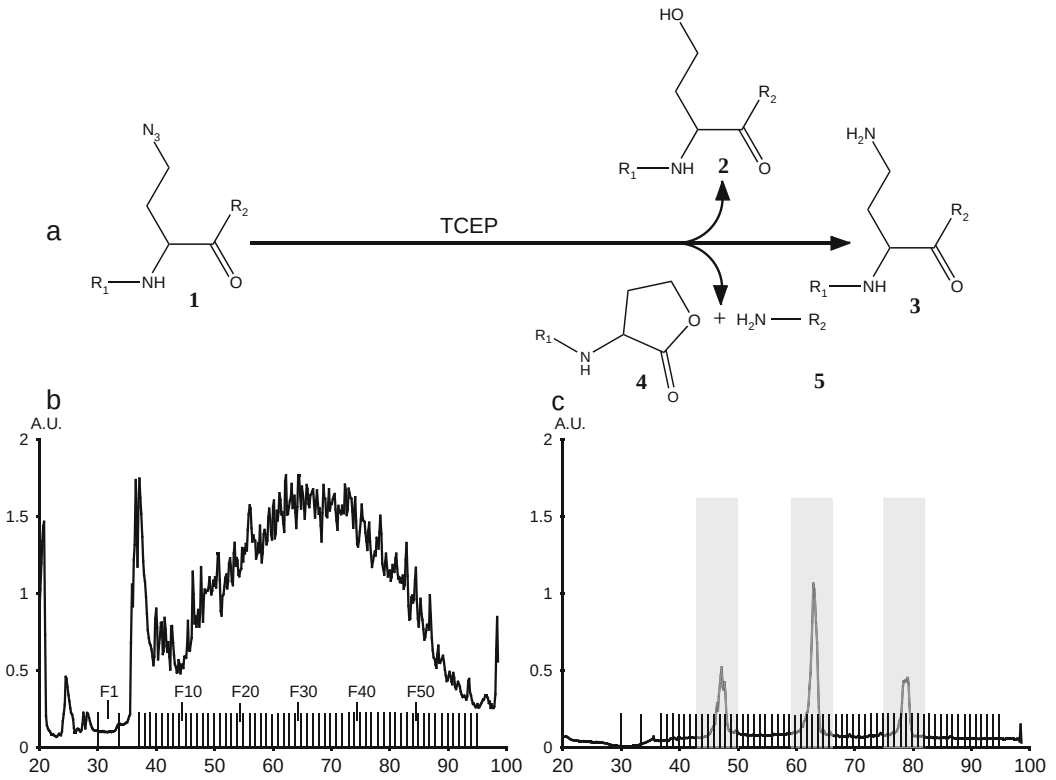


Fig. 12.1. The COFRADIC approach. (**a**) Reaction products resulting from an azhal-containing peptide [**1**] reacting with TCEP are a peptide containing either a homoserine [**2**] or a diaminobutyrate residue [**3**]. Alternatively specific cleavage at the azhal residue results in an N-terminal cleavage product with a homoserine lactone residue at its C-terminus [**4**] and a C-terminal product with a normal N-terminus [**5**]. R_1: N-terminal side of peptide, R_2: C-terminal side of peptide. (**b**) Primary RP-HPLC run of a tryptic digest of a labeled *E. coli* proteome: a total of 58 1-min fractions are collected from 37 to 95 min. In order to diminish the number of secondary runs to 16, primary fractions separated by an interval of 16 min are pooled (**Table 12.1**). Pooled fractions are treated with TCEP and run again under identical conditions (**c**). Unlabeled peptides are not modified and will run at the same retention time, while labeled peptides react with TCEP and shift their retention time. The non-shifted fractions plus three adjacent fractions on the front and back are discarded (*gray* area). The remaining fractions containing shifted peptides are pooled, lyophilized, and analyzed by tandem MS.

3.1. Cell Culture and Lysis

1. Streak out *E. coli* strain MTD123 (17) on an LB plate and grow overnight at 37°C (*see* **Note 4**). Pick colonies from the plate and start four overnight cultures in 2 ml M9 complete medium at 37°C in a shaking incubator.
2. Measure OD_{600} of the overnight cultures and use two of them to inoculate two cultures (designated as cultures A and B) in 100 ml M9 complete medium at OD_{600} 0.01 and grow at 37°C in a shaking incubator.
3. After inoculation prepare four Erlenmeyers with 50 ml M9 minimal medium and add azhal to a final concentration of 400 mg/l. Mark two as cultures A^1 and A^2 and two as cultures B^1 and B^2 and keep at room temperature. In addition

prepare two 2 ml cultures in sterile tubes marked A^c and B^c with M9 minimal medium. These are the negative controls for evaluating the washing efficiency (*see* **Note 5**).

4. Monitor OD_{600}, cells should be harvested when they reach an OD_{600} of 1.0 (*see* **Note 6**) by centrifugation for 5 min at 3,631×*g* and room temperature in marked 50 ml polypropylene tubes (A and B).
5. Decant the supernatant and rinse the pellets twice with M9 minimal medium by adding 50 ml of medium and decanting it again to rinse away residual methionine. Subsequently, re-suspend the pellets in 2 ml of M9 minimal medium and fill the tubes with M9 minimal medium up to 50 ml and spin down for 5 min at 3,631×*g* and room temperature (*see* **Note 7**).
6. Rinse once with M9 minimal medium, re-suspend in 1 ml of M9 minimal medium, and measure OD_{600}.
7. Subsequently inoculate the four cultures A^1, A^2, B^1, and B^2 and the washing control cultures A^c and B^c from their respective pre-cultures at an OD_{600} of about 0.2, take samples for OD_{600} measurement, and immediately start labeling by putting the labeling cultures A^1, B^1, A^c, and B^c at 37°C and the labeling cultures A^2 and B^2 at 44°C for a period of 15 min. During labeling measure the initial OD_{600} and put four 50 ml tubes on ice.
8. After labeling immediately pour the cultures in the chilled 50 ml tubes, take a sample to measure OD_{600}, and spin down the cultures at 3,631×*g* at 4°C for 10 min in the benchtop centrifuge (*see* **Note 8**). Decant the supernatant and re-suspend cells in 0.5 ml of cold lysis solution, and either immediately continue with lysis or place the re-suspended cells at –80°C (*see* **Note 9**).
9. Use OD_{600} changes measured during labeling in azhal-containing cultures to estimate labeling percentage and control cultures to assess washing efficiency (*see* **Note 10**).
10. Lyse cells by sonication for 5 min (output 5, duty cycle 50%) in a 15 ml tube on ice-cold water, transfer to a 1.5 ml Eppendorf tube, and remove cellular debris by centrifugation at 12,000×*g* and 4°C for 45 min.
11. Take supernatant and dialyze overnight against 10 mM HEPES, pH 8.0, to remove urea.

3.2. Digestion and iTRAQ Labeling

1. Take dialyzed lysates and determine protein content with a BCA assay. Take 125 μg of protein for each sample and add 2.5 μg of trypsin, 10% (v/v) acetonitrile, and 25 mM

HEPES, pH 8.0 (end concentrations), in 250 µl (end volume) and digest overnight at 37°C (*see* **Note 11**).

2. After digestion, lyophilize the samples and re-dissolve in 20 µl dissolution buffer provided with the iTRAQ kit (500 mM triethylammonium bicarbonate, pH 8.5), subsequently add 2 vials of iTRAQ reagent (each vial is dissolved in 70 µl ethanol) to each sample (A^1 114, A^2 115, B^1 116, B^2 117). Incubate on a rotary shaker for 2 h at room temperature (*see* **Note 12**).
3. Quench the reaction by adding 320 µl of 0.1% (v/v) formic acid, incubate for a further 5 min at room temperature, and take a 4 µl aliquot (~1 µg) to assess iTRAQ labeling efficiency (*see* **Note 13**).
4. Combine equal amounts of the labeled samples and dilute the 2 ml total volume of combined samples by adding 4 ml of SCX loading buffer (*see* **Note 14)**.
5. Prepare the SCX cartridge by washing with 1 ml of loading buffer, then condition by washing with 2 ml of elution buffer and re-equilibrate with 2 ml of loading buffer. Subsequently load the sample, wash with 1 ml of loading buffer, and elute the sample with 1 ml of elution buffer, then lyophilize the sample overnight and store at –20°C until further use.

3.3. Diagonal Chromatography and Mass Spectrometry

1. Re-dissolve the lyophilized sample (~500 µg protein) in 150 µl of 50 mM HEPES, pH 8.0, add 15 µl of TCEP stock solution (2 mM end concentration), and incubate at room temperature for 5 min to reduce disulfide bridges. Then add 15 µl NaN_3 and IAA and incubate in the dark for 15 min (*see* **Note 15**).
2. Load the sample immediately on the RP-HPLC column for the primary run. First wash with solvent A at a flow of 50 µl/min for 20 min and then apply a linear gradient to 50% solvent B in 75 min, while collecting fractions (*see* **Fig. 12.1**) and monitoring absorbance at 214 nm. Pool peptide fractions collected from 40 until 89 min into 16 pools (labels A through P, *see* **Fig. 12.1** and **Table 12.1**) and lyophilize overnight.
3. Re-dissolve lyophilized pools in 100 µl of 10 mM TCEP, 50 mM HEPES pH 8.0 and incubate overnight at 40°C. Subsequently store at –20°C until starting the secondary RP-HPLC runs (*see* **Note 16)**.
4. Thaw TCEP-treated peptide pools and add 100 µl of water before reinjection for the secondary run. Use the same LC program as for the primary run (*see* **Note 17**). After the secondary run, discard the non-shifted fractions including three fractions before and after the non-shifted fractions (judged

Table 12.1
Pool scheme primary run COFRADIC

Pool	Fractions	Pool	Fractions
A	7, 23, 39	I	15, 31, 47
B	8, 24, 40	J	16, 32, 48
C	9, 25, 41	K	17, 33, 49
D	10, 26, 42	L	18, 34, 50
E	11, 27, 43	M	19, 35, 51
F	12, 28, 44	N	20, 36, 52
G	13, 29, 45	O	21, 37, 53
H	14, 30, 46	P	22, 38, 54

Fractions as shown in **Fig. 12.1**.

on UV absorbance). Pool and lyophilize the remaining fractions containing shifted material (*see* **Fig. 12.1**).

5. In between secondary runs, run a short blank gradient to 99% solvent B in 30 min to eliminate possible carry-over.
6. Re-dissolve lyophilized pooled fractions containing the shifted peptides in 10 μl of 0.1% TFA with 150 pmol of human [Glu1]fibrinopeptide B for internal calibration and inject onto the LC-MS system.
7. Operate the LC-tandem MS setup as follows. Load the sample onto the trapping column by washing for 10 min at a flow rate of 5 μl/min with 98% of solvent A′ and 2% of solvent B′. Switch the trap column in line with the analytical column to elute the trapped peptides onto the analytical column for separation using a linear gradient of 8–30% solvent B′ for 95 min at a flow rate of 125 nl/min. Acquire survey MS scans from m/z 300 to 1,200 and operate the mass spectrometer such that the three most intense ions are selected for fragmentation.

3.4. Data Analysis and Statistics

1. Following LC-MS/MS analysis, generate peak lists (MASCOT generic file format) in Analyst QS 1.1 for subsequent product ion searches in the MASCOT search engine. Concatenate the peak lists of all 16 LC-MS/MS runs before submitting them to MASCOT (*see* **Note 18**).
2. For the MASCOT search use the following parameters: cleavage after lysine or arginine unless followed by proline plus cleavage after methionine, allowing up to two missed cleavages. Allow the following fixed modifications: carbamidomethyl cysteine and iTRAQ (Lys) modifications. And, allow as variable modifications iTRAQ (N-terminal) modification and modifications induced by reaction of

TCEP which include methionine C-terminally converted to a homoserine lactone after cleaving [**4**] and a methionine residue replaced by diaminobutyrate [**3**] ($C_4H_8N_2O$) or by homoserine [**2**] ($C_4H_7NO_2$). Set the peptide mass tolerance to 0.1 Da and MS/MS tolerance to 0.05 Da. Searches with the above settings are performed in a local database of the *E. coli* K12 proteome. Following the search adjust the significance threshold to 0.01 and turn on "MUDPIT" scoring and the "require bold red" option while setting the ion-score cutoff to 35 in the results window and reformat the results to see the significant hits with these threshold settings (*see* **Note 19**).

3. Select the peptides that are identified to be reaction products stemming from azhal-containing peptides manually in MASCOT and resubmit these by using the search selected ions feature of MASCOT, against the same database with the same settings. This gives a MASCOT results file with proteins only identified by azhal-containing peptides (*see* **Note 20**).
4. Subsequently import the MASCOT results file (.dat) into Quant (12) to quantify the iTRAQ reporter ions using the following settings: all four iTRAQ reporters on, report peak areas on, reporter tolerance set at 0.1 Da, intensity range turned off, peak dimensions at 0.025 Da, absolute intensity error set at 0, experimental error set at 0%, use of unique peptides on, *p* value cutoff set at 0.01, and macro language set at English with the macro parameter separator set to comma. Put in the correction factors for the iTRAQ kit used.
5. Quant puts out a comma-separated value file which can be imported into Excel. This file contains ratios of all combinations of the iTRAQ reporter ions on both the peptide and the protein level. Select those relevant for the experiment done (i.e., 115/114 and 117/116 for relative ratios of 44°/37° of cultures A and B, respectively, in this example). By the use of conditional select/delete and automatic fill options of ASAP utilities in combination with the pivot table function of Excel, it is also possible to calculate average protein ratios and their standard deviation in Excel from the Quant output data (*see* **Note 21**).

4. Notes

1. Azhal can be synthesized from L-BOC-DAB (Chem-Impex) using diazo transfer as described before (10) or purchased at AnaSpec.

2. Prepare freshly, do not heat urea to dissolve, keep cold.
3. All solutions need to be prepared using LC-MS-grade solvents unless indicated otherwise; chemicals used must be of analytical grade.
4. MTD123 is a stable methionine-auxotrophic strain; however, other methionine-auxotrophic strains such as CAG18491 and M15MA can also be used and have similar growth characteristics on azhal.
5. Keeping flasks at room temperature assures that cells are under different conditions only after putting them in the water baths.
6. Typically it takes cells about 5 h to reach an OD_{600} of 1.0 after inoculation from overnight culture.
7. Cells are washed in complete medium without methionine and harvested and washed at room temperature to prevent an osmotic or temperature shock, respectively.
8. Sudden cooling quickly stops labeling and faster cooling can be achieved by addition of crushed ice to tubes. Alternatively, translation can be stopped instantaneously by addition of chloramphenicol to the cells.
9. Freezing and thawing cells prior to lysis aid in a more efficient cell lysis. About 50 ml of culture inoculated at OD_{600} 0.2 typically yields about 0.5–1.0 mg of protein after 15 min of labeling, while only 125 μg is used per analysis.
10. Typically cells should have a net increase of about 9% of cellular mass based on increase of OD_{600} after subtracting the increased OD_{600} from the washing control cultures. Some growth in washing control cultures is to be expected, repeating wash steps would not only abolish this but also further stress cells. It is important that the increase in OD_{600} between different washing control cultures is similar, as not to have very different residual methionine concentrations present that would interfere with labeling to different extents between cultures.
11. End volume of the digest is 250 μl, with 25 mM of HEPES, pH 8.0, this results in about 4.2 μmol of sodium ions present in each reaction mixture.
12. Double the amount of iTRAQ reagent and incubation time is used compared to the manufacturer's protocol, to ensure full labeling of peptides.
13. Assess iTRAQ labeling percentage by an LC-MS/MS analysis of the samples separately: typically >95–99% of peptides should have the iTRAQ modification on the N-terminus or

lysine. Ensure that iTRAQ labeling efficiency between samples is comparable for further accurate quantitation.

14. This step is to dilute salts that may impede binding to the SCX cartridge, the estimated residual concentration of sodium ions from reactions after dilution is about 2.8 mM.
15. In COFRADIC, reduction of disulfide bonds in peptides between the primary and secondary run will shift retention times. This would lower the purity of azhal-containing peptides. Therefore, disulfide bonds are reduced and alkylated before the primary run under conditions that prevent extensive modification of azhal-containing peptides. To avoid reaction with the azide group of azhal, it is important in this step to have a TCEP end concentration of 2 mM and use only a short incubation time before quenching the reaction with IAA and NaN_3.
16. At least 50 mM HEPES, pH 8.0, should be used to ensure that the addition of 10 mM of TCEP does not drop the pH reaction mixture below 3; this would promote cleavage over the other two reactions and would result in less protein identifications. This is caused by the more erratic retention time shift that cleavage products have which more often moves them to fractions that are discarded.
17. Pooled fractions should be injected in order (e.g., A, B, C, D), so that possible carry-over between runs from the main chromatographic peaks will be mostly present in fractions that are discarded in the following run.
18. Mascot generic files (.mgf) from different runs can be concatenated by copying and pasting manually in a text editing program such as notepad. When combining mgf files make sure that the resulting concatenated mgf file does not have the two strings "SEARCH=MIS" and "REPTYPE=Peptide" anywhere, but just at the first and second line of the file. This is achieved by not selecting these first two lines from the mgf files you are copying into the concatenated one. Not doing so will result in an error message from MASCOT when attempting to search the concatenated file.
19. Significance settings in the MASCOT search engine and the actual significance level achieved vary with database size and data set quality. Always use a shuffled version of the used database to estimate false-positive rates and adjust significance thresholds to achieve the desired false-positive rate (usually <5% or <1%).
20. When selecting reaction products note that the C-terminal product resulting from cleavage after azhal (shown as

cleavage after M in MASCOT) can only have a lysine modified by iTRAQ and cannot have an N-terminal iTRAQ modification because it is generated after iTRAQ labeling. Any of these peptides which are assigned an N-terminal iTRAQ modification by MASCOT should be discarded as false positives.

21. When using the mixing scheme as described here, the iTRAQ ratios 116/114 and 117/115 represent the biological duplicates grown under identical conditions during the experiment. These ratios should theoretically be one and can be used to assay the combined effect of biological and technical variation of the experiment. These ratios can indicate a general bias in expression ratio in the samples caused by differences in azhal or iTRAQ labeling and can be used to compensate for these biases.

Acknowledgments

The authors wish to acknowledge Dr. A. Bock and Dr. M. Thanbichler for providing *E. coli* strain MTD123 and M.A. Nessen for synthesis of azhal.

References

1. Hoper, D., Bernhardt, J., and Hecker, M. (2006) Salt stress adaptation of Bacillus subtilis: A physiological proteomics approach, *Proteomics 6*, 1550–1562.
2. Pedersen, S., Bloch, P. L., Reeh, S., and Neidhardt, F. C. (1978) Patterns of protein synthesis in E. coli: A catalog of the amount of 140 individual proteins at different growth rates, *Cell 14*, 179–190.
3. Bateman, R. J., Munsell, L. Y., Chen, X., Holtzman, D. M., and Yarasheski, K. E. (2007) Stable isotope labeling tandem mass spectrometry (SILT) to quantify protein production and clearance rates, *J. Am. Soc. Mass Spectrom. 18*, 997–1006.
4. Cargile, B. J., Bundy, J. L., Grunden, A. M., and Stephenson, J. L., Jr. (2004) Synthesis/degradation ratio mass spectrometry for measuring relative dynamic protein turnover, *Anal. Chem. 76*, 86–97.
5. Gustavsson, N., Greber, B., Kreitler, T., Himmelbauer, H., Lehrach, H., and Gobom, J. (2005) A proteomic method for the analysis of changes in protein concentrations in response to systemic perturbations using metabolic incorporation of stable isotopes and mass spectrometry, *Proteomics 5*, 3563–3570.
6. Ong, S. E., Blagoev, B., Kratchmarova, I., Kristensen, D. B., Steen, H., Pandey, A., and Mann, M. (2002) Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics, *Mol. Cell. Proteomics 1*, 376–386.
7. Pratt, J. M., Petty, J., Riba-Garcia, I., Robertson, D. H., Gaskell, S. J., Oliver, S. G., and Beynon, R. J. (2002) Dynamics of protein turnover, a missing dimension in proteomics, *Mol. Cell. Proteomics 1*, 579–591.
8. Kramer, G., Sprenger, R. R., Back, J., Dekker, H. L., Nessen, M. A., van Maarseveen, J. H., de Koning, L. J., Hellingwerf, K. J., de Jong, L., and de Koster, C. G. (2009) Identification and quantitation of newly synthesized proteins in *E. coli* by enrichment of azido homoalanine-labeled peptides with diagonal chromatography, *Mol. Cell. Proteomics 8*, 1599–1611.
9. Dieterich, D. C., Link, A. J., Graumann, J., Tirrell, D. A., and Schuman, E. M.

(2006) Selective identification of newly synthesized proteins in mammalian cells using bioorthogonal noncanonical amino acid tagging (BONCAT), *Proc. Natl. Acad. Sci. USA 103*, 9482–9487.
10. Nessen, M. A., Kramer, G., Back, J., Baskin, J. M., Smeenk, L. E. J., de Koning, L. J., van Maarseveen, J. H., de Jong, L., Bertozzi, C. R., Hiemstra, H., and de Koster, C. G. (2009) Selective enrichment of azide-containing peptides from complex mixtures, *J. Proteome Res. 8*, 3702–3711.
11. Meiring, H. D., van der Heeft, E., ten Hove, G. J., and de Jong, A. P. J. M. (2002) Nanoscale LC-MS(n): Technical design and applications to peptide and protein analysis, *J. Sep. Sci. 25*, 557–568.
12. Boehm, A. M., Putz, S., Altenhofer, D., Sickmann, A., and Falk, M. (2007) Precise protein quantification based on peptide quantification using iTRAQ, *BMC Bioinformatics 8*, 214.
13. Kiick, K. L., Saxon, E., Tirrell, D. A., and Bertozzi, C. R. (2002) Incorporation of azides into recombinant proteins for chemoselective modification by the Staudinger ligation, *Proc. Natl. Acad. Sci. USA 99*, 19–24.
14. Gevaert, K., Van Damme, J., Goethals, M., Thomas, G. R., Hoorelbeke, B., Demol, H., Martens, L., Puype, M., Staes, A., and Vandekerckhove, J. (2002) Chromatographic isolation of methionine-containing peptides for gel-free proteome analysis: identification of more than 800 Escherichia coli proteins, *Mol. Cell. Proteomics 1*, 896–903.
15. Back, J. W., David, O., Kramer, G., Masson, G., Kasper, P. T., de Koning, L. J., de Jong, L., van Maarseveen, J. H., and de Koster, C. G. (2005) Mild and chemoselective peptide-bond cleavage of peptides and proteins at azido homoalanine, *Angew. Chem. Int. Ed. 44*, 7946–7950.
16. Kasper, P. T., Back, J. W., Vitale, M., Hartog, A. F., Roseboom, W., de Koning, L. J., van Maarseveen, J. H., Muijsers, A. O., de Koster, C. G., and de Jong, L. (2007) An aptly positioned azido group in the spacer of a protein cross-linker for facile mapping of lysines in close proximity, *Chembiochem 8*, 1281–1292.
17. Thanbichler, M., Neuhierl, B., and Bock, A. (1999) S-methylmethionine metabolism in Escherichia coli, *J. Bacteriol. 181*, 662–665.

Chapter 13

Analytical Strategies in Mass Spectrometry-Based Phosphoproteomics

Heidi Rosenqvist, Juanying Ye, and Ole N. Jensen

Abstract

Phosphoproteomics, the systematic study of protein phosphorylation events and cell signaling networks in cells and tissues, is a rapidly evolving branch of functional proteomics. Current phosphoproteomics research provides a large toolbox of strategies and protocols that may assist researchers to reveal key regulatory events and phosphorylation-mediated processes in the cell and in whole organisms. We present an overview of sensitive and robust analytical methods for phosphopeptide analysis, including calcium phosphate precipitation and affinity enrichment methods such as IMAC and TiO_2. We then discuss various tandem mass spectrometry approaches for phosphopeptide sequencing and quantification, and we consider aspects of phosphoproteome data analysis and interpretation. Efficient integration of these stages of phosphoproteome analysis is highly important to ensure a successful outcome of large-scale experiments for studies of phosphorylation-mediated protein regulation.

Key words: Protein phosphorylation, cell signaling, IMAC, TiO_2, tandem mass spectrometry, phosphopeptide sequencing.

1. Introduction

Since the discovery of reversible protein phosphorylation as a regulatory mechanism for enzymes in the 1950s (1, 2), the ubiquitous and versatile role of transient protein phosphorylation for cellular homeostasis has been established. The advent of highly sensitive tandem mass spectrometers and robust sample preparation methods now enables detailed analysis of protein

K. Gevaert, J. Vandekerckhove (eds.), *Gel-Free Proteomics*, Methods in Molecular Biology 753,
DOI 10.1007/978-1-61779-148-2_13,

phosphorylation events in individual purified proteins, in isolated protein complexes or organelles, or in whole tissues and organisms (3–5). Thus, 'functional phosphoproteomics' has emerged as a very powerful approach that is currently applied in many areas of biology and biomedicine to elucidate regulatory mechanisms and signaling networks in the context of stem cell biology, cellular metabolism and defense, epigenetics, cancer progression and metastasis, etc.

Phosphoproteomic workflows consist of several levels of analysis, biochemical treatment, and protein isolation, measurement, and interpretation (**Fig. 13.1**). First, the biological system of interest is identified and the relevant questions and hypotheses are raised. Typically, a perturbed biological state is compared to one or more control states. Perturbations may involve stimulation of cells or tissues with growth factors, hormones or drugs, deletion or attenuation of specific genes (knock-out, knock-in), or induction of disease processes, such as infection, mutation, or other types of stress.

Isolation of proteins from cell lines or tissues involves lysis and often also organelle isolation to address questions that relate to regulation in the context of (intra)cellular organization of proteins and their interaction partners (**Fig. 13.1**, right-hand side). Once proteins are extracted, they are proteolytically digested to generate peptides that are amenable to mass spectrometry analysis and sequencing up to this point, there are few differences between a regular proteomics study and a phosphoproteomics study. However, as will be discussed in the following sections, phosphoproteomics relies on efficient enrichment, detection, sequencing, and quantification of phosphopeptides. A variety of phosphoprotein and phosphopeptide handling protocols are available, and the choice of method is highly dependent on the overall aim of the study. Numerous phosphopeptide enrichment approaches have been established and successfully employed to reveal phosphoprotein and regulated phosphorylation sites in complex biological samples. **Table 13.1** displays an overview of the phosphopeptide enrichment methods that are widely used in phosphoproteomic studies. These enrichment methods will be discussed below. Likewise, there are numerous mass spectrometry-based strategies available to the phosphoproteomics researcher. In the following sections we provide descriptions and discussions of the most widely used methods and we comment on emerging or promising techniques that are likely to have an impact on future phosphoproteomics studies.

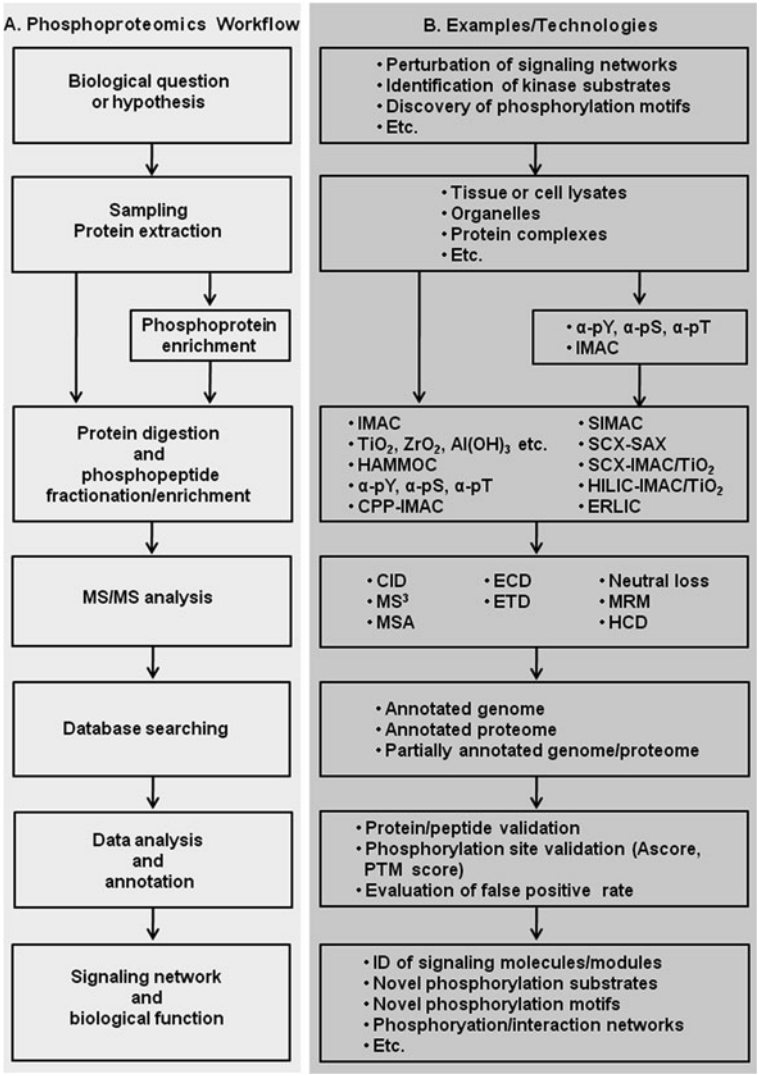


Fig. 13.1. Phosphoproteomic workflow. (**a**) Outline of the workflow. (**b**) Examples/technologies employed in phosphoproteomic studies. The biological question(s) of interest lead(s) to a decision on model system and sample workup procedure. Enrichment of phosphorylated species can take place at the protein and/or peptide level using various protocols. At the mass spectrometry level, MS^2, MS^3, or pseudo-MS^3 is often applied to phosphopeptide analyses. The subsequent protein sequence database search will depend on the type of sample being studied. Data analysis and annotation of the phosphoproteins/phosphopeptides and their phosphorylation sites include the validation of site assignments and evaluation of the false-positive rate. Finally, the identified phosphoproteins may be placed in specific molecular networks based on knowledge bases, and new substrates may be discovered. Phosphorylation sites can be grouped as either known, previously reported sites or as novel phosphorylation sites and possibly also as phosphorylation sequence motifs in substrate proteins.

Table 13.1
Overview of methods for separation, isolation, or enrichment of phosphopeptides prior to mass spectrometry analysis

Phosphopeptide enrichment method	Mechanism	Advantages	Disadvantages	References
IMAC	Affinity between positively charged metal ions and negatively charged phosphate ions	Simple, fast, cheap, and efficient for multiply phosphorylated peptides or acidic phosphopeptide enrichment; only low microgram levels of starting material are needed	Regular acidic peptides can also bind to IMAC resin; sensitive to various buffers and detergents used in biochemical procedures	(9–19)
Immunoprecipitation	Immuno-affinity between antigens and antibodies	Antibodies for phosphotyrosine is very efficient for phosphotyrosine enrichment	Limited specificity of the phosphoserine and phosphothreonine antibodies; milligram levels of starting material are required	(42, 45–47, 51, 158)
TiO_2 ZrO_2 HAMMOC	Phosphate anions and the surface of TiO_2 form a bridging bidentate surface complex	Simple, fast, tolerant to most buffers and salts used in biochemical or biological experiments; low level of starting material is required	Preference to singly phosphorylated peptides; multiply phosphorylated peptides are difficult to elute from TiO_2 column	(21, 29–34, 159)
SCX SAX SCX-SAX	Ion (anion or cation) exchange	Good for large-scale phosphoproteomic studies (phosphopeptide enrichment and fractionation in a single step)	Usually more than 200 μg of starting material is required; insufficient specificity	(63, 64, 71, 72)
CPP Ba^{2+} precipitation	Co-precipitation of phosphopeptides with calcium phosphate at basic conditions	Simple, fast, cheap, tolerant to most detergents used in biochemical experiments; efficient for low complexity protein samples; compatible with IMAC/TiO_2 enrichment	Bias toward acidic phosphopeptides; less efficient for very complex samples unless combined with IMAC/TiO_2 enrichment	(56–59)
HILIC	Fractionate biomolecules based on their polarity (hydrophilicity)	Achieves both phosphopeptide enrichment and fractionation in a single step	Insufficient specificity unless combined with further IMAC/TiO_2 enrichment	(75, 160)

Table 13.1 (continued)

ERLIC	Hydrophilic interaction and electrostatic repulsion	Achieves both phosphopeptide enrichment and fractionation in a single step; higher specificity than HILIC	Recently established method and the properties need to be characterized	(76, 77)
SIMAC IMAC-IMAC	Combines the preference of IMAC for multiply phosphorylated peptides with the preference of TiO_2 for singly phosphorylated peptides	Separation of multiply phosphorylated peptides from singly phosphorylated peptides; needs only low microgram levels of starting material	Requires careful optimization of loading and elution conditions	(61)
CPP-IMAC/TiO_2	As above	Prefractionation with CPP can efficiently remove salts and detergents prior to IMAC/TiO_2 enrichment; very good specificity; only low microgram levels of starting material are required	Bias toward acidic phosphopeptides	(56)
SCX-IMAC/TiO_2 SAX-IMAC/TiO_2	As above	Complementary separation mechanisms reduce the complexity of the sample, thus increasing the efficiency of IMAC/TiO_2 enrichment	Hundreds of microgram levels of starting material are required; low reproducibility due to the multistep enrichment	(65–68)
HILIC-IMAC/TiO_2	As above	HILIC reduces the complexity of the samples and IMAC/TiO_2 provides specificity for phosphopeptide enrichment	Up to hundreds of microgram levels of starting material are required; low reproducibility due to the multistep enrichment approach	(75, 160)

2. Strategies to Enrich Phosphopeptides from Complex Mixtures

2.1. Immobilized Metal Affinity Chromatography (IMAC)

Immobilized metal affinity chromatography was initially used for affinity purification of proteins with N-terminal or C-terminal his-tags (6–8). Andersson and Porath found that phosphorylated proteins and phosphorylated amino acids would bind to ferric ions immobilized on iminodiacetate-agarose gel and could be eluted by increasing pH or by adding phosphate salts to the eluant (9). They successfully applied the procedure to the purification of hen egg albumin and porcine pepsin. Now, IMAC has become one of the most widely used phosphopeptide enrichment methods (10–13). Both Ga(III)- and Fe(III)-based methods have been successfully applied in different large-scale phosphoproteomic studies (12, 14, 15). Although IMAC can predominantly enrich phosphopeptides, it may suffer from insufficient specificity, as acidic peptides may also interact with this resin (13). Many efforts have been made to improve the specificity of the IMAC method. For instance, the use of 1,1,1,3,3,3-hexafluoroisopropanol as both loading and washing solvents enhanced the efficiency of IMAC enrichment (16). Imanishi et al. optimized the phosphopeptide elution conditions in Fe(III)-IMAC utilizing the combination of phosphoric acid and acetonitrile (ACN) (17). The combination of 0.1% TFA in 50% ACN as both loading and washing solvents increased the selectivity of Phos-select IMAC material (18). Ficarro and coworkers established a method to overcome the nonspecific binding of acidic peptides by performing *O*-methyl esterification of peptide carboxyl groups. Thereby, the carboxylate groups are neutralized and the only negatively charged groups remaining are the phosphate groups. In this manner O-esterification may improve the specificity of phosphopeptide enrichment considerably (19, 20). However, substoichiometric or unspecific O-esterification, hydrolyzation of the carboxylic esters, and other side reactions may lead to increased sample complexity and artifacts, which are the major disadvantages of this method (21, 22). Recently, we optimized the Fe(III)-NTA IMAC enrichment protocol for phosphopeptides based on the differences in the degree of ionization of carboxylate groups and phosphate groups in different loading solutions (23). This protocol proved to be very selective for enrichment of phosphorylated peptides from a highly diluted standard sample (1:1,000). Furthermore, our method was applied to phosphoproteome analyses of mouse cell lysates, *Drosophila melanogaster* cell lysates, and plant tissue extracts. From either 20 μg of mouse sample or 50 μg of *D. melanogaster* sample, more than 1,000 phosphorylation sites were identified. Similarly, from 50 μg of plant plasma membrane protein preparation, a total of 876 phosphorylation sites were identified (23). We demonstrated that it is possible

to separate multiply phosphorylated peptides from singly phosphorylated peptides with successive IMAC enrichments. **Figure 13.2** shows the LC-MS ion chromatograms of the plant sample before and after the successive IMAC enrichments. **Figure 13.2a** shows the LC-MS ion chromatogram of the sample prior to any phosphoenrichment, whereas **Fig. 13.2b, c** shows the chromatograms after the first and second IMAC enrichments. It is obvious that there is a significant difference in the LC-MS profiles before and after IMAC enrichments, as well as between the first and second IMAC enrichment. From the first IMAC enrichment, half of the identified phosphopeptides were multiply phosphorylated, whereas these only accounted for less than 4% in the second enrichment, which recovered mostly singly phosphorylated peptides.

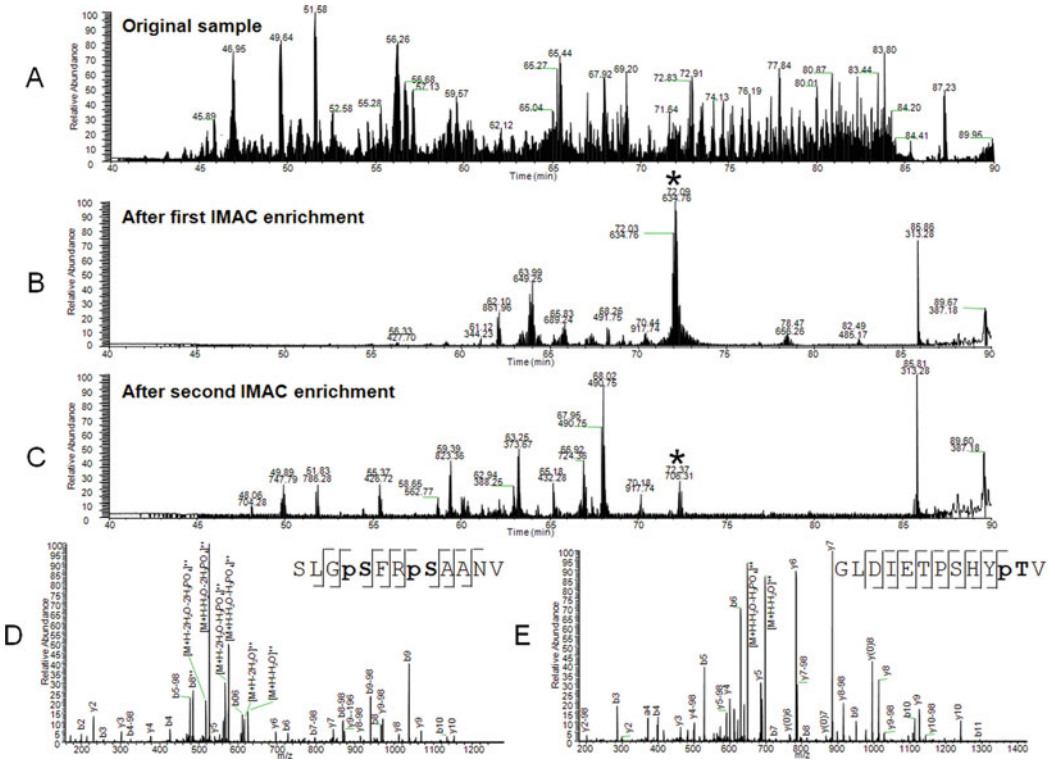


Fig. 13.2. The LC-MS chromatograms of a plant protein sample before and after successive IMAC enrichments for phosphopeptides. (**a**) The LC-MS chromatogram of the crude plant plasma membrane protein preparation after trypsin digestion; (**b**) the LC-MS chromatogram after the first IMAC enrichment of phosphopeptides; (**c**) the LC-MS chromatogram after the second IMAC enrichment of phosphopeptides; (**d**) MS/MS spectrum of a doubly phosphorylated peptide SLGpSFRpSAANV recovered in the first IMAC enrichment (indicated with *bold star* in LC-MS chromatogram **b**); (**e**) MS/MS spectrum of singly phosphorylated peptide **GLDIETPSHYpTV** derived from the second IMAC enrichment (indicated with *bold star* in chromatogram **c**). In MS/MS spectra (**d**) and (**e**), ⌈ indicates observed y-ions, ⌋ indicates observed b-ion, and ⌈⌋ indicates observed y- and b-ions.

2.2. Phosphopeptide Enrichment by TiO_2 and Other Metal Oxides

A range of different metal oxides/hydroxides have been applied for phosphopeptide enrichment (24–29). Of these, TiO_2 was the first metal oxide found to have affinity for phosphate groups of phosphopeptides (30). TiO_2 was shown to efficiently retain organic phosphates at acidic conditions, while alkaline conditions caused elution of these (30). These properties were also exploited by Pinkse et al. who in 2004 introduced an LC-MS/MS-based procedure for online phosphopeptide enrichment, separation, and sequencing by using TiO_2 (26). Phosphopeptide enrichment by TiO_2 protocols has since been assessed, improved, and modified by our laboratory and several other research groups (21, 27, 31, 32). Sano and Nakamura evaluated Ti, TiO, TiO_2, Ti_2O_3, Ti_3O_5, and $TiO_{1.98}$ for their phosphate affinity and found Ti to be more selective than titania (27). Binding of phosphopeptides to TiO_2 can be attributed to the ion exchange properties of the metal dioxide (26), and this may also affect the retention of singly versus multiply phosphorylated peptides. While TiO_2 clearly binds both types of phosphopeptides, the conditions at which they can be eluted seem to differ (33). Under typical elution conditions (elution buffer pH 10–11.5), the majority of the phosphopeptides released from their TiO_2 binding will be singly phosphorylated (21), but change of pH and elution buffer may cause the release of the multiphosphorylated peptides as well (33). Zirconium dioxide (ZrO_2) has also been successfully applied for phosphopeptide enrichment prior to mass spectrometry (29, 34, 35), either in a column (microtips) setup (34–37) or as nanoparticles (29, 38). Less attention has been given to aluminum hydroxide, $Al(OH)_3$ (28, 35), but more recently tin dioxide, SnO_2 (25), niobium dioxide, Nb_2O_5 (24), and tantalum dioxide, Ta_2O_5 (39), have shown some potential for phosphopeptide enrichment.

2.3. Improved TiO_2 Performance by Using Hydroxy Acid Additives

The use of TiO_2 alone and with different additives to enhance phosphopeptide enrichment has been tested in several studies. These include, for example, addition of 2,5-dihydroxybenzoic acid and glycolic acid to the sample solution (40, 41). In 2007, Sugiyama et al. included a range of different metal oxides along with a range of aliphatic hydroxy acids as additives (35). They found that addition of lactic acid and β-hydroxypropanoic acid to the loading buffer significantly reduced co-enrichment of non-phosphopeptides in TiO_2 and ZrO_2 chromatography (35). The procedure was termed HAMMOC for hydroxy acid-modified metal oxide chromatography and subsequently used for large-scale phosphorylation mapping in *Arabidopsis* (37). Further optimization by successive elution with series of 5% ammonium hydroxide solution, 5% piperidine solution, and 5% pyrrolidine solution was recently tested on HeLa cell lysates, showing

identification of 1,803 non-redundant phosphopeptides from 100 μg cytoplasmic protein fractions (36).

2.4. Phosphorylation-Specific Antibodies

One of the traditional procedures for phosphoprotein/peptide enrichment relies on immuno-affinity. Antibodies raised against phosphotyrosine, phosphoserine, or phosphothreonine are available to use for immunoprecipitation and/or Western blotting (42–45). However, the rather low specificity of some of these antibodies limits their general use in phosphoproteomics (46). Several highly specific antibodies exist for phosphotyrosine (44, 45, 47, 48). Antibodies for phosphoserine and phosphothreonine generally suffer from specificity issues (46), yet have been successfully used to identify novel phosphoproteins (42). Because antibodies are primarily used to recover or detect intact proteins, immunoprecipitation with phospho-specific antibodies can be used as an effective protein prefractionation step prior to proteolysis and phosphopeptide-specific enrichment procedures such as IMAC (49, 50). Immunoprecipitation can also be performed at the peptide level, though, with subsequent mass spectrometric analysis to identify phosphorylation sites in the peptides (47, 51–53). As an alternative to phospho-specific antibodies, protein-specific antibodies can be used to enrich for a specific protein of interest prior to phosphorylation analysis (54).

2.5. Calcium Phosphate Precipitation (CPP)

Reynolds et al. reported that the phosphoseryl-containing peptides of tryptic digests of casein were selectively precipitated using calcium ions (20 mol/mol protein) and 50% (v/v) ethanol at pH 3.5, 4.6, and 8.0 (55). In their study, only peptides containing the cluster sequence -Ser(P)-Ser(P)-Ser(P)- were precipitated by Ca^{2+}/ethanol at pH 3.5. At pH 4.6 the precipitate contained all the cluster peptides plus two diphosphorylated peptides containing -Ser(P)-Glu-Ser(P)- and -Ser(P)-Thr-Ser(P)-, while at pH 8.0, a singly phosphorylated peptide containing -Ser(P)-Glu-Glu- was also present in the precipitate (55). In 2007, our group presented a method for phosphopeptide enrichment by calcium phosphate precipitation (CPP) (56). We used disodium phosphate (Na_2HPO_4) and calcium chloride ($CaCl_2$) for phosphopeptide precipitation at alkaline conditions (the pH was adjusted by ammonia solution, $NH_3{\cdot}H_2O$). We assume that the phosphate groups on phosphopeptides incorporate into the calcium phosphate lattice and thereby precipitate out of peptide mixtures. Our CPP method was successfully applied to the phosphopeptide enrichment from α-casein peptide mixtures and a more complex phosphoprotein-containing sample (a tryptic digest of five proteins: α,β-caseins, BSA, carbonic anhydrase, and β-lactoglobulin). When combined with IMAC, the CPP protocol has exhibited high efficiency and selectivity in ongoing plant phosphoproteomic studies. A total of 227 unique phosphorylation sites were

identified from 200 μg of protein that was isolated from dormant rice embryos (56), 147 unique phosphopeptides were obtained from 100 μg spinach thylakoids (Li et al., submitted), and 1,098 unique phosphopeptides were acquired from 50 μg *Arabidopsis* plasma membranes (Ye, manuscript in preparation). Using CPP, Kametani et al. identified 29 phosphorylation sites on recombinant TDP-43 that is phosphorylated by casein kinase-1 (57). Xia and coworkers described the phosphoproteome analysis of postmortem Alzheimer's disease brain tissue using CPP prior to LC-MS/MS analysis. A total of 551 phosphopeptides (466 phosphorylation sites) were identified from the brain tissue, including 379 phosphorylation sites on serine residues and 87 on threonine residues (58). The Yates group recently described a method that is based on Ba^{2+}-binding properties of amino acids (59). According to the previous study by Reynolds et al. (55), three values of Ba^{2+} ion molarity (1, 6, and 7.5 μmol Ba^{2+} ions) at three different pH conditions (pH 3.5, 4.6, and 8.0) were tested. They found that precipitation with 7.5 μmol Ba^{2+} ions at pH 3.5 identified the highest number of unique phosphopeptides. Combining Ba^{2+}/pH/acetone-based precipitations with multidimensional protein identification technology (MudPIT), a total of 1,037 phosphopeptides were identified from 250 μg nuclear extracts of HeLa cells (59). All these examples and applications clearly demonstrate that phosphopeptide precipitation by CPP and related methods are viable new tools in phosphoproteomic research.

2.6. Sequential Elution from IMAC (SIMAC)

Although numerous phosphopeptide enrichment approaches have been developed in recent years, the identification of multiply phosphorylated peptides by MS is still a challenge. Multiply phosphorylated peptides suffer from poor ionization efficiency and are suppressed in the presence of both singly and non-phosphorylated peptides. Therefore, in order to recover more multiply phosphorylated peptides, it is necessary to separate these from the singly and non-phosphorylated peptides prior to MS analysis. In 2008, our laboratory demonstrated sequential elution from IMAC (SIMAC), which combines the strengths of IMAC (for multiply phosphorylated peptides) with the strengths of TiO_2 (for singly phosphorylated peptides) (20, 60, 61). An optimized version of IMAC (18) was exploited in SIMAC, where the flow-through (void volume) containing both unmodified and phosphorylated peptides was collected for further enrichment using TiO_2 chromatography. Singly phosphorylated peptides were eluted from the IMAC beads using acidic conditions (1% TFA), while the multiply phosphorylated peptides retained on IMAC beads as well as the phosphopeptides bound with TiO_2 chromatography were eluted using alkaline conditions (pH 11.3). By this strategy, singly and multiply phosphorylated peptides

could be enriched and recovered in separate fractions and analyzed separately by MALDI MS/MS or LC-MS/MS. Thus, suppression of multiply phosphorylated peptides by monophosphorylated peptides or unmodified peptides during ionization and MS detection could be reduced, and as a result more multiply phosphorylated peptide identifications were obtained. The SIMAC strategy was applied to the analysis of 120 μg of whole-cell extract from human mesenchymal stem cells and compared to the phosphopeptide enrichment provided by TiO_2 chromatography alone. The SIMAC procedure resulted in the identification of 716 phosphorylation sites as opposed to 350 by TiO_2 chromatography alone, mainly due to a significant increase in the number of multiply phosphorylated peptides identified (61). A similar performance can be achieved by optimization of IMAC for sequential elution (IMAC-IMAC) (23).

2.7. Ion Exchange Chromatography

Strong ion exchange chromatography has shown great potential for fractionation of phosphorylated peptides from non-phosphorylated species and has been used as a stand-alone method as well as in combination with complementary phosphopeptide enrichment techniques. Our group was the first to publish the use of strong anion exchange (SAX) in a phosphoproteomic study (62). We combined SAX with IMAC enrichment for the identification of more than 200 phosphorylated peptides in plasma membrane fractions from *Arabidopsis thaliana*. Fractions eluting from the SAX column at low salt concentrations contained an abundance of singly phosphorylated peptides, while multiply phosphorylated peptides were more strongly retained by SAX and eluted at higher salt concentrations (62). A recent study by Han et al. identified 274 phosphorylation sites in human liver tissue using SAX without further enrichment steps (63).

In 2004, Beausoleil et al. described the use of strong cation exchange (SCX) alone and in combination with SDS-PAGE for phosphoproteomics (64). By SCX alone they identified 2,002 phosphorylation sites from 300 μg of the nuclear fraction of HeLa cell lysate, and by SDS-PAGE combined with SCX, more than 500 phosphorylation sites were determined in the developing mouse brain (64). Separation of peptides by SCX is charge dependent, and thus in an acidic solvent (pH~2.7), singly charged phosphopeptides are expected to elute in the beginning of the high salt gradient (64, 65). The charge dependence limits the use of SCX in phosphoproteomics somewhat, though, as not only non-phosphorylated peptides but also multiply phosphorylated peptides with only two basic sites will be retained in the stationary phase, and thus not enriched (64). Yet, combined with IMAC or TiO_2, large numbers of phosphopeptides and phosphorylation sites can still be identified. In 2005, our group identified 729 unique phosphopeptides in yeast by an SCX-IMAC

procedure (66). Trinidad et al. later compared SCX alone, IMAC alone, and the combination of SCX and IMAC in a phosphoproteomic study, managing to identify 998 unique phosphorylated peptides in postsynaptic density samples from mouse brains (67). They found that by performing IMAC on each SCX fraction, a threefold increase in phosphopeptide identifications was achieved relative to either approach alone (67). Villen et al. studied the mouse liver phosphoproteome and identified 8,527 phosphorylated peptides with 5,250 non-redundant phosphorylation sites using the SCX-IMAC approach (51). The combination of SCX and TiO_2 has revealed just as impressive results. With SCX-TiO_2, Olsen et al. identified phosphopeptides from 2,244 proteins in HeLa cells, when using milligram levels of protein starting material (68). Wu et al. detected more than 4,000 distinct phosphopeptides in breast cancer cells (69), and Pinkse et al. were able to enrich 2,152 phosphopeptides from 250 μg protein derived from lysates of *D. melanogaster S2* cells (70).

A combination of SAX and SCX has also been employed. In 2007, Dai et al. introduced multidimensional liquid chromatography (MDLC) with a combination of SCX, SAX, and reverse phase methods for large-scale protein characterization and phosphorylation site mapping from complex samples (71). They termed this procedure Yin-Yang MDLC and managed to identify 14,105 unmodified peptides and 849 phosphorylated peptides with 809 phosphorylation sites (71). Recently, Dai et al. also compared the performance of pH continuous online SAX (pCOG-SAX) to the SCX-TiO_2 approach and found the two procedures to be complementary with different affinities toward acidic and basic peptides (72).

2.8. Hydrophilic Interaction Chromatography (HILIC)

Hydrophilic interaction chromatography (HILIC, also known as normal-phase chromatography) is a separation method, which fractionates biomolecules based on their polarity (hydrophilicity). HILIC was initially used for small, polar molecules (e.g., carbohydrates, oligonucleotides, and amino acids). Since 1990, HILIC has been used to reduce the complexity of peptide/glycopeptide mixtures through depletion of hydrophobic peptides and retention of hydrophilic peptides/glycopeptides (73). In HILIC, samples are loaded at high organic solvent concentration and eluted by increasing the polarity of the mobile phase (e.g., an inverse gradient of ACN in water). This can be considered the opposite of reverse phase chromatography. Therefore, the more hydrophilic the biomolecule is, the longer it is retained in the stationary phase and the later it elutes from the column. In 2007, Boersema et al. described a zwitterionic HILIC system and used it in a 2D LC scheme for proteomic applications (74). They demonstrated that ZIC-HILIC is a very good alternative for the more conventional SCX in multidimensional peptide separation strategies. McNulty

and Annan applied HILIC to a phosphoproteomic study in 2008 (75). In their study, HILIC was used as a first-dimension separation for 2D LC proteomics. When compared to SCX, HILIC provided an equivalent number of peptide and protein identifications for cell lysates. They also found that phosphopeptides exhibited increased retention relative to non-phosphorylated peptides. Using a TSKgel Amide-80 column with an optimal gradient, they identified over 1,000 unique phosphorylation sites in twenty 50-min LC-MS/MS experiments from a 300 μg equivalent of HeLa cell lysate. They further demonstrated that the volatile, salt-free TFA/acetonitrile buffer system used with HILIC was fully compatible with direct IMAC enrichment without any intermediate desalting steps, and the subsequent IMAC enrichment of phosphopeptides from HILIC fractions showed better than 99% selectivity. They compared the IMAC enrichment prior to HILIC fractionation and after HILIC fractionation and found that non-phosphorylated peptides accounted for 43% of the total in the IMAC-HILIC experiment, whereas in the reverse approach they only accounted for less than 1%. Therefore, it seems that HILIC can be used as a highly efficient prefractionation procedure for biological complex samples prior to phosphopeptide enrichment by affinity-based methods, such as IMAC and TiO_2.

2.9. Electrostatic Repulsion–Hydrophilic Interaction Chromatography (ERLIC)

Electrostatic repulsion–hydrophilic interaction chromatography (ERLIC) was recently introduced by Alpert (76) as a potential phosphopeptide enrichment method. It is based on hydrophilic interaction and electrostatic repulsion. At low pH (pH~2), the ionization efficiencies of carboxyl groups at Asp and Glu residues as well as the C-terminus are largely suppressed, alike weak anion exchange (WAX) chromatography, and peptides are generally electrostatically repulsed by the column at their positively charged N-termini. However, phosphopeptides with negatively charged phosphate groups will be electrostatically attracted to WAX and this interaction increases their retention time compared to non-phosphopeptides. The retention time of phosphopeptides is further enhanced by a high concentration of organic solvent (such as 70% acetonitrile), since it promotes hydrophilic interaction of the phosphate group with the column (76). Gan and colleagues compared ERLIC with SCX and with SCX-IMAC for identifying phosphopeptides in EGF-treated A431 cells (77). The efficiency of the procedure, the specificity of phosphopeptide enrichment, the number of identified phosphopeptides and detected phosphorylation sites, as well as the number of unique phosphorylated proteins detected by each approach were studied. In comparison with SCX only (641) and SCX-IMAC strategies (4,850), the ERLIC approach detected the highest number of phosphopeptides (17,311). However, only 926 unique phosphopeptides (representing 761 unique phosphorylation sites) were identified

in ERLIC as compared to 194 in SCX only (representing 202 unique phosphorylation sites) and 1,315 in SCX-IMAC (representing 984 unique phosphorylation sites). Altogether, the three methods identified a total of 2,058 unique phosphopeptides, of which 1,801 (88%) were identified by only one of the methods. Of these 1,801 phosphopeptides, SCX-IMAC accounted for 57%, ERLIC for 38%, and SCX only for 5%. Compared to SCX-IMAC, ERLIC achieves both sufficient phosphopeptide enrichment and fractionation in a single step. It is interesting that only 2.4% unique phosphopeptides were common to all three approaches, which indicated that the ERLIC approach could potentially complement the SCX or SCX-IMAC approaches to generate a more complete phosphoproteome.

3. Phosphorylation Site Mapping by Tandem Mass Spectrometry

To understand a biological event, it is important to be able to assign the phosphate group to a specific amino acid. In contrast to unmodified peptides, phosphorylated peptides have different characteristics which can make them challenging to analyze by mass spectrometry (5). Analysis of protein phosphorylations can be performed by various mass spectrometric techniques, some of which will be described here. For a more thorough introduction, Boersema et al. published an excellent review on MS/MS sequencing of phosphopeptides (78).

3.1. Collision-Induced Dissociation (CID)

Collision-induced dissociation (CID) is the most common MS/MS method used to fragment peptide ions in the gas phase. Peptide ions are usually accelerated by an electrical potential to high kinetic energy in the vacuum of a mass spectrometer and then allowed to collide with neutral gas molecules (e.g., argon) to produce a series of y- or b-type peptide fragment ions (79), and the peptide sequence can be interpreted according to this. When phosphopeptide ions are fragmented with CID, cleavage of the *o*-phosphoric acid bond is the preferred fragmentation channel since it is much more labile than the peptide bond. As a consequence, the major MS/MS fragmentation pathway of phosphopeptides is usually the loss/elimination of the phosphate group and a water molecule (66). There are intense signals at $[M+H]^+$-98 Da (loss of phosphate and water) for phosphoserine/threonine-containing peptides and partial neutral losses ($[M+H]^+$-80 Da, loss of HPO_3) are also observed for phosphotyrosine-containing peptides (75, 80, 81). Due to the aromatic nature of the side chain, the phosphate group on tyrosine residues is much more stable than on

the serine/threonine residues. Therefore, phosphorylation sites can be determined by the observation of 'indicator' signals in MS/MS spectra. β-Elimination of the phosphate group on phosphoserine and phosphothreonine produces dehydroalanine (69 Da) and dehydro-2-amino butyric acid (83 Da), respectively (82). Since phosphotyrosine does not undergo β-elimination due to its aromatic side chain, the intact phosphotyrosine residue is observed in peak patterns of MS/MS spectra (243 Da) (83). Moreover, the phosphotyrosine immonium ion at m/z 216 can be used as a diagnostic ion for phosphotyrosine-containing peptides when using high-accuracy MS and MS/MS (83, 84). However, these neutral losses or diagnostic ion fragments are often of low intensity or masked by other intense fragment ions and it is sometimes a challenge to assign the correct phosphorylation sites.

Many efforts have been directed toward the development of robust and reliable methods for interpretation of phosphopeptide MS/MS spectra to accurately and unambiguously annotate phosphorylation sites. The method of data-dependent neutral loss-triggered MS^3 (MS/MS/MS) analysis has proved to be a useful tool in detecting and sequencing of phosphopeptides by ion trap-type mass analyzers (64, 66, 85). The principle is the following: a peptide is selected for MS/MS fragmentation and when a neutral loss of phosphoric acid is detected (–98 Da) in MS/MS, the ion produced by the neutral loss will be automatically isolated and further fragmented by MS/MS/MS. This method results in relatively more phosphopeptide sequence information and helps the localization of the phosphorylation sites (66). It has been successfully applied to several large-scale phosphoproteomic studies (64, 66, 68, 86, 87). However, the efficiency of MS^3 is only good for abundant phosphopeptide ions and the additional fragmentation step reduces the MS duty cycle and may reduce the total number of fragmented (phospho)peptides when studying complex samples. Also, amino acid sequence information for multiply phosphorylated peptides will still be scarce, as only the neutral loss ion derived from the loss of the first phosphate group will be further fragmented (85, 88).

Multistage activation (MSA), also known as pseudo-MS^3, is a combination of MS^2 and MS^3 that takes advantage of the ion isolation features of ion traps (89). MSA allows for recognition of the neutral loss peak resulting from the loss of the phosphate group, enabling a second stage of activation and fragmentation of this ion. In MSA, the fragmentation of the precursor ion occurs simultaneously with the fragmentation of the ion originating from the neutral loss, and the resulting pseudo-MS^3 spectrum is added to the original MS^2 spectrum. The hybrid spectrum, which combines the features of MS^2 and MS^3, therefore results in more MS/MS information for a better identification from

database searching and an improved phosphorylation site assignment. However, Palumbo and Reid have found that the CID and multistage tandem mass spectrometry (MS^2 and MS^3) fragmentations in a linear ion trap can induce the transfer of a phosphate group from the phosphorylated residue to an unmodified hydroxyl-containing amino acid residue (90). Alternative fragmentation methods, electron capture dissociation (ECD) and the related electron transfer dissociation (ETD), have been reported to successfully sequence phosphopeptides with preservation of the phosphate moiety on most c- and z-fragment ions (91, 92), as discussed in the next section.

3.2. Electron Capture Dissociation (ECD) and Electron Transfer Dissociation (ETD)

Electron capture dissociation (ECD) was introduced by Zubarev and coworkers in 1998 (93). The principle of this method is that low-energy electrons are introduced to trapped gas phase ions and this results in cleavage of the peptide's amine bonds to form c- and z-fragments. The peptide ion fragment patterns generated by ECD are quite different from other techniques. A major advantage of ECD is that fragmentation only occurs on the peptide backbone, thus the labile side chain modifications, such as phosphate groups and N- or O-linked glycan groups, will stay intact on the resulting c- or z-fragment ions, which enables the localization of the specific modification sites and the characterization of the glycan structures (13, 94, 95). It is not depending on the peptide sequence except for its selectivity for disulfide bonds and the inefficient cleavage at the amide bond N-terminal to proline residues (93). Whereas ECD is excellent for phosphopeptide sequencing (96, 97), the main drawbacks are that the cleavage efficiency is not particularly high and that ECD presently is only compatible with FT-ICR instruments although it has shown some promise for peptide sequencing in quadrupole-type instruments (98).

Electron transfer dissociation (ETD) was developed with the aim to perform ion–electron reactions in quadrupole ion trap analyzers. This method was initially implemented in a linear ion trap (LTQ) (91, 99) while it has recently been coupled with high-resolution and high mass accuracy analyzers, such as Orbitrap and Q-TOF instruments (100, 101). ETD fragments peptides in a similar way as ECD (primarily backbone cleavages and no loss of the labile modifications) (91), but as ETD is compatible with less expensive ion trap instruments as well as instruments with high resolution and mass accuracy, it has become more widespread. For example, we have used ETD and ECD to sequence highly labile phosphohistidine-containing peptides (102). Chi et al. combined IMAC enrichment with ETD MS/MS in a yeast phosphoproteomic study. From 30 μg of the sample, they could identify 1,252 phosphorylation sites on 629 proteins (91). Meanwhile, Molina et al. combined TiO_2 enrichment with

ETD MS/MS in a large-scale study of human embryonic kidney 293T cells and obtained 1,435 phosphorylation sites from 500 μg of sample, of which approximately 80% sites were novel (103). ETD fragmentation is more efficient for highly charged peptide and phosphopeptide ions (3^+ or more) than MSA. **Figure 13.3** shows an example of a doubly phosphorylated peptide, RPSSSAS**pT**KD**pS**ESPRHFIPADYLESTEEFIR (mass: 3,698.629 Da, charge: 4+), with high molecular weight analyzed by CID and ETD. It is rather straightforward to read the sequence and localize the phosphorylation sites from the ETD spectrum, while it is not possible to unambiguously assign the phosphorylation sites based on the CID spectrum. Since CID performs well for doubly charged peptides, it has been reported that a combination of ETD and CID by an automatic alternating mode generated improved results (100, 103, 104); however, this method is not

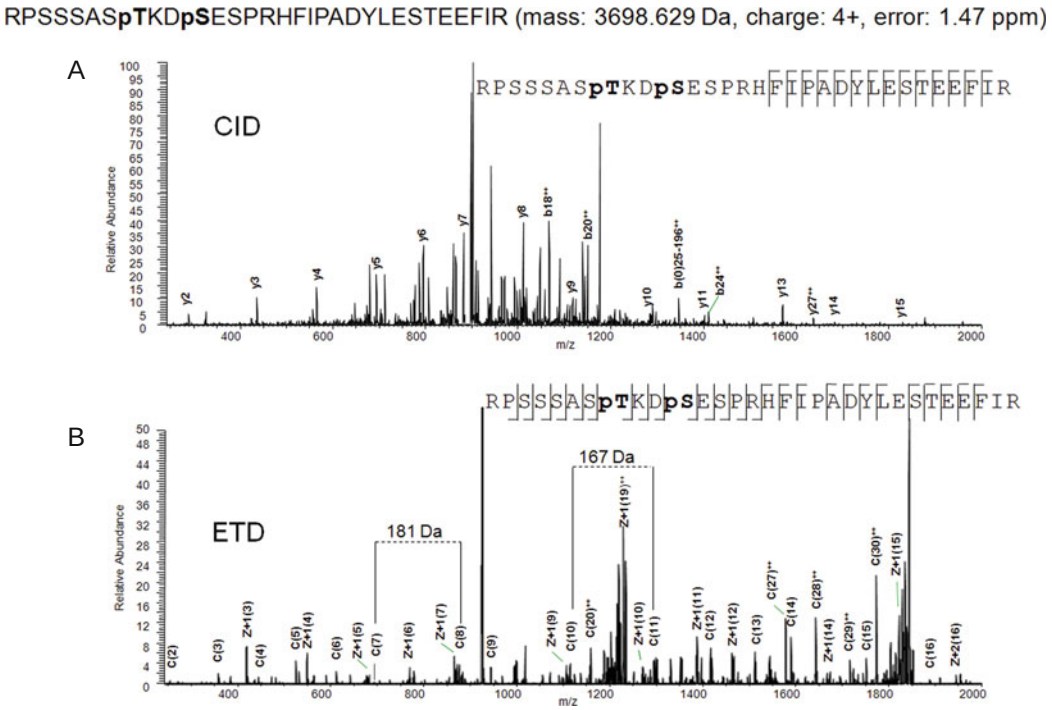


Fig. 13.3. MS/MS spectra of quadruply charged and doubly phosphorylated peptide RPSSSAS**pT**KD**pS**ESPRHFIPADYLESTEEFIR obtained from CID and ETD MS/MS. (**a**) MS/MS spectrum of peptide RPSSSAS**pT**KD**pS**ESPRHFIPADYLESTEEFIR from CID fragmentation mode. (**b**) MS/MS spectrum of peptide RPSSSAS**pT**KD**pS**ESPRHFIPADYLESTEEFIR from ETD fragmentation mode. In MS/MS spectra (**a**), ⌈ indicates y-ion was observed, ⌋ indicates b-ion was observed, and ⌈⌋ indicates both y- and b-ions were observed. In MS/MS spectra (**b**), ⌈ indicates z-ion was observed, ⌋ indicates c-ion was observed, and ⌈⌋ indicates both z- and c-ions were observed.

yet widely used. Since ECD/ETD performs best for multiply charged ions ($>2^+$), efforts are made to generate larger and therefore more highly charged (phospho)peptides, for example, using alternative proteases such as Lys-C (103), Glu-C (103), chymotrypsin (105), and Lys-N (106–108). The most frequently used protease, i.e., trypsin, mainly produces doubly charged ions in MS, which are more suitable for CID but less well suited for ECD or ETD MS/MS. Endoproteinase Lys-C cleaves C-terminal to lysine residues, thus generally creating rather long peptides. However, the ETD performance for phosphopeptide sequencing using either trypsin or Lys-C digests showed no major improvement by using the latter enzyme (103). This might be explained by a high level of missed cleavage sites in the tryptic digest, which is caused by the presence of phosphate groups or highly acidic residues near the lysine substrate sites that in turn inhibits enzyme–substrate binding and cleavage (103). The major limitation of using endoproteinase Glu-C and chymotrypsin was the poor cleavage efficiency and low specificity, which in turn increased the sample complexity and reduced the number of phosphopeptide identifications (103, 105). The Heck group explored the utility of endoproteinase Lys-N digestion for peptide sequencing by ETD MS/MS. They found the peptide sequence easy to interpret from the MS/MS spectrum and the exact phosphorylation sites could often be readily assigned (106–108). However, peptides with internal arginine residues are less readily analyzed and annotated. In 2007, we reported that the predominant charge state of tryptic peptides and phosphopeptides increased from 2^+ to 3^+ or higher, by the addition of 0.1% *m*-nitrobenzyl alcohol (*m*-NBA) to the mobile phase for LC-MS. This improved peptide sequence assignments by ETD MS/MS and database searching (109). In order to improve the fragmentation efficiency of ETD for doubly charged peptide ions, electron transfer with collisionally activated dissociation (ETcaD) was introduced by Coon's group (110). However, this approach may result in the loss of the phosphate group and complicate phosphopeptide analysis. Wu and colleagues reported a combination of CID with ETD by incorporating an additional CID activation step for a charge-reduced species, isolated from ETD fragment ions (111).

3.3. Advanced MS/MS Methods for Phosphopeptide Analysis

In the mass spectrometric setup, characteristics of modified peptides can be utilized in different ways. Marker ions for phosphorylated peptides have been sought and utilized for method development for years (112–114). Steven Carr's group introduced precursor ion scans for phosphopeptides, combining the features of negative ion mode for detection with the strengths of positive ion mode for subsequent sequencing of the phosphorylated peptides (112). In the initial experiments, m/z 79

(PO_3^-) was the main marker used to selectively detect serine-, threonine-, and tyrosine-phosphorylated peptides (112, 114). Edelson-Averbukh et al. later found that CID of multiply charged anions of phosphorylated peptides produces an additional type of specific marker ions, $[M\text{-}nH\text{-}79]^{(n-1)-}$, which also carried the potential of assisting the determination of the number of phosphorylation sites in a peptide when subsequent neutral losses of H_3PO_4 from these marker ions were counted (113).

