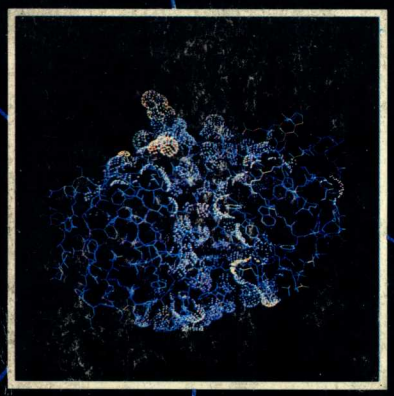
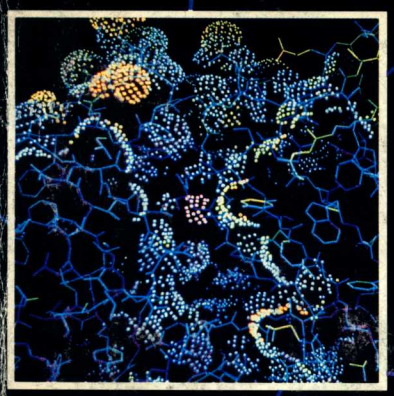
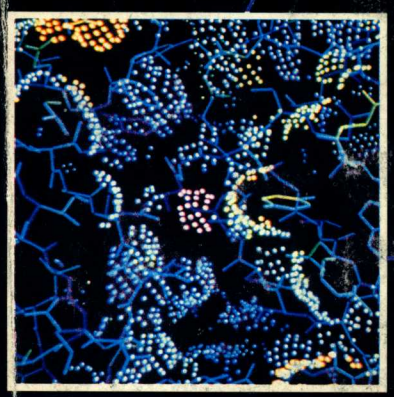
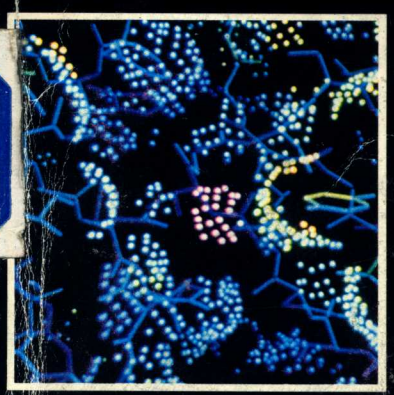
ELSEVIER TRENDS BOOKS









PROTEINS: FORM AND FUNCTION

edited by

RALPH A. BRADSHAW

and

MARY PURTON

PROTEINS: FORM AND FUNCTION

Edited by

RALPH A. BRADSHAW AND MARY PURTON

1990

ELSEVIER TRENDS JOURNALS

Cambridge

ISBN: 1 85166 512 9

Published by:

Elsevier Trends Journals
68 Hills Road
Cambridge
CB2 1LA

Sole distributors worldwide except for the USA and Canada

Elsevier Science Publishers Ltd
Crown House
Linton Road
Barking
Essex IG11 8JU
UK

Sole distributors for the USA and Canada

Elsevier Science Publishing Company Inc.
655 Avenue of the Americas
New York
NY 10010
USA

Cover illustration: Computer graphic representation of the structure of the aspartic proteinase endothiapepsin, courtesy of Lawrence Pearl and Tom Blundell

Design by In-Toto Services
Typeset by Goodfellow & Egan Ltd, Cambridge
Printed in England by Black Bear Press, Cambridge

PROTEINS: FORM AND FUNCTION

蛋白质：形式与功能

Proteins: Form and Function

edited by **Ralph A. Bradshaw** *and* **Mary Purton**

Reprinted from Trends in Biochemical Sciences (TIBS)

1990 About 300 pages
Price: US $37.80 / £ 21.00
ISBN 1-85166-512-9

PUBLICATION: 2nd QUARTER 1990

蛋白质对于细胞结构和功能所有各方面都有根本性的重要意义。蛋白质的研究常常成为生物化学的一个中心部份；近来各种技术的范围和复杂性的增加带来了丰富的新资料。

本书描述了进一步了解蛋白质结构的当代方法和研究趋向；此外有些文章则集中论述细胞中蛋白质的作用。

大部分文章曾在《生物化学科学趋向》(TIBS)评论性月刊上刊登过，有些文章则是特写的。对蛋白质形式和功能感兴趣的学生、讲师、研究人员将喜爱本书。

PREFACE

Protein chemistry is the term generally used to describe studies of the structure and function of proteins. In many respects, however, this term is too limiting. Clearly, the successful study of proteins requires the knowledge and techniques of many fields of science ranging from physics to biology: it is important not only to understand the three-dimensional structure of proteins in atomic detail but also to understand the manner in which they function within cells, organs and whole animals. Biophysics, biochemistry and biology are all equally important in this task.

