1958

Stanford Moore and William H. Stein, 1957

The Rockefeller University

DETERMINATION OF THE STRUCTURE OF PROTEINS: STUDIES ON RIBONUCLEASE*

STANFORD MOORE AND WILLIAM H. STEIN
The Rockefeller Institute for Medical Research, New York, New York

THE most recent Harvey Lecture that dealt to a considerable degree with the present subject, the chemical structure of proteins, was delivered in 1935 by our former chief, the late Dr. Max Bergmann.[1] At that time, as a result of the efforts of many investigators, Dr. Bergmann himself included, proteins had come to be recognized as large molecules built up of hundreds of amino acid residues of about eighteen varieties joined to one another by peptide bonds to form long peptide chains. The peptide bonds, formed by the elimination of a molecule of water between the amino group of one amino acid and the carboxyl group of the next, are separated by $-CH<$ groups that bear the side chains of the amino acids. The side chains in a protein may be aliphatic, aromatic, or heterocyclic; they may contain a hydroxyl, a sulfhydryl, or a thio-ether group; they may be neutral, basic, or acidic. It has been generally believed that it is primarily the difference in the nature and arrangement of these groups that determines the physical, the chemical, and the biological properties of each individual protein.

There are two principal aspects of the structure of proteins to be considered: the chemical composition of the molecule, with which we are concerned in this discussion, and the physical structure, or orientation of the atoms in space, knowledge of which is also essential to a full understanding of proteins and enzymes and how they function.

Although the questions posed by the chemical structure of proteins were recognized for many decades, it has been only within the past ten years that the evolution of new techniques has made possible rapid progress toward the solution of some of these prob-

* Lecture delivered February 21, 1957

lems. Beginning with the introduction of the chromatographic method into protein chemistry by Martin and Synge, and aided by countercurrent distribution, as developed by Craig, and zone electrophoresis, as devised by Wieland, Durrum, and Tiselius and

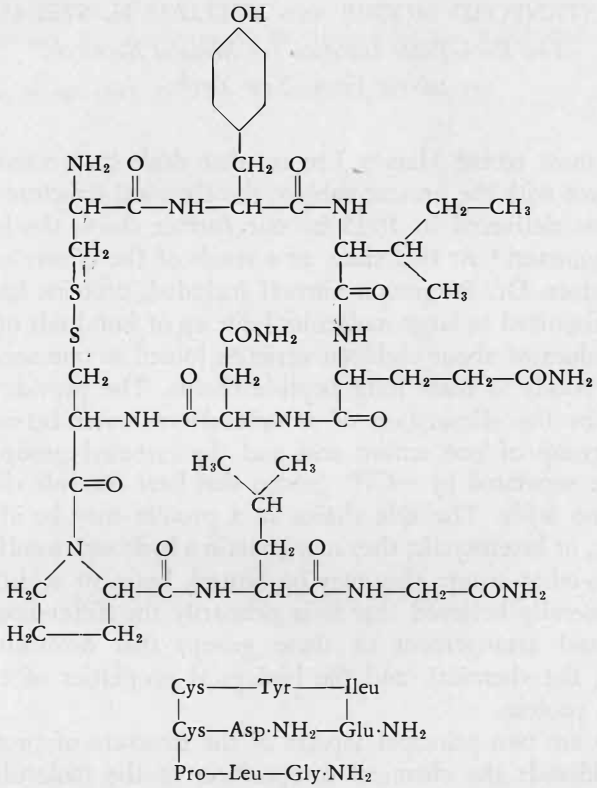


FIG. 1. The structure of oxytocin from beef pituitary glands (du Vigneaud and associates[2]).

