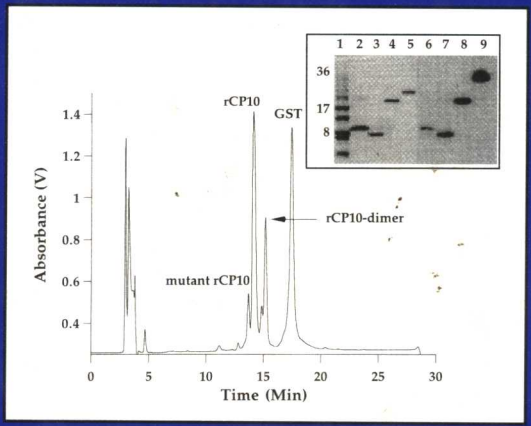
Methods in Molecular Biology™

VOLUME 146

Mass Spectrometry of Proteins and Peptides

Edited by

John R. Chapman



HUMANA PRESS

METHODS IN MOLECULAR BIOLOGY™

Mass Spectrometry of Proteins and Peptides

Edited by

John R. Chapman

Sale, Manchester, UK

Humana Press  **Totowa, New Jersey**

999 Riverview Drive, Suite 208
Totowa, New Jersey 07512

This publication is printed on acid-free paper. ∞
ANSI Z39.48-1984 (American Standards Institute) Permanence of Paper for Printed Library Materials.

Cover design by Patricia F. Cleary.

Cover art from: Figure 2, Chapter 3, *Characterization of a Mutant Recombinant S100 Protein Using Electrospray Ionization Mass Spectrometry*, by M. J. Raftery.

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

Printed in the United States of America. 10 9 8 7 6 5 4 3 2

Library of Congress Cataloging-in-Publication Data

Mass Spectrometry of Proteins and Peptides / edited by John R. Chapman.
p. cm. -- (Methods in molecular biology ; 146)
Includes bibliographical references and index.
ISBN 0-89603-609-x (alk. paper)
1. Proteins--Analysis. 2. Peptides--Analysis. 3. Peptide sequence. 4. Mass spectrometry. I. Chapman, J. R. II. Methods in molecular biology (Totowa, N.J.) ; v. 146.

QP551.P69565 2000
572'.633--dc21 99-41621

Preface

Little more than three years down the line and I am already writing the Preface to a second volume to follow *Protein and Peptide Analysis by Mass Spectrometry*. What has happened in between these times to make this second venture worthwhile?

New types of mass spectrometric instrumentation have appeared so that new techniques have become possible and existing techniques have become much more feasible. More particularly, however, the newer ionization techniques, introduced for the analysis of high molecular weight materials, have now been thoroughly used and studied. As a result, there has been an enormous improvement in the associated sample handling technology so that these methods are now routinely applied to much smaller sample amounts as well as to more intractable samples. Again, this particular community of mass spectrometry users has both increased in number and diversified. And, riding this wave of acceptance, leaders in the field have set their sights on more complex problems: molecular interaction, ion structures, quantitation, and kinetics are just a few of the newer areas reported in *Mass Spectrometry of Proteins and Peptides*.

As with the first volume, one purpose of this collection, *Mass Spectrometry of Proteins and Peptides*, is to show the reader what can be done by the application of mass spectrometry, and perhaps even to encourage the reader to venture down new paths. More important, another purpose is to demonstrate how these analyses are carried out in practice by guiding the reader, in a step-by-step manner, through the pitfalls and nuances of apparently straightforward techniques. It is the earnest hope of the editor that the reader, as a user of these techniques, will profit from the concise details that each of the authors has striven to provide. To parody Dr. Johnson: "what is written with effort is, in general, read with pleasure."

It has been suggested that "Writers, like teeth, are divided into incisors and grinders."* So, to spirit you from the Preface to the pleasures of the book, away with the preface grinder and on with the incisors and their chapters.

John R. Chapman

*Walter Bagehot, *Estimates of Some Englishmen and Scotchmen*.