This volume comprises a series of articles and short reviews that covers this broad spectrum of protein science. Many of the chapters first appeared in a special issue devoted to proteins which appeared as the July 1989 issue of *Trends in Biochemical Sciences*. That issue proved to be highly popular, particularly for senior undergraduate and graduate teaching and has prompted us to reprint and expand this collection, including additional articles that have or will shortly appear in *TIBS* or were specially commissioned for it. In this manner, greater balance and some additional depth in certain areas has been achieved. However, comprehensive coverage in any volume dealing with such a diverse topic is, like beauty, 'in the eye of the beholder'. Clearly, it is not possible to treat every topic in the detail that is required to render the reader a complete understanding. Rather, like the journal from which the articles are taken, the chapters are meant to provide a reasonable introduction, and to provide appropriate leads and direction for further reading to the level desired. Its suitability as instructional material will depend upon the course, the instructor, and the level of expertise to be achieved.

To facilitate its use, the volume is divided into sections that deal with both structure and function. The sections on structure are heavily methodological, focusing on how data are collected and analysed, but are supplemented with articles describing the application of these methods. The sections of the book devoted to function emphasize how structure dictates and controls activity. Considerable attention is devoted to the modification of proteins, both those generated by the powerful arsenal of techniques derived from molecular biological manipulations and those which occur naturally. These latter alterations in protein structure begin at the time of synthesis as co-translational events and ultimately terminate in the proteolytic degradation of the molecule. In between, a large variety of covalent modifications of both a reversible and irreversible character are utilized by cells to ensure their correct translocation, the regulation of their activity, and ultimately their turnover.

It is important to point out that certain functional properties are not developed in great detail. For example, there are not chapters on the properties of enzymes, since many excellent volumes already exist dealing with this material. Similarly specialized structures such as membrane-bound receptors are not addressed directly. Their omission

is not meant to minimize their importance; only to indicate the limitations in space and size of a volume of this type.

Having stressed these shortcomings, it is worth pointing out a few strengths. Importantly the articles are quite recent and therefore represent some of the most up-to-date information on the subject. Some of the material covered is indeed already well known, but most of it focuses on the cutting edge of protein science, as we enter the last decade of this century. It emphasizes the importance of bringing together all of the germane techniques from the aforementioned branches of science if the study of proteins is to advance at the speed and intensity that one expects will characterize the 1990s. We hope that this volume will prove a useful tool for both researchers and students.

We would like to thank the many authors whose original articles appeared in *TIBS*, and those who have provided either updated articles or new contributions for this book. Their willingness to participate in this venture is greatly appreciated.

RALPH A. BRADSHAW AND MARY PURTON

Contents

Preface VII

Protein Primary Structure

Highlights of Protein Structural Analysis 1
John E. Shively, Raymond J. Paxton and Terry D. Lee

Micro-level Protein and Peptide Separations 15
Kenneth J. Wilson

Site-Directed Mutagenesis and Protein Engineering 21
Michael Smith

From Analysis to Synthesis: New Ligand Binding Sites on the Lactate Dehydrogenase Framework. Part I 31
Anthony R. Clarke, Tony Atkinson and J. John Holbrook

From Analysis to Synthesis: New Ligand Binding Sites on the Lactate Dehydrogenase Framework. Part II 39
Anthony R. Clarke, Tony Atkinson and J. John Holbrook

Subtilisin: An Enzyme Designed to be Engineered 45
James A. Wells and David A. Estell

β-Galactoside Transport in *E. coli:* A Functional Dissection of *Lac* Permease 59
H. Ronald Kaback, Eitan Bibi and Paul D. Roepe

Similar Amino Acid Sequences Revisited 71
Russell F. Doolittle

Protein Conformation

Protein Crystallography: More Surprises Ahead 75
David Eisenberg and Christopher P. Hill

Catalysis in Enzyme Crystals 83
J. Hajdu, K. R. Acharya, D. I. Stewart, D. Barford and L. N. Johnson

