蛋白质测序方法

Protein Sequencing Protocols

Second Edition

Bryan John Smith

Humana Press
Totowa,New Jersey

世界图书出版公司

—— 英文重印版 ——

蛋白质测序方法

Protein Sequencing Protocols

Second Edition

Edited by

Bryan John Smith

Celltech Chiroscience,

Slough, UK

Humana Press  **Totowa, New Jersey**

世界图书出版公司

西安 北京 广州 上海

陕版出图字：25－2003－043

图书在版编目（CIP）数据

蛋白测序方法/(美)史密斯(Smith,B.J.)著.2版. —西安：世界图书出版西安公司，2004.6
ISBN 7－5062－5976－1

Ⅰ.蛋… Ⅱ.史… Ⅲ.蛋白质—序列—测试—方法—英文 Ⅳ.Q510.1

中国版本图书馆CIP数据核字（2004）第060121号

蛋白测序方法

策　　划　世图医学出版中心
责任编辑　任卫军
封面设计　赵华龙

出版发行　世界图书出版西安公司
地　　址　西安市南大街17号　　邮编　710001
电　　话　029－87279676　87233647(发行部)
传　　真　029－87279675
E－mail　wmcrxian@public.xa.sn.cn
经　　销　各地新华书店
印　　刷　世界图书出版西安公司印刷厂
开　　本　850×1168　大32K
印　　张　15
字　　数　300千字

版　　次　2004年8月第1次印刷
书　　号　ISBN 7－5062－5976－1
定　　价　206.00元

METHODS IN MOLECULAR BIOLOGY™

Protein Sequencing Protocols

Second Edition

Edited by

Bryan John Smith

Celltech Chiroscience,
Slough, UK

Humana Press Totowa, New Jersey

999 Riverview Drive, Suite 208
Totowa, New Jersey 07512

www.humanapress.com

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

Cover Illustration: Foreground: Diagramatic protein sequence. Illustration provided by Bryan John Smith. Background: PVDF Western Electroblot Transfer of 300 µg human heart proteins. See Fig. 1 on page 46. Illustration provided by Michael J. Dunn.

Cover design by Patricia F. Cleary.

Production Editor: Mark J. Breaugh.

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: http://humanapress.com

Printed in the United States of America. 10 9 8 7 6 5 4 3 2 1

Library of Congress Cataloging in Publication Data

Protein sequencing protocols / edited by Bryan John Smith.-- 2nd ed.
p. cm. -- (Methods in molecular biology ; v. 211)
Includes bibliographical references and index.
ISBN 0-89603-975-7 (alk. paper)
1. Nucleotide sequence--Laboratory manuals. I. Smith, Bryan John. II. Methods in molecular biology (Clifton, N.J.) ; v. 211.

QP625.N89 P76 2002
572.8'633-dc21

2002022740

Preface

Determination of the protein sequence is as important today as it was a half century ago, even though the techniques and purposes have changed over time. Mass spectrometry has continued its recent rapid development to find notable application in the characterization of small amounts of protein, for example, in the field of proteomics. The "traditional" chemical N-terminal sequencing is still of great value in quality assurance of the increasing number of biopharmaceuticals that are to be found in the clinic, checking processing events of recombinant proteins, and so on. It is joined in the armory of methods of protein analysis by such techniques as C-terminal sequencing and amino acid analysis. These methods are continually developing. The first edition of *Protein Sequencing Protocols* was a "snapshot" of methods in use in protein biochemistry laboratories at the time, and this, the second edition, is likewise. Methods have evolved in the intervening period, and the content of this book has similarly changed, the content of some chapters having been superceded and replaced by other approaches. Thus, in this edition, there is inclusion of approaches to validation of methods for quality assurance work, reflecting the current importance of biopharmaceuticals, and also a guide to further analysis of protein sequence information, acknowledging the importance of bioinformatics. Some of the areas are also the subjects of other volumes in the *Methods in Molecular Biology*™ series published by Humana, and the interested reader is directed to such volumes as **59** (*Protein Purification Protocols*), **76** (*Glycoanalysis*), **112** (*2-D Proteome Analysis Protocols*), **143** (*Protein Structure Prediction*), **146** (*Mass Spectrometry of Proteins and Peptides*), and **159** (*Amino Acid Analysis Protocols*).