While precursor ion scans for analysis of phosphorylated peptides are conducted on triple quadrupole types of instruments, the dissemination of Orbitraps during the past few years has enabled new strategies to be evolved. In 2007, Olsen et al. experimented with an increase of the radiofrequency voltage of the C-trap in an LTQ Orbitrap to produce phosphotyrosine-specific immonium ions, much like in a quadrupole or Q-TOF instrument. This process led to the definition of high-collision dissociation (HCD) and introduction of the LTQ Orbitrap XL, which has a dedicated octopole collision cell attached to the C-trap for HCD fragmentation (115). Analysis of phosphorylated peptides by HCD enables phosphotyrosine immonium ion detection and may also produce clearer fragment ion spectra for phospho-S- and phospho-T-containing peptides due to reduced intensity of the neutral loss peak as it undergoes further collisions in the C-trap (115). Additionally, the ability to detect low-mass fragment ions makes HCD useful for iTRAQ-based quantitative analyses of phosphorylated peptides as is discussed later in this review.

Revelation of more phosphorylation sites that become available in public databases concurrently increases the interest in monitoring specific phosphorylation events in selected proteins of interest in a particular cell type or tissue. Toward this aim, a targeted approach using selected reaction monitoring (SRM) may be more practical and feasible. SRM increases the amount of information that can be obtained from one or more analytes in a complex sample. Highly specific MS/MS scans (precursor-to-product transitions) can be carried out on a single analyte, focusing on a single precursor and product ion or on multiple precursors and product ions from single or multiple analytes (116–118). The analyses are usually carried out by triple quadrupole mass spectrometers where the first quadrupole (Q1) only allows transmission of peptide ions with a predefined mass (precursor ions), which are then fragmented in the quadrupole collision cell (Q2). The third quadrupole (Q3) is 'locked' to selectively transmit a specific, predefined product ion, i.e., a peptide fragment ion that is specific for the peptide or post-translational modification of interest (117, 118). In phosphoproteomics, SRM is not only used for highly confident identifications of phosphorylation sites

in one or several specific proteins (119–123), but also provides the ability to perform quantitative analyses (118, 124). Zappacosta et al. used SRM for stoichiometric measures on selected Pho4 phosphorylation sites in their search for functionally relevant phosphorylations and found that ~50-fold less material was required compared to classical isotope-labeling experiments with LC-MS/MS (124).

4. Database Searching, Annotation, Data Analysis, and Interpretation in Phosphoproteomics

Evaluation of phosphoproteomic data is a multistep process, typically initiated by a protein sequence database search. This step is highly dependent on the status of genome sequencing and annotation of the organism of interest, whether it is available in either a fully validated or a preliminary state. Choosing the appropriate sequence database(s) as well as suitable settings for modifications, enzyme(s), mass error tolerance, etc., is very important. Working with samples derived from well-characterized organisms using established procedures and MS instruments with high mass accuracy eases the effort of data analysis. However, the choice of sequence database may still have significant impact on the result. Lysenko et al. recently published a study of data integration for plant genomics, where they included an example from protein interaction data for *A. thaliana*, showing only limited overlap between results obtained from different databases (125), thus indicating that the use of only a single sequence database may not yield sufficient results. Numerous organisms are poorly annotated at the gene or protein level and may not be adequately represented in the most frequently used sequence databases. Custom databases can be generated, e.g., using translated DNA sequences or proteins derived from related species, in order to increase the chance of obtaining 'true' identifications from the organism in question (126). In such cases it may also be advantageous to search the data against several different sequence databases (127), e.g., one sequence database derived from translated DNA sequences from the organism in question and one from one or more close relative(s) with an annotated genome/proteome, to be able to identify homologous proteins/genes (128).

Once sequence database searching has been performed, the output needs to be further analyzed and validated. This is especially important when working with identification of phosphorylation sites in proteins, as the assignment made by the search engines cannot be blindly trusted (129). Use of more than a single database, or alternative search algorithms, may aid in achieving

higher confidence results (130, 131), and for correct phosphorylation site location, applications like the ambiguity score (Ascore) (132) or Post-Translational Modification (PTM) scoring (68, 133) may be useful. Having completed all the above-mentioned steps, interpretation of the data in a biological context can be initiated. The identified phosphoproteins can be grouped by function, cellular location, or pathways by gene ontology (GO) annotations (134) or one or more suitable freeware tools focusing on protein families, pathways, etc., e.g., (135, 136).

5. Quantitative Phosphoproteome Analysis

Protein phosphorylation is reversible, transient, and therefore highly dynamic. Protein phosphorylation states change significantly over time, for example, not only through the cell cycle (137), but also through life stages of an organism (138) and as a consequence of external perturbations (139) or disease (140). Hence, quantitative phosphoprotein analysis is a prerequisite to fully assess the function(s) and role(s) of specific protein phosphorylation sites as well as any effectors tested in the experimental setup. A range of strategies are available for quantitative phosphoprotein analyses, all depending somewhat on the type of sample to be analyzed, and therefore only a few will be mentioned here. More information on quantitative mass spectrometry analyses for proteomics and phosphoproteomics is available in other reviews (141, 142).

Quantitative analysis by mass spectrometry is based on comparative analysis of peptide ion intensities from series of samples. The so-called label-free approaches rely on reproducible, comparative LC-MS/MS analyses of samples and are applicable to all types of protein samples, whether from cell lines or tissues. The extracted ion current (XIC) method and protein abundance index (PAI) strategies are compatible with LC-MS and LC-MS/MS, respectively (143, 144). Label-free quantitative procedures are becoming more common in proteomics (145), but have yet to be fully evaluated for use in phosphoproteomics, despite allowing for a larger dynamic range of the recorded protein/peptide changes (146). Niittyla et al. combined IMAC and MS^3 to identify 67 unique phosphopeptides from sucrose-induced plasma membrane proteins in *Arabidopsis*, the majority of which were then quantitatively assessed over five time points (147). Hoffert et al. applied IMAC combined with LC-MS^n neutral loss scanning for label-free quantification, identifying 714 phosphorylation sites on 223 unique phosphoproteins in rat renal cells (148). Multiplexed

analysis of two to eight samples is possible by using stable isotope-labeling approaches. Stable isotope labeling by amino acids in cell culture (SILAC) (149) with metabolic incorporation of the label(s) at the protein translational level and isobaric tags for relative and absolute quantification (iTRAQ) (150) with chemical labeling at the peptide level have gained popularity, also for phosphoproteomic analyses (52, 66, 151–153). While SILAC only allows multiplexed comparison of two to three samples, it has the advantage of proteins being biologically labeled during synthesis, and labeled and unlabeled peptides will co-elute during liquid chromatography, thus facilitating subsequent data analysis (66, 149). In contrast, iTRAQ derivatization takes place at the peptide level (i.e., after protease digestion of protein), using amine-reactive isobaric tags (150, 153). Quantification is achieved by measuring the abundance of MS/MS reporter ions from the isobaric tags (150). Labeling at the peptide level enables quantification of samples that may not be easily/at all obtained from cell cultures. We have adapted iTRAQ for proteomic and phosphoproteomic strategies that use SDS-PAGE for protein separation followed by in-gel digestion and iTRAQ labeling (153, 154). Gel-free approaches, where complex iTRAQ-labeled mixtures are analyzed on HCD-capable Orbitrap instruments, are also employed by us and other groups (155–157).

6. Conclusions

Mass spectrometry is a key tool in contemporary phosphoproteomics as it enables detailed analysis of phosphoproteins and annotation and quantification of phosphorylation sites in large-scale studies. However, it is important to realize that mass spectrometry relies on efficient and sensitive 'upstream' sample preparation methods for phosphoprotein and phosphopeptide recovery, including affinity-based methods such as IMAC and TiO_2. The combination of biochemical and immunological methods for sample preparation with quantitative strategies for mass spectrometric sequencing of phosphopeptides has already provided insights into the molecular mechanisms of cellular differentiation, growth, and gene regulation. The efforts required to perform data analysis in large-scale phosphoproteomic experiments should not be underestimated and rely on custom-made computational tools that are used in combination with commercial software packages. Ongoing efforts in many laboratories contribute to the maturation of 'functional phosphoproteomics' in the context of cell biology and molecular medicine.

Acknowledgments

Research in O.N.J.'s laboratory is funded by grants from the Danish Research Councils, The Danish National Research Foundation, and The Lundbeck Foundation. H.R. was funded by a Ph.D. fellowship from the German Research Council (DFG, grant number WI 2044/5-1).

References

1. Fischer, E. H., and Krebs, E. G. (1955) Conversion of phosphorylase b to phosphorylase a in muscle extracts. *J Biol Chem 216*, 121–32.
2. Krebs, E. G., and Fischer, E. H. (1955) Phosphorylase activity of skeletal muscle extracts. *J Biol Chem 216*, 113–20.
3. Larsen, M. R., Trelle, M. B., Thingholm, T. E., and Jensen, O. N. (2006) Analysis of posttranslational modifications of proteins by tandem mass spectrometry. *Biotechniques 40*, 790–8.
4. Preisinger, C., von Kriegsheim, A., Matallanas, D., and Kolch, W. (2008) Proteomics and phosphoproteomics for the mapping of cellular signalling networks. *Proteomics 8*, 4402–15.
5. Thingholm, T. E., Jensen, O. N., and Larsen, M. R. (2009) Analytical strategies for phosphoproteomics. *Proteomics 9*, 1451–68.
6. Hochuli, E., Bannwarth, W., Doebeli, H., Gentz, R., Stueber, D. (1988) Genetic approach to facilitate purification of recombinant proteins with a novel metal chelate affinity chromatography. *Bio/Technology 6*, 1321–5.
7. Hochuli, E., Dobeli, H., and Schacher, A. (1987) New metal chelate adsorbent selective for proteins and peptides containing neighbouring histidine residues. *J Chromatogr 411*, 177–84.
8. Porath, J., Carlsson, J., Olsson, I., and Belfrage, G. (1975) Metal chelate affinity chromatography, a new approach to protein fractionation. *Nature 258*, 598–9.
9. Andersson, L., and Porath, J. (1986) Isolation of phosphoproteins by immobilized metal (Fe^{3+}) affinity chromatography. *Anal Biochem 154*, 250–4.
10. Feng, S., Ye, M., Zhou, H., Jiang, X., Jiang, X., Zou, H., and Gong, B. (2007) Immobilized zirconium ion affinity chromatography for specific enrichment of phosphopeptides in phosphoproteome analysis. *Mol Cell Proteomics 6*, 1656–65.
11. Hollender, D., Karoly-Lakatos, A., Forgo, P., Kortvelyesi, T., Dombi, G., Majer, Z., Hollosi, M., Kiss, T., and Odani, A. (2006) Al(III)-binding ability of an octapeptide and its phosphorylated derivative. *J Inorg Biochem 100*, 351–61.
12. Posewitz, M. C., and Tempst, P. (1999) Immobilized gallium(III) affinity chromatography of phosphopeptides. *Anal Chem 71*, 2883–92.
13. Stensballe, A., Andersen, S., and Jensen, O. N. (2001) Characterization of phosphoproteins from electrophoretic gels by nanoscale Fe(III) affinity chromatography with off-line mass spectrometry analysis. *Proteomics 1*, 207–22.
14. Machida, M., Kosako, H., Shirakabe, K., Kobayashi, M., Ushiyama, M., Inagawa, J., Hirano, J., Nakano, T., Bando, Y., Nishida, E., and Hattori, S. (2007) Purification of phosphoproteins by immobilized metal affinity chromatography and its application to phosphoproteome analysis. *FEBS J 274*, 1576–87.
15. Seeley, E. H., Riggs, L. D., and Regnier, F. E. (2005) Reduction of non-specific binding in Ga(III) immobilized metal affinity chromatography for phosphopeptides by using endoproteinase glu-C as the digestive enzyme. *J Chromatogr B Analyt Technol Biomed Life Sci 817*, 81–8.
16. Barnouin, K. N., Hart, S. R., Thompson, A. J., Okuyama, M., Waterfield, M., and Cramer, R. (2005) Enhanced phosphopeptide isolation by Fe(III)-IMAC using 1,1,1,3,3,3-hexafluoroisopropanol. *Proteomics 5*, 4376–88.
17. Imanishi, S. Y., Kochin, V., and Eriksson, J. E. (2007) Optimization of phosphopeptide elution conditions in immobilized Fe(III) affinity chromatography. *Proteomics 7*, 174–6.
18. Kokubu, M., Ishihama, Y., Sato, T., Nagasu, T., and Oda, Y. (2005) Specificity of immobilized metal affinity-based IMAC/C18 tip

enrichment of phosphopeptides for protein phosphorylation analysis. *Anal Chem 77*, 5144–54.
19. Brill, L. M., Salomon, A. R., Ficarro, S. B., Mukherji, M., Stettler-Gill, M., and Peters, E. C. (2004) Robust phosphoproteomic profiling of tyrosine phosphorylation sites from human T cells using immobilized metal affinity chromatography and tandem mass spectrometry. *Anal Chem 76*, 2763–72.
20. Ficarro, S. B., McCleland, M. L., Stukenberg, P. T., Burke, D. J., Ross, M. M., Shabanowitz, J., Hunt, D. F., and White, F. M. (2002) Phosphoproteome analysis by mass spectrometry and its application to *Saccharomyces cerevisiae*. *Nat Biotechnol 20*, 301–5.
21. Larsen, M. R., Thingholm, T. E., Jensen, O. N., Roepstorff, P., and Jorgensen, T. J. (2005) Highly selective enrichment of phosphorylated peptides from peptide mixtures using titanium dioxide microcolumns. *Mol Cell Proteomics 4*, 873–86.
22. Speicher, K., Kolbas, O., Harper, S., and Speicher, D. (2000) Systematic analysis of peptide recoveries from in-gel digestions for protein identifications in proteome studies. *J Biomol Tech 11*, 74–86.
23. Ye, J., Zhang, X., Young, C., Zhao, X., Hao, Q., Cheng, L., and Jensen, O. N. (2010) *J Proteome Res 9*(7), 3561–73.
24. Ficarro, S. B., Parikh, J. R., Blank, N. C., and Marto, J. A. (2008) Niobium(V) oxide (Nb_2O_5): application to phosphoproteomics. *Anal Chem 80*, 4606–13.
25. Leitner, A., Sturm, M., Smatt, J. H., Jarn, M., Linden, M., Mechtler, K., and Lindner, W. (2009) Optimizing the performance of tin dioxide microspheres for phosphopeptide enrichment. *Anal Chim Acta 638*, 51–7.
26. Pinkse, M. W., Uitto, P. M., Hilhorst, M. J., Ooms, B., and Heck, A. J. (2004) Selective isolation at the femtomole level of phosphopeptides from proteolytic digests using 2D-NanoLC-ESI-MS/MS and titanium oxide precolumns. *Anal Chem 76*, 3935–43.
27. Sano, A., and Nakamura, H. (2007) Evaluation of titanium and titanium oxides as chemo-affinity sorbents for the selective enrichment of organic phosphates. *Anal Sci 23*, 1285–9.
28. Wolschin, F., Wienkoop, S., and Weckwerth, W. (2005) Enrichment of phosphorylated proteins and peptides from complex mixtures using metal oxide/hydroxide affinity chromatography (MOAC). *Proteomics 5*, 4389–97.
29. Zhou, H., Tian, R., Ye, M., Xu, S., Feng, S., Pan, C., Jiang, X., Li, X., and Zou, H. (2007) Highly specific enrichment of phosphopeptides by zirconium dioxide nanoparticles for phosphoproteome analysis. *Electrophoresis 28*, 2201–15.
30. Ikeguchi, Y., and Nakamua, H. (1997) Determination of organic phosphates by column-switching high performance anion-exchange chromatography using on-line preconcentration on titania. *Anal Sci 13*, 479–483.
31. Cantin, G. T., Shock, T. R., Park, S. K., Madhani, H. D., and Yates, J. R., 3rd. (2007) Optimizing TiO_2-based phosphopeptide enrichment for automated multidimensional liquid chromatography coupled to tandem mass spectrometry. *Anal Chem 79*, 4666–73.
32. Yu, L. R., Zhu, Z., Chan, K. C., Issaq, H. J., Dimitrov, D. S., and Veenstra, T. D. (2007) Improved titanium dioxide enrichment of phosphopeptides from HeLa cells and high confident phosphopeptide identification by cross-validation of MS/MS and MS/MS/MS spectra. *J Proteome Res 6*, 4150–62.
33. Simon, E. S., Young, M., Chan, A., Bao, Z. Q., and Andrews, P. C. (2008) Improved enrichment strategies for phosphorylated peptides on titanium dioxide using methyl esterification and pH gradient elution. *Anal Biochem 377*, 234–42.
34. Kweon, H. K., and Hakansson, K. (2006) Selective zirconium dioxide-based enrichment of phosphorylated peptides for mass spectrometric analysis. *Anal Chem 78*, 1743–9.
35. Sugiyama, N., Masuda, T., Shinoda, K., Nakamura, A., Tomita, M., and Ishihama, Y. (2007) Phosphopeptide enrichment by aliphatic hydroxy acid-modified metal oxide chromatography for nano-LC-MS/MS in proteomics applications. *Mol Cell Proteomics 6*, 1103–9.
36. Kyono, Y., Sugiyama, N., Imami, K., Tomita, M., and Ishihama, Y. (2008) Successive and selective release of phosphorylated peptides captured by hydroxy acid-modified metal oxide chromatography. *J Proteome Res 7*, 4585–93.
37. Sugiyama, N., Nakagami, H., Mochida, K., Daudi, A., Tomita, M., Shirasu, K., and Ishihama, Y. (2008) Large-scale phosphorylation mapping reveals the extent of tyrosine phosphorylation in Arabidopsis. *Mol Syst Biol 4*, 193.
38. Li, Y., Leng, T., Lin, H., Deng, C., Xu, X., Yao, N., Yang, P., and Zhang, X. (2007) Preparation of Fe_3O_4@ZrO_2 core-shell microspheres as affinity probes for selec-

tive enrichment and direct determination of phosphopeptides using matrix-assisted laser desorption ionization mass spectrometry. *J Proteome Res 6*, 4498–510.
39. Qi, D., Lu, J., Deng, C., and Zhang, X. (2009) Development of core-shell structure Fe_3O_4@Ta_2O_5 microspheres for selective enrichment of phosphopeptides for mass spectrometry analysis. *J Chromatogr A 1216*, 5533–9.
40. Jensen, S. S., and Larsen, M. R. (2007) Evaluation of the impact of some experimental procedures on different phosphopeptide enrichment techniques. *Rapid Commun Mass Spectrom 21*, 3635–45.
41. Thingholm, T. E., Jorgensen, T. J., Jensen, O. N., and Larsen, M. R. (2006) Highly selective enrichment of phosphorylated peptides using titanium dioxide. *Nat Protoc 1*, 1929–35.
42. Gronborg, M., Kristiansen, T. Z., Stensballe, A., Andersen, J. S., Ohara, O., Mann, M., Jensen, O. N., and Pandey, A. (2002) A mass spectrometry-based proteomic approach for identification of serine/threonine-phosphorylated proteins by enrichment with phospho-specific antibodies: identification of a novel protein, Frigg, as a protein kinase A substrate. *Mol Cell Proteomics 1*, 517–27.
43. Kaufmann, H., Bailey, J. E., and Fussenegger, M. (2001) Use of antibodies for detection of phosphorylated proteins separated by two-dimensional gel electrophoresis. *Proteomics 1*, 194–9.
44. Pandey, A., Fernandez, M. M., Steen, H., Blagoev, B., Nielsen, M. M., Roche, S., Mann, M., and Lodish, H. F. (2000) Identification of a novel immunoreceptor tyrosine-based activation motif-containing molecule, STAM2, by mass spectrometry and its involvement in growth factor and cytokine receptor signaling pathways. *J Biol Chem 275*, 38633–9.
45. Steen, H., Kuster, B., Fernandez, M., Pandey, A., and Mann, M. (2002) Tyrosine phosphorylation mapping of the epidermal growth factor receptor signaling pathway. *J Biol Chem 277*, 1031–9.
46. Collins, M. O., Yu, L., and Choudhary, J. S. (2007) Analysis of protein phosphorylation on a proteome-scale. *Proteomics 7*, 2751–68.
47. Rush, J., Moritz, A., Lee, K. A., Guo, A., Goss, V. L., Spek, E. J., Zhang, H., Zha, X. M., Polakiewicz, R. D., and Comb, M. J. (2005) Immunoaffinity profiling of tyrosine phosphorylation in cancer cells. *Nat Biotechnol 23*, 94–101.
48. Sevecka, M., and MacBeath, G. (2006) State-based discovery: a multidimensional screen for small-molecule modulators of EGF signaling. *Nat Methods 3*, 825–31.
49. Ficarro, S., Chertihin, O., Westbrook, V. A., White, F., Jayes, F., Kalab, P., Marto, J. A., Shabanowitz, J., Herr, J. C., Hunt, D. F., and Visconti, P. E. (2003) Phosphoproteome analysis of capacitated human sperm. Evidence of tyrosine phosphorylation of a kinase-anchoring protein 3 and valosin-containing protein/p97 during capacitation. *J Biol Chem 278*, 11579–89.
50. Zheng, H., Hu, P., Quinn, D. F., and Wang, Y. K. (2005) Phosphotyrosine proteomic study of interferon alpha signaling pathway using a combination of immunoprecipitation and immobilized metal affinity chromatography. *Mol Cell Proteomics 4*, 721–30.
51. Villen, J., Beausoleil, S. A., Gerber, S. A., and Gygi, S. P. (2007) Large-scale phosphorylation analysis of mouse liver. *Proc Natl Acad Sci USA 104*, 1488–93.
52. Zhang, Y., Wolf-Yadlin, A., Ross, P. L., Pappin, D. J., Rush, J., Lauffenburger, D. A., and White, F. M. (2005) Time-resolved mass spectrometry of tyrosine phosphorylation sites in the epidermal growth factor receptor signaling network reveals dynamic modules. *Mol Cell Proteomics 4*, 1240–50.
53. Zhang, Y., Wolf-Yadlin, A., and White, F. M. (2007) Quantitative proteomic analysis of phosphotyrosine-mediated cellular signaling networks. *Methods Mol Biol 359*, 203–12.
54. Peck, S. C., Nuhse, T. S., Hess, D., Iglesias, A., Meins, F., and Boller, T. (2001) Directed proteomics identifies a plant-specific protein rapidly phosphorylated in response to bacterial and fungal elicitors. *Plant Cell 13*, 1467–75.
55. Reynolds, E. C., Riley, P. F., and Adamson, N. J. (1994) A selective precipitation purification procedure for multiple phosphoseryl-containing peptides and methods for their identification. *Anal Biochem 217*, 277–84.
56. Zhang, X., Ye, J., Jensen, O. N., and Roepstorff, P. (2007) Highly efficient phosphopeptide enrichment by calcium phosphate precipitation combined with subsequent IMAC enrichment. *Mol Cell Proteomics 6*, 2032–42.
57. Kametani, F., Nonaka, T., Suzuki, T., Arai, T., Dohmae, N., Akiyama, H., and Hasegawa, M. (2009) Identification of casein kinase-1 phosphorylation sites on TDP-43. *Biochem Biophys Res Commun 382*, 405–9.
58. Xia, Q., Cheng, D., Duong, D. M., Gearing, M., Lah, J. J., Levey, A. I., and Peng, J.

(2008) Phosphoproteomic analysis of human brain by calcium phosphate precipitation and mass spectrometry. *J Proteome Res 7*, 2845–51.
59. Ruse, C. I., McClatchy, D. B., Lu, B., Cociorva, D., Motoyama, A., Park, S. K., and Yates, J. R., 3rd. (2008) Motif-specific sampling of phosphoproteomes. *J Proteome Res 7*, 2140–50.
60. Bodenmiller, B., Mueller, L. N., Mueller, M., Domon, B., and Aebersold, R. (2007) Reproducible isolation of distinct, overlapping segments of the phosphoproteome. *Nat Methods 4*, 231–7.
61. Thingholm, T. E., Jensen, O. N., Robinson, P. J., and Larsen, M. R. (2008) SIMAC (sequential elution from IMAC), a phosphoproteomics strategy for the rapid separation of monophosphorylated from multiply phosphorylated peptides. *Mol Cell Proteomics 7*, 661–71.
62. Nuhse, T. S., Stensballe, A., Jensen, O. N., and Peck, S. C. (2003) Large-scale analysis of in vivo phosphorylated membrane proteins by immobilized metal ion affinity chromatography and mass spectrometry. *Mol Cell Proteomics 2*, 1234–43.
63. Han, G., Ye, M., Zhou, H., Jiang, X., Feng, S., Jiang, X., Tian, R., Wan, D., Zou, H., and Gu, J. (2008) Large-scale phosphoproteome analysis of human liver tissue by enrichment and fractionation of phosphopeptides with strong anion exchange chromatography. *Proteomics 8*, 1346–61.
64. Beausoleil, S. A., Jedrychowski, M., Schwartz, D., Elias, J. E., Villen, J., Li, J., Cohn, M. A., Cantley, L. C., and Gygi, S. P. (2004) Large-scale characterization of HeLa cell nuclear phosphoproteins. *Proc Natl Acad Sci USA 101*, 12130–5.
65. Villen, J., and Gygi, S. P. (2008) The SCX/IMAC enrichment approach for global phosphorylation analysis by mass spectrometry. *Nat Protoc 3*, 1630–8.
66. Gruhler, A., Olsen, J. V., Mohammed, S., Mortensen, P., Faergeman, N. J., Mann, M., and Jensen, O. N. (2005) Quantitative phosphoproteomics applied to the yeast pheromone signaling pathway. *Mol Cell Proteomics 4*, 310–27.
67. Trinidad, J. C., Specht, C. G., Thalhammer, A., Schoepfer, R., and Burlingame, A. L. (2006) Comprehensive identification of phosphorylation sites in postsynaptic density preparations. *Mol Cell Proteomics 5*, 914–22.
68. Olsen, J. V., Blagoev, B., Gnad, F., Macek, B., Kumar, C., Mortensen, P., and Mann, M. (2006) Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. *Cell 127*, 635–48.
69. Wu, J., Shakey, Q., Liu, W., Schuller, A., and Follettie, M. T. (2007) Global profiling of phosphopeptides by titania affinity enrichment. *J Proteome Res 6*, 4684–9.
70. Pinkse, M. W., Mohammed, S., Gouw, J. W., van Breukelen, B., Vos, H. R., and Heck, A. J. (2008) Highly robust, automated, and sensitive online TiO_2-based phosphoproteomics applied to study endogenous phosphorylation in Drosophila melanogaster. *J Proteome Res 7*, 687–97.
71. Dai, J., Jin, W. H., Sheng, Q. H., Shieh, C. H., Wu, J. R., and Zeng, R. (2007) Protein phosphorylation and expression profiling by Yin-yang multidimensional liquid chromatography (Yin-yang MDLC) mass spectrometry. *J Proteome Res 6*, 250–62.
72. Dai, J., Wang, L. S., Wu, Y. B., Sheng, Q. H., Wu, J. R., Shieh, C. H., and Zeng, R. (2009) Fully automatic separation and identification of phosphopeptides by continuous pH-gradient anion exchange online coupled with reversed-phase liquid chromatography mass spectrometry. *J Proteome Res 8*, 133–41.
73. Nashabeh, W., and el Rassi, Z. (1991) Capillary zone electrophoresis of alpha 1-acid glycoprotein fragments from trypsin and endoglycosidase digestions. *J Chromatogr 536*, 31–42.
74. Boersema, P. J., Divecha, N., Heck, A. J., and Mohammed, S. (2007) Evaluation and optimization of ZIC-HILIC-RP as an alternative MudPIT strategy. *J Proteome Res 6*, 937–46.
75. McNulty, D. E., and Annan, R. S. (2008) Hydrophilic interaction chromatography reduces the complexity of the phosphoproteome and improves global phosphopeptide isolation and detection. *Mol Cell Proteomics 7*, 971–80.
76. Alpert, A. J. (2008) Electrostatic repulsion hydrophilic interaction chromatography for isocratic separation of charged solutes and selective isolation of phosphopeptides. *Anal Chem 80*, 62–76.
77. Gan, C. S., Guo, T., Zhang, H., Lim, S. K., and Sze, S. K. (2008) A comparative study of electrostatic repulsion-hydrophilic interaction chromatography (ERLIC) versus SCX-IMAC-based methods for phosphopeptide isolation/enrichment. *J Proteome Res 7*, 4869–77.
78. Boersema, P. J., Mohammed, S., and Heck, A. J. (2009) Phosphopeptide fragmentation and analysis by mass spectrometry. *J Mass Spectrom 44*, 861–78.

79. Roepstorff, P., and Fohlman, J. (1984) Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomed Mass Spectrom 11*, 601.
80. Flora, J. W., and Muddiman, D. C. (2004) Determination of the relative energies of activation for the dissociation of aromatic versus aliphatic phosphopeptides by ESI-FTICR-MS and IRMPD. *J Am Soc Mass Spectrom 15*, 121–7.
81. Reinders, J., Lewandrowski, U., Moebius, J., Wagner, Y., and Sickmann, A. (2004) Challenges in mass spectrometry-based proteomics. *Proteomics 4*, 3686–703.
82. Rinalducci, S., Larsen, M. R., Mohammed, S., and Zolla, L. (2006) Novel protein phosphorylation site identification in spinach stroma membranes by titanium dioxide microcolumns and tandem mass spectrometry. *J Proteome Res 5*, 973–82.
83. Nichols, A. M., and White, F. M. (2009) Manual validation of peptide sequence and sites of tyrosine phosphorylation from MS/MS spectra. *Methods Mol Biol 492*, 143–60.
84. Steen, H., Kuster, B., and Mann, M. (2001) Quadrupole time-of-flight versus triple-quadrupole mass spectrometry for the determination of phosphopeptides by precursor ion scanning. *J Mass Spectrom 36*, 782–90.
85. Villen, J., Beausoleil, S. A., and Gygi, S. P. (2008) Evaluation of the utility of neutral-loss-dependent MS3 strategies in large-scale phosphorylation analysis. *Proteomics 8*, 4444–52.
86. Benschop, J. J., Mohammed, S., O'Flaherty, M., Heck, A. J., Slijper, M., and Menke, F. L. (2007) Quantitative phosphoproteomics of early elicitor signaling in Arabidopsis. *Mol Cell Proteomics 6*, 1198–214.
87. Lemeer, S., Pinkse, M. W., Mohammed, S., van Breukelen, B., den Hertog, J., Slijper, M., and Heck, A. J. (2008) Online automated in vivo zebrafish phosphoproteomics: from large-scale analysis down to a single embryo. *J Proteome Res 7*, 1555–64.
88. Ulintz, P. J., Yocum, A. K., Bodenmiller, B., Aebersold, R., Andrews, P. C., and Nesvizhskii, A. I. (2009) Comparison of MS(2)-only, MSA, and MS(2)/MS(3) methodologies for phosphopeptide identification. *J Proteome Res 8*, 887–99.
89. Schroeder, M. J., Shabanowitz, J., Schwartz, J. C., Hunt, D. F., and Coon, J. J. (2004) A neutral loss activation method for improved phosphopeptide sequence analysis by quadrupole ion trap mass spectrometry. *Anal Chem 76*, 3590–8.
90. Palumbo, A. M., and Reid, G. E. (2008) Evaluation of gas-phase rearrangement and competing fragmentation reactions on protein phosphorylation site assignment using collision induced dissociation-MS/MS and MS3. *Anal Chem 80*, 9735–47.
91. Chi, A., Huttenhower, C., Geer, L. Y., Coon, J. J., Syka, J. E., Bai, D. L., Shabanowitz, J., Burke, D. J., Troyanskaya, O. G., and Hunt, D. F. (2007) Analysis of phosphorylation sites on proteins from *Saccharomyces cerevisiae* by electron transfer dissociation (ETD) mass spectrometry. *Proc Natl Acad Sci USA 104*, 2193–8.
92. Hogan, J. M., Pitteri, S. J., and McLuckey, S. A. (2003) Phosphorylation site identification via ion trap tandem mass spectrometry of whole protein and peptide ions: bovine alpha-crystallin A chain. *Anal Chem 75*, 6509–16.
93. Zubarev, R. A., Kelleher, N. L., and McLafferty, F. W. (1998) Electron capture dissociation of multiply charged protein cations. A nonergodic process. *J Am Chem Soc 120*, 3265–3266.
94. Hakansson, K., Cooper, H. J., Emmett, M. R., Costello, C. E., Marshall, A. G., and Nilsson, C. L. (2001) Electron capture dissociation and infrared multiphoton dissociation MS/MS of an N-glycosylated tryptic peptic to yield complementary sequence information. *Anal Chem 73*, 4530–6.
95. Mirgorodskaya, E., Roepstorff, P., and Zubarev, R. A. (1999) Localization of O-glycosylation sites in peptides by electron capture dissociation in a Fourier transform mass spectrometer. *Anal Chem 71*, 4431–6.
96. Shi, S. D., Hemling, M. E., Carr, S. A., Horn, D. M., Lindh, I., and McLafferty, F. W. (2001) Phosphopeptide/phosphoprotein mapping by electron capture dissociation mass spectrometry. *Anal Chem 73*, 19–22.
97. Stensballe, A., Jensen, O. N., Olsen, J. V., Haselmann, K. F., and Zubarev, R. A. (2000) Electron capture dissociation of singly and multiply phosphorylated peptides. *Rapid Commun Mass Spectrom 14*, 1793–800.
98. Voinov, V. G., Deinzer, M. L., and Barofsky, D. F. (2009) Radio-frequency-free cell for electron capture dissociation in tandem mass spectrometry. *Anal Chem 81*, 1238–43.
99. Reid, G. E., Wells, J. M., Badman, E. R., and McLuckey, S. A. (2003) Performance of a quadrupole ion trap mass spectrometer adapted for ion/ion reaction studies. *Int J Mass Spectrom 222*, 243–258.
100. Good, D. M., Wirtala, M., McAlister, G. C., and Coon, J. J. (2007) Performance

characteristics of electron transfer dissociation mass spectrometry. *Mol Cell Proteomics 6*, 1942–51.
101. Xia, Y., Chrisman, P. A., Erickson, D. E., Liu, J., Liang, X., Londry, F. A., Yang, M. J., and McLuckey, S. A. (2006) Implementation of ion/ion reactions in a quadrupole/time-of-flight tandem mass spectrometer. *Anal Chem 78*, 4146–54.
102. Kleinnijenhuis, A. J., Kjeldsen, F., Kallipolitis, B., Haselmann, K. F., and Jensen, O. N. (2007) Analysis of histidine phosphorylation using tandem MS and ion-electron reactions. *Anal Chem 79*, 7450–6.
103. Molina, H., Horn, D. M., Tang, N., Mathivanan, S., and Pandey, A. (2007) Global proteomic profiling of phosphopeptides using electron transfer dissociation tandem mass spectrometry. *Proc Natl Acad Sci USA 104*, 2199–204.
104. Swaney, D. L., McAlister, G. C., and Coon, J. J. (2008) Decision tree-driven tandem mass spectrometry for shotgun proteomics. *Nat Methods 5*, 959–64.
105. Lu, H., Zong, C., Wang, Y., Young, G. W., Deng, N., Souda, P., Li, X., Whitelegge, J., Drews, O., Yang, P. Y., and Ping, P. (2008) Revealing the dynamics of the 20 S proteasome phosphoproteome: a combined CID and electron transfer dissociation approach. *Mol Cell Proteomics 7*, 2073–89.
106. Hennrich, M. L., Boersema, P. J., van den Toorn, H., Mischerikow, N., Heck, A. J., and Mohammed, S. (2009) Effect of chemical modifications on peptide fragmentation behavior upon electron transfer induced dissociation. *Anal Chem 81*, 7814–22.
107. Taouatas, N., Altelaar, A. F., Drugan, M. M., Helbig, A. O., Mohammed, S., and Heck, A. J. (2009) Strong cation exchange-based fractionation of Lys-N-generated peptides facilitates the targeted analysis of post-translational modifications. *Mol Cell Proteomics 8*, 190–200.
108. Taouatas, N., Drugan, M. M., Heck, A. J., and Mohammed, S. (2008) Straightforward ladder sequencing of peptides using a Lys-N metalloendopeptidase. *Nat Methods 5*, 405–7.
109. Kjeldsen, F., Giessing, A. M., Ingrell, C. R., and Jensen, O. N. (2007) Peptide sequencing and characterization of post-translational modifications by enhanced ion-charging and liquid chromatography electron-transfer dissociation tandem mass spectrometry. *Anal Chem 79*, 9243–52.
110. Swaney, D. L., McAlister, G. C., Wirtala, M., Schwartz, J. C., Syka, J. E., and Coon, J. J. (2007) Supplemental activation method for high-efficiency electron-transfer dissociation of doubly protonated peptide precursors. *Anal Chem 79*, 477–85.
111. Wu, S. L., Huhmer, A. F., Hao, Z., and Karger, B. L. (2007) On-line LC-MS approach combining collision-induced dissociation (CID), electron-transfer dissociation (ETD), and CID of an isolated charge-reduced species for the trace-level characterization of proteins with post-translational modifications. *J Proteome Res 6*, 4230–44.
112. Carr, S. A., Huddleston, M. J., and Annan, R. S. (1996) Selective detection and sequencing of phosphopeptides at the femtomole level by mass spectrometry. *Anal Biochem 239*, 180–92.
113. Edelson-Averbukh, M., Pipkorn, R., and Lehmann, W. D. (2006) Phosphate group-driven fragmentation of multiply charged phosphopeptide anions. Improved recognition of peptides phosphorylated at serine, threonine, or tyrosine by negative ion electrospray tandem mass spectrometry. *Anal Chem 78*, 1249–56.
114. Huddleston, M. J., Annan, R. S., Bean, M. F., and Carr, S. A. (1993) Selective detection of phosphopeptides in complex-mixtures by electrospray liquid-chromatography mass-spectrometry. *J Am Soc Mass Spectrom 4*, 710–717.
115. Olsen, J. V., Macek, B., Lange, O., Makarov, A., Horning, S., and Mann, M. (2007) Higher-energy C-trap dissociation for peptide modification analysis. *Nat Methods 4*, 709–12.
116. Blackburn, K., and Goshe, M. B. (2009) Challenges and strategies for targeted phosphorylation site identification and quantification using mass spectrometry analysis. *Brief Funct Genomic Proteomic 8*, 90–103.
117. Unwin, R. D., Griffiths, J. R., Leverentz, M. K., Grallert, A., Hagan, I. M., and Whetton, A. D. (2005) Multiple reaction monitoring to identify sites of protein phosphorylation with high sensitivity. *Mol Cell Proteomics 4*, 1134–44.
118. Yocum, A. K., and Chinnaiyan, A. M. (2009) Current affairs in quantitative targeted proteomics: multiple reaction monitoring-mass spectrometry. *Brief Funct Genomic Proteomic 8*, 145–57.
119. Cirulli, C., Chiappetta, G., Marino, G., Mauri, P., and Amoresano, A. (2008) Identification of free phosphopeptides in different biological fluids by a mass spectrometry approach. *Anal Bioanal Chem 392*, 147–59.
120. Cox, D. M., Zhong, F., Du, M., Duchoslav, E., Sakuma, T., and McDermott, J. C.

(2005) Multiple reaction monitoring as a method for identifying protein posttranslational modifications. *J Biomol Tech 16*, 83–90.
121. Glinski, M., and Weckwerth, W. (2005) Differential multisite phosphorylation of the trehalose-6-phosphate synthase gene family in Arabidopsis thaliana: a mass spectrometry-based process for multiparallel peptide library phosphorylation analysis. *Mol Cell Proteomics 4*, 1614–25.
122. Johnson, R. P., El-Yazbi, A. F., Hughes, M. F., Schriemer, D. C., Walsh, E. J., Walsh, M. P., and Cole, W. C. (2009) Identification and functional characterization of protein kinase A-catalyzed phosphorylation of potassium channel Kv1.2 at serine 449. *J Biol Chem 284*, 16562–74.
123. Wolf-Yadlin, A., Hautaniemi, S., Lauffenburger, D. A., and White, F. M. (2007) Multiple reaction monitoring for robust quantitative proteomic analysis of cellular signaling networks. *Proc Natl Acad Sci USA 104*, 5860–5.
124. Zappacosta, F., Collingwood, T. S., Huddleston, M. J., and Annan, R. S. (2006) A quantitative results-driven approach to analyzing multisite protein phosphorylation: the phosphate-dependent phosphorylation profile of the transcription factor Pho4. *Mol Cell Proteomics 5*, 2019–30.
125. Lysenko, A., Hindle, M. M., Taubert, J., Saqi, M., and Rawlings, C. J. (2009) Data integration for plant genomics–exemplars from the integration of Arabidopsis thaliana databases. *Brief Bioinform 10*, 676–93.
126. Lucker, J., Laszczak, M., Smith, D., and Lund, S. T. (2009) Generation of a predicted protein database from EST data and application to iTRAQ analyses in grape (Vitis vinifera cv. Cabernet Sauvignon) berries at ripening initiation. *BMC Genomics 10*, 50.
127. Ishii, A., Dutta, R., Wark, G. M., Hwang, S. I., Han, D. K., Trapp, B. D., Pfeiffer, S. E., and Bansal, R. (2009) Human myelin proteome and comparative analysis with mouse myelin. *Proc Natl Acad Sci USA 106*, 14605–10.
128. Vissers, J. P., Pons, S., Hulin, A., Tissier, R., Berdeaux, A., Connolly, J. B., Langridge, J. I., Geromanos, S. J., and Ghaleb, B. (2009) The use of proteome similarity for the qualitative and quantitative profiling of reperfused myocardium. *J Chromatogr B Analyt Technol Biomed Life Sci 877*, 1317–26.
129. Chen, Y., Zhang, J., Xing, G., and Zhao, Y. (2009) Mascot-derived false positive peptide identifications revealed by manual analysis of tandem mass spectra. *J Proteome Res 8*, 3141–7.
130. Alves, G., Wu, W. W., Wang, G., Shen, R. F., and Yu, Y. K. (2008) Enhancing peptide identification confidence by combining search methods. *J Proteome Res 7*, 3102–13.
131. Xu, H., Wang, L., Sallans, L., and Freitas, M. A. (2009) A hierarchical MS2/MS3 database search algorithm for automated analysis of phosphopeptide tandem mass spectra. *Proteomics 9*, 1763–70.
132. Beausoleil, S. A., Villen, J., Gerber, S. A., Rush, J., and Gygi, S. P. (2006) A probability-based approach for high-throughput protein phosphorylation analysis and site localization. *Nat Biotechnol 24*, 1285–92.
133. Mortensen, P., Gouw, J. W., Olsen, J. V., Ong, S. E., Rigbolt, K. T., Bunkenborg, J., Cox, J., Foster, L. J., Heck, A. J., Blagoev, B., Andersen, J. S., and Mann, M. (2010) MSQuant, an open source platform for mass spectrometry-based quantitative proteomics. *J Proteome Res 9*, 393–403.
134. Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., and Sherlock, G. (2000) Gene ontology: tool for the unification of biology. The gene ontology consortium. *Nat Genet 25*, 25–9.
135. *KEGG Pathway Database* (http://www.genome.jp/kegg/pathway.html), Kanehisha Laboratories.
136. *Pfam* (http://pfam.sanger.ac.uk/), The Welcome Trust Sanger Institute.
137. Huang, C. Y., Chang, C. P., Huang, C. L., and Ferrell, J. E., Jr. (1999) M phase phosphorylation of cytoplasmic dynein intermediate chain and p150(Glued). *J Biol Chem 274*, 14262–9.
138. Morales, M. A., Watanabe, R., Laurent, C., Lenormand, P., Rousselle, J. C., Namane, A., and Spath, G. F. (2008) Phosphoproteomic analysis of Leishmania donovani pro- and amastigote stages. *Proteomics 8*, 350–63.
139. El Benna, J., Han, J., Park, J. W., Schmid, E., Ulevitch, R. J., and Babior, B. M. (1996) Activation of p38 in stimulated human neutrophils: phosphorylation of the oxidase component p47phox by p38 and ERK but not by JNK. *Arch Biochem Biophys 334*, 395–400.
140. Lee, H. J., Na, K., Kwon, M. S., Kim, H., Kim, K. S., and Paik, Y. K. (2009)

Quantitative analysis of phosphopeptides in search of the disease biomarker from the hepatocellular carcinoma specimen. *Proteomics 9*, 3395–408.
141. Bantscheff, M., Schirle, M., Sweetman, G., Rick, J., and Kuster, B. (2007) Quantitative mass spectrometry in proteomics: a critical review. *Anal Bioanal Chem 389*, 1017–31.
142. Tedford, N. C., Hall, A. B., Graham, J. R., Murphy, C. E., Gordon, N. F., and Radding, J. A. (2009) Quantitative analysis of cell signaling and drug action via mass spectrometry-based systems level phosphoproteomics. *Proteomics 9*, 1469–87.
143. Ono, M., Shitashige, M., Honda, K., Isobe, T., Kuwabara, H., Matsuzuki, H., Hirohashi, S., and Yamada, T. (2006) Label-free quantitative proteomics using large peptide data sets generated by nanoflow liquid chromatography and mass spectrometry. *Mol Cell Proteomics 5*, 1338–47.
144. Rappsilber, J., Ryder, U., Lamond, A. I., and Mann, M. (2002) Large-scale proteomic analysis of the human spliceosome. *Genome Res 12*, 1231–45.
145. Zhu, W., Smith, J. W., and Huang, C. M. (2010) Mass spectrometry-based label-free quantitative proteomics. *J Biomed Biotechnol 2010, 840518*, 6. doi:10.1155/2010/840518.
146. Asara, J. M., Christofk, H. R., Freimark, L. M., and Cantley, L. C. (2008) A label-free quantification method by MS/MS TIC compared to SILAC and spectral counting in a proteomics screen. *Proteomics 8*, 994–9.
147. Niittyla, T., Fuglsang, A. T., Palmgren, M. G., Frommer, W. B., and Schulze, W. X. (2007) Temporal analysis of sucrose-induced phosphorylation changes in plasma membrane proteins of Arabidopsis. *Mol Cell Proteomics 6*, 1711–26.
148. Hoffert, J. D., Pisitkun, T., Wang, G., Shen, R. F., and Knepper, M. A. (2006) Quantitative phosphoproteomics of vasopressin-sensitive renal cells: regulation of aquaporin-2 phosphorylation at two sites. *Proc Natl Acad Sci USA 103*, 7159–64.
149. Ong, S. E., Blagoev, B., Kratchmarova, I., Kristensen, D. B., Steen, H., Pandey, A., and Mann, M. (2002) Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol Cell Proteomics 1*, 376–86.
150. Ross, P. L., Huang, Y. N., Marchese, J. N., Williamson, B., Parker, K., Hattan, S., Khainovski, N., Pillai, S., Dey, S., Daniels, S., Purkayastha, S., Juhasz, P., Martin, S., Bartlet-Jones, M., He, F., Jacobson, A., and Pappin, D. J. (2004) Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. *Mol Cell Proteomics 3*, 1154–69.
151. Bantscheff, M., Eberhard, D., Abraham, Y., Bastuck, S., Boesche, M., Hobson, S., Mathieson, T., Perrin, J., Raida, M., Rau, C., Reader, V., Sweetman, G., Bauer, A., Bouwmeester, T., Hopf, C., Kruse, U., Neubauer, G., Ramsden, N., Rick, J., Kuster, B., and Drewes, G. (2007) Quantitative chemical proteomics reveals mechanisms of action of clinical ABL kinase inhibitors. *Nat Biotechnol 25*, 1035–44.
152. Boja, E. S., Phillips, D., French, S. A., Harris, R. A., and Balaban, R. S. (2009) Quantitative mitochondrial phosphoproteomics using iTRAQ on an LTQ-Orbitrap with high energy collision dissociation. *J Proteome Res 8*, 4665–75.
153. Sachon, E., Mohammed, S., Bache, N., and Jensen, O. N. (2006) Phosphopeptide quantitation using amine-reactive isobaric tagging reagents and tandem mass spectrometry: application to proteins isolated by gel electrophoresis. *Rapid Commun Mass Spectrom 20*, 1127–34.
154. Manteca, A., Sanchez, J., Jung, H. R., Schwämmle, V., and Jensen, O. N. (2010) Quantitative proteomic analysis of *Streptomyces coelicolor* development demonstrates that onset of secondary metabolism coincides with hyphae differentiation. *Mol Cell Proteomics 9*, 1423–36.
155. Kocher, T., Pichler, P., Schutzbier, M., Stingl, C., Kaul, A., Teucher, N., Hasenfuss, G., Penninger, J. M., and Mechtler, K. (2009) High precision quantitative proteomics using iTRAQ on an LTQ Orbitrap: a new mass spectrometric method combining the benefits of all. *J Proteome Res 8*, 4743–52.
156. Rewitz, K. F., Larsen, M. R., Lobner-Olesen, A., Rybczynski, R., O'Connor, M. B., and Gilbert, L. I. (2009) A phosphoproteomics approach to elucidate neuropeptide signal transduction controlling insect metamorphosis. *Insect Biochem Mol Biol 39*, 475–83.
157. Zhang, Y., Ficarro, S. B., Li, S., and Marto, J. A. (2009) Optimized orbitrap HCD for quantitative analysis of phosphopeptides. *J Am Soc Mass Spectrom 20*, 1425–34.
158. Maguire, P. B., Wynne, K. J., Harney, D. F., O'Donoghue, N. M., Stephens, G., and Fitzgerald, D. J. (2002) Identification of the phosphotyrosine proteome from thrombin activated platelets. *Proteomics 2*, 642–8.

159. Li, Q. R., Ning, Z. B., Tang, J. S., Nie, S., and Zeng, R. (2009) Effect of peptide-to-TiO_2 beads ratio on phosphopeptide enrichment selectivity. *J Proteome Res 8*, 5375–81.

160. Saleem, R. A., and Aitchison, J. D. (2009) Quantitative phosphoproteomics in fatty acid stimulated *Saccharomyces cerevisiae*. *J Vis Exp 32*, e1474. doi:10.3791/1474.

Chapter 14

A Protocol on the Use of Titanium Dioxide Chromatography for Phosphoproteomics

Martijn W.H. Pinkse, Simone Lemeer, and Albert J.R. Heck

Abstract

Over the past decade phosphoproteomics has become an emerging discipline within proteomics research, focusing on detection of the reversible modification of proteins by phosphorylation of serine, threonine, and tyrosine residues. For successful analysis, phosphopeptide enrichment is often a prerequisite due to their low stoichiometry, heterogeneity, and low abundance. The enrichment of phosphopeptides is often performed manually, which is inherently labor intensive and a major hindrance in large-scale analyses. Automation of the enrichment method would vastly improve reproducibility and thereby facilitate "high-throughput" phosphoproteomics research. Here, we describe the setup of a simple, robust, and automated online TiO_2-based nanoscale chromatographic approach to selectively enrich and separate phosphorylated peptides from proteolytic digests of moderate and high complexity.

Key words: Phosphopeptides, online enrichment, titanium dioxide, 2D chromatography, nanoscale liquid chromatography.

1. Introduction

Reversible protein phosphorylation is a ubiquitous mechanism for the control in signaling networks that regulate diverse biological processes. Due to its fundamental role in biology, the identification of phosphorylation events in different biological contexts is a growing area of research (1, 2). Liquid chromatography coupled to shotgun-based tandem mass spectrometry (LC-MS/MS) has become the predominant means of identifying phosphorylation sites in simple or complex protein mixtures (3). However, phosphorylation is often sub-stoichiometric, and an enrichment procedure of phosphorylated peptides, derived from phosphorylated

K. Gevaert, J. Vandekerckhove (eds.), *Gel-Free Proteomics*, Methods in Molecular Biology 753,
DOI 10.1007/978-1-61779-148-2_14,

proteins, is a necessary prerequisite for the characterization of such modified peptides by modern mass spectrometric methods (4). Several approaches have been developed that can be used to isolate and/or enrich phosphorylated peptides and proteins, all with their own advantages and disadvantages (5). To make phosphoproteomics amendable to non-specialists, the development of routine automated enrichment protocols is highly desirable. Liquid chromatography in combination with electrospray ionization offers the possibility to do online mass spectrometric analysis. Nanoscale liquid chromatography is often the preferred analytical strategy in proteomics due to the resolution and sensitivity it offers. We have developed an automated online TiO_2-based liquid chromatography approach that is relatively simple and does not require complex plumbing and column switching schemes. The approach is based on the absorption of phosphorylated peptides onto the surface of spherical particles of titanium dioxide packed in a small capillary column. Absorption is done under acidic conditions at very low flow rates with high efficiency. Elution is performed under alkaline conditions using a special phosphopeptide elution buffer. The method is configured in a fully automated manner and operates with minimal manual interference, offering high throughput and high sensitivity (6–9).

2. Materials

2.1. Preparing Trap-Columns and Separation Columns

1. A capillary column packing high-pressure vessel operating at 50 bar helium pressure.
2. Column packing material for trap-columns (*see* **Note 1**): Aqua C18, 5 μm, 200 Å (Phenomenex), and Titansphere TiO_2, 10 μm (GL Sciences).
3. Column packing material for separation columns: Reprosil-Pur C18-AQ, 3 μm, 120 Å (Dr. Maisch, GmbH).
4. Potassium silicate (Kasil® 1624, PQ Europe) and formamide (*see* **Note 2**).
5. Undeactivated fused silica of 50 and 100 μm internal diameter and 375 μm outer diameter (*see* **Note 3**).
6. Two MicroTight® microtees and two MicroTight® True ZDV unions (Upchurch Scientific) and compatible tubing sleeves to connect 375 μm OD fused silica.
7. 0.3 mm ID and 1/16 in. OD Teflon sleeves for butt-connecting 375 μm OD fused silica capillaries.

2.2. HPLC System and Solvents

1. A binary HPLC pump module and autosampler with a 30 μl sample loop and a six-port switching valve with programmable electronic actuator (*see* **Note 4**).
2. HPLC solvent A: 0.1 M acetic acid and 0.46 M formic acid in Milli-Q water.
3. HPLC solvent B: 0.1 M acetic acid and 0.46 M formic acid in 8/2 (v/v) acetonitrile/Milli-Q water.
4. High-purity (>99%) ammonium bicarbonate, sodium phosphate, sodium orthovanadate, and potassium fluoride.
5. Phosphopeptide elution buffer: 250 mM ammonium bicarbonate, pH 9.0 (adjusted with ammonia), 10 mM sodium phosphate, 1 mM potassium fluoride, and 5 mM sodium orthovanadate.

2.3. Sample Preparation

1. Lysis buffer: 8 M urea in 25 mM ammonium bicarbonate, 5 mM sodium phosphate, 1 mM potassium fluoride, and 1 mM sodium orthovanadate, pH 8.0. Freshly prepare before use. Possible phosphatase activity in samples should be minimized as much as possible by the use of phosphatase inhibitors.
2. Reduction buffer: 100 mM dithiothreitol in 25 mM ammonium bicarbonate, pH 8.0, and alkylation buffer – 200 mM iodoacetamide in 25 mM ammonium bicarbonate, pH 8.0. Both solutions are prepared freshly prior to use (*see* **Note 5**).
3. Trypsin and endoproteinase Lys-C, both sequencing grade (Roche Diagnostics) and dissolved in 1 mM HCl at a concentration of 1 μg/μl (store at –20°C).
4. For offline desalting and concentration of peptide use Eppendorf GELoader pipette tip and a 3 M Empore C18 extraction disk as previously described (10).

3. Methods

Titanium dioxide has a very high affinity for phosphorylated peptides under acidic conditions, but the absorption of phosphorylated peptides onto TiO_2 is a slow process. In typical offline TiO_2 enrichment strategies, the low flow required for absorption of phosphorylated peptides onto the TiO_2 surface either is tedious to control manually or requires additional long sample handling time (8, 11). Here we describe a method for the efficient binding of phosphorylated peptides onto TiO_2 at a very slow and controllable flow. This occurs not at the cost of long sample

handling times and is performed in an automated manner that allows for unattended operation. The method uses a triple trap-column, consisting of a small TiO_2 trap-column "sandwiched" in between two C18 trap-columns in a nanoscale vented column system (12). In the first stage, loading of the peptide sample occurs on the first C18 trap-column, at a regular loading flow of several microliters per minute, which is subsequently switched in line with an analytical nanoscale column. In this stage, the gradient flow through trap and analytical columns is much lower and phosphorylated peptides are loaded onto the TiO_2 trap-column at a flow of ~100 nl/min, while non-phosphorylated peptides pass the TiO_2 trap and are separated and detected within this first gradient. Due to this very low loading speed, phosphorylated peptides are highly efficiently trapped, with almost no breakthrough. Eventually, phosphorylated peptides are eluted from the small TiO_2 trap-column, by means of an injection plug of an alkaline buffer with additives. When desorbing from the TiO_2 material, phosphopeptides are trapped again on the second C18 trap-column and finally they are separated and detected by running a second water/acetonitrile gradient.

3.1. Preparation of Frits

1. To prepare fused silica capillaries with porous ceramic frits, prepare a mixture of 75% potassium silicate and 25% formamide (typically 200 μl) in an empty 2 ml glass sample vial and mix rapidly.
2. Place bundles of 10–15 undeactivated fused silica capillaries in this solution for a few seconds. For trapping columns, prepare porous frits in capillaries of 100 μm ID and 10 cm length. For separation columns, prepare porous frits in capillaries of 50 μm ID and 30–40 cm length.
3. Use a tissue to remove the liquid on the outside of the capillaries. Hold the capillaries vertically at all times, with the plug of liquid facing down to avoid that it transcends to the middle of the capillary. Place the capillaries inside an oven pre-heated to 100°C for 1 h.
4. Cut down the frit of the capillary to 1 or 2 mm prior to use by using a fused silica tubing cutter. Ensure that the edges of the fritted capillary are cut clean (i.e., with an angle of 90°) on both sides by inspecting it under a microscope.

3.2. Preparing Capillary Columns

This procedure requires the use of a high-pressure column packing vessel connected to a helium gas cylinder. Prior to using such a setup, be sure to understand the potential hazards of the column packing vessel and the high-pressure helium gas cylinder and apply correct manual cylinder handling.

1. Connect the capillary to an open piece of fused silica of ~40 cm length that is attached to the column packer. Flush

the empty frit with 2-propanol to ensure that the frit is open.

2. Prepare a slurry of 1 mg/ml TiO_2 or C18 material in 2-propanol in a 2 ml glass vial. Briefly sonicate (for a few seconds) the slurry to assure that the slurry is homogeneous. Additionally, use a micro-stir bar and a magnetic stirrer to maintain the slurry homogeneous during packing.
3. Pack the TiO_2 trap-column to a length of 5 mm. Pack C18 trap-columns to a length of 3 cm (*see* **Note 6**).
4. For packing of separations columns, pre-fill a fritted 100 μm ID capillary of 40 cm length with 5–8 cm of packing material. Then, connect the open end of this capillary to the open end of a fritted 50 μm ID capillary using a Teflon sleeve (1/16 in. OD, 0.3 mm ID, and 2 cm length, *see* **Note 7**). Backflush the column material into the 50 μm ID capillary using the high-pressure vessel to create a ~20 cm separation column.

3.3. Assembling the "Sandwich" Trap-Column and the Vented Column Setup

1. Prior to assembling and using the three trap-columns, wash each individual trap-column five times by injecting 30 μl of phosphopeptide elution buffer and subsequent flushing for 10 min with solvent B, all at a flow rate of 5 μl/min.
2. Connect a 3 cm C18 trap-column to a 5 mm TiO_2 trap-column using a single 0.4 mm ID and 1.7 mm OD Teflon sleeve inside a MicroTight® True ZDV union. Slide the fitting of the union over the Teflon sleeve until the ferrule is positioned in the middle. Insert the C18 column and make sure that about 0.5 mm of the fused silica passes the tip of the ferrule (this is best seen by holding the Teflon sleeve in front of a bright light source). Slide the union from the other side over the Teflon sleeve and tighten it with the fitting. Next, insert the TiO_2 trap-column into the other side of the Teflon sleeve, push it down tightly against the other trap-column, position the fitting, and tighten it in the union. Finally, check if both columns are correctly tightened by gently pulling on both pieces of fused silica that are sticking outside the Teflon sleeve. On the other side of the TiO_2 trap-column, the second C18 trap-column is mounted in the same way as described above. Make sure the flow direction of all trap-columns is identical.

The schematics of a vented column setup are given in **Fig. 14.1** and more detailed guidelines can be found in Meiring et al. (12). The triple trap-column is connected in between two PEEK microtees. A fused silica tubing (typically 30 cm) is connected from the injector valve of the autosampler to the first microtee. From this microtee, a fused silica capillary, acting as a

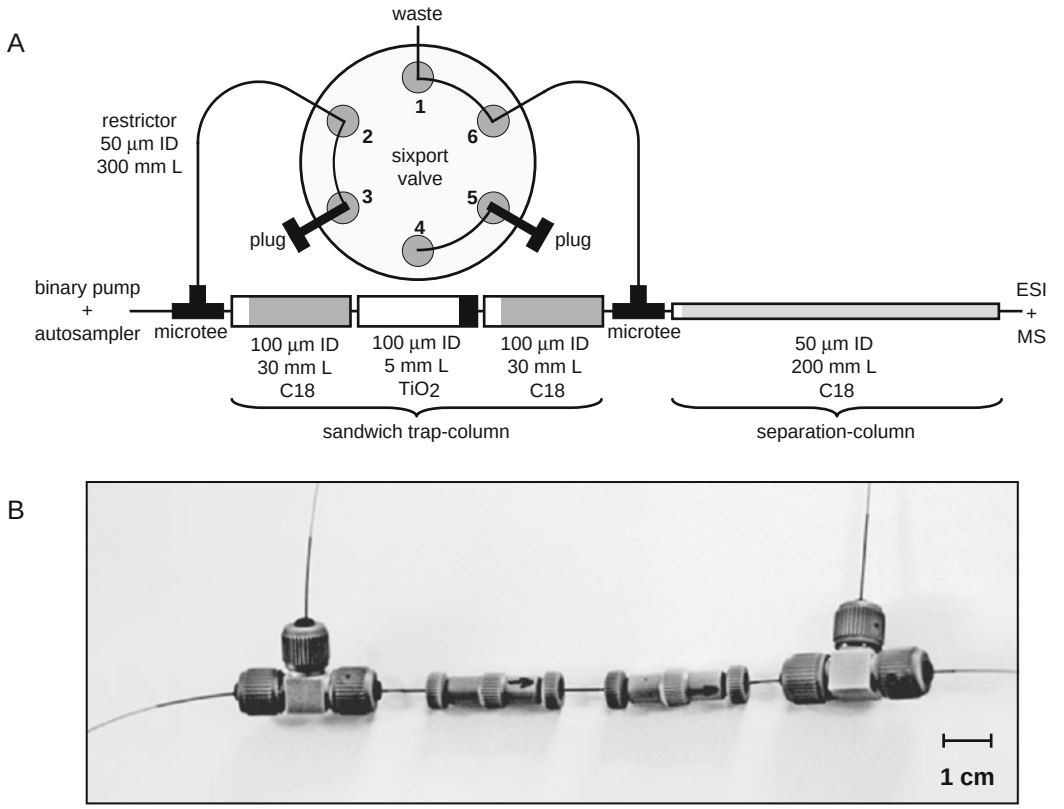


Fig. 14.1. (**a**) Schematic representation of the vented column system with a two-phase sandwich trap-column. (**b**) Picture of the assembled sandwich trap-column.

restrictor (50 µm ID and 30 cm length), is connected to port 2 of the six-port switching valve. The separation column is directly connected with the second microtee. In addition, a 50 µm ID capillary of 20 cm length is connected from this second microtee to port 6 of the switching valve. A waste line is connected to port 1 of the switching valve and ports 3 and 5 are blocked using a blind nut capable of resisting high pressures with no leakage.

3.4. HPLC Program

The operation of the HPLC valve described below is summarized in **Table 14.1** and schematically depicted in **Fig. 14.2**. Prior to loading real samples onto the 2D column system, run several blanks and ensure that the system is clean.

1. Program the six-port switching valve in the 1:6 position, connecting the fluid line coming from the second microtee to the waste line and blocking the line from the first microtee to the switching valve. Set the flow at 5 µl/min and inject the sample into the system using the autosampler.
2. After 10 min, program the switching valve such that it is switched to the 1:2 position and simultaneously increase

Table 14.1
Time program for phosphopeptide enrichment

Time (min)	Pump speed (μl/min)	% Solvent A	Six port	Description
0–10	5	100	1:6	Injection sample + loading
10–35	350	100–60	1:2	Linear H_2O/MeCN gradient
36–40	350	0	1:2	High MeCN wash
40–50	350	100	1:2	Reconditioning C18 material
50–70	3	100	1:6	Injection 30 μl elution buffer
70–80	5	100	1:6	Injection 10 μl 5% formic acid
80–115	350	100–60	1:2	Linear H_2O/MeCN gradient
115–120	350	0	1:6	High MeCN wash
120–130	350	100	1:2	Reconditioning C18 material
130	5	100	1:6	Returning to start conditions

the flow from the binary pump from 5 to 350 μl/min. A pressure of about 100–150 bar is reached and this will result in a flow over the separation column of about 100–150 nl/min. Check that the flow of the separation column is indeed around 100 nl/min (*see* **Note 8**).

3. Program a linear gradient of 0.5–1.5% of solvent B increase per minute. Phosphorylated peptides and non-phosphorylated peptides previously trapped onto the first C18 trap-column will be released at higher acetonitrile concentrations and travel through the TiO_2 trap-column at 100 nl/min. In this manner, phosphorylated peptides travel at a reduced speed through the TiO_2 trapping column and are given enough time to absorb to the TiO_2 material. Simultaneously, all non-phosphorylated peptides pass through the TiO_2 trapping column and are subsequently separated on the separation column and detected by the mass spectrometer.
4. Re-condition both trap and separations column by flushing with 100% of solvent A. Afterwards the pump speed must be reduced to 5 μl/min and the switching valve must be programmed to switch back to the 1:6 position. Using

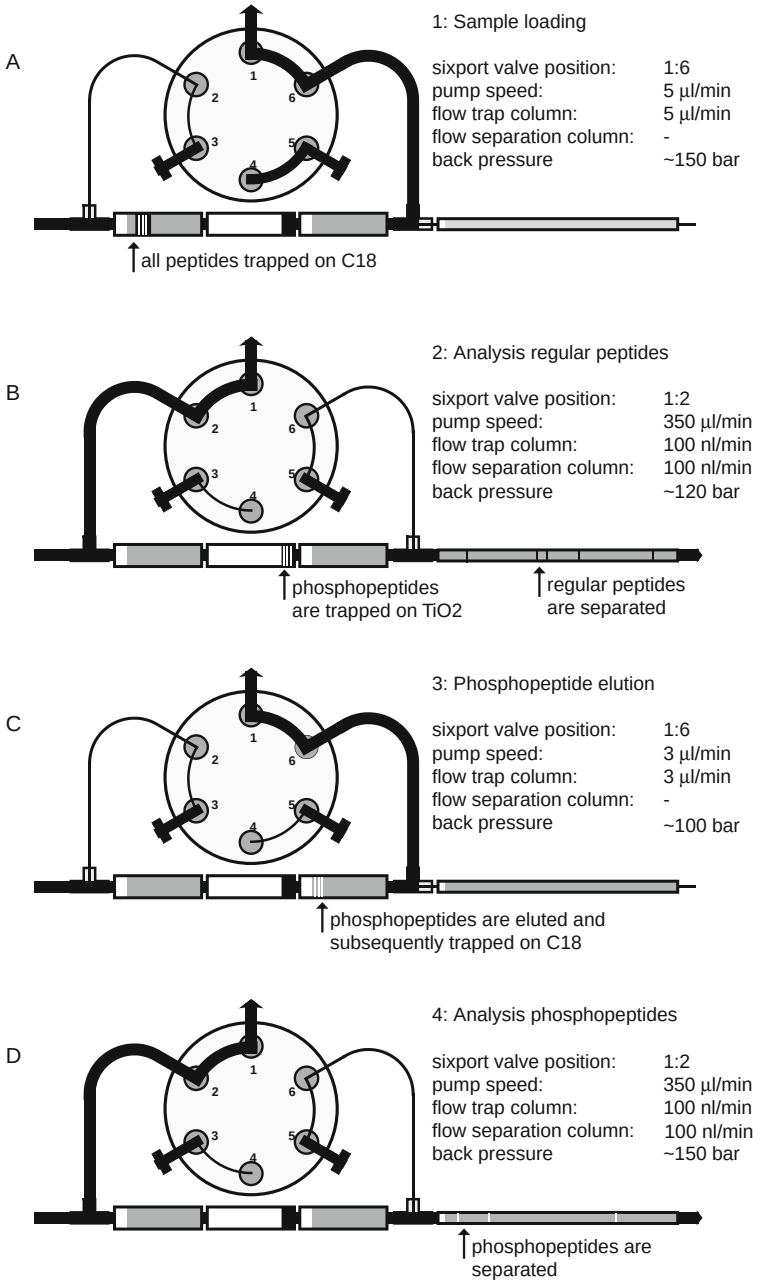


Fig. 14.2. Schematic diagram showing the principle of "sandwich" trapping valve positions given for each individual steps. (**a**) During sample loading all peptides and phosphopeptides are trapped on the first C18 trap-column with a flow of 5 µl/min. (**b**) The first water/acetonitrile gradient transports all peptides over the TiO_2 trap-column at a flow of 100 nl/min. Phosphopeptides are highly efficiently trapped at this flow, while all other peptides pass through and are separated on the separation column. (**c**) Injection of an alkaline phosphopeptide elution buffer desorbs the phosphopeptides, which are trapped again on the second C18 trap-column. (**d**) The second water/acetonitrile gradient is used to separate the phosphorylated peptides.

the autosampler, inject 30 µl of phosphopeptide elution buffer directly after the first water/acetonitrile gradient at 3 µl/min for 20 min.

