Laboratory Methodology in Biochemistry

Amino Acid Analysis and Protein Sequencing

Carlo Fini

Ardesio Floridi Vincent N. Finelli

Brigitte Wittman-Liebold

Laboratory Methodology in Biochemistry

Amino Acid Analysis and Protein Sequencing

Editors

Carlo Fini, Ph.D.
Professor of Biochemical Methodology
Faculty of Pharmacy
University of Perugia
Perugia, Italy

Ardesio Floridi, Ph.D.
Professor of Applied Biochemistry
Faculty of Pharmacy
University of Perugia
Perugia, Italy

Vincent N. Finelli, Ph.D.
Adjunct Professor of
Environmental Chemistry
Department of Chemistry
Florida Atlantic University
Boca Raton, Florida

Guest Editor
Brigitte Wittman-Liebold, Ph.D.
Professor of Molecular Biology
Max-Planck Institute for Molecular Genetics
Berlin, West Germany

CRC Press, Inc.
Boca Raton, Florida

PREFACE

This volume meets the increasing demand for information about the most useful techniques for amino acid analyses and protein sequencing. It includes the fundamental aspects as well as the most recent developments in the fast moving field of biochemical methodology. Chapters of this volume encompass all of the important aspects of laboratory methods and computerized data processing for protein sequencing. They were contributed by internationally renowned investigators who have hands-on knowledge and experience in the development and/or applicaton of these methods.

The brief historical introduction not only permits the reader to appreciate all the progress made, starting from the determination of the sequence of insulin (Sanger, 1953) to the introduction of sophisticated fast-atom-bombardment methodologies in the 1980s, but also informs the reader of the potentiality of the various techniques. Details of various methodologies are described and wide bibliographical documentation is given. Above all, the methodologies are illustrated by experiments carried out in the laboratories of the respective authors.

HPLC has proved to be a very flexible technique and is particularly valid not only for the purification and isolation of protein and peptides, but also for the amino acids analysis of protein hydrolysates up to the level of femtomoles. Moreover, it has achieved a remarkable importance when combined with automated sequencers. For this reason, HPLC has been discussed extensively in this volume. Although automatic sequencers are regarded as state-of-the-art in protein sequencing, the basic and essential knowledge of nonautomated sequencing is still more useful to investigators who are in need of procedures optimized for sample concentrations in the range of 0.5 to 10 nmol. This volume includes a detailed description of the nonautomated DABITC/PITC method of sequencing which can be easily adopted in every laboratory (even those with modest instrumentation) and which can be particularly useful for detecting amino acid substitution in mutant proteins.

The two chapters dealing with the use of computers in protein sequencing point out the importance of the elaboration of different algorithms and show how the use of an appropriate algorithm allows the extraction of maximum information from the available experimental data, not only for the definition of a primary structure, but also for predicting the structures of higher orders from knowledge of the primary structure. Finally, the possibility of pointing out functional correlations among proteins of different origins, even with the use of less sophisticated microcomputers, is shown.

The use of various chemical reagents and enzymes for the cleavage of polypeptide chains is reported, and the potentiality of the various methods is illustrated by detailed experiments, thus providing the reader with a good deal of information on the use of various methods. This information together with a comparative evaluation of the discussed methods will assist the investigator in defining an experimental protocol suitable for his/her research problem. The great versatility in blotting techniques in the purification and characterization of proteins is skillfully presented here, introducing the reader to the state-of-the-art method useful not only for purification purposes but also for the determination of terminus amino acid and internal microsequencing of blotted proteins.

The latest approach in the determination of primary structures and sequences of peptides and proteins through mass spectrometric analysis is dealt with in two chapters. The first one reporting the determination of covalently modified peptides, points out the validity of combining gas-phase microsequencing with fast-atom-bombardment mass spectrometry. On the other hand, the second article focuses on the state-of-the-art in mass spectrometry for the determination of sequences of proteins as well as the confirmation of primary structures derived from cDNA sequencing or the confirmation of synthetic peptides which have become important in biochemistry as well as in molecular biology and immunology. With the wide range of topics reported, this volume may certainly be considered a useful reference not only for studies strictly

dealing with protein sequencing but also in many other fields in which amino acid determination is concerned.

We are deeply grateful to all the authors for their valuable contribution and interest. We also would like to express our gratitude to the members of the Advisory Board for their valuable suggestions and to Guest Editor Dr. Wittman-Liebold for her knowledgeable effort. Our sincere thanks go to the staff of CRC Press for their patient cooperation and support.