Contributors

JEAN-FRANÇOIS ARRIGHI • *Department of Medical Biochemistry, University Medical Centre, Geneva, Switzerland*
AYEDA AYED • *Department of Chemistry, University of Manitoba, Winnipeg, Canada*
MARCUS BANTSCHEFF • *Faculty of Chemistry, University of Konstanz, Konstanz, Germany*
STEPHANIE BARBIRZ • *Faculty of Chemistry, University of Konstanz, Konstanz, Germany*
KEIRYN L. BENNETT • *Protein Research Group, Department of Molecular Biology, Odense University, Odense, Denmark*
RODERICK S. BLACK • *Athletic Drug Testing and Toxicology Laboratory, Department of Pathology and Laboratory Medicine, Indiana University School of Medicine, Indianapolis, IN*
K. OLAF BÖRNSEN • *Drug Metabolism and Pharmacokinetics, Novartis Pharma AG, Basel, Switzerland*
JEAN-CLAUDE BOULAIN • *Département d'Ingénierie et d'Etude des Protéines, CEA-Saclay, France*
LARRY D. BOWERS • *Athletic Drug Testing and Toxicology Laboratory, Department of Pathology and Laboratory Medicine, Indiana University School of Medicine, Indianapolis, IN*
EMILIO CAMAFEITA • *Unidad de Análisis Estructural de Proteínas, Centro Nacional de Biotecnología, Cantoblanco, Madrid, Spain*
FABRICE CERINI • *Department of Medical Biochemistry, University Medical Centre, Geneva, Switzerland*
BRIAN T. CHAIT • *Laboratory for Mass Spectrometry, The Rockefeller University, New York, NY*
JOHN R. CHAPMAN • *Sale, Manchester, UK*
IGOR CHERNUSHEVICH • *PE Sciex, Concord, Canada*
HANS-ULRICH DEMUTH • *Probiodrug Research, Halle, Germany*
LYNDA J. DONALD • *Department of Chemistry, University of Manitoba, Winnipeg, Canada*
HARRY W. DUCKWORTH • *Department of Chemistry, University of Manitoba, Winnipeg, Canada*

JOHN R. ENGEN • *Department of Chemistry, University of Nebraska-Lincoln, Lincoln, NE*
WERNER ENS • *Department of Physics, University of Manitoba, Winnipeg, Canada*
CATHERINE FENSELAU • *Department of Chemistry and Biochemistry, University of Maryland, College Park, MD*
PASQUALE FERRANTI • *International Mass Spectrometry Facility Centre and Mass Spectrometry Network-CNR, Naples, Italy*
ROGER GENET • *Département d'Ingénierie et d'Etude des Protéines, CEA-Saclay, France*
M. ERIC GERSHWIN, MD • *Division of Rheumatology, Allergy, and Clinical Immunology, School of Medicine, University of California, Davis, CA*
CÉCILE GILLARD-FACTOR • *Département des Procédés d'Enrichissement, DCC/DPE/SPCP/LASO, CEA-Saclay, France*
NICOLAS GILLES • *Département d'Ingénierie et d'Etude des Protéines, CEA-Saclay, France*
MICHAEL O. GLOCKER • *Faculty of Chemistry, University of Konstanz, Konstanz, Germany*
H. PETER HAPPERSBERGER • *Faculty of Chemistry, University of Konstanz, Konstanz, Germany*
KHAN HOANG-VAN • *Department of Medical Biochemistry, University Medical Centre, Geneva, Switzerland*
TORSTEN HOFFMANN • *Probiodrug Research, Halle, Germany*
RICKY D. HOLLAND • *Division of Chemistry, National Center for Toxicological Research, Food and Drug Administration, Jefferson, AR*
IAN HUMPHERY-SMITH • *Department of Pharmaceutical Proteomics, Universiteit Utrecht, Utrecht, The Netherlands*
RICHARD S. JOHNSON • *Immunex Corporation, Seattle, WA*
IGOR A. KALTASHOV • *Polymer Science Department, University of Massachusetts, Amherst, MA*
HAYDAR KARAOGLU • *Department of Microbiology, University of Sydney, Sydney, Australia*
RICHARD W. KRIWACKI • *Department of Structural Biology, St. Jude Children's Research Hospital, Memphis, TN*
THOMAS N. KROGH • *Department of Molecular Biology, Odense University, Odense, Denmark*
ANDREW N. KRUTCHINSKY • *Department of Physics, University of Manitoba, Winnipeg, Canada*
MARTIN KUSSMANN • *Cerbios-Pharma SA, Barbengo, Switzerland*
JACKSON O. LAY, JR. • *Division of Chemistry, National Center for Toxicological Research, Food and Drug Administration, Jefferson, AR*

