

Chapter 1

Development of the Chemistry and Macromolecular Structures of Proteins

Summary

Amino acids began to be identified as constituents of proteins in the eighteenth and nineteenth centuries. The first important advance was the proposal of a unifying theory of protein structure in 1902. This stated that amino acids were joined to one another through amide (peptide) linkages that connected the α -amino group of one amino acid to the carboxyl group of the next amino acid. Progress towards the next two steps, namely complete amino acid analysis of proteins and the determination of amino acid sequences in them, was slow in the first forty years of this century. However, the invention of partition chromatography and, somewhat later, ion exchange chromatography revolutionized work in the field. The first amino acid sequence of a protein, insulin, was determined in the early fifties and the sequences of large proteins soon followed with increasing speed. The next climactic advance was the three-dimensional X-ray analysis (including isomorphous replacements) in the early sixties, to be followed later by protein structures in solution by NMR. Automated synthesis of polypeptides in the solid phase provided the next high point. The next two major advances in the protein field were made possible by the developing DNA technology: (1) derivation of the amino acid sequences of classically intractable membrane proteins from the corresponding gene sequences; (2) site-specific mutagenesis, namely amino acid replacements by mutagenesis at the gene level followed by expression of the mutant genes. Most recent are the dramatic advances in mass spectrometry for determining molecular weights and amino acid sequences in proteins.

Early History: Discovery of Amino Acids

Early contributions, made in the eighteenth and nineteenth centuries, were from workers interested in medicine or nutrition. Beccari, a physician and natural philosopher, prepared the first protein of vegetable origin in 1747. Fourcroy, also a physician, coined the terms “albumin” and “gelatin,” and Scheele, an apothecary, carried out the first noteworthy investigation of casein. Braconnot, a pharmacist, discovered glycine and leucine in 1820 and Wollaston, the discoverer of cysteine, was a physician, physicist and chemist. Berzelius provided the names “cystine” and “glycine.” Mulder, who coined the word “protein,” speculated on what these substances were and stimulated much research. Liebig discovered tyrosine and his student Ritthausen became the first protein chemist of distinction. In the next phase, essentially throughout the nineteenth century, organic chemists and early biochemists, among them Kuhne, Kossel, Hofmeister and Schulze, made great strides by discovering a variety of amino acids as constituents of proteins.

With the discoveries of amino acids in different proteins and also in other natural products, a number of questions began to be formulated. Could one characterize a protein in terms of its amino acid composition? What amino acids were the common and important constituents in proteins? For example, in his classic book,¹ Ritthausen described analyses of amino acids in the three main types of proteins (of cereals, legumes and oilseeds) and produced the first table on the amino acid composition of proteins. Yet, there were only a very few amino acids in the table. Into the twentieth century, and well into the 1940's and even 1950's, the above questions persisted. How many amino acids are really of significance in the chemistry of proteins? Which amino acids are, *bona fide*, the naturally occurring constituents of proteins and which ones are not? As late as 1941, Vickery² who had dedicated his life to the chemistry and analysis of amino acids in proteins, presented comprehensive lists of amino acids, dividing them into four groups. The first, and main, group consisted of eighteen amino acids; this group undoubtedly was produced by hydrolysis of proteins. Except for hydroxyproline, the remaining seventeen in Vickery's list retained their authenticity and the three missing out of the total list of the “magic” twenty established later (see below) were cysteine, glutamine and asparagine (all these three are now known to suffer degradation or hydrolysis during the acidic treatment used for proteins). The second group in Vickery's list occupied a special position because of its narrow range of distribution, while the third group was identified as plant constituents that could also be expected to be present in proteins. The fourth group contained amino acids whose source identification was ambiguous. This last group actually contained some interesting amino acids, such as amino-butyric acid, hydroxyvaline, hydroxylysine, norvaline and diamino-adipic acid. Hydroxyproline also belonged to this list.

After the advent of partition chromatography (see below), which made quantitation feasible and far easier, Syngé³ drew up lists of amino acids that were judged to be authentic protein constituents. Table 1 lists and includes structures of the twenty accepted amino acids in proteins. These are also now known to be coded genetically (Chapter 8).

We also know now that there are a number of additional amino acids that are found in proteins and these are produced by posttranslational modification, i.e. after synthesis of the

Table 1. Twenty naturally occurring amino acids in proteins grouped into four classes, (a)–(d).