What Can Two-dimensional NMRT Tell us About Proteins? 95
Peter E. Wright

Circular Dichroism of Proteins 107
N. Neumann and G. Snatzke

Conformational Stability of Globular Proteins 117
C. Nick Pace

How Does Protein Folding Get Started? 125
Robert L. Baldwin

Protein Conformational Prediction 135
Gerald D. Fasman

Membrane Protein Models: Possibilities and Probabilities 147
Jean-Luc Popot and Donald M. Engelman

Protein Motifs and Data-base Searching 153
Janet M. Thornton and Stephen P. Gardner

From Comparisons of Protein Sequences and Structures to Protein Modelling and Design 163
A. Šali, J. P. Overington, M. S. Johnson and T. L. Blundell

The *De Novo* Design of Protein Structures 173
Jane S. Richardson and David C. Richardson

CO- AND POST-TRANSLATIONAL MODIFICATIONS

Protein Translocation and Turnover in Eukaryotic Cells 183
Ralph A. Bradshaw

Post-translational Modifications of Proteins: Some Problems Left to Solve 191
Sou Chi B. Yan, Brian W. Grinnell and Finn Wold

Proteolytic Processing and Physiological Regulation 201
Hans Neurath

Glycoproteins: What are the Sugar Chains For? 209
James C. Paulson

Nuclear and Cytoplasmic Glycosylation: Novel Saccharide Linkages In Unexpected Places 219
Gerald W. Hart, Gordon D. Holt and Robert S. Haltiwanger

MOLECULAR RECOGNITION

Secretion and Membrane Assembly 227
William Wickner

A Novel Pathway for Secretory Proteins? 235
Anne Muesch, Enno Hartmann, Klaus Rohde, Anna Rubartelli, Roberto Sitia and Tom A. Rapoport

The Regulation of Protein Polymerization 241
Carl Frieden

Structural Basis of DNA–Protein Recognition 249
Richard G. Brennan and Brian W. Matthews

Helix-turn-helix, Zinc-finger, and Leucine-zipper Motifs for Eukaryotic Transcriptional Regulatory Proteins 259
Kevin Struhl

INDEX 267

HIGHLIGHTS OF PROTEIN STRUCTURAL ANALYSIS

JOHN E. SHIVELY, RAYMOND J. PAXTON AND TERRY D. LEE

This article describes highlights of the state of the art in protein structural analysis, and comments on the current trends toward increased sensitivity and integrated isolation–structure methodologies.

The biological activity of a protein depends on its three-dimensional structure and its interaction with other molecules. In order to understand the fundamental relationship of a protein's structure to its function, one must start at the beginning, namely, the determination of its structure. Since proteins comprise linear polymers of amino acids, structural analysis of proteins begins with amino acid sequence determination. With the advent of molecular cloning of genes encoding for proteins, and rapid DNA sequencing methods, it has been possible to shorten, or in the opinion of some, to bypass protein sequence analysis by simply deducing amino acid sequence information from nucleotide sequences. However, proteins often undergo extensive post-translational modifications which are not predicted from the gene sequence and so a strong argument can be made for extensive protein structural analysis. In addition, preliminary protein structural information is often used as the basis for gene identification, or for generating oligonucleotides which can be used to clone the gene of interest. In this respect, the emphasis is usually on 'rare' proteins. Recombinant DNA technology allows the production of recombinant proteins which makes rare proteins available in large amounts, or allows the production of site specific modifications in proteins. With recombinant proteins, it is important to confirm that the predicted structure conforms to the expressed product. In fact, it can be said that the recent growth of modern protein structural analysis and molecular biology have been complementary, and that molecular biologists and protein chemists have become dependent on each other for the isolation and characterization of rare proteins and the confirmation of the structure of recombinant proteins.

The science of protein structural analysis is considerably complicated by the need to integrate protein and peptide isolation techniques with amino acid sequence determination. One of the most common complaints is that the sample was lost between the isolation and sequence determination steps. Another problem is concerned with the sequence analysis itself. With 20 naturally occurring amino acids, and hundreds possible by post-translational modifications, the challenge inherent in protein sequence determination is obvious. Nonetheless, the problems are solvable and the sensitivity and speed of the approaches are impressive.

A comprehensive overview of modern protein sequence analysis is beyond the scope of this article. Books have been written on the

subject, and contain valuable practical hints, but are usually out of date before they appear in print[1–5]. A recent review by Simpson and co-workers[6] can be recommended for the integration of isolation and structural determination methods. This article will cover the broad scope of the field with an emphasis towards expected new developments over the next few years.