Kunkel, knowledge of peptides and proteins has been accumulating at an increasing rate. One important milestone is the synthesis of the peptide hormones vasopressin and oxytocin by du Vigneaud and his group, an accomplishment described in a Harvey Lecture two years ago.[2] In Fig. 1 is shown the formula of oxytocin, a peptide containing 9 amino acid residues and possessing a molecular weight of 1007. The full chemical structure of the side chains

of the amino acids can be represented in the figure, as can the location of the disulfide bond that holds the peptide chain in the form of a ring. When much larger molecules are portrayed, the formulas of the amino acids usually need to be abbreviated, as is now customary, by the use of the first three letters of their names, in the manner indicated in the lower part of the figure. Du Vigneaud's research has many facets of interest, but of particular significance for the subject of this lecture is the fundamental support that it has furnished for the peptide theory. The structure of the hormone was deduced from degradative experiments, and a peptide possessing this structure was synthesized and found to exhibit full biological activity.

Another classic achievement is the elucidation, by Sanger and his colleagues,[3,4] of the chemical structure of insulin, a protein of 51 amino acid residues and a molecular weight of 5733. The formula, which has been derived through research carried out over the past ten years, is shown in Fig. 2 in abbreviated form. Insulin is the largest complex organic molecule for which a structural formula can be written. Sanger's brilliant achievement has been followed by several notable advances in other laboratories. These include the working out of the structure of the adrenocorticotropic hormones (ACTH), which have 39 residues, by Bell and associates,[5] Li *et al.*,[6,7] and White and Landmann;[8] the determination of the structure of glucagon, which has 29 residues, by Behrens and colleagues;[9] and the elucidation of the structure of the melanocyte stimulating hormone, which has 18 residues, by Harris and Roos,[10] and Geschwind, Li, and Barnafi.[11] Structural analysis of the more typical proteins having molecular weights in the range 10,000 to 100,000 still lies ahead. Even this goal is in sight, however, and in order to illustrate how it may be approached, we should like to take you with us as far as we have gone toward unraveling the structure of one such protein of 124 amino acid residues, the enzyme ribonuclease.

At the outset, it is a pleasure to emphasize the fact that this research has been carried out jointly with Dr. C. H. W. Hirs. Through his extensive and skillful experimentation, he is directly responsible for a large part of the data on the structure of ribonuclease presented in this lecture.

Ribonuclease was first described in 1920 by Jones,[12] who showed that there was present in beef pancreas a relatively heat-stable enzyme, capable of digesting yeast nucleic acid. Dubos and Thompson[13] partially purified the enzyme some eighteen years

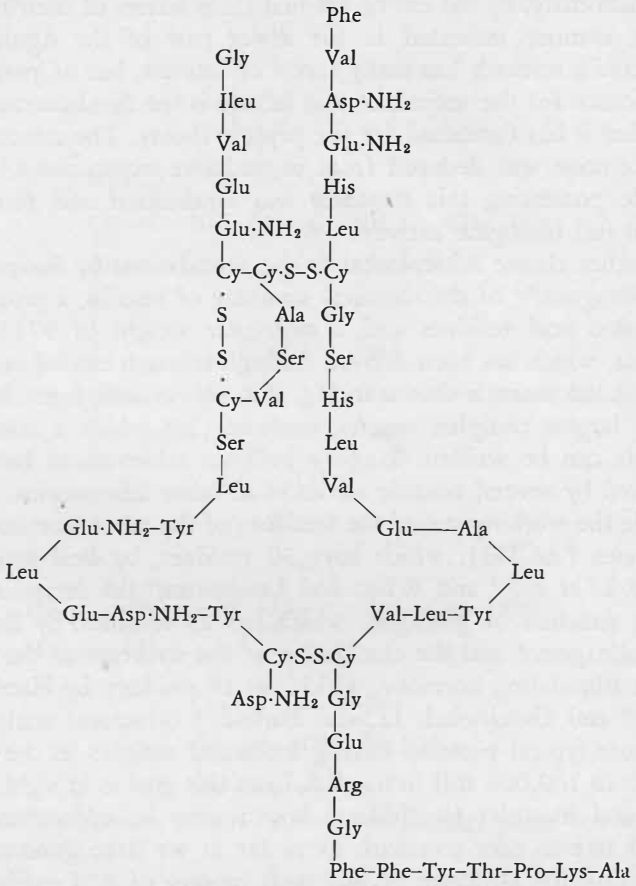


FIG. 2. The structure of beef insulin (Sanger and associates[3,4]). The first three letters of the names of the amino acids are used as abbreviations.