(a) Non-polar amino acids with aliphatic sidechains			(b) Polar amino acids with aliphatic sidechains		
Name	Structure	Three (one) letter abbreviation	Name	Structure	Three (one) letter abbreviation
Glycine	$\text{H}_3\text{N}^+ - \underset{\text{H}}{\overset{\text{H}}{\text{C}}} - \text{COO}^-$	Gly (G)	Serine	$\text{H}_3\text{N}^+ - \underset{\text{H}}{\overset{\text{CH}_2\text{OH}}{\text{C}}} - \text{COO}^-$	Ser (S)
Alanine	$\text{H}_3\text{N}^+ - \underset{\text{H}}{\overset{\text{CH}_3}{\text{C}}} - \text{COO}^-$	Ala (A)	Threonine	$\text{H}_3\text{N}^+ - \underset{\text{H}}{\overset{\text{CH}(\text{OH})\text{CH}_3}{\text{C}}} - \text{COO}^-$	Thr (T)
Valine	$\text{H}_3\text{N}^+ - \underset{\text{H}}{\overset{\text{CH}(\text{CH}_3)_2}{\text{C}}} - \text{COO}^-$	Val (V)	Cysteine	$\text{H}_3\text{N}^+ - \underset{\text{H}}{\overset{\text{CH}_2\text{SH}}{\text{C}}} - \text{COO}^-$	Cys (C)
Leucine	$\text{H}_3\text{N}^+ - \underset{\text{H}}{\overset{\text{CH}_2\text{CH}(\text{CH}_3)_2}{\text{C}}} - \text{COO}^-$	Leu (L)	Methionine	$\text{H}_3\text{N}^+ - \underset{\text{H}}{\overset{\text{CH}_2\text{CH}_2\text{S-CH}_3}{\text{C}}} - \text{COO}^-$	Met (M)
Isoleucine	$\text{H}_3\text{N}^+ - \underset{\text{H}}{\overset{\text{CH}(\text{CH}_3)\text{CH}_2\text{CH}_3}{\text{C}}} - \text{COO}^-$	Ile (I)	Asparagine	$\text{H}_3\text{N}^+ - \underset{\text{H}}{\text{C}} - \text{COO}^-$	Asn (N)
Proline	$\text{H}_3\text{N}^+ - \underset{\text{H}}{\text{C}} - \text{COO}^-$ (where the nitrogen is part of a five-membered ring with the side chain)	Pro (P)	Glutamine	$\text{H}_3\text{N}^+ - \underset{\text{H}}{\overset{\text{CH}_2\text{C}(=\text{O})\text{NH}_2}{\text{C}}} - \text{COO}^-$	Gln (Q)
(c) Charged amino acids with aliphatic sidechains			(d) Amino acids with aromatic sidechains		
Name	Structure	Three (one) letter abbreviation	Name	Structure	Three (one) letter abbreviation
Aspartic Acid	$\text{H}_3\text{N}^+ - \underset{\text{H}}{\overset{\text{CH}_2\text{COO}^-}{\text{C}}} - \text{COO}^-$	Asp (D)	Histidine	$\text{H}_3\text{N}^+ - \underset{\text{H}}{\overset{\text{CH}_2\text{C}_5\text{H}_4\text{N}_2}{\text{C}}} - \text{COO}^-$	His (H)
Glutamic Acid	$\text{H}_3\text{N}^+ - \underset{\text{H}}{\overset{\text{CH}_2\text{CH}_2\text{C}(=\text{O})\text{O}^-}{\text{C}}} - \text{COO}^-$	Glu (E)	Phenylalanine	$\text{H}_3\text{N}^+ - \underset{\text{H}}{\overset{\text{CH}_2\text{C}_6\text{H}_5}{\text{C}}} - \text{COO}^-$	Phe (F)
Lysine	$\text{H}_3\text{N}^+ - \underset{\text{H}}{\overset{\text{CH}_2\text{CH}_2\text{CH}_2\text{CH}_2\text{NH}_3^+}{\text{C}}} - \text{COO}^-$	Lys (K)	Tyrosine	$\text{H}_3\text{N}^+ - \underset{\text{H}}{\overset{\text{CH}_2\text{C}_6\text{H}_4\text{OH}}{\text{C}}} - \text{COO}^-$	Tyr (Y)
Arginine	$\text{H}_3\text{N}^+ - \underset{\text{H}}{\overset{\text{CH}_2\text{CH}_2\text{CH}_2\text{N}(\text{H})\text{C}(=\text{NH}_2)\text{NH}_2}{\text{C}}} - \text{COO}^-$	Arg (A)	Tryptophan	$\text{H}_3\text{N}^+ - \underset{\text{H}}{\overset{\text{CH}_2\text{C}_8\text{H}_6\text{N}_2}{\text{C}}} - \text{COO}^-$	Trp (W)