The Editors Fini, Floridi, and Finelli

THE EDITORS

Carlo Fini, Ph.D., is Professor of Biochemical Methodology in the Faculty of Pharmacy of the University of Perugia, Perugia, Italy. Dr. Fini obtained his degree in chemistry at the University of Perugia in 1968 and joined the faculty in 1969.

Dr. Fini's current research interests include enzymology and metalloproteins. He has published more than 40 research papers.

Ardesio Floridi, Ph.D., is Professor of Applied Biochemistry in the Faculty of Pharmacy of the University of Perugia and is Director of the postgraduate school of Clinical Biochemistry of the Department of Experimental Medicine and Biochemical Sciences, also at the University of Perugia, Perugia, Italy.

Dr. Floridi is a member of many scientific societies and has authored more than 110 research papers, including monographs and a book. His current research interest is focused on the purification and the molecular characterization of the enzymes involved in nucleic acid and nucleotide metabolism and also on the development and refinement of HPLC methodologies for the analysis of amino acids, peptides, sugars, and biogenic amines.

Vincent N. Finelli, Ph.D., is Director of Hazardous Waste Management and also Adjunct Professor in the Department of Chemistry at Florida Atlantic University, Boca Raton. Dr. Finelli received his doctorate in chemistry at the University of Rome, Italy in 1968 and was awarded a postdoctoral fellowship in Environmental Health at the College of Medicine, University of Cincinnati (1969 to 1971), where he remained as a faculty member until 1984. During a sabbatical leave in 1980—1981, he was Visiting Professor at the Institute of Occupational Medicine at the Catholic University of Rome.

Dr. Finelli's major interests include the interaction of essential and toxic metals, especially zinc and lead, the health effects of pollutants in relation to the nutritional status of the organism, and the evaluation of the catalytic converter in detoxification of automobile emissions. Dr. Finelli has published more than 100 papers in scientific journals, symposia proceedings, and book chapters, including presentations at national and international meetings, invited lectures, and seminars. He also participated as coauthor and reviewer in the preparation of several environmental quality criteria documents for the U.S. Environmental Protection Agency (EPA) and other government agencies.

GUEST EDITOR

Brigitte Wittmann-Liebold, Dr. rer. nat., Diplom Chemist, is Research Leader of the Protein Chemical Group, Max-Planck Institute of Molecular Genetics, Berlin, West Germany, in the Department Wittmann. Dr. Wittmann-Liebold received her Vordiplom and Diplom in chemistry from the University of Tuebingen in 1953 and 1958, and her Ph.D. (Dr. rer. nat.) in chemistry from the University of Munich, West Germany in 1960, respectively.

She is an honorary member of the Faculty of Medicine at the Free University of Berlin in the Department of Clinical and Experimental Virology. Dr. Wittmann-Liebold instructs courses for advanced students of Biochemistry, Chemistry, Medicine, and Biology and organized International Courses on modern protein structure analysis. Dr. Wittmann-Liebold has published more than 150 papers on the structure analysis of various proteins (human hemoglobin, virus coat proteins, phages, and ribosomal proteins of various sources — *E. coli*, *B. stearothermophilus*, Archaebacteria).

She developed micromethods for the purification of peptides and proteins in minute amounts, and for manual and automatic protein and peptide sequence analysis. Her current research interests include topographical investigations of ribosomes by chemical protein-protein cross-linking, and the evolution of the ribosomal constituents. She is engaged in the design of automates of modular construction and appropriate programs for the sequence analysis and synthesis of biomolecules.

ADVISORY BOARD

CONTRIBUTORS

Wolfgang Ade, M.D.
Project Molecular Biology of Mitosis
German Cancer Research Center
Heidelberg, West Germany

Alastair Aitken, Ph.D.
Staff Scientist
Laboratory of Protein Structure
National Institute for Medical Research
London, England

Guy Bauw
Research Assistant
Laboratory of Genetics
State University of Ghent
Ghent, Belgium

Suzanne Benjannet, M.Sc.
Research Assistant
J.A. De Sève Laboratory of Molecular Neuroendocrinology
Clinical Research Institute of Montreal
Montreal, Quebec, Canada

Francesco Bossa, Ph.D.
Professor
Department of Biochemical Sciences
University of Rome La Sapienza
Rome, Italy