AIQUN LI • *Department of Chemistry and Biochemistry, University of Maryland, College Park, MD*
ANDREW J. LINK • *Department of Microbiology and Immunology, Vanderbilt University Medical Center, Nashville, TN*
ANTONIO MALORNI • *International Mass Spectrometry Facility Centre and Mass Spectrometry Network-CNR, Naples, Italy*
GIANFRANCO MAMONE • *International Mass Spectrometry Facility Centre and Mass Spectrometry Network-CNR, Naples, Italy*
MATTHIAS MANN • *Protein Interaction Laboratory, University of Southern Denmark, Odense, Denmark*
JAN MARKUSSEN • *Novo Nordisk, Bagsvaerd, Denmark*
THOMAS MATTHIESEN • *Protein Research Group, Department of Molecular Biology, Odense University, Odense, Denmark*
LOUIS MAZZARELLI • *Laboratory for Mass Spectrometry, The Rockefeller University, New York, NY*
CHRISTOPHER H. S. MCINTOSH • *Department of Physiology, University of British Columbia, Vancouver, Canada*
DIETRICH MEBS • *Zentrum der Rechtsmedizin, Frankfurt, Germany*
ENRIQUE MÉNDEZ • *Unidad de Análisis Estructural de Proteínas, Centro Nacional de Biotecnología, Cantoblanco, Madrid, Spain*
EKATERINA MIRGORODSKAYA • *Department of Molecular Biology, Odense University, Odense, Denmark*
RUSSELL P. NEWTON • *Biochemistry Group, School of Biological Sciences, University of Wales, Swansea, UK*
ROBIN E. OFFORD • *Department of Medical Biochemistry, University Medical Centre, Geneva, Switzerland*
PIERRE-ALAIN PANCHAUD • *Reptiles du Monde, Grandvaux, Switzerland*
CAROL E. PARKER • *Laboratory of Structural Biology, National Institute of Environmental Health Sciences, Research Triangle Park, NC*
ROBERT P. PAULY • *Department of Physiology, University of British Columbia, Vancouver, Canada*
RAYMOND A. PEDERSON • *Department of Physiology, University of British Columbia, Vancouver, Canada*
MARK J. RAFTERY • *Cytokine Research Unit, School of Pathology, University of New South Wales, Kensington, Australia*
PETER ROEPSTORFF • *Department of Molecular Biology, Odense University, Odense, Denmark*
FRED ROSCHE • *Probiodrug Research, Halle, Germany*
KEITH ROSE • *Department of Medical Biochemistry, University Medical Centre, Geneva, Switzerland*