nascent protein chains on ribosomes (Chapter 7). Some of these are diaminopimelic acid, hydroxyproline, hydroxylysine, phosphoserine, γ -carboxyglutamic acid.

Peptide (Amide) Theory of Protein Structure

Fischer⁴ and Hofmeister⁵ independently proposed in 1902 that proteins consist of polypeptide chains in which the individual amino acids are linked to one another through amide bonds formed between the carboxyl group of one amino acid and the α -amino group of the next amino acid (Fig. 1).

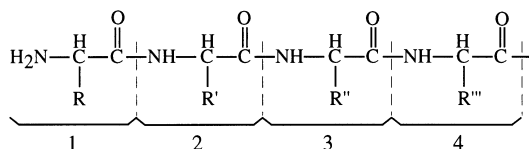


Fig. 1. R, R', R'', etc. stand for different side chains.

This proposal represented a unifying concept for all the preceding work on amino acids for more than a century. All the subsequent work showed without exception that the manner of linkage shown in Fig. 1 is universally used in proteins. Only in the case of certain naturally occurring peptides, not synthesized on ribosomes, may there be exceptions. For example, glutathione is γ -glutamyl cysteinyl-glycine, with the γ -carboxyl group of glutamic acid linked to the amino group of cysteine. Further, all the nineteen chiral amino acids (glycine is nonchiral) occurring in proteins have the *L*- α configuration at the chiral carbon. D-amino acids are found only in certain naturally occurring polypeptides, such as gramicidins and bacitracin. The manner of their biosynthesis is quite different from the ribosome-based protein biosynthesis (Chapter 7).

Determination of the Amino Acid Composition of Proteins

The methods used for many decades, before the advent of partition chromatography in 1942, for the determination of amino acid composition were laborious and time-consuming in the extreme.^{2,6} The invention of partition chromatography on a filter paper by Martin and Synge⁷ effectively brought about a revolution in the task of separation of amino acids, a necessity for the determination of composition. Some years later, column-chromatographic methods using starch and later ion exchange resins were developed by Moore and Stein, and these culminated in complete automation of determination of the amino acid composition of proteins. A commentary by Tristram in 1974, cited below, highlighted the revolution that had taken place in the field.

“The revolution brought about by the pioneering work of Moore and Stein, Martin and Synge and others has meant that we no longer require 100 g samples and months of effort to analyse a protein. To carry out even a partial amino acid analysis of a protein it was necessary to hydrolyse 100 g of protein. It was then possible to estimate glutamic acid,

aspartic acid, arginine, histidine, lysine, tyrosine, tryptophan, serine and threonine, cysteine and methionine. The remainder of the amino acids could be estimated by distillation of the methyl esters but the results never inspired any confidence. Many of the methods involved the removal of metal ions (e.g. Ba^{2+} , Ag^+ , Ca^{2+} , Hg^{2+}) or anions (e.g. SO_4^{2-} , Cl^-) by precipitation methods and the precipitates, such as barium sulfate, adsorbed amino acids very tenaciously. One story common to us all was the experience of Leavenworth, in the laboratory of H. B. Vickery, who precipitated BaSO_4 in the presence of amino acids and spent 6 months removing the last traces of amino acids from the precipitate.”