J.-Y. Chang, Ph.D.
Scientific Specialist
Department of Biotechnology
Ciba-Geigy Ltd.
Basel, Switzerland

Michel Chrétien, Ph.D.
Director
J.A. Sève Laboratory of Molecular Neuroendocrinology
Clinical Research Institute of Montreal
Montreal, Quebec, Canada

Alfredo Colosimo, Ph.D.
Professor
Department of Experimental Medicine
University of Rome
Rome, Italy and
Professor
Institute of Biochemical Sciences
University of Chieti
Chieti Scalo, Italy

Séverine Frutiger, Ph.D.
Research Associate
Department of Biochemical Medicine
University of Geneva Medical Center
Geneva, Switzerland

Mark A. Hermodson, Ph.D.
Professor and Head
Department of Biochemistry
Purdue University
West Lafayette, Indiana

Agnes Hotz, Dr. rer. nat.
Institute of Experimental Pathology
German Cancer Research Center
Heidelberg, West Germany

Graham John Hughes, Ph.D.
Senior Research Scientist
Department of Biochemical Medicine
University of Geneva Medical Center
Geneva, Switzerland

Paul Jenö, Ph.D.
Department of Biotechnology
Ciba-Geigy Ltd.
Basel, Switzerland

Tore Kempf, Dr. rer. nat.
Project Molecular Biology of Mitosis
German Cancer Research Center
Heidelberg, West Germany

Rene Knecht
Chief Assistant
Department of Biotechnology
Ciba-Geigy Ltd.
Basel, Switzerland

Erika Krauhs, Ph.D.
Department of Molecular Biology
Max-Planck Institute for Biophysical Chemistry
Gottingen-Nikolausberg, West Germany

Claude Lazure, Ph.D.
Associate Director
J.A. Sève Laboratory of Molecular Neuroendocrinology Laboratory
Clinical Research Institute of Montreal
Montreal, Quebec, Canada

Melvyn Little, Ph.D.
Institute of Cell and Tumor Biology
German Cancer Research Center
Heidelberg, West Germany

Mary B. LoPresti
Associate in Research
Department of Molecular Biophysics and Biochemistry
Yale University
New Haven, Connecticut

Gernot Maier, Dr. rer. nat.
Project Molecular Biology of Mitosis
Institute for Cell and Tumor Biology
German Cancer Research Center
Heidelberg, West Germany

Stefano Pascarella, Ph.D.
Research Assistant
Department of Biochemical Sciences
University of Rome La Sapienza
Rome, Italy

Pasquale Petrilli, Ph.D.
Research Assistant
Istituto di Industrie Agrarie
Universita di Napoli
Naples, Italy

Herwig Ponstingl, Dr. phil.
Professor
Project Molecular Biology of Mitosis
German Cancer Research Center
Heidelberg, West Germany

Magda Puype
Technician
Laboratory for Genetics
State University of Ghent
Ghent, Belgium

James A. Rochemont
Research Associate
Instrumentation and Methods
J.A. Sève Laboratory of Biochemical and Molecular Neuroendocrinology
Clinical Research Institute of Montreal
Montreal, Quebec, Canada

Nabil G. Seidah, Ph.D.
Director
J.A. Sève Laboratory of Biochemical Neuroendocrinology
Clinical Research Institute
Montreal, Quebec, Canada

Yasutsugu Shimonishi, Ph.D.
Professor
Institute for Protein Research
Osaka University
Suita, Osaka, Japan

Kathryn L. Stone
Laboratory Manager
Department of Molecular Biophysics and Biochemistry
Yale University
New Haven, Connecticut

Toshifumi Takao, Ph.D.
Instructor
Institute for Protein Research
Osaka University
Suita, Osaka, Japan

Gaétan Thibault, Ph.D.
Associate Director
Laboratory of Pathobiology
Clinical Research Institute of Montreal
Montreal, Quebec, Canada

Jozef Van Damme
Researcher
Laboratory of Genetics
State University of Ghent
Ghent, Belgium

Joel Vandekerckhove, Ph.I
Professor
Laboratory of Genetics
State University of Ghent
Ghent, Belgium

Marc Van Den Bulcke
Research Assistant
Laboratory of Genetics
State University of Ghent
Ghent, Belgium

Marc Van Montagu, Ph.D.
Professor
Laboratory of Genetics
State University of Ghent
Ghent, Belgium