JOHANN SCHALLER • *Department für Chemie und Biochemie, Universität Bern, Freiestrasse, Bern, Switzerland*
DAVID SCHIELTZ • *Department of Molecular Biotechnology, University of Washington, Seattle, WA*
JÖRN SCHMIDT • *Probiodrug Research, Halle, Germany*
ANDREJ SHEVCHENKO • *European Molecular Biology Laboratory, Heidelberg, Germany*
GARY SIUZDAK • *Beckman Center for Chemical Sciences, The Scripps Research Institute, La Jolla, CA*
DAVID L. SMITH • *Department of Chemistry, University of Nebraska-Lincoln, Lincoln, NE*
KENNETH G. STANDING • *Department of Physics, University of Manitoba, Winnipeg, Canada*
RETO STÖCKLIN • *Atheris Laboratories, Research and Development, Geneva, Switzerland*
ZOLTÁN SZILÁGYI • *Institute of Chemistry, Chemical Research Center, Hungarian Academy of Sciences, Budapest, Hungary*
J. ALEX TAYLOR • *Immunex Corporation, Seattle, WA*
KENNETH B. TOMER • *Laboratory of Structural Biology, National Institute of Environmental Health Sciences, Research Triangle Park, NC*
SACHA N. ULJON • *Laboratory for Mass Spectometry, The Rockefeller University, New York, NY*
ISRAEL VALDÉS • *Unidad de Análisis Estructural de Proteínas, Centro Nacional de Biotecnología, Madrid, Spain*
KÁROLY VÉKEY • *Institute of Chemistry, Chemical Research Center, Hungarian Academy of Sciences, Budapest, Hungary*
HENRI VIRELIZIER • *CEA Saclay, SPEA-SAIS, Cedex, France*
DAVID J. VOCADLO • *Department of Chemistry, University of British Columbia, Vancouver, Canada*
LAN VU • *Department of Medical Biochemistry, University Medical Centre, Geneva, Switzerland*
RONG WANG • *Laboratory for Mass Spectometry, The Rockefeller University, New York, NY*
JUDY VAN DE WATER • *Division of Rheumatology, Allergy and Clinical Immunology, School of Medicine, University of California, Davis, CA*
MATTHIAS WILM • *European Molecular Biology Laboratory, Heidelberg, Germany*
STEPHEN G. WITHERS • *Department of Chemistry, University of British Columbia, Vancouver, Canada*
JOHN R. YATES, III • *Department of Molecular Biotechnology, University of Washington, Seattle, WA*

Contents

1

De Novo Peptide Sequencing by Nanoelectrospray Tandem Mass Spectrometry Using Triple Quadrupole and Quadrupole/Time-of-Flight Instruments

Andrej Shevchenko, Igor Chernushevich, Matthias Wilm, and Matthias Mann

1. Introduction

Recent developments in technology and instrumentation have made mass spectrometry the method of choice for the identification of gel-separated proteins using rapidly growing sequence databases *(1)*. Proteins with a full-length sequence present in a database can be identified with high certainty and high throughput using the accurate masses obtained by matrix-assisted laser desorption/ionization (MALDI) mass spectrometry peptide mapping *(2)*. Simple protein mixtures can also be deciphered by MALDI peptide mapping *(3)* and the entire identification process, starting from in-gel digestion *(4)* and finishing with acquisition of mass spectra and database search, can be automated *(5)*. Only 1–3% of a total digest are consumed for MALDI analysis even if the protein of interest is present on a gel in a subpicomole amount. If no conclusive identification is achieved by MALDI peptide mapping, the remaining protein digest can be analyzed by nanoelectrospray tandem mass spectrometry (Nano ESI-MS/MS) *(6)*. Nano ESI-MS/MS produces data that allow highly specific database searches so that proteins that are only partially present in a database, or relevant clones in an EST database, can be identified *(7)*. It is important to point out that there is no need to determine the complete sequence of peptides in order to search a database—a short sequence stretch consisting of three to four amino acid residues provides enough search specificity when combined with the mass of the intact peptide and the masses of corresponding fragment

From: *Methods in Molecular Biology, vol. 146:*
Protein and Peptide Analysis: New Mass Spectrometric Applications
Edited by: J. R. Chapman