— G. R. Tristram, “The Way It Was,” *Trends in Biochemical Science*, February 1979

Identification of the N-Terminal and C-Terminal Amino Acids in Peptides and Proteins

The selected paper in this chapter reviews in detail the work on this problem up to about 1951. The introduction of 2,4-dinitrophenyl fluorobenzene as a reagent for labeling and identifying the N-terminal groups by formation of the dinitrophenyl derivatives by Sanger was a landmark. Later, modifications to form fluorescent derivatives (dansylation and fluorescamine) increased the sensitivity.^{8,9}

Determination of the Sequence of Amino Acids in Peptides and Proteins

Syngé in 1943³ provided a comprehensive review of all the earlier work on sequence determination in which partial hydrolysis of proteins had been the key step. It was clear that despite enormous efforts by a large number of protein chemists in the first forty years of the twentieth century, there had been no definitive progress in the sequencing of amino acids in proteins. Then the situation changed very rapidly. While the most decisive factor was the introduction of chromatographic methods for amino acid analysis, there was also the discovery of a large number of naturally occurring biologically active polypeptides in the 1940's. These peptides were much shorter than the smallest proteins and they included hormones such as vasopressin and oxytocin, growth factors and several families of antibiotics, such as gramicidins, tyrocidins, bacitracin and polymyxins. The polypeptides served as models for structural work on proteins. The first successful application of the rapid analytical methods described above was in the determination of the sequence of gramicidin S, a cyclic decapeptide (dimeric form of the same pentapeptide sequence), by Consden, Gordon and Martin¹⁰ As they wrote: “We report here a study of the products of partial hydrolysis of gramicidin S by mineral acid acting at low temperature. Identification of these products has permitted an unequivocal formulation of the sequence of the amino acid residues in this compound.” This work was soon to be followed by the epoch-making amino acid sequence determination of insulin by Sanger.¹¹

Stepwise Degradation of Polypeptides from the α -Amino End

The method developed by Edman in 1950¹² was also an important chemical development, which found very wide application in the sequencing of proteins and continues to be applied in the current microsequencing technique, which had already been completely automated by the seventies.

Selective Cleavage of Proteins to Form Relatively Large Fragments

While partial chemical hydrolysis and proteolysis by enzymes (trypsin, chymotrypsin, V8 and many other specific endo- and exopeptidases) were the main tools for a long time for fragmentation of proteins, the introduction of cyanogen bromide for selective cleavage at methionine sites by Witkop¹³ proved to be a particularly useful method for preparation of larger polypeptide fragments as intermediates in sequencing work.

Chemical Synthesis of Polypeptides

Emil Fischer, a very great bio-organic chemist, favored in all his work the synthetic approach to the study of biologically important macromolecules. His Nobel Prize in 1902 had been for the elucidation of the structure of glucose. The same year he enunciated the polypeptide theory of protein structure and focussed on methods for the synthesis of polypeptides. His efforts took him as far as the linking of nineteen glycines in a polypeptide chain. The general requirements in organochemical synthesis were: (1) suitable protection of the α -amino group of the first amino acid and of the carboxyl group of the second amino acid (usually by esterification), thus leaving only the amino group free for condensation; (2) activation of the carboxyl group of the amino acid carrying the protected amino group to form a peptide bond and, finally, removal of the protecting groups. In collaboration with one of his brilliant students, Bergmann, he developed the use of the ethoxycarbonyl group for protection of the α -amino group. An important advance in the methodology followed later in 1932 from Bergmann and Zervas,¹⁴ who introduced the benzyloxycarbonyl (or carbobenzyloxy) group for the protection of the α -amino group. The group could be removed by hydrogenolysis under mild conditions. Subsequently, this strategy with additional innovations led to the first total synthesis of the pituitary hormones, vasopressin and oxytocin, by du Vigneau and his colleagues.¹⁵

Polypeptide synthesis on insoluble polymer supports was pioneered by Merrifield.¹⁶ The method, which has been automated for many years, has facilitated the synthesis of polypeptides enormously.

The 1955 Gordon Conference on Proteins: A High Point Signaling the Beginning of a New Era in Protein Chemistry

In retrospect, this conference marked the end of one era and the beginning of a new one. A group photograph of the attendees is on the following page. Most of the active and leading protein chemists of the period attended. Putnam, the Chairman of the meeting, was sequencing the Ben Jones proteins. Sanger had completed the amino acid sequences of the A and B chains of insulin somewhat earlier but had now carefully defined the location of the disulfide bridges within and between the two chains.¹¹ Anfinsen and Moore and Stein, the pioneers in ribonuclease A research,¹⁷⁻¹⁹ reported progress in studies of the protein and, indeed, the meeting was characterized by reports on a number of additional sequences of biologically interesting polypeptides. Of particular note was the adrenocorticotrophic hormone (ACTH) sequence by Bell.