The style of *Protein Sequencing Protocols* is like that of others in the series, that of a laboratory manual. The aim is to permit ready replication of the methods. It is recognized, however, that various of the techniques require expensive equipment. This can be a barrier to those without such facilities, but who might wish to collaborate with those who do have them. This volume is intended to indicate to those workers what is required by way of sample preparation, and what can be achieved by the techniques, and so be an aid to collaboration. Appendices are provided as handy references to the molecular weights of amino acids and their derivatives, and to consensus sequences.

I would like to thank all the authors for their expert contributions to this volume—their efforts have made the book what it is. I would also like to thank Stef, for her continued support.

Bryan John Smith

Contributors

ALASTAIR AITKEN • *Department of Biomedical Sciences, University of Edinburgh, Edinburgh, UK*
NEBOJSA AVDALOVIC • *Dionex Corporation, Sunnyvale, CA*
PEER BORK • *European Molecular Biology Laboratory, Heidelberg, Germany*
FRANCESCO BOSSA • *Dipartimento Scienze Biomediche, Università La Sapienza, Rome, Italy*
ALEX F. CARNE • *Institute of Cancer Research, London, UK*
JUN CHENG • *Dionex Corporation, Sunnyvale, CA*
STEVEN A. COHEN • *Waters Corporation, Milford, MA*
RICHARD G. COOK • *Protein Chemistry Core Lab, Baylor College of Medicine, Houston, TX*
RICHARD R. COPLEY • *EMBL, Heidelberg, Germany*
MARK W. CRANKSHAW • *Department of Molecular Biology and Pharmacology, Washington University School of Medicine, St. Louis, MO*
IAN DAVIDSON • *Department of Molecular & Cell Biology, University of Aberdeen, UK*
NEIL DODSWORTH • *Delta Biotechnology, Nottingham, UK*
BRYAN DUNBAR • *Department of Molecular & Cell Biology, University of Aberdeen, UK*
MICHAEL J. DUNN • *Institute of Psychiatry, Kings College London, London, UK*
DAVID R. DUPONT • *Applied Biosystems, Foster City, CA*
KENNETH S. GRAHAM • *Regeneron Pharmaceuticals, Tarrytown, NY*
GREGORY A. GRANT • *Department of Molecular Biology and Pharmacology, Washington University School of Medicine, St. Louis, MO*
FIONA M. GREER • *M-Scan Ltd, Wokingham, UK*
DAVID J. HARVEY • *Oxford Glycobiology Institute, Oxford University, UK*
HISASHI HIRANO • *Yokohama City University, Yokohama, Japan*
G. BRENT IRVINE • *Division of Biochemistry, School of Biology and Biochemistry, Medical Biology Centre, Queen's University of Belfast, Belfast, UK*
PHILIP J. JACKSON • *Applied Biosystems, Warrington, Cheshire, UK*
PETR JANDIK • *Dionex Corporation, Sunnyvale, CA*
ROZA M. KAMP • *Department of Biotechnology, University of Applied Science and Technology, Berlin, Germany*

MICHELE LEARMONTH • *Department of Biomedical Sciences, University of Edinburgh, UK*
ADAM LISKA • *MPI of Molecular Cell Biology and Genetics, Dresden, Germany*
HOWARD R. MORRIS • *M-Scan Ltd., Wokingham, UK; Department of Biological Sciences, Imperial College, London, UK*
JACEK MOZDZANOWSKI • *Analytical Methods Development, Glaxosmithkline Pharmaceuticals, King of Prussia, PA*
DARRYL J. C. PAPPIN • *Imperial College, University of London, London, UK*
ANDREW J. REASON • *M-Scan Ltd, Wokingham, UK*
ROBERT B. RUSSELL • *European Molecular Biology Laboratory, Heidelberg, Germany*
ANDREA SCALONI • *Proteomics and Mass Spectrometry Laboratory, IABBAM, National Research Council, Naples, Italy*
DAVID M. SCHIELTZ • *Torrey Mesa Research Institute, San Diego, CA*
ANDREJ SHEVCHENKO • *European Molecular Biology Laboratory, Heidelberg, Germany*
ANNA SHEVCHENKO • *MPI of Molecular Cell Biology and Genetics, Dresden, Germany*
MAURIZIO SIMMACO • *Dipartimento Scienze Biomediche, Università G. D'Annunzio, Chieti, Italy*
ALAN J. SMITH • *Stanford University Medical Center, Stanford, CA*
BRYAN JOHN SMITH • *Celltech Chiroscience, Slough, UK*
LAUREY STEINKE • *Protein Structure Core Facility, University of Nebraska Medical Center, Lincoln, NE*
SHAMIL SUNYAEV • *European Molecular Biology Laboratory, Heidelberg, Germany*
PAUL TEMPST • *Protein Center, Memorial Sloan-Kettering Cancer Center, New York, NY*
CHRISTOPH W. TURCK • *Max Planck Institute of Psychiatry, Molecular, Cellular, Clinical Proteomics, Munich, Germany*
JAMES P. TURNER • *Celltech Chiroscience, Slough, UK*
JOHN R. YATES, III • *Torrey Mesa Research Institute, San Diego, CA*
SYLVIA W. YUEN • *Applied Biosystems, Foster City, CA*