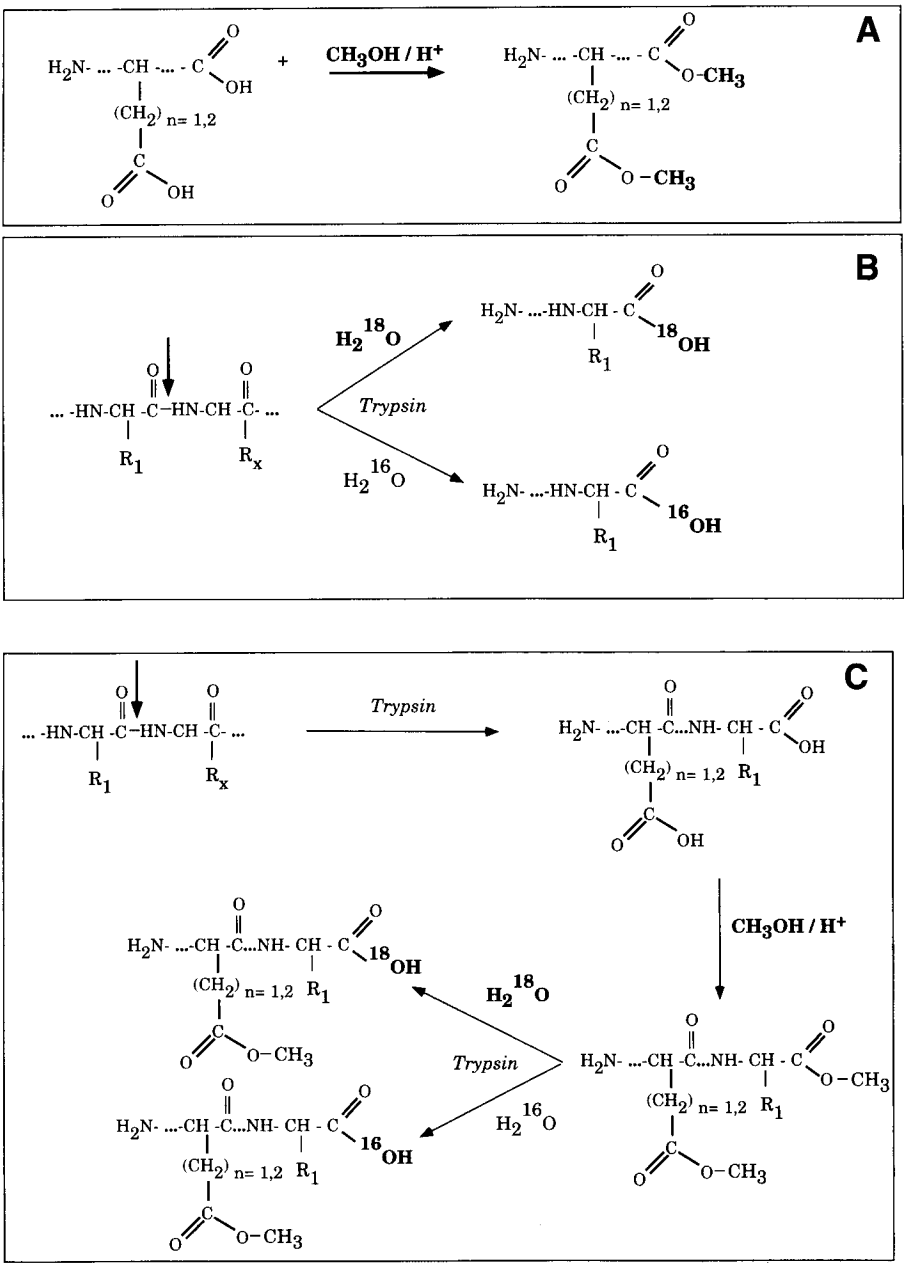
ions in a peptide sequence tag *(8)* (*see* **Subheading 3.4.**). Furthermore, proteins not present in a database that are, however, strongly homologous to a known protein can be identified by an error-tolerant search *(9)*.

Despite the success of ongoing genomic sequencing projects, the demand for *de novo* peptide sequencing has not been eliminated. Long and accurate peptide sequences are required for protein identification by homology search and for the cloning of new genes. Degenerate oligonucleotide probes are designed on the basis of peptide sequences obtained in this way, and subsequently used in polymerase chain reaction-based cloning strategies.