later, and in 1940, Kunitz[14] described the isolation of bovine ribonuclease in crystalline form. It proved to be a relatively small molecule as proteins go, having a molecular weight of about 14,000. The isoelectric point was shown to be near pH 8, and the

analyses of Brand[15] revealed the presence of all of the usual amino acids, with the significant exception of tryptophan. Recently the specificity of the enzyme has been studied in several laboratories, notably those of Todd, of Cohn, and of Markham,[cf.16] and it has been found to hydrolyze the large nucleic acid molecule at the phosphate bond of pyrimidine, but not of purine nucleotides (Fig. 3). For precise work it is preferable to use substrates of lower molecular weight than RNA and of known structure, and

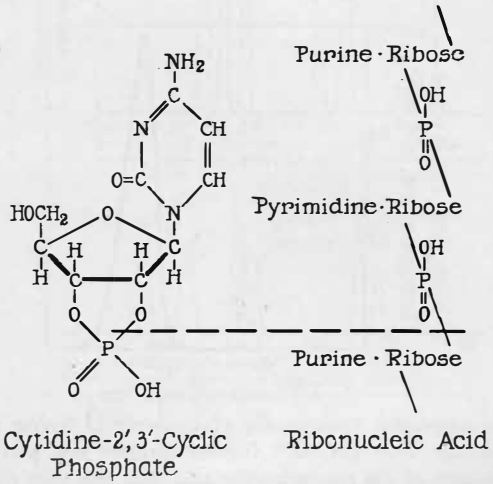


FIG. 3. The enzymatic specificity of ribonuclease.[16] The types of phosphate ester bonds of pyrimidine nucleotides cleaved by the enzyme are indicated by the dashed line.

fortunately it has been found that ribonuclease hydrolyzes the ester phosphate bond in the simple synthetic pyrimidine nucleotides, 2′,3′-cyclic cytidylic and uridylic acids, as illustrated in Fig. 3.

All of these properties of ribonuclease rendered it a favorable protein for structural analysis. Since such research is not possible unless there is available a highly purified preparation of the protein with which to work, it is fortunate that ribonuclease has been one of the first proteins the homogeneity of which has been studied by the improved chromatographic techniques devised in the last decade. Martin and Porter[17] in England employed liquid-

liquid chromatography for this purpose, and Dr. Hirs in our laboratory used columns of the polycarboxylic acid resin, IRC-50.[18] Both methods of chromatography were capable of removing detectable amounts of impurity from crystalline ribonuclease. Evidence was also obtained for the occurrence in pancreas of a small amount of a second protein possessing ribonuclease activity. The predominant enzyme, ribonuclease A, was prepared in chromatographically homogeneous form, as can be seen in Fig. 4, and it is this purified preparation that furnished the starting material for

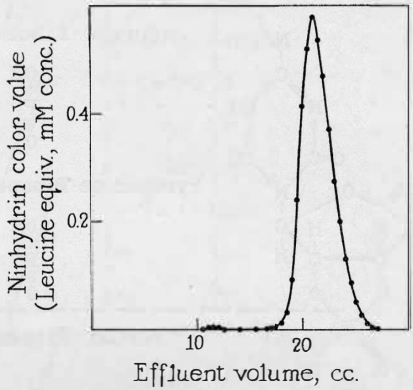


FIG. 4. Chromatographic homogeneity of a sample of bovine ribonuclease A prepared by passage over IRC-50.[18] Chromatography was performed with a 0.9 × 30-cm. column of the polycarboxylic acid resin and with phosphate buffer at pH 6.47 as the eluent.