Thus, this was a meeting that highlighted the success and maturity in primary sequence determination of proteins. Further elaborations of techniques and methodologies that continued to accelerate work in the field were soon to come. Automation of procedures was also to follow later.

Three-Dimensional Crystal Structures of Proteins

Perutz in Cambridge (England) had pioneered X-ray study of hemoglobin crystals, and after many years' work he was able to use successfully, in 1953, the isomorphous heavy atom replacement method, which had been discovered by Hodgkin. In the mid-forties, Kendrew had joined Perutz, starting on myoglobin. In early 1962, they reported on the three-dimensional structures of the two proteins.^{20,21} Perutz kindly provided the structure of haemoglobin (Fig. 2).

The work of Perutz and Kendrew indeed represented another landmark in protein science in that it opened a new door to understanding how proteins really did their job. Shortly after, David Phillips described the first three-dimensional structure of an enzyme, the hen lysozyme.²² Since then, very many water-soluble proteins have been crystallized and their structures described. It is difficult to find a current issue of scientific journals without a 3D structure of a protein in it.

Three-Dimensional Structures of Proteins in Solution by NMR

Criticism used to be made that a crystal structure gives a static picture, being perhaps a selection of a structure out of multiple possibilities. On the whole, this has proved not to be true, and X-ray structures of proteins have been the source of most of our structural information. However, a complementary approach to protein structure in solution was pioneered particularly by Wuthrich.²³ While there is still restriction on the size of water-soluble proteins that can be studied, the method has in most cases provided confirmation



Frank W. Putnam	G. B. Brown	H. G. Khorana	R. G. Shepherd	R. Hofman	David Lipkin	H. Fraenkel-Conrat	Paul H. Bell
P. Desnuelle	W. F. White	William H. Stein	W. A. Schroeder	F. Sanger	C. B. Anfinsen	A. M. Michelson	R. E. F. Matthews
J. J. Fox	E. C. Taylor	Kurt C. Stern	C. H. W. Hirs	R. R. Redfield	I. R. Hooper	W. E. Cohn	Aaron Bendich
M. P. Drake	R. L. Slushyheimer	A. Koch	J. H. Hunter	J. T. Yang	J. W. Rowen	T. P. King	B. R. Baker
T. D. Kroner	Donald Acker	W. W. Bromer	V. E. Shashoua	R. S. Tipson	D. I. Magrath	W. A. Landmann	I. Goodman
R. A. Brown	L. F. Cavalieri	S. Timasheff	J. Gergely	H. B. Pahl	W. R. Middlebrook	E. L. Grinnan	R. Cubiles
H. Van Vunakis	M. Seraydarian	C. A. Thomas, Jr.	John Waterlow	J. F. Scott	Franz Meyer	M. A. Mitz	Max V. Sigal, Jr.
A. Munck	J. Leggett Bailey	Julius Schultz	J. B. Weisiger	Charles M. Apt	M. Rosoff	W. Hausmann	B. Roth
M. Bossman	M. Liss	J. R. Vaughan, Jr.	J. F. Lenney	Milton E.	Getzendaner	M. L. Groves	R. H. Locker
Charles A. Zittle	J. Casel	R. H. Hall	G. Gamov	C. A. Hackethal	T. J. Bardos	M. P. Gordon	D. B. Smith
Arnold H. Schein	Charles A. Dekker	G. Kalnitsky	R. S. Baker	H. Heymann	E. B. Taft	J. S. Dixon	N. B. Kurnick
C. Miller	J. Zamenhof	G. Schmidt	E. Goldwasser	W. Ritschard	E. J. Schantz	S. A. Morrell	G. C. Butler
E. H. Sakal	Lew Cunningham	K. Seraydarian	K. Schmid	K. Miller	Milton Levy	J. Bourdillon	

Gordon Conference "Proteins and Nucleic Acids" (New Hampton School, New Hampton, New Hampshire, USA; June 27–July 1, 1955)