Contents

1

Strategies for Handling Polypeptides on a Micro-Scale

Bryan John Smith and Paul Tempst

1. Introduction

Samples for sequence analysis frequently are in far from plentiful supply. Preparation of protein without loss, contamination or modification becomes more problematical as the amount of the sample decreases. The most successful approach is likely to include the minimum number of steps, at any of which a problem might arise. The strategy for preparation of a given protein will depend on its own particular properties, but several points of advice apply. These are:

- Minimize sample loss: *see* **Note 1**.
- Minimize contamination of the sample: *see* **Note 2**.
- Minimize artificial modification of the sample: *see* **Note 3**.

When it comes to sample purification, polyacrylamide gel electrophoresis is a common method of choice, since it is suited to sub-μg amounts of sample, entails minimal sample handling, is quick, and has high resolving power. Proteins may be fragmented while in the gel (*see* Chapters 5 and 6), or electroeluted from it using commercially available equipment. Commonly, however, proteins and peptides are transferred onto membranes prior to analysis by various strategies as described in Chapter 4. Capillary electrophoresis (Chapter 8) and high-performance liquid chromatography (HPLC) are alternative separation techniques. Capillary electrophoresis has sufficient sensitivity to be useful for few μg or sub- μg amounts of sample. For maximum sensitivity on HPLC, columns of 1 mm or less inside diameter (id) may be used, but for doing so there are considerations extra to those that apply to use of larger-bore columns. These are discussed below.

From: *Methods in Molecular Biology, vol. 211: Protein Sequencing Protocols, 2nd ed.*
Edited by: B. J. Smith

Although desirable to minimize the amount of handling of a sample, it is frequently necessary to manipulate the sample prior to further purification or analysis, in order to concentrate the sample or to change the buffer, for instance. Some examples of methods for the handling of small samples follow below. They do not form an exhaustive list, but illustrate the type of approach that it may be necessary to adopt.

2. Materials

2.1. Microbore HPLC

1. An HPLC system able to operate at low flow rates (of the order of 30 µL/min) while giving a steady chromatogram baseline, with minimal mixing and dilution of sample peaks in the postcolumn plumbing (notably at the flow cell) and with minimal volume between flow cell and outflow (to minimize time delay, so to ease collection of sample peaks).

   An example design is described by Elicone et al *(1)*. These authors used a 140B Solvent Delivery System from Applied Biosystems. The system was equipped with a 75 µL dynamic mixer and a precolumn filter with a 0.5 µm frit (Upchurch Scientific, Oak Harbor, WA) was plumbed between the mixer and a Rheodyne 7125 injector (from Rainin, Ridgefield, NJ) using two pieces (0.007 inch ID, 27 cm long [1 in. = 2.54 cm]) of PEEK tubing. The injector was fitted with a 50 µL loop and connected to the column inlet with PEEK tubing (0.005 inch × 30 cm). The outlet of the column was connected directly to a glass capillary (280 µm OD/ 75 cm ID × 20 cm; 0.88 µL), which is the leading portion of an U-Z view flow cell (35 nL volume, 8-mm path length; LC Packings, San Francisco, CA), fitted into an Applied Biosystems 783 detector. The trailing portion of the capillary cell was trimmed to a 15 cm length and threaded out of the detector head, resulting in a post flow cell volume of 0.66 µL and a collection delay of 1.3 s (at a flow rate of 30 µL/ min). Alternatively, various HPLC systems suitable for microbore work are available from commercial sources.
2. Clean glassware, syringe, and tubes for collection (polypropylene, such as the 0.5 µL or 1.5 µL Eppendorf type).
3. Solvents: use only HPLC-grade reagents (Fisons or other supplier), including distilled water (commercial HPLC-grade or Milli-Q water). A typical solvent system would be an increasing gradient of acetonitrile in 0.1% (v/v) trifluoroacetic acid (TFA) in water. The TFA acts as an ion-pairing reagent, interacting with positive charges on the polypeptide and generally improving chromatography. If TFA is not added to the acetonitrile stock, the baseline will decrease (owing to decreasing overall content of TFA), which makes identification of sample peaks more difficult. A level baseline can be maintained by adding TFA to the acetonitrile stock, in sufficient concentration (usually about 0.09% v/v) to make its absorbency at 214 or 220 nm equal to that of the other gradient component, 0.1% TFA in water. Check this by spectrophotometry. The absorbency remains stable for days.