Kenneth R. Williams, Ph.D.
Senior Research Scientist
Department of Molecular Biophysics and Biochemistry
Yale University
New Haven, Connecticut

TABLE OF CONTENTS

Chapter 1

A SHORT HISTORY OF PROTEIN SEQUENCE ANALYSIS

Mark A. Hermodson

TABLE OF CONTENTS

I. ANCIENT HISTORY: 1950—1970

Amino acid sequence analysis of proteins has progressed through a number of stages since the structure of insulin was determined in 1953 by Frederick Sanger. Dr. Sanger did that analysis at a time when very few of the chemical and instrumental tools of protein chemistry were developed. He had to isolate and characterize more than 150 short peptides from the 51-residue protein; the anaylsis was extremely labor-intensive and required huge amounts of protein. But it proved once and for all that each protein has a unique amino acid sequence.

Two developments in the mid-1950s made sequence analysis of small (less than M_r 40,000) proteins possible: (1) the development of a quantitative amino acid analyzer by Stanford Moore and William Stein; and (2) Per Edman's contribution of a sequential chemical-degradation method capable of removing one amino acid at a time cleanly from the amino terminus of a polypeptide. Various enzymatic and chemical cleavage methods were developed to generate peptides 5 to 15 residues in length, and the newly-developing science (art) of column chromatography (mainly Dowex ion exchangers) made it possible to purify the peptides for sequence analyses. A general strategy developed which was used in the 1950s and 1960s to sequence dozens of proteins.

1. The protein was cleaved into peptides averaging about 8 to 10 residues in length.
2. The peptides were isolated by chromatographic and paper electrophoretic methods.
3. A portion of each was acid hydrolyzed to determine its amino acid composition (using the amino acid analyzer), and the rest was subjected to Edman degradation for as many cycles as definitive sequence could be determined (by hand methods that was usually 5 to 15 cycles).
4. A different cleavage method was employed on the whole protein and Steps 1 to 3 repeated for that set of peptides.
5. Overlapping sequences were aligned to give extended sequence.
6. "Holes" in the sequence were filled in by generating yet more sets of small peptides until a complete sequence was obtained.

The above approach was still very labor-intensive and required gram quantities of protein. The size of the protein was limited because large proteins gave more small peptides than the separation methods could resolve.

II. THE RECENT PAST: 1970—1985

Automated sequencers became available in 1970. They perform the Edman degradation under rigidly controlled conditions in an inert atmosphere with highly purified reagents. Consequently, the length of readable sequence per degradation increased from an average of 10 or fewer residues to between 30 and 40 (sometimes more). This meant that far fewer peptides had to be isolated and each gave a substantial stretch of sequence, thus reducing the total number of overlapping sequences which were necessary. Roughly a tenfold increase in efficiency, both in terms of labor expended and in protein used, was realized. Continued improvements in peptide isolation techniques and sequencer technology in the 1970s and early 1980s increased the speed of analysis by a further factor of severalfold and vastly reduced the amounts of peptides required in the sequencer reaction chamber (from about 100 nmol in 1970 to 100 pmol in 1985). Much larger proteins could be sequenced by these methods — up to 1,000 residues or so.

In spite of the significant improvements in speed and sensitivity which have been realized over the past two decades, a radically new approach to protein sequence analysis was developed by the mid-1980s (see following section), and no one should contemplate sequencing a whole protein more than 200 residues long by protein sequencer technology any more. Nevertheless,

ENZYMATIC CLEAVAGE METHODS

Enzyme	Site	Conditions	Comments	Ref.
Staphylococcus aureus V8	Glu-X	pH 8 or pH 4	Glu-Pro is resistant Peptides average ~20 to 25	11
Clostripain	Arg-X	pH 8	Variable average size — usually 30 to 50	12
Endoproteinase Lys-C	Lys-X	pH 8	Peptides average ~15 to 20	
Trypsin	Arg-X Lys-X	pH 8	Peptides average ~10. Can be restricted by modifying Lys. Arg-Pro or Lys-Pro resistant.	