STRATEGY

The strategy begins with assurance of a steady supply of the 'purified' protein. As analysis proceeds, the results may dictate further purification. The chief criteria of purity include single bands or peaks using *several* chromatographic and/or electrophoretic methods (these are discussed by Ken Wilson, see p. 15). Researchers who wish to sequence proteins directly from one- or two-dimensional gel electrophoretograms, should elute the sample and verify its homogeneity by an additional technique. Proteins which are composed of subunits may exhibit several bands or peaks in the pure state, but will be evident by their reproducible stoichiometry and specific biological activity.

Paramount to the opening gambit is the knowledge of how much sample is present. Although proteins can be detected on electrophoresis gels by various staining techniques, they cannot be quantitated unless they are standards of already known structure. In all cases the sample should be eluted or directly hydrolysed and quantitated by amino acid compositional analysis. This composition also provides a valuable piece of data which should ultimately match that predicted for the total structure. It can be used as a criteria of purity if it is reproducible and unique. From the quantitation and molecular weight (most commonly obtained from SDS gel electrophoresis), one can calculate molar amounts. Although there are occasional examples of obtaining structural information from subpicomole amounts of proteins, it is fair to say that current technology requires at least 10 pmol to begin the analysis. The actual amount used at each step will also depend on the skill and resources of the protein chemist.

Assuming that sample quantitation is known and supply lines are assured, one can plan strategies aimed at three levels of increasing involvement:

(1) 10–100 pmol for several N-terminal sequence and amino acid analyses.
(2) 100–200 pmol for single peptide maps, sequence and mass spectrometric analyses. Important information gained on post-translational modifications.
(3) 100–1000 pmol for complete (or >90%) sequence analysis. Involves multiple peptide maps and complete analysis of post-translational modifications.

If the sample is so rare and difficult to prepare that you can only hope for an N-terminal sequence analysis, then the method of choice may be to sequence the protein following SDS gel electrophoresis and transfer-

ence to a polyvinylidiene (PVDF) membrane, or microbore (1 mm ID) HPLC purification. Even in desperate circumstances, the sample mass should be quantitated by amino acid analysis, by splitting the sample (best), or by running a parallel sample (second best). If no N-terminal sequence is obtained, then the quantitative analysis will allow one to conclude if it is blocked or simply below the level of analysis. If a low yield of the N-terminal group relative to the mass analysis is obtained, it is possible that only a trace impurity was sequenced. The latter possibility may have dire consequences, for many an impurity has been mistaken for the real sample. Thus, even at level (1) above, there is no excuse for omitting quantitative amino acid analysis.

Approaches (2) or (3) are the most desirable, since they will provide enough information to relate the structure to other proteins, reveal post-translational modifications, and with reasonable luck (and skill) provide the entire structure. In these approaches, the key is to obtain peptide maps of the protein by digestion with chemicals such as cyanogen bromide, or proteases such as trypsin. The peptide maps (usually obtained by reversed phase HPLC) provide multiple chances to obtain amino acid sequence information (unlike the N-terminal sequence of the intact protein), but ultimately have to be aligned with either overlapping peptides or the cDNA deduced sequence information. Thus, level (2) assumes a cooperative approach with the molecular biologist, and level (3) offers the chance to 'go it alone'. The speed of peptide mapping and sequence analysis is fast: 20–30 amino acids per day can be sequenced in the average lab. However, it is a general rule that some portion of the protein structure will be refractory, and one usually obtains 90% of the structure within the first few weeks or months, with the remaining 10% requiring an indefinite period of time.

Structural approaches must be integrated with protein-purification strategies in at least three areas. The first is that most analytical methods are adversely affected by salts and require the use of volatile buffers. Thus, the sample may have to be desalted by reversed phase HPLC or precipitation prior to sequence analysis. The second is that the structural methods require concentrated samples (0.1–10 $\mu g\ \mu l^{-1}$), thus the last step of purification may also involve concentration. These strategies are discussed by Ken Wilson (see p. 15). The third is often the trickiest, namely, that the sample may precipitate or be adsorbed to a surface just prior to analysis. Problems with protein precipitation or adsorption can be prevented: (1) by collecting or immediately transferring the sample to a membrane which will be directly analysed; (2) by adding a carrier or substance which prevents protein adsorption to surfaces; or (3) by integrating protein purification and sample analysis. Usually, when these problems become evident, researchers are content to 'scale up', concentrate their samples, or wash the sample tube with solvents of extremes of polarities or pH. Super solvents such as formic acid or hexafluoroacetone trihydrate are recommended for recovery of adsorbed samples. The addition of carriers is still an art, and since proteins differ widely in their adsorptive properties, no universal carrier has

been devised. Organic solvents such as propanol have become popular for preventing protein adsorption to polyethylene or polypropylene tubes. Integrated approaches, for the most part, are still in the future, but the PVDF electrotransfer technology is a first approach.