4. Microbore HPLC columns of internal diameter 2.1 mm, 1 mm or less, are available from various commercial sources.

2.2. Concentration and Desalting of Sample Solutions

1. HPLC system: not necessarily as described above for microbore HPLC, but capable of delivering a flow rate of a few hundred µL to 1 mL per min. Monitor elution at 220 nm or 214 nm.
2. Clean syringe, tubes, HPLC-grade solvents, and so on as described in **Subheadings 2.1., steps 2** and **3**.
3. Reverse-phase HPLC column, of alkyl chain length C2 or C4. Since analysis and resolution of mixtures of polypeptides is not the aim here, relatively cheap HPLC columns may be used (and reused). The method described employs the 2.1 mm ID × 10 mm C2 guard column. (Brownlee, from Applied Biosystems), available in cartridge format.

2.3. Small Scale Sample Clean-Up Using Reverse-Phase "Micro-tips"

1. Pipet tip: Eppendorf "gel loader" tip (cat. no. 2235165-6, Brinkman, Westbury, NY).
2. Glass fiber, such as the TFA-washed glass fibre disks used in Applied Biosystems automated protein sequencers (Applied Biosystems, cat. no. 499379).
3. Reverse-phase chromatography matrix, such as Poros 50 R2 (PerSeptive Biosystems, Framingham, MA). Make as a slurry in ethanol, 4:1::ethanol:beads (v/v).
4. Wash buffer: formic acid (0.1%, v/v in water). Elution buffer: acetonitrile in 0.1% formic acid, e.g., 30% acetonitrile (v/v).
5. Argon gas supply, at about 10–15 psi pressure, with line suited to attach to the pipet tip.
6. Micro-tubes: small volume, capped, e.g., 0.2 mL (United Scientific Products, San Leandro CA, Cat. no. PCR-02).

3. Methods

3.1. Microbore HPLC (see Notes 4–13)

3.1.1. Establishment of Baseline (see Notes 4 and 7)

A flat baseline at high-sensitivity setting (e.g., 15 mAUFs at 214 mm) is required for optimal peak detection. The use of an optimized HPLC and clean and UV absorbency-balanced solvents should generate a level baseline with little noise and peaks of contamination. A small degree of baseline noise originates from the UV detector. Beware that this may get worse as the detector lamp ages. Some baseline fluctuation may arise from the action of pumps and/or solvent mixer. Slow flow rates seem to accentuate such problems that can go unnoticed at higher flows. Thorough sparging of solvents by helium may

reduce these problems. New or recently unused columns require thorough washing before a reliable baseline is obtained. To do this, run several gradients and then run the starting solvent mixture until the baseline settles (this may take an hour or more). Such problems are reduced if the column is used continuously, and to achieve this in between runs, an isocratic mixture of solvents (e.g., 60% acetonitrile) may be run at low flow rate (e.g., 10 µL/min). Check system performance by running standard samples (e.g., a tryptic digest of 5 pmole of cytochrome C).

3.1.2. Identification of Sample Peaks (see **Notes 4, 7, and 8**)

1. Peaks that do not derive from the sample protein(s), may arise from other sample constituents, such as added buffers or enzymes. To identify these contaminants, run controls lacking sample protein. Once the sample has been injected, run the system isocratically in the starting solvent mixture until the baseline is level and has returned to its pre-inject position. This can take up to 1 h in case of peptide mixtures that have been reacted with UV-absorbing chemicals (4-vinyl pyridine for example) before chromatography.
2. Peaks may be large enough to permit on-line spectroscopy where a diode array is available. Some analysis of amino acid content by second derivative spectroscopy may then be undertaken, identifying tryptophan-containing polypeptides, for instance, as described in Chapter 9.
3. Polypeptides containing tryptophan, tyrosine, or pyridylethylcysteine may be identified by monitoring elution at just three wavelengths (253, 277, 297 nm) in addition to 214 nm. Ratios of peak heights at these wavelengths indicate content of the polypeptides as described in **Note 8**. This approach can be used at the few pmole level.
4. Flow from the HPLC may be split and a small fraction diverted to an on-line electrospray mass spectrograph, so as to generate information on sample mass as well as possible identification of contaminants.