5. Next, inject 10 µl of 5% formic acid at 5 µl/min for 10 min, and directly after this, start another water/acetonitrile gradient to separate and detect the phosphorylated peptides trapped on the second C18 trap-column. Now, the switching valve must again be in the 1:2 position.

3.5. Sample Preparation

1. In order to test the performance of the phosphopeptide enrichment method, prepare a standard protein digest. Reduce a solution of 100 µM of bovine serum albumin, bovine alpha and beta casein, and glyceraldehyde dehydrogenase with 5 mM DTT at 50°C for 30 min and subsequently alkylate with 10 mM iodoacetamide for 30 min at room temperature and in the dark. Add sequencing-grade trypsin in a protein:protease ratio of 50:1 (w/w) and incubate the mixture overnight at 37°C. Store 50 pmol aliquots of this digest at –20°C. Prior to use, dilute one such aliquot to 50 fmol/µl using 0.5% formic acid.
2. Depending on the type of cell or tissue from which phosphorylated peptides will be characterized, proper care has to be taken to prevent undesired dephosphorylation by phosphatase activity. The use of phosphatase inhibitors is strongly recommended. The following example is applicable to mammalian cells in suspension. Harvest the cells and subsequently lyse them in lysis buffer and centrifuge the lysate at 13,000×*g* for 20 min to pellet cellular debris. Add endoproteinase Lys-C to the supernatant in a protein:protease ratio of 200:1 (w/w) and incubate for 4 h at 37°C.
3. Reduce peptides with 2 mM DTT at 50°C for 30 min and alkylate with 4 mM iodoacetamide for 1 h at room temperature in the dark. Dilute this solution fourfold with 25 mM ammonium bicarbonate, pH 8.0, to reduce the urea concentration to 2 M, add sequencing-grade trypsin (in a ratio of 50:1, w/w), and digest the sample overnight at 37°C.
4. Prior to online phosphopeptide analysis, desalt the proteolytic digest by concentrating by solid phase extraction or another means of peptide desalting fractionation (*see* **Note 9**). Place a small plug of C18 material from a 3 M Empore C18 extraction disk into an Eppendorf GELoader pipette tip. Place onto this C18 plug 10 µl of 5 µm C18 column material. Wash the tip with 50 µl of solvent B and equilibrate it using 50 µl of solvent A (repeat the latter step once). Then, load the sample (typically 0.5 ml) onto the tip, wash it three times with 50 µl of solvent A, and elute the peptides with 50 µl of solvent B. Dry the eluate in a vacuum centrifuge.

3.6. Monitoring Performance

1. In order to check the performance of the system, inject 25 fmol of the standard protein digest on a C18-only column system.
2. Measure the retention time and base peak intensity of the phosphorylated peptides from the tryptic digests by making extraction ion chromatograms.

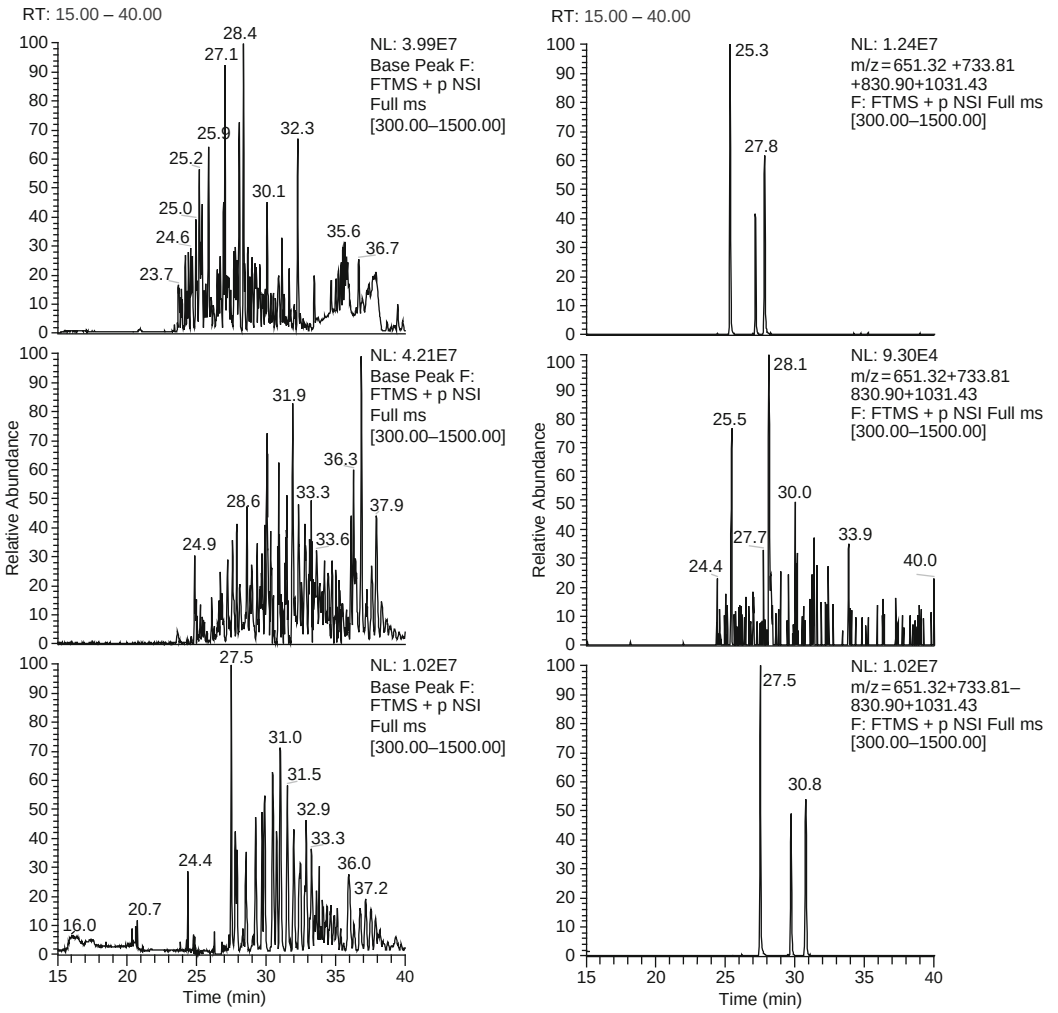


Fig. 14.3. Typical chromatograms of a phosphopeptide enrichment from 25 fmol of the digested protein standard. Chromatograms on the *left* are base peak intensity chromatograms of (*top*) analysis on a normal C18 capillary column (*middle*), the first gradient on the 2D system, and (*bottom*) the second gradient on the 2D system. On the *right* the complementary extracted ion chromatogram of three phosphorylated peptides from the protein mixture is shown. The three peptides are not detected in the first 2D system gradient and are exclusively found in the second gradient, with high recovery. The *m/z* values of the alpha and beta casein phosphopeptides shown in this example are the $[M+2H]^{2+}$ at *m/z* 1,031.3 of FQsEEQQQTEDELQDK, the $[M+2H]^{2+}$ at *m/z* 733.8 of TVDMEsTEVFTK, the $[M+3H]^{3+}$ at *m/z* 651.3 of YKVPQLEIVPNsAEER, and the $[M+2H]^{2+}$ at *m/z* 830.9 of VPQLEIVPNsAEER. The last two peptides co-eluted in this example. Data were acquired on an LTQ-Orbitrap mass spectrometer (Thermo Electron, Bremen, Germany).

3. Inject 25 fmol of the standard protein mixture on the system using the triple trap-column and run the first gradient.
4. Make extracted ion chromatograms using the *m/z* values of the phosphorylated peptides previously used in Step 2 and inspect the chromatogram to check if phosphorylated peptides are absent (i.e., completely trapped on the TiO_2 material) (*see* **Note 10**).
5. Inject 30 μl of phosphopeptide elution buffer at 3 μl/min for 20 min (*see* **Note 11**).
6. Inject 10 μl of 5% formic acid at 5 μl/min and start the second gradient.
7. Repeat Step 4 and determine whether the height or area of the observed peaks is of the same order as observed in the C18-only analysis (*see* **Note 12**). An example of the performance of the enrichment of phosphopeptides from the 25 fmol standard protein digests is given in **Fig. 14.3**.
8. Perform another cycle of phosphopeptide elution to check if sufficient numbers of phosphopeptides were eluted the first time. Typically, less than 5% of the phosphopeptide signals should be observed in a second elution cycle.

4. Notes

1. The C18 material should be fully compatible to work under full water conditions and should be stable against exposure to high and low pH.
2. Formamide is a teratogenic compound and possible carcinogenic. Use proper protection and safety when handling.
3. In combination with suitable connectors and sleeves, it is also possible to use fused silica of 360 μm outer diameter.
4. The use of metal parts in the sample flow path might cause phosphopeptides to get immobilized. Hence, metal surfaces in the autosampler needle, tubing, and/or metal frits should be avoided as much as possible.
5. Iodoacetamide is a carcinogenic substance; use suitable protection when handling this substance.
6. Packing of trap-columns and separation columns is best monitored using a stereo-microscope with sufficient enlargement.
7. During column packing, two pieces of fused silica can be tightly connected against each other using a 2 cm piece

of Teflon tubing (1/16 in. OD, 0.3 mm ID) that is prepared by forcing a third piece of fused silica through the Teflon tubing a single time, creating a slightly larger internal diameter. The two pieces of fused silica are inserted into this pre-treated Teflon sleeve so that they meet in the middle of the Teflon sleeve. This connection cannot withstand too high pressures, but is strong enough during column packing.

8. The flow over the separation column can be measured by pumping water through the system and collecting the column effluent in a 10–20 μl pipette tip that has been accu-

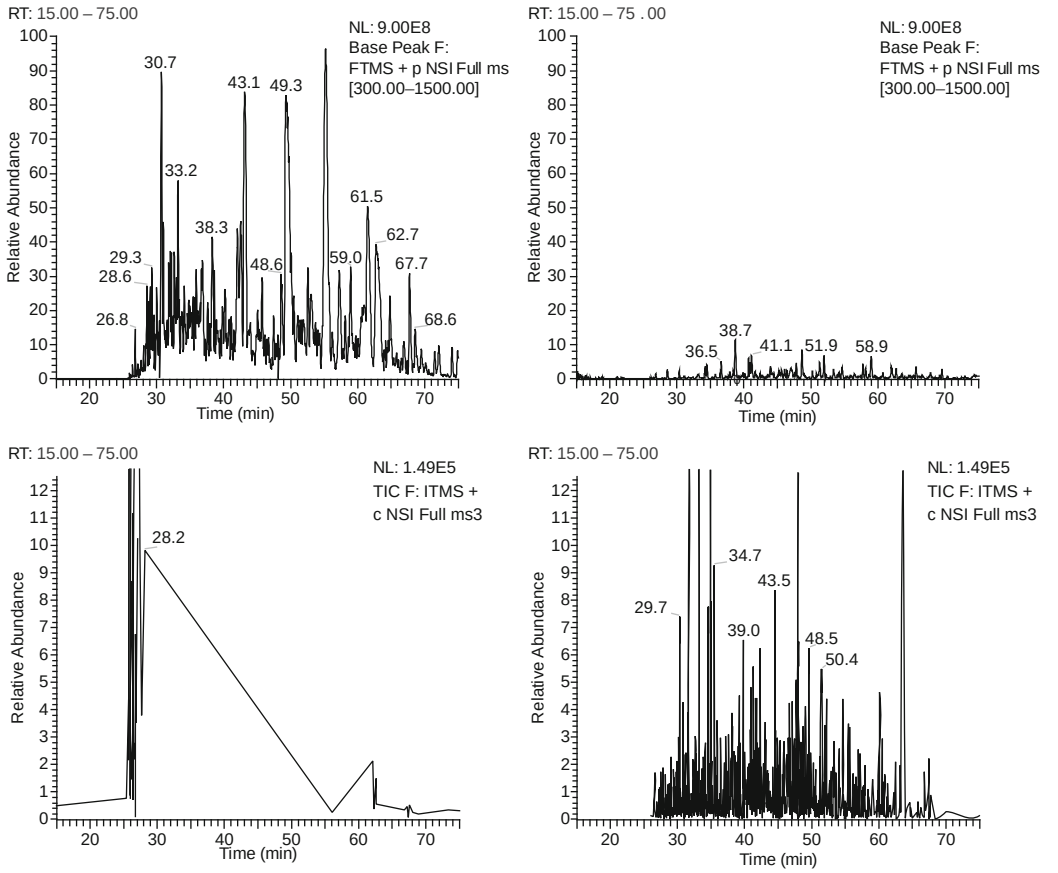


Fig. 14.4. Example from a large-scale analysis of phosphorylation site in human embryonic stem cells (8). Shown are (*left*) the first and (*right*) second gradient of one single strong cation exchange fraction containing both phosphorylated peptides and regular C-terminal peptides and N-terminally acetylated peptides. The top traces are the base peak intensity chromatograms of the full MS scan and the lower traces are the neutral loss-driven MS3 scan (loss of 98 Da for phosphoric acid). Although the base peak intensity signal is ~10 lower in the second gradient, the number of MS3 scans (indicative for the presence of phosphorylated peptides) is much bigger. The MS3 signals that are seen in the first gradient are due to non-phosphorylated peptides (for example, neutral loss of N-terminal valine residues or side chain loss of oxidized methionines). Data were acquired on an LTQ-Orbitrap mass spectrometer (Thermo Electron, Bremen, Germany).

rately weighted. After collecting the column effluent for 1 min the pipette tip is weighted again and the increase in mass per minute (in μg) is equal to the flow in nanoliters per minute. In all steps handle the pipette tip using a pair of tweezers.

9. Depending on the complexity of the sample, pre-fractionation of the proteolytic digest is advisable. For example, strong cation exchange of the peptide mixture is very beneficial for pre-fractionation of phosphorylated peptides (**Fig. 14.4**) (13).
10. If breakthrough of phosphorylated peptides is observed in the first gradient, this could be caused by bad performance of the first C18 trap-column and needs to be replaced by a new C18 trap-column. In general, the first C18 trap-column is damaged the most often, due to small particles and/or other compounds from injected samples. Using a microscope the end of the column can be examined, and damage is often visible as a dark color on the white column material.
11. When large amounts of peptides are injected on the nanoscale column, carry-over or memory effects can occur. If this is the case, it is advisable to perform a blank run in between the first sample injection and first phosphopeptide elution injection.
12. Performance and efficiency of the sandwich trap-column can be monitored on a regular basis using the standard protein mixture.

Acknowledgments

The authors would like to acknowledge GL Sciences Inc., Tokyo, for the TiO_2 Titansphere material. This work was supported by the Netherlands Proteomics Centre (www.netherlandsproteomicscentre.nl), a program embedded in the Netherlands Genomics Initiative.

References

1. Mukherji M (2005) Phosphoproteomics in analyzing signaling pathways. *Expert Review of Proteomics.* **2**, 117–128.
2. Rogers LD, Foster LJ (2009) Phosphoproteomics–finally fulfilling the promise? *Molecular Biosystems.* **5**, 1122–1129.
3. Gafken PR (2009) An overview of the qualitative analysis of phosphoproteins by mass spectrometry. *Methods in Molecular Biology.* **527**, 159–172.
4. Thingholm TE, Jensen ON, Larsen MR (2009) Analytical strategies for phosphoproteomics. *Proteomics.* **9**, 1451–1468.
5. Reinders J, Sickmann A (2005) State-of-the-art in phosphoproteomics. *Proteomics.* **5**, 4052–4061.

6. Lemeer S, Pinkse MW, Mohammed S, et al. (2008) Online automated in vivo zebrafish phosphoproteomics: from large-scale analysis down to a single embryo. *Journal of Proteome Research*. **7**, 1555–1564.
7. Mohammed S, Kraiczek K, Pinkse MW, Lemeer S, Benschop JJ, Heck AJ (2008) Chip-based enrichment and nanoLC-MS/MS analysis of phosphopeptides from whole lysates. *Journal of Proteome Research*. **7**, 1565–1571.
8. Pinkse MW, Mohammed S, Gouw JW, van Breukelen B, Vos HR, Heck AJ. (2008) Highly robust, automated, and sensitive online TiO2-based phosphoproteomics applied to study endogenous phosphorylation in Drosophila melanogaster. *Journal of Proteome Research*. **7**, 687–697.
9. Van Hoof D, Munoz J, Braam SR, et al. (2009) Phosphorylation dynamics during early differentiation of human embryonic stem cells. *Cell Stem Cell*. **5**, 214–226.
10. Rappsilber J, Ishihama Y, Mann M (2003) Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Analytical Chemistry*. **75**, 663–670.
11. Schlosser A, Vanselow JT, Kramer A. (2005) Mapping of phosphorylation sites by a multi-protease approach with specific phosphopeptide enrichment and NanoLC-MS/MS analysis. *Analytical Chemistry*. **77**, 5243–5250.
12. Meiring HD, van der Heeft E, ten Hove GJ, de Jong APJM (2002) Nanoscale LC-$MS^{(n)}$; technical design and applications to peptide and protein analysis. *Journal of Separation Science*. **25**, 557–568.
13. Beausoleil SA, Jedrychowski M, Schwartz D, et al. (2004) Large-scale characterization of HeLa cell nuclear phosphoproteins. *Proceedings of the National Academy of Sciences of the United States of America*. **101**, 12130–12135.

Chapter 15

Positional Proteomics at the N-Terminus as a Means of Proteome Simplification

Gemma R. Davidson, Stuart D. Armstrong, and Robert J. Beynon

Abstract

One strategy to reduce complexity in proteome analysis is through rational reduction of the proteolytic peptides that constitute the analyte for mass spectrometric analysis. Methods for selective isolation of C- and N-terminal peptides have been developed. In this chapter, we outline the context and variety of methods for selective isolation of N-terminal peptides and detail one method based on negative selection through differential removal of internal peptides.

Key words: Proteomics, proteome simplification, N-terminal peptide isolation, biotinylation, positional proteomics, mass spectrometry.

1. Introduction

It is rare for proteome analysis to be conducted on unfractionated, complex mixtures of intact proteins. Some strategies aim to reduce complexity at the protein level (by, for example, gel-based separation). In other instances, the first analytical step in many experimental workflows in 'bottom-up' proteomics is based on selective hydrolysis into an even more complex mixture of peptides. Usually, the fragmentation reagent is trypsin, which cleaves peptide bonds C-terminal to lysine and arginine residues. Together, these two amino acids constitute about 8–10% of the amino acids in a proteome, and thus, tryptic fragments are short (typically about 15 amino acids long). Since each protein (~50 kDa average) will generate on the order of 40 tryptic peptides, this is a substantial increase in analyte complexity. This might therefore seem counterproductive, but against this

K. Gevaert, J. Vandekerckhove (eds.), *Gel-Free Proteomics*, Methods in Molecular Biology 753,
DOI 10.1007/978-1-61779-148-2_15,

increase in complexity are tensioned two positive factors: the more uniform and predictable behaviour of the peptides on high-resolution chromatographic separation and the reduction in size relative to the parent protein, which bring peptides into the optimal mass range for most mass spectrometers. This increase in complexity brings further advantages – identification or quantification of the parent protein can be attained by tandem mass spectrometric analysis of relatively few of the peptides. Indeed, it can be argued that many proteomic analyses suffer from 'over-determinism'; we analyse more peptides than are required to attain the goal of the experiment. This is particularly true not only for abundant proteins but for those proteins that are less abundant; not only are the peptides present at lower concentrations, but their analysis can be impeded by crowding or ion suppression in the analyte stream. It is not surprising, therefore, that there have been multiple attempts to achieve a reduction in proteome complexity. Some, based on the amino composition of each protein, are directed at specific chemistries associated with particular amino acid residues (for example, cysteine residues). Others attempt a rational reduction in proteome complexity by targeting one of the features that is common to every protein in the cell – an N-terminus and a C-terminus – so-called positional proteomics. Such proteomic simplification strategies have the aim of enhancement of proteome analysis by the selective removal or isolation of specific peptides, leading to a reduction in sample complexity without compromising information for analysis (1).

Positional simplification strategies for global proteome characterization are based upon the concept that peptide subsets provide substantial information with respect to the parent protein. Utilizing one peptide as a surrogate for its parent protein requires a degree of confidence in the information provided by the peptide of interest. Against this must be tensioned the resistance of the proteomics community to the use of a single peptide hit to confirm protein identification, thus the term 'one hit wonders' has been coined. Much of the controversy associated with 'one hit wonders' relates to the origin of the peptide and its connectivity to the parent protein (2). Traditional proteomic workflows often begin with protein separation of a sample by SDS-PAGE followed by in-gel proteolysis of a single protein and often mass spectrometric analysis. This technique generates a set of peptides, each of which can assume connection to the starting protein. However, when dealing with a complex biological sample, such as human plasma, an in-solution/shotgun proteolytic approach will yield an even more complex peptide mixture where connectivity between peptides and parent proteins can no longer be assumed. Therefore, it becomes increasingly difficult to confirm protein identifications from a complex biological sample based on a single peptide. However, the general resistance to single peptide hits is

based on an experimental workflow in which the peptide could have been derived from anywhere in the protein. Simplification strategies in which the positional location of a peptide within the parent protein is known gain assurance. Effectively, the identification becomes a question of matching a peptide to the positionally defined single peptide derived from each protein. For a search against a 6,000 protein database, the search is no longer a single peptide against 300,000 possibilities in the entire digested proteome but a single positional peptide against a predicted set of 6,000.

Two positional locations are consistent amongst all proteins: the extremities (N-terminus and C-terminus). N-terminal and C-terminal peptide isolation can reduce sample complexity and also provide sequence information essential for understanding key processing events, such as in vivo modifications, removal of signal peptide and cleavage of methionine during protein maturation. Terminal sequence information is extremely valuable for protein identification, which can take place when four or five amino acids within the N-terminal sequence are known (3). Complementary methods have been developed for C-terminal analysis and isolation but these will not be discussed here.

A complexity in the analysis of protein N-termini lies in their diversity, a consequence of the substantial post-translational processing that can succeed de novo biosynthesis. When first translated, a nascent polypeptide chain can undergo extensive N-terminal processing (4), removal of signal peptides (endoproteolytic), excision of single amino acids (exoproteolytic) and N-terminal amino group modification by, for example, N-α-acetylation. Of course, most protein sequences in databases are derived from in silico translation of the corresponding genomic or cDNA sequence, and thus, the analyte, which is the true N-terminus in all its complexity, is likely to fail to be precisely represented in any sequence database, which poses a significant challenge for post-analytic bioinformatics. Indeed, positional proteomics can enhance protein databases by defining, unambiguously, the true N-terminus of an uncharacterized open reading frame.

There is a strong historical precedent to the analysis of the N-terminus of proteins based on the early development of Edman degradation. As mass spectrometry has reached into the arena of protein and peptide chemistry, several strategies for specific exploration of protein termini have been published. In this chapter, we address the methodologies for analysis of the N-terminus, more extensively developed as the chemical reactivity of the free amino group makes it easier to derivatize. One method in particular is detailed here.

In an analytical workflow for positional proteomics, the N-terminal peptide will be excised almost inevitably from the

parent protein by a proteolytic enzyme, and this will usually be trypsin. Moreover, many protocols block lysine residues (for example, by acetylation) and trypsin is not able to cleave peptide bonds adjacent to N-ε-acetyl lysine residues, reducing the cleavage sites to those adjacent to arginine residues only. The number of sites that can be digested in any protein is thus reduced by a factor of 2. Even then, the disposition of arginine residues is random and there will be some N-terminal peptides that are too small for analysis and do not contain sufficient sequence specificity to permit identification of the source proteins. Other peptides will be too large for chromatographic separation and mass spectrometric analysis. Inevitably then, all positional methods are selective, and greater coverage of the proteome would probably require orthogonal methods (e.g. multiple proteinases of differing specificities or a combination of N-terminal and C-terminal isolation).

Although there are a large number of published methods for analysis and isolation of N-terminal peptides, they reduce to simple analytical principles that predominantly make use of the nucleophilicity of the α-amino group as a reactive centre. The second source of amino groups in proteins is the side chain of lysine residues that are only marginally different in reactivity when compared to α-amino groups. However, targeting to N-terminal α-amino groups or side chain ε-amino groups is feasible (by careful control of pH and reaction time and by selective chemistries such as guanidinylation).

Proteomic strategies employed for the analysis of protein N-termini often involve chemical derivatization, affinity enrichment and labelling or selective tagging (for recent reviews *see* 5–7). Most derivatization and enrichment strategies are based upon the manipulation of differences induced by the addition or removal of chemical moieties. It is estimated that up to 80% of eukaryotic proteins have an acetylated α-amino group (8, 9), the function of which is not yet fully understood (10). Typically, N-α-blocked peptides will exhibit a single positive charge at acidic pH, if histidine is not present, unlike tryptic peptides that have at least two positive charges due to free α-amino group and C-terminal lysine or arginine. Cation exchange exploits the difference in charge between blocked N-terminal peptides and internal tryptic peptides for isolation (11, 12). Further development of cation exchange methods for isolation of N-α-blocked peptide using successive endoproteolysis and exoproteolysis has been reported in the literature (13). Strong cation exchange has proven beneficial for the simultaneous analysis of blocked (for example, by acetylation) N-terminal peptides and protein C-terminal peptides, due to their reduced basicity when compared to non-acetylated tryptic peptides (11, 14). Coupled with novel database searching strategies (15), strong cation exchange remains one of

the few methods observed in the literature for the simultaneous enrichment of C-terminal and blocked N-terminal peptides.

Selective isolation and enrichment of N-terminal peptides routinely requires derivatization. Derivatization using acetylation blocks the α-amino group of the N-terminal peptide and ε-amino group of lysine residues. Upon enzymatic cleavage internal peptides yield free α-amino groups. Affinity removal of the internal peptides containing free α-amino groups using amine reactive matrices such as NHS-activated Sepharose (16), cyanogen bromide-activated Sepharose (17, 18) or isocyanate-coupled resin (19) leads to enrichment of α-amino blocked (N-terminal) peptides. Derivatization of free amino groups using biotinylating reagents followed by affinity removal with avidin or streptavidin (20) has been used to remove internal peptides for N-terminal peptide enrichment (21) and for N-terminal peptide removal and enrichment (22).

Gevaert and Vandekerckhove (23) pioneered a technique known as combined fractional diagonal chromatography (COFRADIC), based on the principle of chemical derivatization with a sorting regent that elicited a change in chromatographic behaviour of a subset of peptides. This approach, developed for selective isolation of α-amino blocked peptides has been reported in the literature (7, 24, 25). After reversed-phase chromatographic separation of a peptide mixture the blocked α-amino peptides or N-terminal peptides are segregated from internal peptides by the reaction of the free amino groups of the internal peptides with 2,4,6-trinitrobenzenesulfonic acid (TNBS) that induces a strong hydrophobic shift allowing separation of internal peptides from N-terminal peptides in a second chromatographic separation step. Recent developments of this method involve the use of strong cation exchange for the reduction of peptide noise observed by internal peptides and enzymatic treatment to enhance exposure of true N-terminal peptides (25).

Utilizing positional proteomics for the selective enrichment of N-terminal peptides of proteins found within complex biological samples has some advantages over global, 'bottom-up' proteomic workflows. The reduction in sample complexity observed following the isolation of N-terminal peptides reduces the peptide 'noise' and could therefore solve the challenge imposed by analysis of complex biological samples using shotgun strategies. The protocol described in this chapter selectively enriches N-terminally blocked peptides via the selective removal of internal peptides (16, 21) with changes in the chemical derivatization and removal protocols. Reagents are readily available and can be easily modified for the study of naturally acetylated N-terminal peptides. From sample preparation to collection of mass spectrometric data requires 3–4 days (*see* **Fig. 15.1** for workflow outline).

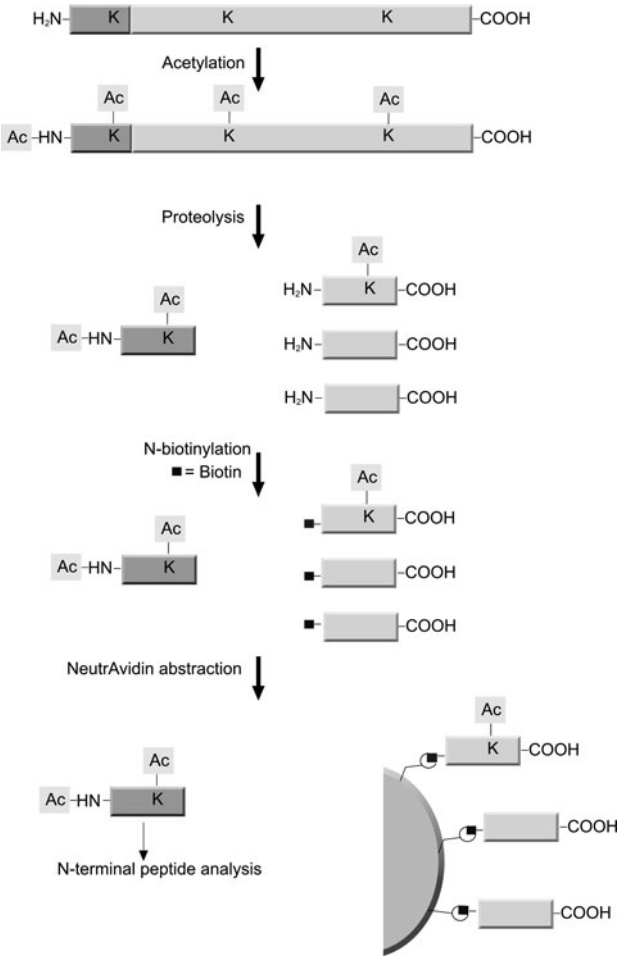


Fig. 15.1. Overview of N-terminal simplification strategy.

2. Materials

2.1. Equipment

1. StageTips, C18 pipette tips (Proxeon)
2. Microcentrifuge tubes, 0.5 and 1.5 ml
3. Glass vials, 0.1–1 ml
4. MALDI-TOF mass spectrometer
5. High-resolution nanoflow chromatography system, fitted with a reversed-phase C18 column, coupled to an electrospray ionization tandem mass spectrometer (optional)

2.2. Reagents

1. Coomassie Plus® protein assay reagent (Thermo Scientific)

2. HPLC-grade water
3. HPLC-grade acetonitrile
4. Dithiothreitol
5. Iodoacetamide
6. Formic acid
7. Dimethylformamide
8. Acetylation reagent: acetic anhydride
9. Acetylation buffer: 1 M sodium carbonate, pH 8.5 (store at room temperature)
10. Quenching reagent: Tris(2-aminoethyl)amine, polymer bound (Aldrich)
11. Trichloroacetic acid
12. Diethyl ether (flammable, use in fume hood)
13. Digestion buffer: 20 mM sodium phosphate buffer, pH 7.5 (store at 4°C)
14. Trypsin (or Arg C), sequencing grade (Roche), 0.1 μg/μl reconstitute in 50 mM acetic acid (store at 4°C)
15. EZ-Link NHS-biotin (Thermo Scientific, store desiccated at room temperature)
16. NeutrAvidin (or streptavidin) agarose, high capacity (Thermo Scientific, store at 4°C in ethanol)
17. Binding buffer: 20 mM sodium phosphate buffer, pH 7.5 (store at 4°C)
18. MALDI matrix: 10 mg/ml α-cyano-4-hydroxycinnamic in 50% acetonitrile, 0.2% formic acid (v/v)
19. HPLC Buffer A: 0.1% formic acid (v/v)
20. HPLC Buffer B: 90% acetonitrile, 0.1% formic acid (v/v)

3. Methods

3.1. Selective Enrichment of the N-Terminal Peptides of Proteins

1. Measure the protein concentration of the sample using, for example, the Coomassie Plus® protein assay. The protein concentration should ideally be 2–4 mg/ml. Ensure the protein sample is in a buffer compatible with the protocol, i.e. is non-amine containing. If not, dialyse the protein preparation against a suitable buffer such as sodium carbonate, sodium phosphate or HEPES (pH 7–9) (*see* **Note 1**).

2. If required, reduce and alkylate the protein sample using dithiothreitol and iodoacetamide at final concentrations of 3 and 9 mM, respectively (*see* **Note 2**). Incubate with dithiothreitol for 30 min at 50°C, followed by incubation with iodoacetamide for 1 h in the dark at room temperature.
3. Acetylate intact proteins by adding 50 μl of 1 M Na_2CO_3, pH 8.5, to 50 μl of 100 μg of the soluble protein solution (this buffers acid formation) in a 0.5 ml microcentrifuge tube. Add 1 μl of acetic anhydride to the 100 μl solution. Vortex for 20 s. Incubate at room temperature for 1 h. Add an additional 1 μl of acetic anhydride (>600-fold molar excess of acetic anhydride, reduce accordingly for less complex protein mixtures), vortex and incubate a further 1 h at room temperature (*see* **Note 3**).
4. Add approximately 5 mg of the free amine quenching reagent (Tris(2-aminoethyl)amine, polymer bound, 10-fold molar excess) directly to the acetylated protein solution, vortex for 30 s and incubate with gentle agitation at room temperature for 1 h (*see* **Note 4**). Recover acetylated proteins by using spin columns; using a needle carefully pierce a small hole in the bottom of the 0.5 ml microcentrifuge tube containing the sample and place inside a larger 1.5 ml microcentrifuge tube. Centrifuge at 2,000×*g* for 1 min at room temperature. Discard resin and retain flow-through (acetylated proteins).
5. Add 150 μl of 30% TCA (w/v) to the acetylated protein sample, vortex for 20 s and incubate for 2 h on ice (*see* **Note 5**).
6. Centrifuge at 13,000×*g* for 10 min at room temperature to pellet acetylated proteins. Remove TCA using a pipette and discard. In a fume hood wash the protein pellet with 200 μl of diethyl ether, centrifuge for 10 s at 13,000×*g* at room temperature. Using a pipette, carefully remove diethyl ether without disturbing the protein pellet. Repeat diethyl wash steps a further two times. Allow pellet to air-dry.
7. Re-solubilize the protein pellet in 50 μl of 20 mM sodium phosphate, pH 7.5.
8. Digest the acetylated protein with sequencing-grade trypsin (50:1 substrate:enzyme) overnight at 37°C. This is effectively an endopeptidase ArgC digest as lysine residues are acetylated and not compatible with the specificity of trypsin.
9. If there is enough protein digest available, verify that complete protein digestion has occurred using SDS-PAGE (*see* **Note 6**). To check acetylation, dilute a small aliquot of

sample 1:20 with matrix. Spot 1 μl onto a clean MALDI-TOF target and allow to air-dry. Analyse the peptide mixture using MALDI-TOF over the range of *m/z* 800–3,500. The spectra obtained will vary according to the complexity of the initial protein sample. For more complex samples (e.g. *Escherichia coli* lysates, plasma) nLC-MS/MS analysis is preferable and will allow for a more extensive and definitive survey of the extent of peptide acetylation. N-terminal- and/or lysine-containing peptides will shift mass by +42 Da in the acetylated digest (**Fig. 15.2**, *see* **Note 7**). This acetylated peptide mixture can be stored frozen for several weeks.

10. Add 7 μl of EZ-Link NHS-biotin (1 mg in 50 μl DMF, prepared in a glass vial) to the acetylated peptide mixture (gives approximately 7 nmol of biotin/μl digest, a 20-fold molar excess). Vortex 20 s. Incubate at room temperature for 1–2 h or overnight at 4°C (*see* **Note 8**).
11. To check biotinylation, dilute the digest with matrix solution (1:20) and spot 1 μl onto a MALDI-TOF target and analyse as in Step 9. Again, more complex digests will require nLC-MS/MS analysis. Peptides with a free α-amino group ('internal' peptides) will have shifted mass by +226 Da (due to incorporation of biotin). N-terminal peptides will have no mass shift (**Fig. 15.2**). Once biotinylation is confirmed the 'internal' peptides can be removed using neutrAvidin (*see* **Note 9**).
12. Take 6 μl of the digest and dilute with 100 μl of 20 mM sodium phosphate buffer, 0.15 M NaCl, pH 7.5.
13. Add digest mixture to approximately 350 μl of neutrAvidin (washed three times with 200 μl 20 mM sodium phosphate buffer, 0.15 M NaCl, pH 7.5, remove excess buffer from resin by using spin columns as described in Step 4) in a 0.5 ml microcentrifuge tube. Ensure resin is just swollen, with no excess buffer. Incubate at room temperature for 1 h.
14. Spin neutrAvidin mixture as described in Step 4 and collect the flow-through (N-terminal peptides). Wash the neutrAvidin resin with an additional 100 μl of binding buffer. Pool N-terminal peptide eluates in a 0.5 ml Eppendorf tube.
15. Use a StageTip or a similar C18 reversed-phase column to concentrate and desalt the N-terminal peptide mixture (*see* **Note 10**).
16. Using the entire N-terminal preparation in one injection, separate N-terminal peptides using a reversed-phase column in line with an ESI-MS/MS capable mass spectrometer (*see* **Note 11**).

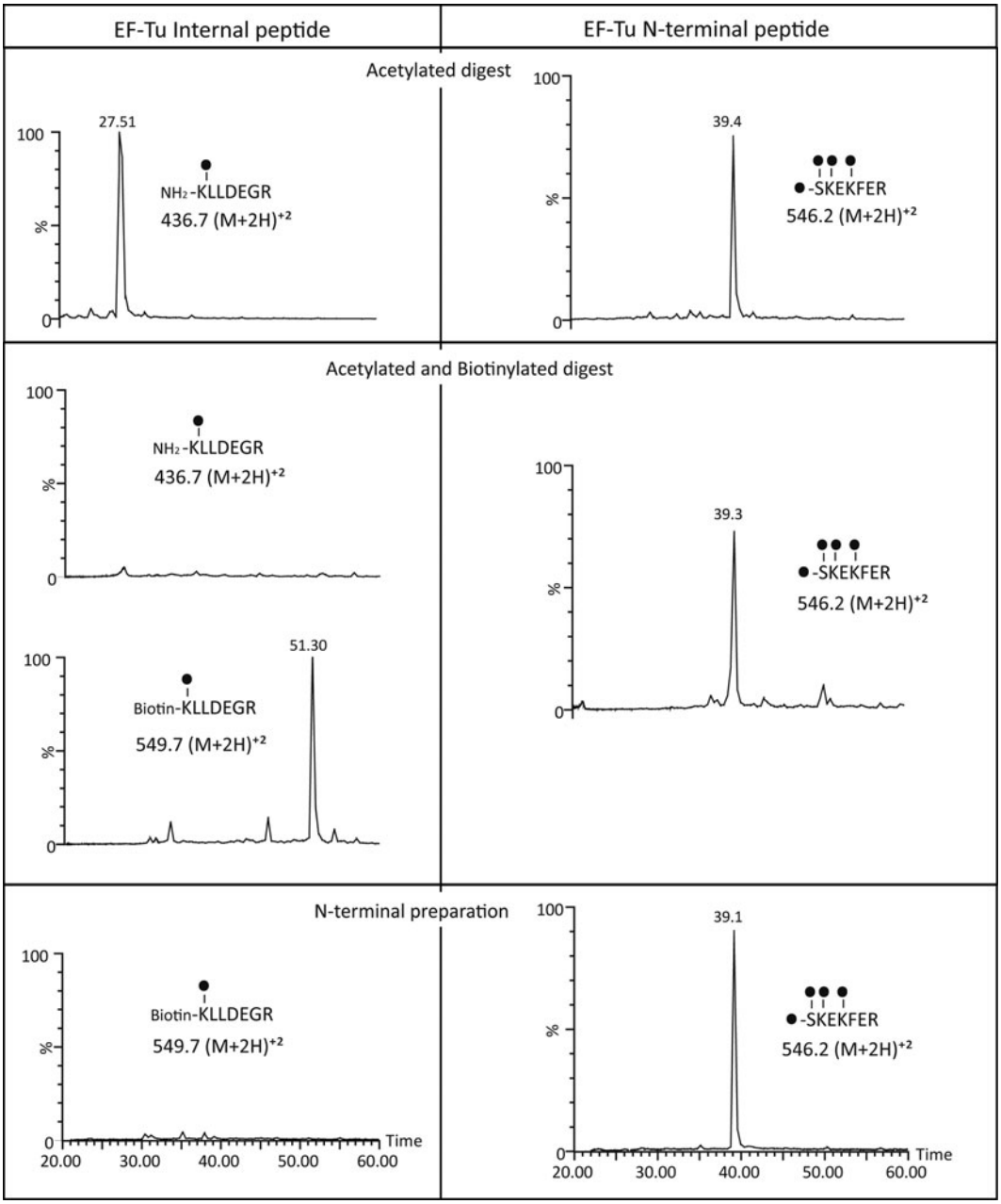


Fig. 15.2. Isolation of N-terminal peptides. A protein sample from *Escherichia coli* was passed through the N-terminal peptide protocol (**Fig. 15.1**). At each stage, samples were removed and analysed on a Waters Synapt QTOF instrument coupled to a nanoAcquity chromatography system. Extracted ion chromatograms were prepared to show the recovery of the N-terminal peptide (*right-hand panes*) and loss of an internal peptide (*left-hand panes*) for an abundant protein (elongation factor Tu) through the process. All chromatograms are normalized to the same scale.

17. Search the tandem mass spectra against Swiss-Prot (or database of choice) using MASCOT. Use search parameters that include fixed modifications of N-terminal acetylation and lysine acetylation (and carbamidomethyl if sample was reduced and blocked) and the variable modifications of O-acetylated serine (*see* **Note 12**) and methionine oxidation. To check that no internal peptides have managed to leak through the system use search parameters that include the fixed modifications of lysine acetylation and N-terminal biotinylation and the variable modifications, O-acetylated serine and methionine oxidation.

4. Notes

1. Amine-containing buffers (e.g. Tris or ammonium bicarbonate) will compete with proteins for acetylation. In the preparation of any protein lysate ensure that a proteolytic inhibitor cocktail (e.g. complete inhibitor cocktail tablets, from Roche Diagnostics) is included to limit degradation by endogenous proteases. This is important as uncontrolled proteolytic trimming may create artefactual protein N-termini.
2. Acetylation of some primary amino groups may be impaired by protein structure. Increasing the acetylation incubation time or using reagents that denature the protein structure, e.g. chaotropes, detergents and reducing agents, may increase primary amino group accessibility. These reagents may need to be removed prior to the acetylation reaction and/or to any mass spectrometry steps.
3. If a precipitate forms on addition of acetic anhydride, check that the correct buffer was added prior to the acetylation reaction. It is important to keep the pH at about 8.5 for the reaction.
4. This step ensures that any excess acetylation reagent is removed before the digestion step. Make sure that mixing is continuous by using a rotating or end-over-end mixer. If the proteins are subsequently precipitated using TCA, a Tris buffer solution (approximately 50 μl of 1.5 M Tris solution) can also be used to mop up excess acetylation reagent.
5. The precipitation step will denature proteins and will also remove unwanted reagents (e.g. reducing agents, any remaining acetylation reagent), prior to proteolysis, biotinylation and neutrAvidin steps.

6. Take an aliquot of digest (approximately 10 μg) and analyse using SDS-PAGE. No protein bands should be visible on the Coomassie stained gel. Incomplete digestion can be due to incorrect pH of the reaction mixture due to inadequately washed protein pellets. Wash pellets carefully to remove all traces of TCA.
7. If poor MALDI-TOF spectra are obtained, desalt the peptide mixture using a StageTip before analysis. If incomplete acetylation is observed it could be due to several factors. The reaction needs a longer incubation time. The acidification of the protein sample by acetic anhydride shifts the pH from the optimal range. The accessibility of amine groups and the presence of competing amines will also hinder acetylation (*see* **Notes 1** and **2**).
8. Use NHS-ester solutions immediately as they readily hydrolyse and become un-reactive. Do not store or reuse the solution. Ensure organic solvent concentration does not exceed 20% of final reaction volume. DMF may strip polymers from plastic tubes and result in contaminant peaks in mass spectrometry analysis. Use glass vials to contain the NHS-biotin stock solution and perform the biotinylation reaction.
9. If biotinylation is incomplete, make sure that the NHS-biotin solution is fresh and/or increase the incubation time.
10. The relatively low capacity of neutrAvidin for biotin means that a large amount is needed to bind excess NHS-biotin as well as the 'internal' biotin-labelled peptides. Another approach is to load a greater amount (above the biotin-binding capacity of the neutrAvidin) of the acetylated and biotinylated peptide preparation onto a neutrAvidin packed column and slowly pump the peptide sample through (e.g. 100 μl/h with a syringe pump), collecting fractions. Monitor each fraction for N-terminal peptides using MALDI-TOF or nLC-MS/MS. When biotinylated peptides elute, cease taking fractions. Desalt and concentrate the N-terminal fraction(s) using StageTips and analyse using nLC-MS/MS.
11. If biotinylated peptides are present in the N-terminal peptide mixture, increase the volume of neutrAvidin used and increase the coupling time.
12. There are a number of methods reported in the literature for the reversal of O-acetylation. Popular strategies include treating the sample with hydroxylamine (25) or heating the sample in a hot water bath.

References

1. Gevaert, K., Ghesquiere, B., Staes, A., Martens, L., Van Damme, J., Thomas, G. R., and Vandekerckhove, J. (2004) Reversible labeling of cysteine-containing peptides allows their specific chromatographic isolation for non-gel proteome studies, *Proteomics 4*, 897–908.
2. Veenstra, T. D., Conrads, T. P., and Issaq, H. J. (2004) Commentary: What to do with "one-hit wonders"? *Electrophoresis 25*, 1278–1279.
3. Wilkins, M. R., Gasteiger, E., Tonella, L., Ou, K., Tyler, M., Sanchez, J. C., Gooley, A. A., Walsh, B. J., Bairoch, A., Appel, R. D., Williams, K. L., and Hochstrasser, D. F. (1998) Protein identification with N and C-terminal sequence tags in proteome projects, *Journal of Molecular Biology 278*, 599–608.
4. Meinnel, T., and Giglione, C. (2008) Tools for analyzing and predicting N-terminal protein modifications, *Proteomics 8*, 626–649.
5. Agard, N. J., and Wells, J. A. (2009) Methods for the proteomic identification of protease substrates, *Current Opinion in Chemical Biology 13*, 503–509.
6. Doucet, A., Butler, G. S., Rodriguez, D., Prudova, A., and Overall, C. M. (2008) Metadegradomics: Toward in vivo quantitative degradomics of proteolytic post-translational modifications of the cancer proteome, *Molecular & Cellular Proteomics 7*, 1925–1951.
7. Van Damme, P., Van Damme, J., Demol, H., Staes, A., Vandekerckhove, J., and Gevaert, K. (2009) A review of COFRADIC techniques targeting protein N-terminal acetylation, *BMC Proceedings 3*(Suppl 6), S6.
8. Brown, J. L., and Roberts, W. K. (1976) Evidence that approximately 80 per cent of soluble-proteins from Ehrlich ascites-cells are N-alpha-acetylated, *Journal of Biological Chemistry 251*, 1009–1014.
9. Polevoda, B., and Sherman, F. (2000) N-α-terminal acetylation of eukaryotic proteins, *Journal of Biological Chemistry 275*, 36479–36482.
10. Hwang, C. S., Shemorry, A., and Varshavsky, A. (2010) N-terminal acetylation of cellular proteins creates specific degradation signals, *Science 327*, 973–977.
11. Kawasaki, H., Imajoh, S., and Suzuki, K. (1987) Separation of peptides on the basis of the difference in positive charge – simultaneous isolation of C-terminal and blocked N-terminal peptides from tryptic digests, *Journal of Biochemistry 102*, 393–400.
12. Titani, K., Okunuki, K., and Narita, K. (1962) N-Terminal sequence in beef- and horse-heart cytochrome C, *Journal of Biochemistry 51*, 350–358.
13. Betancourt, L., Besada, V., Gonzalez, L. J., Morera, V., Padron, G., Takao, T., and Shimonishi, Y. (2001) Selective isolation and identification of N-terminal blocked peptides from tryptic protein digests, *Journal of Peptide Research 57*, 345–353.
14. Gorman, J. J., and Shiell, B. J. (1993) Isolation of carboxyl-termini and blocked amino-termini of viral-proteins by high-performance cation-exchange chromatography, *Journal of Chromatography 646*, 193–205.
15. Dormeyer, W., Mohammed, S., van Breukelen, B., Krijgsveld, J., and Heck, A. J. R. (2007) Targeted analysis of protein termini, *Journal of Proteome Research 6*, 4634–4645.
16. McDonald, L., and Beynon, R. J. (2006) Positional proteomics: Preparation of amino-terminal peptides as a strategy for proteome simplification and characterization, *Nature Protocols 1*, 1790–1798.
17. Akiyama, T. H., Sasagawa, T., Suzuki, M., and Titani, K. (1994) A method for selective isolation of the amino-terminal peptide from [alpha]-amino-blocked proteins, *Analytical Biochemistry 222*, 210–216.
18. Zhang, X., Ye, J., and Højrup, P. (2009) A proteomics approach to study in vivo protein N-α-modifications, *Journal of Proteomics 73*, 240–251.
19. Mikami, T., and Takao, T. (2007) Selective isolation of N-blocked peptides by isocyanate-coupled resin, *Analytical Chemistry 79*, 7910–7915.
20. Miller, B. T., Collins, T. J., Rogers, M. E., and Kurosky, A. (1997) Peptide biotinylation with amine-reactive esters: Differential side chain reactivity, *Peptides 18*, 1585–1595.
21. McDonald, L., Robertson, D. H., Hurst, J. L., and Beynon, R. J. (2005) Positional proteomics: Selective recovery and analysis of N-terminal proteolytic peptides, *Nature Methods 2*, 955–957.
22. Yamaguchi, M., Nakazawa, T., Kuyama, H., Obama, T., Ando, E., Okamura, T., Ueyama, N., and Norioka, S. (2005) High-throughput method for N-terminal sequencing of proteins by MALDI mass spectrometry, *Analytical Chemistry 77*, 645–651.
23. Gevaert, K., and Vandekerckhove, J. (2004) COFRADIC(TM): The Hubble telescope of proteomics, *Drug Discovery Today: TARGETS 3*, 16–22.

24. Gevaert, K., Goethals, M., Martens, L., Van Damme, J., Staes, A., Thomas, G. R., and Vandekerckhove, J. (2003) Exploring proteomes and analyzing protein processing by mass spectrometric identification of sorted N-terminal peptides, *Nature Biotechnology 21*, 566–569.

25. Staes, A., Van Damme, P., Helsens, K., Demol, H., Vandekerckhove, J., and Gevaert, K. (2008) Improved recovery of proteome-informative, protein N-terminal peptides by combined fractional diagonal chromatography (COFRADIC), *Proteomics 8*, 1362–1370.

Chapter 16

N-Terminomics: A High-Content Screen for Protease Substrates and Their Cleavage Sites

John C. Timmer and Guy S. Salvesen

Abstract

Proteases play vital roles in many cellular processes and signaling cascades through specific limited cleavage of their targets. It is important to identify what proteins are substrates of proteases and where their cleavage sites are so as to reveal the molecular mechanisms and specificity of signaling. We have developed a method to achieve this goal using a strategy that chemically tags the substrate's alpha amine generated by proteolysis, enriches for tagged peptides, and identifies them using liquid chromatography-coupled tandem mass spectrometry (LC-MS/MS). Peptide MS/MS data are searched against a database to reveal what proteins are cleaved, whereby peptide N-termini demarcate sites of protease cleavage.

Key words: Cleavage site, protease, substrate, N-terminomics, biotin labeling, signaling.

1. Introduction

Proteases participate in cellular signaling by cleaving protein substrates. The functional outcome of substrate cleavage is particular to the nature of the substrate itself, and the site of proteolysis (1–3). Proteases generally have more than one physiological substrate, and they cleave these proteins at specific sites. Therefore, it is of vital importance to identify the biological substrates and cleavage sites of proteases so as to elucidate the molecular mechanisms of proteolytic processes. To this end, we have developed N-terminomics, a method to identify proteolytic substrates and cleavage sites from biologically rich samples using mass spectrometry. This approach focuses the power of proteomics on the complement of cleaved substrates, bringing biologically relevant proteolytic signaling within reach.

K. Gevaert, J. Vandekerckhove (eds.), *Gel-Free Proteomics*, Methods in Molecular Biology 753,
DOI 10.1007/978-1-61779-148-2_16,

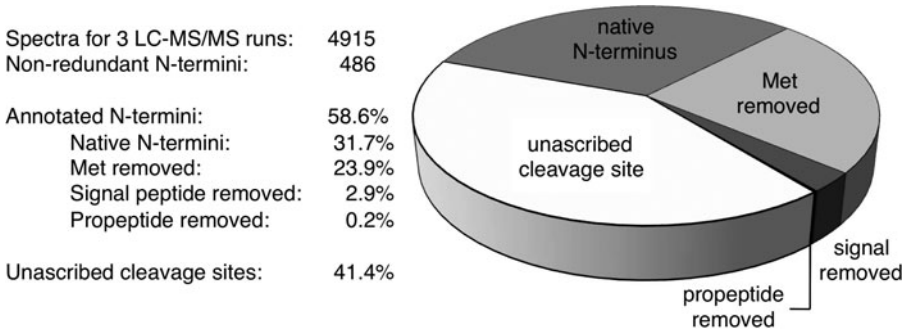


Fig. 16.1. Naturally occurring N-termini from intact and cleaved proteins in *Escherichia coli*. Approximately 60% of the non-redundant N-termini identified from a typical N-terminomic analysis of proteins in *E. coli* lysate correspond to native N-termini or cleavage sites from housekeeping proteases. The other 40% of N-termini originate from cleavage sites without a known function or instigating protease and are referred to as unascribed cleavage sites. Novel protease cleavage sites will be found in the unascribed group, while previously identified cleavage sites annotated in the Swiss-Prot database will fall into the propeptide removed group.

N-Terminomics can identify cleavage sites from nearly any type of sample containing proteases and substrates. This technique has been applied successfully to investigate the scope and magnitude of housekeeping proteolysis in various organisms (**Fig. 16.1**) (4), and more recently was used to probe the structural tolerance of proteases. In the latter study, the proteases human caspase-3 and staphylococcal GluC were applied to an *Escherichia coli* lysate, representing a natively structured complex proteome, and cleavage sites were identified by N-terminomics (5). These results revealed that in contrast to the longstanding dogma in the field stating that only unstructured loops are cleaved, both of these proteases cleaved *E. coli* proteins in alpha helices as well as flexible loops.

The strategy of N-terminomics centers on tagging the N-terminal amine of substrate cleavage sites without modifying the epsilon amine on lysine side chains. This approach utilizes *o*-methylisourea, which under defined conditions specifically converts lysines to homoarginine, a derivative that is unreactive toward the NHS group of the tag. Proteins are digested into peptides, and tagged peptides are enriched by affinity chromatography using neutravidin resin. The disulfide linker within the tag permits bound peptides to be released quantitatively upon reduction after untagged peptides are washed away. Eluted peptides are then cleaned up, analyzed by LC-MS/MS, and searched against a database for peptide matches with the anticipated modifications. Unlike traditional proteomics that weights protein identifications on the coverage of multiple different peptides, N-terminomics utilizes only cleavage site peptides to identify protein substrates.

The output list of cleavage site peptides is more easily interpreted using software that can compare several experimental and control samples to focus in on the cleavage sites of interest, and in this capacity we have developed a series of scripts called N-TerProt. The procedure is relatively simple, inexpensive, uses commonly available reagents, and thus can be carried out in any competent proteomic facility.

2. Materials

2.1. Reagents

1. *o*-Methylisourea hemisulfate (Acros Organics)
2. AG 501-X8 resin (BioRad)
3. 8 M urea buffer: 8 M urea, 50 mM HEPES pH 7.8, 100 mM NaCl
4. PD-10 columns G-25 M (GE Healthcare), single use
5. Sulfo NHS-SS-biotin (Pierce). Prepare the stock NHS-SS-biotin reagent by dissolving 100 mg of sulfo NHS-SS-biotin with 1 ml of DMF making a 165 mM stock, and store at –20°C.
6. 10x TCEP solution: 50 mM TCEP
7. High-capacity neutravidin agarose resin (Thermo)
8. Disposable 2 ml polystyrene columns (Pierce)
9. Micro-Bio chromatography columns (BioRad)
10. Sep-Pak wash buffer: 4.25% acetonitrile, 0.1% TFA
11. Sep-Pak elution buffer: 50% acetonitrile, 0.1% TFA
12. C_{18} Sep-Pak Vac 6 cc cartridges (Waters), single use
13. Solvent A: 2% acetonitrile, 0.1% TFA
14. Solvent B: 80% acetonitrile, 0.1% TFA

2.2. Equipment

1. LC-MS/MS system
2. SEQUEST Sorcerer (required for N-TerProt analysis)
3. Macintosh computer with OS 10.4 or 10.5 (required for N-TerProt analysis)
4. Microsoft Excel for Mac 2004, 2006, or 2008 (required for N-TerProt analysis)
5. Script Editor version 2.1.2 or 2.2.1 (required for N-TerProt analysis)

3. Methods

3.1. Sample Preparation

Prepare appropriate experimental and control conditions in which the protease(s) of interest are active or inactive/inhibited, respectively. The amount of starting material will depend on the experimental system; however, a reasonable starting range is 1–10 mg of protein for a complex sample, such as an *E. coli* lysate.

1. Inactivate proteases in the sample by adding dry guanidinium hydrochloride to 6 M final concentration (*see* **Note 1**).
2. Add DTT to 10 mM final concentration to reduce disulfides. Vortex the sample well and incubate on a 95°C heat block for 10 min. Cool denatured sample to room temperature (*see* **Note 2**).
3. Alkylate cysteines with 30 mM iodoacetamide (final concentration) in the dark at room temperature for 30 min (*see* **Note 3**).
4. Treat the sample with dry *o*-methylisourea hemisulfate to 0.5 M final concentration to convert lysine residues to homoarginines by guanidination.
5. Add 5 N NaOH to adjust the pH to ≥10.3 and monitor with a small electrode probe (*see* **Note 4**).
6. Incubate at 4°C for about 16 h to allow for complete lysine modification without reacting with protein N-termini.
7. Load 25 ml of 8 M urea buffer to equilibrate PD-10 columns (*see* **Note 5**).
8. Load the sample onto the PD-10 column (load ≤1.5 ml of sample), then add 8 M urea buffer to 2.5 ml total volume including the sample (*see* **Notes 6**, **7**, and **8**).
9. Collect the desalted sample by adding 2.5 ml of 8 M urea buffer and saving the elution.
10. Tag N-terminal amines with 5 mM sulfo NHS-SS-biotin for 1 h at 37°C. Terminate the tagging reaction by adding 1 M stock ammonium bicarbonate, pH 7.8, to reach a 50 mM final concentration.
11. Remove the low-frequency side reactions of the biotin tag with serine and threonine residues by adding 1 M stock hydroxylamine to 40 mM final concentration, and incubate at 37°C for 15 min.
12. Buffer exchange the sample into 8 M urea buffer with a PD-10 column again (*see* steps 7, 8, and 9 of **Section 3.1**). Since the sample volume will be greater than 1.5 ml, it will require two separate buffer exchange runs on the PD-10

column. Load the sample and add buffer to 2.5 ml. Collect the next 2.5 ml (*see* **Note 9**).

13. Dilute the sample to 2 M urea with 50 mM ammonium bicarbonate buffer (*see* **Note 10**).
14. Digest proteins with 5–10 μg of modified sequence grade trypsin. Incubate at 37°C overnight (*see* **Note 11**).
15. Spin down any insoluble material at high speed (10,000×*g*) for 5 min at room temperature.
16. Estimate the amount of high-capacity neutravidin resin needed for all samples (*see* **Note 12**).
17. Equilibrate the high-capacity neutravidin resin with 2 M urea buffer.
18. Add high-capacity neutravidin resin to each sample. Incubate at 4°C with rocking for 30 min to bind biotinylated peptides.
19. Load samples on disposable 2 ml polystyrene columns.
20. Wash extensively (10 column volumes) with 2 M urea buffer 10 times.
21. Wash with 3 column volumes of 50 mM ammonium bicarbonate, pH 7.8, buffer 3 times.
22. Pipet the resin into a 2 ml microcentrifuge tube using 1 ml of 50 mM ammonium bicarbonate buffer.
23. Spin down the resin at 500×*g* for 30 s at room temperature and remove buffer to the 1 ml mark.
24. Cleave the biotin linker by adding TCEP to 5 mM final concentration from a 10x stock solution. Incubate at 37°C with rocking for 30 min.
25. Collect eluted peptides by loading the sample onto a Micro-Bio chromatography column. Spin eluent through at 500×*g* for 30 s at room temperature. Repeat until all of the sample has been loaded on the spin column. Flush the resin with 500 μl of 50 mM ammonium bicarbonate buffer. The eluted peptides are in the flow-through fraction.
26. Cleanup peptides using C_{18} Sep-Pak Vac 6cc cartridges. Condition the columns with 6 ml of isopropanol, equilibrate with 6 ml of 0.1% TFA, and load the samples.
27. Wash five times with 2 ml of Sep-Pak wash buffer.
28. Elute five times with 1 ml of Sep-Pak elution buffer. Samples are usually only in the third elution (*see* **Note 13**).
29. Dry samples in a vacuum centrifuge.
30. Dissolve peptides in 40 μl of solvent A and store at 4°C until analysis by LC-MS/MS.

3.2. LC-MS/MS Analysis

Peptides are first separated by reverse phase liquid chromatography prior to MS/MS. In principle, multi-dimensional chromatography approaches may improve the number of non-redundant peptide identifications, yet our preliminary studies with an initial ion exchange step and a following reverse phase separation have not improved the overall number of peptides identified.

1. Load 5–8 μl of the sample onto the reverse phase column at a flow rate of 10 μl/min, then wash with solvent A for 4 min, and elute over a 2 h linear gradient from 10 to 60% solvent B.
2. Peptides eluting from the column go directly into the mass spectrometer for MS/MS analysis. Run the instrument with a duty cycle consisting of a single MS scan followed by four MS/MS scans and enable dynamic exclusion to manage highly abundant peptides (*see* **Notes 14** and **15**).

3.3. Database Searching

1. Combine the MS/MS spectra from repeat runs of the same sample. Analyze these spectra with SEQUEST Sorcerer using the following modifications: N-terminal cleaved biotin stub = 88 Da (variable), Cys carboxyamidomethylation = 57 Da (fixed), Lys guanidination = 42 Da (variable), and Met oxidation = 16 Da (variable).
2. Search a concatenated forward and reverse database to estimate the false discovery rate (6), and be sure to derive the database from the Swiss-Prot proteome database of the relevant organism. Use a semi-enzyme (usually semi-tryptic) database to allow for protease cleavage sites that are not limited to Lys or Arg residues.
3. The database search results from SEQUEST Sorcerer are reported by protein (ProteinProphet) or by peptide (PeptideProphet). The peptide-centric nature of N-terminomics focuses on the position of peptides within proteins, and not in the coverage of each protein. Therefore, export the PeptideProphet output into Microsoft Excel format. This file lists the peptides identified and the proteins they came from, as well as various other information including probability score and cross-correlation.

3.4. N-TerProt Analysis

1. Export the PeptideProphet results into Excel spreadsheets.
2. Add three additional columns to each sample data set with the following headers and content: “sorcerer” and fill down the unique numerical sample identifier number, “sample” and fill down the unique sample description (shorter is better), and “enzyme” and fill down the enzyme used to generate peptides for MS/MS analysis (usually trypsin).

3. Download the N-TerProt folder from http://www.burnham.org/labs/Salvesen/Salvesen%20Lab%20-%20Links.html.
4. Move the folder "N-TerProt" to "Macintosh HD/users/(username)/library/scripts" and note that this includes a test data set "Sample data."
5. Open the file "N-TerProt V3.0" and run the scripts indicated below.
6. Run "Currate Probability & Xcorr." This condenses the data set to peptides with a probability score of $\geq$0.8 and cross-correlation of $\geq$2.0 and copies the resulting information into a new spreadsheet. These cutoff values can be modified in the script by going to the N-TerProt folder and opening the file "2009.07.19 curate prob & xcorr." Simply change the value for the probability score cutoff variable "probCutOff" or the cross-correlation cutoff variable "xcorrCutoff."
7. Run "Fix unlabelled N-terminal artifacts." This corrects an error resulting from the database search of lower resolution LTQ data that is not seen with Orbitrap LTQ data. Some peptides from protein N-termini are incorrectly assigned by SEQUEST as unlabeled, with an initiator methionine, and also with an unmodified lysine near the N-terminus of the peptide. These peptides are actually labeled with the cleaved biotin at the N-terminus, have their initiator methionine removed, and have a modified lysine. This error is due to the similarity in the molecular weight of these two peptide species, which is not distinguishable within the accuracy of the LTQ instrument.
8. Run "Peptide Stuff." This script creates some new columns with more peptide information that the script will use later.
9. Run "Weird NT & CT Analysis." This part looks for non-enzyme (usually non-tryptic) ends of peptides.
10. Run "NT Index." This script makes an index of the non-redundant N-termini, the non-redundant peptides, and the spectral counts per non-redundant N-terminus.
11. Run "New ES." This part returns columns displaying the samples that each peptide is identified in, as well as the enzyme(s) (trypsin, GluC, etc.), and the number of spectral counts for each respective sample and enzyme.
12. Run "Enzyme & Sample column separation." This will separate the columns derived in the last script into individual columns for each sample/enzyme with the respective spectral counts.

13. Run "Swiss-Prot Stuff." This script gathers information about the substrate and cleavage site from Swiss-Prot and deposits it in several new columns. The "protein name" is hyperlinked to each Swiss-Prot entry, "FT & P1 list" is a list of all the potential annotated proteolytic events for that protein, "protein sequence" is the amino acid sequence of that protein, "obs P1" is the amino acid number of the P1 residue, "obs FT" is the proteolytic feature that matches the "obs P1" residue (if any), "cut site" is the P4–P4′ amino acid sequence of the cleavage site with a hyphen at the scissile bond, "group" shows the annotated cleavage features as they are or as "Unannotated," and "annotation" is the confidence level that Swiss-Prot has designated for the annotated proteolytic event observed (if any).
14. Run "Get Category." This will further categorize the "Unannotated" cleavage sites. For example, a sample that has been treated with trypsin and searched against a semi-tryptic database will produce three types of unascribed peptides. "Bonus-1" peptides have a non-tryptic N-terminus and a tryptic C-terminus, "Bonus-2" peptides are fully tryptic, and "Mystery-1" peptides are tryptic at the N-terminus and non-tryptic at the C-terminus. When searching a non-enzyme database an additional category emerges, "Mystery-2," that is non-tryptic at both ends. When analyzing samples with two different proteases, such as trypsin and GluC, then a small fraction of peptides will be derived from the same cleavage site at the N-terminus, but have a different C-terminus. These N-termini are called "Bonus-0." These sub-categories help to assess the confidence in unascribed cleavage sites.
15. Run "Find unascribed sites." This part simplifies the annotation to "Unascribed" or "Annotated."
16. Run "Negative Control." This shows which cleavage sites are also found in the negative control samples and which ones are specific to the protease-treated samples. The default controls are "L" for lysate, "M" for catalytic mutant, and "I" for inhibited protease; however, customized control samples can be added to the script. Go to the N-TerProt folder and open the file named "2008.09.02 No Neg Control." Add new control descriptions to the variable list "negList" keeping the text in quotes.
17. Run "Unascribed Frequency Distribution." This script creates a new worksheet with the percent frequency distribution of each amino acid for unascribed cleavage sites in the experimental vs shared and control samples. These data can then be plotted in a bar graph format to distinguish amino

acids enriched in the protease-treated sample. The script default is to analyze the P1 position, but can be modified to scan any position from P4 to P4′ by going to the N-TerProt folder and open the file named "2009.07.17 UFD." Change the variable "myPosition" in the third line to the desired position keeping the text in quotes.

18. Run "Statistics." This script makes a new spreadsheet with the number of spectra and non-redundant N-termini identified for each category, either annotated or unascribed, and the total tally as well. Typical MS/MS results from an N-terminomic sample run three times are shown in **Fig. 16.1**.

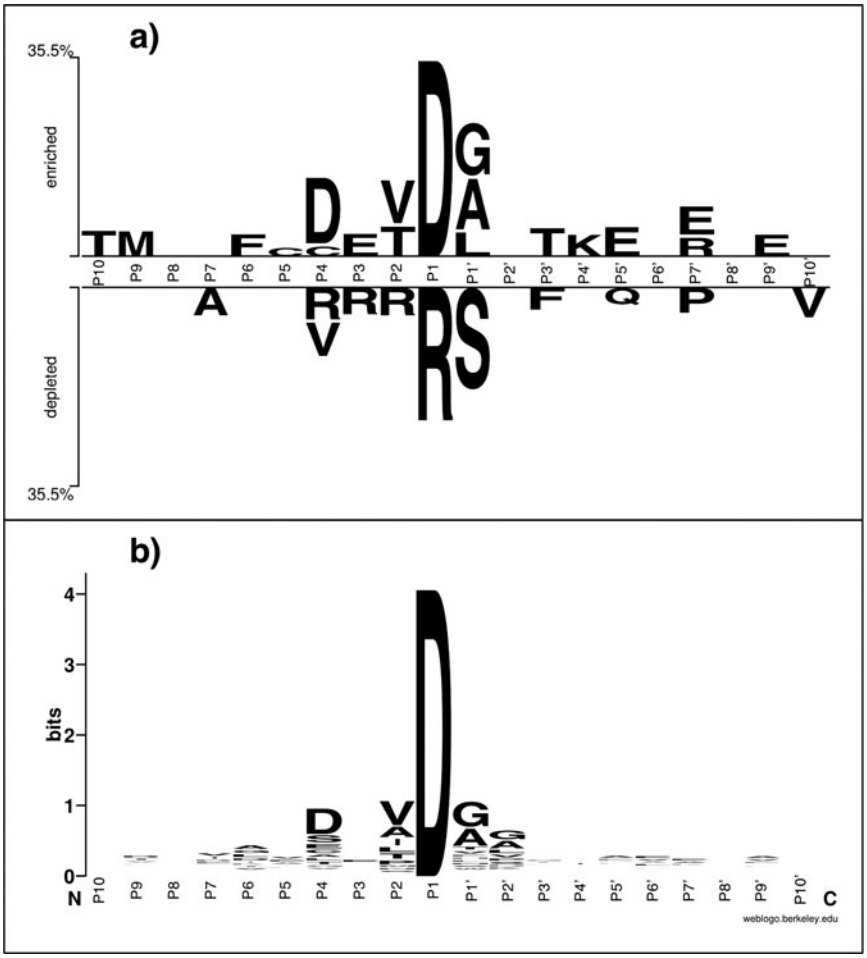


Fig. 16.2. Profiling protease specificity from N-terminomic data. The specificity of active protease(s) in N-terminomic samples can be visualized from the background of cellular proteolysis by using online software such as (**a**) Two Sample Logo or (**b**) WebLogo. In this case, unascribed cleavage sites from the caspase-3-treated lysate were compared with the control lysate and analyzed by Two Sample Logo, which reveals the classical caspase-3 specificity DEVD↓G from P4 to P1′. A refined specificity profile was generated using WebLogo from the unascribed cleavage sites with aspartic acid in position P1 found in the caspase-3-treated lysate.