Amino acid analysis

This all important analytical tool has been recently re-examined in terms of sensitivity and sample preparation. Nonetheless, the analysis of submicrogram amounts of proteins is a daunting problem because even traces of dust or contamination on hydrolysis tubes, for example, will lead to backgrounds greater than the sample mass itself. A recent review on the very popular PTC (phenylthiocarbamyl) method is recommended as a starting point[7]. Amino acid analysis has always been a critical tool in protein chemistry, but has only recently been re-examined in terms of sensitivity and sample handling. We can predict that integrated methods for sample handling, hydrolysis and analysis will ultimately solve these challenging problems.

Reduction and alkylation of cysteine residues

Prior to sequence analysis or peptide mapping the sample may require reduction and *S*-alkylation of cysteine residues. If the protein has cysteines involved in disulfide bonds, it will be necessary to establish their location by peptide mapping of the intact protein (usually involving more aggressive proteases). Standard denaturing conditions for proteins are dissolution in 6 M guanidinium HCl (best) or 8 M urea (requires careful deionization and minimization of sample exposure to prevent blockage of the N-terminal group). The reduction step is performed under nitrogen or argon with DTT (dithiothreitol), mercaptoethanol or tributyl phosphine. Tributyl phosphine is popular because it is volatile and may be used directly on sequencer supports[8]. Popular alkylating agents include iodoacetic acid which can be obtained radiolabeled to trace *S*-carboxymethyl cysteine derivatives[9], and vinyl pyridine which imparts a unique UV absorbance to the *S*-ethylpyridyl cysteine derivatives[10] and is volatile[8]. The references cited demonstrate the methods and discuss strategies for removing the denaturing agent prior to peptide mapping. Denatured proteins are even trickier to handle than the native proteins, often precipitating during the desalting steps. In cases where low microgram amounts of sample are handled, the precipitates are invisible. Care must be taken to follow the sample quantitatively, never discarding tubes in which the sample has resided. Strategies to retain protein solubility include the use of organic solvents such as propanol or detergents such as SDS; however, these additives will frequently affect subsequent mapping and chromatographic steps.

Peptide mapping

Two major options are open for peptide mapping. One is treatment

with cyanogen bromide which cleaves at the C-terminal side of methionine residues, and usually results in very large, relatively insoluble peptides. Before choosing this option, ascertain that the protein has methionine residues by amino acid analysis. If the sample has already precipitated, it can be redissolved in 70–98% formic acid (10–100 µl) and treated with cyanogen bromide[11]. The sample can be dried, redissolved in formic acid for reversed phase HPLC analysis, or in SDS for gel electrophoresis. Remembering that large peptides often aggregate and elute together in low yields on reversed phase HPLC, a safer approach may be to separate the peptides by SDS gel electrophoresis adapted for the resolution of low molecular weight peptides[12]. The peptides can be electrotransferred to PVDF, stained and sequenced. For preservation of disulfide bonds, 6 M guanidinium HCl can be used as the denaturant, and the cyanogen bromide treated samples directly analysed by reversed phase HPLC[13]. If poor yields or poorly resolved peaks are observed for reversed phase HPLC, the sample can be dissolved in 40–60% acetic acid and subjected to gel permeation chromatography. Acidic conditions will promote the formation of homoserine lactone at the C-terminus of the methionyl cleaved peptides. The C-terminus of the protein will be the only peptide not ending in homoserine (lactone) unless nonspecific peptide bond cleavages occur.

The second, more widely used approach, is peptide mapping with proteases. The most useful proteases are trypsin (cleavage at C-terminal side of Lys or Arg), *S. aureus* V8 protease (cleavage at C-terminal side of Glu and sometimes Asp), and Asp-N-protease (cleavage at N-terminal side of Asp, cysteic acid and sometimes Glu). The samples are usually digested in ammonium bicarbonate at pH 8, and the resulting peptides resolved by reversed phase HPLC. Many proteins produce an insoluble peptide core which has to be solubilized and redigested with another protease. Trypsin or Asp-N-protease are good choices for first peptide maps. *S. aureus* V8 protease is a good second choice for overlapping peptides or for digesting peptides in the presence of SDS. Since proteases are inhibited by organic solvents to varying degrees, it is advisable to perform test runs on standards before settling on optimal conditions for maintaining protein solubility and proteolytic activity[13]. In addition to reversed phase HPLC, excellent peptide maps can also be obtained on ion exchange columns such as the sulfoethyl aspartamide column[14,15]. This recent work is of interest because the samples were further analysed without the need for desalting.