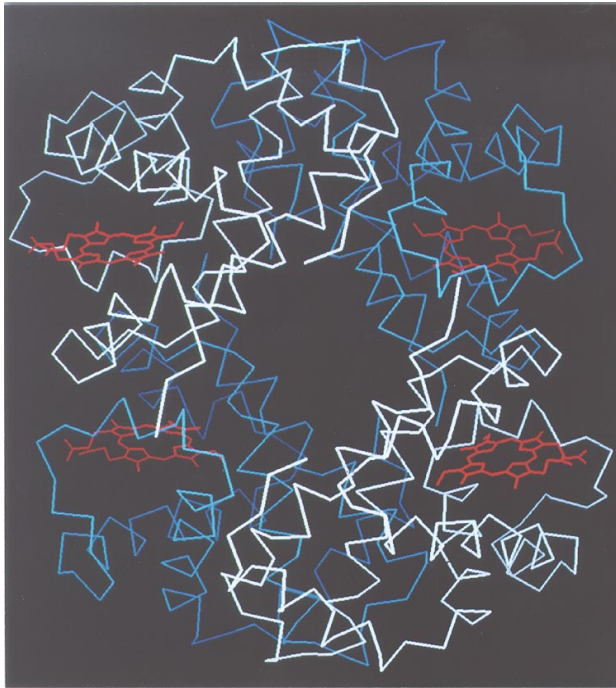


Fig. 2. Course of the four polypeptide chains with the four hemes in human deoxyhemoglobin. (Courtesy of Dr. M. Perutz.²⁰)

of structures derived from X-ray analysis. In more recent years, NMR techniques, used both in solution and in the solid state, have emerged as powerful tools in studies of protein structure and dynamics. An increasing number of applications are also being made to studies of membrane proteins.

The Second Revolution in Protein Sequencing: Sequences of Membrane Proteins Deduced from the Corresponding Gene Sequences

Essentially all of the protein sequencing work until the 1970's had been with water-soluble proteins. Straightforward application of the methods developed for the water-soluble proteins continued to be hampered by the difficulties of working with the hydrophobic integral membrane proteins. The development of recombinant DNA methods, resulting in cloning and sequencing of genes corresponding to membrane proteins, made it possible to deduce amino acid sequences from the gene sequences. Thus sequences of an increasing number of ion channel proteins, cell surface receptors, and transporters all began to be available with increasing rapidity. Since the early 1980's, the technology of gene sequencing has indeed resulted in a new revolution in studying a large variety of membrane proteins of importance in the emerging field of Molecular Neurobiology.

Site-Specific Mutagenesis in Protein Structure-Function Studies

The main approach to chemical and physicochemical studies of enzymes and proteins, in general, up to about 1980 involved chemical modifications of the available functional groups, particularly the sulfhydryl groups of cysteines and terminal α -amino groups or ε -amino groups of internal lysine residues. Indeed, over the years, this general approach resulted in a variety of methods that formed the topics of books that appeared from time to time.

Lederberg wrote in 1959:²⁴ “The *ignis fatuus* of genetics has been the specific mutagen, the reagent that would penetrate to a given gene, recognize it, and modify it in a specific way.” Within two decades this dream had been fulfilled. Site-specific mutagenesis at the gene level as directed by short oligodeoxyribonucleotides was developed by Smith, a former colleague of this author in British Columbia.²⁵ The first mutagenic experiment was reported on the bacteriophage $\phi \times 174$ bacteriophage protein in 1976.²⁶ Site-directed mutagenesis of proteins, especially in conjunction with three-dimensional analysis of protein structures, indeed introduced the next era in the study of protein structure and function.

Mass Spectrometry in Protein Sequencing and Protein-Protein Interactions

Mass spectrometry is now an important method for the characterization and determination of the primary structures of proteins. In the late 1950's, when Klaus Biemann started his pioneering work, it was still necessary to use volatile derivatives of peptides which were obtained by reduction of the carbonyl and the carboxyl groups to polyaminoalcohols. For two decades this methodology was used chiefly to sequence peptides not amenable to the Edman degradation, such as N-terminally blocked or very hydrophobic ones.²⁷ It culminated, in combination with the Edman method, in the determination of the primary structure of the transmembrane protein bacteriorhodopsin.²⁸