19. All of the unascribed cleavage sites found only in the experimental samples and not in the control samples ("OK" in the "neg control" column) are candidate substrate cleavage sites. These cleavage sites can be analyzed by online software such as WebLogo (http://weblogo.berkeley.edu/logo.cgi) (7) and Two Sample Logo (http://www.twosamplelogo.org/cgi-bin/tsl/tsl.cgi) (8) to profile the activity of an experimental protease (**Fig. 16.2**). This analysis can also help to identify the protease(s) specifically activated in the experimental samples by matching the observed cleavage preferences with the known specificity for select proteases.

4. Notes

1. When adding dry guanidinium hydrochloride to the sample, assume that the final volume will be about 1.65 times the original volume (for example, a 1 ml sample will become 1.65 ml after the addition of 0.946 g of guanidine HCl).
2. Sample denaturation in 6 M guanidinium hydrochloride and heating to 95°C is required to immediately inactivate proteases in the sample that may elicit artificial cleavages if the cell integrity is disrupted without complete protease inactivation.
3. Keep the stock iodoacetamide protected from light exposure and avoid exposure, as it is toxic.
4. The addition of 100 μl of 5 N NaOH is usually sufficient to adjust the pH of the sample to 10.3 or greater after the addition of *o*-methylisourea hemisulfate.
5. To prepare the 8 M urea buffer, first make an appropriate volume of 9.1 M urea (440 ml is typical). Then add two or three large scoops of AG 501-X8 resin to the urea solution and stir for about 30 min. Deionizing the urea solution removes contaminating salts that may interfere with the NHS-SS-biotin labeling reaction. Next, clarify the urea over a 0.22 μm filter to remove AG 501-X8 resin. Finally, add HEPES, pH 7.8, to 50 mM and NaCl to 100 mM final concentration (50 ml of 1 M HEPES and 10 ml of 5 M NaCl into 440 ml of urea solution results in 50 mM HEPES, 100 mM NaCl, and 8 M urea).
6. As an example of how to load the PD-10 column, if the sample is 1.5 ml, then let that load onto the column, then add 1 ml of 8 M urea buffer.

7. Protein recovery from the PD-10 column can be monitored by SDS-PAGE. However, the starting sample in 6 M guanidinium hydrochloride is not compatible with SDS and must be precipitated by TCA.
8. Protein and peptide recovery can also be quantitated by the Bradford assay, using reagents such as Coomassie Plus (Thermo). The assay is done in a Costar 96-well flat bottom plate and requires 50 μl of sample, or a dilution thereof, and 150 μl of Coomassie Plus reagent. A bovine serum albumin standard should be run each time in the range of 66.7 μg/ml, with dilutions down to 5.9 μg/ml (2/3 dilution series). Leave the last well as a blank. A dilution series for each sample is also useful to ensure that values fall within the linear range of the standard. After mixing, the absorbance is measured in a plate reader at 595 nm. Calculate the concentration of protein/peptide in the samples using the standard curve equation and the appropriate sample dilution factor.
9. The second buffer exchange gets rid of any excess biotin tag that has not reacted with protein N-terminal amines.
10. Although diluting the 8 M urea sample to 2 M urea results in a large sample volume, it minimizes the precipitation of proteins in the sample as they transition back to less denaturing conditions.
11. The extent of trypsin digestion can be confirmed by running equivalent amounts of sample on SDS-PAGE before and after digestion.
12. High-capacity neutravidin binds 75 μg of biotin *para*-nitrophenyl ester per ml of resin (*p*-NPE has a molecular weight of 365) or 0.21 μmol/ml of resin. Assuming the average peptide has a molecular weight of 2,500, this equates to 514 μg of biotinylated peptide per ml of resin. Generally 500 μl is sufficient for each sample.
13. Eluted peptides from the Sep-Pak can be quantitated by using 5,5′-dithio-bis(2-nitrobenzoic acid) (DTNB) to measure free sulfhydryls resulting from cleavage of the biotin tag at the N-terminus of each peptide. For the DTNB assay, make the following stock reagent: 50 mM sodium acetate, 2 mM DTNB in water, and store at 4°C. Also, make 1 M Tris-HCl pH 8.0. Each assay reaction is done in a final volume of 100 μl and is composed of 80 μl water, 10 μl of 1 M Tris-HCl pH 8.0, 5 μl of DTNB stock reagent, and 5 μl of sample. Use a Costar 96-well plate and read the absorbance at 412 nm. A standard curve of TCEP should be used, ranging from 100 μM down to 5 μM. The total microgram of peptide can be estimated from the

results of the DTNB assay that fall within the linear range of the TCEP standard curve and by assuming a molecular weight of 2,500. It is typical to recover 50–100 μg of peptide from a sample of *E. coli* lysate; however, eukaryotic samples may result in lower quantities as a result of protein N-terminal acetylation.

14. Although we have had success with the LTQ and the Orbitrap LTQ MS/MS instruments, other comparable mass spectrometers will almost certainly work as well.
15. Due to the incomplete nature of peptide sampling, repeated runs of each sample improve the number of non-redundant peptides. Three replicate runs of each sample is generally enough to cover 90% of the peptides in a sample, while minimizing instrument time and computational resources (9). This is a nice compromise that will increase the number of non-redundant peptide identifications and improve the number of spectral counts, which serve as a semi-quantitative reference when comparing samples.

Acknowledgments

We would like to thank Mari Enoksson, Wenhong Zhu, and Eric Wildfang for their effort and expertise, which was essential for the development of N-terminomics. This work was supported by the US National Institutes of Health (NIH) Roadmap Initiative National Biotechnology Resource Center grant RR20843 for the Center on Proteolytic Pathways, CA69381 from the National Cancer Institute (NCI), RPG0024/2006 from the Human Frontiers Science Program, and by Training Grant 5T32CA77109-9 from the NCI.

References

1. auf dem Keller, U., Doucet, A., and Overall, C. M. (2007) Protease research in the era of systems biology, *Biol Chem* ***388***, 1159–1162.
2. Fuentes-Prior, P., and Salvesen, G. S. (2004) The protein structures that shape caspase activity, specificity, activation and inhibition, *Biochem J* ***384***, 201–232.
3. Timmer, J. C., and Salvesen, G. S. (2007) Caspase substrates, *Cell Death Differ* ***14***, 66–72.
4. Timmer, J. C., Enoksson, M., Wildfang, E., Zhu, W., Igarashi, Y., Denault, J. B., Ma, Y., Dummitt, B., Chang, Y. H., Mast, A. E., Eroshkin, A., Smith, J., Tao, W. A., and Salvesen, G. S. (2007) Profiling constitutive proteolytic events in vivo, *Biochem J* ***407***, 41–48.
5. Timmer, J. C., Zhu, W., Pop, C., Regan, T., Snipas, S. J., Eroshkin, A. M., Riedl, S. J., and Salvesen, G. S. (2009) Structural and kinetic determinants of protease substrates, *Nat Struct Mol Biol* ***16***, 1101–1108.
6. Elias, J. E., and Gygi, S. P. (2007) Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry, *Nat Methods* ***4***, 207–214.

7. Crooks, G. E., Hon, G., Chandonia, J. M., and Brenner, S. E. (2004) WebLogo: a sequence logo generator, *Genome Res 14*, 1188–1190.
8. Vacic, V., Iakoucheva, L. M., and Radivojac, P. (2006) Two sample logo: a graphical representation of the differences between two sets of sequence alignments, *Bioinformatics (Oxford, England) 22*, 1536–1537.
9. Elias, J. E., Haas, W., Faherty, B. K., and Gygi, S. P. (2005) Comparative evaluation of mass spectrometry platforms used in large-scale proteomics investigations, *Nat Methods 2*, 667–675.

Chapter 17

Protease Specificity Profiling by Tandem Mass Spectrometry Using Proteome-Derived Peptide Libraries

Oliver Schilling, Ulrich auf dem Keller, and Christopher M. Overall

Abstract

Protease specificity profiling using proteome-derived, database-searchable peptide libraries is a novel approach to define the active site specificity of proteolytic enzymes we call PICS (Proteomic Identification of protease Cleavage Sites). Proteome-derived peptide libraries are generated by trypsin, GluC, or chymotrypsin digestion of biologically relevant proteomes, such as cytosolic lysates, to generate three separate libraries that each differ from the others in their C-terminal amino acid residues according to the protease specificity. Primary amines of all peptides are then chemically protected so that after incubation with a test protease, the neo-N-termini of the prime-side cleavage products with exposed α-amines can be specifically biotinylated, enriched, and identified by liquid chromatography-tandem mass spectrometry. The corresponding nonprime-side sequences are derived bioinformatically. Suited for all protease classes except carboxyproteases and those aminoproteases and dipeptidases requiring a free α-amine for cleavage, PICS simultaneously profiles the specificity of prime and nonprime positions and directly determines scissile peptide bonds of up to hundreds of cleavage site sequences in a single experiment. This wealth of sequence specificity information also allows for the investigation of subsite cooperativity. Herein we describe a simplified procedure to produce PICS peptide libraries, the methods to perform a PICS assay, and a new method of data analysis.

Key words: Protease profiling, active site specificity, protease specificity, peptide library, subsite cooperativity.

1. Introduction

Active site specificity profiling of proteases is fundamental for biochemical enzyme characterization and is a prerequisite for the design of specific peptide cleavage assays and inhibitor drugs. Genetic and biochemical profiling techniques – including substrate phage display, positional scanning synthetic combinatorial

K. Gevaert, J. Vandekerckhove (eds.), *Gel-Free Proteomics*, Methods in Molecular Biology 753,
DOI 10.1007/978-1-61779-148-2_17,

peptide libraries, and variants thereof (reviewed in (1)) – have been widely used to study protease specificity. However, substrate phage display is subject to iterative rounds of selection that restrict throughput and typically requires follow-up experiments to locate the exact cleavage site in a cleavable sequence. On the other hand the chemical architecture of positional scanning synthetic combinatorial peptide libraries typically restricts profiling to either prime or nonprime specificity. Hence, none of these techniques can provide both the prime- and the nonprime-side sequences of a cleavage site in the same experiment. To address this deficiency, we recently introduced proteome-derived, database-searchable peptide libraries with the aim to profile a large sequence space while screening simultaneously for prime and nonprime specificity and directly determining the position of the proteolytic cleavage site (2).

Termed PICS (Proteomic Identification of protease Cleavage Sites), this technique generates biologically relevant and diverse peptide libraries from proteomes that are then digested with the test protease. To do so, proteomes such as cell lysates are digested into peptides by proteases such as trypsin, followed by chemical protection of sulfhydryls and dimethylation of primary amines. This proteome-derived peptide library is employed as substrate for a protease activity assay. Prime-side cleavage products possess a free α-amine at the newly formed (neo)-amino (N)-termini. The neo-N-termini of the prime-side cleavage products are then specifically biotinylated, affinity isolated, and identified by liquid chromatography-tandem mass spectrometry (LC-MS/MS). The corresponding nonprime-side sequences are derived bioinformatically by database searches, hence reconstructing the sequence information of the full-length tryptic peptidic substrate. The large number of cleavage site sequences obtained enables robust statistics to be performed to determine the preferred amino acid sequence of the cleavage site and for the study of potential subsite cooperativity. For example, using HIV protease 1, positive cooperativity between leucine in P3 and alanine in P1 was detected in agreement with a previous study (3). PICS has been successfully applied to serine, cysteine, aspartate, and metalloprotease families. Proteases profiled include matrix metalloprotease (MMP)-2, HIV protease 1, caspases 3 and 7, cathepsins K and G, elastase, and thrombin. In all cases, PICS corroborated and refined previous specificity studies. For MMP-2 alone, >1,200 peptidic cleavage sites were identified. Typical results vary between tens and hundreds of cleavage sites.

PICS libraries are generated with trypsin (cleaves C-terminal to lysine and arginine), chymotrypsin (cleaves C-terminal to large hydrophobic residues), or GluC (cleaves C-terminal to glutamate and, to a lesser extent, aspartate). While PICS has been used with proteases from almost all mechanistic classes, it yields best

results for proteases that can be profiled with tryptic peptide libraries. The application of libraries made with different endoproteases enhances sequence coverage, provides validation of cleaved sequences, and allows for the profiling of residues that are otherwise only present at carboxy-terminal positions in libraries generated with a particular endoprotease. For example, tryptic peptides lack internal arginine and lysine residues and so specificity for basic amino acids is profiled using chymotryptic or GluC PICS libraries. PICS libraries contain dimethylated lysine residues and so specificity for lysine cannot be profiled with the current PICS scheme. However, dimethylation preserves the basic character of lysine and some proteases such as MMPs recognize dimethylated lysine as a basic amino acid (2).

An important caveat is that the cleavage sites detected from PICS libraries do not represent physiological cleavage events since most of the PICS cleavage sites are masked in native, folded protein structures. Since PICS employs short peptide sequences it also does not assess exosite contribution to substrate selectivity. Hence, it is unwise to infer physiological substrates from the peptides and proteins identified using PICS.

In this protocol we describe how to generate PICS peptide libraries from proteomes harvested from cultured adherent cells, how to perform a PICS specificity assay, and how to analyze the resulting data. Mass spectrometry is not described in detail as this is routine and involves machine-specific modifications. In comparison to the original PICS publication (2), we present a simplified protocol for peptide library purification and data analysis. While the original protocol employed detergent-assisted cell lysis and a three-step chromatographic purification scheme, the simplified protocol avoids detergents and combines protein precipitation and solid phase extraction for peptide purification. Moreover, a novel Perl computer script is introduced that unifies and automates tasks that were previously handled with multiple computer programs. An overview of the workflow is provided in **Fig. 17.1** and a timeline in **Fig. 17.2**. Exemplary specificity profiles for two proteases (caspases 3 and 7) are shown in **Fig. 17.3**. Recent PICS developments that were published after preparation of this manuscript are described (**Note 1**).

2. Materials

2.1. Library Preparation and Purification

1. Adherent cell lines such as the human fibrosarcoma cell line HT1080 or the 293 human embryonic kidney cell line (**Note 2**).
2. Standard cell culture equipment, including adequate cell growth medium and supplements such as serum and antibiotics.

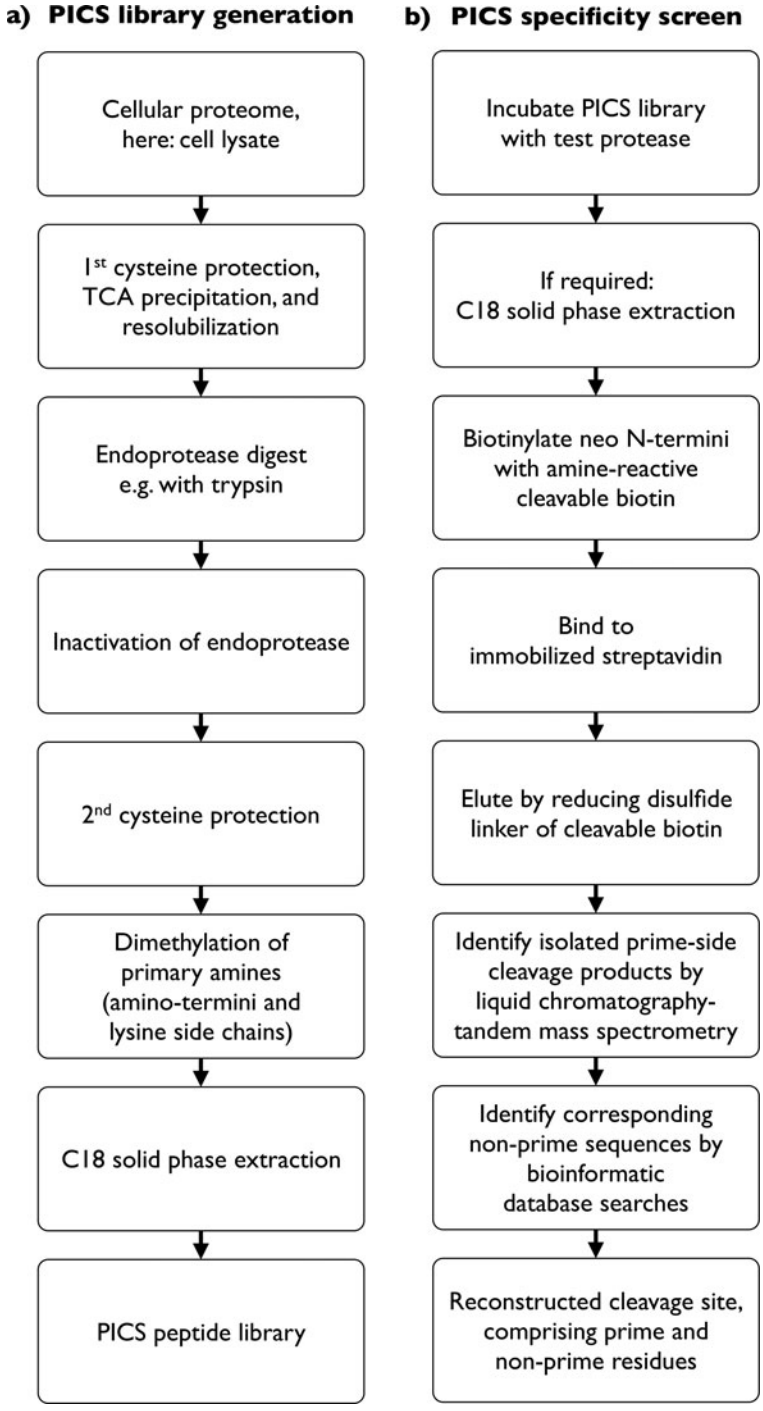


Fig. 17.1. Schematic workflow for PICS library generation from cell lysate (**a**) and PICS specificity assay with a test protease (**b**). For details *see* main text.

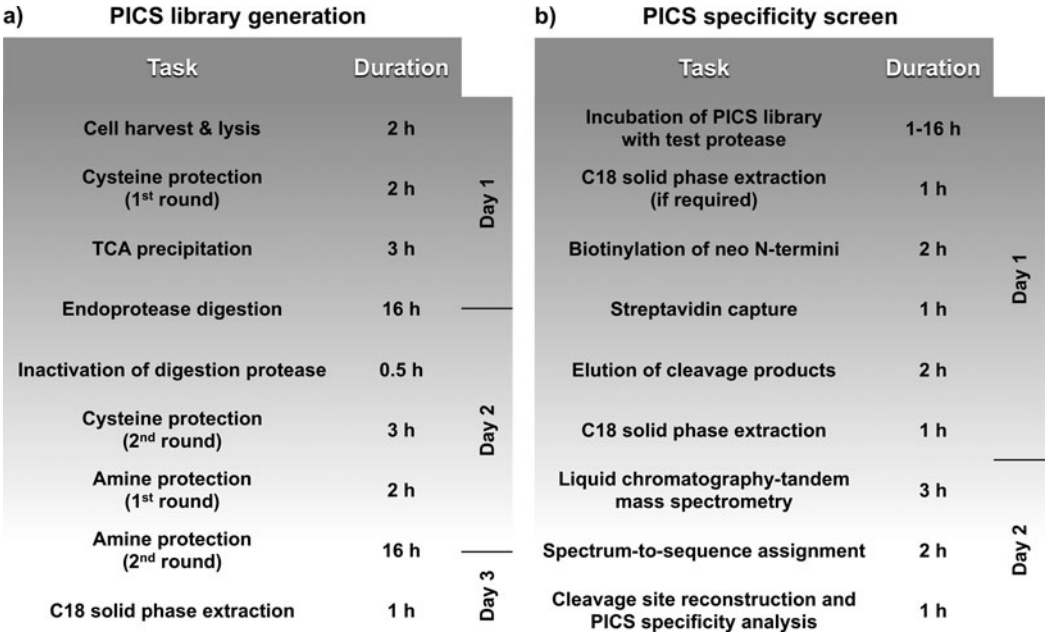


Fig. 17.2. Timeline for PICS library generation (**a**) and PICS specificity assay with a test protease (**b**). For details see main text.

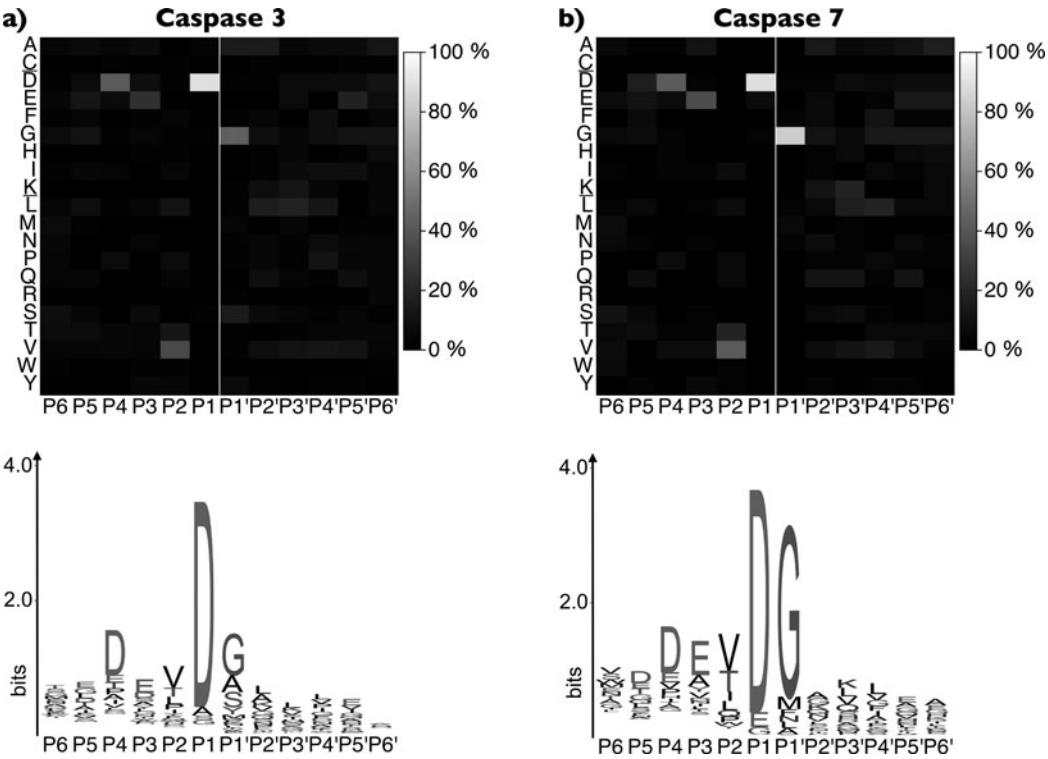


Fig. 17.3. Specificity profiles, displayed as heat maps and sequence logos (12), for caspases 3 and 7 as derived with tryptic PICS libraries (2). The PICS specificity data corroborated the caspase specificity for the nonprime-side motif DEVD and additionally showed a preference for small amino acids (glycine, alanine, serine) in P1′.

3. 0.1 M phenylmethylsulfonyl fluoride (PMSF) stock solution in dimethyl sulfoxide (DMSO).
4. 1.0 M 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES), pH 7.5, stock solution.
5. Hypotonic lysis buffer consisting of 10 mM HEPES, pH 7.5, 1.0 mM PMSF, 10 μM E-64, and 10 mM EDTA. Since PMSF quickly degrades, it must be added directly before use.
6. 1.0 M dithiothreitol (DTT) stock solution.
7. 0.5 M iodoacetamide stock solution.
8. 100% (w/v) trichloroacetic acid (TCA) stock solution.
9. Methanol, pre-cooled to –20°C.
10. 20 mM NaOH, ice cold.
11. Bradford protein assay (Bio-Rad).
12. Chymotrypsin (1-chloro-3-tosylamido-7-amino-2-heptanone treated, Worthington), GluC (V8 protease, Worthington), trypsin (1-chloro-3-tosylamido-4-phenyl-2-butanone treated, Worthington).
13. Formaldehyde, biochemical grade.
14. 1.0 M sodium cyanoborohydride stock solution ("ALD coupling solution," Sterogene).
15. 1.0 M glycine stock solution.
16. Trifluoroacetic acid (TFA).
17. C18 reversed-phase solid phase extraction cartridges (e.g., Sep-Pak, Waters).
18. Acetonitrile.
19. Bicinchoninic acid (BCA) peptide assay (Pierce).

2.2. Enrichment of Cleavage Products

1. Test protease
2. Assay buffer for test protease
3. C18 reversed-phase solid phase extraction cartridges (e.g., Sep-Pak, Waters)
4. 1.0 M HEPES, pH 7.5, stock solution
5. Sulfosuccinimidyl 2-(biotinamido)-ethyl-1,3-dithiopropionate (sulfo-NHS-SS-biotin, Pierce, **Note 3**)
6. High-capacity streptavidin Sepharose (GE Healthcare)
7. Spin column with a filter of ~10 μm pore size
8. Washing buffer: 50 mM HEPES, 150 mM NaCl, pH 7.5
9. Elution buffer: 50 mM HEPES, 10 mM DTT, pH 7.5

10. C18 reversed-phase solid phase extraction cartridges with binding capacity >10 μg, e.g., OMIX tips (Varian)
11. BCA peptide assay

2.3. Identification of Cleavage Products by Liquid Chromatography-Tandem Mass Spectrometry

1. Tandem mass spectrometer in conjunction with a capillary liquid chromatography system
2. C18 resin capillary column
3. Acetonitrile
4. Formic acid

2.4. Bioinformatic Data Analysis

1. MASCOT (4) or X!TANDEM (5) software for spectrum-to-sequence assignment
2. Conversion utility for mzXML format (6)
3. Trans-Proteomic Pipeline software (7), in particular the PeptideProphet (8) utility
4. Perl interpreter software
5. PICS analysis Perl script (downloadable at www.clip.ubc.ca/resources/clippics.html)
6. Proteome data file of the organism from which the peptide library was generated; the PICS Perl script is known to function with data files of the International Protein Index (9)

3. Methods

3.1. Library Preparation and Purification

1. Grow cells to confluence. Depending on cell type and lysis efficiency, a 150–175 cm^2 cell culture flask yields 2–6 mg of final, purified peptide library and we recommend to start this procedure with cell lysate from at least three 150–175 cm^2 cell culture flasks.
2. Detach cells using protease-free detachment methods, such as PBS supplemented with 0.2% (w/v) EDTA. Avoid scraping of cells since this can result in mechanical cell lysis and reduced yields of final purified peptide library.
3. Remove detachment buffer by centrifuging cells at 400×*g*, 4°C for 5 min; decant supernatant and keep cell pellet.
4. Resuspend cell pellet in hypotonic lysis buffer.
5. Lyse cells by repeated aspiration (15 or more times) of the cell suspension with a 27-gauge needle.
6. Centrifuge lysate at 20,000×*g* for 20 min and collect the supernatant.

7. Determine protein concentration and total protein amount using the Bradford assay method according to the manufacturer's instructions.
8. Adjust to 100 mM HEPES, pH 7.5.
9. Add 5 mM DTT (final concentration) and incubate for 60 min at 25°C (**Note 4**).
10. Add 20 mM iodoacetamide (final concentration) and incubate in the dark for 60 min at 25°C.
11. Add another 5 mM DTT (final concentration, accumulated DTT concentration is now 10 mM) and incubate for 5 min at 25°C for 1 h to quench excess iodoacetamide.
12. Adjust to 15% (v/v) TCA and incubate on ice for 1 h.
13. Centrifuge at 20,000×*g* at 4°C for 10 min or at 8,500×*g* at 4°C for 1 h.
14. Wash pellet twice with minute amounts of –20°C cold methanol. If pellet loosens, briefly centrifuge. Briefly air-dry pellet after the second wash step.
15. Overlay pellet with ice-cold 20 mM NaOH. Based on the total protein amount determined in **Section 3.1**, Step 7, assume all protein was precipitated and will entirely re-dissolve and so use sufficient volume to reach a 2.0 mg/ml protein concentration.
16. Ultrasonicate in a cooled ultrasonication bath (use ice slurry to keep sample cold) for up to 60 min. The ultrasonication time can be doubled if the protein pellet is hard to re-dissolve.
17. Bring to 200 mM HEPES, pH 7.5.
18. Centrifuge at 20,000×*g* for 10 min or at 8,500×*g* for 30 min and collect the supernatant.
19. Determine the protein concentration and total protein amount using the Bradford assay method.
20. Digest with trypsin, chymotrypsin, or endoprotease GluC. Use a protease to proteome ratio of 1:100 (w/w) and incubate at 37°C for 16 h.
21. After protease digest, incubate at 70°C for 20 min to stop digestion. Let it cool to 25°C before continuation.
22. Add 1 mM PMSF (final concentration) to entirely abolish activity of the digestion protease.
23. Add 1.0 M guanidine hydrochloride (final concentration) and centrifuge at 20,000×*g* at 4°C for 10 min or at 8,500×*g* for 1 h. Continue with supernatant.
24. Add 5 mM DTT (final concentration) and incubate at 37°C for 1 h.

25. Add 40 mM iodoacetamide (final concentration) and incubate at 37°C for 1.5 h.
26. Add 15 mM DTT (final concentration, accumulated concentration is now 20 mM) and incubate at 37°C for 10 min.
27. Add 30 mM formaldehyde and 30 mM sodium cyanoborohydride (final concentrations) and incubate at 25°C for 2 h.
28. Add another 30 mM formaldehyde and 30 mM sodium cyanoborohydride (final concentrations, accumulated concentrations are now 60 mM) and incubate at 25°C for 16 h.
29. Add 100 mM glycine (final concentration) and incubate at 25°C for 0.5 h.
30. Acidify to 0.5% TFA and let it degas.
31. Purify by C18 solid phase extraction (SPE, **Note 5**).
32. After elution, determine peptide concentration by the bicinchoninic acid assay (**Note 6**).
33. Remove acetonitrile from SPE eluate by vacuum evaporation. Check remaining volume regularly. Based on the total peptide amount determined in **Section 3.1**, Step 32, do not exceed a peptide concentration of 2 mg/ml. Once a volume resulting in a peptide concentration of 2 mg/ml has been reached, refill with water to approximately half of the original volume. Repeat this step four times and continue with a calculated peptide concentration of 1.5–2.0 mg/ml.
34. If peptide precipitation occurs, incubate in an ultrasonication bath for up to 3 h.
35. Centrifuge at 20,000×*g* for 10 min or at 8,500×*g* for 30 min and continue with supernatant.
36. Determine final peptide concentration.
37. Store peptide libraries in aliquots of 200–300 μg at –80°C.

3.2. Test Protease Assay and Enrichment of Cleavage Products

1. Thaw peptide library; 200 μg library per protease if no cleanup is required before biotinylation and 400 μg per protease if cleanup is required before biotinylation (*see* **Section 3.2.7**, **Notes 7, 8**, and **9**).
2. Adjust appropriate buffer conditions for the test protease (pH, type of buffer, reducing agents, co-factors). Do not use detergents, carrier protein, or peptide additives.
3. Adjust peptide library concentration to approximately 1 mg/ml.
4. Add active test protease; typical protease to library ratios are 1:1,000–1:50 (w/w) with 1:100 (w/w) usually applied for a first enzyme characterization.

5. Incubate for 1–16 h at a temperature suitable for the protease under investigation (**Note 10**).
6. Heat de-activate test protease.
7. If PICS assay buffer and protease preparation contain neither primary amines such as Tris(hydroxymethyl)aminomethane (Tris) nor reducing substances such as DTT, skip this step. If primary amines or reducing substances are present, perform C18 SPE using cartridges with sufficient peptide binding capacity and follow the manufacturer's instructions. After elution, vacuum-evaporate C18 SPE eluate to near-dryness and reconstitute sample in 200 μl of 200 mM HEPES, pH 7.5.
8. Check that pH is 7–8, adjust if required.
9. Prepare a 10 mM sulfo-NHS-SS-biotin stock solution in DMSO immediately prior to mixing with the PICS assay and biotinylate cleavage products with 0.5 mM sulfo-NHS-SS-biotin (final concentration, **Note 3**).
10. Incubate at 25°C for 2 h.
11. Equilibrate high-capacity streptavidin Sepharose. High-capacity streptavidin Sepharose is typically supplied as a diluted slurry (the commercially available 5 ml unit is provided as a 1:4 dilution resulting in a total volume of 20 ml). Equilibrate Sepharose resin by repeated centrifugation and resuspension in 50 mM HEPES, 150 mM NaCl, pH 7.5.
12. Add buffer-equilibrated streptavidin Sepharose to the biotinylated PICS assay; the required volume of undiluted slurry is 1.5 times the volume of the PICS assay.
13. Incubate for 0.5 h at 22°C. Apply mild agitation to keep the slurry in suspension.
14. Pour slurry in a spin column of sufficient volume with a filter of ~10 μm pore size (in a typical PICS experiment, the total volume at this point is 550–600 μl).
15. Place the column in a 2 ml reaction tube. Centrifuge the column at a centrifugation speed that is sufficiently high to pass the PICS assay solution through the column without letting the column run completely dry. After centrifugation, the resin should still be wet but without supernatant. Since resin packing can vary, we recommend first centrifuging the column at 150×*g* for 1 min prior to sample addition. If this centrifugation speed proves to be insufficient, increase centrifugation time or speed in small increments. During every centrifugation, cover column top loosely with the supplied screw cap to stop any cross contamination.
16. Re-apply the first flow-through to the streptavidin Sepharose resin and centrifuge a second time.

17. Discard flow-through.
18. Wash resin by applying 500 μl of washing buffer.
19. Centrifuge column and discard flow-through.
20. Repeat washing eight times.
21. Prepare fresh elution buffer.
22. Press bottom plug-in column.
23. Apply 500 μl of elution buffer.
24. Close column top with screw cap.
25. Incubate 2 h at 25°C with mild agitation.
26. Place column in a clean 2 ml reaction tube.
27. Remove bottom plug and loosen screw cap, then centrifuge and keep eluate.
28. Place column in a second clean 2 ml reaction tube and apply 500 μl of elution buffer.
29. Centrifuge and keep eluate and pool the two elution fractions: the estimated eluate amount is 1–5% of library amount (2–10 μg for 200 μg library) depending on biotinylation efficiency as well as protease activity and specificity.
30. Perform a SPE of eluate using C18 cartridges or tips with peptide binding capacity >10 μg. Proceed according to the manufacturer's instructions. Optionally, overlay dried sample with 5 μl of 1 mM DTT. Note that the peptide concentration can only be determined in the absence of DTT (*see* next step).
31. Determine peptide concentration by BCA assay (**Note 6**).
32. Samples can be stored at –80°C for several weeks.

3.3. Identification of Cleavage Products by Liquid Chromatography-Tandem Mass Spectrometry

Analyze the PICS pullout by liquid chromatography-tandem mass spectrometry (LC-MS/MS) (**Note 11**). PICS experiments have been successfully analyzed with typical settings and devices used for LC-MS/MS-based analysis of peptide mixtures derived from biological samples. Electrospray quadrupole time-of-flight (Q-TOF) mass spectrometers and Orbitrap mass spectrometers coupled to capillary liquid chromatography systems have been employed. The peptide mixture is typically separated on a C18 resin capillary column using a 5.0–40% gradient of 80% of acetonitrile in 0.1% formic acid over a period of 90 min. Typical settings for a Q-TOF mass spectrometer include information-dependent acquisition based on a 1 s MS survey scan followed by up to three MS/MS scans of 3 s each. To increase the number of peptide ion spectra collected, each sample can be measured twice to cover complementary mass ranges, e.g., 300–700 and

700–1,500 *m/z*. Inclusion of singly charged precursor ions for MS/MS analysis can be a possibility to increase the number of peptide identifications from PICS experiments employing non-tryptic peptide libraries.

3.4. Bioinformatic Data Analysis

1. Convert raw LC-MS/MS data to a format compatible with search engines such as Mascot (4) or X!Tandem (5) (**Note 12**). Vendor-specific converters are described in detail at http://tools.proteomecenter.org/wiki/index.php?title=Formats:mzXML.
2. Perform spectrum-to-sequence assignment searches with the following parameters (Parameter files for Mascot and X!Tandem can be downloaded at www.clip.ubc.ca/resources/clippics.html (**Note 13**).): carboxyamidomethylation of cysteine residues (+57.02 Da), dimethylation of lysine amines (+28.03 Da), and thioacylation of amino termini (+88.00 Da) as static modifications. Semi-style cleavage searches using the specificity of the enzyme employed for peptide library generation. Allow up to two missed cleavages and set trypsin to cleave C-terminal to either lysine or arginine and chymotrypsin to cleave C-terminal to tryptophan, tyrosine, leucine, and phenylalanine. For GluC, combine results from two searches with definition of cleavage C-terminal to glutamate or glutamate and aspartate (*see* below).
3. Perform secondary validation of search results with the PeptideProphet algorithm (8) provided by the TPP (7). Convert search result files to pep.xml format (http://tools.proteomecenter.org/wiki/index.php?title=Formats:pepXML) and analyze peptides by PeptideProphet, allowing an error (false discovery rate) of 0.05. Export the peptide list as tab-delimited text file ("Export Spreadsheet") with default column settings (interact.pep.xls). If GluC was used for library generation, combine the results from both searches within one tab-delimited text file for downstream analysis.
4. Derive nonprime-side sequences and generate input files for specificity heat maps and sequence logos by processing peptide lists with PICS-Analyzer v2.0, a Perl software script that is available for download at www.clip.ubc.ca/resources/clippics.html. PICS-Analyzer takes interact.pep.xls files and IPI-style Fasta databases as input to derive nonprime-side sequences and subsequently text files (interact.heat.txt, interact.logo.txt) that are suitable for generation of heat maps and sequence logos for visualization of cleavage specificity of the test protease (*see* below). Thereby, cleavage site sequences of up to 12 amino acids spanning the scissile

peptide bond (P6 – P6′) can be analyzed. Details on the software can be found in the accompanying documentation.

5. Generate protease specificity heat maps and sequence logos by using the PICS-Analyzer output file "interact.heat.txt" to generate heat maps with microarray analysis software such as TM4:MeV (www.tm4.org). For meaningful results, we recommend the following settings (**Note 14**): Color Scheme = Rainbow; Color Scale Limits: Lower Limit = min. data value, Midpoint Value = median data value, Upper Limit = max. data value; Element Size = 25×25; Draw Borders = yes. To generate protease specificity sequence logos, paste data from interact.logo.txt into the web form provided at http://www.cbs.dtu.dk/~gorodkin/appl/plogo.html (10, 11) and generate the logo with default settings. Alternatively, use the given alignment as input for the iceLogo software package (http://iomics.ugent.be/icelogoserver/main.html) (12).
6. For subsite cooperativity analysis, run PICS-Analyzer in "cooperativity" mode and set a specific amino acid as fixed in a defined position (**Note 15**). Manually analyze heat maps and sequence logos derived from this analysis for interdependent changes in abundances of specific amino acids in other positions. Set the parameter "Upper Color Scale Limit" in TM4:MeV to the same value for both the initial and the cooperativity analyses. Perform a "vice-versa cross-check" for interdependent amino acids. For example, the subsite cooperativity for alanine in P1 and leucine in P3 for HIV-1 protease (2) involves alanine in P1 increasing the occurrence of leucine in P3 and leucine in P3 increasing the rate of alanine in P1.

4. Notes

1. Recently, the application of bacterial proteomes for PICS specificity screens and a web-based PICS analysis platform have been presented (14).
2. The cell strains used for PICS library generation should be easy to grow and detach with protease-free detachment methods, such as by 0.2% (w/v) ethylenediaminetetraacetic acid (EDTA) phosphate-buffered saline (PBS). Cells must originate from an organism with a completely sequenced genome to allow for a searchable database.
3. Sulfo-NHS-SS-biotin readily hydrolyzes in aqueous solutions. Prepare a 10 mM stock solution in DMSO directly

before use and follow the manufacturer's instructions to preserve integrity of the dry powder.

4. **Section 3.1**, Steps 9–11 of library preparation facilitate TCA pellet resolubilization in **Section 3.1**, Step 16. Cysteine reduction and carboxyamidomethylation can remain incomplete at this point since they are repeated in **Section 3.1**, Steps 24–26.
5. For purification by C18 SPE, several SPE columns can be coupled in tandem to increase binding capacity.
6. For the bicinchoninic acid assay, note that acetonitrile can interfere with this assay and must be removed from the test aliquot by vacuum evaporation.
7. C18 SPE cleanup of the PICS assay is required if the assay buffer contains primary amines or reducing substances.
8. PICS libraries contain modified lysine and cysteine residues and specificity for these amino acids cannot be profiled with the present protocol.
9. PICS libraries must be generated with a digestion protease that has different specificity than the test protease. For example, specificity for acidic residues cannot be profiled with GluC PICS libraries nor can many serine proteases using a tryptic library alone.
10. The extent of proteolysis for a PICS assay can be assessed by using a small aliquot for a fluoraldehyde assay according to the manufacturer's instructions. A negative control should be included, consisting of a PICS library incubated under conditions identical to the PICS test protease assay but without added test protease.
11. Protocols and settings for LC-MS/MS analysis vary widely between different mass spectrometers. The settings presented in this protocol should be adjusted to the actual mass spectrometer used for LC-MS/MS analysis.
12. Data from PICS experiments can be analyzed with any search engine. Secondary validation of search results is not mandatory but highly recommended to avoid false positives.
13. Parameter files provided are suitable for output from a QSTAR (Applied Biosystems) tandem mass spectrometer and settings have to be adjusted to the specific system and database used. From our experience best results are obtained when using native X!Tandem scoring for data from QSTAR systems but the *k*-score plug-in (13) when analyzing Orbitrap output. The scoring algorithm for X!Tandem searches should be changed to "*k*-score" in the provided parameter files when analyzing data generated on an Orbitrap mass spectrometer.

14. The Upper Color Scale Limit has to be adjusted depending on actual results. For example, it could be lower than the maximum data value if predominant preferences in specific positions are observed but more subtle differences in other positions should be visualized.
15. Subsite cooperativity analysis typically requires data sets comprising >300 cleavage sites and should be confirmed in replicate PICS experiments.

Acknowledgments

The authors thank Bettina Mayer for technical assistance in establishing the simplified purification protocol. O.S. acknowledges support from the Deutsche Forschungsgemeinschaft (DFG) (grants SCHI 871/1-1 and 871/2-1) and the Michael Smith Foundation for Health Research (MSFHR). U.a.d.K. was supported by a DFG research fellowship. C.M.O. is supported by a Canada Research Chair in Metalloproteinase Proteomics and Systems Biology. This work was supported by a grant from the Canadian Institutes of Health Research (CIHR) and from a program project grant in Breast Cancer Metastases from the Canadian Breast Cancer Research Alliance (CBCRA) with funds from the Canadian Breast Cancer Foundation and the Cancer Research Society as well as with an Infrastructure Grant from MSHFR.

References

1. Schilling, O., and Overall, C. M. (2007) Proteomic discovery of protease substrates, *Curr. Opin. Chem. Biol.* ***11***, 36–45.
2. Schilling, O., and Overall, C. M. (2008) Proteome-derived, database-searchable peptide libraries for identifying protease cleavage sites, *Nat. Biotechnol.* ***26***, 685–694.
3. Ridky, T. W., Cameron, C. E., Cameron, J., Leis, J., Copeland, T., Wlodawer, A., Weber, I. T., and Harrison, R. W. (1996) Human immunodeficiency virus, type 1 protease substrate specificity is limited by interactions between substrate amino acids bound in adjacent enzyme subsites, *J. Biol. Chem.* ***271***, 4709–4717.
4. Perkins, D. N., Pappin, D. J., Creasy, D. M., and Cottrell, J. S. (1999) Probability-based protein identification by searching sequence databases using mass spectrometry data, *Electrophoresis* ***20***, 3551–3567.
5. Craig, R., and Beavis, R. C. (2004) TANDEM: matching proteins with tandem mass spectra, *Bioinformatics* ***20***, 1466–1467.
6. Pedrioli, P. G. A., Eng, J. K., Hubley, R., Vogelzang, M., Deutsch, E. W., Raught, B., Pratt, B., Nilsson, E., Angeletti, R. H., Apweiler, R., Cheung, K., Costello, C. E., Hermjakob, H., Huang, S., Julian, R. K., Kapp, E., McComb, M. E., Oliver, S. G., Omenn, G., Paton, N. W., Simpson, R., Smith, R., Taylor, C. F., Zhu, W., and Aebersold, R. (2004) A common open representation of mass spectrometry data and its application to proteomics research, *Nat. Biotechnol.* ***22***, 1459–1466.
7. Pedrioli, P. G. A. (2010) Trans-proteomic pipeline: a pipeline for proteomic analysis, *Methods Mol. Biol.* ***604***, 213–238.
8. Keller, A., Nesvizhskii, A. I., Kolker, E., and Aebersold, R. (2002) Empirical statistical

model to estimate the accuracy of peptide identifications made by MS/MS and database search, *Anal. Chem.* ***74***, 5383–5392.
9. Kersey, P. J., Duarte, J., Williams, A., Karavidopoulou, Y., Birney, E., and Apweiler, R. (2004) The international protein index: an integrated database for proteomics experiments, *Proteomics* ***4***, 1985–1988.
10. Gorodkin, J., Heyer, L. J., Brunak, S., and Stormo, G. D. (1997) Displaying the information contents of structural RNA alignments: the structure logos, *Comput. Appl. Biosci.* ***13***, 583–586.
11. Schneider, T. D., and Stephens, R. M. (1990) Sequence logos: a new way to display consensus sequences, *Nucleic Acids Res.* ***18***, 6097–6100.
12. Colaert, N., Helsens, K., Martens, L., Vandekerckhove, J., and Gevaert, K. (2009) Improved visualization of protein consensus sequences by iceLogo, *Nat. Methods* ***6***, 786–787.
13. MacLean, B., Eng, J. K., Beavis, R. C., and McIntosh, M. (2006) General framework for developing and evaluating database scoring algorithms using the TANDEM search engine, *Bioinformatics* ***22***, 2830–2832.
14. Schilling, O., Huesgen, P. F., Barré, O., Auf dem Keller, U., and Overall, C. M. (2011) Characterization of the prime and non-prime active site specificities of proteases by proteomederived peptide libraries and tandem mass spectrometry. *Nat. Protoc.* ***6***, 111–120.

Chapter 18

Identification of Proteolytic Products and Natural Protein N-Termini by Terminal Amine Isotopic Labeling of Substrates (TAILS)

Alain Doucet, Oded Kleifeld, Jayachandran N. Kizhakkedathu, and Christopher M. Overall

Abstract

Determining the sequence of protein N-termini and their modifications functionally annotates proteins since translation isoforms, posttranslational modifications, and proteolytic truncations direct localization, activity, and the half-life of most proteins. Here we present in detail the steps required to perform our recently described approach we call *T*erminal *A*mine *I*sotopic *L*abeling of *S*ubstrates (TAILS), a combined N-terminomics and protease substrate discovery degradomics platform for the simultaneous quantitative and global analysis of the N-terminome and proteolysis in one MS/MS experiment. By a 3-day procedure with flexible α- and ε-amine labeling and blocking options, TAILS removes internal tryptic and C-terminal peptides by binding to a dendritic polyglycerol aldehyde polymer. Therefore, by negative selection, this enriches for both the N-terminal-labeled peptides and all forms of naturally blocked N-terminal peptides. In addition to providing valuable proteome annotation, the simultaneous analysis of the original mature N-terminal peptides enables these peptides to be used for higher confidence protein substrate identification by two or more different and unique peptides. Second, the analysis of the N-terminal peptides forms a statistical classifier to determine valid isotope ratio cutoffs in order to identify with high-confidence protease-generated neo-N-terminal peptides. Third, quantifying the loss of acetylated or cyclized N-terminal peptides that have been cleaved extends overall substrate coverage. Hence, TAILS allows for the global analysis of the N-terminome and determination of cleavage site motifs and substrates for protease including those with unknown or broad specificity.

Key words: Protein N-termini, N-terminome, degradomics, N-terminomics, functional proteomics, proteolysis, protease substrates, acetylated proteins.

K. Gevaert, J. Vandekerckhove (eds.), *Gel-Free Proteomics*, Methods in Molecular Biology 753,
DOI 10.1007/978-1-61779-148-2_18,

1. Introduction

The nature and amount of proteins expressed by a cell, called its proteome, defines the cell type and its functional state. Identifying and quantifying the proteome of cells, tissues, or whole organisms is a challenging task considering the variety – up to several thousands of proteins in each proteome – and large dynamic range of protein concentration in proteomes that sometimes can be up to 12 orders of magnitude in plasma (1). This problem has been tackled by the mass spectrometry-based proteomics community, making great progress in recent years using directed (2) and non-directed (3) approaches. However, the identification and quantification of proteins is not sufficient to define their functional state and other layers of information are required, including knowledge of their posttranslational modifications. This is critically important as posttranslational modifications are many and varied, often switching protein activity or location in a cell.

Proteolysis is a major, irreversible posttranslational modification affecting protein activity, functions, localization, and life span (4). The extent of proteolysis ranges from degradation-to-completion to finely regulated and specific proteolysis such as initiator methionine removal, peptide chain generation from single translation products during protein maturation, export sequence cleavage, elimination of the propeptide for protein activation, and protein processing of mature proteins to switch activity. Proteolysis generates proteins with new N- and/or C-termini not originally present in the initially translated polypeptide. This characteristic can be exploited to identify protease-generated cleavage products and so provides an important layer of functional annotation of the proteome (5). Modification of protein N-termini (other than proteolytic processing) also affects the protein behavior. For instance, about 80% of intracellular proteins possess an acetylated N-terminus (6, 7) that very recently is shown to modify protein half-life in yeast (8), and a variety of other protein N-terminal modifications exist (9).

We have developed a technique we call terminal amine isotopic labeling of substrates (TAILS) to enrich and identify both natural and proteolysis-generated protein N-termini in a single experiment (10). We optimized TAILS using proteins secreted by embryonic fibroblasts in culture, but we have used proteomes from other systems including other cell types, intracellular proteins, tissues, tumors, bronchoalveolar fluid, and platelets. TAILS identifies protease substrates by comparing one proteome that has been subjected to proteolysis to a control proteome, as shown in the workflow depicted in **Fig. 18.1**. In TAILS, protein N-termini from protease-treated and control samples are labeled separately

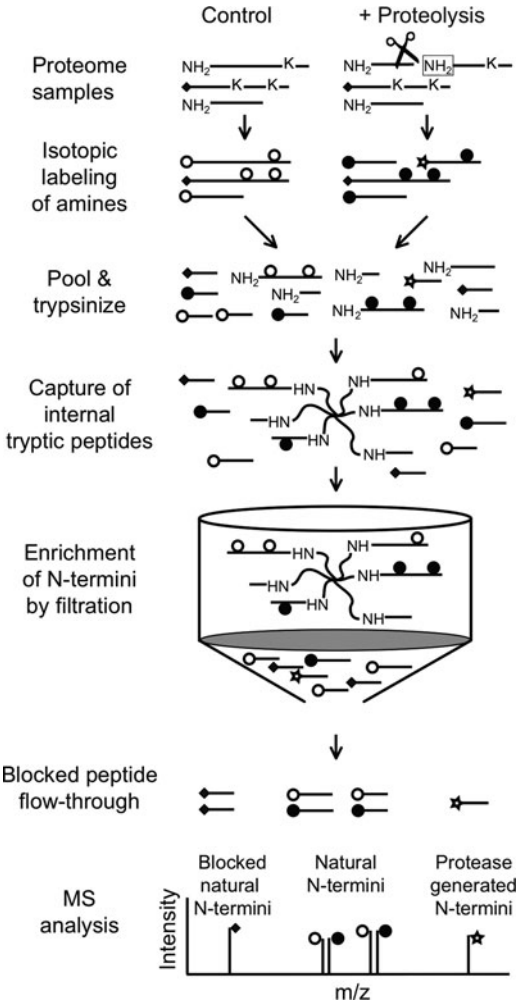


Fig. 18.1. **The TAILS workflow.** Proteins derived from control and protease-treated samples are labeled on the primary amines of the N-termini (NH_2) and lysines (K) via dimethylation using light formaldehyde ($^{12}C^1H_2O$) (*open circles*) and heavy formaldehyde ($^{13}C^2H_2O$) (*solid circles and stars*), respectively. *Black diamond* represents naturally blocked N-terminus. Protease cleavage (represented by *scissors*) generates a proteolytic fragment with a neo-N-terminus labeled by dimethylation with heavy formaldehyde (*star*). After protease inactivation the protease-treated and control proteins are combined and digested with trypsin. The newly formed internal tryptic peptides are removed by reaction of their free N-termini with the amine reactive polymer. Following ultrafiltration separation by centrifugation, the naturally blocked and isotopically labeled/blocked N-terminal peptides are collected in the unbound fraction and are analyzed and quantified in MS1 mode by high-accuracy LC-MS/MS. Protease-generated neo-N-termini are represented by heavy-labeled singletons in the mass spectra (spectral peak labeled with the *star*). These can be easily distinguished from background proteolysis products and original mature N-terminal peptides equally present in both samples and found in doublets having an isotope ratio centered on 1.0 (spectral peak pairs labeled with *open* and *solid circles*). Naturally blocked N-terminal peptides are also found as doublet of equal intensities if lysine residues are present in their sequence. Naturally blocked N-terminal peptides without lysine are found as an unlabeled singleton (peak labeled with a *diamond*). Loss of a naturally acetylated lysine-labeled N-terminal peptide to give a low ratio peptide indirectly indicates cleavage in this sequence.

by reductive dimethylation using formaldehyde-containing stable isotopes. Samples are then combined and digested with trypsin. We circumvented a serious shortfall of other degradomic methods that cannot identify many cleaved proteins from the short prime-side cleavage product. Blocking lysine residues by dimethylation is an important feature that is key to increasing coverage of the truncated shorter semi-tryptic peptides generated by proteolysis; blocked lysine residues are not recognized by trypsin leading to longer peptides, resembling digestion by ArgC. The increased length of many of the neo-N-terminal peptides greatly increases the probability that they can be ionized for MS and identified thereafter with high confidence. Notable of the other substrate degradomic approaches that are N-terminal centric, only COFRADIC also blocks lysines in its workflow (11). Protein N-termini are separated from the internal tryptic peptides using a novel, aldehyde-derivatized, high molecular weight and soluble polymer named HPG-ALD. This negative selection results in an enrichment of the N-terminal peptides. Liquid chromatography-tandem mass spectrometry and bioinformatics analysis of the data identify the protease-generated proteins and reveal the other modifications of the N-terminome. Quantification of the isotopic ratios of peptides enables discrimination of protease-cleaved neo-N-terminal peptides from the background proteolysis products present in the sample and the original mature N-terminal peptides. Here we present a detailed protocol of the TAILS method.

2. Materials

2.1. Proteome Preparation Prior to TAILS

1. The mixture of proteins of interest (the proteome).
2. Millipore Amicon Ultra-15 concentrator with a 3 kDa molecular cutoff.
3. 1.0 M 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) buffer, pH 7.0, stock solution.
4. Protein quantification assay such as bicinchoninic acid titration assay.

2.2. Isotopic Labeling of Samples and Digestion

1. 8.0 M guanidine hydrochloride (GuHCl) in water.
2. 1.0 M NaOH and 1.0 M HCl stock solutions.
3. pH paper strips with a pH range between 5 and 10.
4. 1.0 M dithiothreitol (DTT) in water. This solution can be aliquoted and stored at –20°C.
5. 0.5 M iodoacetamide in water, freshly prepared and protected from light.

6. 12.3 M light formaldehyde ($^{12}C^{1}H_2O$) and 6.6 M heavy formaldehyde ($^{13}C^{2}H_2O$) stock solutions.
7. 1.0 M sodium cyanoborohydride ($NaCNBH_3$) in water stock solution (*see* **Note 1**).
8. 1.0 M ammonium bicarbonate buffer in water.
9. 100 ml of acetone and 100 ml of methanol pre-cooled to –20°C. Organic solvents cannot be stored in a regular freezer due to the explosion hazard.
10. 1.0 M HEPES buffer pH 8.0 stock solution.
11. Dialysis cassette with a 10 kDa molecular weight cutoff.
12. HPG-ALDII polymer at 35 mg/ml: HPG-ALDII is the second of a series of five aldehyde-derivatized polymers with different molecular weights and aldehyde content. HPG-ALDII has a molecular mass of 90 kDa and 516 aldehyde groups/polymer (10). HPG-ALD polymers are available through Flintbox (www.flintbox.ca).
13. Mass spectrometry-grade trypsin.
14. 10% SDS-PAGE gel, Laemmli loading buffer, and SDS-PAGE silver staining solutions.

2.3. Polymer Selection of Peptides with Blocked N-Termini

1. Dialyzed HPG-ALDII polymer.
2. 1.0 M $NaCNBH_3$.
3. 1.0 M NaOH and 1.0 M HCl stock solutions.
4. pH paper strips with a pH range 5–10.
5. 1.0 M ammonium bicarbonate buffer.
6. Microcon spin-filter device with a 10 kDa molecular weight cutoff.
7. Formic acid (99.6%).
8. C_{18} reverse-phase solid phase extraction cartridge with a binding capacity of at least 100 μg. We suggest the Sep-Pack light cartridges from Waters.
9. A solution of 80% acetonitrile, 19.5% water, and 0.5% formic acid (*see* **Note 2**).
10. 0.1% formic acid in water.
11. Liquid nitrogen.
12. Argon gas.

2.4. Identification of N-Terminal Peptides by Liquid Chromatography-Tandem Mass Spectrometry

1. Buffer A: 0.5% acetic acid in water.
2. Buffer B: 80% acetonitrile and 0.5% acetic acid in water.

3. Methods

3.1. Proteome Preparation Prior to TAILS

TAILS is based on the relative quantification of N-terminal peptides from a protease-treated sample and a control sample. This section describes the use of a centrifugal filter device for concentration of proteins and removal of low molecular weight molecules, a method chosen to maintain proteins in a native, folded conformation. Exogenous protease can be added to the concentrated proteome for native substrate identification. If maintaining native protein folding is not a concern, alternative protein concentration methods such as trichloroacetic acid precipitation or solid phase extraction can be used.

1. Collect a minimum of 100 μg of proteins for each sample to be analyzed by TAILS (*see* **Notes 3** and **4**).
2. Apply the protein solution to an ultrafiltration centrifugal filter device such as the Amicon Ultra-15 centrifugal filter device with a 3 kDa molecular cutoff and concentrate the proteins according to the manufacturer's protocol. Never dry the proteins or reduce the volume to less than 1 ml.
3. Add 14 ml of 100 mM HEPES buffer, pH 7.0, to the sample in the concentrator (*see* **Notes 5** and **6**).
4. Centrifuge the sample again to approximately 1 ml.
5. Repeat **Section 3.1**, Step 3, and **Section 3.1**, Step 4, twice (*see* **Note 7**).
6. Measure the protein concentration of each sample.
7. Adjust the protein concentration to 1 mg/ml by concentrating further or diluting the samples with 100 mM HEPES buffer, pH 7.0.

3.2. Isotopic Labeling of Samples and Digestion

Isotopic labeling of N-terminal peptides is required to distinguish between the peptides generated by the protease under study from other N-terminal peptides (protein natural N-terminal and the peptides generated by endogenous proteases in the sample). In this protocol we use a reductive dimethylation labeling strategy (12) that employs normal (light) and isotope-containing (heavy) formaldehyde. This reaction also labels lysine side chains.

1. Add 100 μl of 8.0 M GuHCl to 100 μl (100 μg) of each proteome sample (protease-treated and control) to denature the proteins (*see* **Note 8**).
2. Adjust pH of the samples to 7.0 using 100 mM NaOH or HCl (*see* **Note 9**).
3. Add 1 μl of 1.0 M DTT (5 mM final concentration) to reduce protein disulfide bridges.

4. Incubate at 65°C for 1 h.
5. Add 6 μl of 0.5 M iodoacetamide (15 mM final concentration) to alkylate cysteines.
6. Incubate at room temperature in the dark for 2 h.
7. Add 6 μl of 1.0 M DTT (30 mM final concentration) to quench the excess iodoacetamide.
8. Incubate at room temperature for 30 min.
9. Put the samples in a fume hood (*see* **Note 10**) and add 7 μl of 1.2 M light formaldehyde (regular formaldehyde ($^{12}C^{1}H_2O$)) to the control sample and 7 μl of 1.2 M heavy formaldehyde (formaldehyde containing the isotope ^{13}C and deuterium ^{2}H ($^{13}C^{2}H_2O$)) (40 mM final concentration) to the protease-treated sample (*see* **Note 11**). Labeling swaps can be performed as a control.
10. Add 4.4 μl of 1.0 M $NaCNBH_3$ (20 mM final concentration, *see* **Note 1**).
11. Adjust pH between 6 and 7.
12. Incubate at 37°C for a minimum of 4 h. Incubation overnight is highly recommended.
13. Add 22.5 μl of 1.0 M ammonium bicarbonate (100 mM final concentration) to quench the excess formaldehyde.
14. Adjust pH between 6 and 7.
15. Incubate at 37°C for 4 h (*see* **Note 12**).
16. Combine the light- and heavy-labeled samples in a 15 ml polypropylene tube and vortex.
17. Withdraw 15 μl of the protein solution before precipitation and store this sample at –20°C.
18. Precipitate the remaining labeled proteins by addition of 3.85 ml of cold acetone (8 sample volumes) and 480 μl of cold methanol (1 sample volume), vortex, and incubate at –80°C for a minimum of 1 h.
19. Divide the sample equally between four 2 ml tubes (1.20 ml/tube).
20. Centrifuge the samples at 14,000×*g* for 10 min and discard the supernatant.
21. Add 1 ml of cold methanol to each tube, vortex the samples, centrifuge and discard the supernatants.
22. Repeat **Section 3.2**, Step 21, once (*see* **Note 13**).
23. Air-dry the samples.
24. Re-suspend the dried pellets in 20 μl of 8.0 M GuHCl per each tube. A higher volume can be used if the pellet does

not completely dissolve, but the volume should be kept to a minimum.

25. Add 180 µl (9 volumes) of 50 mM HEPES, pH 8.0, per tube and combine the four tubes. This dilutes the GuHCl and permits trypsin activity in the following steps.
26. Withdraw 25 µl of each sample (after precipitation sample) and store at –20°C.
27. Adjust pH to 8.0 if required.
28. Add 4 µg of mass spectrometry-grade trypsin.
29. Incubate overnight at 37°C.
30. Dialyze 0.5 ml of HPG-ALDII polymer against 4 l of water overnight at room temperature with agitation. This is a preparative step for **Section 3.3**.
31. The next morning, withdraw 25 µl of the sample (after digestion sample).
32. Analyze the "before precipitation," "after precipitation," and "after digestion" samples on a 10% SDS-PAGE gel. Stain gel with silver nitrate. Verify that the same bands appear before and after precipitation, which indicates efficient precipitation and re-solubilization of the sample. No band of molecular weight greater than 10 kDa should be visible after tryptic digestion. If such bands persist, add 2 µg of trypsin to the sample and digest for 6 h. Verify digestion on SDS-PAGE again (*see* **Note 14**).

3.3. Negative Selection of Peptides with a Blocked N-Terminus

This step allows an important enrichment of the peptides with dimethylated and naturally blocked N-termini via the elimination of the non-blocked internal tryptic peptides found in the peptide solution after trypsinization. The non-blocked, N-terminal primary amine group of internal tryptic peptides reacts with and covalently binds to the aldehyde groups on the HPG-ALDII polymer (90 kDa aldehyde-derivatized polymer II). The HPG-ALDII internal tryptic peptide complex is separated from the N-terminally blocked peptides by filtration.

1. Add 400 µg of dialyzed HPG-ALDII (12 µl if the polymer concentration is 35 mg/ml) to the trypsinized sample (*see* **Notes 15** and **16**).
2. Add 32 µl of 1.0 M $NaCNBH_3$ (20 mM final concentration).
3. Adjust pH between 6 and 7.
4. Incubate overnight at 37°C.
5. Add 170 µl of 1.0 M ammonium bicarbonate (100 mM final concentration) to block unreacted aldehyde groups on the polymer.

6. Adjust pH to 6–7.
7. Incubate at 37°C for 30 min.
8. Rinse a Microcon spin-filter with a 10 kDa molecular cutoff with 400 µl of water.
9. Apply 400 µl of the sample to the filter and centrifuge at 14,000×*g* for 15 min. Monitor the volume left in the filter and centrifuge until only a few microliters are left on the filter.
10. Collect the filtrate, which contains the enriched N-terminal peptides. The internal tryptic peptides covalently bound to the polymer are retained on the filter.
11. Add more sample to the same Microcon spin-filter and repeat until all the sample is filtered. Using the same Microcon spin-filter improves recoveries.
12. Wash the filter by adding 200 µl of 100 mM ammonium bicarbonate buffer and centrifuge again.
13. Collect the filtrate and combine it with the filtrate of **Section 3.3**, Step 10 (*see* **Note 17**).
14. Acidify the sample to pH 3 with formic acid and dilute it to 3 ml with 0.1% formic acid in water.
15. Remove all salts and other reagents from the peptides with a C_{18} reverse-phase solid phase extraction cartridge by first conditioning a Sep-Pack light C_{18} cartridge by injecting 5 ml of 80% acetonitrile, 19.5% water, and 0.5% formic acid with a syringe (*see* **Note 2**). Discard the flow-through. Do not introduce air in the cartridge at the end of the injection as this dries the cartridge (always keep the cartridge wet).
16. Rinse the Sep-Pack light C_{18} cartridge with 5 ml of 0.1% formic acid in water and discard the flow-through.
17. Apply the sample to the cartridge at a maximum of 1 ml/min and collect the flow-through. Measure the flow with a timer and the graduation on the syringe.
18. Reapply the sample to the cartridge for maximum peptide binding.
19. Wash the Sep-Pack light C_{18} cartridge with 5 ml of 0.1% formic acid in water and discard the flow-through. Repeat this wash.
20. Elute the peptides with 1.5 ml of 80% acetonitrile, 19.5% water, and 0.5% formic acid at a maximum of 1 ml/min. Collect the eluate.
21. Evaporate the organic solvent under vacuum.
22. Re-suspend the peptides in 20 µl of 3% acetonitrile, 97% water, and 0.1% formic acid. Store the samples at –80°C until mass spectrometry analysis.

3.4. Identification of N-Terminal Peptides by Liquid Chromatography-Tandem Mass Spectrometry

We analyzed the N-terminal peptides enriched by TAILS on quadrupole time-of-flight and LTQ-Orbitrap (Thermo Fisher) mass spectrometers. We prefer the LTQ-Orbitrap mass spectrometer because of its fast duty cycle time and high mass accuracy. Here we describe the conditions used with the LTQ-Orbitrap, but these steps can be easily adapted to any other tandem mass spectrometer. The TAILS protocol was developed without sample pre-fractionation prior to liquid chromatography-mass spectrometry analysis, but such a step can be introduced into the protocol if required to increase coverage.

1. Load peptides on a C_{18} reverse-phase (3 μm ReproSil Pur C18 beads) capillary column (15 cm, 75 mm inner diameter fused silica emitter with a 8 mm diameter opening) with a nanoflow HPLC in line with the mass spectrometer.
2. Elute the peptides from the reverse-phase column with a gradient of Buffer A and Buffer B and inject directly into the mass spectrometer by electrospray ionization. The gradient is formed with 6–30% Buffer B in 60 min, then from 30 to 80% Buffer B in 10 min, and held at 80% of Buffer B for 5 min.
3. Acquire MS1 scans between 350 and 1,500 *m/z* at a resolution of 60,000 and select the five most intense ions for fragmentation. Repeat this cycle for the period of the gradient.

3.5. Bioinformatics Analysis of the Tandem Mass Spectrometry Data

This section covers the analysis of mass spectrometry data at the peptide level, which leads to protease substrate and protein natural N-termini identification. The bioinformatics analysis is influenced by the type of mass spectrometer used as well as the software available. Bioinformatics software and tools evolve rapidly and so different bioinformatics analysis tools might become available. Currently, we perform a database search with two search engines (Mascot and X!Tandem) to select those peptides only identified by two algorithms, and analysis of the data and peptide relative quantification are conducted using the Trans-Proteomics Pipeline (TPP) software from the Systems Biology Institute in Seattle (13). The TPP and X!Tandem are available online, free of charge.

1. Convert LTQ-Orbitrap raw data to mzXML format in profile mode (no centroiding and no deisotoping) with the Petunia interface of the Trans-Proteomics Pipeline software (*see* **Note 18**).
2. Convert the mzXML files to Mascot generic format (.mgf extension) using the TPP.
3. Search the data with the.mgf files against an appropriate database using the Mascot search engine. Separate searches

are performed for the light- and heavy-labeled samples. The search criteria are as follows: (a) semi-ArgC cleavage specificity; (b) three missed cleavages allowed; (c) cysteine carbamidomethylation; (d) peptide N-terminal and lysine dimethylation (light or heavy depending on the search); (e) methionine oxidation defined as a variable (optional) modification; (f) precursor ion mass tolerance of 10 ppm; (g) fragment mass tolerance of 0.8 Da; and (h) scoring scheme is ESI-TRAP.