FIGURE 1. All other proteinases are too nonspecific (e.g., chymotrypsin, pepsin, papain) or too restricted (e.g., thrombin) to be of general use. All four of the enzymes above can be used in urea solution, which increase the rate of digestion and prevents precipitation of partially digested protein. This is crucial for obtaining complete digestion. If >4 *M* urea is required, small aliquots of proteinase should be added in one hour intervals to compensate for autolysis. The V8 proteinase also cleaves Asp-X bonds if, but only if, phosphate buffers are used, usually not a desirable feature. The reactions are normally run at 5 to 10 mg/ml of protein with 1 to 2% by weight of proteinase. Dropping the pH to about 1 with formic acid terminates the reaction. Upon injection into a reverse-phase HPLC column, the urea, formic acid, and buffer all come out in the breakthrough with peptides emerging in the gradient.[1]

the methods which were developed to accomplish that task are still needed for the new approach and the following paragraphs outline the most important cleavage and peptide separation techniques available to the protein chemist at this time.

Application of the automated sequenator to the determination of the total amino acid sequence of a protein required the investigator to generate and purify appropriate fragments of the protein in order to obtain overlapping stretches of sequence covering the whole molecule. Since the sequenator yielded 35 to 50 residues of sequence from a protein fragment, it was most efficient to generate fragments in the 25- to 80-residue size range for the analyses. This required that the cleavage points be rather infrequent in the protein, a situation which also simplified the purification of the fragments by reducing the number of fragments in the digest. The fragmentation procedure also needed to be highly specific for a particular type of bond, so that side reactions (which make fragment purification and identification of certain residues difficult) were not encountered. Finally, the desired bond cleavage needed to proceed in near quantitative yield in order to minimize the number and amounts of fragments generated by incomplete cleavage of a particular bond.

The cyanogen bromide degradation meets these criteria almost perfectly for most proteins. Cyanogen bromide cleaves most Met-X bonds in almost quantitative yield. It is highly specific to Met-X without modifying other residues. Methionyl residues are rare enough in most proteins to yield fragments of ideal size (averaging about 60 residues).

Digestion at arginyl bonds with trypsin can be accomplished by blocking the lysyl residues, preferably with citraconic or succinic anhydride in order to change the charge of the residue from positive to negative. This change makes the denatured protein highly acidic and thus more soluble at pH 8 where the tryptic digestion must be run. Modification with citraconic anhydride has the advantage of reversibility. Thus, merely acidifying the mixture stops the trypsin and removes the blocking groups. Again, this cleavage method is highly quantitative, highly

CHEMICAL CLEAVAGE METHODS

Reagent	Site	Conditions	Comments	Ref.
CNBr	Met-X	70% HCOOH	Peptides average ~50	
o-Iodosobenzoic acid	Trp-X	4 *M* guanidine in 80% HOAC	Peptides average ~60. Messy workup	2
Hydroxylamine	Asn-Gly	NH_2OH in 4 to 6 *M* guanidine, pH 9.5	Peptides average >80 yield not always good	14
Mild acid	Asp-Pro	(see ref.)	Peptides average >80	15
2-Nitro-5-thiocyanobenzoic acid	X-Cys	6 *M* guanidine, pH 9.0	Peptides usually large. N-terminus is blocked.	16

FIGURE 2. The methods in Figure 1 all produce very large peptides from most proteins. Consequently, the peptides tend to be very difficult to dissolve; they aggregate, and they usually give low yields on HPLC. If sufficient protein is available to use gel filtration procedures, these methods can be very good means of "cracking" a molecule into a small number of pieces; gel filtration should be run in 10 to 20% formic or acetic acids in order to dissolve the peptides and prevent aggregation.[2,14-16]

specific, and for many proteins, yields fragments of ideal size. Very often 70 to 80% of a protein sequence could be obtained from sequenator degradations of the cyanogen bromide and arginyl peptides alone. An alternative way to cleave at arginyl residues is to use the enzyme clostripain. It is quite specific for arginyl residues and does not cleave at lysyl residues. The protein to be digested must be dissolved at pH 8, which, for most denatured proteins, requires high concentrations of urea. Clostripain is active in urea solutions, so this is not a serious limitation.

Digestion of polypeptides at tryptophanyl residues with *o*-iodosobenzoic acid generates fragments averaging about 60 residues from most proteins.[1,2] The yield of cleavage at most Trp-X bonds is close to quantitative. Methionyl and alkylated cysteinyl residues are oxidized to the sulfoxides which can be reduced later with thiol reagents. Free cysteine is oxidized to cystine.