Proteins in the size range of 10–50 000 daltons will produce a moderate number of peptides (10–40) which can be resolved by a single HPLC run plus a few selected rechromatographic runs (usually on the same column). The peptides can be preselected by FAB/MS prior to sequence analysis to analyse for purity and molecular weight. Larger proteins give complicated peptide maps requiring extensive rechromatography, but with the existing narrow bore

(2 mm ID) and microbore (1 mm ID) columns, the yields and sensitivities are equal to the task.

EDMAN CHEMISTRY

N-terminal sequence analysis involves the reaction of a free N-terminal amino acid with phenylisothiocyanate (PITC) in base to form a phenythiocarbamyl (PTC) derivative which under acidic conditions forms the cyclic anilinothiazolinone (ATZ) amino acid derivative and a new free N-terminal residue. The unstable ATZ is usually converted by acid hydrolysis to the more stable phenylthiohydantoin (PTH) derivative which is identified by reversed phase HPLC. The chemistry remains unchanged from the original formulation by Edman in 1950[16]. Major improvements include automation[17], and the use of Polybrene to immobilize the sample, especially small peptides[18,19]. The modern automated instrument is similar to that described by Hewick *et al.* in 1981[20], and has a sensitivity in the range of 10–100 pmol when connected on-line to a narrow-bore HPLC. The limit of sensitivity of analysis depends on the detection of PTH amino acid derivatives (currently about 0.1 pmol), initial yields (~30–80%), and repetitive yields (~92–95%). Thus, starting with 10 pmol of sample and (assuming initial and repetitive yields of 50% and 95%, respectively, and no injection losses) the amount of PTH derivative detected after 20 cycles would be about 1.8 pmol. Practically speaking, less than 100% of the sample is analysed, repetitive yields are lower for real samples (compared to standards), and some PTH derivatives are recovered in low yields (e.g. Ser, Thr, His and Arg are detected at about 30–50% the level of the other derivatives). In order to obtain reasonably reliable data with the current methods available, one should analyse about 10 pmol of sample. Since the confidence of assigning amino acid derivatives falls with increasing cycle number, the method has inherent limitations. A determination of the amino acid composition or molecular weight of the analysed peptide is an important check on sequence data. Given the need for high sensitivity, mass spectrometry is the current method of choice for providing molecular weight information. If sequence data and molecular weight data agree, there is an extremely high likelihood that the sequence is correct. If the data don't agree, either an error has been made in the sequence, or a modified amino acid is present.

ELECTROBLOTTING TECHNIQUES

Since SDS gel electrophoresis is the mainstay of protein molecular weight analysis, it is not surprising that many researchers have directed their efforts towards the direct sequencing of proteins from one- or two-dimensional gels. The early work of Hunkapiller and co-workers[21] which describes the elution and sequence analysis of stained protein bands from gels was fraught with problems of recovery and background noise. This approach has now been updated and automated with a continuous elution system commercially available from Applied Biosys-

tems. Highly sensitive electroblotting techniques have now become the mainstay in the field. The initial approach reported by Abersold and co-workers[22], who performed electrotransfer of proteins from SDS gels to derivatized glass fiber paper, later gave way to the more reproducible method of Matsudaira[23] in which proteins were electrotransferred to PVDF membranes. Importantly, the transferred proteins could be stained with Coomassie blue and sequenced without interfering background peaks on HPLC analysis of the PTH derivatives. There has been a flurry of papers describing improvements on the method, including the use of Polybrene[24], reducing agents[25], and detergents[6] to improve transfer and sequencer recoveries. It is also possible to elute proteins from PVDF and recover enzyme activity[26]. Other advantages of the PVDF membrane include its hydrophobic, protein-adsorbing properties, and its relative chemical inertness towards strong acid and base, and organic solvents. For example, samples can be subjected to amino acid analysis or cyanogen bromide cleavage on PVDF membranes without prior elution[6]. PVDF can be chemically derivatized with a wide variety of functional groups[27,28] leading to the possibility of covalent immobilization of peptides and proteins. This concept has revitalized interest in solid-phase Edman chemistry originally described by Laursen in 1971[29], and has recently been adapted to PVDF as a solid phase support[30]. It is too early to evaluate the realized advantages of this technology, but in theory it promises increased sensitivity and cycle number.