4. Import the search result files (.dat extension) into the file folder.
5. Convert the .dat file to a pepXML file for both the light- and the heavy-labeled searches.
6. The two peptide lists are then merged and the peptides are scored and quantified using the XInteract, PeptideProphet, and XPRESS tools of the TPP in a single step.
7. Perform the database search with X!Tandem directly from the Petunia interface of the TPP using the *k*-scoring option and repeat **Section 3.5**, Steps 5 and 6, for these searches.
8. Combine the analyzed and quantified data from the two search engines using the iProphet tool of the TPP. This will result in a list of identified and quantified peptides.
9. Select peptides with a peptide probability score corresponding to a ≥99% confidence and eliminate all the others.
10. Select the peptides with an experimental mass within 5 ppm of their theoretical mass and eliminate all the others.
11. Manually verify the quantification data (extracted ion chromatograms from XPRESS) of peptides with undefined ratios and heavy and light singletons and correct if required.
12. Export the final list of peptides to a Microsoft Excel sheet.
13. Perform these steps on two or more biological replicates, which can include technical replicates and labeling swaps.
14. Compare peptide lists and extract the peptides appearing in two replicates or that were identified by two different tandem mass spectra (different charge states, different labels, and oxidized and non-oxidized methionine).
15. Identify peptides with high and low ratios, which are considered as potential substrates of the protease of interest.
16. Perform separate searches to define the natural N-termini of proteins in the samples by replacing the N-terminal dimethylation modification by N-terminal acetylation or N-terminal glutamine to pyroglutamate conversion or

other N-terminal modifications of interest at **Section 3.5**, Step 3.

17. Peptide mixtures obtained before the negative selection can also be analyzed by tandem mass spectrometry to define the labeling efficiency, proteome coverage, and enrichment factor of the negative selection step. Search the tandem mass spectrometry data using N-terminal dimethylation as variable (optional) modification to allow identification of non-labeled tryptic and semi-tryptic peptides.

4. Notes

1. For safety, we recommend buying $NaCNBH_3$ already in solution instead of the solid form. Sterogene Bioseparations Inc. sells this under the product name "ALD coupling solution."
2. Organic solutions should always be prepared and stored in glass containers. Organic solvents extract plasticizers from plastic containers, which results in a massive contamination of plastic polymers in the mass spectrometry analysis. This hampers peptide ionization and reduces proteome coverage. Short contact time (a few minutes) with microtubes, solid extraction cartridges, and syringes are not an issue.
3. The proteome can be from numerous sources such as animal tissue lysates, cultured cell lysates, or cell culture supernatant. The proteome should be from an organism with a sequenced genome.
4. Serum-free, phenol red-free culture media is preferred when cell culture supernatant is used. Albumin is the major protein of cell culture supernatant with serum and interferes with the proteomic identification of other proteins.
5. TAILS is based on the labeling of peptide primary amines and thus molecules with primary amines from other sources will interfere with the labeling step, resulting in incomplete labeling of peptides. Ammonium bicarbonate (used in many proteomic protocols) and Tris buffers contain primary amines and should be avoided. Triethylammonium acetate (TEAA) and triethylammonium bicarbonate (TEAB) buffers do not contain primary amines, but they decompose and generate primary amines when heated.
6. Culture medium contains free amino acids (source of primary amines) that are removed from the samples upon concentration with a filtration device.

7. This step is essential for the elimination of primary amine-containing small molecules that interfere with protein quantification and peptide labeling. More washes can be performed if needed.
8. Plasticizers and other plastic polymers are major contaminants in mass spectrometry-based experiments. The amount of plastic released from microtubes depends on the batch and brand of the tubes. Microtubes from Eppendorf release minimal amount of plastic polymer contaminants and we suggest using this brand for samples to be analyzed by mass spectrometry.
9. To measure pH, apply 1 μl of the sample to a pH paper strip covering pH 5–10. Add 1 μl of NaOH or HCl to the sample and measure the pH. Repeat until pH close to 7 is reached.
10. Formaldehyde is a known carcinogen and $NaCNBH_3$ releases hydrogen cyanide gas during reductive dimethylation, thus sample tubes must be opened in a fume hood.
11. The isotope content of the heavy formaldehyde should be higher than 98% for this experiment to work properly. Formaldehyde with a 99% ^{13}C and 98% ^{2}H content can be purchased from Cambridge Isotope Laboratories, Inc. ^{13}C and ^{2}H are stable isotopes and do not produce radioactivity.
12. This is imperative to quench excess formaldehyde to prevent cross-reaction when samples are combined and with the internal tryptic peptides generated in future steps. Ammonium bicarbonate can be replaced by Tris buffer, hydroxylamine, or other amine-containing molecules, as long as they are not precipitated by acetone/methanol.
13. Washing the acetone with methanol prevents unwanted acetylation of tryptic peptide N-termini in case of a carry-over of $NaCNBH_3$.
14. The GuHCl in the sample may precipitate when mixed with SDS from the Laemmli buffer. Diluting, boiling, and loading warm samples on the electrophoretic gel can reduce the problem. Alternatively, acetone/methanol precipitation of the samples can be performed prior to mixing with the Laemmli buffer. The precipitated protein pellet can be solubilized directly in the Laemmli buffer and loaded on the gel.
15. Prepare 20 μl aliquots of unused dialyzed HPG-ALDII polymer in microtubes, flow argon gas over the liquid, close the microtubes, and freeze the polymer solution in liquid

nitrogen. Stored at –80°C, these polymer aliquots are ready to be used for other experiments. If liquid nitrogen is not used to freeze the polymer solution, it will form a gel-like, opaque solution upon thawing and it will take about 1 h to form a clear, usable solution.

16. HPG-ALDII polymer has a binding capacity of 2.5 mg of peptide per mg of polymer (10). Here, we capture 200 μg of peptide with 400 μg of HPG-ALDII, translating in a fivefold excess of polymer. Other versions of the HPG-ALD polymer have different binding capacities and thus the amount of polymer for the capture should be modified accordingly.
17. Removal of the internal tryptic peptides results in a 90% decrease of the total peptide content, thus a maximum of 20 μg of peptides can be found in the N-terminal enriched sample. The peptide content is in fact lower due to sample loss through the different steps.
18. The mzXML files and all other files generated during the analysis of a data set should be kept in the same folder on the computer for the TPP software to work properly.

Acknowledgments

A.D. acknowledges the Fonds Quebecois de la Recherche sur la Nature et les Technologies and the Michael Smith Foundation for Health Research (MSFHR) for research fellowships. O.K. acknowledge the Centre for Blood Research for a postdoctoral fellowship. C.M.O. is supported by a Canada Research Chair in Metalloproteinase Proteomics and Systems Biology. J.N.K. is a recipient of a CIHR/Canadian Blood Services new investigator award in transfusion science. This work was supported by grants from the CIHR and from a program project grant in Breast Cancer Metastases from the Canadian Breast Cancer Research Alliance with funds from the Canadian Breast Cancer Foundation and The Cancer Research Society as well as with an Infrastructure Grant from the Canada Foundation for Innovation (CFI) and MSHFR to the UBC Centre of Blood Research. We thank Dr. Wei Chen and the CBR mass spectrometry core facility, and Dr. Leonard Foster for proteomic analysis and Dr. Georgina Butler for critical reading and editing of the manuscript.

References

1. Anderson, N. L., and Anderson, N. G. (2002) The human plasma proteome: history, character, and diagnostic prospects. *Mol Cell Proteomics* **1**, 845–67.
2. Picotti, P., Bodenmiller, B., Mueller, L. N., Domon, B., and Aebersold, R. (2009) Full dynamic range proteome analysis of S. cerevisiae by targeted proteomics. *Cell* **138**, 795–806.
3. Washburn, M. P., Wolters, D., and Yates, J. R. (2001) Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat Biotechnol* **19**, 242–47.
4. Walsh, C. T., Garneau-Tsodikova, S., and Gatto, G. J. J. (2005) Protein posttranslational modifications: the chemistry of proteome diversifications. *Angew Chem Int Ed Engl* **44**, 7342–372.
5. Doucet, A., Butler, G. S., Rodriguez, D., Prudova, A., and Overall, C. M. (2008) Metadegradomics: toward in vivo quantitative degradomics of proteolytic posttranslational modifications of the cancer proteome. *Mol Cell Proteomics* **7**, 1925–951.
6. Brown, J. L., and Roberts, W. K. (1976) Evidence that approximately eighty percent of the soluble proteins from Ehrlich ascites cells are Nalpha-acetylated. *J Biol Chem* **251**, 1009–14.
7. Driessen, H. P., de Jong, W. W., Tesser, G. I., and Bloemendal, H. (1985) The mechanism of N-terminal acetylation of proteins. *CRC Crit Rev Biochem* **18**, 281–325.
8. Hwang, C. S., Shemorry, A., and Varshavsky, A. (2010) N-terminal acetylation of cellular proteins creates specific degradation signals. *Science* **327**, 973–77.
9. Meinnel, T., and Giglione, C. (2008) Tools for analyzing and predicting N-terminal protein modifications. *Proteomics* **8**, 626–49.
10. Kleifeld, O., Doucet, A., auf dem Keller, U., Prudova, A., Schilling, O., Kainthan, R. K., Starr, A. E., Foster, L., Kizhakkedathu, J. N., and Overall, C. M. (2010) Isotopic labeling of terminal amines in complex samples identifies protein N-termini and protease cleavage products *Nat Biotechnol* **28**, 281–88.
11. Gevaert, K., Goethals, M., Martens, L., Van Damme, J., Staes, A., Thomas, G. R., and Vandekerckhove, J. (2003) Exploring proteomes and analyzing protein processing by mass spectrometric identification of sorted N-terminal peptides. *Nat Biotechnol* **21**, 566–69.
12. Hsu, J. L., Huang, S. Y., Chow, N. H., and Chen, S. H. (2003) Stable-isotope dimethyl labeling for quantitative proteomics. *Anal Chem* **75**, 6843–852.
13. Keller, A., Eng, J., Zhang, N., Li, X. J., and Aebersold, R. (2005) A uniform proteomics MS/MS analysis platform utilizing open XML file formats. *Mol Syst Biol* **1**, 2005.0017.

Chapter 19

Lectins as Tools to Select for Glycosylated Proteins

Els J.M. Van Damme

Abstract

Glycosylation has been recognized as one of the most important modifications on proteins. The interactions between proteins and glycans are known to play an important role in many biological processes. Lectins are carbohydrate-binding proteins that can specifically interact with and select for carbohydrate structures. The technique of lectin affinity chromatography takes advantage of this specific interaction and enables the selection and purification of glycoproteins with carbohydrate structures complementary to the lectin-binding site. Depending on the carbohydrate specificity of the lectin glycoprotein fractions enriched for example, high mannose or complex N-glycans or O-glycans can be obtained. Afterward both the protein part and the glycan part can be analyzed in more detail allowing the identification of the interacting partners and the type of glycans involved.

Key words: Affinity chromatography, carbohydrate specificity, glycan, glycoprotein, lectin.

1. Introduction

Lectins are a group of non-enzymatic proteins which recognize and bind mono- and oligosaccharides. Since their discovery in plants in the late 1880s, lectins or carbohydrate-binding proteins have been found not only in many plant species but also in animals, microorganisms, and viruses. Hundreds of lectins have been purified and in many cases their carbohydrate-binding properties have been determined. Consequently several lectins with interesting sugar specificities have been identified and some of them have been developed into powerful tools for the purification, separation, and structural analysis of glycoproteins and the analysis of carbohydrate structures on tissues, cells, and proteins (1–3).

K. Gevaert, J. Vandekerckhove (eds.), *Gel-Free Proteomics*, Methods in Molecular Biology 753,
DOI 10.1007/978-1-61779-148-2_19,

Although lectins are available from several sources, plant lectins have attracted most interest, simply because some plant lectins are present in reasonable to (very) high concentrations in seeds or vegetative tissues, which allowed easy purification of the lectin in sufficient amounts. Most of the lectins currently used as tools in glycobiology come from plants and are commercially available. However, some animal lectins are also widely used such as, e.g., the *Helix pomatia* agglutinin (HPA) from the snail and the *Anguilla anguilla* agglutinin (AAA) from the freshwater eel.

At present the group of plant carbohydrate-binding proteins can be subdivided into 12 families (4). Each lectin family is characterized by its own three-dimensional fold for the lectin and the carbohydrate-binding site. Representatives of each lectin family have been studied for their carbohydrate-binding properties. Initially the carbohydrate specificity of most lectins was characterized using hapten inhibition assays, in which monosaccharides, monosaccharide derivatives, or small oligosaccharides are used to block lectin binding to some glycan-coated target such as, e.g., red blood cells. For most lectin families the structure–function relationships – in terms of specific recognition of monosaccharides – are reasonably well understood. However, specificity studies clearly demonstrated that most plant lectins exhibit a much higher affinity for oligosaccharides/complex glycans than for monosaccharides. Whereas the binding affinity of lectins for complex glycans is often in the range of 1–10 μM, the affinity for monosaccharides is only in the millimolar range. The interactions of lectins with glycans are complex and not fully understood. Many lectins recognize terminal nonreducing saccharides, while others also recognize internal sugar sequences. Furthermore most lectins do not have an absolute specificity and therefore can bind to similar carbohydrate structures with different affinities.

Structural analyses not only confirmed that the monosaccharide-binding site of plant lectins usually accommodates a single well-defined sugar unit of a bulky N-glycan chain but also indicated that other amino acid residues located in the vicinity of the primary site participate in the binding of other sugar units so that a more extended carbohydrate-binding site is created.

In recent years the study of proteins (lectins) interacting with carbohydrate structures has been greatly enhanced by the introduction of microarray technology into the field of glycobiology. Different carbohydrate or glycan microarrays have been developed which allow a rapid and comprehensive screening of (a) carbohydrate-binding protein(s) for interaction with a large set of carbohydrate structures and characterization of their carbohydrate-binding properties (5–7). In addition, the glycan array technology also allows quantification of the relative binding of carbohydrate-binding proteins to glycans (8, 9).

Glycan array analyses of plant lectins allowed – by virtue of an unprecedented broad range of test glycans – refining the specificity of many plant lectins and confirmed the preferential binding of most lectins to oligosaccharides and glycans rather than to monosaccharides (http://www.functionalglycomics.org/static/consortium/resources/resourcecoreh8.shtml) (5, 6).

The increasing information related to specificity of lectins in turn led to the development of lectin microarrays for high-throughput analysis of glycans and glycoproteins (10, 11).

The specific binding of a lectin to a carbohydrate structure can be exploited in lectin affinity chromatography. This technique allows to select, for example, glycosylated proteins from different tissue or cell extracts and provides the basis for the use of lectins as tools for the separation and structural analysis of glycoproteins and oligosaccharides. The principle of lectin affinity chromatography is shown in **Fig. 19.1**. In a first step lectins need to be coupled to a matrix taking care not to block the lectin-binding sites. Subsequently this matrix with immobilized lectin can be used for affinity chromatography. Analysis of a protein extract on the column allows for the enrichment of glycosylated structures. Depending on the specificity of the lectin, different selections for glycoconjugates can be achieved. The use of a series of lectin columns with different carbohydrate-binding properties allows the fractionation of oligosaccharides into structurally related subsets and can help to determine the glycan profile for different cells/tissues. Hence, lectin affinity chromatography is a versatile tool for the fractionation of heterogeneous glycan mixtures.

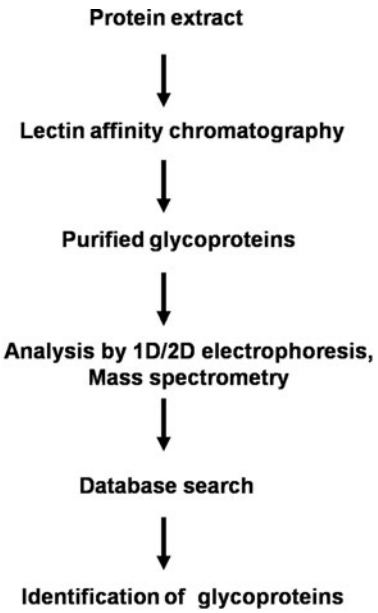


Fig. 19.1. Schematic outline of the procedure for lectin affinity chromatography.

2. Materials

2.1. Lectins for Affinity Chromatography

1. *Galanthus nivalis* (snowdrop) agglutinin (GNA) (12)
2. *Sambucus nigra* (elderberry) agglutinin I (SNA-I) (13)
3. *Maackia amurensis* agglutinin (MAA) (14)
4. *Nicotiana tabacum* (tobacco) agglutinin (Nictaba) (15)
5. *Rhizoctonia solani* agglutinin (RSA) (16)

Lectins can be purified according to the references indicated above. Alternatively, many of these lectins are sold by companies such as EY Laboratories (San Mateo, CA, USA), Sigma-Aldrich Corp. (St. Louis, MO, USA), and Vector Laboratories (Burlingame, CA, USA). In addition to crude and purified lectin preparations, immobilized lectins or lectins conjugated to enzymatic and fluorescent labels, colloidal gold, or biotin are also commercially available (*see* **Note 1**).

2.2. Material and Chemicals for Lectin Coupling

1. Sepharose 4B (GE Healthcare).
2. Carbonate buffers: 0.5 M Na_2CO_3 (pH 11) and 0.5 M $NaHCO_3$ (pH 10).
3. Divinyl sulfone.
4. Ethanolamine.

Warning: Divinyl sulfone is extremely poisonous. All operations should be performed under a fume hood.

2.3. Lectin Affinity Chromatography

1. Immobilized lectin: lectin-conjugated Sepharose 4B packed in an appropriate column.
2. Equilibration buffer: phosphate-buffered saline (PBS) – 135 mM NaCl, 3 mM KCl, 1.5 mM KH_2PO_4, and 8 mM Na_2HPO_4, pH 7.5, or 0.1 M Tris–HCl, pH 7.6, containing 0.2 M NaCl.
3. Elution buffer: 20 mM 1,3-diaminopropane.

3. Methods

Although glycosylated proteins can also be selected using typical immunoprecipitation procedures, the use of a column gives cleaner data and is more appropriate for larger (diluted) extracts or large-scale purifications. The protocol described here only allows for an enrichment of different classes of N-linked and/or O-linked glycoproteins. It is important to select an appropriate (set of) lectin(s) for enrichment of glycoconjugate structures

Table 19.1
Selection of lectins that can be used for isolating glycoproteins and glycopeptides

Lectin	Monosaccharide specificity	Oligosaccharide specificity	References
GNA, *Galanthus nivalis* agglutinin	Terminal mannose	High mannose N-glycans	(19)
MAA, *Maackia amurensis* agglutinin	NeuAc	Complex glycans containing NeuAc(α-2,3)Gal	(14)
RSA, *Rhizoctonia solani* agglutinin	GalNAc	O-glycans and complex N-glycans	(16)
Nictaba, *Nicotiana tabacum* agglutinin	(GlcNAc)n	High mannose and complex N-glycans	(15); (20)
SNA-I, *Sambucus nigra* agglutinin I	NeuAc	Complex glycans containing NeuAc(α-2,6)Gal	(21); (22)

of interest. A selection of lectins that can be used is shown in **Table 19.1**. New lectins with interesting specificities are still being discovered (*see* **Note 2**). The outline of the procedure is shown in **Fig. 19.1**.

3.1. Lectin Coupling to Sepharose 4B

1. Wash Sepharose 4B with 20 volumes of water to allow the matrix to swell and equilibrate with 0.5 M Na_2CO_3 (pH 11) in a sintered glass funnel (*see* **Note 3**).
2. Transfer 50 ml of the settled Sepharose 4B into 50 ml of 0.5 M Na_2CO_3 (pH 11) and add 5 ml of divinyl sulfone to activate the vinyl groups of Sepharose 4B (*see* **Note 4**). Use glass bottles or glass beakers.
3. Incubate at room temperature for 3 h with occasional stirring with a glass rod.
4. Transfer the matrix to the sintered glass funnel.
5. Wash the gel with at least 10 volumes of water and 2 volumes of 0.5 M $NaHCO_3$ (pH 10).
6. Dry the gel (by suction on the funnel) and transfer the required quantity into an equal volume of a solution of the lectin to be coupled (typically use a 5–10 mg lectin solution per ml of gel matrix) in 0.5 M $NaHCO_3$ (pH 10). The complementary sugar can also be included in the lectin solution (*see* **Note 5**).
7. Incubate overnight at 37°C under constant agitation (on a shaker) to allow the proteins to react with the activated Sepharose.

8. Wash the gel with 10 volumes of water and block the unreacted groups with 5 volumes of ethanolamine (*see* **Notes 6** and **7**).

3.2. Sample Preparation

Sample preparation depends on the type of tissue and protein to be extracted. Protein extracts are typically made in phosphate-buffered saline or Tris buffer at neutral pH. For extraction of membrane proteins, detergents should be added to the extraction buffer and of note is that lectin affinity chromatography is compatible with the presence of 1% Triton in the extraction buffer. It is advisable to add a protease inhibitor cocktail to the extraction buffer in order to avoid protein degradation due to the presence of proteases present in the tissue. Large volumes of extract are not a problem for affinity chromatography; however, the sample should be cleared by, e.g., different centrifugation and/or filtration steps to avoid clogging of the lectin column.

3.3. Lectin Affinity Chromatography

1. Equilibrate the lectin-Sepharose column in 5 volumes of PBS or Tris buffer.
2. Apply the extract on the column and collect the flow-through (*see* **Notes 8, 9, 10,** and **11**).
3. Wash the column with 5–10 volumes of PBS (containing 0.1% Triton) (*see* **Note 12**).
4. Elute the glycoproteins from the column with 10 volumes of 1,3-diaminopropane. Collect fractions individually (*see* **Notes 13, 14,** and **15**).

4. Notes

1. Lyophilized lectin preparations can be stored in a freezer at –20°C for several years. Lectin solutions are stable at –20°C for several months.
2. If no information is available on the presence or type of glycan structures on the proteins under study, glycosylation profiles can first be analyzed using the DIG Glycan Detection Kit (Roche Applied Bioscience). The use of this kit allows to determine whether a protein is glycosylated or not. It does, however, not allow to determine the type of glycan involved. The DIG Glycan Differentiation Kit (Roche Applied Bioscience) contains a selection of DIG-labeled lectins that then allow characterization and analysis of carbohydrate structures by the specific binding of selected lectins. Alternatively the determination of the glycan profile can be performed by Far-Western blotting (17).

These techniques allow to select the appropriate lectin to be used in lectin affinity chromatography.

3. For the washing steps, a Büchner filter with very fine pores can be used.
4. Coupling of proteins to a divinyl sulfone-activated matrix has been shown to produce affinity matrices with low leakage (18). Generally, divinyl sulfone-activated agarose reacts with amino, hydroxyl, and sulfhydryl groups, thus allowing immobilization of a wide spectrum of ligands. Unlike CNBr-activated gels, it does not leak the immobilized protein at high pH.
5. The ratio between liquid and ligand solution should be around 1. The sugar in the coupling buffer is added in order to protect the binding site of the lectin during coupling. Usually a monosaccharide that will interact with the lectin is added at a concentration of 0.1 M.
6. After completion of the coupling step, the suspension can be transferred into, e.g., single fritted 25 ml columns (International Sorbent Technology Ltd., IST) and allowed to settle by draining the coupling medium under gravity. An aliquot of the flow-through can be collected and its A_{280} measured to check the coupling efficiency. Under the conditions described here, >98% of the lectins were immobilized to the Sepharose 4B matrix.
7. To block the remaining free activated vinyl groups, the matrix is equilibrated with a 0.2 M solution of Tris–HCl (pH 8.7), kept in this buffer for at least 6 h, and washed with 50 ml 0.2 M NaCl containing 0.01% Na-azide as an antimicrobial agent. Columns can be used immediately or stored at 4°C until use.
8. Analysis of extracts on the lectin affinity columns can be performed under gravity. Alternatively, lectin columns can be prepared in special columns which allow using a chromatography system such as, e.g., AKTA system (GE Healthcare).
9. For optimal performance a suitable ratio of matrix volume to the amount of extracted protein should be used. When using consecutive lectin columns to select for different glycan populations, the binding capacity of each lectin matrix should be large enough. If the binding capacity of the matrix is too small, a significant amount of the glycans will not bind and will contaminate the flow-through fraction.
10. Before application to the lectin column the extract should be cleared by extensive centrifugation and/or filtration to remove all particles from the extract. It is important to start

with a very clear extract, if not the column will be clogged. Typically at least 5 mg of protein should be applied to a 1–2 ml column.

11. If desirable, the flow-through fraction of the first lectin column can be transferred to a second lectin column.
12. *Note*: Triton X-100 can be removed from subsequent steps (washing and beyond) for samples being prepared for mass spectrometry. It is important to wash the column thoroughly to remove aspecifically bound glycoconjugates.
13. Elution with 20 mM diaminopropane will allow to elute all proteins in a small elution volume, simply by increasing the pH. Alternatively, the bound glycoproteins can be eluted using carbohydrate solutions. However, this requires a high concentration of the carbohydrate or glycoprotein to compete with binding on the lectin column.
14. Peak fractions can be pooled after setting the pH to neutral and addition of 0.2 M NaCl can be chromatographed a second time on the same lectin column to ensure selection of only those glycoproteins that react most strongly with the lectin column.
15. The collection of individual fractions allows for the selection (absorption at 280 nm, SDS-PAGE) of the most concentrated fraction for subsequent analysis. Protein concentrations in the eluted fractions can be determined. Eluted fractions can be analyzed by one- or two-dimensional electrophoresis and mass spectrometry for subsequent identification of the protein or characterization of the glycan.

Acknowledgments

The financial support of the Research Council of Ghent University and Fund for Scientific Research-Flanders (G.0022.08) is gratefully acknowledged.

References

1. Van Damme, E. J. M., Peumans, W. J., Barre, A., and Rougé, P. (1998) Plant lectins: a composite of several distinct families of structurally and evolutionary related proteins with diverse biological roles. *Crit Rev Plant Sci* **17**, 575–692.
2. Van Damme, E. J. M., Peumans, W. J., Pusztai, A., and Bardocz, S. (1998) *Handbook of Plant Lectins: Properties and Biomedical Applications*, John Wiley & Sons, Chichester, p. 452.
3. Van Damme, E. J. M., Rougé, P., and Peumans, W. J. (2007) Plant lectins. In: *Comprehensive Glycoscience – From Chemistry to Systems Biology*, J. P. Kamerling, G. J. Boons, Y. C. Lee, A. Suzuki, N. Taniguchi,

and A. J. G. Voragen, eds., Elsevier, New York, volume 3, pp. 563–99.
4. Van Damme, E. J. M., Lannoo, N., and Peumans, W. J. (2008) Plant lectins. *Adv Bot Res* **48**, 107–209.
5. Van Damme, E. J. M., Smith, D. F., Cummings, R., and Peumans, W. J. (2011) Glycan arrays to decipher the specificity of plant lectins. In: *The Molecular Immunology of Complex Carbohydrates*, A. M. Wu, ed., Kluwer Academic/Plenum Publishers, New York, 841–854.
6. Taylor, M. E. and Drickamer, K. (2009) Structural insights into what glycan arrays tell us about how glycan-binding proteins interact with their ligands. *Glycobiology* **19**, 1155–62.
7. Yue, T. and Haab, B. B. (2009) Microarrays in glycoproteomics research. *Clin Lab Med* **29**, 15–29.
8. Blixt, O., Head, S., Mondala, T., Scanlan, C., Huflejt, M. E., Alvarez, R., et al. (2004) Printed covalent glycan array for ligand profiling of diverse glycan binding proteins. *Proc Natl Acad Sci USA* **101**, 17033–38.
9. Paulson, J. C., Blixt, O., and Collins, B. E. (2006) Sweet spots in functional glycomics. *Nat Chem Biol* **2**, 238–48.
10. Hirabayashi, J. (2008) Concept, strategy and realization of lectin-based glycan profiling. *J Biochem* **144**, 139–47.
11. Pilobello, K. T., Krishnamoorthy, L., Slawek, D., and Mahal, L. K. (2005) Development of a lectin microarray for the rapid analysis of protein glycopatterns. *Chembiochem* **6**, 985–9.
12. Van Damme, E. J. M., Allen, A. K., and Peumans, W. J. (1987) Isolation and characterization of a lectin with exclusive specificity towards mannose from snowdrop (*Galanthus nivalis*) bulbs. *FEBS Lett* **215**, 140–4.
13. Van Damme, E. J. M., Barre, A., Rougé, P., Van Leuven, F., and Peumans, W. J. (1996) The NeuAc (α-2,6)-Gal/GalNAc binding lectin from elderberry (*Sambucus nigra*) bark, a type 2 ribosome inactivating protein with an unusual specificity and structure. *Eur J Biochem* **235**, 128–37.
14. Wang, W.-C. and Cummings, R. D. (1988) The immobilized leukoagglutinin from the seeds of *Maackia amurensis* binds with high affinity to complex type Asn-linked oligosaccharides containing terminal sialic acid-linked α-2,3 to penultimate galactose residues. *J Biol Chem* **261**, 4576–85.
15. Chen, Y., Peumans, W. J., Hause, B., Bras, J., Kumar, M., Proost, P., et al. (2002) Jasmonic acid methyl ester induces the synthesis of a cytoplasmic/nuclear chitooligosaccharide-binding lectin in tobacco leaves. *FASEB J* **16**, 905–7.
16. Vranken, A. M., Van Damme, E. J. M., Allen, K., and Peumans, W. J. (1987) Purification and properties of an N-acetylgalactosamine specific lectin from the plant pathogenic fungus *Rhizoctonia solani*. *FEBS Lett* **216**, 67–72.
17. Bador, M., Cabrera, G., Stadlmann, J., Lerouge, P., Cremata, J. A., Gomord, V., et al. (2009) N-glycosylation of plant recombinant pharmaceuticals. In: *Recombinant Proteins from Plants*, Methods and Protocols, Springer Protocols, L. Faye and V. Gomord, eds., Humana Press, New York, pp. 239–64.
18. Lihme, A., Schafer-Nielsen, C., Larsen, K. P., Müller, K. G., and Bøg-Hansen, T. C. (1986) Divinylsulphone-activated agarose. Formation of stable and non-leaking affinity matrices by immobilization of immunoglobulins and other proteins. *J Chromatogr* **376**, 299–305.
19. Shibuya, N., Goldstein, I. J., Van Damme, E. J. M., and Peumans, W. J. (1988) Binding properties of a mannose-specific lectin from the snowdrop (*Galanthus nivalis*) bulb. *J Biol Chem* 263, 728–34.
20. Lannoo, N., Peumans, W. J., Van Pamel, E., Alvarez, R., Xiong, T.-C., Hause, G., et al. (2006) Localization and in vitro binding studies suggest that the cytoplasmic/nuclear tobacco lectin can interact in situ with high-mannose and complex N-glycans. *FEBS Lett* **580**, 6329–37.
21. Shibuya, N., Goldstein, I. J., Broekaert, W. F., Nsimba-Lubaki, M., Peeters, B., and Peumans, W. J. (1987) The elderberry (*Sambucus nigra*) bark lectin recognizes the Neu5Ac (α2-6)/GalNAc sequence. *J Biol Chem* **262**, 1596–601.
22. Shibuya, N., Goldstein, I. J., Broekaert, W. F., Nsimba-Lubaki, M., Peeters, B., and Peumans, W. J. (1987) Fractionation of sialylated oligosaccharides, glycopeptides and glycoproteins on immobilized elderberry (*Sambucus nigra*) bark lectin. *Arch Biochem Biophys* **254**, 1–8.

Chapter 20

Strong Cation Exchange Chromatography for Analysis of Sialylated Glycopeptides

Katharina Lohrig, Albert Sickmann, and Urs Lewandrowski

Abstract

Glycosylations represent major and essential co- and post-translational modification forms of proteins and facilitate a multitude of functions such as cell–cell interactions as well as protein folding and stability. The analysis of protein glycosylation is still an enormous task due to the vast heterogeneity and multitude of different possible carbohydrate structures. The elucidation of glycosylation sites – the attachment points of carbohydrate structures to the polypeptide backbone – is often among the first necessary steps of analysis. Therefore, we here present a simple protocol for charge-based enrichment of sialylated glycopeptides by strong cation exchange chromatography and subsequent analysis of glycosylation sites by mass spectrometry.

Key words: Glycosylation, platelet, strong cation exchange chromatography, mass spectrometry, affinity capture.

1. Introduction

Glycosylations – enzymatic sugar additions to the polypeptide backbone – are among the most abundant modifications of proteins. Indeed, about 50% of all proteins is estimated to exhibit glycosylations in various forms ranging from single residues to larger substructures (1). The role of these modifications in cellular processes is manifold and varies from cell–cell interaction, protein maturation, stability and activity to subcellular trafficking and signal transduction properties to name but a few. This large scope of biomolecular tasks is enabled by a large number of different glycosylation types and forms. Most of current studies focus nevertheless on so-called N- and O-glycosylations

K. Gevaert, J. Vandekerckhove (eds.), *Gel-Free Proteomics*, Methods in Molecular Biology 753,
DOI 10.1007/978-1-61779-148-2_20, © Springer Science+Business Media, LLC 2011

bound either to asparagine or serine and threonine residues. In general, the study of glycosylations comprises different levels of analysis to elucidate the detailed structure of attached carbohydrates, possible attachment sites to the polypeptide backbone, and the distribution of carbohydrate forms within a complete protein (macroheterogeneity) or for individual glycosylation sites (microheterogeneity).

The here presented method focuses on the identification of N-glycosylation sites exemplified for human platelet proteins. However, the parallelized analysis of glycopeptides within complex peptide mixtures is still a major obstacle for mass spectrometric detection due to suppression effects and lower abundance in comparison to non-modified peptides. Therefore, an enrichment step for glycosylated peptides is usually employed to reduce the overall complexity. For this purpose a wide range of methods has been proposed in the past based on several properties and characteristics of the glycan structure itself: lectin affinity chromatography (use of structure-specific carbohydrate–protein interactions) (2, 3), oxidative hydrazide coupling (permanent immobilization following periodate oxidation of sugar residues) (3, 4), and hydrophilic liquid interaction chromatography (predominant interaction of hydrophilic carbohydrate moieties) (5) are among the most common ones. However, recently also methods were proposed for specific enrichment of sialylated carbohydrate structures, e.g., titanium dioxide (6) and strong cation exchange chromatography (7). The latter method relies on the assumption of a reduced net charge for sialic acid bearing peptides at low pH. Commonly, tryptic peptides show net charges greater than +2 since at least the N-terminus and the lysine or arginine at the C-terminus are positively charged, while carboxyl groups are neutral at pH 2.7. Additional sialic acid residues will reduce this net charge to +1 or even lower because of their partially negatively charged carboxyl group and thus those peptides are no longer or less retained by the negatively charged strong cation exchange resin.

After enrichment, the carbohydrate moieties are cleaved from the glycopeptides by PNGase F (8), a glycosidase with broad substrate specificity for nearly all forms of N-glycans. In this process, the asparagine within the consensus sequence (NXS/T) is deamidated and thus the mass of the residue is increased by 1 Da. After mass spectrometric sequencing of the former glycopeptides, the respective glycosylation site can be specified due to the Asn to Asp shift. However, the concrete information of the carbohydrate part is no longer present.

Here, a simple and robust workflow is presented covering sample preparation for human platelets, proteolytic digest followed by enrichment of sialylated glycopeptides using SCX, deglycosylation, and subsequent mass spectrometric sequencing including data interpretation.

2. Materials

In general, solvents and chemicals should be of the highest purity available, and at least p.a. grade. Solvents for HPLC separations prior to MS analysis should be LC gradient grade or higher. High purity water with 18.2 MΩ is recommended. Fused silica capillaries for custom-made separation columns are available from a range of suppliers such as Polymicro and Agilent. The use of non-deactivated polyimide-coated fused silica capillaries is recommended.

2.1. Sample Preparation of Human Platelets

1. Resuspension buffer: 10 mM citric acid, 5 mM KCl, 145 mM NaCl, 14 mM glucose, 1 mM $MgCl_2$, pH 6.4, stored at 4°C.
2. Use plastic labware throughout sample processing (*see* **Note 1**).
3. Liquid nitrogen.

2.2. Proteolytic Digest

1. Lysis buffer: 2.5% (w/v) SDS, 50 mM NH_4HCO_3, pH 7.8, freshly prepared.
2. Reduction solution: 1 mM DTT, 50 mM NH_4HCO_3, pH 7.8, freshly prepared.
3. Alkylation solution: 3 mM iodoacetamide (IAA), 50 mM NH_4HCO_3, pH 7.8, freshly prepared.
4. Ethanol, pre-cooled at –30°C.
5. Digestion buffer: 50 mM NH_4HCO_3, pH 7.8, freshly prepared.
6. Trypsin, modified, sequencing grade, 20 μg lyophilized trypsin per vial (Promega).

2.3. Strong Cation Exchange Chromatography (SCX)

1. SCX solvent A: 5 mM sodium phosphate buffer, pH 2.7.
2. SCX solvent B: 5 mM sodium phosphate buffer, 15% (v/v) acetonitrile, 500 mM NaCl, pH 2.7. Bioinert Famos™/Ultimate™-nano-HPLC system (Dionex) (*see* **Note 2**).
3. SCX column: custom-made 550 μm inner diameter × 15 cm length SCX column (Polysulfoethyl A, 200 Å pore size, 5 μm particle size, PolyLC).

2.4. Mass Spectrometric Analysis of Sialylated Glycopeptides

1. Digestion buffer: 50 mM NH_4HCO_3, pH 7.8, freshly prepared.
2. PNGase F, 100 U, *Flavobacterium meningosepticum*, lyophilized (Roche).
3. Formic acid, 98–100%.

4. LC-MS loading buffer: 0.1% (v/v) trifluoroacetic acid.
5. LC-MS solvent A′: 0.1% (v/v) formic acid.
6. LC-MS solvent B′: 0.1% (v/v) formic acid, 84% (v/v) acetonitrile.
7. Nano-HPLC system: Ultimate 3,000 (Dionex) or equivalent.
8. Pre-column: 75 μm ID × 2 cm length, ODS-3, 5 μm particle size (GL Sciences).
9. Main column: 75 μm ID × 25 cm length, ODS-3, 3 μm particle size (GL Sciences).
10. Mass spectrometer: HCT Ultra ETD II system (Bruker Daltonik GmbH).

3. Methods

3.1. Sample Preparation of Human Platelets

To obtain sufficient amounts of sample material for a set of reproducible experiments, aphaeresis-derived platelet concentrates should be used. Although these preparations ('buffy coats') are already enriched in platelets, they may still contain smaller portions of leukocytes, erythrocytes, and large amounts of plasma proteins. Therefore, platelets need to be further purified to avoid contaminations. Here, human platelets were from fresh aphaeresis concentrates (leukocyte depleted, ~2–4×10^{11} platelets/250 ml, Department of Transfusion Medicine, University Würzburg, Germany) (3, 7, 9).

1. Divide aphaeresis concentrates into 50 ml reaction tubes (*see* **Note 1**) and centrifuge twice at 310×*g* for 15 min at room temperature to remove remaining leukocytes or erythrocytes. Discard both pellets.
2. Centrifuge the supernatant of the second centrifugation step at 380×*g* for 20 min at room temperature to pellet the remaining platelets (*see* **Note 3**).
3. Resuspend the platelet pellets in 25 ml of resuspension buffer and centrifuge again at 380×*g* for 20 min at room temperature to remove residual plasma proteins. Repeat this step once.
4. Resuspend the platelet pellets in 12 ml of resuspension buffer and take 1 ml aliquots. Centrifuge briefly at 12,000×*g* for 30 s at room temperature to pellet the platelets.
5. Freeze the platelet pellets directly in liquid nitrogen and keep them at –80°C until further use.

3.2. Proteolytic Digest

1. Resuspend pellets of purified platelets in lysis buffer (*see* **Note 4**).
2. Incubate the samples at 56°C for 30 min with 1 mM DTT, followed by an incubation step at room temperature with 3 mM IAA in darkness to reduce disulfide bridges and alkylate the free cysteine residues, respectively.
3. Precipitate proteins with a 10-fold excess (volume) of precooled ethanol for at least 3 h at –30°C to remove contaminants, such as salts and detergents (*see* **Note 5**).
4. Resuspend the pellets in an appropriate volume of 50 mM ammonium bicarbonate. Resolubilize the precipitate thoroughly, e.g., by help of ultrasonication, until no particulate matter is visible. Dilute to higher volume if necessary.
5. Perform tryptic digestion overnight at 37°C by adding trypsin to a ratio of 1/50 (w/w, trypsin/protein).

3.3. Strong Cation Exchange Chromatography

Following tryptic digestion, the generated peptide mixtures contain a large set of different peptides holding all kinds of modifications as well as unmodified peptides. In order to enrich especially for sialic acid-containing peptide fractions, a charge-based enrichment step in form of a strong cation exchange chromatography is applied.

1. Dilute samples in an appropriate volume of SCX solvent A (*see* **Note 6**).
2. After injection of sample apply a prolonged washing period (20 min) with SCX solvent A (**Fig. 20.1**) to collect for sialic acid-containing peptides not retained by the resin (*see* **Note 7**).
3. Perform a short linear gradient up to 20% B within 15 min followed by a linear gradient up to 95% B within 30 min to remove the bulk of remaining peptides, which exhibit a higher net charge (**Fig. 20.1**).
4. Re-equilibrate the column (*see* **Note 8**).

3.4. Mass Spectrometric Analysis of Sialylated Glycopeptides

Prior to mass spectrometric analysis, individual fractions are digested with PNGase F.

1. Dilute samples sufficiently with 50 mM ammonium bicarbonate or desalt them prior digestion in order to meet appropriate digest conditions regarding pH and organic solvent content (*see* **Note 6**).
2. Add 4 units of PNGase F to each fraction and incubate overnight at 37°C.
3. Desalt the sample prior to mass spectrometric analysis (*see* **Note 6**) and reconstitute the final sample in 60 µl 0.1% (v/v) trifluoroacetic acid.

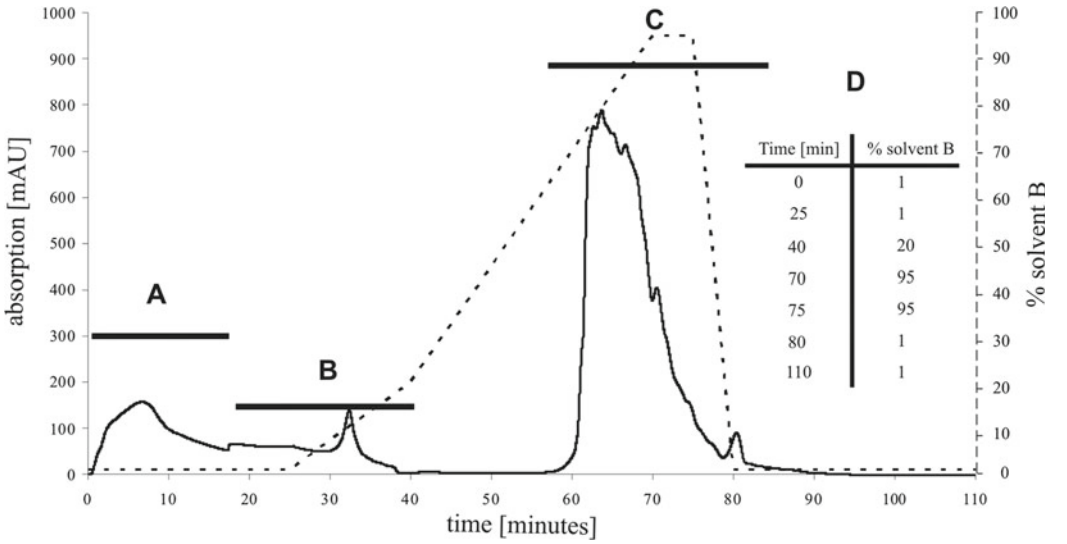


Fig. 20.1. Exemplary UV chromatogram of 200 μg platelet peptides separated by SCX. (**a**) The flow-through fraction including the first washing phase contains peptides with low net charge being either sialylated (glycosylated), phosphorylated, N-terminal acetylated, or free C-terminal peptides. (**b**) In the following fraction peptides with a slightly higher net charge are detected within the short gradient up to 20% of solvent B. (**c**) The last fraction contains the bulk of peptides with a higher net charge. (**d**) The table/graphic shows simultaneously the analytical SCX gradient by a *dashed line*.

4. Load the sample onto the pre-column using 0.1% (v/v) trifluoroacetic acid at a flow rate of 10 μl/min for 5 min using the loading pump of the HPLC system.
5. After switching the pre-column in-line with the main column, separate the peptides according to their hydrophobicity by applying a binary gradient of organic solvent with a slope of 1.5% solvent B′ per minute, followed by a washing step with 95% of solvent B′ for 5 min and re-equilibration with 5% of solvent A′ for 11 min. Keep the flow rate at 230 nl/min throughout the separation.
6. For MS survey scans, use the standard enhanced modus with a spray voltage of 4 kV and a scan range of 380–2,000. Perform MS/MS fragmentation of the five most abundant precursors (charge state 2 or higher) with an ICC target value of 2×10^5. Enable dynamic exclusion of already fragmented precursor ions after a repeat count of one for a duration of 60 s.
7. Export MS/MS spectra by the Data Analysis software and perform database searches with the MASCOT search algorithm against a human subset of the Swiss-Prot database (www.uniprot.org). Set the following search parameters: peptide mass tolerance of 0.2 Da for the monoisotopic precursor, fragment mass tolerance of 0.4 Da, trypsin as protease (with one missed cleavage site allowed), and

inclusion of the first ^{13}C isotopic peak for precursor mass evaluation. Besides the fixed modification of alkylated cysteine, the following variable modifications need to be set: oxidation of methionine, deamidation of asparagine and glutamine, acetylation of the protein N-terminus, and phosphorylation (*see* **Note 9**).

3.5. Alternative Enrichment Strategy

To reduce time and costs and in case an HPLC system is not available for SCX prefractionation, a simplified enrichment protocol of the presented SCX chromatography based on spin columns can be applied.

1. Make a spin column from a gel loader tip closed at the tip with a small glass fiber filter.
2. Fill the column with SCX stationary phase resuspended in methanol up to a length of 5 mm (*see* **Note 10**).
3. Equilibrate the SCX material by applying several column volumes of SCX solvent A (a column length of 5 mm usually requires a threefold washing step with 50 μl of SCX solvent A).
4. Apply the sample dissolved in SCX solvent A and carefully collect the flow-through containing the enriched sialylated glycopeptides.
5. Wash the spin column once with 25 μl of SCX solvent A prior to elution with SCX solvent B if further analysis of non-glycosylated peptides bound to the column is required.

4. Notes

1. In order to obtain platelets in their non-activated form the use of plastic labware is mandatory. Contact to glass surfaces may lead to activation and subsequent aggregation of the platelets.
2. Equivalent HPLC systems may be used as long as their flow rate is compatible with the applied column hardware.
3. This step may be performed twice to maximize the yield of platelets. Meanwhile, resuspend the platelets in buffer as described in the next step and proceed as instructed.
4. The volume of lysis buffer depends on the pellet volume. In general, after cell lysis by SDS the solution should be yellowish but clear.
5. To clean up the samples, other kinds of precipitation procedures may also be used depending on the sample and its

characteristics. The ethanol precipitation step can optionally be prolonged to an overnight incubation.

6. To remove salts and to achieve appropriate loading conditions, pre-cleaning of peptide samples prior to SCX separation is recommended. For this purpose, ZipTip (Millipore) or several other C18 material-packed spin columns or SPE cartridges may be used. In general, the manufacturer's instructions can be applied. Usually, the material is first activated by applying organic solvent (e.g., acetonitrile) followed by equilibration with water. After loading the sample, the columns are washed by applying 2–3 column volumes of water. Peptides are released from the column by applying a column volume of 60% (v/v) acetonitrile. Subsequently, organic solvent is removed by evaporation. Afterward, the sample can be diluted with appropriate SCX solvent A to the final loading volume. In this context, the correct pH of sample solution is of utmost importance for the following separation quality. The appropriate sample loading amount strongly depends on the dimensions of the applied SCX column hardware. For the current approach, a self-packed 15 cm length SCX column (Polysulfoethyl A) with an internal diameter of 550 μm was used to separate ~200 μg peptide sample material. For SCX-based enrichment, an HPLC system from Dionex consisting of a Famos™ autosampler, an Ultimate™ pump, and an UV detector was used. The autosampler was equipped with a 100 μl loop and was directly coupled to the Ultimate system via the included six-port valve as shown in **Fig. 20.2**. A flow rate of

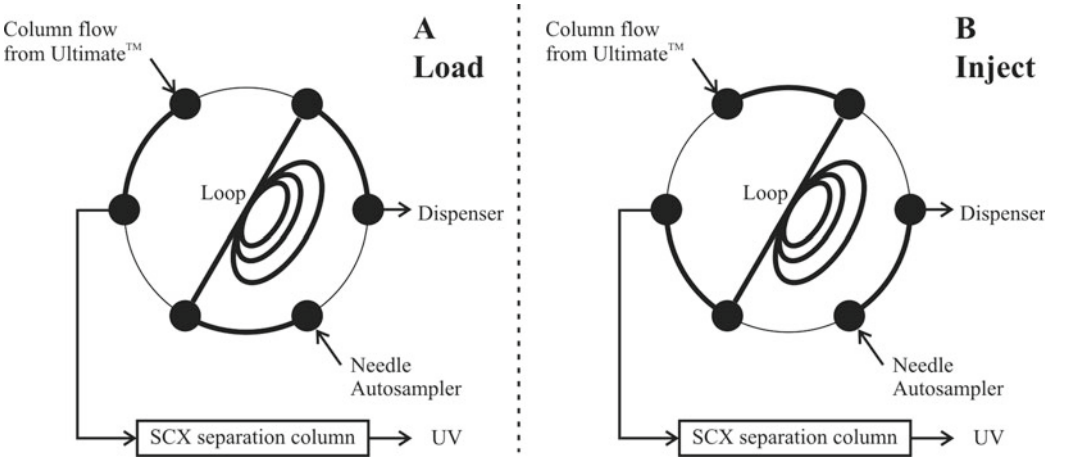


Fig. 20.2. Schematic view of a six-port valve. (**a**) In the load position, the sample is drawn up into the loop by the dispenser (syringe). (**b**) The sample is applied to the SCX separation column by the pump flow via the sample loop.

15µl/min should result in a column pressure of approx. 80 bar.

7. Besides sialylated glycopeptides, other peptides with reduced net charge may be identified in the flow-through fractions of the gradient. For example, not only N-terminal acetylated peptides and native C-terminal peptides but also phosphorylated peptides may elute early or are not retained at all by the stationary phase.
8. To enhance separation quality, a sufficient equilibration time of the SCX column has to be ensured. Therefore, the SCX column should be equilibrated with solvent A (or with 1–5% of solvent B) for at least 30 min before sample loading. Also, a sufficiently long regeneration step after the separation gradient has to be included.
9. For evaluation of MASCOT hits and identification of glycosylation sites, a critical assessment of the data is necessary. First, the deamidation site (former attachment site of N-glycan chain) should be on an asparagine residue within the proper consensus sequence (NXS/T; X no proline;

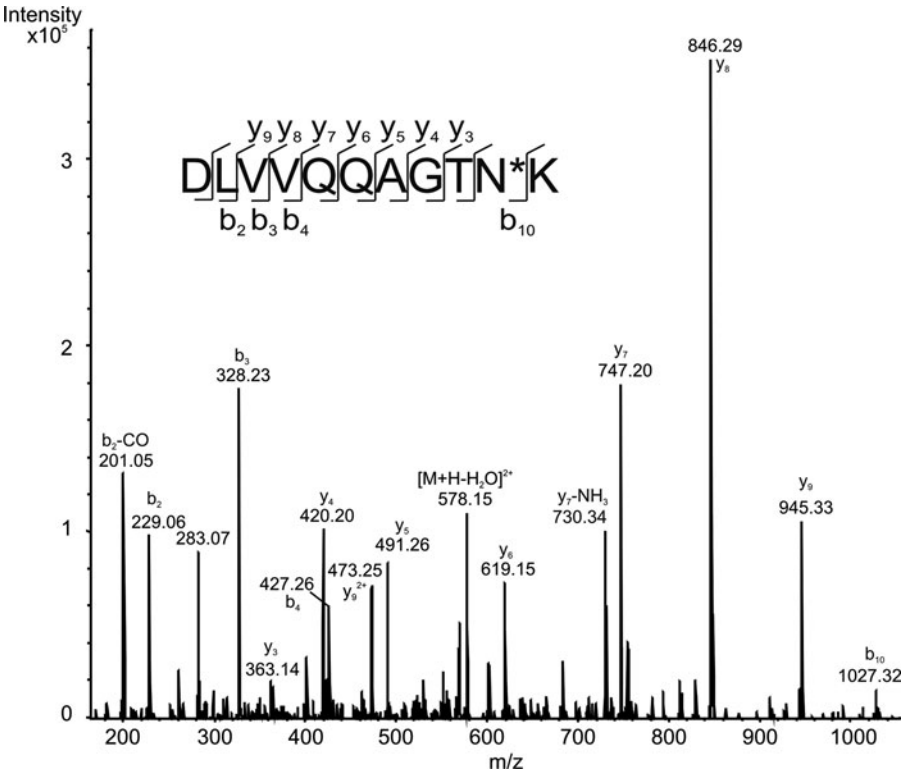


Fig. 20.3. MS/MS spectrum of tumor necrosis factor receptor superfamily member 5 (P25942) peptide (K.)DLVVQQAGTNK(.T). Besides a continuous series of y-ions this spectrum shows the deamidation of asparagine to aspartate (marked by an *asterisk*) and thus determines the glycosylation site to be N180KT.

only few exceptions have been found for NXC (10, 11)). Second, the deamidation site should be unambiguously identified within the peptide sequence to exclude alternative deamidation sites (due to experimental procedures or natural deamidations). In this context, NGT/S is often prone to artificial deamidation. Third, the peptide sequence should be identified by a continuous series of b- or y-ions and the major peaks of the respective spectrum have to be explained by the algorithm (**Fig. 20.3**).

10. The length of the spin column and the amount of SCX resin is again depending on the respective sample amount.

References

1. Apweiler, R., Hermjakob, H., and Sharon, N. (1999) On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database, *Biochim Biophys Acta* ***1473***, 4–8.
2. Kaji, H., Saito, H., Yamauchi, Y., Shinkawa, T., Taoka, M., Hirabayashi, J., Kasai, K., Takahashi, N., and Isobe, T. (2003) Lectin affinity capture, isotope-coded tagging and mass spectrometry to identify N-linked glycoproteins, *Nat Biotechnol* ***21***, 667–672.
3. Lewandrowski, U., Moebius, J., Walter, U., and Sickmann, A. (2006) Elucidation of N-glycosylation sites on human platelet proteins: A glycoproteomic approach, *Mol Cell Proteomics* ***5***, 226–233.
4. Zhang, H., Li, X. J., Martin, D. B., and Aebersold, R. (2003) Identification and quantification of N-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry, *Nat Biotechnol* ***21***, 660–666.
5. Hagglund, P., Bunkenborg, J., Elortza, F., Jensen, O. N., and Roepstorff, P. (2004) A new strategy for identification of N-glycosylated proteins and unambiguous assignment of their glycosylation sites using HILIC enrichment and partial deglycosylation, *J Proteome Res* ***3***, 556–566.
6. Larsen, M. R., Jensen, S. S., Jakobsen, L. A., and Heegaard, N. H. (2007) Exploring the sialiome using titanium dioxide chromatography and mass spectrometry, *Mol Cell Proteomics* ***6***(10), 1778–1787.
7. Lewandrowski, U., Zahedi, R. P., Moebius, J., Walter, U., and Sickmann, A. (2007) Enhanced N-glycosylation site analysis of sialoglycopeptides by strong cation exchange prefractionation applied to platelet plasma membranes, *Mol Cell Proteomics* ***6***, 1933–1941.
8. Tarentino, A. L., Gomez, C. M., and Plummer, T. H., Jr. (1985) Deglycosylation of asparagine-linked glycans by peptide:N-glycosidase F, *Biochemistry* ***24***, 4665–4671.
9. Tandon, N. N., Lipsky, R. H., Burgess, W. H., and Jamieson, G. A. (1989) Isolation and characterization of platelet glycoprotein IV (CD36), *J Biol Chem* ***264***, 7570–7575.
10. Bause, E., and Legler, G. (1981) The role of the hydroxy amino acid in the triplet sequence Asn-Xaa-Thr(Ser) for the N-glycosylation step during glycoprotein biosynthesis, *Biochem J* ***195***, 639–644.
11. Miletich, J. P., and Broze, G. J., Jr. (1990) Beta protein C is not glycosylated at asparagine 329. The rate of translation may influence the frequency of usage at asparagine-X-cysteine sites, *J Biol Chem* ***265***, 11397–11404.

Chapter 21

Titanium Dioxide Enrichment of Sialic Acid-Containing Glycopeptides

Giuseppe Palmisano, Sara E. Lendal, and Martin R. Larsen

Abstract

Glycosylation is one of the many post-translational protein modifications that regulate several biological processes of proteins and lipids. In particular aberrant sialylation, at the terminal position of the glycan structures of cell surface proteins, occurs in numerous diseases such as cancer metastasis and viral infections. Methodological improvements in the sample preparation and analysis currently enable the detailed identification of the glycosylation sites and glycan structure characterization. In this context, the aim of this chapter is to describe a methodology to identify the glycosylation site of N-linked sialylated glycoproteins. The method relies on the specificity of titanium dioxide affinity chromatography to isolate sialic acid-containing glycopeptides. After enzymatic release of the glycans, the enriched sialylated glycopeptides are analyzed by mass spectrometry. This strategy was applied to a crude membrane fraction of EGF-stimulated HeLa cells metabolically labeled with SILAC enabling both qualitative and quantitative analyses of sialoglycopeptides.

Key words: Sialylation, glycosylation, titanium dioxide, quantitative proteomics, mass spectrometry.

1. Introduction

The analytical improvements in the characterization of biological molecules by genomics and proteomics have revealed a close relationship between certain disease states and the post-translational modification (PTM) state of proteins. In particular, protein glycosylation, which is the most common PTM (1, 2), regulates several biological processes (development, growth, or survival (3–5)) and is involved in different physiopathological conditions (cancer, congenital disorders of glycosylation, infections, and diabetes (6–10)). There are two main types of protein glycosylation:

K. Gevaert, J. Vandekerckhove (eds.), *Gel-Free Proteomics*, Methods in Molecular Biology 753,
DOI 10.1007/978-1-61779-148-2_21,

N-linked and O-linked glycosylation. N-Linked is referred to the glycan structures attached to the polypeptide chain via an amide bond at asparagine (Asn) residues in the consensus sequence -Asn-X-Thr/Ser/Cys (where X can be any amino acid except Pro (11)). O-Linked glycans are attached to the oxygen of a hydroxylated amino acid, commonly Ser or Thr via an ether bond (12). Other less abundant glycosylation types are glycosylphosphatidylinositol (GPI) anchors (13), P-glycosylation (14), and C-glycosylation (15).

Sialic acids (SAs) are a group of neuraminic acid (5-amido-3,5-dideoxy-D-glycero-D-galacto-nonulosonic acid) ubiquitously distributed as terminal sugars on oligosaccharides attached to protein or lipid moieties. In particular, *N*-acetylneuraminic acid (NANA) and its derivatives occupy the terminal, non-reducing position of N-linked membrane and secreted glycoproteins (16).

Aberrant sialylation has been associated with several diseases such as cancer metastatic progression (17, 18), cardiac arrhythmias (19), cystic fibrosis (20), coronary artery disease (21), IgA nephropathy (22), and viral infections (23).

Different strategies can be used to characterize the SA-containing glycoproteins as shown in **Table 21.1**. Lectins are a broad class of glycan-binding proteins that recognize glycan epitopes and SA-binding lectins such as *Sambucus nigra* agglutinin and *Maackia amurensis* leukoagglutinin I/II have been used to enrich SA glycoproteins (24, 25). A labeling strategy for cell surface SA-containing glycans has been described (26, 27). This method uses mild periodate oxidation to generate an aldehydic group on SAs and a chemical coupling step with specific tags prior to release and analysis. Strong cation exchange has been described to enrich sialoglycopeptides from human platelet membranes (28). Another approach described by our group (29) is based on the selective enrichment of SA-containing glycopeptides using titanium dioxide (TiO_2) as affinity material toward charged molecules such as phosphorylated and SA-containing peptides.

Table 21.1
Methods for SA-containing glycoproteins characterization

Methods	Enrichment principle	References
Lectins	Glycan epitopes recognition	(24, 25)
Hydrazide	Selective chemical oxidation and coupling	(26, 27)
Strong cation exchange	Electrostatic interactions	(28)
Titanium dioxide chromatography	Bi/multi-dentate chelating complexes	(29)

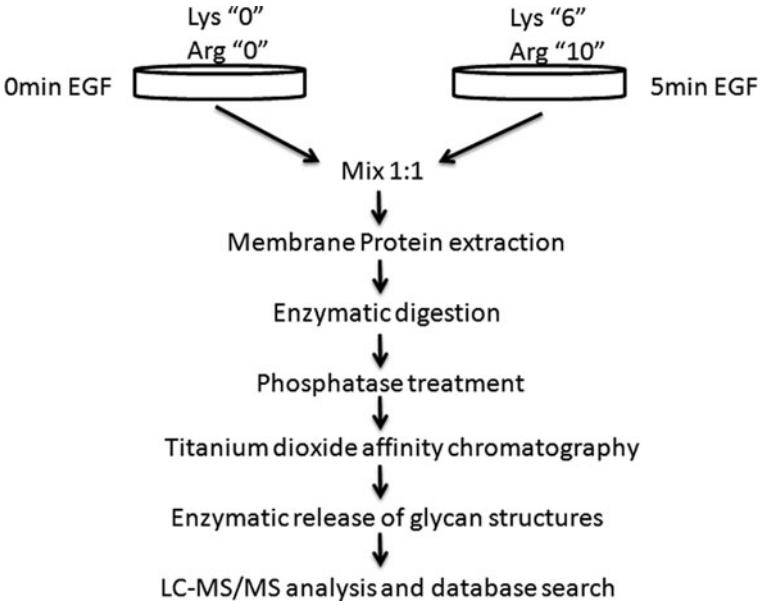


Fig. 21.1. Strategy for the qualitative and quantitative identification of *N*-linked sialic acid-containing glycopeptides in the membrane fraction of EGF-stimulated HeLa cells. HeLa cells were grown in medium supplemented with "light" ($^{12}C_6$)-lysine "0" and ($^{12}C_6$)-arginine "0" and "heavy" ($^{13}C_6$)-lysine "6" and ($^{13}C_6{}^{15}N_4$)-arginine "10" and, respectively, stimulated with EGF for 0 and 5 min. After mixing equal numbers of cells and membrane protein extraction, proteins were proteolytically digested with trypsin and treated with alkaline phosphatase. Tryptic peptides were subsequently loaded on TiO_2 to enrich the sialoglycopeptides. After *N*-glycosidase F treatment the resulting deglycosylated peptides were analyzed by LC-MS/MS and database searching.

In this chapter we describe a robust, selective, and simple method used in our laboratory to enrich sialoglycopeptides using TiO_2 affinity chromatography and we apply this to the membrane fraction from epidermal growth factor (EGF)-stimulated HeLa cells (**Fig. 21.1**). HeLa cells were grown in medium containing "light" lysine and arginine ($^{12}C_6$) and "heavy" lysine ($^{13}C_6$) and arginine ($^{13}C_6{}^{15}N_4$), respectively, and the labeled cell population was stimulated with EGF for 5 min. After mixing equal amounts of cells, the membrane protein fraction was enriched using Na_2CO_3 and ultracentrifugation as previously described (30) and the membrane proteins were subsequently enzymatically digested with trypsin. Prior to TiO_2 enrichment the peptide mixture was treated with alkaline phosphatase to avoid co-purification of phosphopeptides. The enriched SA-containing glycopeptide fraction was treated with *N*-glycosidase F and subjected to LC-MS/MS analysis. The raw data were processed in order to obtain qualitative and quantitative identification of the SA-containing glycopeptides. Using this strategy we identified 418 sialylated glycosites mapped to 190 glycoproteins (**Fig. 21.2a**); 75 of the identified glycoproteins contain more than one transmembrane domain (**Fig. 21.2b**).

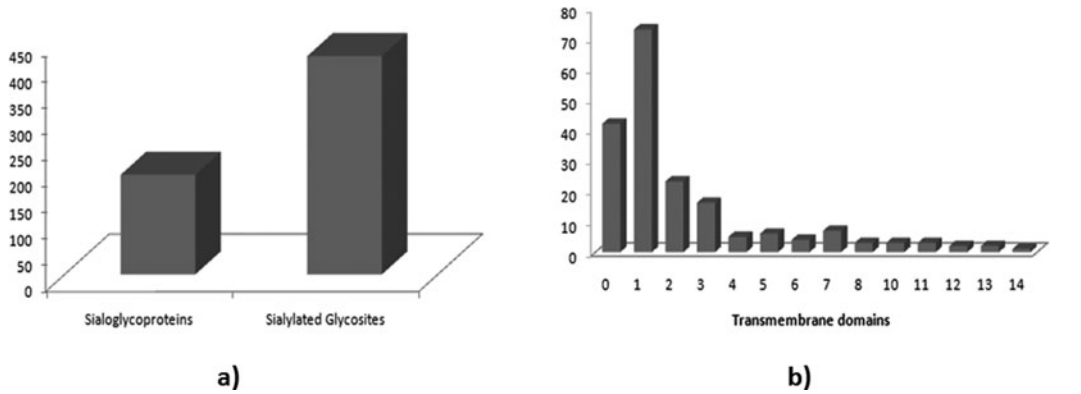


Fig. 21.2. Sialoglycoproteins and glycosites identified using TiO_2 enrichment. (**a**) Number of sialoglycoproteins and sialylated glycosites identified using TiO_2-based enrichment of membrane fraction from HeLa cells and (**b**) transmembrane domains of the identified glycoproteins.

2. Materials

2.1. Membrane Proteins Extraction

1. HeLa cells (human cervix epithelial adenocarcinoma cells), grown in Dulbecco's modified Eagle's medium (GIBCO BRL Invitrogen) containing "light" ($^{12}C_6$)-lysine and ($^{12}C_6$)-arginine (both Sigma) or "heavy" ($^{13}C_6$)-lysine and ($^{13}C_6{}^{15}N_4$)-arginine (both Cambridge Isotope Laboratories) (10^7 cells per condition). The two conditions were stimulated with epidermal growth factor (Sigma) at 150 ng/ml for 0 and 5 min, respectively (*see* **Note 1**).
2. Lysis buffer: 100 mM Na_2CO_3 (pH 11), protease inhibitor cocktail (Roche Diagnostics), added according to the manufacturer' instructions (*see* **Note 2**).
3. MEGAFUGE 1.0R centrifuge (Thermo).
4. Ultracentrifuge M150GX (Sorvall) and 4 ml PC tube Assy (Hitachi).

2.2. In-Solution Enzymatic Digestion and Phosphatase Treatment of Membrane Proteins

1. Denaturing buffer: 6 M urea, 2 M thiourea in 100 mM NH_4HCO_3, pH 7.8
2. Reduction buffer (10x): 100 mM dithiothreitol (DDT) in 50 mM NH_4HCO_3, pH 7.8.
3. Alkylating buffer (10x): 500 mM iodoacetamide in 50 mM NH_4HCO_3, pH 7.8. This solution must be kept in the dark.
4. Trypsin solution: 0.5 mg/ml trypsin (sequence-grade, modified trypsin (Promega)) in 50 mM NH_4HCO_3, pH 7.8 (*see* **Note 3**).

5. Alkaline phosphatase from calf intestinal (Roche Diagnostics, *see* **Note 4**).

2.3. Titanium Dioxide Affinity Chromatography

1. Titansphere TiO_2, 5 μm chromatographic material (GL Sciences) suspended in acetonitrile (*see* **Note 5**).
2. TiO_2 loading buffer: 1 M glycolic acid, 80% acetonitrile, 5% trifluoroacetic acid (TFA). This solution can be stored in glass bottle for 1 week at 4°C.
3. TiO_2 washing buffer 1: 80% acetonitrile, 19% water (MilliQ 18 MΩ cm), 1% TFA (can be stored at 4°C for up to 1 week).
4. TiO_2 washing buffer 2: 2% acetonitrile, 97.9% water, 0.1% TFA (can be stored at 4°C for up to 1 week).
5. TiO_2 elution buffer: 40 μl 25% ammonia in 960 μl water, pH 11.3 (this solution can be stored at 4°C for up to 1 week).
6. Thermomixer.
7. Low binding "SafeSeal" Microcentrifuge Tubes, 1.7 ml (Sorenson Bioscience).

2.4. Enzymatic Deglycosylation

1. *N*-Glycosidase F (PNGase F) in glycerol containing solution (Roche Diagnostics).
2. Sialidase A recombinant from *Arthrobacter ureafaciens* (Glyko, Porozymes)
3. Deglycosylation buffer:50 mM NH_4HCO_3, pH 7.8 (*see* **Note 6**).

2.5. Desalting and Mass Spectrometry Analysis of Deglycosylated N-Linked Sialoglycopeptides

All chemicals must be of the highest degree of purity.

1. Chromatographic material Poros OLIGO R3 (Applied Biosystems), approximately 1.5 mg/100 μl suspended in 70% acetonitrile and 30% water.
2. Desalting wash: 2% formic acid (this solution can be stored at 4°C for 1 week).
3. Desalting elution: 70% acetonitrile, 28% water, and 2% formic acid (this solution can be stored at 4°C for 1 week).
4. C-18 Stage Tips (Proxeon Biosystems).
5. Disposable 1 ml syringe with a cut down 200 μl tip to fit to the C-18 Stage Tips.
6. EasynLC Nanoflow HPLC (Proxeon Biosystems).
7. Nanoelectrospray ion source (Proxeon Biosystems).
8. Reversed phase column (15 cm long, 100 μm inner diameter packed in-house with ReproSil-Pur C_{18}-AQ 3 μm resin, Dr. Maisch GmbH, Germany).

9. Solvent A: 0.1% formic acid in water.
10. Solvent B: 0.1% formic acid and 90% acetonitrile in water.
11. LTQ-Orbitrap XL (Thermo) is used in our laboratory. Other tandem mass spectrometers capable of automated acquisition of tandem mass spectra should also be suitable.

2.6. Software Analysis

1. MaxQuant Software Package (http://www.maxquant.org/) (*see* **Note 7**).

3. Methods

3.1. Membrane Proteins Extraction

All procedures must be carried out at 4°C.