The presence of a continuous series of mass spectrometric fragment ions containing the C terminus (y″ ions) *(10)* has been successfully used to determine *de novo* sequences using fragment ion spectra of peptides from a tryptic digest *(11)*. The peptide sequence can be deduced by considering precise mass differences between adjacent y″ ions. However, it is necessary to obtain additional evidence that the particular fragment ion does indeed belong to the y″ series. To this end, a separate portion of the unseparated digest is esterified using 2 *M* HCl in anhydrous methanol (**Fig. 1A**) (*see* **Subheading 3.2.**). Upon esterification, a methyl group is attached to the C-terminal carboxyl group of each peptide, as well as to the carboxyl group in the side chain of aspartic and glutamic acid residues. Therefore the *m/z* value of each peptide ion is shifted by $14(n + 1)/z$, where n is the number of aspartic and glutamic acid residues in the peptide, and z is the charge of the peptide ion. The derivatized digest is then also analyzed by Nano ESI-MS/MS, and, for each peptide, fragment ion spectra acquired from underivatized and derivatized forms are matched. An accurate peptide sequence is determined by software-assisted comparison of these two fragment spectra by considering precise mass differences between the adjacent y″ ions as well as characteristic mass shifts induced by esterification (*see* **Subheading 3.4.1.**) (**Fig. 2**). Since esterification with methanol significantly shifts the masses of y″ ions (by 14, 28, 42, ... mass units), it is possible to use low-resolution settings when sequencing is performed on a triple quadrupole mass spectrometer, thus attaining high sensitivity on the instrument. This sequencing approach employing esterification is laborious and time consuming and requires much expertise in the interpretation of tandem mass spec-

---

Fig. 1. Chemical derivatization for mass spectrometric *de novo* sequencing of peptides recovered from digests of gel separated proteins. **(A)** A protein is digested in-gel (*see* **Subheading 3.1.**) with trypsin and a portion of the unseparated digest is esterified by 2 *M* HCl in anhydrous methanol (*see* **Subheading 3.2.**). **(B)** A protein is digested in-gel with trypsin in a buffer containing 50% (v/v) $H_2{}^{18}O$ and 50% (v/v) $H_2{}^{16}O$ (*see* **Subheading 3.1.**). **(C)** A protein is digested in-gel with trypsin, and the digest is esterified and subsequently treated with trypsin in the buffer containing 50% (v/v)



$H_2{}^{18}O$ and 50% (v/v) $H_2{}^{16}O$ (*see* **Note 22**). Here, R_1 repesents the side chain of arginine or lysine amino acid residues (these are trypsin cleavage sites) whereas R_x represents the side chain of any other amino acid residue except for proline.