Membrane proteins pose special problems for protein solubilization, recovery, and sequence analysis. A recent approach describes solubilization of samples in neat trifluoroacetic acid, disaggregation in SDS, and separation by SDS gel electrophoresis[31]. For complex mixtures, a single dimension separation is inadequate. Successful approaches include two-dimensional gel electrophoresis (isoelectric focusing in one dimension followed by SDS electrophoresis in the second). In this approach sample sizes of about 0.5–1 μg can be electrophoresed, electrotransferred, and sequenced[32]. The obvious problem is sensitivity: 0.5 μg of a 50 kDa protein is only 10 pmol, which is at the limit of sensitivity of current sequencing methods (especially considering transfer and sequencing yields). A second approach is tandem gel electrophoresis or tandem HPLC/SDS gel electrophoresis. One example uses first isoelectric focusing and then acetic acid/triton/urea gels[33], and a second uses first microbore HPLC and then SDS gel electrophoresis[34]. These methods are limited to N-terminal sequence information only. In order to obtain internal sequence information, the sample must be eluted and chemically or proteolytically digested. One approach is described by Abersold *et al.*[35] who transferred gel electrophoresis samples to nitrocellulose and digested reversibly stained proteins directly on the membrane. Due to limitations of sensitivity, the approach often requires multiple samples, especially for samples separated by two-dimensional gel electrophoresis. These techniques point the way to integrated isolation/sequence analysis methods, and demand a further increase in the sensitivity of sequence analysis.

MASS SPECTROMETRY

New developments in mass spectrometry have revolutionized peptide analysis in the last ten years. Among the most widely used is fast atom bombardment mass spectrometry (FAB/MS), in which underivatized peptides are placed in a liquid matrix and desorbed and ionized with a neutral atomic beam (e.g. xenon) or metal ion beam (e.g. cesium), giving rise to high yields of the protonated molecular ion. The high field magnetic instruments are capable of scanning to m/z of 10 000 at full ionizing potential, thus allowing analysis of the majority of peptide fragments obtained from peptide maps. The resolution of these instruments is sufficiently high to give molecular formulae below mass 1000 and are accurate to within one mass until even up to mass 10 000. Thus, one can distinguish amides from free carboxyl derivatives in peptides. Although both negative and positive ions can be detected, greater sensitivity is usually observed in the positive ion mode. However, negative FAB/MS may be the method of choice for phosphorylated peptides. The sensitivity of the analysis is in the low picomole range, competing well with microsequence and amino acid analysis. The results are based on direct mass measurements and are not complicated by the identification of chemical derivatives or measuring retention times. Knowledge of the molecular weight of a peptide allows rapid confirmation of structure or provides evidence for a post-translational modification. Many post-translational modifications have known masses, thus guiding the researcher towards the proper type of analysis. In the case of complex glycosyl units, the majority have well-defined structures, and thus can be predicted with high confidence from mass alone.

An example of the analysis of a glycopeptide is shown in Figs 1 and 2. When a tryptic fragment of the heavy chain (50 kDa) of an immunoglobulin was subjected to microsequence analysis (Fig. 1), clear signals were seen for PTH amino acids at each cycle except cycle 5, suggesting a post-translational modification. The suspicion was further strengthened when FAB/MS analysis revealed a protonated molecular ion of 2603 (monoisotopic mass of 2602.00). This mass matches the structure shown in Fig. 2. Fragmentation ions are consistent with much of the proposed glycosyl structure. This analysis cannot be considered a proof of structure, but based on our knowledge of immunoglobulin structure including glycosylation sites and structure, it is a highly likely structure. Further structural analysis could be initiated at this point if necessary. When mixtures of peptides are analysed the protonated molecular ion intensities vary and usually do not give an indication of the relative amounts present. In some cases, a major species is suppressed, with only the minor species observed. In the case of glycopeptides, because of their low intensities and competition for the matrix surface, it is a good idea to purify them first[36] or convert them to more hydrophobic derivatives[37].

