R.M. Kamp
T. Choli-Papadopoulou
B. Wittmann-Liebold
(Eds.)

Protein Structure Analysis

Preparation, Characterization, and Microsequencing

Springer
Lab Manual

R. M. Kamp T. Choli-Papadopoulou
B. Wittmann-Liebold (Eds.)

Protein Structure Analysis

Preparation, Characterization, and Microsequencing

With 64 Figures

Springer

Professor Dr. ROZA MARIA KAMP
Technische Fachhochschule Berlin
FB3 Chemie- und Biotechnik
Seestraße 64
13347 Berlin
Germany

Ass. Professor Dr. THEODORA CHOLI-PAPADOPOULOU
Aristotle University Thessaloniki
School of Biochemistry
Laboratory of Biochemistry
54006 Thessaloniki
Greece

Professor BRIGITTE WITTMANN-LIEBOLD
Max-Delbrück-Centrum
für Molekulare Medizin
Robert-Rössle-Straße 10
13125 Berlin-Buch
Germany

ISBN 3-540-61500-8 Springer-Verlag Berlin Heidelberg New York

Library of Cogress Cataloging-in-Pulication Data

Protein structure analysis : preparation, characterization, and
microsequencing / Roza Maria Kamp, Theodora Choli-Papadoupoulou,
Brigitte Wittmann-Liebold (eds.).
p. cm. -- (Springer lab manual)
Includes bibliographical references and index.
ISBN 3-540-61500-8 (alk. paper)
1. Proteins--Structure--Laboratory manuals. I. Kamp, Roza Maria,
1951- . II. Choli-Papadopoulou, Theodora, 1956- III. Wittmann
-Liebold, B. (Brigitte), 1931- IV. Series.
QP551.P697643 1997 547.7'5046--DC20 96-35790

Cover design: Design & Production GmbH, Heidelberg
Typesetting: M. Masson-Scheurer, Homburg, Saar
SPIN 10527054 31/3137 5 4 3 2 1 0 - Printed on acid-free papier

Preface

"Protein Structure Analysis – Preparation and Characterization" is a compilation of practical approaches to the structural analysis of proteins and peptides. Here, about 20 authors describe and comment on techniques for sensitive protein purification and analysis. These methods are used worldwide in biochemical and biotechnical research currently being carried out in pharmaceutical and biomedical laboratories or protein sequencing facilities. The chapters have been written by scientists with extensive experience in these fields, and the practical parts are well documented so that the reader should be able to easily reproduce the described techniques. The methods compiled in this book were demonstrated in student courses and in the EMBO Practical Course on "Microsequence Analysis of Proteins" held in Berlin September 10–15, 1995. The topics also derived from a FEBS Workshop, held in Halkidiki, Thessaloniki, Greece, in April, 1995. Most of the authors participated in these courses as lecturers and tutors and made these courses extremely lively and successful. Since polypeptides greatly vary depending on their specific structure and function, strategies for their structural analysis must for the most part be adapted to each individual protein. Therefore, advantages and limitations of the experimental approaches are discussed here critically, so that the reader becomes familiar with problems that might be encountered.

Young scientists who are newcomers to the field of protein chemistry may ask why it is necessary to learn all about purifying and characterizing individual proteins and peptides, processes which often are tricky and tedious. Cloning and sequencing the protein's gene is a commonly used and modern method to deduce the sequence of the amino acid chain; hence it may not be necessary to study the protein directly. We often hear the sentence: "we can do it all using molecular biology techniques." However, recent work on yeast and human genome projects and the subsequent discovery of many intron and exon sequences whose biological functions are unknown make the search for and

direct study of the protein itself mandatory. Moreover, the shift, 10–15 years ago, from protein studies to gene analysis left many open questions behind, which could be addressed by direct protein isolation and characterization. Young scientists interested in natural science research are strongly advised to learn as much as possible about protein and peptide structural analysis, so that problems encountered in their research can be approached by investigating the proteins involved.

We would like to present a compilation of commonly used protein analysis techniques, but we cannot include all the techniques currently in use. The variability in protein and peptide structure is enormous, ranging from very small polypeptide chains of just a few amino acids to large ones with more than 1000 residues, from very hydrophilic globular structures to largely hydrophobic membrane proteins which may be elongated. A single cell contains thousands of different proteins and peptides; indeed, in the human body the biomass of polypeptides is approximately 17%. These biopolymers are involved in all essential processes in the living cell, such as transport within the body fluids and transfer of molecules through membranes; signal transduction; transcription of the genetic information and its translation into amino acid sequences at the ribosome; and mediators of the immune response. Proteins (enzymes) and peptides (hormones) control and also regulate the biosynthetic pathways. Needless to say correlation of their structure and function is of great importance in understanding biological mechanisms at the molecular level.