the next step in the structural analysis, namely, the determination of the amino acid composition.[19] This was done by chromatographing the hydrolyzate from 2 mg.-samples of ribonuclease A on buffered columns of the sulfonated polystyrene resin, Dowex 50-X4,[20] as is illustrated by the effluent curve shown in Fig. 5. In this curve, and in others somewhat like it that will be discussed later, the horizontal axis gives the volume of eluent that has passed through the column. The vertical axis gives the concentration of amino acid in each small fraction of the effluent from the column, as determined by a photometric ninhydrin method.[21] The area of each peak on the curve is proportional to the quantity of amino acid in the original mixture.

The results of this chromatographic analysis were, in general, in good agreement with the earlier findings of Brand, except for the values for alanine and leucine. Assuming a molecular weight of about 14,000, integral molar ratios were obtained for most of

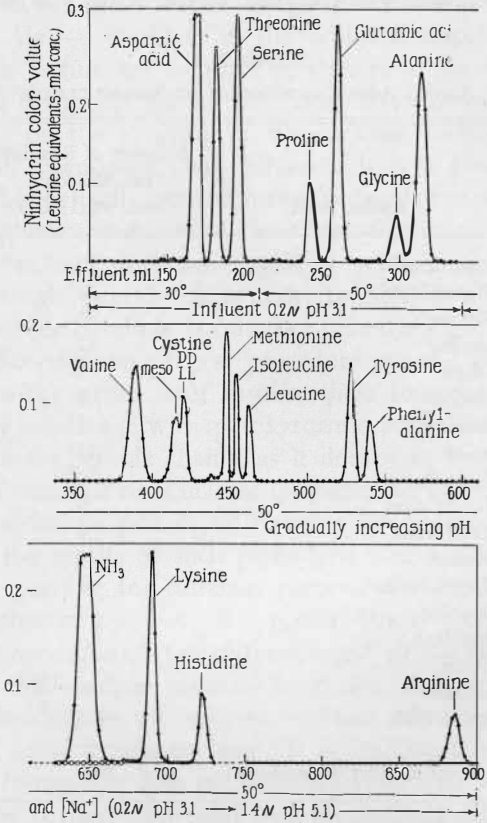


FIG. 5. Chromatographic determination of the amino acid composition of an acid hydrolyzate of 2 mg. of ribonuclease A.[19] The analysis was performed with a 0.9 × 150-cm. column of Dowex 50-X4 and eluents of the indicated pH and molarity.

the amino acids. The analytical results have had to be revised slightly in the light of subsequent structural work,[22] which has indicated that the original estimates for the decomposition during hydrolysis of aspartic acid and of proline led to values that were

high by one residue each. The empirical formula for ribonuclease in terms of amino acid residues (rather than the elements C, H, and N) is given in Table 1. Such a formula is useful as a first step in the determination of the structure of a protein. Note that there are 14 residues of the basic amino acids, 4 arginines and 10

TABLE 1

AMINO ACID COMPOSITION OF RIBONUCLEASE*

Amino Acid	Number of Residues per Molecule (mol. wt. 13,683)
Aspartic acid	15
Glutamic acid	12
Glycine	3
Alanine	12
Valine	9
Leucine	2
Isoleucine	3
Serine	15
Threonine	10
Half-cystine	8
Methionine	4
Proline	4
Phenylalanine	3
Tyrosine	6
Histidine	4
Lysine	10
Arginine	4
Amide NH_3	(17)
Total number of residues	124

* Ribonuclease from beef pancreas.[21,22]

lysines; that there are 8 half-cystine residues; and that several of the amino acids—glycine, phenylalanine, leucine, and isoleucine, for example—are present to the extent of only 2 to 4 residues per molecule. The total number of amino acid residues is 124, which, with 17 amide groups, and assuming the 124 residues to be present in one peptide chain, leads to a calculated minimal molecular weight of 13,683.