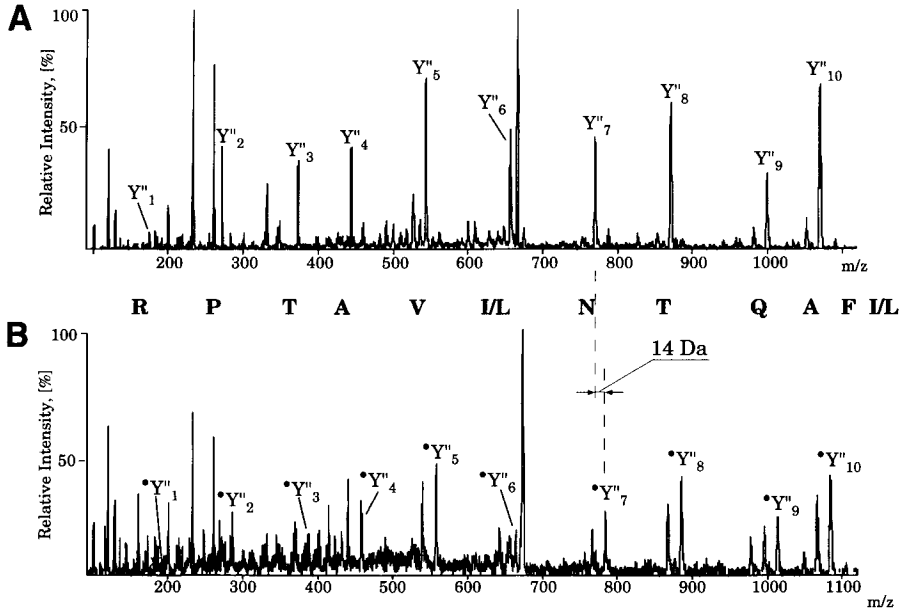


Fig. 2. Peptide *de novo* sequencing by comparison of tandem mass spectra acquired from intact and esterified peptide. A 120-kDa protein from *E. aediculatis* was purified by one-dimensional gel electrophoresis ***(24)*** and digested in-gel with trypsin; a part of the digest was analyzed by Nano ESI-MS/MS on an API III triple quadrupole mass spectrometer (PE Sciex, Ontario, Canada). A separate part of the digest was esterified and then also analyzed by Nano ESI-MS/MS. **(A)** Tandem (fragment-ion) mass spectrum recorded from the doubly charged ion with *m/z* 666.0 observed in the conventional (Q1) spectrum of the original digest. **(B)** Matching tandem spectrum acquired from the ion with *m/z* 673.0 (Δ mass = [673–666] × 2 = 14) in the conventional (Q1) spectrum of the esterified digest. The peptide sequence was determined by software-assisted comparison of spectra **A** and **B**. The only methyl group was attached to the C-terminal carboxyl of the peptide (designated by a filled circle) and therefore the masses of the singly charged y″ ions in spectrum **B** are shifted by 14 mass units compared with the corresponding y″ ions in spectrum **A**.

tra. However, it allows the determination of accurate peptide sequences even from protein spots that can only be visualized by staining with silver ***(12,13)***.

An alternative approach to *de novo* sequencing became feasible after a novel type of mass spectrometer—a hybrid quadrupole/time-of-flight instrument (Q/TOF ***[14]*** or QqTOF ***[15]***) was introduced. QqTOF instruments allow the acquisition of tandem mass spectra with very high mass resolution (>8000 full-width at half-maximum height [FWHM]) without compromising sensitivity. These instruments also benefit from the use of a nonscanning TOF analyzer that

records all ions simultaneously in both conventional and MS/MS modes and therefore increases sensitivity. These features make it possible and practical to apply selective isotopic labeling of the peptide C-terminal carboxyl group in order to distinguish y″ ions from other fragment ions in tandem mass spectra (*see* **Subheading 3.4.2.**). Proteins are digested with trypsin in a buffer containing 50% $H_2{}^{16}O$ and 50% $H_2{}^{18}O$ (v/v) (*see* **Subheading 3.1.**) so that half of the resulting tryptic peptide molecules incorporate ^{18}O atoms in their C-terminal carboxyl group, whereas the other half incorporate ^{16}O atoms (**Fig. 1B**). During subsequent sequencing by MS/MS, the entire isotopic cluster of each peptide ion, in turn, is selected by the quadrupole mass filter (Q) and fragmented in the collision cell *(9)*. Since only the fragments containing the C-terminal carboxyl group of the peptide appear to be partially (50%) isotopically labeled, y″ ions are distinguished by a characteristic isotopic pattern, viz. doublet peaks split by 2 mass units (*see* **Subheading 3.4.2.**) (**Fig. 3**); other fragment ions have a normal isotopic distribution. Thus, only a single analysis is required, peptide sequence readout is much faster and the approach lends itself to automation *(15)*.

2. Materials

For general instructions, *see* **Note 1**.

2.1. In-Gel Digestion

For contamination precautions, *see* **Note 2**.

1. 100 m*M* ammonium bicarbonate in water (high-performance liquid chromatography [HPLC] grade [LabScan, Dublin, Ireland]).
2. Acetonitrile (HPLC grade [LabScan]).
3. 10 m*M* dithiothreitol in 100 m*M* ammonium bicarbonate.
4. 55 m*M* iodoacetamide in 100 m*M* ammonium bicarbonate.
5. 100 m*M* $CaCl_2$ in water.
6. 15 µL aliquots of trypsin, unmodified, sequencing grade (Boerhringer Mannheim, Germany) in 1 m*M* HCl (*see* **Note 3**).
7. 5% (v/v) formic acid in water.
8. Heating blocks at 56°C and at 37°C.
9. Ice bucket.
10. Laminar flow hood (optional) (*see* **Note 2**).

2.2. Esterification with Methanol

1. Methanol (HPLC grade), distilled shortly before the derivatization process.
2. Acetyl chloride (reagent grade), distilled shortly before the derivatization (*see* **Note 4**).