1. Wash the cell pellets two times with ice-cold phosphate buffered saline (10^7 cells per condition, equal numbers combined) and spin down at 300×*g* in MEGAFUGE 1.0R.
2. Re-suspend the pellet in 2 ml of lysis buffer for 60 min on ice to introduce discontinuities into membrane vesicles and to remove the proteins loosely associated with artificial vesicles of microsomes and intact organelles such as mitochondria and lysosomes of the membrane proteins (*see* **Note 8**).
3. Transfer the lysate to a 4 ml PC tube and centrifuge in an ultracentrifuge at 150,000×*g* for 1 h.
4. Remove the supernatant (cytoplasmic proteins).
5. Wash the pellet gently with 500 μl of 500 mM NH_4HCO_3, pH 7.8.

3.2. In-Solution Enzymatic Digestion and Phosphatase Treatment of Membrane Proteins

1. Re-solubilize the pellet containing the membrane proteins in a 50 μl of denaturing buffer using sonication to improve solubilization (*see* **Note 9**).
2. Add reduction buffer at 10 mM final concentration and incubate at 30°C for 40 min (*see* **Note 10**).
3. Add alkylating buffer at 50 mM final concentration and incubate at room temperature for 40 min in the dark.
4. Dilute the solution with 50 mM NH_4HCO_3, pH 7.8, in order to obtain a urea concentration less than 1 M.
5. Add trypsin solution and incubate overnight at 30°C (*see* **Note 11**).
6. Upon digestion, freeze at –80°C and thaw the peptide mixture to inactivate the proteases or add protease inhibitors.
7. Treat the tryptic peptides with 2 units of alkaline phosphatase at 30°C for 2 h to avoid co-purification of phosphorylated peptides in the TiO_2 enrichment step.

3.3. Titanium Dioxide Enrichment

1. Prepare TiO_2 bead slurry by weighing 10 mg of TiO_2 beads and adding 100 μl of acetonitrile. Stir this slurry for 2–3 min. Transfer 10 μl of this solution to a low binding tube and vacuum dry in a Speedvac. This tube contains 1 mg of TiO_2 beads.
2. Dilute the dephosphorylated peptide solution with the TiO_2 loading buffer 1:10 and add this to the 1 mg of TiO_2 beads.
3. Incubate for 30 min on a thermomixer (1,400 rpm, room temperature) and settle the beads by centrifugation.
4. Remove the supernatant (*see* **Note 12**).
5. Add 50 μl of loading buffer to the beads.
6. Incubate for 1 min on a thermomixer (1,400 rpm, room temperature) and settle the beads down by centrifugation. Remove the supernatant.
7. Wash the beads with 50 μl of TiO_2 washing buffer 1 and repeat **Section 3.3**, step 6.
8. Wash the beads with 50 μl of TiO_2 washing buffer 2 and repeat **Section 3.3**, step 6.
9. Vacuum dry the beads for 5–10 min in a Speedvac.
10. Incubate the beads with TiO_2 elution buffer solution for 20 min on a thermomixer (1,400 rpm, room temperature) to elute the sialoglycopeptides.
11. Transfer the eluted sialoglycopeptides to a new tube and lyophilize these peptides by vacuum centrifugation (*see* **Note 13**).

3.4. Enzymatic Deglycosylation

1. Re-dissolve the sialoglycopeptides in 20 μl of 50 mM NH_4HCO_3 and add 0.5 U of PNGase F and 0.1 U of sialidase A (*see* **Note 14**).
2. Incubate at 37°C from 3 h to overnight.

3.5. Desalting and Mass Spectrometric Analysis of Deglycosylated N-Linked Sialoglycopeptides

1. Prepare a slurry of 100–200 μl of POROS Oligo R3 beads in 70% acetonitrile (approximately 1.5 mg/100 μl, *see* **Note 15**).
2. Load 20 μl of 70% acetonitrile in the top of a C18 Stage Tip (p10) and add 5 μl of the POROS Oligo R3 resin slurry on top of the acetonitrile. Use a 1 ml syringe with a cut down 200 μl tip to fit to the diameter of the C18 stage Tip and press the liquid gently and completely down to create a Poros R3 microcolumn at the end of the tip (0.5–1 cm long).
3. Apply 40 μl of desalting wash solution to the column and use 30 μl of this solution to equilibrate the column. Leave the remaining 10 μl on top of the column bed.

4. Re-dissolve the lyophilized deglycosylated peptides in 1 μl of formic acid and dilute using 20 μl of UHQ water. Apply the sample on top of the remaining 10 μl of desalting wash solution.
5. Load the sample through the column by applying air pressure using a syringe. Do not let the column turn completely dry.
6. Wash the column with 20 μl of desalting wash solution to remove salts and contaminants and leave the column dry.
7. Elute the desalted, deglycosylated peptides using 40 μl of desalting elution buffer (*see* **Note 13**).
8. Lyophilize the eluted peptide mixture in a vacuum centrifuge.
9. Dissolve the deglycosylated peptides in 5 μl of solvent A and analyze using nanoflow HPLC coupled online via a nanoelectrospray ion source to an LTQ-OrbitrapXL mass spectrometer (this is the analytical setup used in our laboratory).
10. Load the dissolved peptide mixture onto the reverse phase column at a flow rate of 550 nl/min of solvent A.
11. Elute the peptides by applying a linear gradient from 0 to 40% of solvent B in 50 min at a flow rate of 200 nl/min. Operate the mass spectrometer in data-dependent mode, whereby multiply charged ions trigger the acquisition of tandem mass spectra of the five most abundant ions per MS scan. As an example, a base peak ion chromatogram is shown in **Fig. 21.3a** and the SILAC pair of the galectin-3-binding protein peptide GLN(de)LTEDTYKPR, where N(de) indicates the sialylated glycosite, is shown in **Fig. 21.3b**. The precursor ion intensities clearly indicate a down-regulation of this glycopeptide after EGF stimulation.

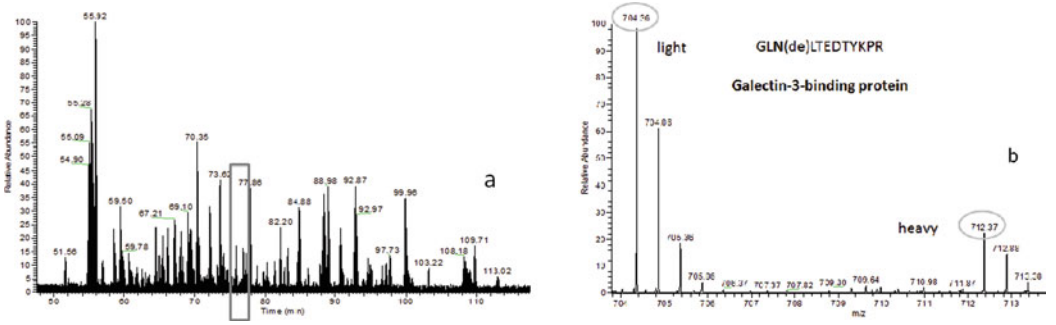


Fig. 21.3. Base peak ion chromatogram (**a**) and zoomed mass spectrum at 75.4 min (**b**).

Table 21.2
List of sialoglycoproteins that are regulated upon EGF stimulation

Protein ID	Protein name	Peptide sequence[a]	Ratio H/L	Significance B
IPI:IPI00554481.1	4F2 cell-surface antigen heavy chain	_SLVTQYLN(de)ATGNR_	19.152	1.49E–69
IPI:IPI00221224.6	Aminopeptidase N	_GPSTPLPEDPNWN(de)VTEFHTTPK_	0.51984	7.42E–09
IPI:IPI00021536.1	Calmodulin-like protein 5	_AFSAVDTDGN(de)GTINAQELGAALK_	0.098948	1.39E–100
IPI:IPI00789155.1	Calumenin	_N(de)ATYGYVLDDPDPDDGFNYK_	0.76706	1.62E–05
IPI:IPI00027078.3	Carboxypeptidase D	_FANEYPN(de)ITR_	11.299	6.37E–30
IPI:IPI00289819.4	Cation-independent mannose-6-phosphate receptor	_N(de)GSSIVDLSPLIHR_	0.13371	4.06E–51
IPI:IPI00011302.1	CD59 glycoprotein	_TAVN(de)CSSDFDACLITK_	0.11214	3.99E–13
IPI:IPI00215998.5	CD63 antigen	_QQMENYPKNN(de)HTASILDR_	0.14135	1.31E–62
IPI:IPI00400826.1	Clusterin	_LAN(de)LTQGEDQYYLR_	0.28569	1.20E–202
IPI:IPI00400826.1	Clusterin	_HN(de)STGCLR_	0.067979	6.60E–108
IPI:IPI00022937.4	Coagulation factor V	_TWN(de)QSIALR_	0.020455	2.12E–104
IPI:IPI00397229.4	EGF-like module-containing mucin-like hormone receptor-like 2	_TFKN(de)ESENTCQDVDECQQNPR_	0.68906	0.00021715
IPI:IPI00550382.2	Equilibrative nucleoside transporter 1	_LDMSQN(de)VSLVTAELSK_	0.082769	1.67E–16
IPI:IPI00855785.1	Fibronectin	_HEEGHMLN(de)CTCFGQGR_	0.00049198	0

(continued)

Table 21.2 (continued)

Protein ID	Protein name	Peptide sequence[a]	Ratio H/L	Significance B
IPI:IPI00029723.1	Follistatin-related protein 1	_GSN(de)YSEILDK_	0.18161	0.00038199
IPI:IPI00023673.1	Galectin-3-binding protein	_ALGFEN(de)ATQALGR_	0.37125	3.15E–47
IPI:IPI00023673.1	Galectin-3-binding protein	_GLN(de)LTEDTYKPR_	0.38522	5.11E–44
IPI:IPI00023673.1	Galectin-3-binding protein	_AAIPSALDTN(de)SSK_	0.29638	1.14E–08
IPI:IPI00009890.1	Glia-derived nexin	_N(de)ASEIEVPFVTR_	0.24005	1.95E–10
IPI:IPI00290043.1	Integrin alpha-3	_TSIPTINMEN(de)K_T	0.63546	0.00019238
IPI:IPI00004503.5	Lysosome-associated membrane glycoprotein 1	_SSCGKEN(de)TSDPSLVIAFGR_	0.54926	3.34E–08
IPI:IPI00219111.7	Translocation-associated membrane protein 1	_KGTENGVN(de)GTLTSNVADSPR_	0.12604	6.61E–33
IPI:IPI00298971.1	Vitronectin	_N(de)GSLFAFR_	0.32263	1.29E–25

Protein ID, protein name, peptide sequence, ratio heavy/light, and significance B are given (37).
[a]N(de) indicates deamidated asparagine converted to aspartic acid due to PNGase F reaction.

3.6. Software Analysis

1. Process the experimental raw data using the MaxQuant software package (31, 32) (*see* **Note 7**). The N-linked sialoglycoproteins that are statistically changing upon EGF stimulation are reported in **Table 21.2.**

4. Notes

1. For a detailed protocol about SILAC label, refer to (33) and http://www.silac.org. It should be noted that other quantitative strategies such as iTRAQ, ICAT, ^{18}O, GIST, ICPL, TMT or a label-free approach can be combined with this protocol to retrieve quantitative informations. The choice of a specific quantitative strategy should be evaluated and subsequently the protocol should be optimized.
2. The protease inhibitor should be added just before use. If it is intended to perform a phosphorylation study, phosphatase inhibitors should be added to the lysis buffer. The phosphatase treatment is important to dephosphorylate the phosphopeptides and reduce the number of binding peptides to titanium dioxide improving the binding efficiency.
3. Reduction buffer, alkylating buffer, and trypsin solution should be prepared just before use.
4. To improve the efficiency of phosphatase treatment, it is possible to use lambda phosphatase.
5. Different synthesis conditions (e.g., temperature and particle size) have been found to strongly influence the material characteristics (34) and could influence the selectivity of the method even though there are no reports regarding the influence of the TiO_2 characteristics on SA-containing glycopeptides enrichment.
6. The PNGase F reaction converts the asparagine N-linked to glycan structures to aspartic acid resulting in the increase of 0.9840 Da. Alternatively, this reaction can be carried out in deglycosylation buffer prepared in ^{18}O-enriched water to further confirm the glycosylation site resulting in the increase of 2.9880 Da. For a protocol to avoid potential pitfalls in ^{18}O-based N-linked deglycosylation, refer to (35).
7. Other software platforms could be used to analyze quantitative data depending on the chemical, enzymatic, or label-free approach used. For an overview, refer to (31).

8. After collecting the membrane fraction pellet, the sodium carbonate incubation can be repeated two or three times if the fraction of membrane proteins is less than 50%.
9. Measure the protein concentration to calculate the amount of the enzyme to perform the digestion.
10. Digestions in high concentration of urea can often lead to carbamylations of lysine and N-terminal amines. Due to that elevated temperatures should be avoided to minimize this side reaction.
11. Before adding the trypsin solution, the protein solution in urea can be incubated with lysyl endopeptidase (LysC) for 3 h at room temperature at 1:50 (enzyme:protein) ratio, as this enzyme is active in high urea concentrations.
12. The supernatant can be dried down and desalted using reverse phase solid phase extraction for further analysis.
13. A small aliquot of this solution can be spotted on MALDI target plate and co-crystallized with a matrix like 2,5-dihydroxybenzoic acid (DHB) or 2,6-dihydroxyacetophenone (DHAP)/di-ammonium hydrogen citrate (DHAC) matrix that are "cold matrixes" that reduce in-source fragmentation and metastable decay. Moreover, DHAC is useful to avoid cation adduction to sialylated glycopeptides.
14. PNGase F is an amidase that cleaves between the innermost GlcNAc and asparagine residues of high mannose, hybrid, and complex oligosaccharides but it does not cleave N-linked glycans containing core α1,3-fucose. These structures can be cleaved using fucosidase or *N*-glycosidase A (PNGase A).
15. POROS R2 can be used for desalting peptides. POROS OligoR3 is designed to bind more hydrophilic species than R2 and the choice is sample dependent. Combining microcolumns packed with materials with increasing hydrophobicity can result in a further increase in the number of detected peptides in a mixture (36, 37).

References

1. Lowe, J. B., and Marth, J. D. (2003) A genetic approach to mammalian glycan function, *Annu Rev Biochem 72*, 643–691.
2. Apweiler, R., Hermjakob, H., and Sharon, N. (1999) On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database, *Biochim Biophys Acta 1473*, 4–8.
3. Ohtsubo, K., and Marth, J. D. (2006) Glycosylation in cellular mechanisms of health and disease, *Cell 126*, 855–867.
4. Varki, A., Cummings, R., Esko, J. D., Freeze, H., Stanley, P., Bertozzi, C. R., Hart, G. W., and Etzler, M. E. (2008) Essentials in Glycobiology, Cold Spring Harbor Press, Cold Spring Harbor, NY.

5. Varki, A. (1993) Biological roles of oligosaccharides: all of the theories are correct, *Glycobiology* ***3***, 97–130.
6. Hart, G. W., Housley, M. P., and Slawson, C. (2007) Cycling of O-linked beta-N-acetylglucosamine on nucleocytoplasmic proteins, *Nature* ***446***, 1017–1022.
7. Rudd, P. M., Elliott, T., Cresswell, P., Wilson, I. A., and Dwek, R. A. (2001) Glycosylation and the immune system, *Science* ***291***, 2370–2376.
8. Freeze, H. H. (2006) Genetic defects in the human glycome, *Nat Rev Genet 7*, 537–551.
9. Freeze, H. H., and Aebi, M. (2005) Altered glycan structures: the molecular basis of congenital disorders of glycosylation, *Curr Opin Struct Biol* ***15***, 490–498.
10. Kim, Y. J., and Varki, A. (1997) Perspectives on the significance of altered glycosylation of glycoproteins in cancer, *Glycoconj J* ***14***, 569–576.
11. Kornfeld, R., and Kornfeld, S. (1985) Assembly of asparagine-linked oligosaccharides, *Annu Rev Biochem* ***54***, 631–664.
12. Hart, G. W. (1992) Glycosylation, *Curr Opin Cell Biol* ***4***, 1017–1023.
13. Low, M. G. (1989) Glycosylphosphatidylinositol: a versatile anchor for cell surface proteins, *FASEB J* ***3***, 1600–1608.
14. Haynes, P. A. (1998) Phosphoglycosylation: a new structural class of glycosylation? *Glycobiology* ***8***, 1–5.
15. Hartmann, S., and Hofsteenge, J. (2000) Properdin, the positive regulator of complement, is highly C-mannosylated, *J Biol Chem* ***275***, 28569–28574.
16. Schauer, R. (2000) Achievements and challenges of sialic acid research, *Glycoconj J* ***17***, 485–499.
17. Varki, N. M., and Varki, A. (2007) Diversity in cell surface sialic acid presentations: implications for biology and disease, *Lab Invest* ***87***, 851–857.
18. Thomas, P. (1996) Cell surface sialic acid as a mediator of metastatic potential in colorectal cancer, *Cancer J* ***9***, 1–10.
19. Fozzard, H. A., and Kyle, J. W. (2002) Do defects in ion channel glycosylation set the stage for lethal cardiac arrhythmias? *Sci STKE* ***2002***, pe19.
20. Rhim, A. D., Stoykova, L. I., Trindade, A. J., Glick, M. C., and Scanlin, T. F. (2004) Altered terminal glycosylation and the pathophysiology of CF lung disease, *J Cyst Fibros* ***3*** (Suppl 2), 95–96.
21. Goodarzi, M. T. (2008) Changes in sialylation of low-density lipoprotein in coronary artery disease, *Eur J Lipid Sci Tech* ***110***, 302–306.
22. Coppo, R., and Amore, A. (2004) Aberrant glycosylation in IgA nephropathy (IgAN), *Kidney Int* ***65***, 1544–1547.
23. Stray, S. J., Cummings, R. D., and Air, G. M. (2000) Influenza virus infection of desialylated cells, *Glycobiology* ***10***, 649–658.
24. Zhao, J., Simeone, D. M., Heidt, D., Anderson, M. A., and Lubman, D. M. (2006) Comparative serum glycoproteomics using lectin selected sialic acid glycoproteins with mass spectrometric analysis: application to pancreatic cancer serum, *J Proteome Res* ***5***, 1792–1802.
25. Yang, Z., and Hancock, W. S. (2005) Monitoring glycosylation pattern changes of glycoproteins using multi-lectin affinity chromatography, *J Chromatogr A* ***1070***, 57–64.
26. Zeng, Y., Ramya, T. N., Dirksen, A., Dawson, P. E., and Paulson, J. C. (2009) High-efficiency labeling of sialylated glycoproteins on living cells, *Nat Methods* ***6***, 207–209.
27. Nilsson, J., Ruetschi, U., Halim, A., Hesse, C., Carlsohn, E., Brinkmalm, G., and Larson, G. (2009) Enrichment of glycopeptides for glycan structure and attachment site identification, *Nat Methods* ***6***, 809–811.
28. Lewandrowski, U., Zahedi, R. P., Moebius, J., Walter, U., and Sickmann, A. (2007) Enhanced N-glycosylation site analysis of sialoglycopeptides by strong cation exchange prefractionation applied to platelet plasma membranes, *Mol Cell Proteomics* ***6***, 1933–1941.
29. Larsen, M. R., Jensen, S. S., Jakobsen, L. A., and Heegaard, N. H. (2007) Exploring the sialiome using titanium dioxide chromatography and mass spectrometry, *Mol Cell Proteomics* ***6***, 1778–1787.
30. Fujiki, Y., Hubbard, A. L., Fowler, S., and Lazarow, P. B. (1982) Isolation of intracellular membranes by means of sodium carbonate treatment: application to endoplasmic reticulum, *J Cell Biol* ***93***, 97–102.
31. Mueller, L. N., Brusniak, M. Y., Mani, D. R., and Aebersold, R. (2008) An assessment of software solutions for the analysis of mass spectrometry based quantitative proteomics data, *J Proteome Res 7*, 51–61.
32. Cox, J., and Mann, M. (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification, *Nat Biotechnol* ***26***, 1367–1372.
33. Ong, S. E., and Mann, M. (2006) A practical recipe for stable isotope labeling by amino acids in cell culture (SILAC), *Nat Protoc* ***1***, 2650–2660.

34. Pizzio, L. R. (2005) Mesoporous titania: effect of thermal treatment on the texture and acidic properties, *Mater Lett.* ***59***, 994.
35. Angel, P. M., Lim, J. M., Wells, L., Bergmann, C., and Orlando, R. (2007) A potential pitfall in ^{18}O-based N-linked glycosylation site mapping, *Rapid Commun Mass Spectrom* ***21***, 674–682.
36. Larsen, M. R. (2003) Mass spectrometric characterization of posttranslationally modified proteins-phosphorylation, *Methods Mol Biol* ***251***, 245.
37. Larsen, M. R., Cordwell, S. J., Roepstorff, P. (2002) Graphite powder as an alternative to reversed phase material for desalting and concentration of peptide mixtures prior to mass spectrometric analysis, *Proteomics* ***2***, 1277–1287.

Chapter 22

Chemical De-O-glycosylation of Glycoproteins for Applications in LC-Based Proteomics

Franz-Georg Hanisch

Abstract

This paper describes a cyclic on-column procedure for the sequential degradation of complex *O*-glycans on proteins by periodate oxidation of sugars and cleavage of oxidation products by elimination. Glycoproteins are immobilized to alkali-stable, reversed-phase Poros 20 beads, desialylated by treatment with dilute trifluoroacetic acid, and de-O-glycosylated by two degradation cycles before the eluted apoproteins are digested with trypsin for analysis by liquid chromatography electrospray ionization-mass spectrometry. Even complex glycan moieties are removed under mild conditions with only minimal effects on structural integrity of the peptide core by fragmentation, dehydration, or racemization of lysine and arginine residues. The protocol is also applicable on gel-immobilized glycoproteins after 1D or 2D gel electrophoresis. Conversion of *O*-glycoproteins into their corresponding apoproteins results in facilitated accessibility of tryptic cleavage sites, increases the numbers of peptide fragments, and accordingly enhances protein coverage and identification rates within the subproteome of mucin-type *O*-glycoproteins. The protocol is suitable for automatization, but due to partial elution from the Poros 20 columns it is not recommended for applications on the glycopeptide level.

Key words: Chemical deglycosylation, *O*-glycoproteins, *O*-glycoproteome, liquid chromatography-mass spectrometry, proteomics.

Abbreviations: HGA, human glycophorin A; TFMSA, trifluoromethanesulfonic acid;

1. Introduction

O-Glycoproteins, in particular those with clustered mucin-type glycans, still represent a challenge for scientists interested in the development of robust protocols, which allow a highly parallel analysis of complex proteomes (comprising glycosylated and non-glycosylated proteins). In particular, proteomic approaches

K. Gevaert, J. Vandekerckhove (eds.), *Gel-Free Proteomics*, Methods in Molecular Biology 753,
DOI 10.1007/978-1-61779-148-2_22, © Springer Science+Business Media, LLC 2011

based on 2D gel electrophoresis still encounter a variety of problems. These are caused by the charge heterogeneity of highly sialylated or sulfated glycoproteins and by their unusually high molecular masses. There is still no sensitive stain for the selective detection of glycoproteins in gels, and in particular no stain which would work equally well with both glycosylated and non-glycosylated proteins, and over a broad dynamic range. Densely O-glycosylated proteins are highly resistant to proteases, in particular to proteolysis with site-specific endopeptidases, like trypsin. Finally, the identification rates in proteome analyses via mass spectrometric peptide mapping are strongly affected by the reduction of protein coverage caused by complex O-glycosylation.

Numerous attempts were made to solve the above outlined problems by reducing the structural complexity of the O-linked glycans, either by their removal or by substituting these with simple, low molecular mass tags. As a matter of fact, there is no endoglycosidase which would cleave O-linked glycans in a similar, structurally largely independent way as PNGase F in case of N-linked chains. The existence of such an enzyme in a *Streptomyces* strain, although claimed in publications (1), could not be confirmed by others (Razawi, Schwientek, and Hanisch, unpublished results). A commercially available *O*-glycanase from *Diplococcus*, on the other hand, exhibits a high degree of specificity, which restricts its application to the cleavage of O-linked disaccharide Gal1-3GalNAc (2). Chemical protocols suffer from a variety of drawbacks. First of all, anhydrous hydrazine, which is known to cleave selectively O-linked chains at 60°C within several hours (3), simultaneously degrades the protein backbone. By contrast, anhydrous trifluoromethane sulfonic acid (TFMSA), which can be used at 0°C for partial deglycosylation of O- and N-linked proteins (4), has proven to be incompatible with the need for high sensitivity in proteomic protocols. TFMSA can find application in larger scales, but has severe limitations arising from low protein recovery. A combination of partial de-O-glycosylation with TFMSA and mild periodate oxidation of resistant core GalNAc followed by β-elimination of the sugar as a dehydro derivative indicated that the peptide-linked hexosamine can be cleaved from the protein at pH 10.5 without changing the protein structure (5). Ammonia has been reported in an application of β-elimination/Michael addition on O-linked glycans and their replacement by the nucleophilic base (6). A disadvantage of this protocol is the long reaction time in 25% ammonium hydroxide at high temperature, which causes partial degradation of some proteins and dehydration at non-glycosylated serine and threonine residues. A variant of this method uses alkylamines, like ethylamine or methylamine, which both catalyze the β-elimination process and add to the dehydro intermediates as nucleophiles (7). This technique is less affected by unspecific dehydration and

offers the advantage that uniform mass tags with increments of 27 or 13 Da are introduced, which are better discriminated from non-glycosylated positions than the tag decline of 1 Da in case of ammonia. However, also this approach suffers from side reactions, like the addition of alkylamines to asparagine and glutamine side chains, and several proteins were found partially or totally degraded on prolonged treatments.

The principle of the protocol presented here is based on mild selective periodate oxidation of vicinally hydroxylated sugar carbons resulting in aldehydic moieties that are readily cleaved by base-catalyzed elimination from the subterminal sugar. Part of the reaction scheme, the degradation of the core GalNAc, has previously been applied to remove sialyl-Tn and Tn epitopes from mucin-type glycoproteins (5) and to improve thereby immunochemical accessibility of the protein core (8). Our observation that not only the core sugar (Tn, GalNAc) but also sugars of the glycan chains can be consecutively cleaved from the non-reducing ends has led to the development of a robust and sensitive proteomic application (9). The mild conditions of quantitative sugar cleavage avoid unspecific peptide fragmentation and dehydration observed during prolonged base treatments at higher temperatures. Only two cycles of treatment are principally required even for the removal of longer and complex glycans, if the periodate-resistant polylactosamines extend from the C6 branch of the core GalNAc. Up to three cycles were run to demonstrate feasibility of the protocol with respect to the structural integrity of the protein cores. Since the proteins are immobilized on a chemically resistant reversed-phase matrix, all treatments and washing steps can be performed on-column making the procedure applicable for automatization.

2. Materials

2.1. Glycopeptides and Glycoproteins

1. Glycopeptides used in this study were based on the 20-meric MUC1 tandem repeat sequence and were chemically synthesized by Prof. Hans Paulsen, Institute of Organic Chemistry, University of Hamburg, Germany.
2. Glycoproteins such as human glycophorin A and bovine fetuin are commercially available or were recombinantly expressed fusion proteins isolated from culture supernatants of EBNA-293 cells (10, 11).

2.2. Labware, Reagents, Buffers, and Solutions

1. SelfPack Poros 20R1 and R2 (Applied Biosystems) suspended in 8 ml of aqueous ethanol and stored at 4–8°C.

2. Mobicol columns including screw caps and Luer-lock caps and filter insertion tool with lower filter (2.7 mm, 35 μm pore size) and upper filter (6.8 mm, 35 μm pore size) (MoBiTec).
3. Activation solution: 80% acetonitrile and 0.1% aqueous trifluoroacetic acid.
4. Equilibration solution: 0.1% trifluoroacetic acid.
5. Oxidation buffer: 10 mM sodium metaperiodate in 0.1 M sodium acetate, pH 4.5.
6. Elimination solution: 25% aqueous ammonia.
7. Washing solution: 0.1% trifluoroacetic acid or bidistilled water.
8. Elution solution: 90% isopropanol and 0.1% trifluoroacetic acid.
9. Digestion buffer: 50 mM ammonium hydrogen carbonate, pH 8.0.
10. Sequencing-grade modified trypsin (porcine, Promega).
11. Resolubilization solution: 1% formic acid.

2.3. Instruments, Mass Spectrometers

1. Liquid chromatography (LC)-MS data can be acquired on a Q-TOF2 quadrupole-TOF mass spectrometer (Micromass) equipped with a Z-spray source. Samples are introduced by an Ultimate Nano-LC system (LC Packings) equipped with a Famos autosampler and a Switchos column-switching module. The column setup comprises a 0.3 mm by 10 mm trapping column and a 0.075 by 150 mm analytical column, both packed with 3 μm Atlantis dC18 (Waters). The electrospray ionization (ESI) interface comprises a 20 μm internal diameter and 90 μm outer diameter tapered spray emitter (Carbotec) linked to the HPLC flow path using a 7 nl dead volume stainless steel glass capillary mounted onto the PicoTip holder assembly (New Objective). Of course, the described LC-MS parameters represent only an example that can be replaced by other related setups.
2. Liquid chromatography (LC)-MS data can also be acquired on an HCT ETD II ion trap mass spectrometer (Bruker Daltonics, Bremen, Germany) equipped with a nano-ESI source. Samples are introduced via an EASY nano-LC system (Proxeon, Odense, Denmark) using a vented column setup comprising a 0.1 × 20 mm trapping column and a 0.075 × 100 mm analytical column, both self-packed with ReproSil-Pur C18-AQ, 5 μm (Dr. Maisch, Ammerbuch, Germany).

3. Methods

3.1. On-Column De-O-glycosylation

The principle of cyclic sugar oxidation and elimination of *O*-glycans is shown in **Fig. 22.1** (9).

1. Pack 50 µl of Poros 20 R1 beads (binding characteristics similar to C4) or R2 beads (similar to C8) into Mobicol columns, activate with 80% acetonitrile in 0.1% aqueous trifluoroacetic acid (TFA) (two aliquots of 0.5 ml) and equilibrate with 0.1% aq. TFA (three aliquots of 0.5 ml). Press the volumes through the column bed using a syringe with Luer adapter.
2. Apply proteins and glycoproteins (0.1–30 µg) in water, aqueous buffer, or 0.1 M TFA for immobilization on the resins and wash with three aliquots (0.5 ml) of water.
3. Desialylate sialoglycoproteins chemically, if necessary, by treatment with 0.1 M trifluoroacetic acid at 80°C for 1 h (*see* **Note 1**).
4. Wash with two 0.5 ml aliquots of water and oxidize sugars with oxidation buffer by passing approximately half of a 0.5 ml aliquot through the column, closing the outlet and incubating the column at 37°C for 1 h.
5. After three washing cycles (0.5 ml water) apply a 0.5 ml aliquot of 25% ammonium hydroxide in the same way and

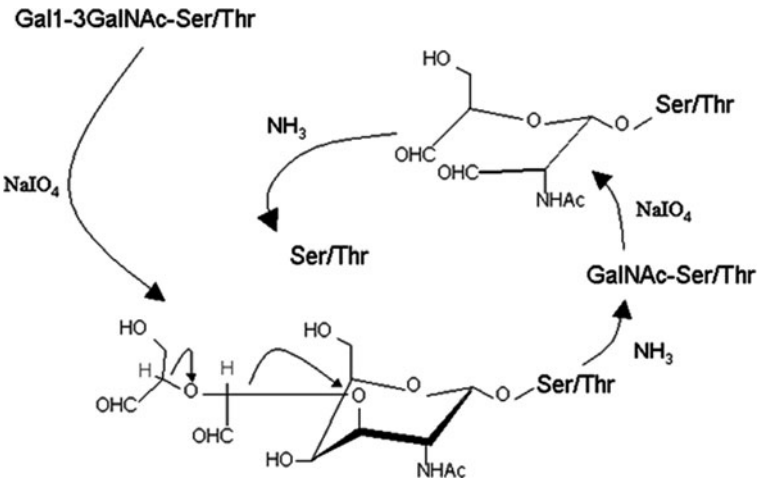


Fig. 22.1. Reaction scheme of cyclic glycan degradation by consecutive periodate oxidation and base-catalyzed hydrolysis. The sequential elimination of peptide-linked Gal-GalNAc by two cycles of periodate oxidation and base-catalyzed removal of oxidized sugar derivatives is shown.

incubate at 22°C for 1 h. The number of cycles before removal of core GalNAc(*n*) is determined by the length (number of monosaccharides) of the GalNAc-C3-branch (*see* **Notes 2** and **3** and (9)).

6. To remove core GalNAc in the second (or last) cycle, perform the base treatment with 25% ammonia at 37°C for 2 h (*see* **Note 4**).
7. After a final washing with water, elute the proteins with 90% isopropanol in 0.1% TFA (1 ml) and dry by vacuum centrifugation (*see* **Note 5**). Samples are either used for 1-DE or further processed by proteolysis.
8. Prior to digestion with trypsin, reduce and alkylate the proteins using standard protocols, resolubilize in 50 mM ammonium hydrogen carbonate, pH 8.0 (20 μl), and add 1 μg/μl of trypsin prior to overnight incubation at 37°C.
9. Dry the sample by vacuum centrifugation and add 10–50 μl of 1% formic acid for resolubilization of peptides.

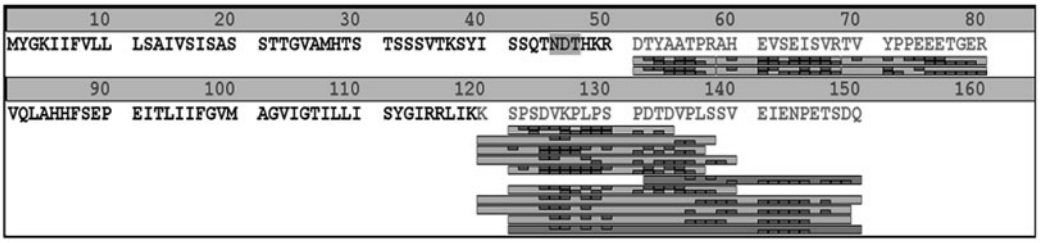
3.2. In-Gel Experiments

1. Excise protein bands from the gel and transfer into Eppendorf vials.
2. Wash protein bands twice with 0.5 ml of 50% acetonitrile for 10 min at room temperature.
3. Wash protein bands twice with 0.5 ml of water for 10 min at room temperature.
4. Vacuum dry the gel pieces.
5. Treat the vacuum-dried gel by two oxidation/β-elimination cycles (refer to **Section 3.1**, steps 4–6) with intermediate washing steps as above.
6. Equilibrate deglycosylated proteins in digestion buffer and in-gel digest with trypsin (1 μg in 20 μl). After heat inactivation of the protease, the generated peptide products can be optionally treated with V8 (1 U in PBS, at 37°C for 18 h).
7. Elute peptides (each for 1 h) into 100 μl aliquots of water, and acetonitrile:water (1:1), dry and prepare samples for MALDI-MS by PepClean C18 spin column application. An example of an in-gel application is shown in **Fig. 22.2** for human glycophorin A (*see* **Notes 6** and 7).

3.3. LC-ESI-Tandem Mass Spectrometry

3.3.1. QTOF2 Analysis

1. Dilute samples in 1% formic acid and inject a total of 10 μl onto the trap column.
2. After desalting for 1 min with 1% formic acid and a flow rate of 10 μl/min, switch the trap column into the analytical flow path and elute peptides onto the analytical column by a gradient of 5% acetonitrile in 1% formic acid to 40% acetonitrile in 1% formic acid over 35 min and a column flow



Cmpd.	No. of Cmpds.	m/z meas.	Δ m/z [ppm]	z	Rt [min]	Score	P	Range	Sequence	Modification
15	2	447.7219	5.53	2	9.12	32.29	0	51-58	R.DTYAATPR.A	
6	2	549.2519	-13.128	2	8.09	16.75	0	51-58	R.DTYAATPR.A	HexNAc: 6
22	1	563.8126	29.098	2	10.02	60.49	0	59-68	R.AHEVSEISVR.T	
25	1	665.3238	-18.158	2	10.10	15.67	0	59-68	R.AHEVSEISVR.T	HexNAc: 5
50	2	469.5096	-91.696	3	11.47	38.22	0	69-80	R.TVYPPEEETGER.V	
47	1	537.2094	-67.717	3	11.18	14.50	0	69-80	R.TVYPPEEETGER.V	HexNAc: 1
182	1	946.4547	-52.304	2	16.33	33.07	1	120-137	K.KSPSDVKPLPSPDTDVPL.S	
146	1	660.3406	-13.082	3	15.21	11.39	1	120-138	K.KSPSDVKPLPSPDTDVPLS.S	
195	1	722.3501	-45.156	3	16.82	26.95	1	120-140	K.KSPSDVKPLPSPDTDVPLSSV.E	
226	1	1060.4349	-80.98	3	18.17	30.10	1	120-149	K.KSPSDVKPLPSPDTDVPLSSVEIENPETSD.Q	
223	1	1103.0938	-102.586	3	18.15	61.27	1	120-150	K.KSPSDVKPLPSPDTDVPLSSVEIENPETSDQ.-	
71	1	727.7925	-84.637	2	12.32	38.54	0	121-134	K.SPSDVKPLPSPDTD.V	
213	1	882.4290	-31.417	2	17.58	80.70	0	121-137	K.SPSDVKPLPSPDTDVPL.S	
228	1	1018.9551	-66.593	2	18.22	41.06	0	121-140	K.SPSDVKPLPSPDTDVPLSSV.E	
250	1	1017.7661	-55.373	3	19.34	36.29	0	121-149	K.SPSDVKPLPSPDTDVPLSSVEIENPETSD.Q	
252	1	1060.7273	206.172	3	19.42	44.30	0	121-150	K.SPSDVKPLPSPDTDVPLSSVEIENPETSDQ.-	
222	1	827.6755	-30.88	3	17.93	32.39	0	132-150	P.DTDVPLSSVEIENPETSDQ.-	HexNAc: 7, 8

Fig. 22.2. Nano-LC-electrospray MS/MS of tryptic fragments from human glycophorin A after de-O-glycosylation. Human glycophorin A was run in a 1D SDS gel prior to in-gel de-O-glycosylation. Tryptic peptides generated by in-gel digestion were analyzed by LC-MS on an HCTultra (Bruker Daltonics, Bremen, Germany). The SwissProt database (release from 2009-06-03) was searched allowing HexNAc modifications. Human glycophorin A was identified with a score of 530.7 based on 17 unique peptides. The sequence coverage was 40.67%. Note in particular the coverage of the O-glycosylated region p51–80. Four incompletely de-O-glycosylated peptides with GalNAc modification were identified and the sites assigned on the basis of MS/MS spectra.

rate of 200 nl/min, resulting from a 1:1,000 split of the 200 μl/min flow delivered by the pump. One LC cycle is finished within an hour. Apply 1.7–2.4 kV to the stainless steel union for establishing a stable nanospray.

3. The data-dependent acquisition of MS and tandem MS (MS/MS) spectra is controlled by the Masslynx 4.0 software. Survey scans of 1 s generally cover the range from m/z 400 to 1,400. Select doubly and triply charged ions rising above a given threshold for MS/MS experiments. In MS/MS mode, scan the mass range from m/z 40 to 1,400 in 1 s and add up five scans for each experiment. Use the Proteinlynx software module in Masslynx 4.0 to generate Micromass-formatted peak lists from the raw data.
4. Upload the peak lists to MASCOT Deamon (version 1.9), which triggers batch searches in NCBInr using a local installation of MASCOT version 1.9. Restrict searches to the respective species and use trypsin specificity with one missed

cleavage allowed. Set the maximum mass error to 0.5 Da for peptide and fragment spectra, the state of the cysteine to carbamidomethyl, and allow an optional oxidation of methionine.

3.3.2. HCTultra Analysis

1. Aspirate 5–18 µl of the samples into the sample loop and load a total of 25 µl onto the trap column using a flow rate of 6 µl/min. Loading pump buffer is 0.1% formic acid (FA).
2. Elute peptides with a gradient of 0–35% acetonitrile in 0.1% formic acid over 20 min and a column flow rate of 300 nl/min.
3. Subsequently raise the acetonitrile content to 100% over 2 min and regenerate the column in 100% acetonitrile for additional 8 min.
4. Control data-dependent acquisition of MS and tandem MS (MS/MS) spectra using the Compass 3.0 software.
5. Acquire MS1 scans in standard enhanced mode. Combine five single scans in the mass range from m/z 400 to 1,400 for one survey scan. Select up to three doubly and triply charged ions rising above a given threshold for MS/MS experiments. Use the ultrascan mode for the acquisition of MS2 scans in the mass range from m/z 100 to 1,600 and add up three single scans. Set the ion charge control value to 250,000 for all scan types.
6. Use the Data Analysis software module to generate peak lists in Mascot generic format (mgf).

4. Notes

1. Desialylation of the glycans prior to their degradation is necessary, because the pyranosidic ring structure of sialic acids remains stable to oxidation with mild periodate and base-catalyzed elimination of the oxidized C7 product does not take place. Desialylation can easily be achieved by chemical hydrolysis under mild conditions using 0.1 M TFA at 80°C for 1 h. In solution sialoglycoproteins undergo considerable fragmentation under these conditions. However, due to the immobilization on Poros beads glycoproteins remain largely stable during chemical hydrolysis. Alternatively, sialoglycoproteins can be desialylated enzymatically by using *Clostridium perfringens* sialidase prior to application onto the Poros column.
2. One limitation arises from partial elution of peptides from Poros R1 and R2 columns during base treatment and

washing off from the column after treatment. Under alkaline conditions peptides become more polar by increasing charge states at acidic and basic amino acid side chains. Depending on the length and amino acid composition of the peptides they can hence elute partially or completely from the reversed-phase column, when the pH reaches values above 12.

3. Core 2-based glycans are characterized by the common trisaccharide core Gal1-3(GlcNAc1-6)GalNAc. This trisaccharide can be elongated at both branches. Generally, the number of cycles needed for complete removal of O-linked glycans is determined by the number of monosaccharides *n* linked to C3 of the core GalNAc and is equal to *n* +1. However, the majority of O-linked glycans found on mammalian glycoproteins is core 2-based and elongations occur preponderantly at the C6-linked branch. Accordingly, for a majority of *O*-glycoproteins with non-elongated Gal1-3 branches the desialylated glycans can be liberated with only two cycles. This feature allows for the ready de-O-glycosylation in two degradation cycles of even complex core 2-based glycans as found on many mucins.
4. The core GalNAc residue is more stable during oxidation and elimination and requires hence more drastic conditions by enhancing the temperature and extending the reaction time during base treatment. Under these conditions besides the oxidized substrate and intermediates, the major products are generally accompanied by minor degradation products (water elimination products) and peptide fragments (N- or C-terminal fragmentation). A feasible compromise between quantitative hydrolysis on the one hand, which was achieved after overnight treatment, and increasing degradation on the other hand was found in 1 or 2 h treatments at 37°C, after which the deglycosylated product represents at least 70%. Multiply O-glycosylated peptides may escape detection due to accumulative effects caused by the incomplete core GalNAc removal. To overcome the problem, the deglycosylation procedure can be elongated by a further degradation cycle. In the presence of strong bases the peptide core could undergo unspecific fragmentation and racemization of the amino acids. For this reason, different protocols for the elimination of 3,4-dialdehydic moieties from the peptide were evaluated for various time periods and at different temperatures. To assess the degree of degradation the samples were run on reversed-phase HPLC and compared to a control sample treated in water (9). All proteins tested were stable during 2 h treatments with aq. ammonia at 37°C. However, on prolongation of the base treatment using aq. ammonia

for 18 h a considerable degradation became evident. Racemization of L-amino acids during base treatment could represent a limitation for the following trypsin cleavage at Arg and Lys. For this reason peptide samples were digested with the endopeptidase prior to or after base treatment, but no reduction in cleavage efficiency could be observed (9).

5. Using acetonitrile, the elution efficiency is slightly lower compared to isopropanol at 90%. Under the latter conditions the elution efficiency is apparently close to 100% referring to mixtures of standard proteins.
6. The de-O-glycosylation protocol can be applied on polyacrylamide gel-immobilized glycoproteins after 1D or 2D electrophoresis. The bands can be cut and treated as described above or the entire gel be run through the two deglycosylation cycles with intermediate washing and equilibration steps (Hanisch F.-G., manuscript in preparation). After restaining with coomassie the protein spots, which lost their stain during the ammonia treatment, reappear with no apparent loss in staining intensity or even at higher intensity, if glycans were removed from *O*-glycoproteins.
7. The described method suffers from its apparent inapplicability in 2-DE gel-based proteomics (not shown). 2-DE applications revealed severe affection of the performance, in particular with respect to the separation of treated proteins by isoelectric focussing. The reasons for these technical problems are not known yet and will have to be identified in further studies. While protein degradation under strong alkaline conditions does not play a major role according to quantitative HPLC data, the apparent molecular masses of the de-O-glycosylated proteins were generally larger than the expected nominal masses of the apoproteins. This can be explained in part by incomplete removal of the glycans, but other reasons cannot be ruled out.

References

1. Ishii-Karakasa, I., Iwase, H., Hotta, K. (1997) Structural determination of the O-linked sialyl oligosaccharides liberated from fetuin with endo-α-N-acetylgalactosaminidase-S by HPLC analysis and 600-MHz 1H-NMR spectroscopy. *Eur. J. Biochem.* **247**, 709–715
2. Umemoto, J., Bhavanandan, V.P., Davidson, E.A. (1977) Purification and properties of an endo-α-N-acetyl-D-galactosaminidase from Diplococcus pneumoniae. *J. Biol. Chem.* **252**, 8609–8614
3. Patel, T.P., Parekh, R.B. (1994) Release of oligosaccharides from glycoproteins by hydrazinolysis. *Methods Enzymol.* **230**, 57–66
4. Sojar, H.T., Bahl, O.M.P. (1987) Chemical deglycosylation of glycoproteins. *Arch. Biochem. Biophys.* **259**, 52–57
5. Gerken, T.A., Gupta, R., Jentoft, N. (1992) A novel approach for chemically deglycosylating O-linked glycoproteins. The deglycosylation of submaxillary and respiratory mucins. *Biochemistry* **31**, 639–648

6. Rademaker, G.J., Pergantis, S.A., Blok-Tip, L.A., Langridge, J.I., et al. (1998) Mass spectrometric determination of the sites of O-glycan attachment with low picomolar sensitivity. *Anal. Biochem.* **257**, 149–160
7. Hanisch, F.-G., Jovanovic, M., Peter-Katalinic, J. (2001) Glycoprotein identification and localization of O-glycosylation sites by mass spectrometric analysis of deglycosylated/alkylaminylated peptide fragments. *Anal. Biochem.* **290**, 47–59
8. Hong, J.C., Kim, Y.S. (2000) Alkali-catalyzed beta-elimination of periodate-oxidized glycans: a novel method of chemical deglycosylation of mucin gene products in paraffin-embedded sections. *Glycoconj. J.* **17**, 691–703
9. Hanisch, F.-G., Teitz, S., Schwientek, T., Müller, S. (2009) Chemical de-O-glycosylation of glycoproteins for application in LC-based proteomics. *Proteomics* **9**, 710–719
10. Müller, S., Hanisch, F.G. (2002) Recombinant MUC1 probe authentically reflects cell-specific O-glycosylation profiles of endogenous breast cancer mucin. *J. Biol. Chem.* **277**, 26103–26112
11. Engelmann, K., Kinlough, C.L., Müller, S., Razawi, H., et al. (2005) Transmembrane and secreted MUC1 probes show trafficking-dependent changes in O-glycan core profiles. *Glycobiology* **15**, 1111–1124

Chapter 23

Ubiquitination and Degradation of Proteins

Yelena Kravtsova-Ivantsiv and Aaron Ciechanover

Abstract

Modification by ubiquitin (Ub) and ubiquitin-like proteins (UbLs) is involved in the regulation of numerous cellular processes and has therefore become an important subject of research in various areas of biomedicine. The large number of components of the system (~1,500), most of them being ligases (~800) that recognize their target substrates specifically, along with the complexity of the ubiquitination process, mostly the synthesis of the hallmark polyubiquitin chains, has rendered studies of many of the processes related to the activity of the system resistant to detailed mechanistic analysis. Thus, our knowledge of the modes of recognition of target substrates by ligases and of consensus ubiquitination sites is sparse. We also lack basic tools such as antibodies directed against specific internal polyubiquitin chain linkages and analytical methods to decipher the structure of intact chains and their formation. All these tools are essential in order to understand the mechanisms that underlie the diverse activities of the system, proteolytic as well as non-proteolytic, and the manner in which it exerts its high specificity and selectivity toward its myriad substrates. Here we describe selected basic procedures that allow one to become acquainted with this rapidly evolving field, realizing that one cannot provide a comprehensive coverage of all or even a small part of the methodologies related to this research area. We provide information on how to set up a cell-free system for ubiquitination – a powerful tool that enables researchers to reconstitute the modification from purified components – and how to identify ubiquitin adducts in cells. Additionally, we describe methods to follow stability (degradation) of proteins in cell-free systems and in cells.

Key words: Ubiquitin, ubiquitin aldehyde, ubiquitination, degradation, reticulocyte lysate, proteolytic substrates.

1. Introduction

Covalent modification of proteins by ubiquitin (Ub) and ubiquitin-like proteins (UbL) is involved in the regulation of numerous cellular pathways. Among them are cell cycle and division, growth and differentiation, apoptosis, maintenance of the

K. Gevaert, J. Vandekerckhove (eds.), *Gel-Free Proteomics*, Methods in Molecular Biology 753,
DOI 10.1007/978-1-61779-148-2_23, © Springer Science+Business Media, LLC 2011

cell's quality control, and the response to stress. The ability of the ubiquitin system to regulate such a broad array of processes is due to the fact that the covalent tagging generates an extremely diverse form of a protein–protein interaction module that, according to its structure, enables it to associate with numerous different downstream effectors. Furthermore, the same target protein can be modified by different modules under distinct pathophysiological conditions, thus being targeted to disparate fates. In many cases, the modification is followed by targeting of the tagged proteins to proteasomal or lysosomal degradation, thus terminating their function (1–4). Yet, in many other processes, the modification does not lead to destruction of the target and is therefore reversible. Thus, mono-, oligo-, multiple mono-, and polyubiquitination along with targeting of non-lysine residues in the substrates and generation of chains based on linkages to different internal lysines have converted this novel mode of post-translational modification into an immense regulatory platform. Adding to this complexity is obviously the same set of modifications mediated by UbLs. Thus, ubiquitin and UbLs can be regarded as modules that mediate interaction in trans of the modified substrates with downstream effectors, an interaction that is responsible for the execution of the specific function. In the case of degradation, this downstream effector is the 26S proteasome complex. With the myriad of targeted substrates and numerous diverse processes regulated, it has not been surprising to find that aberrations in the system have been implicated in the pathogenesis of many diseases – among them malignancies, inflammatory and immune disorders, and neurodegeneration. This has consequently led to efforts to develop mechanism- and processes-based drugs; one successful drug to combat plasma cell leukemia (multiple myeloma) is already in the market. The present protocols describe methods to monitor protein ubiquitination and degradation both in cell-free reconstituted systems and in intact cells.

2. Materials

All solutions are made in ddH_2O unless otherwise indicated. Enzo® Life Sciences and Boston Biochem Inc. produce a large selection of ubiquitin system reagents which we have used successfully along the years; however, other suppliers are also available.

2.1. Preparation of Crude Cell Extracts for Monitoring Conjugation and Degradation of Protein Substrates in a Cell-Free System

2.1.1. Preparation of Reticulocyte Lysate

1. New Zealand white rabbits (preferably females) of approximately 2 kg body weight (2–3 months old) for preparation of reticulocyte lysate. Experiments involving live animals must be carried out according to international, national, and institutional regulations.
2. Phenylhydrazine (toxic, dangerous for the environment).
3. Phosphate-buffered saline (PBS): 150 mM NaCl, 10 mM potassium phosphate buffer, pH 7.4.
4. Staining solution: 1 g methylene blue, 0.4 g sodium citrate, 0.85 g sodium chloride per 100 ml of aqueous solution. Dissolve the sodium citrate first in the saline buffer and then dissolve the dye. Filter this solution and store at 4°C.
5. Krebs–Ringer phosphate buffer (KRP): 130 mM NaCl, 5 mM KCl, 1.3 mM $CaCl_2$, 1.3 mM $MgSO_4$, 10 mM potassium phosphate buffer, pH 7.4.
6. KRP with ATP-depleting reagents: KRP that contains 20 mM 2-deoxy-D-glucose (diluted from a 1 M stock solution) and 0.2 mM 2,4-dinitrophenol (2,4-DNP, diluted from a 20 mM stock solution). To prepare the stock solution of 2,4-DNP, add the volume of H_2O yielding the 20 mM concentration. Then, gradually add solid $NaHCO_3$ until all 2,4-DNP gets dissolved.
7. 1, 4-Dithiothreitol (DTT).

2.1.2. Preparation of Extract from Cultured Cells

1. Cultured cells in a monolayer or in suspension (tested for the absence of *Mycoplasma* contamination).
2. HEPES-saline: 20 mM HEPES, 124 mM NaCl, 4 mM KCl, 1.2 mM $MgSO_4$, 1 mM $CaCl_2$, pH 7.5.
3. 1 M stock solution of 1, 4-dithiothreitol.
4. Krebs–Ringer phosphate buffer (KRP): 130 mM NaCl, 5 mM KCl, 1.3 mM $CaCl_2$, 1.3 mM $MgSO_4$, 10 mM potassium phosphate buffer, pH 7.4.
5. KRP with ATP-depleting reagents: KRP that contains 20 mM 2-deoxy-D-glucose (diluted from a 1 M stock solution) and 0.2 mM 2,4-dinitrophenol (2,4-DNP, diluted from a 20 mM stock solution (*see* step 6 of **Section 2.1.1**)), 20 mM NaF, 10 mM sodium azide.
6. Nitrogen cavitation bomb in a volume of 45 ml (Parr Instrument Company).

2.2. Fractionation of Cell Extract to Fraction I and Fraction II

1. Diethylaminoethyl cellulose (DEAE cellulose, DE-52).
2. 300 mM potassium phosphate buffer, pH 7.0.

3. Buffer A: 3 mM potassium phosphate buffer, pH 7.0, 1 mM DTT.
4. Buffer B: 3 mM potassium phosphate buffer, pH 7.0, 20 mM KCl, 1 mM DTT.
5. Buffer C: 20 mM Tris-HCl, pH 7.2, 500 mM KCl, 1 mM DTT.
6. Buffer D: 20 mM Tris-HCl, pH 7.2, 1 mM DTT.
7. Ammonium sulfate (ACS reagent, ≥99.0).
8. Dialysis tubing.

2.3. Labeling of Proteolytic Substrates

2.3.1. Radioiodination of Proteins

1. 100 mM sodium phosphate, pH 7.5.
2. 10 mM NaI.
3. $Na^{125}I$, specific activity of 100–350 mCi/ml (PerkinElmer®) (*see* **Note 1**).
4. 10 mg/ml chloramine T, freshly dissolved in 50 mM sodium phosphate, pH 7.5.
5. 20 mg/ml sodium metabisulfite, freshly dissolved in 50 mM sodium phosphate, pH 7.5.
6. PD MiniTrap G-25 desalting column (GE Healthcare) equilibrated in 150 mM NaCl, 10 mM Tris-HCl, pH 7.6.
7. 1 M of Tris-hydroxymethyl aminomethane (Trizma® base; Tris buffer) stock solution, brought to the desired pH with concentrated HCl (*see* **Note 2**).
8. 3 M NaCl.

2.3.2. Biosynthetic Labeling of Proteins

1. Wheat Germ Flexi® Vectors (Promega).
2. RiboMAX™ Express Systems for generation of mRNA from linearized cDNA (RiboMAX™ Large Scale RNA Production System SP6 or T7, Promega).
3. 3 M sodium acetate, pH 4.5.
4. Phenol:chloroform:isoamyl alcohol (25:24:1), centrifuge for 5 min at 4,000 rpm and use the lower phase.
5. Chloroform:isoamyl alcohol (24:1).
6. Ethanol.
7. Nuclease-free water (Promega).
8. L-[^{35}S]Methionine – for in vitro translation (specific activity 1,000 Ci/mmol at 43.3 mCi/ml; PerkinElmer®) (*see* **Note 1**).
9. Wheat Germ Extract Plus or Rabbit Reticulocyte Lysate (Promega).

2.4. Conjugation of Ubiquitin to Proteolytic Substrates in a Cell-Free System

1. 1 M of Tris-hydroxymethyl aminomethane (Trizma® base; Tris buffer) stock solution, brought to the desired pH with concentrated HCl (*see* step 7 of **Section 2.3.1**).
2. 2.5 M magnesium chloride.
3. 1 M 1, 4-dithiothreitol.
4. Reticulocyte lysate or complete cell extract.
5. 1 M 2-deoxy-D-glucose.
6. 10 mg/ml suspension of hexokinase (HK) in 3.2 M ammonium sulfate solution (pH ~6.5) (Roche). Centrifuge the ammonium sulfate slurry and re-suspend to the original volume in 20 mM Tris-HCl, pH 7.6. Dilute further in the same buffer as necessary before setting up the reaction. The re-suspended enzyme can be stored at 4°C for ~4 weeks.
7. 25 mg/ml ubiquitin.
8. Ubiquitin aldehyde (UbAl; Enzo® Life Sciences or Boston Biochem Inc.). Use according to the manufacturer's instructions. Typically, the modified protein should be dissolved in aqueous, slightly acidic solution in a concentration of 10 mg/ml.
9. 0.1 M adenosine 5′-[γ-thio]triphosphate tetralithium salt (ATPγS) dissolved in ddH_2O and neutralized with NaOH.
10. 0.1 M adenosine 5′-triphosphate disodium salt (ATP) dissolved in ddH_2O and neutralized with NaOH.
11. 1 M phosphocreatine.
12. 10 mg/ml phosphocreatine kinase in 50 mM Tris-HCl, pH 7.6.

2.5. Degradation of Proteolytic Substrates in a Cell-Free System

Same as **Section 2.4**, but additionally

1. 100 mg/ml bovine serum albumin (BSA)
2. Trichloroacetic acid (TCA)

2.6. Ubiquitination of Proteolytic Substrates in Cells

1. MG 132 (*Z*-Leu-Leu-Leu-CHO), or epoxomicin, or *Z*-Leu-Leu-Leu-vinyl sulfone, or MG 262 [*Z*-Leu-Leu-Leu-B$(OH)_2$]or *clasto*-lactacystin β-lactone (Enzo® Life Sciences or Boston Biochem Inc.). Use according to the manufacturer's instructions. Typically, the inhibitors should be dissolved in DMSO to a concentration of 10–20 mM.
2. Phosphate-buffered saline (PBS).
3. "Hot" lysis buffer I: 1% SDS, 1 mM EDTA in PBS.

4. "Hot" lysis buffer II: 2% Triton® X-100, 0.5% sodium deoxycholate (DOC), 1% BSA, 1 mM EDTA, protease inhibitors mixture (Protease Inhibitor Cocktail Set 1, Calbiochem; diluted according to the manufacturer's instructions). Protease inhibitors mixture should be added to the lysis buffer just before use.
5. "Hot" lysis buffer III: 1% Triton® X-100, 1% SDS, 0.5% sodium deoxycholate, 1% BSA, 1 mM EDTA in PBS.
6. Immobilized proteins A and G suspended according to the manufacturer's instructions.
7. 2x sample buffer: 125 mM Tris-HCl, pH 6.8, 4% SDS, 20% glycerol, 1.4 M β_2-mercaptoethanol.

2.7. Degradation of Proteolytic Substrates in Cells

2.7.1. Cycloheximide Chase

1. Cultured cells in a monolayer or in suspension (tested for the absence of *Mycoplasma* contamination).
2. Cell culture medium.
3. Cycloheximide, freshly dissolved in ethanol to 100 mg/ml or in ddH_2O to 20 mg/ml.
4. MG 132 (*Z*-Leu-Leu-Leu-CHO), or epoxomicin, or *Z*-Leu-Leu-Leu-vinyl sulfone, or MG 262 (*Z*-Leu-Leu-Leu-B$(OH)_2$), or *clasto*-lactacystin β-lactone (*see* step 1 of **Section 2.6**).
5. RIPA lysis buffer: 150 mM NaCl, 0.5% sodium deoxycholate, 50 mM Tris-HCl, pH 8.0, 0.1% SDS, 1% NP-40, and protease inhibitors mixture.
6. 2x sample buffer: 125 mM Tris-HCl, pH 6.8, 4% SDS, 20% glycerol, 1.4 M β_2-mercaptoethanol.

2.7.2. Pulse-Chase Labeling and Immunoprecipitation

1. Cultured cells in a monolayer or in suspension (tested for the absence of *Mycoplasma* contamination).
2. Cell culture medium.
3. Methionine-free medium: Dulbecco's modified Eagle's medium (DMEM), minimum essential medium (MEM), or RPMI-1640 medium are the media that are most frequently used.
4. Dialyzed serum.
5. L-[^{35}S]Methionine for pulse-chase labeling (EasyTag; specific activity 1,000 Ci/mmol at 10.2 mCi/ml; PerkinElmer®) (*see* **Note 1**).
6. MG 132 (*Z*-Leu-Leu-Leu-CHO), or epoxomicin, or *Z*-Leu-Leu-Leu-vinyl sulfone, or MG 262 (*Z*-Leu-Leu-Leu-B$(OH)_2$) or *clasto*-lactacystin β-lactone (*see* step 1 of **Section 2.6**).
7. 200 mM L-methionine.

8. RIPA lysis buffer: 150 mM NaCl, 0.5% sodium deoxycholate, 50 mM Tris-HCl, pH 8.0, 0.1% SDS, 1% NP-40, and protease inhibitors mixture.
9. Immobilized proteins A and G.
10. 2x sample buffer: 125 mM Tris-HCl, pH 6.8, 4% SDS, 20% glycerol, 1.4 M β_2-mercaptoethanol.

3. Methods

3.1. Preparation of Crude Cell Extracts (see Note 3) for Monitoring Conjugation and Degradation of Protein Substrates in a Cell-Free System

3.1.1. Preparation of Reticulocyte Lysate

1. Inject rabbits subcutaneously with 10 mg/kg of phenylhydrazine (dissolved in PBS) on days 1, 2, 4, and 6.
2. Bleed the rabbits from the ear artery or vein, or from the heart (following anesthesia) on day 8.
3. Determine the fraction of reticulocytes in the peripheral blood (should be >90%) by incubating 0.3 ml of blood with 0.2 ml of staining solution for 15 min at 37°C. To accomplish this, proceed with steps 4–5.
4. Prepare a thin smear of the stained blood using a spreader slide and air dry the smear.
5. Count reticulocytes under an immersion objective in about 10 fields. Reticulocytes are identified by fine deep violet filaments and granules that are arranged in a network.
6. Wash the cells by centrifugation at 1,000×*g* for 10 min with ice-cold PBS (*see* **Note 4**). Repeat this step twice.
7. If it is necessary to deplete ATP (*see* **Note 5**), proceed as described under this step, otherwise, move to the next step. Re-suspend the pellet in 1 volume (i.e., pelleted cell volume) of Krebs–Ringer phosphate buffer containing 20 mM of 2-deoxy-D-glucose and 0.2 mM of 2,4-dinitrophenol. Following incubation accompanied by gentle shaking for 90 min at 37°C, wash the cells twice with ice-cold PBS (*see* step 6 of **Section 3.1.1**).
8. Break cells open in 1.6 volumes of ice-cold water containing 1 mM DTT.
9. Centrifuge at 80,000×*g* for 1 h at 4°C to remove particulate material.
10. Collect the supernatant and freeze at –70°C (*see* **Note 6**).

3.1.2. Preparation of Extract from Cultured Cells

1. Seed cells and grow them to almost confluency (in case of adhering cells) or to a density of 10^5–10^6 cells/ml (in case of cells in suspension). All procedures from here on are carried out at 4°C or on ice.

2. Wash cells three times in 20 mM HEPES, pH 7.5, saline buffer, and re-suspend to a concentration of 10^7–10^8 cells/ml in 20 mM HEPES, pH 7.5, containing 1 mM DTT.
3. If it is necessary to deplete ATP (*see* **Note 5**), proceed as described under this step, otherwise, move to the next step. Re-suspend the cells to a density of 10^7 cells/ml in Krebs–Ringer phosphate buffer containing 20 mM 2-deoxy-D-glucose, 0.2 mM 2,4-dinitrophenol, 20 mM NaF, and 10 mM NaN_3. In case of adhering cells, add the same solution to the plate. Following incubation for 60 min at 37°C, wash the cells twice in 20 mM HEPES, pH 7.5, saline buffer, and re-suspend to a density of 10^7 cells/ml in 20 mM HEPES, pH 7.5, containing 1 mM DTT.
4. Cavitate the cells in a high-pressure nitrogen chamber (*see* **Note 7**). For HeLa cells, the best conditions are 1,000 psi for 30 min. However, these conditions can be different for different cells (*see* **Note 8**).
5. Centrifuge the homogenate successively at 3,000×*g* and 10,000×*g* for 15 min each, and then at 80,000×*g* for 60 min. Collect and freeze the supernatant at –70°C (*see* **Note 6**). This extract should contain 5–10 mg/ml of protein.

3.2. Fractionation of Cell Extract to Fraction I and Fraction II

All procedures are carried out at 4°C (5).

1. Swell the DEAE cellulose (DE-52) resin in 0.3 M potassium phosphate buffer, pH 7.0, for several hours (this slurry can be prepared a day before fractionation and kept at 4°C). Use enough resin to absorb proteins in the extract: as a rule, use 0.6 resin volume per volume of reticulocyte lysate or 1 ml resin/~5 mg of protein of nucleated cell extract (*see* **Note 9**).
2. Load the resin onto a column and wash with 9 column volumes of buffer A without DTT (*see* **Note 10**).
3. Wash the resin with 1.5 column volumes of buffer A containing 1 mM DTT.
4. Load the extract. Once all the material is loaded, elute fraction I with 2.5 column volumes of buffer A. When resolving reticulocyte lysate, collect only the dark red fraction. When resolving other cell extracts, collect only the fractions with the highest absorption at 280 nm. Freeze Fraction I in aliquots at –70°C (*see* **Note 11**).
5. Wash the column extensively with buffer B (*see* **Note 12**).
6. Elute Fraction II with 2.5 column volumes of a buffer C. Collect Fraction II into a flask immersed in ice.

7. Add ammonium sulfate to saturation (~70 g/l of solution) and swirl on ice for 30 min.
8. Centrifuge at 20,000×*g* for 20 min.
9. Re-suspend the pellet in buffer D using 0.2–0.3 volumes of the original extract (*see* **Note 13**).
10. Dialyze overnight against 100–500 volumes of buffer D (change buffer in the morning for an additional several hours).
11. Remove particulate material by centrifugation at 20,000×*g* for 15 min. Freeze in aliquots at –70°C (*see* **Notes 11 and 14**).

3.3. Labeling of Proteolytic Substrates

*3.3.1. Radioiodination of Proteins (See **Note 15**)*

1. Add reagents in the following order to a 1.5 ml microcentrifuge tube (the volume of the reaction mixture can vary from 20 to 100 μl): 100 mM sodium phosphate buffer, pH 7.5, 10–500 μg of protein substrate dissolved in water or buffer (*see* **Note 16**), 50 nmol of unlabeled NaI, 0.1–2.0 mCi of radiolabeled $Na^{125}I$, and 10–50 μg of chloramine T in 50 mM sodium phosphate buffer, pH 7.5. Mix gently and incubate for 1–2 min at room temperature. Add 20–100 μg of sodium metabisulfite (double the amount of the chloramine T used) solution in 50 mM sodium phosphate buffer, pH 7.5. Mix gently.
2. To remove unreacted radioactive iodine, resolve the mixture over a desalting column equilibrated with 10 mM Tris-HCl, pH 7.6, and 150 mM NaCl (*see* **Note 17**).
3. Collect fractions (in a fraction collector or manually), each of approximately 10% of the column volume. The radioactive protein is typically eluted in fraction 4 (void volume of the column which is about 35% of the column's total volume).
4. Store in aliquots at –20°C (*see* **Notes 11 and 18**).

*3.3.2. Biosynthetic Labeling of Proteins (See **Notes 19 and 20**)*

1. Clone the cDNA of the target protein into a vector that contains T7, T3, or SP6 RNA polymerase promoters. To increase translation efficiency in wheat germ extract, it is recommended to use Wheat Germ Flexi® Vectors (these vectors have T7 and SP6 promoters but lack T3).
2. Linearize the DNA template by restriction digestion (*see* **Note 21**).
3. Transcribe the linear cDNA using systems that generate sufficient amount of mRNA (typically 6–12 μg of mRNA is required for a standard translation reaction), such as RiboMAX™ Large Scale RNA Production System SP6 or T7.

4. Purify the mRNA from the transcription reaction using phenol:chloroform extraction (*see* **Note 22**). Extract RNA with 1 volume of 3 M sodium acetate (pH 4.5)-saturated phenol:chloroform:isoamyl alcohol (25:24:1). Mix the sample for 1 min and spin it at top speed in a microcentrifuge for 2 min. Transfer the upper, aqueous phase to a fresh tube and add 1 volume of chloroform:isoamyl alcohol (24:1). Mix for 1 min and re-centrifuge. Transfer the upper, aqueous phase to a fresh tube. Any transferred chloroform can be removed by a quick spin (10 s) in a microcentrifuge followed by removal of the bottom phase with a micropipette. Add 0.1 volume of 3 M sodium acetate (pH 4.5) and 1 volume of isopropanol or 2.5 volumes of 95% ethanol. Mix and place on ice for 2–5 min. Spin at top speed in a microcentrifuge for 10 min. Pour off or aspirate the supernatant and wash the pellet carefully with 1 ml of 70% ethanol. Aspirate the ethanol, dry the pellet under vacuum, and dissolve the RNA in nuclease-free water in a volume identical to that of the transcription reaction. Store in aliquots at –70°C (*see* **Note 11**).
5. To translate the protein, add the following reagents to 1.5 ml microcentrifuge tube: 6–12 μg of G-25-purified or phenol:chloroform-extracted mRNA in nuclease-free water, 40 μCi of [^{35}S]methionine, 30 μl of Wheat Germ Extract Plus or Rabbit Reticulocyte Lysate (*see* **Note 23**), and nuclease-free water to a final volume of 50 μl.
6. Incubate at 30°C for 1 h.