3.1.3. Peak Collection (see **Notes 4, 9–12**)

1. While programmable fraction collectors are available, peak collection is most reliably and flexibly done by hand. This operation is best done with detection of peaks on a flatbed chart recorder in real time. The use of flatbed chart recorder allows notation of collected fractions on the chart recording for future reference. The delay between peak detection and peak emergence at the outflow must be accurately known (*see* **Note 5**).
2. When the beginning of a peak is observed, remove the forming droplet with a paper tissue. Collect the outflow by touching the end of the outflow tubing against the side of the collection tube, so that the liquid flows continuously into the tube and drops are not formed. Typical volumes of collected peaks are 40–60 µL (from a 2.1 mm ID column) and 15–30 µL (from a 1 mm ID column). *See* **Note 9**.
3. Cap tubes to prevent evaporation of solvent. Store collected fractions for a short term on ice, and transfer to freezer (–20°C or –70°C) for long-term storage (*see* **Notes 10** and **11**).

4. Retrieval of sample following storage in polypropylene tubes is improved by acidification of the thawed sample, by addition of neat TFA to a final TFA concentration of 10% (v/v).

3.2. Concentration and Desalting of Sample Solutions (see Notes 14–24)

1. Equilibrate the C2 or C4 reverse-phase HPLC column in 1% acetonitrile (or other organic solvent of choice) in 0.1% TFA (v/v) in water, at a flow rate of 0.5 mL/min at ambient temperature.
2. Load the sample on to the column. If the sample is in organic solvent of concentration greater than 1% (v/v), dilute it with water or aqueous buffer (to ensure that the protein binds to the reverse-phase column) but do this just before loading (to minimize losses by adsorption from aqueous solution onto vessel walls). If the sample volume is greater than the HPLC loop size, simply repeat the loading process until the entire sample has been loaded.
3. Wash the column with isocratic 1% (v/v) acetonitrile in 0.1% TFA in water. Monitor elution of salts and/or other hydrophilic species that do not bind to the column. When absorbency at 220 nm has returned to baseline a gradient is applied to as to elute polypeptides from the column. The gradient is a simple, linear increase of acetonitrile content from the original 1% to 95%, flow rate 0.5mL/min, ambient temperature, over 20 min. Collect and store emerging peaks as described above (*see* **Subheading 3.1.2.** and *see* **Note 9**).
4. The column may be washed by isocratic 95% acetonitrile in 0.1% TFA in water, 0.5 mL/min, 5 min before being re-equilibrated to 1% acetonitrile for subsequent use.

3.3. Small Scale Sample Clean-up Using "Micro-tips" (see Notes 25–28).

1. Using a pipet tip, core out a small disk from the glass-fiber disk. Push it down the inside of the gel-loader tip (containing 20 µL of ethanol), until it is stuck. Pipet onto this frit 10 µL of reverse-phase matrix slurry (equivalent to about 2 µL of packed beads). Apply argon gas to the top of the tip, to force liquid through the tip and pack the beads. Wash the beads by applying 3 lots of 20 µL of 0.1% formic acid, forcing the liquid through the micro-column with argon, but never allowing the column to run dry. Use a magnifying glass to check this, if necessary. Leave about 5 mm of final wash above the micro-column. The column is ready to use.
2. Apply the sample solution to the micro-column and wash with 3 lots of 20 µL 0.1% formic acid, leaving a minimum of the final wash solution above the micro-column. Pipet 3–4 µL (i.e., about 2 column volumes) of elution buffer into the micro-tip, leaving a bubble of air between the elution buffer and the micro-column in ash buffer. The elution buffer is then forced into the micro-column (but without mixing with the wash buffer, for clearly, this would alter the composition of the buffer and possibly adversely affect elution). Collect the buffer containing the eluted sample. If further elution steps are required, do not let the