Proteins are complicated structures, they form three-dimensional folds of the linear polypeptide chain in a manner which is specific and unique to each of the various proteins. The specificity is fully imprinted in the amino acid sequence; yet, how folding of the linear sequence into the higher order structure is accomplished is still one of the great puzzles in biochemistry. Firstly, the primary structure (amino acid sequence) has to be established; secondly, the secondary structural elements (α-helices, β-sheets, extended structures, turns or coil) must be correlated with the linear sequence, and thirdly, folding of the protein according to the specific higher order (tertiary) structure has to be studied. Finally, interactions with other polypeptide chains or other molecules (quarternary structure) must be identified. Although X-ray structural analysis of protein crystals and NMR studies on proteins in solution are powerful tools to determine tertiary protein structure, the basis for the correct interpretation of these results definitely is knowledge of the amino acid sequence. Any type of modification, such as phosphorylation,

methylation, or acetylation, adds to the character of the protein and alters its properties. Furthermore, many proteins are constituents of complexes composed of various proteins and RNA or DNA; they may be bound to lipids and glycosides. Therefore, while it is important to know the sequence and the tertiary structure, and to identify amino acid modifications and the presence of S-S-bonds, it is also necessary to establish whether ligands, cofactors and/or other molecules are complexed to the protein being studied. Finally, the binding sites and binding forces which lead to formation of the stable and fully active complex have to be evaluated. These are the reasons why it is not simple to determine the structure of a protein and to explain its functional role on the molecular level. Many complicated methods must be applied to study the protein's role(s) in the biological process.

The study of proteins includes the use of physical, physicochemical, chemical and biological methods. Although not all of these methods can be described here, we nonetheless hope that we have assembled a suitable collection of modern and sensitive techniques that are easy to perform for beginners and which will facilitate study of a protein's structure and function. The presented methods are well established and were found by experienced investigators to be useful. This book includes purification methods under native and denaturing conditions for structural and functional studies; it describes methods for generation and separation of complex protein mixtures in minute amounts by one- and two-dimensional polyacrylamide gels; it contains highly sensitive amino acid compositional analysis procedures as well as manually and automatically performed sequencing techniques. In addition, database searches are recommended that allow alignment of two sequences or multi-alignments of many proteins, thus enabling identification of proteins by both partial NH_2-terminal and internal sequences obtained by Edman degradation. MALDI-mass spectrometry, which has turned out to be a highly sensitive and fast tool to characterize proteins and fragments by mass analysis, adds to the state of the art methods of protein characterization offered here.

If any questions remain, the authors and editors of this book will be glad to offer their help and assistance on request.

Berlin, September 1996

ROZA MARIA KAMP
THEODORA CHOLI-PAPADOPOULOU
BRIGITTE WITTMANN-LIEBOLD

List of Contributors

THEODORA CHOLI-PAPADOPOULOU
Aristotle University
of Thessaloniki
School of Chemistry
Laboratory of Biochemistry
54006 Thessaloniki, Greece

TSEZI EGOROV
Engelhardt Institute
of Molecular Biology
Analytical Protein Chemistry
Research Unit
ul. Vavilova
32, Moscow 117984
Russian Federation

FRANÇOIS FRANCESCHI
Max-Planck Institut
für molekulare, Genetik
Ihnestraße 73
14195 Berlin, Germany

THOMAS G. GIANNAKOUROS
Laboratory of Biochemistry
School of Chemistry
Aristotle University
of Thessaloniki
Thessaloniki 54006, Greece

FRANZ GODT
Technische Fachhochschule
Berlin, FB3
(Chemie und Biotechnik)
Seestraße 64
13347 Berlin, Germany

ULF HELLMAN
Ludwig Institute for Cancer
Research, Box 595
75124 Uppsala, Sweden

HISASHI HIRANO
Yokohoma City University
Kihara Institute
for Biological Research
Majoka-cho 641–12
Totsuka-ku, Yokohama 244
Japan

LOTHAR JACOB
E. Merck
Abt. Bio- und
Produktionschromatographie
64271 Darmstadt, Germany

PETER JUNGBLUT
Max-Planck Institut
für Infektionsbiologie, AG
Proteinanalytik
Monbijoustraße 2
10117 Berlin, Germany