3.4. Conjugation of Proteolytic Substrates in a Cell-Free System

1. Add the following reagents (on ice) to a 1.5 ml microcentrifuge tube (*see* **Note 24**): 50 mM Tris-HCl, pH 7.6, 5 mM $MgCl_2$, 2 mM DTT, 2–4 μl of reticulocyte lysate or 50–100 μg of complete cell extract (*see* **Notes 25** and **26**), 5 μg of ubiquitin, 0.5–1.0 μg of UbAl (*see* **Note 27**), and 2 mM of ATPγS (*see* **Notes 28** and **29**) and substrate, being either a labeled protein (*see* **Section 3.3**, 25,000–100,000 cpm) or an unlabeled substrate in an amount that is sufficient for detection by Western blot analysis (50–2,000 ng).
2. Incubate the mixture for 30 min at 37°C and resolve via SDS-PAGE.
3. Detect conjugates using a PhosphorImager (labeled proteins) or via enhanced chemiluminescence (ECL) following Western blot (for unlabeled substrates) using a specific primary antibody against the test protein and a secondary tagged antibody. An example result is shown in **Fig. 23.1**.

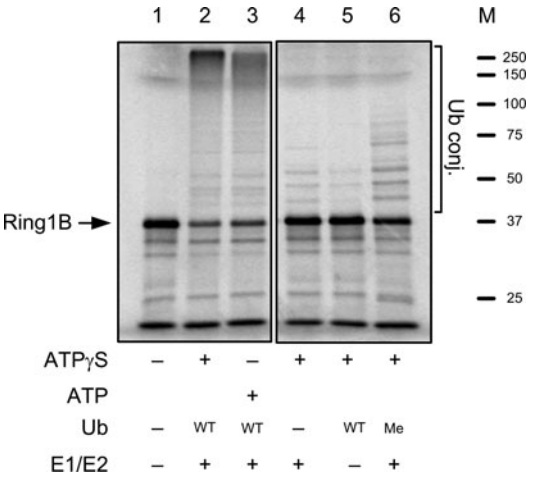


Fig. 23.1. Ubiquitination of Ring1B in a cell-free system. In vitro-translated and ^{35}S-labeled Ring1B was self-ubiquitinated in the absence of an energy source, Ub and E1/E2 (*lane 1*), in the presence of ATPγS (*lane 2*), or ATP and ATP-regenerating system (*lane 3*), in the absence of E1/E2 (*lane 4*), or in the presence of methylated ubiquitin (*lane 5*; methylated ubiquitin cannot polymerize and generate chains, as all its internal lysines and N-terminal residue are modified. It can be conjugated, however, once to lysines, and potentially to threonines, serines, and the N-terminal residue of the substrate, generating a multiply monoubiquitinated substrate). Please note that the conjugates observed in the presence of ATP are scattered over a larger range of molecular weights (they are "fuzzier") than those generated in the presence of ATPgS. That is due to the fact that with ATP, the conjugates are also degraded (*see* **Note 28**).

3.5. Degradation of Proteolytic Substrates in a Cell-Free System (See Note 30)

3.5.1. Monitoring of the Disappearance of the Substrate

1. Follow step 1 of **Section 3.4**, except for the addition of UbAl and ATPγS. Use 0.5 mM ATP instead of ATPγS as a source of energy. Its level is maintained by the addition of ATP-regenerating system (10 mM of phosphocreatine and 0.5 μg of phosphocreatine kinase). For depletion of endogenous ATP, the system should contain, instead of ATP and the ATP-regenerating system, 10 mM 2-deoxy-D-glucose and 0.5 μg hexokinase.
2. Incubate the reaction mixture for 2–3 h at 37°C and resolve it via SDS-PAGE.
3. Monitor the disappearance of the substrate either using a PhosphorImager (in case the protein substrate is radioactively labeled) or via Western blot analysis (in case of unlabeled substrate). An example result is shown in **Fig. 23.2**.

3.5.2. Monitoring of the Appearance of Acid-Soluble Radioactivity

1. Label the proteolytic substrate as described in **Section 3.3** (*see* **Note 31**).

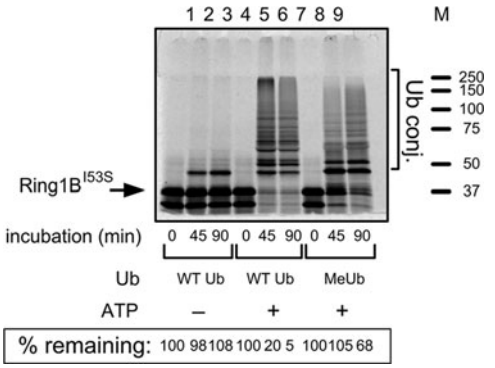


Fig. 23.2. Degradation of Ring1B^{I53S} in a cell-free system. In vitro-translated and ^{35}S-labeled Ring1B^{I53S} was subjected to degradation in a cell-free system in the presence of WT or methylated ubiquitins, and in the presence or absence of ATP as indicated. Degradation was calculated based on the radioactivity remained in the lane along time relative to time 0. Please note that methylated Ub does not support proteolysis as it does not generate polyubiquitin chains necessary for the recognition by the 26S proteasome.

2. For setting the degradation reaction mixture, follow steps 1 and 2 of **Section 3.5.1**.
3. At the end of the incubation, add a carrier protein (10–25 μl of 100 mg/ml solution of bovine serum albumin, BSA).
4. Add 0.6 ml of ice-cold 20% TCA.
5. Mix the reaction and incubate on ice for 10 min.
6. Centrifuge for 5 min at 15,000×*g*.
7. Collect 0.5 ml of the supernatant.
8. Count the radioactivity in either a γ-counter (for iodine-labeled substrates; **Section 3.3.1**) or a β-scintillation counter (for methionine-labeled substrates; **Section 3.3.2**).

3.6. Monitoring Ubiquitination of Proteolytic Substrates in Cells

1. Transform the cells with cDNA coding for the studied substrate alone or with tagged ubiquitin (HA-, Flag-, or Myc-) (*see* **Note 32**).
2. After 24–48 h, add for 2–3 h a proteasome inhibitor such as 20 μM MG132, *clasto*-lactacystin β-lactone, or epoxomicin.
3. Aspirate medium from the dishes.
4. Wash cells with ice-cold PBS.
5. To avoid deubiquitination, use "hot" lysis to dissolve the cells (*see* **Note 33**).
   Add "hot" lysis buffer I to cells (250 μl/6 cm^2 plate) and scrape them into a microcentrifuge tube.
6. Seal the tube (use a cap locker or cap-locked tubes) and boil for 5 min at 100°C.

7. Shear lysate with 25 G needle (three to five times).
8. Boil for 3 min at 100°C.
9. Mix and centrifuge the sample at 15,000×*g* for 5 min at room temperature.
10. Transfer the supernatant to a clean tube.
11. Add 1 volume of "hot" lysis buffer II and antibody against the target protein (using antibodies directed against the substrate or against its fused tag, or alternatively, an antibody against the tag fused to ubiquitin).
12. Shake gently (rotate) overnight.
13. Add 20 μl of a mixture of equal amounts of immobilized proteins A and G suspended (50% beads/volume) in "hot" lysis buffer II.
14. Shake gently (rotate) for 1 h.
15. Spin beads at 1,000 rpm for 1 min.
16. Aspirate supernatant with a 30 G needle.
17. Wash beads twice with "hot" lysis buffer III.
18. Wash beads twice with PBS.
19. Boil beads for 5 min at 100°C with 20 μl of sample buffer.
20. Resolve the proteins via SDS-PAGE followed by blotting onto a nitrocellulose membrane (*see* **Note 34**).
21. In case proteins were precipitated with an anti-substrate antibody, conjugates can be visualized using antibodies against the substrate or ubiquitin, or against the tag fused to either of them. In case anti-ubiquitin tag antibodies were used for immunoprecipitation, conjugates can be visualized only by using antibodies directed against the substrate or its fused tag.

3.7. Degradation of Proteolytic Substrates in Cells

*3.7.1. Cycloheximide Chase (See **Note 35**)*

1. Transform cells with cDNA coding for the substrate (*see* **Note 36**).
2. In case of cDNA-transfected cells, incubate the cells for 24–48 h before adding 20–100 μM of cycloheximide. Otherwise (when testing the stability of an endogenous protein(s)), add cycloheximide when the cells reach the desired density.
3. Add proteasome inhibitor to the control dish for 2–4 h (as a control for stabilization).
4. Harvest the cells at the desired times after addition of cycloheximide.
5. Monitor degradation/stabilization of the target protein via Western blot analysis. An example result is shown in **Fig. 23.3**.

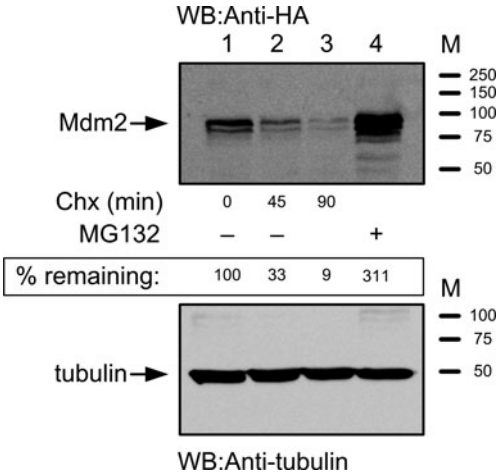


Fig. 23.3. Monitoring the kinetics of degradation of Mdm2 in cells using cycloheximide chase. HEK 293 cells were grown to desired 75% density. Degradation of endogenous Mdm2 was monitored following the addition of cycloheximide (*lanes 1–3*). Inhibition of the proteasome by MG132 (added for 2 h) resulted in stabilization and accumulation of Mdm2 (*lane 4*). Tubulin was used as a marker to ascertain equal protein loading.

3.7.2. Pulse-Chase Labeling and Immunoprecipitation

1. Follow steps 1 and 2 of **Section 3.7.1**.
2. Wash the cells twice in a methionine-free medium at 4°C.
3. Add methionine-free medium that contains dialyzed serum (serum is added in the concentration used for growing the cells).
4. Incubate for 1 h (to remove endogenous methionine), remove the medium (by aspiration for adhering cells and following centrifugation at 800×*g* for 10 min for cells in suspension), and add fresh methionine-free medium with dialyzed serum. To save on labeled methionine, for adherent cells add medium to barely cover the cells' layer (1–1.5 ml for a 60 mm dish; cells can be rocked in the incubator, and for suspension cells, re-suspend them to 2×10^6/ml).
5. Add labeled methionine (50–250 μCi/ml) and continue the incubation for 0.5–1.0 h (pulse).
6. Add proteasome inhibitor to the control dish. The inhibitor should be added for the last 30 min of the labeling period and should remain throughout the entire experiment.
7. Remove the labeling medium.
8. Add ice-cold complete medium that contains, in addition to the inhibitor (as needed), also 2 mM of unlabeled methionine and wash the cells twice in the same medium.

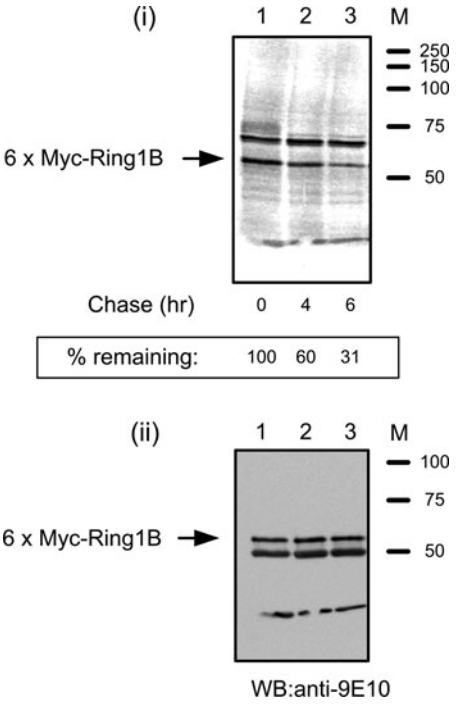


Fig. 23.4. Monitoring the kinetics of degradation of Myc-Ring1B using pulse-chase labeling and immunoprecipitation. Cos7 cells were transfected with a cDNA coding for Myc-Ring1B (*lanes 1–3*). Stability of the protein was monitored in a pulse-chase labeling and immunoprecipitation experiment. Immunoprecipitation was carried out using anti-Myc antibody, and the proteins were visualized following SDS-PAGE and blotting onto nitrocellulose membrane. Radioactive Ring1B was detected using PhosphorImaging (i) and total Ring1B in cells was detected using anti-Myc antibody (ii). Note that only the radioactive protein disappears, whereas there is no effect on the level of the total protein.

9. Add pre-warmed complete medium (containing the inhibitor (as needed) and 2 mM of unlabeled methionine) and continue the incubation for the desired times (chase).
10. Withdraw samples at various time points and monitor degradation/stabilization of the target protein(s) by immunoprecipitation followed by SDS-PAGE and phosphorImage analysis. An example result is shown in **Fig. 23.4**.

4. Notes

1. When handling radioactive materials, appropriate safety precautions must be followed.
2. Please refer to the supplier's catalog for change in pH between 4, 25, and 37°C. As a rule, for all procedures

carried out in cold environment, one should use Tris buffer prepared at 25°C at pH 7.2 and for reactions carried out at 37°C, Tris buffer prepared at 25°C at pH 7.6.

3. Conjugation can be monitored in a crude cell extract (for example, rabbit reticulocyte lysate or extract from cultured cells). Among the advantages of rabbit reticulocyte lysate are the following: (a) it contains a relatively small number of proteins which makes purification of components of the system relatively easy. Also, it is easier to monitor conjugation and degradation of substrates in this extract as the competition by endogenous substrates is less fierce; (b) since reticulocytes do not contain lysosomes, the extract is basically devoid of proteases that could have leaked during its preparation and obscure the ubiquitin proteolytic activity; (c) it can be obtained in a relatively large amount; (d) the lack of requirement for tissue culture media and sera make this lysate significantly less expensive than its nucleated cultured cells counterpart. Nucleated cell extract can be useful when (a) ubiquitin system components and/or auxiliary enzymes have to be activated/modified or (b) the level of certain proteins has to be modulated (overexpressed or inactivated (e.g., siRNA)).
4. Aspirate carefully the thin layer of white blood cells ("buffy coat") that overlays the pelleted red blood cells.
5. The extract prepared according to the protocol can be used also to monitor ATP dependence of degradation (and obviously ubiquitination), as ATP can be depleted from the cells prior to their disruption by using inhibitors of anaerobic and aerobic respiration that are later removed during dialysis. The extract can be re-supplemented with ATP (and ATP-regenerating system to counteract the activity of ATPases). ATP can be also depleted directly from the crude extract, though here, because respiration does not occur, we use only a trap composed of hexokinase and 2-deoxy-D-glucose. ATP depletion from cells will be also important if one plans to fractionate the extract in order to monitor ubiquitin dependence of degradation or to follow the fate of exogenously added (e.g., tagged) ubiquitin. Typically, the cell extract is fractionated over an anion exchange resin (e.g., diethylaminoethyl (DEAE) cellulose), where ubiquitin is eluted in fraction I, the unabsorbed, flow-through material that contains also certain E2 enzymes. Fraction II, the high salt eluate, contains E1, the remaining E2 s, all the E3 s, and the 26S proteasome, but not free ubiquitin. Depletion of ATP from cells inhibits ubiquitination, whereas deubiquitination continues. This leads to release of ubiquitin from all conjugated substrates and resolution

of the free ubiquitin with fraction I. If ATP is not depleted, ubiquitin-conjugated proteins are resolved in fraction II. The ubiquitin moiety will be released upon incubation in the studied extract and will be conjugated to substrates, among them to the test substrate. Consequently, it will be impossible to monitor ubiquitin dependence of conjugation in fraction II prepared without prior depletion of ATP.

6. The lysate is stable at –70°C for several years.
7. Alternatively, cells can be frozen (in liquid N_2) and thawed several times.
8. Make sure that most of the cells are disrupted by visualizing the suspension in a light microscope before and after cavitation. Following disruption, one should observe mostly intact nuclei and cell debris.
9. One can use also a chromatographic system such as fast protein liquid chromatography (FPLC, GE Healthcare) with an anion exchange resin-loaded column (MonoQ or Q Sepharose), though, for resolution of large quantities, the DEAE resin procedure is advantageous.
10. Can be done a day before fractionation if the column is kept at 4°C.
11. Choose the volume of aliquots according to the needs of your planned experiments to reduce repeated refrosting.
12. When resolving reticulocyte lysate, make sure all the hemoglobin is eluted. When resolving nucleated cell extract, wash until the absorbance at 280 nm returns to baseline.
13. At times, it will be impossible to dissolve all the ammonium sulfate-precipitated proteins. This is not essential. Collect, however, the slurry of the pelleted proteins into the dialysis tubing; it will be dissolved during dialysis.
14. Fraction I and fraction II are stable at –70°C for several years.
15. Iodination is utilized mostly when a purified recombinant protein is available. The main advantage of the method is the high specific radioactivity that can be obtained and the low cost. The disadvantage of the method is that during iodination, unless it is carried out using the Bolton–Hunter reagent, the protein can be damaged from chloramine T, the reagent used to oxidize the iodide anion converting it to a free radical. In addition, during storage, the labeled substrate may be subjected to radiochemical damage from the isotope.

16. Make sure that the buffer does not contain free amino or hydroxyl groups, Tris-HCl, for example, as this may result in iodination of these groups. As the buffer is in large molar excess over the protein, the protein will not be labeled.
17. In case the amount of the iodinated protein is less than 10 μg, it is recommended to add to the resolving buffer a protein carrier to protect the labeled protein from absorbance to the tube and from radiochemical damage. One can use 1 mg/ml of cytochrome C or ovalbumin. These proteins are not ubiquitinated and are not degraded by the UPS, and therefore do not compete with the labeled substrate on components of the system.
18. Labeled proteins are stable at –20°C for several weeks.
19. It is possible to use a coupled transcription–translation cell-free extract that synthesizes the mRNA from its cognate cDNA and translates it in a coupled manner. Such systems require only the addition of the cDNA and the labeled amino acid and are available commercially (TNT® Reticulocyte Lysate system or TNT® Wheat Germ Extract system (Promega) with SP6, T7, or T3 polymerase promoters, and S30 which is a bacterially derived extract (Promega)). Biosynthesis is carried out according to the manufacturer's instructions. The kit must be based on an RNA polymerase promoter (T7, T3, or SP6) identical to that present in the cDNA.
20. The generated protein is native and its specific activity can be high. However, the chemical amount of the protein is low. Since the labeled protein is contained in the crude extract in which it is synthesized, it is also not pure. From these reasons, proteins prepared this way are not amenable to isolation, analysis, characterization, and chemical modifications.
21. Supercoiled DNA templates can be used for transcription, although translation efficiency from the resulting RNAs may be lower.
22. Alternatively, it is possible to purify the mRNA from the transcription reaction using gel filtration (for example, PD MiniTrap G-25 columns (GE Healthcare)).
23. It is preferred to use wheat germ extract to translate substrates for studying the ubiquitin system. This extract lacks many, although not all, of the mammalian E3 enzymes. Therefore, in most cases, a protein synthesized in this extract can be used in experiments in which a cell-free system is reconstituted from purified enzymes, and in particular, when the role of a specific E3 is tested. A protein synthesized in reticulocyte lysate may be "contaminated"

in many cases with endogenous E2 and/or E3 enzyme(s) derived from the lysate. The enzymes, which are being carried to the reconstituted conjugation/degradation assay, may interfere with the examination of the role of an exogenously added E2 or E3 in these processes. Yet, at times, one must use the reticulocyte lysate, as the translation efficiency in the wheat germ extract can be low. In that case, if needed, the "contaminating" E2 or E3 in the lysate can be inactivated after translation by *N*-ethylmaleimide (NEM; 10 min incubation at room temperature in a final concentration of 10 mM of freshly prepared solution). Because E1, all known E2 s, and some of the E3 s (HECT domain containing) have an essential -SH group, the alkylating agent inactivates them. The NEM is then neutralized by the addition of DTT (final concentration of 10 mM). It should be noted that this procedure can also denature/inactivate the substrate. In most cases, however, the substrate can still be utilized and its behavior reproduces faithfully that of the native substrate.

24. The volume of the reaction mixture can vary from 10 to 1,000 μl. For demonstration of conjugates, the volume is typically on the low range, while for analytical purposes, such as mass spectrometry, the volume should be larger.

25. For monitoring ubiquitin dependence of conjugation, fraction II (from which ubiquitin was removed by anion exchange chromatography following also ATP depletion) derived from reticulocyte lysate (25–50 μg) or nucleated cell extract (50–100 μg) should be added rather than reticulocyte lysate or complete cell extract. Different ubiquitin species (WT; in which all, all but one, or a single lysine residue were substituted with Arg; or methylated ubiquitin in which all internal lysines and the N-terminal residue were blocked) should be added (2.5–5.0 μg). In addition, the reaction mixture should be supplemented with E1 (0.25 μg) and E2 (~0.5 μg). In most cases, UbcH5c will be sufficient, but certain reactions will require UbcH7 or UbcH8.

26. In case all the components of the ubiquitination reaction are known, one can design the reaction using purified enzymes. This reaction mixture contains instead of the lysate, crude cell extract or fraction II, purified E1 (~1 μg), E2 (0.5 μg), and E3 (should be titrated). In case of monitoring self-ubiquitination of an E3, no substrate is needed, as once ubiquitinated, the conjugated E3 can be traced by following its ladder (it should be added metabolically labeled or it can be followed via Western blotting).

27. Ubiquitin aldehyde (UbAl), a specific inhibitor of certain ubiquitin C-terminal hydrolases, isopeptidases (6), is used to increase the amount of the conjugates in a cell-free system by preventing their deubiquitination.
28. The nonhydrolyzable ATP analog, adenosine-5′-*O*-(3-thiotriphosphate) (ATPγS), can be used instead of ATP (7) to increase the amount of the adducts generated. The ubiquitin-activating enzyme, E1, can catalyze activation of ubiquitin in the presence of the analog, as it utilizes the α–β high-energy bond of the nucleotide that is cleavable also in this derivative. In contrast, assembly and activity of the 26S proteasome complex requires the β-γ bond of ATP that cannot be cleaved in the analog. Thus, the conjugates generated cannot be degraded by the proteasome, and their amount is increased. However, caution should be exercised when utilizing the ATP analog. Often, phosphorylation of the target protein is required in order for the ubiquitin ligase to recognize it and conjugate it with ubiquitin (8). In these cases, the analog cannot substitute the hydrolyzable native ATP.
29. For monitoring of ATP dependence, ATPγS is not added, and residual ATP in the extract is depleted using 10 mM 2-deoxy-D-glucose and 0.5 μg hexokinase. If phosphorylation is required for substrate recognition, 0.5 mM ATP and ATP-regenerating system (10 mM phosphocreatine and 0.5 μg phosphocreatine kinase) substitute for ATPγS.
30. For monitoring degradation in a cell-free system (**Section 3.5.1**), one can use – with a few modifications – a similar assay described for monitoring of their conjugation (**Section 3.4**). For proteins that are degraded inefficiently or slowly, it is difficult to follow the reduction in the density of a protein band in gel analysis. In this case it is recommended to monitor the appearance of radioactivity in trichloroacetic acid (TCA)-soluble fraction (**Section 3.5.2**). Control reactions are complete mixtures that have been incubated on ice or mixtures that were incubated at 37°C in the absence of ATP (ATP should be depleted) or ubiquitin. (Fraction II is not completely depleted of ubiquitin. Therefore, some proteolytic activity – the level of which is dependent on the substrate – can still be observed even in mixtures to which ubiquitin was not added.)
31. In case of a protein substrate that was labeled synthetically (**Section 3.3.2**), one needs to remove the excess of unincorporated label as well as of free ubiquitin (if this is required). To remove the labeling amino acid, extensive dialysis against a buffer that contains also 1 mM of the

corresponding unlabeled methionine is sufficient (20 mM Tris-HCl, pH 7.2, 1 mM DTT). To remove both the labeling amino acid and ubiquitin, fractionation on DEAE is most efficient (**Section 3.2**). Typically, for a translation mixture of 50 μl, a 200 μl column of DE-52 is sufficient (poured into 1 ml insulin syringe plugged with glass wool). Wash the column after loading of the translation mix with 25 column volumes (5 ml) of buffer B. Elute the labeled protein in 3 column volumes of buffer C, dialyze extensively against buffer D, and concentrate in a centrifugation filtration device (GE Healthcare) of a molecular weight cutoff (MWCO) lower by at least threefold than that of the labeled protein. Instead of dialyzing, one can remove the eluting salt by repeated concentration–dilution cycles in this device.

32. If you examine the ubiquitination of an endogenous protein, you may use only tagged ubiquitin.
33. Alternatively, the cells can be lysed in RIPA buffer (150 mM NaCl, 0.5% sodium deoxycholate, 50 mM Tris-HCl, pH 8, 0.1% SDS, 1% NP-40, and protease inhibitors cocktail) containing freshly dissolved iodoacetamide and *N*-ethylmaleimide (5 mM each; to inhibit deubiquitinating enzymes). Following lysis, immunoprecipitate the tested protein using antibodies against the substrate, its tag, or the tag fused to ubiquitin. Shake gently (rotate) overnight and add ~20 μl of a mixture of equal amounts of immobilized proteins A and G suspended (50% beads/volume) in RIPA buffer. Shake gently (rotate) for additional 1 h and centrifuge the samples at 1,000 rpm for 1 min. Aspirate the supernatant. Wash the beads (add 1 ml of RIPA buffer, centrifuge the samples at 1,000 rpm for 1 min, and aspirate the supernatant) for five to six times. Boil the samples for 5 min at 100°C with 20 μl of sample buffer, then resolve the proteins via SDS-PAGE, and blot onto nitrocellulose membrane (*see* **Note 34**). Visualize the proteins as described in step 21 of **Section 3.6**. An example result is shown in **Fig. 23.5**.
34. In order to visualize conjugates in the entire range of molecular weights, do not remove the stacking gel while transferring proteins to the nitrocellulose membrane.
35. The advantage of this method is that it does not necessitate the use of radioactive material and immunoprecipitation, and one can resolve a whole cell extract via SDS-PAGE. The disadvantage is the potential interference of the drug in the proteolytic process. Thus, if cycloheximide inhibits the synthesis of a short-lived ubiquitin ligase (E3),

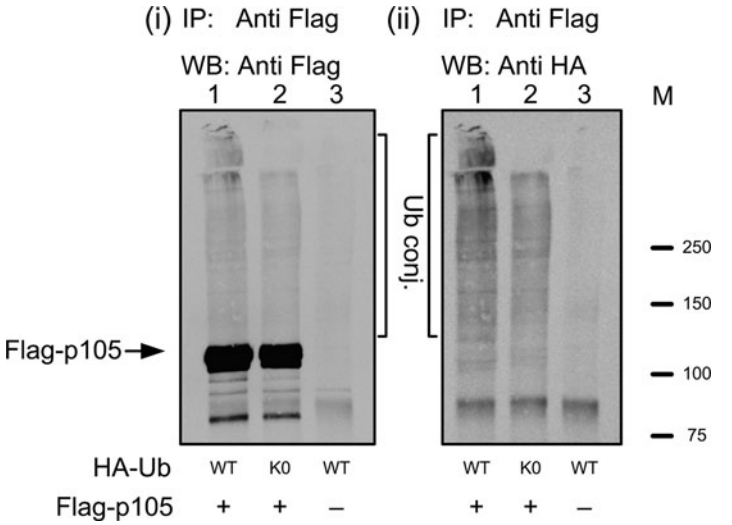


Fig. 23.5. Ubiquitination of p105 in HEK 293 cells. HEK 293 cells were transfected with cDNAs coding for Flag-p105 (*lanes 1–2*), and HA-WT Ub (*lanes 1 and 3*) or HA-UbK0 (*lane 2*). p105 and its conjugates were immunoprecipitated using anti-Flag and detected by either anti-Flag (i) or anti-HA (ii).

an inhibitor, or an activator involved in the process, the test protein can be stabilized or further destabilized, dependent on the role of the component affected.

36. If you examine an endogenous protein, this step is not necessary.

Acknowledgments

Studies in the laboratory of AC are supported by the Angelman Syndrome Foundation (ASF) and by grants from the Dr. Miriam and Sheldon Adelson Foundation for Medical research (AMRF), the Israel Science Foundation (ISF), the German-Israeli Foundation for Research and Scientific Development (G.I.F.), the European Union (EU) Network of Excellence Rubicon, an Israel Cancer Research Fund (ICRF) USA Professorship, and a grant from the Foundation for Promotion of Research in the Technion.

References

1. Mayer, R. J., Ciechanover, A., and Rechsteiner, M. (2005) *Protein Degradation Handbook*, Vol. 1, Weinheim: Wiley-VCH.
2. Mayer, R. J., Ciechanover, A., and Rechsteiner, M. (2006) *Protein Degradation Handbook*, Vol. 2, Weinheim: Wiley-VCH.

3. Mayer, R. J., Ciechanover, A., and Rechsteiner, M. (2006) *Protein Degradation Handbook*, Vol. 3, Weinheim: Wiley-VCH.
4. Mayer, R. J., Ciechanover, A., and Rechsteiner, M. (2008) *Protein Degradation Handbook*, Vol. 4, Weinheim: Wiley-VCH.
5. Hershko, A., Heller, H., Elias, S., and Ciechanover, A. (1983) Components of ubiquitin-protein ligase system. Resolution, affinity purification, and role in protein breakdown, *J Biol Chem 258*, 8206–8214.
6. Hershko, A., and Rose, I. A. (1987) Ubiquitin-aldehyde: a general inhibitor of ubiquitin-recycling processes, *Proc Natl Acad Sci USA 84*, 1829–1833.
7. Johnston, N. L., and Cohen, R. E. (1991) Uncoupling ubiquitin-protein conjugation from ubiquitin-dependent proteolysis by use of beta, gamma-nonhydrolyzable ATP analogues, *Biochemistry 30*, 7514–7522.
8. Petroski, M. D., and Deshaies, R. J. (2005) Function and regulation of cullin-RING ubiquitin ligases, *Nat Rev 6*, 9–20.

Chapter 24

Bioinformatics Challenges in Mass Spectrometry-Driven Proteomics

Lennart Martens

Abstract

Mass spectrometry-based proteomics has become an essential part of the analytical toolbox of the life sciences. With the ability to identify and quantify hundreds to thousands of proteins in high throughput, the field has contributed its fair share to the data avalanche coming from the so-called *omics* fields. As a result, the challenges involved in processing and managing this flood of data have grown as well. This chapter will point out and discuss these challenges, starting from the processing of raw mass spectrometry data into peaks, over the identification of peptides and proteins, to the quantification of the identified molecules. Finally, the informatics aspects of the nascent field of targeted proteomics are outlined as well.

Key words: Bioinformatics, databases, mass spectrometry, identification, quantification.

Abbreviations: FDR, False discovery rate; *m/z*, Mass-to-charge ratio; MS, Mass spectrometry; MS/MS, Tandem-MS; SRM, Selected reaction monitoring;

1. Introduction

The field of mass spectrometry-based proteomics has matured quite considerably over the past 10 years. Fueled by substantial advances in instrumentation (1), sequence databases (2, 3), specialized software (4), and innovative methodologies (5), the field has quickly transformed into a high-throughput analytical tool for the identification and quantification of hundreds to thousands of proteins per experiment. This ability to generate continuously large volumes of information has correspondingly increased the pressure on the downstream data processing algorithms and pipelines. Furthermore, the innovative techniques used in sample preparation have brought several lingering issues in proteomics

K. Gevaert, J. Vandekerckhove (eds.), *Gel-Free Proteomics*, Methods in Molecular Biology 753,
DOI 10.1007/978-1-61779-148-2_24,

data processing into sharper contrast. Indeed, the data processing now quite often provides a considerable bottleneck in proteomics experiments, with the development of robust algorithms and production-grade software to address the management and interpretation of the acquired data lagging behind developments in the other areas of proteomics research. This chapter therefore aims to outline the various challenges and issues in data processing that can be encountered when performing a typical proteomics experiment. The following sections follow the overall workflow in proteomics analyses, starting immediately after the acquisition of raw mass spectral data by the mass spectrometer. Each subsequent stage of data processing will then be individually discussed in detail, moving from the conversion of the raw data signal to peaks, over the identification of peptides from these peak lists and the inference of proteins from these peptides, to the quantification of the proteins. Finally, the informatics aspects relevant to the fast developing field of targeted proteomics using selected reaction monitoring will also be discussed.

2. From Raw Data to Peaks

There is a consistent confusion in the field as to the meaning of the term *raw data*, so it is important to explicitly state here what is meant when reference is made to raw data. Raw data is here considered to be the proprietary, binary output provided by the instrument at the end of an analysis (6). Very few downstream approaches actually use the raw data directly, however. The sheer size of the data precludes fast access, and the level of detail stored in these files can overwhelm downstream algorithms. The typical first step in data processing is therefore the use of so-called signal processing algorithms to analyze and reduce the raw data to obtain a much more manageable set of peaks instead (7). Despite the significant size reduction typically obtained by this type of processing (6), possible errors are relatively small (7). It is important, however, to distinguish two different aspects of this processing. The first goal is to reduce a set of measurements by the detector into defined ion mass-over-charge ratios (m/z). Each m/z is then represented by a so-called centroid, a single peak that reflects the center of the m/z distributions as measured by the detector. The width of the original distribution can be used to determine the resolution of a mass spectrometer, with a smaller width corresponding to a higher resolution. An efficient centroiding process is important in subsequent identification, as more accurate m/z values are more discriminating than m/z values with larger errors (8). The second goal of raw data reduction is the

determination of signal intensity, output as peak height. Various methods exist, but most often the area under the curve outlined by the detector measurements is used (9). While the accuracy of the signal intensity is less important in identification, the quantification of peptides or proteins is crucially dependent on this metric (9).

While the acquisition software that comes with a mass spectrometer can perform this raw data processing, several improvements have been proposed by various groups (10–13) that result in more accurate determination of m/z or signal intensity. Furthermore, the removal of information from spectra is reported to also aid downstream identification, e.g., by reducing spectra to the 10 most intense peaks (14).

The importance of correct signal processing is of course directly related to its role so early on in the processing pipeline, and while the determination of peak m/z can generally be considered to be quite tolerant to small errors, the quantification of proteins is far less forgiving. It is therefore useful to spend some time evaluating the optimal settings for raw data processing for a particular instrument and workflow and not treat signal processing as a black box (9).

3. From Peaks to Peptides

3.1. Three Types of Identification Algorithms

Once peaks have been inferred from the raw data, the resulting peak lists are typically used to identify peptides using specialized algorithms. The software used for identification can be broadly divided into three categories based on their overall approach to the problem: database search algorithms, de novo algorithms, and tag-based algorithms. The next sections discuss these three types of search algorithm in more detail.

3.2. Database Search Algorithms

This first class of identification software is by far the most commonly used and is therefore often further sub-classified in the literature, for instance, based on their scoring algorithm (4). The search engines all operate on the same basic premise: using a protein sequence database, they perform an in silico proteolytic digest to obtain peptides, which are then fragmented in silico to obtain a theoretical MS/MS spectrum for each of the peptides. The core algorithm then compares and scores the similarity between these theoretical spectra and the experimental spectra (15). Most often, this process yields many (ranked) candidate peptides per spectrum, and the critical last task is then to distinguish correct matches from incorrect matches (4). This process of calling correct and incorrect matches is prone to the classical two types of

statistical error: false positives (type I error or α error) and false negatives (type II error or β error). The former results from calling an incorrect match as correct, while the latter occurs when a correct match is considered incorrect; these errors thus influence the specificity and sensitivity of the algorithm, respectively (16). It is important to realize that it is quite difficult to optimize both these parameters simultaneously: gains in one aspect usually require losses in the other. Depending on the downstream use of the obtained identifications, one can either opt for increased sensitivity at the price of more false positives, or for more specific identifications, with a corresponding loss in sensitivity. Most of the current data processing efforts relating to database searching algorithms are aimed at one of two goals: (i) reducing the amount of false positives while maintaining a good sensitivity and (ii) gaining insight into the amount and nature of false-positive identifications obtained. These two topics will be discussed in more detail in the next paragraphs.

The main search engines in use today, Mascot (17), SEQUEST (18), X!Tandem (19), OMMSA (20), and MyriMatch (21), aim to provide reliable scoring for identifications, allowing the straightforward distinction between correct and incorrect identifications. A comparison of different search algorithms has shown, however, that there is no perfect solution among them as they complement each other quite substantially (22). Additionally, regardless of the search engine used, false positives remain in the results. Several post-processing strategies have therefore been developed to analyze further the output of the search engine, often relying on orthogonal information not used by the search engine. These tools, including the alliterating trio PeptideProphet (23), Percolator (24), and Peptizer (25), essentially attempt to emphasize the score differences between correct and incorrect matches by examining various properties of the peptide-to-spectrum assignments. The three algorithms each present a unique approach to this end, however. PeptideProphet uses mixture modeling to separate correct from incorrect identifications, fitting the models to subpopulations of identifications that are split into correct and incorrect identifications based on a limited set of rules (one of the dominant properties, for instance, is tryptic correctness of the peptide termini). Percolator does not split the actual search results, but rather relies on a combined search against both a normal and a so-called decoy database (*see* next paragraph) for its population of correct and incorrect hits, respectively. Percolator then determines distinguishing properties from these populations in an iterative way through machine learning. Peptizer, finally, is an expert system that relies on user-defined and user-configured expert rules to pick out suspect identifications which can then be manually evaluated or automatically rejected.

As outlined above, many database search algorithms exist today. Since all of them are prone to type I and type II errors, however, a substantial amount of effort has been invested in estimating the actual magnitude of these errors, especially type I errors (false-positive assignments) have been investigated. The main strategy underlying most commonly used approaches is the use of nonsense databases for searching to estimate the number of false-positive matches that search engines report for a given set of MS/MS spectra. These nonsense databases are typically called decoy databases and they can take various forms, with the most popular being reversed and shuffled protein sequence databases, respectively. In the former version, the protein sequences in the database are simply reversed, whereas the latter approach randomly reassigns all residues to new positions within the protein sequence. Such databases can be created easily using freely available software (26, 27) and will work without further modifications with most search engines. A good overview of the basic strategy is given by Elias and Gygi (28), employing a concatenated decoy database in combination with a SEQUEST search to estimate the number of false positives as twice the number of decoy hits at a given set of cutoff scores. The false discovery rate (FDR) can then be easily calculated by dividing the total number of decoy hits by the total number of decoy and normal peptide identifications. While this approach has proven very popular because of its ease of execution, it is important to note several caveats. First of all, some probabilistic search engines (most notably Mascot) determine threshold scores based (in part) on the size of the search database. Concatenated normal–decoy databases will increase the threshold, thus lowering the number of peptide hits compared to a normal database only search. The measurement of the FDR thus directly influences the search results, which is obviously an undesirable situation. The situation can be rescued by performing two parallel searches: one against the normal database and one against the decoy database. Elias and Gygi do show, however, that this can result in an overestimation of the FDR, because certain spectra yielding identifications against the decoy database also provide higher scoring peptide hits in the normal database (28). Counting the decoy hits in such cases as an incorrect identification thus overestimates the prevalence of decoy hits. In order for the parallel search strategy to yield a correct result, more data processing is required; by joining the results from both searches together, and subsequently retaining only a single peptide hit per spectrum (the highest scoring one), a useful list for FDR estimation is obtained once again. In the particular case of Mascot, however, the problem has been mitigated by a built-in FDR estimation using decoy sequences, obviating the need to generate a decoy database or perform any calculations oneself. An inverse problematic situation can be encountered with the Paragon search

algorithm (29). Here the concept behind identifying peptides is based on the assumption that a good match can be found in the database. As a result, very many decoy peptides will be reported if a data set is searched against only a decoy database. Parallel searches will thus result in a vast overestimation of the FDR with this algorithm. It is therefore important to investigate the behavior of an algorithm prior to applying a decoy strategy to estimate FDR. Despite the popularity of these simple decoy-based FDR calculations, improvements have been suggested to both the calculation (30) and the metric calculated (31). The latter study, in particular, highlights the use of the complementary *q*-values and posterior error probabilities in favor of the FDR.

3.3. De Novo Algorithms

The search algorithms discussed in the previous section all suffer from the same fundamental issue, however: their reliance on a priori constructed peptide candidate lists renders them incapable of identifying unexpected peptide sequences or sequences carrying unexpected modifications. When MS-based proteomics is applied to organisms for which no complete genome sequence is available or when the goal of the study is to identify novel or strangely modified sequences, search algorithms are not well suited to the task. In these cases, the only available strategy consists of directly interpreting the sequence information in the acquired MS/MS spectrum. Such sequence extraction relies on the occurrence of ladders of fragments in the spectrum, which occurs when a peak in a spectrum corresponding to the *n*th fragment ion is followed by subsequent peaks corresponding to fragment ions that contain the $n + 1$, $n + 2$, and in general $n + x$ residues. The distance between these consecutive peaks will then correspond to the mass of a single (modified) amino acid residue, allowing the ladder to be sequenced. The (partial) sequence thus obtained from a spectrum can then be used to infer further information about the parent protein, most often using BLAST to find homologous sequences in genome databases of related, sequenced species. This process of reading a sequence directly from a spectrum is called de novo identification.

In practice, however, the extraction of sequence from an MS/MS spectrum is fraught with difficulty, as ladders occur infrequently and usually only cover very few consecutive residues. Furthermore, the peaks comprising a ladder need not be the most prominent peaks in the spectrum, and charge state shifts can also cause breaks in the ladder (e.g., when the y3 fragment ion is singly charged, but the y4 ion carries an additional basic residue and has therefore acquired two charges). A single spectrum can also contain more than one seemingly valid ladder, starting from the same peak of origin or even from a different peak of origin. These issues confound the extraction of long, unambiguous sequence ladders, and the result is often shorter sequence segments with

substantial ambiguity. While de novo sequencing can be automated, the interpretation of the results often requires substantial manual processing, especially in the case of high-throughput spectra with poor mass accuracy. An overall guide to the practice of de novo sequencing can be found in two useful reviews of the performance of commonly used de novo algorithms on high-throughput proteomics data (32, 33). As a result, some researchers have incorporated the ambiguity inherent in de novo sequencing as mass gaps in the obtained sequence, thus maintaining low-level (but correct) mass information instead of inferring high-level (but highly ambiguous or downright incorrect) sequence information (34).

While de novo algorithms can be very useful tools, it is important to emphasize that the quality of the obtained data is paramount to success with this approach, and that mass accuracy at the fragment ion level is even more important here than it is for sequence database searching.

3.4. Tag-Based Algorithms

The previous sections outlined the starkly contrasting database search and de novo algorithms for peptide identifications, but a third, hybrid form exists that relies on short sequence tags (typically 3–6 amino acids in length) extracted de novo from an MS/MS spectrum and then searches a sequence database using this tag. The tag actually consists of both the short sequence stretch and flanking masses on either side. The software thus complements the sequence with compositional measures of the flanking region, which, together with the peptide precursor mass, allows the number of possible matches to be reduced substantially. Interestingly, the first identification algorithm to be published was a tag-based algorithm (35), and despite its initial popularity, it was soon replaced by database search algorithms as the identification tool of preference. Tag-based algorithms do carry substantial advantages in certain situations, however, as they marry the benefits from both database searching and de novo identification without carrying over too many of the corresponding caveats. The initial de novo step is less prone to ambiguity and error due to the nature of the short tags. The software does not need to mine the spectrum overly to extract a short and clear tag from a spectrum, and many spectra will include at least one short ladder. Furthermore, the flanking masses are obtained, and these provide coarse but useful information about the remainder of the sequence, much like the gapped peptides discussed above. Once the tags are extracted, they can be used to search conventional protein databases. In this stage, however, allowances can be made to both the flanking masses and the entire precursor mass, allowing the search to provide results with divergent sequences or carrying modifications. This latter trick is made possible by the superior discriminating power of the sequence bit in the tag

as compared to an MS/MS spectrum proper. Recent implementations of tag-based algorithms can be found in both GutenTag (36) and TagRecon (37) by Tabb and colleagues.

4. From Peptides to Proteins

Once peptides have been identified from the acquired MS/MS spectra, there remains the task of inferring proteins from these peptides. This step is, however, far from trivial (7), and the various issues involved are described in much detail by Nesvizhskii and Aebersold (38). The main issue with protein inference is that it is essentially an *ill-posed* problem: for a given set of peptides, many different protein sets can be inferred, and it is impossible to differentiate between these various sets. There are three general ways of dealing with this problem (7): (i) the construction of minimal explanatory lists, containing as few protein identifiers as possible to explain the observed peptides; (ii) the creation of a maximal exploratory list that simply contains all proteins for which at least one peptide has been observed; and (iii) the minimal set with maximal annotation, reducing the proteins to as small a set as possible while maximizing the amount of information known about the proteins in the set (a metric often equated with their database of origin). Although examples of each of these approaches are found in the literature, by far the most popular approach is the construction of minimal explanatory lists, which has been reported to increase the accuracy of protein identification (39). Several tools, including ProteinProphet (40) and IDPicker (41), are able to extract such lists automatically from the identified peptides. However, indistinguishable protein sets still occur quite frequently, and a choice between such equivalent sets then has to be made on more or less arbitrary grounds, such as the overall coverage of proteins in a set. The complexity involved in protein inference renders the comparison of data sets challenging (42) and can have important effects on protein quantification (9). It is therefore important to judiciously use these algorithms and to assemble quantitative information on the peptide level into protein-level information with considerable care.

5. Protein Quantification

The determination of the (relative) amount of protein in a sample is the final goal of many proteomics experiments. The most commonly applied approach performs relative quantification of

proteins, meaning that a sample is compared against another sample, and relative differences in protein abundance between these two samples are established. The alternative approach of absolute quantification determines the actual amount of a protein in a sample, but this is commonly performed by the relative quantification of a protein as compared to a spike-in standard protein at known concentration.

Protein quantification can be performed using a variety of techniques, which are listed and explained in several relevant review articles (43–45). For the purposes of this section, it is sufficient to divide the approaches into those that rely on the intensity of peaks and those that simply count the number of peptides assigned or the protein sequence covered. The latter approaches are based on the assumption that the more of a protein there is in a sample, the more likely it is that peptides of that protein are seen more than once or the more likely it is that more unique peptides are found for this protein. The diversity of available techniques directly translates into an equally broad spectrum of specialized tools that are available to process the corresponding data that are reviewed and presented in two useful review articles (9, 46).

Since quantification relies on the successful completion of all the previously discussed steps, it is here that errors along the way will be most strongly felt. Incorrect raw data processing will quickly introduce substantial errors in intensity-based methods, for instance, easily contributing 10% or more errors as compared to the original measurements (9). Missed peptide identifications will obviously impact any quantification method, since it reduces the amount of information available for a protein at best, and in many cases the loss of a peptide might also mean the loss of the corresponding protein, as only one peptide yielded an MS/MS spectrum for that protein. For spectral counting or sequence coverage techniques, the loss of a peptide identification translates into a measurement error for that protein that is inversely proportional to the total number of identified (unique) peptides for that protein. The inverse, a false-positive identification of a peptide, will of course create equally disrupting effects in protein quantification.

Errors or ambiguities in protein assembly also create issues, since protein quantities are derived from the quantities of their mapped peptides. The quantification information coming from peptides that could be mapped to more than one precursor protein, for instance, is very difficult to deconvolute into the various contributions made by the different precursor proteins (47), although the simple approach to distribute shared peptide spectral counts according to the counts of unique peptides seems to work quite well for spectral counting (48). Interestingly, the Rover quantification validation tool (49), for instance, offers specific

views for unique and shared peptides in order to aid the correct manual validation of protein-level quantification. Furthermore, both Rover and jTraqX (50), among others, allow you to eliminate shared peptides from the protein-level quantification altogether.

Protein quantification must always be considered in light of a thorough understanding of the performance and possible errors of all the previous data processing steps.

6. Targeted Proteomics

A very recent development in proteomics has been the introduction of targeted approaches using selected reaction monitoring (SRM) as a means to pre-program a specialized mass spectrometer (typically a triple-quadrupole instrument) to analyze only a defined subset of so-called transitions. A transition is defined as a combination of a precursor ion m/z and a fragment ion m/z obtained after fragmentation. The three quadrupoles (Q) perform the following functions in this approach: Q1 is the precursor m/z filter, Q2 serves as the fragmentation cell, and Q3 is the fragment ion m/z filter. The combination of precursor and fragment ion m/z filtering yields relatively specific signals (especially when multiple such transitions are monitored for a single precursor), and as transitions are predefined and actively searched for, sensitivity is increased as well. A thorough introduction to the SRM methodology for quantitative proteomics can be found in a recent article by Lange et al. (51).

This targeted approach has important repercussions for the positioning of bioinformatics in proteomics. Rather than processing the data post-acquisition, which is the default modus operandi throughout the previous sections, the delineation of targets must be done prior to the start of acquisition. As a result, a substantial part of the bioinformatics effort takes place before the sample is actually presented to the mass spectrometer. Several software packages or tools have recently been developed for the automatic prediction of suitable targets for SRM-based proteomics (52–54). Most of these approaches rely on previously acquired empirical data to build a predictor that can then also be used to assess the suitability of novel targets, and some of the software can also be used to interpret the results obtained from SRM experiments.

Because of the relative youth of this field, many of the details are not yet worked out in detail, although the field is actively addressing the relevant issues. Doubtlessly, considerable advances will be made in this field in the near future.

Acknowledgment

LM would like to thank Joël Vandekerckhove for his support.

References

1. Domon B. and Aebersold R. (2006) Mass spectrometry and protein analysis. *Science* 312, 212–217.
2. Hubbard T., Aken B., Ayling S., Ballester B., Beal K., Bragin E. et al. (2009) Ensembl 2009. *Nucleic Acids Res* 37, D690–D697.
3. The UniProt Consortium (2009) The universal protein resource (UniProt) 2009. *Nucleic Acids Res* 37, D169–D174.
4. Sadygov R.G., Cociorva D. and Yates J.R. (2004) Large-scale database searching using tandem mass spectra: Looking up the answer in the back of the book *Nat Methods* 1, 195–202.
5. Gevaert K., Van Damme P., Ghesquière B., Impens F., Martens L., Helsens K. et al. (2007) A la carte proteomics with an emphasis on gel-free techniques. *Proteomics* 7, 2698–2718.
6. Martens L., Nesvizhskii A.I., Hermjakob H., Adamski M., Omenn G.S., Vandekerckhove J. et al. (2005) Do we want our data raw? Including binary mass spectrometry data in public proteomics data repositories. *Proteomics* 5, 3501–3505.
7. Martens L. and Hermjakob H. (2007) Proteomics data validation: Why all must provide data. *Mol Biosyst* 3, 518–522.
8. Boyne M.T., Garcia B.A., Li M., Zamdborg L., Wenger C.D., Babai S. et al. (2009) Tandem mass spectrometry with ultrahigh mass accuracy clarifies peptide identification by database retrieval. *J Proteome Res* 8, 374–379.
9. Vaudel M., Sickmann A. and Martens L. (2009) Peptide and protein quantification: A map of the minefield. *Proteomics* 10, 650–670.
10. Cox J. and Mann M. (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* 26, 1367–1372.
11. Zhang J., Xu D., Gao W., Lin G. and He S. (2009) Isotope pattern vector based tandem mass spectral data calibration for improved peptide and protein identification. *Rapid Commun Mass Spectrom* 23, 3448–3456.
12. Zhang X., Asara J.M., Adamec J., Ouzzani M. and Elmagarmid A.K. (2005) Data pre-processing in liquid chromatography-mass spectrometry-based proteomics. *Bioinformatics* 21, 4054–4059.
13. Kwon D., Vannucci M., Song J.J., Jeong J. and Pfeiffer R.M. (2008) A novel wavelet-based thresholding method for the pre-processing of mass spectrometry data that accounts for heterogeneous noise. *Proteomics* 8, 3019–3029.
14. Renard B.Y., Kirchner M., Monigatti F., Ivanov A.R., Rappsilber J., Winter D. et al. (2009) When less can yield more – Computational preprocessing of MS/MS spectra for peptide identification. *Proteomics* 9, 4978–4984.
15. Matthiesen R. (2007) Methods, algorithms and tools in computational proteomics: A practical point of view. *Proteomics* 7, 2815–2832.
16. Nesvizhskii A.I., Vitek O. and Aebersold R. (2007) Analysis and validation of proteomic data generated by tandem mass spectrometry. *Nat Methods* 4, 787–797.
17. Perkins D.N., Pappin D.J., Creasy D.M. and Cottrell J.S. (1999) Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* 20, 3551–3567.
18. Eng J.K., McCormack A.L. and Yates J.R. (1994) An approach to correlate tandem mass-spectral data of peptides with amino-acid-sequences in a protein database. *J Am Soc Mass Spectrom* 5, 976–989.
19. Craig R. and Beavis R. (2004) TANDEM: matching proteins with tandem mass spectra. *Bioinformatics (Oxford, England)* 20, 1466–1467.
20. Geer L.Y., Markey S.P., Kowalak J.A., Wagner L., Xu M., Maynard D.M. et al. (2004) Open mass spectrometry search algorithm. *J Proteome Res* 3, 958–964.
21. Tabb D.L., Fernando C.G. and Chambers M.C. (2007) MyriMatch: Highly accurate tandem mass spectral peptide identification by multivariate hypergeometric analysis. *J Proteome Res* 6, 654–661.

22. Kapp E.A., Schütz F., Connolly L.M., Chakel J.A., Meza J.E., Miller C.A. et al. (2005) An evaluation, comparison, and accurate benchmarking of several publicly available MS/MS search algorithms: Sensitivity and specificity analysis. *Proteomics* 5, 3475–3490.
23. Keller A., Nesvizhskii A.I., Kolker E. and Aebersold R. (2002) Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search *Anal Chem* 74, 5383–5392.
24. Käll L., Storey J.D. and Noble W.S. (2008) Non-parametric estimation of posterior error probabilities associated with peptides identified by tandem mass spectrometry *Bioinformatics* 24, 42–48.
25. Helsens K., Timmerman E., Vandekerckhove J., Gevaert K. and Martens L. (2008) Peptizer, a tool for assessing false positive Peptide identifications and manually validating selected results *Mol Cell Proteomics* 7, 2364–2372.
26. Martens L., Vandekerckhove J. and Gevaert K. (2005) DBToolkit: Processing protein databases for peptide-centric proteomics *Bioinformatics* 21, 3584–3585.
27. Reidegeld K.A., Eisenacher M., Kohl M., Chamrad D., Körting G., Blüggel M. et al. (2008) An easy-to-use Decoy Database Builder software tool, implementing different decoy strategies for false discovery rate calculation in automated MS/MS protein identifications *Proteomics* 8, 1129–1137.
28. Elias J.E. and Gygi S.P. (2007) Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry *Nat Methods* 4, 207–214.
29. Shilov I.V., Seymour S.L., Patel A.A., Loboda A., Tang W.H., Keating S.P. et al. (2007) The Paragon Algorithm, a next generation search engine that uses sequence temperature values and feature probabilities to identify peptides from tandem mass spectra *Mol Cell Proteomics* 6, 1638–1655.
30. Käll L., Storey J.D., MacCoss M.J. and Noble W.S. (2008) Assigning significance to peptides identified by tandem mass spectrometry using decoy databases *J Proteome Res* 7, 29–34.
31. Käll L., Storey J.D., MacCoss M.J. and Noble W.S. (2008) Posterior error probabilities and false discovery rates: Two sides of the same coin *J Proteome Res* 7, 40–44.
32. Pevtsov S., Fedulova I., Mirzaei H., Buck C. and Zhang X. (2006) Performance evaluation of existing de novo sequencing algorithms *J Proteome Res* 5, 3018–3028.
33. Pitzer E., Masselot A. and Colinge J. (2007) Assessing peptide de novo sequencing algorithms performance on large and diverse data sets *Proteomics* 7, 3051–3054.
34. Kim S., Bandeira N. and Pevzner P.A. (2009) Spectral profiles, a novel representation of tandem mass spectra and their applications for de novo peptide sequencing and identification *Mol Cell Proteomics* 8, 1391–1400.
35. Mann M. and Wilm M. (1994) Error-tolerant identification of peptides in sequence databases by peptide sequence tags *Anal Chem* 66, 4390–4399.
36. Tabb D.L., Saraf A. and Yates J.R. (2003) GutenTag: High-throughput sequence tagging via an empirically derived fragmentation model *Anal Chem* 75, 6415–6421.
37. Dasari S., Chambers M.C., Slebos R.J., Zimmerman L.J., Ham A.L. and Tabb D.L. (2010) TagRecon: High-throughput mutation identification through sequence tagging *J Proteome Res* 9, 1716–1726.
38. Nesvizhskii A.I. and Aebersold R. (2005) Interpretation of shotgun proteomic data: The protein inference problem *Mol Cell Proteomics* 4, 1419–1440.
39. Zhang B., Chambers M.C. and Tabb D.L. (2007) Proteomic parsimony through bipartite graph analysis improves accuracy and transparency *J Proteome Res* 6, 3549–3557.
40. Nesvizhskii A.I., Keller A., Kolker E. and Aebersold R. (2003) A statistical model for identifying proteins by tandem mass spectrometry *Anal Chem* 75, 4646–4658.
41. Ma Z., Dasari S., Chambers M.C., Litton M.D., Sobecki S.M., Zimmerman L.J. et al. (2009) IDPicker 2.0: Improved protein assembly with high discrimination peptide identification filtering *J Proteome Res* 8, 3872–3881.
42. Martens L., Muller M., Stephan C., Hamacher M., Reidegeld K.A., Meyer H.E. et al. (2006) A comparison of the HUPO Brain Proteome Project pilot with other proteomics studies *Proteomics* 6, 5076–5086.
43. Fenselau C. (2007) A review of quantitative methods for proteomic studies. *J Chromatogr B Analyt Technol Biomed Life Sci* 855, 14–20.
44. Bantscheff M., Schirle M., Sweetman G., Rick J. and Kuster B. (2007) Quantitative mass spectrometry in proteomics: A critical review *Anal Bioanal Chem* 389, 1017–1031.
45. Nakamura T. and Oda Y. (2007) Mass spectrometry-based quantitative proteomics *Biotechnol Genet Eng Rev* 24, 147–163.
46. Mueller L.N., Brusniak M., Mani D.R. and Aebersold R. (2008) An assessment of software solutions for the analysis of mass spectrometry based quantitative proteomics data *J Proteome Res* 7, 51–61.

47. Jin S., Daly D.S., Springer D.L. and Miller J.H. (2008) The effects of shared peptides on protein quantitation in label-free proteomics by LC/MS/MS *J Proteome Res* 7, 164–169.
48. Zhang Y., Wen Z., Washburn M.P. and Florens L. (2010) Refinements to label free proteome quantitation: How to deal with peptides shared by multiple proteins *Anal Chem* 82, 2272–2281.
49. Colaert N., Helsens K., Impens F., Vandekerckhove J. and Gevaert K. (2010) Rover: A tool to visualize and validate quantitative proteomics data from different sources *Proteomics* 10, 1226–1229.
50. Muth T., Keller D., Puetz S.M., Martens L., Sickmann A. and Boehm A.M. (2010) jTraqX: A free, platform independent tool for isobaric tag quantitation at the protein level *Proteomics* 10, 1223–1225.
51. Lange V., Picotti P., Domon B. and Aebersold R. (2008) Selected reaction monitoring for quantitative proteomics: A tutorial *Mol Syst Biol* 4, 222.
52. Walsh G.M., Lin S., Evans D.M., Khosrovi-Eghbal A., Beavis R.C. and Kast J. (2009) Implementation of a data repository-driven approach for targeted proteomics experiments by multiple reaction monitoring *J Proteomics* 72, 838–852.
53. Sherwood C.A., Eastham A., Lee L.W., Peterson A., Eng J.K., Shteynberg D. et al. (2009) MaRiMba: A software application for spectral library-based MRM transition list assembly *J Proteome Res* 8, 4396–4405.
54. Mead J.A., Bianco L., Ottone V., Barton C., Kay R.G., Lilley K.S. et al. (2009) MRMaid, the web-based tool for designing multiple reaction monitoring (MRM) transitions *Mol Cell Proteomics* 8, 696–705.

Chapter 25

A Case Study on the Comparison of Different Software Tools for Automated Quantification of Peptides

Niklaas Colaert, Joël Vandekerckhove, Lennart Martens, and Kris Gevaert

Abstract

MS-driven proteomics has evolved over the past two decades to a high tech and high impact research field. Two distinct factors clearly influenced its expansion: the rapid growth of an arsenal of instrument and proteomic techniques that led to an explosion of high quality data and the development of software tools to analyze and interpret these data which boosted the number of scientific discoveries. In analogy with the benchmarking of new instruments and proteomic techniques, such software tools must be thoroughly tested and analyzed. Recently, new tools were developed for automatic peptide quantification in quantitative proteomic experiments. Here we present a case study where the most recent and frequently used tools are analyzed and compared.

Key words: Automated quantification, Mascot Distiller, MaxQuant, Census, MsQuant, Rover.

1. Introduction

Proteomics has recently moved from qualitative to quantitative science. As it relies nearly entirely on mass spectrometry, which on itself is not well suited for measuring absolute amounts of ions, quantitative proteomics is still mainly based on comparative analysis in which peptide ions that are differently labeled with stable isotopes are compared. Compared peptides are chemically identical and as a result behave similar during the entire isolation procedure, whereas the isotopic variants only segregate during the final mass spectrometric step. The mass differences between two isotopic partners are generated by either metabolic labeling (in vivo) or by chemical or enzymatic labeling procedures (in vitro) (1, 2).

K. Gevaert, J. Vandekerckhove (eds.), *Gel-Free Proteomics*, Methods in Molecular Biology 753,
DOI 10.1007/978-1-61779-148-2_25,

Among the former procedures, SILAC labeling has gained a lot of attention because it is relatively easy to incorporate amino acids carrying stable isotopes that show no or limited metabolic conversion into other amino acids in mammalian cells (3, 4). It is therefore not surprising that various tools have been developed that automatically calculate peptide ratios from the analysis of SILAC-labeled proteomes.

In this study we compared Census (5), MsQuant (6), MaxQuant (7), and the Mascot Distiller Quantization Toolbox (http://www.matrixscience.com/distiller.html). The four quantification tools were compared with respect to peptide and protein identification capacities, quality of peptide quantification, and visualization of the quantitative data. We used data obtained from a proteomic experiment on human neuroblastoma SHEP cells in which the synthesis of protein clusters was regulated as the result of the expression of the oncogenic micro-RNA cluster (miR-17-92 (8)). Differential protein expression levels were measured between cells where the micro-RNA cluster was either or not induced. Cells were SILAC labeled with Lys and Arg ($^{12}C_6$ versus $^{13}C_6$) and only the methionine-containing tryptic peptides were sorted and identified by LC-MS/MS on an Orbitrap XL mass spectrometer (this mass spectrometer was operated as described previously (9)) using the previously published methionine COFRADIC proteomic technology (10). This experiment was repeated once with label swapping rendering a "forward" and a "reverse" experiment where the micro-RNA cluster was expressed in the heavy and the light labeled cells, respectively. Given that several previous proteome studies indicated that protein levels were only moderately affected by induction of miRNA expression and that the grand majority of the proteins were not affected (e.g., (11, 12)), we here expected that most proteins would be present in highly similar, if not, equal amounts in both setups, thus facilitating interpretation of quantification data.

2. Materials and Methods

2.1. Installing the Different Software Tools

2.1.1. Census

The Census quantification workflow requires three programs: Census, DTASelect, and RawExtractor (5, 13) that were all developed in the Yates lab (http://fields.scripps.edu/?q=content/software). By unzipping the downloaded file, a folder is created that contains the fully functional Census and DTASelect programs. Installing RawExtractor takes only a few steps in the downloaded installer. Of note, however, is that the Thermo Fisher Scientific Xcalibur software must be installed for RawExtractor

to perform correctly. Census and DTASelect further require Java version 1.5, which is by default installed on most modern PCs but, if not, Java can be downloaded from http://www.java.com/. Both Census and DTASelect can be installed on Windows, Mac, and Linux operating systems, and RawExtractor on the other hand can only be installed on Windows operating systems.

2.1.2. MaxQuant

A download link to MaxQuant, user name, and password are sent by e-mail after a registration form is filled out on http://www.biochem.mpg.de/en/rd/maxquant/Downloads/index.html (7). Unzipping the downloaded file creates a folder with the different fully functional programs. MaxQuant must be run on a 32-bit Microsoft Windows XP or Microsoft Windows Vista PC. Thermo Fisher Scientific Xcalibur software and the .NET framework 2.0 are also required. Of note is that the Windows "Regional and Language Options" should be set to English (14). Following installation, several files that are required for MaxQuant must be copied from the Mascot Server to the /MaxQuant/conf folder. These files are "enzymes.txt," "mod_file.txt," "mascot.dat," and "unimod.xml" and can be found in the folder /inetpub/mascot/config on the Mascot Server. When these files change on the Mascot Server, they should also be updated in the local MaxQuant folder (14).

2.1.3. MsQuant

The programs MsQuant and DTASuperCharge used in the MsQuant quantification workflow are available on the MsQuant web page (http://msquant.alwaysdata.net/msq/download/) (6). The Thermo Fisher Scientific Xcalibur software with the XDK option is required for a fully functional MsQuant installation. Note that the XDK option is by default not selected during installation, so re-installing Xcalibur may be required. MsQuant only runs on computers with Microsoft Windows operating systems.

2.1.4. Mascot Distiller Quantitation Toolbox

In contrast to the other tools Mascot Distiller with the Quantitation Toolbox is not freely available. After payment, a DVD with the software and a license code is sent. Installation is straightforward using an installation wizard. Mascot Distiller requires access to a Mascot Server 2.0. However, if the Quantitation Toolbox is used, Mascot Server has to be version 2.2 or higher. Mascot Distiller can only be installed on Microsoft Windows 2000, XP Professional, Server 2003, Vista, or Server 2008 computers. Although Mascot Distiller can read many different raw file types, it sometimes uses software libraries that are supplied as part of the mass spectrometer software. These libraries and tools must be installed on the computer but this was unnecessary for the Thermo Fisher Scientific Xcalibur software. An installer for the Mascot Daemon tool can be freely downloaded from an installed Mascot Server.

2.2. MS/MS Data Generation

To exclude differences originating from using different MS/MS search engines, all MS/MS spectra were searched and identified with the Mascot database search engine (http://www.matrixscience.com). Mascot generic files (mgf)-containing MS/MS data were therefore generated for every workflow with one mgf file typically containing all MS/MS spectra of one LC-MS/MS run.

2.2.1. Census

RawExtractor 1.8 extracts MS and MS/MS spectra from raw data files from Thermo Fisher Scientific mass spectrometers. These spectra can be saved to MS1 (MS spectra), MS2 (MS/MS spectra), and mzXML (both MS and MS/MS spectra) files (15, 16). Here, the MS2 file format was chosen to extract MS/MS data with default parameters, except for the parameter "data-dependent acquisition" which was selected. MS2 files generated for every LC run were then converted to mgf files by an in-house developed script.

2.2.2. MaxQuant

The "Quant" program was used to generate mgf files. The raw files were first loaded and the numbers of threads – threads indicate the number of parallel processes executed by the program – were selected. This number should not exceed the number of available computing cores minus one (14) since otherwise the computer becomes less responsive. Then, the different parameters on the Mascot database search, SILAC method used, etc., were selected, after which mgf files were generated. In contrast to the other quantification workflows, the MaxQuant algorithm divides all MS/MS spectra in three categories (for double SILAC modus) and thus in three files. A first mgf file contains MS/MS spectra that could not be linked to a SILAC event (a SILAC event is defined as two ion envelopes in an MS spectrum with a specific mass difference corresponding to that of the essential amino acids used for SILAC; here this is 6 Da). The second file contains MS/MS spectra that were linked to the "light ion envelope" in a SILAC event and the third file contains spectra linked to the "heavy ion envelope." MaxQuant additionally creates an index file for every raw data file, a folder with files containing extracted information, a temporary parameter file, etc., all of which are used in the quantification process.

2.2.3. MsQuant

DTASuperCharge 1.37, part of the MsQuant installation, was used to extract mgf files from the raw data files using default settings (one mgf file per LC run). Following installation, the location of the file "extract_msn.exe" must be given in the first tab (Preprocess) on the main screen. This file is part of the Xcalibur program installation and be found in the folder "C:\Xcalibur\system\programs\" if default Xcalibur program installation parameters were used.

2.2.4. Mascot Distiller Quantitation Toolbox

Mascot Deamon 2.2.2 was used to create a search task during which mgf files are generated. Such a task needs several items including raw data files. Several of these can be added to the task in the "data file list" in the "task editor" tab. In addition, a parameter file must be created in the last tab of Mascot Daemon. Another necessary element of a task is a Mascot Distiller Option file, which must be created using an interface shown once the "option" button is clicked on the "task editor" tab in Mascot Daemon and after Mascot Distiller is selected as the "data import filter." However, a previously created.opt file can also be selected and here, the default Orbitrap_low_rs_MS2.opt file was selected.

2.3. MS/MS Spectra Identification

Peptide identification was performed in the same way for the different quantification workflows to minimize its influence on the overall quantification results. The general peptide identification parameters used were the following. MS/MS spectra were identified using a locally installed version of the Mascot database search engine version 2.2.06 (Matrix Science) and the human Swiss-Prot database (version 58.7 of UniProtKB/Swiss-Prot protein database, containing 20,401 human sequence entries). Peptide mass tolerance was set at 10 ppm and peptide fragment mass tolerance was set at 0.5 Da. Trypsin/P was set as the protease, allowing for one missed cleavage. Variable modifications were pyroglutamate formation of N-terminal glutamine, pyrocarbamidomethyl cysteine formation of N-terminal S-alkylated cysteine and acetylation of the alpha-N-terminus. Fixed modifications were *S*-carbamidomethyl cysteine and methionine oxidation (sulfoxide form). Quantitation was set to "SILAC Arg Lys +6"; this allows Mascot to identify arginine- and lysine-containing peptides in their heavy form.