Digestion at glutamyl residues is performed with *Staphylococcus aureus* protease V8.[3] The protease is active at pH 4.0 where many denatured proteins or peptides are soluble. It is specific for Glu-X bonds provided phosphate is absent (in phosphate, Asp-X bonds are also cleaved). It cleaves quantitatively at most susceptible bonds under proper conditions. Peptides average 15 to 30 residues long.

Specific and quantitative digestion of proteins at cysteinyl residues is also possible[4] and would be exceedingly useful for sequencing if the fragments were not blocked. Since there is no practical way to generate a free amino group from the 2-iminothiazolidine ring formed during the cleavage reaction, this method has limited value for protein-sequencing approaches, but with the new approach employing mass spectral analyses (below), it is once again an important cleavage method.

An enzyme, endoproteinase Lys-C, is available which is specific for Lys-X bonds. It is the preferred way to cleave at those sites. It is active in high concentrations of urea which are usually required to dissolve the denatured peptide at pH 8.

Specific digestion at prolyl residues is also possible.[5] The reaction conditions (sodium metal in anhydrous liquid ammonia) are both dangerous and cumbersome. This digest is rarely employed.

Cleavage can be accomplished at Asn-Gly bonds with hydroxylamine.[6] These bonds are rare in proteins, occurring once or twice in the typical protein. Although the cleavage yield is normally only 70%, the low number of susceptible bonds makes fractionation of the mixture relatively easy. This is an excellent procedure to "crack" a molecule into two or three large pieces and thus reduce the complexity of the cyanogen bromide or arginyl digests.

Likewise, mild acid treatment can give good fragments by cleaving Asp-Pro bonds, another very rare sequence.[7] Strong acid will, of course, result in nonspecific cleavage of peptide bonds. Figures 1 and 2 summarize the enzymatic and chemical cleavage methods, respectively.

If 10 mg or more of a protein are subjected to one of the above cleavage procedures, a preliminary separation of the peptides on a Sephadex® G-50 gel permeation column is the best first step toward purification. The column should be run in 10 to 20% formic or acetic acid solutions because most peptides over 20 residues long are very insoluble in neutral aqueous solutions. Peaks of absorbance at 280 nm and/or 255 nm should be pooled, but pools should also be made of the areas between the peaks throughout the column effluent (peptides without Tyr, Phe, Trp, or modified Cys residues do not absorb above 250 nm). Pools in the first $^2/_3$ of the effluent between V_0 and V_s of the column (roughly peptides of >30 residues) should be further purified on reverse phase HPLC using propanol gradients (see below), while the small peptides in the last $^1/_3$ of the column should be run on RP-HPLC in acetonitrile gradients.

If less than 10 mg of protein is digested, it is best to use HPLC for all separations. Reverse-phase, gel permeation, and ion-exchange columns are available. They are limited by the poor solubility of peptides in water buffers, which severely limits ion exchangers (usually high concentrations of urea are necessary for ion exchange of peptides which is messy, tedious, and interferes with peptide detection). HPLC separations are extremely fast (less than 2 h per run) and resolution is extremely good. Separation conditions are easily modified by solvent programming to provide optimal separations. We have found that trifluoroacetic acid solutions are excellent solvents in which to chromatograph peptides of all sizes on reverse phase HPLC columns.[8] This system has the following advantages:

1. It employs a volatile solvent so the sample is easily recovered.
2. The solvent is transparent in the low UV (210 to 230 nm) so detection of all peptides is sensitive and nondestructive.
3. Recoveries are generally good.

New stationary phases have been developed for optimal resolution of large peptides.[9] These methods work best for peptides 5 to 60 residues in size. Larger peptides often give low yields on reverse-phase columns.

III. THE MODERN ERA: 1985 —

While nine different cleavage sites of high specificity and yield (see above section) may seem like a reasonably large collection, nature has a way of distributing them in most proteins to make it very difficult to generate the last overlap or two to complete the sequence. In addition, due to the highly variable amino acid compositions of peptides generated from the whole protein, the physical and chemical characteristics of a given set of peptides are very different, complicating the task of dissolving, separating, and isolating them. Finally, even though the amount of protein needed to perform a complete sequence analysis has decreased by a factor of more than a thousand while the speed of the analysis has increased manyfold, it still takes milligram quantities to even contemplate such a task, and a medium-sized protein (e.g., M_r 40,000) may consume more than a working year of time for a researcher.

A combination of protein and DNA-sequencing technology is a far better approach at this