Since the early 1980's, the advent of new ionization techniques, such as “fast atom bombardment,” “matrix-assisted laser desorption” and “electrospray,” has made it possible to determine the exact molecular weight (to $\pm 0.1 - 0.001\%$) of peptides and intact proteins. Peptide sequence can be determined by “tandem mass spectrometry.”²⁹ The great power of mass spectrometry lies in its high sensitivity (pico- to femtomoles), high accuracy and ability to determine molecular weights in mixtures, without prior separation into the individual components. These characteristics have made it possible to identify proteins representing spots on a two-dimensional gel of a cell extract, by using the ever-increasing protein and gene databases.³⁰

Bibliography

1. H. Ritthausen, *The Proteins of the Cereals, Legumes and Oil Seeds* (Bonn, 1872).
2. H. B. Vickery, “Evidence from Organic Chemistry Regarding the Composition of Protein Molecules,” *Ann. N. Y. Acad. Sci.* **41**, 87 (1941).

3. R. L. M. Synge, "Partial Hydrolysis Products Derived from Proteins and Their Significance for Protein Structure," *Chem. Rev.* **32**, 135 (1943).
4. E. Fischer, *Chem. Ztg.* **26**, 939 (1902).
5. F. Hofmeister, *Ergebn Physiol.* **1**, 159 (1902).
6. A. C. Chibnall, *Proc. R. Soc.* **B131**, 136 (1942); A. C. Chibnall, Wm. W. Rees and E. F. Williams, *Biochem. J.* **37**, 372 (1943).
7. A. J. P. Martin and R. L. M. Synge, *Adv. Protein Chem.* **2**, 1–83 (1945).
8. B. S. Hartley, *Biochem. J.* **80**, 36P (1961).
9. S. Udenfriend, S. Stein, P. Boblen, W. Daairman, W. Leimgruber and M. Weigele, *Science* **178**, 871 (1972).
10. R. Conden, A. H. Gordon, A. J. P. Martin and R. L. M. Synge, *Biochem. J.* **41**, 596 (1947).
11. F. Sanger, in *Selected Papers of Frederick Sanger*, eds. F. Sanger and M. Dowding (World Scientific, Singapore, 1996).
12. P. Edman, *Acta Chem. Scandanavica* **4**, 283 (1950).
13. B. Witkop, *Science* **162**, 318–326 (1968).
14. M. Begmann and L. Zervas, *Ber.* **65**, 1192 (1932).
15. V. du Vigneaud, *Les Prix Nobel en 1955* (Norstedt and Sons, Stockholm).
16. R. B. Merrifield, *Les Prix Nobel en 1984* (Norstedt and Sons, Stockholm).
17. C. B. Anfinsen, "Studies on the Principles That Govern the Folding of Protein Chains," in *Les Prix Nobel en 1972* (1972), pp. 101–119.
18. S. Moore, "The Chemical Structures of Pancreatic Ribonuclease and Deoxyribonuclease," in *Les Prix Nobel en 1972* (1972).
19. W. H. Stein, "The Chemical Structures of Pancreatic Ribonuclease and Deoxyribonuclease," in *Les Prix Nobel en 1972* (1972).
20. M. F. Perutz, *Les Prix Nobel en 1972* (1962).
21. J. C. Kendrew, *Les Prix Nobel en 1962*.
22. C. C. F. Blake, D. F. Koenig, G. A. Mair, A. C. T. North, D. C. Phillips and V. R. Sarma, *Nature* **206**, 757–763 (1965).
23. K. Wuthrich, "NMR in Structural Biology," in *Selected Papers of K. Wuthrich* (World Scientific, Singapore).
24. J. Lederberg, *Les Prix Nobel en 1958*.
25. M. Smith, *Les Prix Nobel en 1993*.
26. C. A. Hutchison, III, S. Phillips, M. H. Edgell, S. Gillam, P. Jahnke and M. Smith, "Mutagenesis at a specific position in a DNA sequence," *J. Biol. Chem.* **253**, 6551–6560 (1978).
27. K. Biemann, *Protein Science* **4**, 1920–1927 (1995).
28. H. G. Khorana *et al.*, *Proc. Natl. Acad. Sci. USA* **76**, 5046–5050 (1979).
29. K. Biemann, *Annu. Rev. Biochemistry* **61**, 977–1010 (1992).
30. A. Shevchenko *et al.*, *Proc. Natl. Acad. Sci. USA* **93**, 14440–14445 (1996).

Selected Paper

1. "Structural Investigation of Peptides and Proteins," H. G. Khorana, *Quarterly Reviews* I. No. 4, 340–357 (1952).