In the MaxQuant workflow, the parameters for identifying MS/MS spectra are set before generating mgf files (see above) and are slightly different from the other quantification workflows. MaxQuant automatically generates three parameter files that can be used for every corresponding type of mgf file (see above). Most of the parameters were shared between the three parameter files and are the following. Peptide mass tolerance was set at 7 ppm and peptide fragment mass tolerance was set at 0.5 Da. Trypsin/P was set as the protease with a maximum number of one missed cleavage. The number of labeled arginines and/or lysines per peptide was set to a maximum of 2. Variable modifications and fixed modifications were the same as for the other quantification methods. The incorporation of the heavy form of an arginine or lysine (both + 6 Da) was set as a variable modification for the mgf file containing the MS/MS spectra that were not matched to a SILAC event. These were set as a fixed modification for the mgf file containing spectra that were matched to the heavy

peptide in a SILAC event. No extra modifications were added to the identification process of spectra in the mgf file containing spectra that were matched to the light peptide in a SILAC event.

MaxQuant requires a species-specific protein database for identification of MS/MS spectra and further requires the addition of reversed protein sequences to calculate a false discovery rate (FDR). A reversed database concatenated with the human Swiss-Prot database (version 57.0 of UniProtKB/Swiss-Prot protein database, containing 20,332 human sequence entries) was created by the "SequenceReverser" program which is part of the MaxQuant installation.

2.4. Quantification

2.4.1. Census

Census requires three data sources. A Census configuration file must first be created in the graphical Census environment and this file contains parameters for the quantification process including parameters for chromatogram extraction, peak finding, chromatogram alignment, and amino acid elemental composition information. Second, identification information for the MS/MS spectra is needed, which can be provided as pepXML or DTAselect-filter files. Of note is that the pepXML files generated by Mascot 2.2.06 in our setup could not be used since errors occurred during file parsing by Census. Therefore, we opted for DTAselect-filter files but these, generated by Mascot 2.2.06, also gave errors upon parsing. The DTAselect 2.0.21 program (http://fields.scripps.edu/?q=content/software) was used to extract and save MS/MS identification data in a DTAselect text file (13). This text file was then used to create a DTAselect-filter file. Many different filters (peptide length, identification score, peptide charge, residues in peptide sequence, etc.) can here be applied. Dat files (files generated by Mascot and containing identification information) can be used to run DTAselect. However, the used version of DTAselect (2.0.21) could not read dat files from Mascot 2.2.06, and these thus needed to be tweaked by an in-house developed script. The MS/MS spectrum filenames in the DTAselect files were also changed by an in-house developed script to names with the following syntax: "raw data filename. start scan. end scan. charge" (e.g., reverseExperiment.1453.1453.2). This was needed since, else Census failed to perform quantification. A DTAselect-filter text file was then created by DTAselect but no filters were here selected. Third, MS information is needed. Census accepts both MS1 and mzXML files. Since RawExtractor 1.8 was used to extract the MS/MS information from the raw data files and save it in MS2 files (see above), it was also used to create MS1 files.

Quantification is started in Census via a simple GUI, after which a list of protein is presented. Here, peptides with their corresponding ratios and XIC are also shown. Census creates a Census_chro.xml file in the folder containing the MS1 or mzXML

files. This file holds all information needed for Census to create the user interface after quantification, and thus further allows the user to view and analyze data. A Microsoft Office Excel readable report can also be generated with the use of specific filters including the determination score and outlier threshold *p*-values.

2.4.2. MaxQuant

Quantification of peptides in the MaxQuant workflow is done by the "Identify" program. In the "Input file" tab, three different files are required. The raw data files must be selected in the first panel. The index and other files generated by the "Quant" program cannot be deleted or moved and should be in the same folder as these raw data files. The peptide identification information stored in different Mascot dat files should be set in the second panel. The protein database used for peptide identification is the last file that is required. The quantification parameters are set in the second tab, and here the default parameters were used. Quantification parameters are divided into two groups and either are linked to peptide identification (e.g., false discovery rate and minimum peptide length) or deal with protein quantification. The number of threads (see above) can also be selected in this panel.

2.4.3. MsQuant

The MsQuant quantification parameters are set by choosing "Options" in the "utility" menu. The quantitation mode was set to "v2.1 Arg13C6 and Lys13C6." "DTASuperCharge, v1.19 and later" was selected as the MGF file generator and "LTQ-FT/LTQ-Orbitrap (.raw)" was selected as raw file type.

MsQuant requires html Mascot result pages (these must be saved by Microsoft Internet Explorer 6) as the source for peptide identification information. Html pages are not the default pages created by the Mascot server and specific parameters must be selected on every page. However, if the date of creation of the Mascot result file and the name of this result file is known, a customized url can be created that will generate html pages that can be used by MsQuant. An example of such an url is http://MASCOT_SERVER_ADRESS/mascot/cgi/master_results.pl?file=..%2Fdata%2FDATE%2FFILENAME.dat&REPTYPE=peptide&_sigthreshold=0.05&REPORT=AUTO&_server_mudpit_switch=99999999&_ignoreionsscorebelow=0&_showsubsets=0&_showpopups=TRUE&_sortunassigned=scoredown&rbrchkbo.

MascotResultLauncher, a program that is installed with MsQuant, should be able to create html pages that can be used by MsQuant; however, in our hands, this program failed.

Following selection of raw data files and html Mascot result pages, MsQuant will find associations between both files. Therefore, it is important that the filename of the html file reflects that of the RAW file.

Prior to starting quantification, the html file is first parsed to acquire peptide identification information, after which these data are visualized in a protein-centric table. The quantification process can then be started for all or a selection of proteins. Peptide identification and quantification results can be stored as text files which can also be loaded in Microsoft Office Excel.

2.4.4. Mascot Distiller Quantitation Toolbox

When a search task is finished in Mascot Daemon, .rov and .mgf files are created. Only one .rov file at a time can be opened in Mascot Distiller. The following parameters concerning extracted ion chromatogram (XIC) detection were changed in the "Format option" panel: XIC threshold was set from 0.1 to 0.3, XIC smooth was set from 3 to 1, and Max XIC width was set to 250. The correlation score, a quantification quality parameter, was set from 0.7 to 0.9 (see below). The other quality parameter, the standard error, was 0.14 and not changed.

2.5. Protein Ratios

Only proteins with two or more distinct identified peptides were further used. To reduce the problem of protein inference (peptides pointing to multiple proteins in the search space), every protein needed to contain a unique peptide or an Occam's razor peptide (17), being a peptide that can be linked to multiple proteins, but was here linked to the protein with the most distinct peptide identifications or highest protein coverage. In this way, a minimal protein set is created. This method of selecting quantified proteins can easily be done for MaxQuant and Mascot Distiller. Peptide identification information in both export files contains links to, if applicable, the different proteins containing this peptide. For the two other quantification workflows, however, this information is not readily available. Therefore, all proteins with two distinct peptide identifications will be selected for MsQuant and Census, and proteins with two distinct, one unique, and one or two Occam's razor peptide(s) will be selected for MaxQuant and Mascot Distiller.

The different quantification workflows all have specific methods to calculate protein ratios, the simplest method being calculating the average of all peptide ratios. MaxQuant uses the median of the peptide ratios. Slightly more complex is the method where the average is taken from peptide ratio averages. This normalizes the contribution of distinct peptide ratios. Another way of creating weighted protein ratios is performed by Census in which individual peptide weights are determined by the inverse square of the standard deviation of the measurement (5). On top of that, the peptide ratios or the calculated protein ratios can be normalized to correct for mixing unequal amounts of protein material. To allow comparison of protein ratios calculated by the different quantifications workflows, the average of the peptide ratios was used as protein ratio and not the protein ratio that was calculated by the individual programs.

The final goal of any quantitative proteomic experiment is to find regulated proteins between different conditions. However, how is a regulated protein defined? In our opinion, regulated proteins should be considered against the background of all identified and quantified proteins in an experiment. The Rover tool created in our lab accepts quantitative data from Mascot Distiller Quantitation Toolbox, MaxQuant, MsQuant, and Census (18). Visualization of these data is done such that the user can select and validate algorithm-suggested regulated proteins.

In Rover, all calculated peptide ratios are used to create a peptide ratio reference set, which is used to compare against peptide ratios from a selected protein. Robust statistics are performed to correct for the influence of outliers on the average and standard deviation of the peptide ratio reference set. The standard deviation is calculated by robust statistics (http://www.rsc.org/images/brief6_tcm18-25948.pdf). The $\log_2$(peptide ratio) will be transformed by a process called winsorization. Here, the $\log_2$(peptide ratio) with a given value (*see* **formula (1) and (2)**) will be changed in (**3**) and (**4**), respectively:

$$\log_2(\text{peptide ratio}) < \text{median} - 1.5\,\sigma \text{ (of the previous winsorization cycle)} \quad [1]$$

$$\log_2(\text{peptide ratio}) > \text{median} + 1.5\,\sigma \text{ (of the previous winsorization cycle)} \quad [2]$$

$$\text{median} - 1.5\,\sigma \text{ (of the previous winsorization cycle)} \quad [3]$$

$$\text{median} + 1.5\,\sigma \text{ (of the previous winsorization cycle)} \quad [4]$$

After each winsorization, the standard deviation is calculated as 1.134 times the standard deviation of the winsorized data. If the difference between two consecutive standard deviations is smaller than 0.000001, the winsorization process stops and the standard deviation (σ_{calc}) will be 1.134 times the standard deviation of the last winsorized data.

The final standard deviation (σ_{fin}) of the reference set is calculated both with the instrument standard deviation (σ_{instr}) and with the winsorized standard deviation (σ_{calc}) (*see* formula (**5**)). The instrument standard deviation is the measured standard deviation for $\log_2$(peptide ratio) values of peptides mixed in equal amounts and analyzed by the mass spectrometer. In this project, 0.14277 was set as the instrument standard deviation:

$$\sigma_{fin} = \sqrt{\sigma_{calc}^2 + \sigma_{instr}^2} \quad [5]$$

This standard deviation, calculated with robust statistics, and the mean of the $\log_2$ peptide ratio reference set are used to calculate a *Z*-score for every peptide ratio. A *Z*-score reflects the distance

between the mean and the peptide ratio in function of the standard deviation. By using a protein ratio *Z*-score (calculated with the protein ratio), proteins are identified that are regulated (a protein ratio *Z*-score must here be larger than 1.96 or smaller than –1.96). The different up- or downregulated proteins in the two experiments and the different quantification workflows were in this way extracted with the Rover tool.

3. Results and Discussion

3.1. Peptide Identification

MS/MS spectrum peptide identifications are required before quantification can start. The different quantification workflows trace back the MS ion envelopes of identified peptides via the scan number of the MS/MS spectrum. All workflows use peptide identification information for quantification and thus do not quantify ion pairs directly in the raw data. Trypsin was here used for generating peptides, hence peptides will generally only carry one SILAC amino acid (Lys or Arg) and peptide pairs are spaced by 6 Da. Missed cleavages (e.g., Lys or Arg followed by Pro) will create larger mass differences. Some peptide quantification workflows extract extra information from the peptide identification. The Mascot Distiller Quantitation Toolbox, for instance, creates a predicted ion envelope for the identified peptide taking into account the natural occurrences of the atomic elements. This prediction is matched with the observed ion envelope and a correlation score is calculated which gives extra information on the validity and quality of the peptide identification and quantification (see below).

Peptide identifications are generally influenced by two parameters. A first parameter is the method used to extract MS/MS spectra from raw data and this was different for every quantification method used. The second set of parameters are introduced by the MS/MS spectrum identification algorithm and among others include allowed precursor mass errors, amino acid modifications, protein database, number of allowed miss cleavages (see above). These parameters were very comparable between the different quantification workflows.

The influence of mgf file generation or MS/MS data extraction on peptide identification was first studied. Peptizer, a tool developed in our lab to analyze Mascot results and find false-positive peptide identifications, was used to extract all peptide identifications with a Mascot ion score higher than the threshold (set at 95% confidence) (19). The number of MS/MS spectra generated by the different workflows was first analyzed. On average, 52.2% more MS/MS spectra were generated by the different

workflows for the reverse experiment (average of 134,763 spectra) compared to the forward experiment (average of 88,549). This large difference in the number of generated MS/MS spectra is further reflected in the number of identified spectra: on average, 12.5% more spectra were identified in the reverse experiment (average of 38,197) compared to the forward experiment (average of 33,937), but only 28.3% of all MS/MS spectra in the reverse experiment were identified, compared to 38.3% in the forward experiment (**Fig. 25.1**). This indicates that a higher number of MS/MS spectra does not linearly increases the number of peptide identifications.

Different workflows generated different numbers of MS/MS spectra. The MsQuant workflow generated the most spectra, Census and MaxQuant performed similar, and Mascot Distiller gave the lowest number of spectra. The latter, however, identified the largest fraction of spectra (35.2% on average for both experiments) compared to MaxQuant (34.2%), Census (33.2%), and MsQuant (31.1%). On the other hand, MaxQuant identified the largest number of unique peptides.

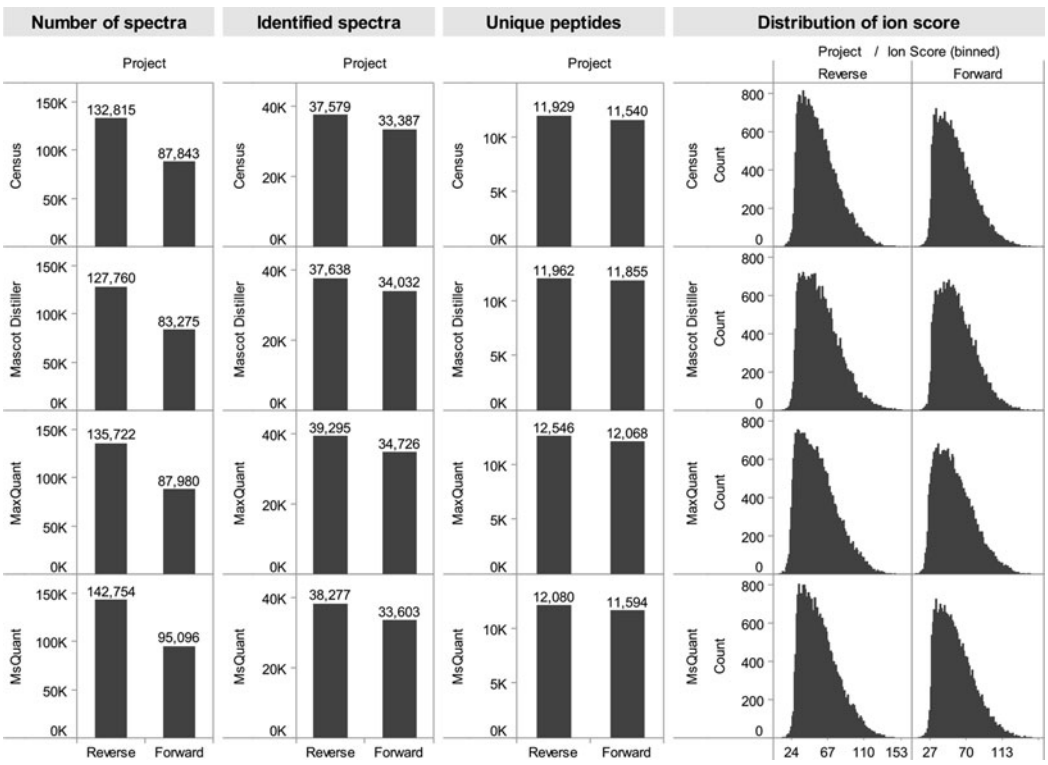


Fig. 25.1. Overview of generated and identified MS/MS spectra, and identified peptides by all tested workflows for both data sets. Peptizer was used to extract confident (threshold 0.05) identified peptides; the number of generated MS/MS spectra, identified spectra, and the number of unique peptides are shown in the first three columns. The distributions of the ion scores of the identified spectra are shown in the *right column*.

The overall identification efficiency can be shown as a distribution of the ion scores of confident peptide identifications which indicated that this score peaked around 29 for the different workflows and the two different projects (**Fig. 25.1**) and also that there were no large efficiency differences between the workflows. About 12,742 and 13,209 distinct peptides were found for the forward and reverse experiments, respectively, when all identifications from all workflows were combined (**Fig. 25.2**). A total of 84.7% of these peptides (averaging over the two experiments) are shared by all four quantification workflows. MaxQuant and Mascot Distiller contributed 3.5 and 2.2%, respectively, to the overall number of distinct (unique) peptides. These numbers are high compared to MsQuant (0.6%) and Census (0.27%) and even if the results of both these workflows were combined, only 1.5% of unique peptides were found.

Our results clearly indicate that raw data processing for creating and extracting MS/MS spectra is an important factor determining peptide identification. However, the final result (the number of identified peptides) was comparable between the different methods, although MaxQuant and Mascot Distiller clearly identified relatively more peptides compared to the other two methods.

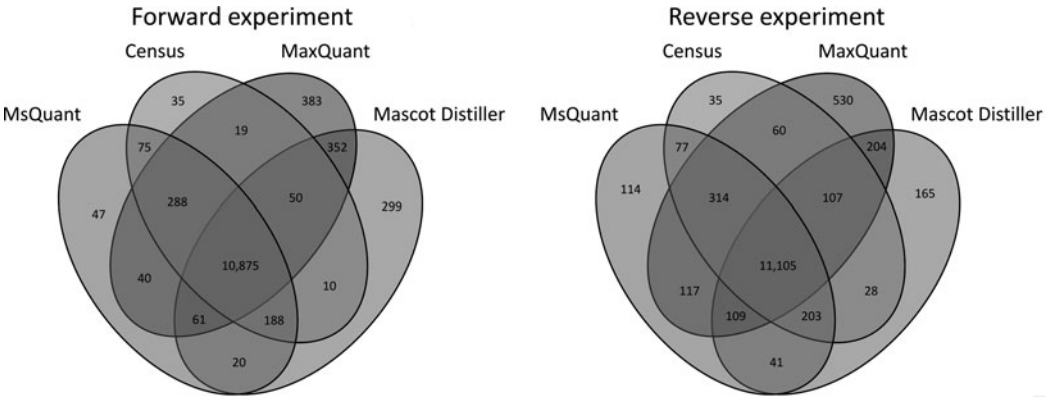


Fig. 25.2. The overlap of unique peptides identified by the different workflows for both data sets.

3.2. Quantification

3.2.1. Analysis of the Assigned Quality Parameters for Peptide Quantifications

The identified peptides were quantified by the different quantification workflows. Both Mascot Distiller and Census have parameters that indicate the quality of calculated peptide ratios. In Census, the linear regression is calculated and given as the regression score or determination factor, which is the square of the regression score. A regression line is plotted between the intensities of the light and the heavy peptide peaks in different scans. The distance between the points and the line is the regression score and reflects the relationship between the intensities

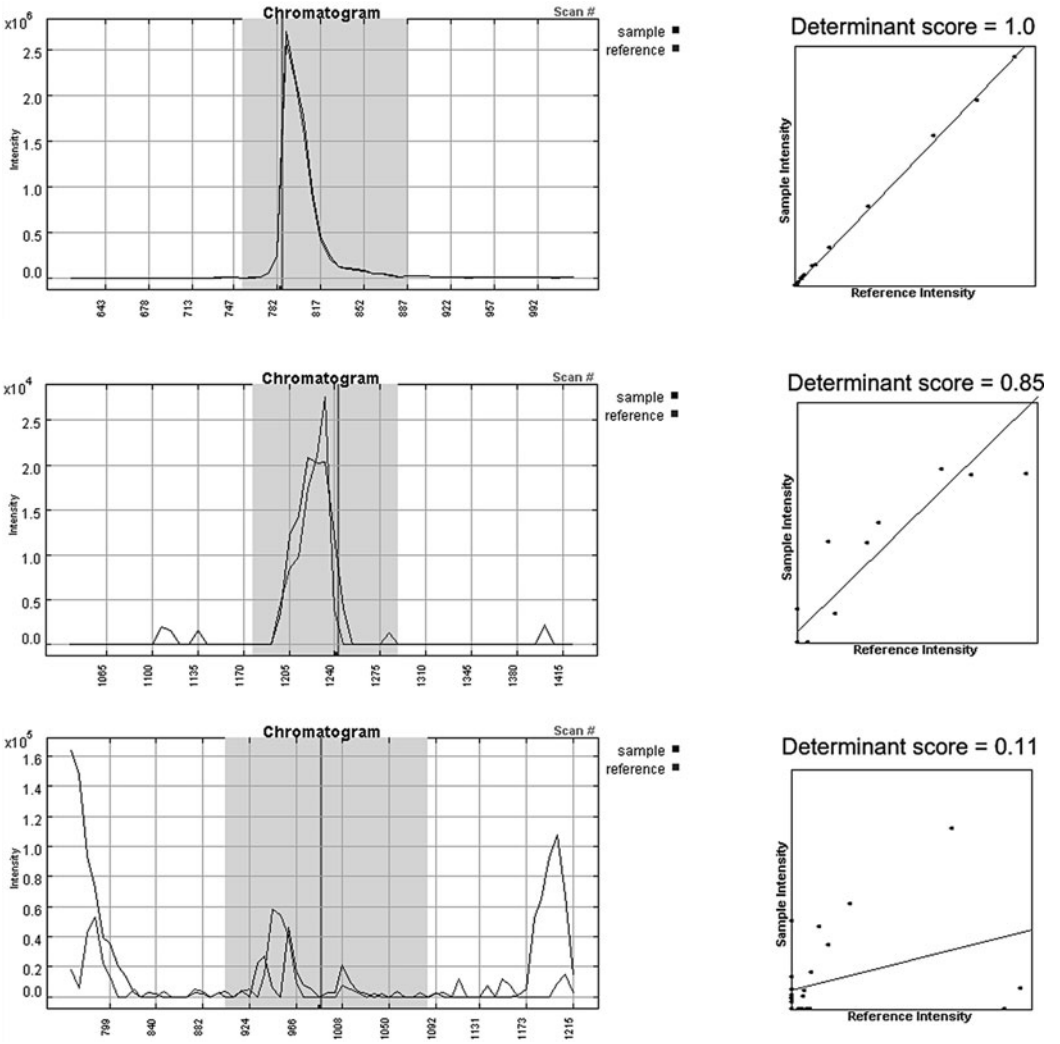


Fig. 25.3. Examples of Census determination factors. XICs for three different peptides visualized by Census are shown in the *left column*. The *gray zones* represent the areas Census used to calculate the peptide ratio. The corresponding regression lines and determination factors are shown in the *right column*. The *upper two rows* represent good determination factors and corresponding XICs and the *lower row* shows a bad XIC and low determination factor.

of the light and heavy peptides at different time points. Good correlation between these intensities gives a determination factor close to 1 and bad correlations gives a score close to 0 (*see* also **Fig. 25.3**). The slope of the regression line is then used to calculate a ratio value. Hence, bad regression scores likely point to incorrect peptide ratios and Census therefore suggests to only use peptide ratios with a determination factor higher than 0.5. The relationship between these determination factors and peptide ratios is shown in **Fig. 25.4** and no obvious correlation was at first sight noticed. However, if a distribution was plotted of the Census valid and discarded peptide ratios, the ratio distribution of valid

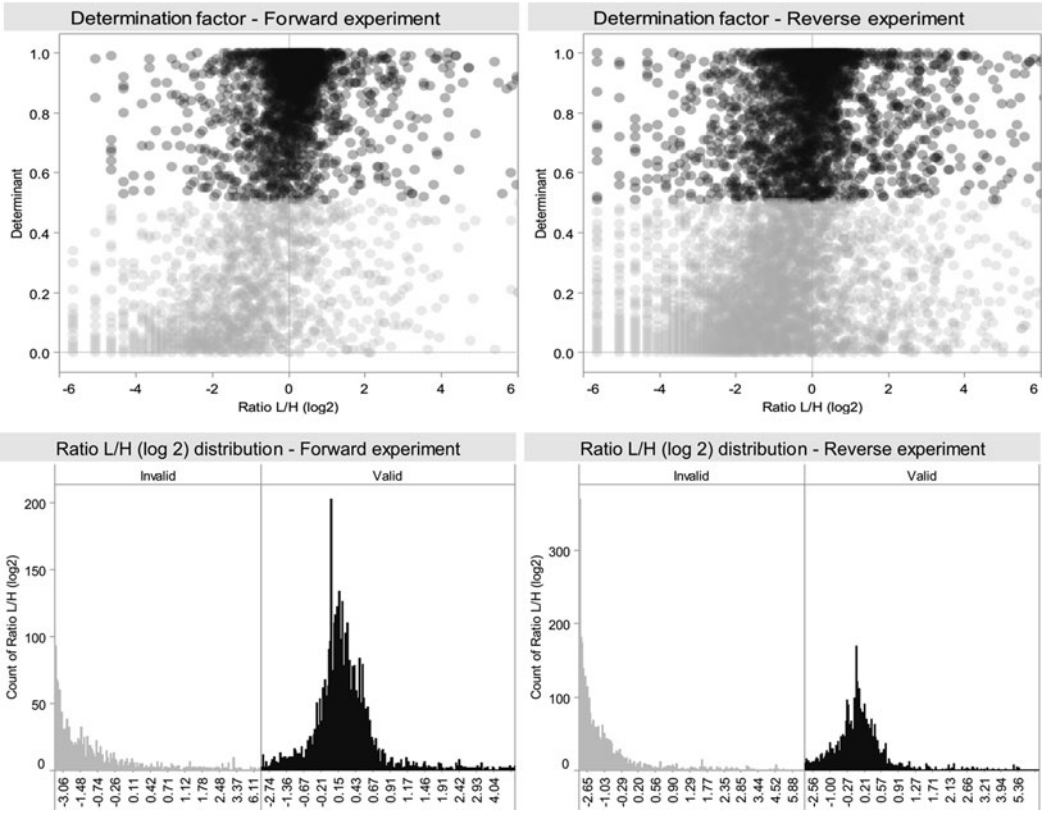


Fig. 25.4. The scatter plots show the determination factor versus the corresponding peptide ratio (log_2) for both data sets. *Black dots* represent peptide ratios suggested by Census to use in further calculations (valid ratios), *gray dots* are peptide ratios suggested to be invalid. The histograms plot log_2 peptide ratios for Census valid and invalid peptide ratios for both data sets.

ratios clearly better reflects the expected distribution than that of the discarded ratios. This further indicates that filtering results are very important when working with the Census quantification workflow.

The Mascot Distiller Quantitation Toolbox has three parameters that describe the quality of peptide ratio quantifications: the standard error is similar to the regression score of Census, the correlation score describes the correlation of the experimental and predicted peptide ion envelopes, and the fraction score is the fraction of MS peaks that were used to calculate peptide ratios. The latter is a good indicator to search for co-eluting peptides. The LTQ-Orbitrap used in our study has a very high resolution in MS mode (20) and the ion envelopes were thus very well resolved, excluding in most cases influences of co-eluting peptides on ratio calculation. Therefore, the fraction score was not a discriminative parameter for the quality of peptide quantifications in our study. By plotting the standard error and correlation score

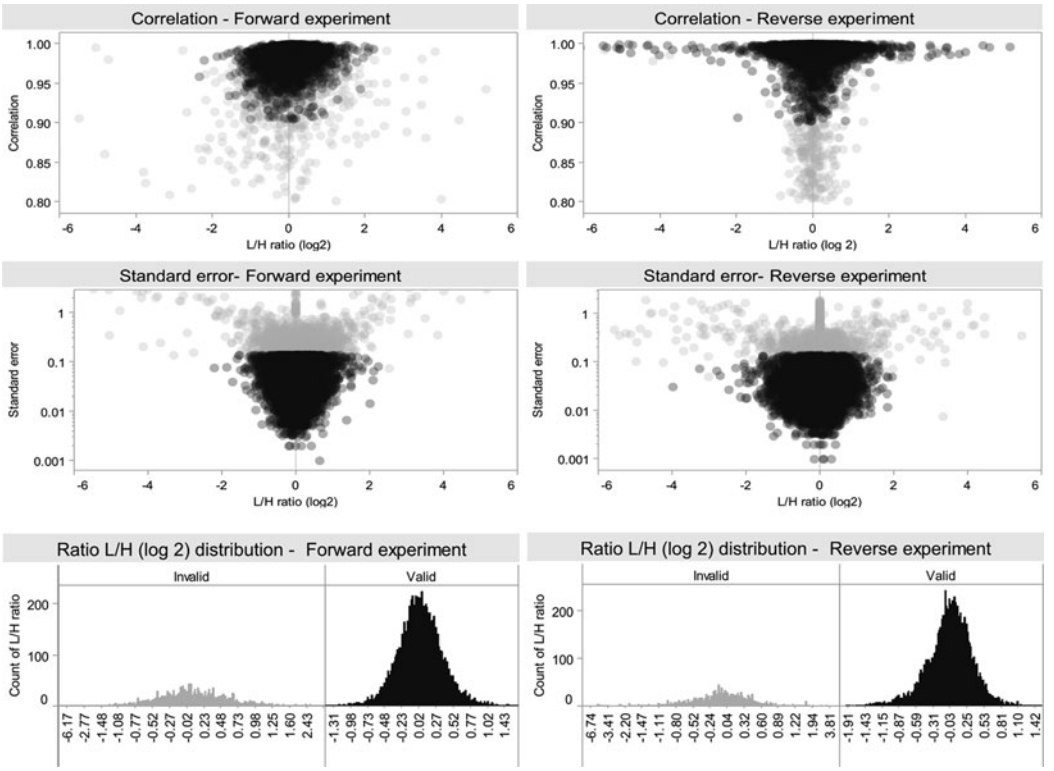


Fig. 25.5. The scatter plots compare the correlation value or standard error to their corresponding peptide ratios ($\log_2$) for both data sets. *Black dots* represent peptide ratios accepted as valid by Mascot Distiller, whereas *gray dots* are peptide ratios suggested to discard. Histograms plot $\log_2$ ratio values for both the valid and the invalid peptide ratios for both data sets.

against the peptide ratio (**Fig. 25.5**) it was found that lower standard errors better correlate with peptide ratio close to the mean ratio. This was much less pronounced for the correlation score parameter and thus only the standard error can indicate outlier peptide ratios in a fast and easy way. The distributions of invalid ratios were roughly similar to those of the valid ratios, indicating that removal of invalid ratios probably also discards numerous good quantifications or that the threshold for accepting peptide quantifications was set too high. Filtering based on peptide ratios is thus less recommended for the Mascot Distiller Quantitation Toolbox compared to Census.

The Mascot Distiller Quantitation Toolbox will remove 16.6% (averaging over the two experiments) of all peptide quantifications, while Census discards on average 36.2%. Furthermore, the ratio distribution from the peptide quantifications discarded by Mascot Distiller is much better than that of the discarded peptide ratios by Census. This again clearly indicates a need for excluding peptide quantifications in the Census results to raise

the general quality level and this in contrast to Mascot Distiller where the excluded peptide quantifications do not lower the overall quality to the same extent as with Census.

3.2.2. Correlations Between Different Quantification Workflows

All calculated peptide ratios by the different quantification workflows were used to calculate a mean, a standard deviation, borders of the 2.5 and 97.5% quantiles, and a confidence interval $[-1.96\times\sigma;\ 1.96\times\sigma]$ (**Table 25.1**). A distribution of the $\log_2$(peptide ratio) values was created for every method and experiment (**Figs. 25.6** and **25.7**). MaxQuant and Mascot Distiller have in both experiments distributions that resemble best the expected distributions, indicating that both quantify SILAC-labeled peptides very well (see below). The distributions of Census' ratios contain spikes and are in general less smooth. This was also apparent in the $\log_2$ ratio region between –6 and –4 when plotting the determinant score versus the $\log_2$ ratio (**Fig. 25.4**): in this area Census did not calculate peptide ratios "fluently" but rather in "individual steps." The distributions of MsQuant ratios show in both experiments a spike at zero, but besides this, the ratio distributions seem rather good. However, the $\log_2$ ratio for the 97.5% quantile (**Table 25.1**) is 8.216 in the forward and 9.324 in the reverse experiments indicating that the ratio distributions are heavily skewed to positive values, which is unexpected since protein levels are only moderately affected in the given biological setup. These 97.5% quantile values are extreme compared to those of the other workflows (average of 0.962 for the forward and 1.306 for the reverse experiments) and, combined with the rather well ratio distribution, this suggests that some peptide ratio quantifications by MsQuant must be incorrect.

The correlation of peptide ratios calculated by the different workflows is shown in scatter plots in **Figs. 25.6** and **25.7**. If a trend line can be observed in the scatter plot, then the correlation is high, meaning that peptides quantified by one method have in general a similar calculated ratio by another method. The presence of such a trend line can be analyzed by calculating the actual correlation coefficient. This coefficient is 1 when the correlation between two data sets is perfect, implying that all points in the scatter plot lay on one diagonal line. If, however, this correlation coefficient is close to zero, correlation between two data sets is purely random. Of note here is that at first sight such trends are more apparent when MaxQuant or Mascot Distiller is compared to MsQuant, and much more vague with Census. The correlation scores, on the other hand, clearly indicate that the comparisons with Census yield on average better correlations which can be explained by the fraction of peptide ratios that are incorrectly calculated by MsQuant. The best correlation coefficient is in both

Table 25.1
Statistical features of the results generated by the different quantification workflows for both data sets

	Forward experiment				Reverse experiment			
	Census	MsQuant	MaxQuant	Mascot distiller	Census	MsQuant	MaxQuant	Mascot distiller
$[-1.96\sigma; 1.96\sigma]$ in $\log_2$	[–0.504; 0.932]	[–0.563; 1.225]	[–0.747; 0.517]	[–0.635; 0.701]	[–0.842; 0.982]	[–1.338; 2.194]	[–0.93; 0.578]	[–0.771; 0.677]
Standard deviation ($\log_2$), σ	0.359	0.447	0.316	0.334	0.456	0.883	0.377	0.362
Median ($\log_2$)	0.214	0.331	–0.115	0.033	0.07	0.428	–0.176	–0.047
Quantified proteins	454	2,066	1,452	2,092	558	2,425	1,689	2,135
Proteins with upregulated ratio	27	718	50	35	44	1,134	59	73
Validated proteins with upregulated ratio	8	38	45	35	158	94	55	71
Proteins with downregulated ratio	51	11	35	37	74	4	66	64
Validated proteins with downregulated ratio	14	5	29	37	25	1	59	63

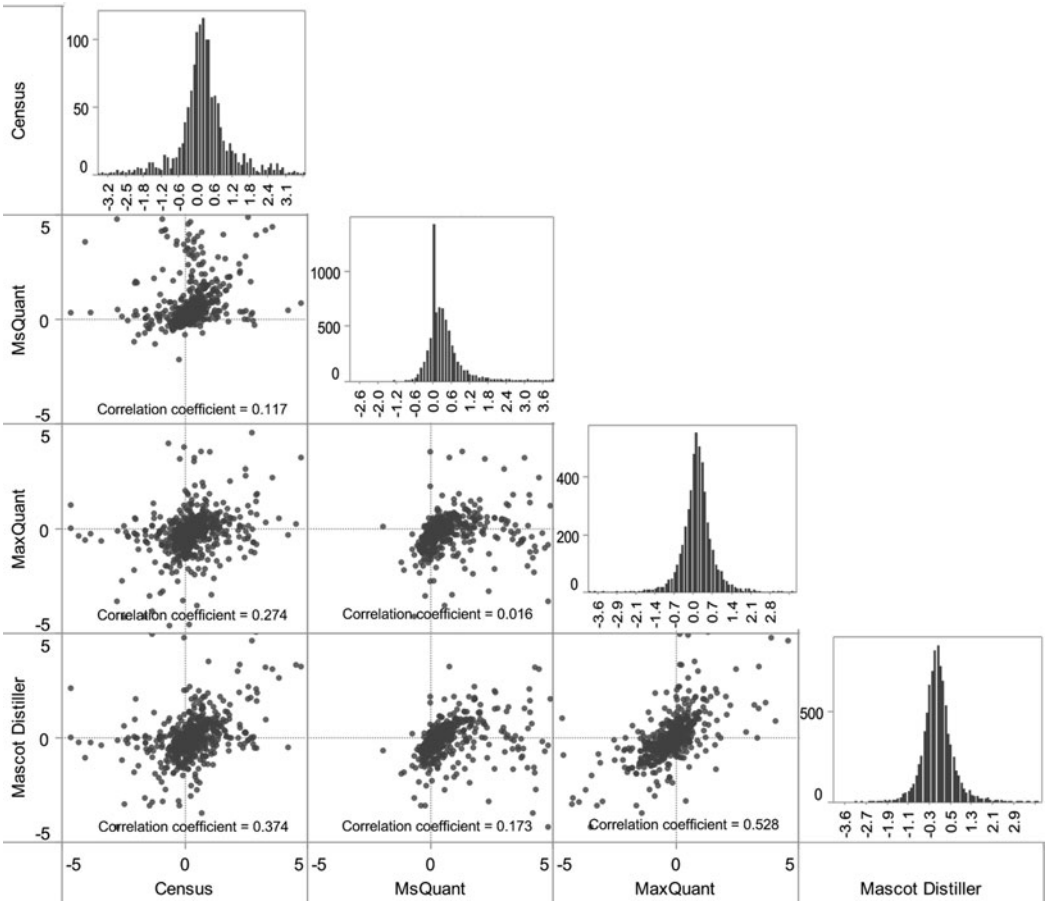


Fig. 25.6. Peptide ratios calculated for the forward experiment. The $\log_2$ peptide ratio distribution for the different quantification workflows are on top of every line. Ratio values calculated by two different quantification workflows in the forward experiment linked to a peptide identified and quantified in both workflows were used to create scatter plots which visualize the correlation between two quantification methods.

experiments found upon comparing results from Mascot Distiller and MaxQuant.

Next, protein ratios were calculated as described above (*see* **Section 2.5**) and the resulting number of regulated proteins is given in **Table 25.1**. The number of upregulated proteins calculated from MsQuant results was enormous: 718 on a total of 2,066 proteins for the forward experiment and 1,134 on a total of 2,425 proteins for the reverse experiment. The protein ratio here is the mean of all peptide ratios and, if one of these peptide ratios is incorrect and has an extreme value, the calculated protein ratio will follow this trend and will thus also be incorrect. Thus, the high numbers of upregulated proteins in MsQuant could be explained by those peptide ratios that were badly quantified as being very large. This clearly demonstrates that suggested regulated proteins need to be manually validated, which was here done

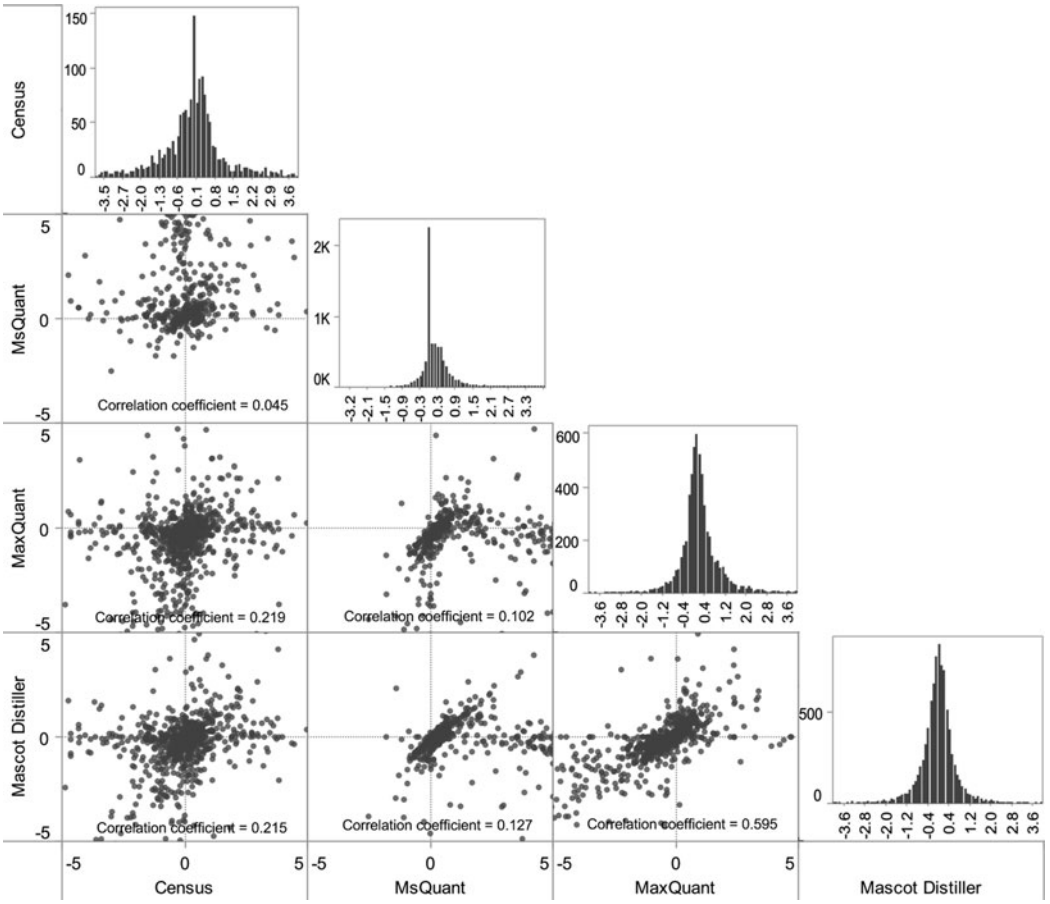


Fig. 25.7. Peptide ratios calculated for the reverse experiment.

with the Rover application (18). In brief, Rover visualizes protein information in an intuitive and informative display and an example of an upregulated protein that was validated by Rover is shown in **Fig. 25.8**. Protein ratios were then compared over the different quantification workflows using scatter plots (**Fig. 25.9**). The best correlation coefficient was again found in both experiments between protein ratios calculated by MaxQuant and Mascot Distiller.

Of further note is that not all proteins were quantified by all workflows (**Fig. 25.10**). Considering only the regulated proteins, the overlap is the highest between results from MaxQuant and Mascot Distiller, supporting the correlation between protein and peptide ratios indicated above. The number of proteins that were found regulated in only one workflow is generally quite high: on average 67.3 and 89.2% of the regulated proteins found by MsQuant and Census, respectively, were such proteins. This is

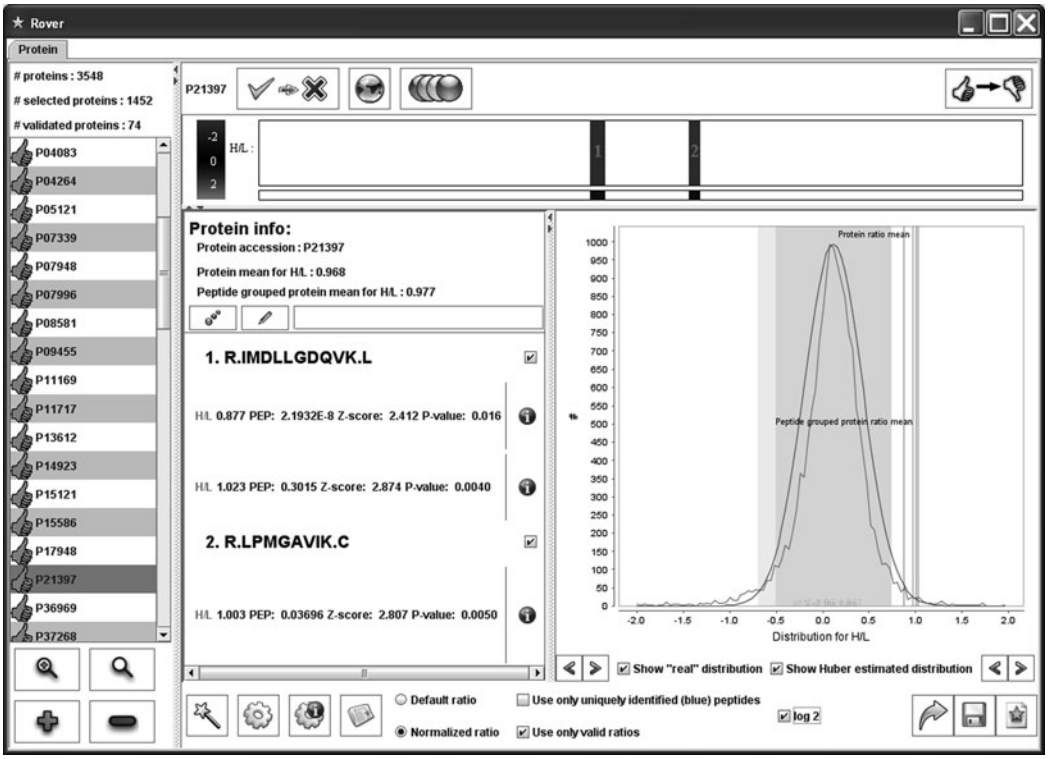


Fig. 25.8. Screenshot of the protein quantification visualization and validation program, Rover. Different panels highlight information from different angles, enabling the user to validate regulated proteins.

much higher than the 44 and 46.7% of uniquely regulated proteins found by MaxQuant and Mascot Distiller.

3.3. Data Analysis, Visualization by and User Experience of the Different Quantification Workflows

The different workflows try to visualize the data in such a way that they explain the calculated peptide or protein ratios to the user. One exception is MaxQuant (7) for which only a beta version of a data viewer is currently available, but we were, however, not able to start this application. This makes the quantification process of MaxQuant more a black box than the other methods but Rover can visualize MaxQuant data in such a way that suggested peptide and protein quantifications can be analyzed and validated (18). Census and MaxQuant quantify different raw data files – which are typically generated per proteome study – simultaneously (5). MsQuant, however, lacks this useful feature (6). Mascot Distiller has an option to quantify different raw files simultaneously, but the output files for peptide identification and quantification become too large for easy handling and parsing. Therefore, this multi-file approach of Mascot Distiller is quite hard to use. Mascot Distiller further does not store quantification settings of a given quantification run as default settings, implying that such

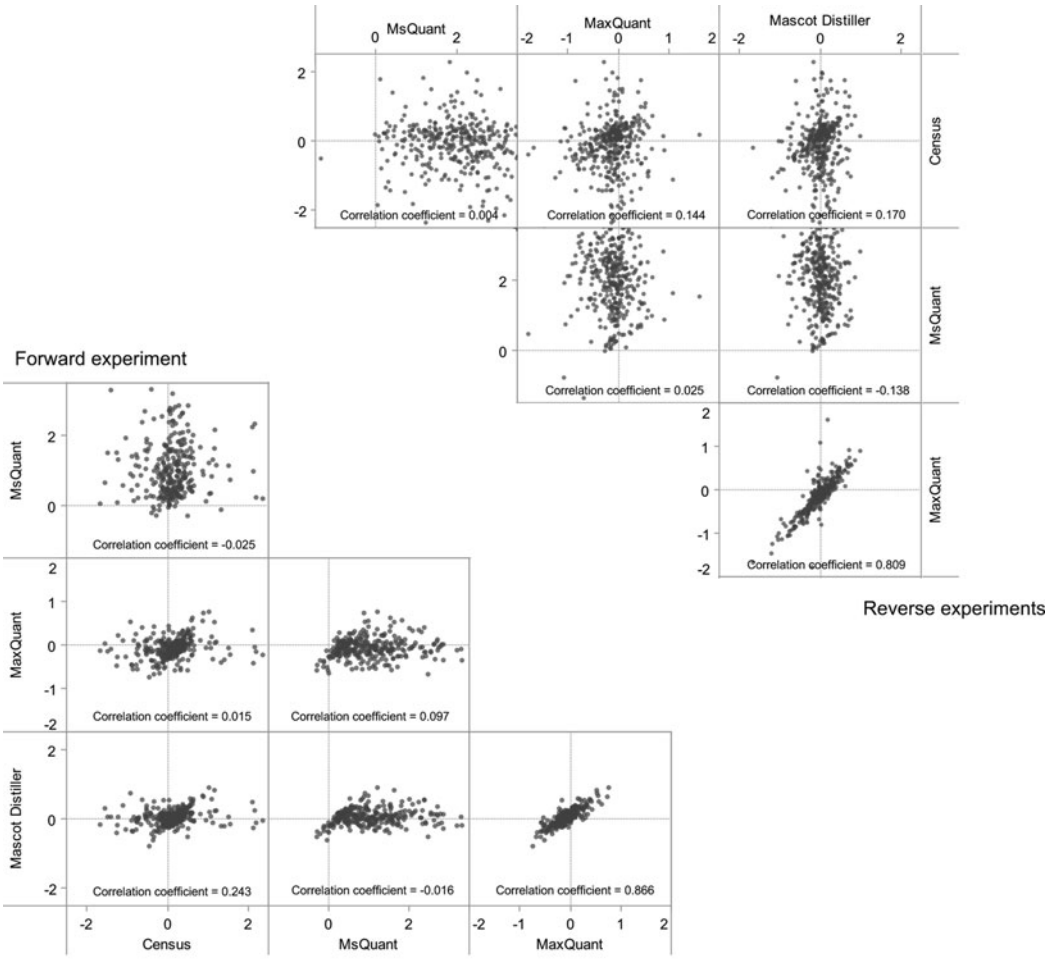


Fig. 25.9. The *upper right triangle* shows six scatter plots visualizing the correlation between protein ratios from different quantification workflows for the reverse experiment. The scatter plots for the forward experiment are shown in the *lower left triangle*.

settings must be set for every raw data file, making it more error prone. MsQuant, on the other hand, stores by default the settings of the last run. User-specific script can be developed to ease the process of starting and saving quantification of peptide ratios by MsQuant and Mascot Distiller.

Following quantification, the census_chro.xml file is read by Census to extract all information necessary to visualize the quantification data. MsQuant creates an mb4 file and Mascot Distiller a rov file. Both contain quantification information and both need raw data files to visualize results. MsQuant also needs the peptide identification html files for visualizing quantification data. Identified peptides can be visualized by MsQuant and Mascot

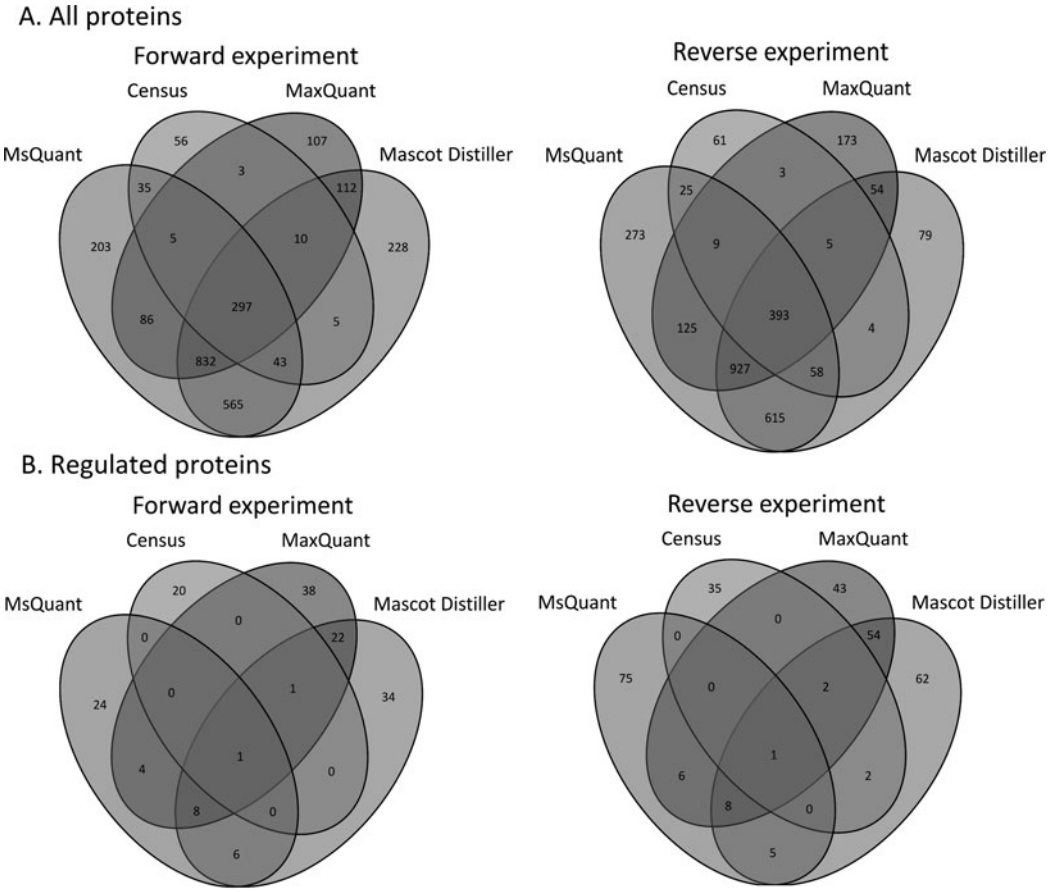


Fig. 25.10. Overlap of all and regulated proteins quantified by the different workflows.

Distiller as annotated MS/MS spectra (**Fig. 25.11**). This is very useful to scan for misidentified peptides and thus inaccurate protein ratios. MsQuant further allows the user to discard peptide ratios with bad peptide identifications, but this is not possible in Mascot Distiller.

The extracted ion chromatogram (XIC) is visualized by MsQuant, Census, and Mascot Distiller. In Census and Mascot Distiller the XIC region used for calculating the peptide ratio is indicated directly on the XIC, in MsQuant this region is indicated in a table and can be adapted there. Such XIC regions further allow checking the quality of peak picking and XICs in general give a good impression of the calculated peptide ratio. Further MS scan information is only visualized by MsQuant and Mascot Distiller, providing more information on the actual quantification (**Fig. 25.11**).

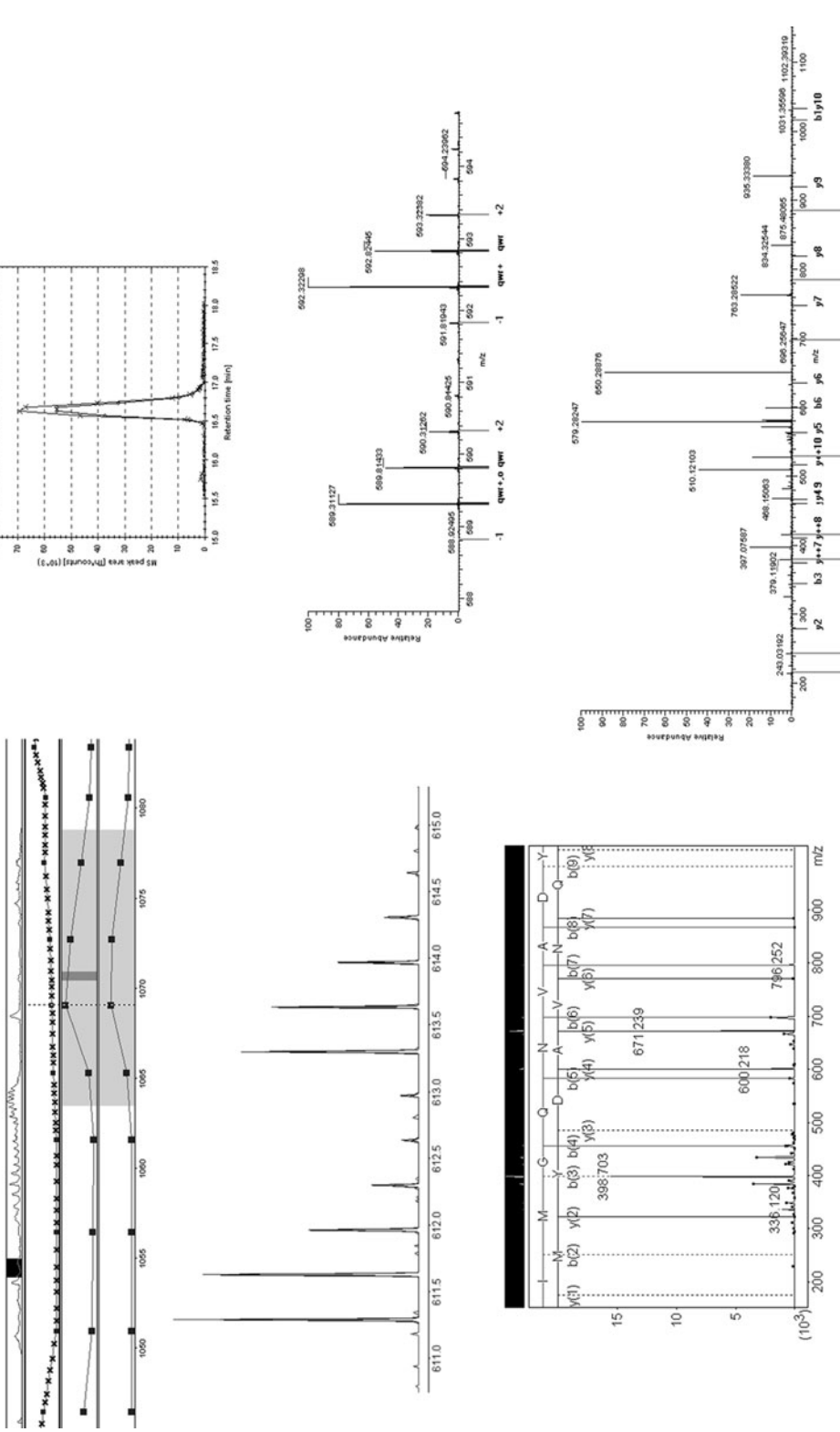


Fig. 25.11. The different data visualization options in Mascot Distiller (*left column*) and MsQuant (*right column*). XICs are shown in the *top panels*, MS scans in the *middle panels*, and annotated MS/MS spectra in the *lower panels*.

4. Conclusions

Since the different methods only quantify MS ion envelopes that are linked to an identified peptide, peptide identification and also MS/MS data generation is of general high importance for the overall peptide and protein quantification process. The different quantification workflows here tested use different methods for creating MS/MS data; nevertheless, they generated comparable results considering numbers of identified peptides. Mascot Distiller and MaxQuant, however, were able to identify a greater portion of unique peptides, indicating that both extract and interpret the MS/MS data in a unique way.

The correlation of peptide and protein ratios was the highest between MaxQuant and the Mascot Distiller Quantitation Toolbox. Further, correlations of the peptide ratios from these two methods with Census were better than the correlations with MsQuant. The correlation of the protein ratios, however, was rather bad for both MsQuant and Census when compared with MaxQuant or Mascot Distiller.

The visualization methods of the quantification workflows, except for MaxQuant, are generally intuitive and informative. The problem is, however, that they do not allow easy and fast analysis of all quantification data in one project, making detection and validation of regulated proteins time consuming. One solution is performing the analysis of quantified peptides and proteins with Rover (18) since this will ease the analysis and validation of regulated proteins as it visualizes all needed information in context of the whole experiment. MaxQuant and the Mascot Distiller Quantitation Toolbox clearly performed best in calculating peptide and protein ratios. Mascot Distiller further calculates quality parameters associated with peptide quantification, supporting and informing the user on the quality of quantification. MaxQuant does not report these parameters and therefore its quantification process is more a black box. A major disadvantage of MaxQuant is that it can only quantify peptides analyzed by Thermo Fisher Scientific LTQ-Orbitraps or LTQ-FTs. Also, the SILAC method is required for isotopic labeling. This is in great contrast to the versatility of instruments and the numerous labeling methods supported by the Mascot Distiller Quantitation Toolbox. This versatility clearly comes with a price; Mascot Distiller and its Quantitation Toolbox plug-in are expensive compared to the MaxQuant software which is completely free of charge.

Acknowledgments

The authors acknowledge the support of research grants from the Fund for Scientific Research – Flanders (Belgium) (project number G.0077.06), the Concerted Research Actions (project BOF07/GOA/012) from the Ghent University, and the Inter University Attraction Poles (IUAP06). We further would like to thank Dr. Bart Ghesquière, Francis Impens, and Evy Timmerman for providing the proteomic data.

References

1. Vaudel, M., Sickmann, A., and Martens, L. (2009) Peptide and protein quantification: a map of the minefield, *Proteomics* **10**, 650–670.
2. Gevaert, K., Impens, F., Ghesquiere, B., Van Damme, P., Lambrechts, A., and Vandekerckhove, J. (2008) Stable isotopic labeling in proteomics, *Proteomics* **8**, 4873–4885.
3. Ong, S. E., Blagoev, B., Kratchmarova, I., Kristensen, D. B., Steen, H., Pandey, A., and Mann, M. (2002) Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics, *Mol Cell Proteomics* **1**, 376–386.
4. Mann, M. (2006) Functional and quantitative proteomics using SILAC, *Nat Rev Mol Cell Biol* **7**, 952–958.
5. Park, S. K., Venable, J. D., Xu, T., and Yates, J. R., 3rd. (2008) A quantitative analysis software tool for mass spectrometry-based proteomics, *Nat Methods* **5**, 319–322.
6. Mortensen, P., Gouw, J. W., Olsen, J. V., Ong, S. E., Rigbolt, K. T., Bunkenborg, J., Cox, J., Foster, L. J., Heck, A. J., Blagoev, B., Andersen, J. S., and Mann, M. (2009) MSQuant, an open source platform for mass spectrometry-based quantitative proteomics, *J Proteome Res* **9**, 393–403.
7. Cox, J., and Mann, M. (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification, *Nat Biotechnol* **26**, 1367–1372.
8. O'Donnell, K. A., Wentzel, E. A., Zeller, K. I., Dang, C. V., and Mendell, J. T. (2005) c-Myc-regulated microRNAs modulate E2F1 expression, *Nature* **435**, 839–843.
9. Ghesquiere, B., Colaert, N., Helsens, K., Dejager, L., Vanhaute, C., Verleysen, K., Kas, K., Timmerman, E., Goethals, M., Libert, C., Vandekerckhove, J., and Gevaert, K. (2009) In vitro and in vivo protein-bound tyrosine nitration characterized by diagonal chromatography, *Mol Cell Proteomics* **8**, 2642–2652.
10. Gevaert, K., Van Damme, J., Goethals, M., Thomas, G. R., Hoorelbeke, B., Demol, H., Martens, L., Puype, M., Staes, A., and Vandekerckhove, J. (2002) Chromatographic isolation of methionine-containing peptides for gel-free proteome analysis: identification of more than 800 Escherichia coli proteins, *Mol Cell Proteomics* **1**, 896–903.
11. Baek, D., Villen, J., Shin, C., Camargo, F. D., Gygi, S. P., and Bartel, D. P. (2008) The impact of microRNAs on protein output, *Nature* **455**, 64–71.
12. Selbach, M., Schwanhausser, B., Thierfelder, N., Fang, Z., Khanin, R., and Rajewsky, N. (2008) Widespread changes in protein synthesis induced by microRNAs, *Nature* **455**, 58–63.
13. Tabb, D. L., McDonald, W. H., and Yates, J. R., 3rd. (2002) DTASelect and Contrast: tools for assembling and comparing protein identifications from shotgun proteomics, *J Proteome Res* **1**, 21–26.
14. Cox, J., Matic, I., Hilger, M., Nagaraj, N., Selbach, M., Olsen, J. V., and Mann, M. (2009) A practical guide to the MaxQuant computational platform for SILAC-based quantitative proteomics, *Nat Protoc* **4**, 698–705.
15. McDonald, W. H., Tabb, D. L., Sadygov, R. G., MacCoss, M. J., Venable, J., Graumann, J., Johnson, J. R., Cociorva, D., and Yates, J. R., 3rd. (2004) MS1, MS2, and SQT-three unified, compact, and easily parsed file formats for the storage of shotgun proteomic spectra and identifications, *Rapid Commun Mass Spectrom* **18**, 2162–2168.
16. Pedrioli, P. G., Eng, J. K., Hubley, R., Vogelzang, M., Deutsch, E. W., Raught, B., Pratt, B., Nilsson, E., Angeletti, R. H.,

Apweiler, R., Cheung, K., Costello, C. E., Hermjakob, H., Huang, S., Julian, R. K., Kapp, E., McComb, M. E., Oliver, S. G., Omenn, G., Paton, N. W., Simpson, R., Smith, R., Taylor, C. F., Zhu, W., and Aebersold, R. (2004) A common open representation of mass spectrometry data and its application to proteomics research, *Nat Biotechnol* **22**, 1459–1466.

17. Nesvizhskii, A. I., and Aebersold, R. (2005) Interpretation of shotgun proteomic data: the protein inference problem, *Mol Cell Proteomics* **4**, 1419–1440.
18. Colaert, N., Helsens, K., Impens, F., Vandekerckhove, J., and Gevaert, K. (2010) Rover: a tool to visualize and validate quantitative proteomics data from different sources, *Proteomics* **10**, 1226–1229.
19. Helsens, K., Timmerman, E., Vandekerckhove, J., Gevaert, K., and Martens, L. (2008) Peptizer, a tool for assessing false positive peptide identifications and manually validating selected results, *Mol Cell Proteomics* **7**, 2364–2372.
20. Olsen, J. V., de Godoy, L. M., Li, G., Macek, B., Mortensen, P., Pesch, R., Makarov, A., Lange, O., Horning, S., and Mann, M. (2005) Parts per million mass accuracy on an Orbitrap mass spectrometer via lock mass injection into a C-trap, *Mol Cell Proteomics* **4**, 2010–2021.

Subject Index

A

Acylation 80
Affinity capture 300
Affinity chromatography 11–12, 81, 105, 111, 188–189, 244, 291–292, 294–295, 300, 311, 313
Amidation 81–85, 87–89, 305, 307–308
Azidohomoalanine (azhal) 169–180

B

Bioinformatics
- automated protein quantification 58
- census 374–376, 378
- ICPLQuant 57–58, 60
- Mascot Distiller 374–375, 377, 379–384, 386–396
- MaxQuant 374–381, 383–384, 388–392, 396
- MsQuant 374–376, 379–381, 383–384, 388–396
- peptide identification 17–20, 35, 48, 51, 55, 145–146, 193–194, 200, 248, 363, 365, 367, 377, 379–380, 382–384, 392–394, 396
- peptide quantification 373–396
- protein identification 10, 17, 30, 38, 40, 66, 71–74, 79, 139, 145, 158, 179, 192, 195, 230–231, 244, 366, 374
- protein quantification 15, 63, 94–96, 108, 113, 126, 276, 285, 366–368, 379, 392, 396
- protein sequence databases 3, 17–20, 146, 165, 185, 202, 361, 363
- Rover 367–368, 381–382, 391–392, 396

Biotinylation 237, 239–240, 265, 267

C

Caenorhabditis elegans 30–31, 33–34, 36–37, 39
Carbohydrate specificity 290
Cell fractionation 118
Cell-free protein expression 103–104
Cell signaling 184
Chemical protein deglycosylation 323–332
Combined fractional diagonal chromatography (COFRADIC) 11–12, 172–173, 176, 179, 233, 276, 374

D

Drosophila melanogaster 34, 40, 188, 194

E

Electron transfer dissociation (ETD) 8, 83, 158–160, 165, 198–200, 302, 326
Endocytosis 120
Endoproteinase Lys-N 200
Escherichia coli 30–31, 33–34, 36, 39, 97, 99–100, 102–104, 170, 172–173, 177, 237–238, 244, 246, 254

F

Fourier transform mass spectrometry (FTMS) 6, 79, 82–84, 224, 226

G

GeLCMS 10, 143–154

I

Immobilized metal ion affinity chromatography (IMAC) 12, 81, 105–107, 111, 113, 119, 126, 186–189, 191–196, 198, 203–204

L

Lectin chromatography 12, 291–292, 294–295
Lipid-based protein immobilization (LPI) 129–141

M

Membrane proteins 127, 129–141, 158, 170, 188–189, 203, 294, 311–314, 318, 320

N

N-terminal peptide isolation 231–232
N-terminomics 243–254, 276

P

Peptide library 258–259, 263, 265, 268
Phagocytosis 118–120
Phosphopeptide sequencing 198, 200
Platelets 274, 300–305, 310
Positional proteomics 229–240
Protease
- active site specificity 257
- cleavage site 258–259, 268
- degradomics 276
- protease-mediated cellular signaling 184
- protein processing 274
- specificity 257–271
- subsite cooperativity 258, 269, 271
- substrate profiling 257–271

Protein digestion
- protein in-gel digestion 135–136
- protein in-solution digestion 135
- sequential protein digestion 130

Protein modification
- protein acetylation 71

K. Gevaert, J. Vandekerckhove (eds.), *Gel-Free Proteomics*, Methods in Molecular Biology 753,
DOI 10.1007/978-1-61779-148-2, © Springer Science+Business Media, LLC 2011

protein degradation 108, 131, 294, 332
protein glycosylation 309
protein O-glycosylation 299–300
protein phosphorylation 12, 95–96, 183, 196, 203, 215
protein processing 10, 12, 274
protein sialylation 310
protein ubiquitination 336
Protein quantification
absolute quantification 15, 95, 108–110, 204, 367
chemical labeling 65, 204
dimethylation 81, 258–259, 268, 275–276, 278, 283–285
enzymatic labeling 15, 373
ICPLQuant 57–58, 60, 62
Isobaric peptide termini labeling (IPTL) 65–75
Isobaric tag for relative and absolute quantitation (iTRAQ) 15, 30, 56, 65, 80, 171–172, 174–179, 180, 201, 204, 319
IsobariQ 67, 71–75
Isotope coded protein labels (ICPL) 55–63, 80, 319
Nitrogen-15 labeling 13–14
post-metabolic oxygen-18 labeling 15
Protein standard absolute quantification (PSAQ) 95–96, 99, 104, 108–111
quantitative proteomics 9, 13–16, 29, 37, 43–53, 56, 79, 81, 93–113, 368, 373
stable isotope labeling 13, 43, 56, 95, 104, 111, 144, 204
Tandem mass tags (TMT) 65, 319
Proteome simplification 229–240
Pulse-chase labeling 340–341, 348–349

R
Reticulocyte lysate 337–339, 341–342, 344, 350–353

S
Strong cation exchange chromatography (SCX) 4, 10, 12, 158–166, 171, 175, 179, 186–187, 193–196, 299–308

T
Titanium dioxide 12, 215–227, 300, 309–320
Translational regulation 169
Two-dimensional chromatography 224

U
Ubiquitin 80, 335–336, 339, 344–347, 350–355
Ubiquitin aldehyde 339, 354