ROZA MARIA KAMP
Technische Fachhochschule
Berlin, FB3
(Chemie und Biotechnik)

Seestraße 64
13347 Berlin, Germany

KATERINA KATSANI
Aristotle University
of Thessaloniki
School of Chemistry
Laboratory of Biochemistry
54006 Thessaloniki, Greece

REGINE KRAFT
Max-Delbrück-Centrum
für Molekulare Medizin
(MDC)
Robert-Rössle-Straße 10
13122 Berlin, Germany

KERSTIN KRAUSE
Technische Fachhochschule
Berlin, FB3
(Chemie und Biotechnik)
Seestrasse 64
13347 Berlin, Germany

D.A. KYRIAKIDIS
Aristotle University
of Thessaloniki
School of Chemistry
Laboratory of Biochemistry
54006 Thessaloniki, Greece

JOHANN SALNIKOW
Technische Universität Berlin
Institut für Biochemie
und Molekulare Biologie
Franklinstraße 29
10587 Berlin, Germany

HORST SCHÜTTE
Technische Fachhochschule
Berlin, FB3
(Chemie und Biotechnik)
Seestraße 64
13347 Berlin, Germany

YANNIS SKENDROS
Aristotle University
of Thessaloniki
School of Chemistry
Laboratory of Biochemistry
54006 Thessaloniki, Greece

ALBRECHT OTTO
Max-Delbrück-Centrum
für Molekulare Medizin
(MDC)
Robert-Rössle-Straße 10
13122 Berlin, Germany

BERND THIEDE
MDC Forschungsgruppe
Proteinchemie
Max-Delbrück-Centrum
für Molekulare Medizin
Robert-Rössle-Straße 10
13125 Berlin-Buch, Germany

DIRK VOLLENBROICH
Technische Universität Berlin
Institut für Biochemie
und Molekular Biologie
Franklinstraße 29
10587 Berlin, Germany

B. WITTMANN-LIEBOLD
Max-Delbrück-Centrum
für Molekulare Medizin
Research Group Protein
Chemistry
Robert-Rössle-Straße 10
13125 Berlin-Buch, Germany

HELMUT WOMBACHER
Technische Fachhochschule
Berlin, FB3
(Chemie und Biotechnik)
Studiengang Biotechnologie
Seestraße 64
13347 Berlin, Germany

Contents

Part I
Separation of Proteins

Chapter 1

Purification of Proteins for Sequencing

D. A. KYRIAKIDIS

1.1 Introduction

During the past decade, biochemical methods of protein purification have become so specialized and sophisticated that it is now difficult for the beginner, whether graduate student or specialist in another field, to follow all the minor but important details which lead to a successful procedure. It should always be kept in mind that there are many alternative ways to purify a particular protein. If a particular procedure is not working another must be tried.

I shall not attempt to discuss all the recent progress in the field, which has been well documented elsewhere (Harris and Angal 1989, 1990; Scopes 1994). Rather, some of the basic criteria for purification of a protein will be cited and some of the recent advances in protein purification will be discussed.

1.2 Criteria for Purification of a Protein

To succesfully purify a protein, a suitable strategy is required that is based on background knowledge and on the following criteria (Harris and Angal 1989):

- What is the best source?

  It is usually worthwhile to spend some time searching for a rich source of the protein. The availability and quantity of the source must be considered as well. In the past, for large quantities of a protein, livers from cows or mice or from animals that could be kept in the laboratory were used. Now most biochemists use cultured cells such as bacteria, yeast or mammalian cells. With the advantages of gene cloning techniques, the desired protein can now be expressed in large

quantities thereby making purification a much more interesting procedure.

- What is known about the protein?

  If a protein has been previously purified from a different source, much of the information can be applied to the protein from the desired source. The size of the molecule, its localization, the pI, its hydrophobicity, the posttranslational modification(s), etc., likely remain the same. If a protein is an enzyme or receptor a successful strategy can be applied based on the relationships of the protein to the substrate or ligand.

- How pure must the protein be?

  The extent of purification usually depends on the final use of the protein. If a protein is prepared for research use it must be very clean, whereas if the protein will be used by industry partial purification is more than sufficient.

- How much of the purified protein is required?

  For studies on activity, relatively small amounts of the active protein are required, whereas for structural studies larger amounts of highly pure protein will be needed.

- How should the protein be assayed?

  To follow a protein during the purification procedure, a quick, reproducible and sensitive assay must exist. This assay should also be cheap, able to be performed in small volumes and not require very expensive instruments.