Mass spectrometric methods also provide structural information for peptides. Sample ionization techniques cause fragmentation of the

peptide backbone and can provide significant if not complete sequence information. The low mass region of FAB/MS spectra are usually dominated by ions due to the matrix and sample fragments. Thus, sequence-related fragments may be obscured by spectral background for sample amounts less than a few nanomoles. In ideal cases, considerable sequence information can be obtained from FAB/MS spectra for samples in the 100 pmol range. An example is provided in Fig. 3. The sample was a tryptic peptide obtained from a branched peptide provided to the Protein Society as an unknown in 1987. The presence of the branch point complicated structural analysis, providing a unique opportunity for solving the structure by FAB/MS. The spectrum provides sequence information from both the N and C termini.

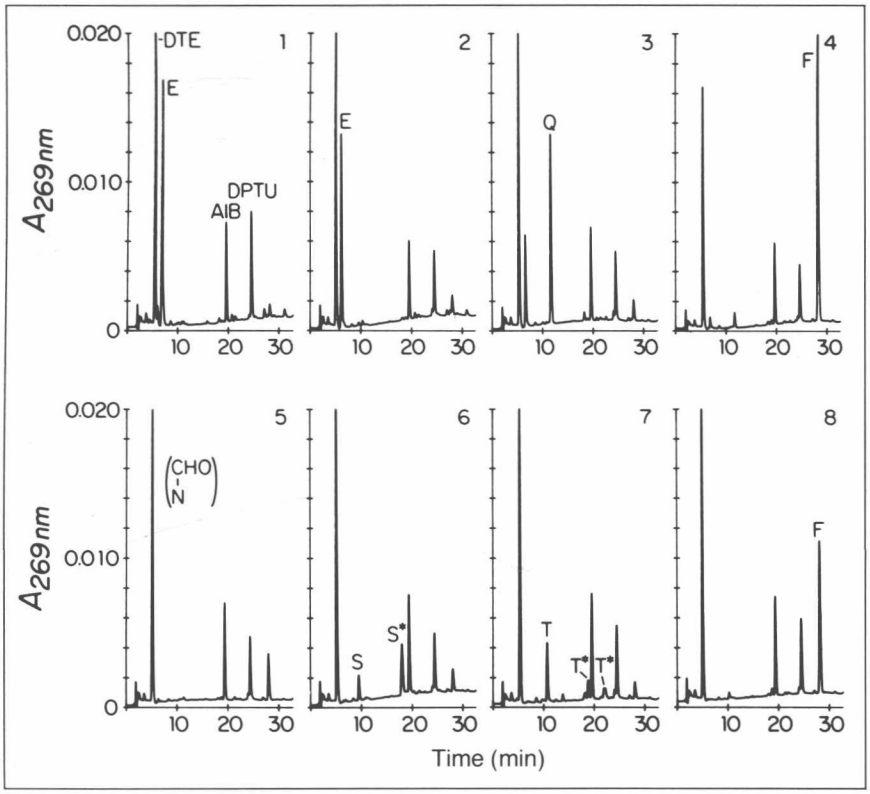


FIGURE 1

Microsequence analysis of a glycosylated peptide. Purified heavy chain from an immunoglobulin was digested with trypsin and mapped by reversed phase HPLC (not shown.) Peptide T13 was sequenced on PVDF using a continuous flow reactor on a gas phase sequencer. The sequence was NH_2-Glu-Glu-Gln-Phe-?-Ser-Thr-Phe-Arg-COOH. (The last cycle is not shown.) Approximately 100 pmol of sample were analysed. Cycle 5, which gave no signal, was suspected to be a glycosylated asparagine residue (N-CHO) by the presence of the sequence (Asn)-XXX-Ser/Thr.

Improvements on sample introduction and sensitivity include the technique of 'flow FAB'[38], in which the sample can be introduced as a continuous stream of liquid. In addition to providing a means of directly coupling FAB/MS to HPLC[39] and capillary electrophoresis, the use of a more volatile flowing matrix greatly reduces the contribution of interfering matrix ions to the mass spectrum. The end result is an order of magnitude increase in the sensitivity of the analysis. A second approach is the use of a focused liquid metal ion gun[40] which gives a similar increase in sensitivity.

Tandem mass spectrometry has the capability to obtain significant structural information on peptides without ancillary analyses. In a tandem experiment, an ion of interest in the spectrum is isolated from all the rest and then fragmented either by collision-activated decomposition or laser photodissociation. The resulting daughter ion spectrum is generally free from interfering ions from the matrix or other peptides present in the sample. Consequently, it is possible to sequence a mixture of peptides without prior separation. All two-sector magnetic instruments have limited tandem MS capabilities by using B/E linked