2.3. Isotopic Labeling Using $H_2{}^{18}O$

1. Reagents as in **Subheading 2.1.**
2. $H_2{}^{18}O$ (Cambridge Isotopic Laboratories, Cambridge, MA), distilled (*see* **Note 5**).

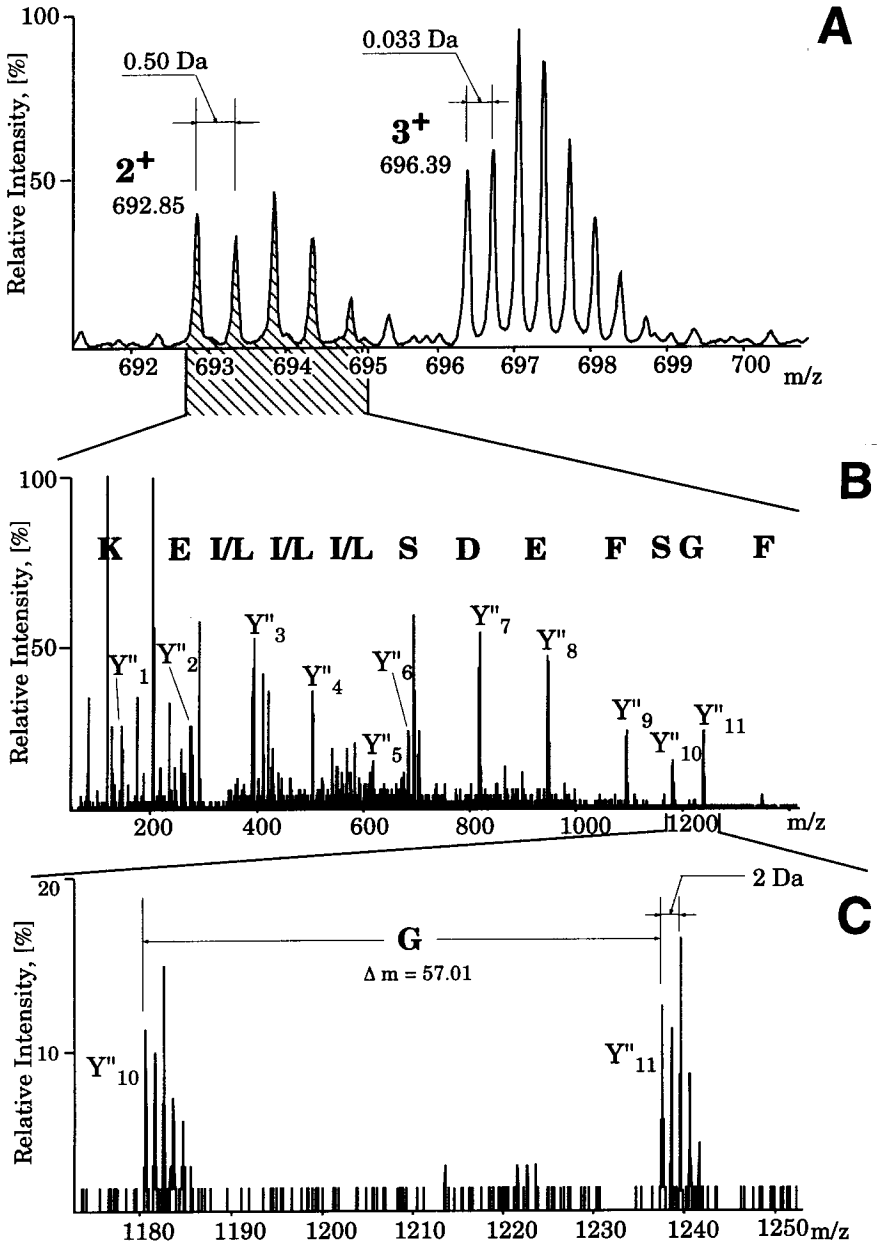


Fig. 3. Sequencing of ^{18}O C-terminally labeled tryptic peptides by Nano ESI-MS/MS. A 35-kDa protein from *Drosophila* was purified by gel electrophoresis, digested in-gel in a buffer containing 50% (v/v) $H_2{}^{18}O$, and analyzed using a QqTOF mass