- How long should purification take?

  The purification procedure should be very quick to minimize activity losses, degradation, etc. Usually protein is lost during the different steps of the purification procedure. Therefore, in order to maximize the yield the number of steps should be minimized.

- What will the final cost be?

  Cost and time are much more important for commercially used proteins.

- How can a protein be purified?

  There are many ways to purify a particular protein. The old procedures included precipitation of proteins with salt, changing the pH of the extract, adsorption, anion exchange, and gel filtration techniques. The new methods include HPLC

and FPLC columns, affinity absorbents, immunoaffinity columns and many other techniques described elsewhere (Harris and Angal 1989, 1990; Scopes 1994). These newer approaches allow even the most difficult protein to be purified. Therefore, before any purification procedure is started all the above critical questions should be addressed.

1.3 Purification of Recombinant Proteins

Purification of proteins from transformed cells has become fashionable these days and almost as common as the purification of proteins from natural sources (Harris and Angal 1990; Scopes 1994). The former procedure includes many steps but the critical points to be considered are:

1. Isolation of the particular gene
2. High level expression of the desired protein (high plasmid number, strong promoters, inducible expression)
3. Suitability of the purification procedure from the recombinant cells

A recombinant protein which is expressed in bacterial host cells is usually either an extracellular protein, a membrane bound protein, an intracellular protein, or in the form of inclusion bodies (Harris and Angal 1990; Scopes 1994). Purification of recombinant proteins offers many advantages since it is possible to modify the level of protein in the initial extract. Moreover, the techniques for separation of unwanted proteins are mainly the same as those used in conventional purification methods.

Unfortunately, recombinant proteins are usually not produced in an active form (Schein 1989). The overexpressed eukaryotic and prokaryotic insoluble proteins, the inclusion bodies, can be extracted using strong chaotropic reagents like 6 M urea or 8 M guanidinium chloride and then correctly refolded during or after removal of the denaturant (Marston 1986; Marston et al. 1988). Also, it is possible to purify inclusion bodies proteins by direct addition of affinity resins, without a denaturing and refolding step (Hoess et al. 1988). Problems with the purification of inclusion bodies protein are: (a) Denaturants (urea or guanidinium) are unpleasant to work with and expensive. They can cause irreversible modification of protein structure that will elude all of the most sophisticated analytical tests except the immune system. (b) Refolding must be usually done in very dilute solutions and the protein reconcentrated (Marston 1986; van Kimmenade et al.

1988). The reconcentrating step is complicated by proteolysis and further precipitation of the protein. (c) Refolding encourages protein isomerization, which leads to precipitation during storage.

A new technique, making fusion proteins, has recently been developed. The advantages of this approach are: (a) improved stability of the protein, (b) improvement of those properties of the protein which allow its isolation from the culture extract, (c) faster isolation and (d) use of the protein by biotechnology industries producing either bulk enzymes or high-purity pharmaceuticals.

When the expression level of a protein is low or the purification process is not suitable, a fusion protein should be prepared (Scopes 1994; Schein 1989). Usually a fusion DNA sequence is connected to the gene encoding protein DNA. To simplify purification, either polyarginine or polyhistidine (in the COOH-terminal), protein A, or glutathione transferase can be used. The purification procedures will therefore involve ion exchange column (with ligand polyArg), immunoaffinity column (with ligand IgG) or substrate affinity (with ligand glutathione) and can be easily applied. Finally, cleavage of the fusion part can be achieved by proteolytic enzymes (e.g., carboxypeptidase B) thus releasing the native protein (Scopes 1994; Sassenfeld 1990; Enfords 1992).

1.4 Purification of Membrane Bound Proteins

The extraction of proteins from membranes presents many difficulties (Harris and Angal 1989, 1990; Scopes 1994). In most processes a detergent is used to solubilize the hydrophobic protein from its membrane structure and then to stabilize the extracted protein. The detergents are normally chosen according to the efficiency of extracting the desired protein; however, some consideration for the next purification step should be made (Scopes 1994). Less detergent in a smaller volume is better. Detergents with high critical micelle concentration such as octyl glucoside may improve the fractionation procedure. The most efficient detergents are the strongly ionic sulfates, such as dodecyl sulfate (SDS), and cationic detergents, such as cetyl-trimethylammonium bromide. Purification in the presence of these detergents is limited mainly to size separation by gel electrophoresis.

Although many membrane proteins must be purified in the presence of detergents, in the end it may be necessary to remove the detergent (Buse et al. 1986). In many cases, this will cause