The fact that ribonuclease does contain but a single peptide chain was shown by Dr. C. B. Anfinsen and his colleagues at the National Heart Institute in Bethesda. By the use of the dinitrophenyl technique of Sanger they succeeded in demonstrating that the first four residues of this single chain had the sequence Lys.Glu.Thr.Ala.—, starting at the amino-terminal end.[23] Included in this lecture are frequent references to the work of Dr. Anfinsen and his group, for his laboratory and our own have both been working on the structure of ribonuclease, although our approaches have been somewhat different. It is a pleasure to acknowledge the helpful interchange of information that has occurred throughout the course of these investigations.

Since ribonuclease does not appear to contain any sulfhydryl groups, the single peptide chain must be held in a folded configuration by four disulfide bonds that join the eight half-cystine residues. Following the general procedure utilized by Sanger in his studies on the structure of insulin, these four disulfide bonds were split by oxidation with peroxyformic acid. This procedure does not separate peptide chains, as it does with insulin, but the single folded chain of ribonuclease is opened up into a form more readily hydrolyzed by proteolytic enzymes. In addition, the interpretation of the results of such proteolytic cleavage is easier if it is no longer possible for different parts of the peptide chain to be held together by the —S—S— bonds. Under properly chosen experimental conditions,[24] peroxyformic acid almost quantitatively converts the half-cystine residues to cysteic acid residues, and methionine residues to the sulfone, without affecting any of the other amino acids in ribonuclease. It is important to note that ribonuclease fortunately does not contain tryptophan, which would be oxidized by peroxyformic acid with the formation of a number of products.

At this point we are ready to begin the determination of the sequence of the 124 amino acid residues in the chain of oxidized ribonuclease. This chain is too large to be studied as such; it must first be reduced to fragments of a more manageable size. In the work of Sanger and his colleagues on the sequence of the amino acids in the oxidized A and B chains of insulin, partial hydrolysis with HCl was the principal method used to split the long chains

into peptides small enough for study. The products of this hydrolysis were then separated by paper chromatography. As Sanger[25] pointed out, however, it is doubtful whether hydrolysis with concentrated HCl, which splits the peptide chain in a more or less random manner, could be used effectively with larger molecules. The complex mixture of peptides which would result from the partial acid hydrolysis of a protein of the size of ribonuclease would be extremely difficult to fractionate. Moreover, the deduction of a unique structure for the parent molecule becomes progressively more difficult as the number of small fragments increases. It was decided, therefore, to use proteolytic enzymes which, because of their specificity, cleave a protein into a limited number of fragments. Enzymatic hydrolysis was used by Sanger and his colleagues for the elucidation of parts of the structure of insulin, and was subsequently employed by Bell and associates[5] and by Li and his group[6,7] in their work on the structure of ACTH.

Of the readily available proteolytic enzymes, trypsin possesses the sharpest specificity, as was demonstrated in the classic studies of Bergmann, Fruton, and Hofmann[cf.26] carried out with the aid of synthetic substrates. On the basis of its specificity, trypsin might be expected to hydrolyze oxidized ribonuclease at the peptide bonds involving the carboxyl groups of the 10 lysine and 4 arginine residues present in the molecule, to yield 14 or 15 fragments. Whether or not free lysine would be cleaved from the amino-terminal position could not be foretold with assurance, nor could the probability of the occurrence of trypsin-resistant lysyl-prolyl bonds be predicted.

When about 200 mg. of oxidized ribonuclease were hydrolyzed by trypsin and the products formed were separated by ion exchange chromatography on columns of Dowex 50-X2, the results shown in Fig. 6 were obtained.[22] Columns were used in these studies in preference to paper chromatography, because quantitative data were desired, and because it was necessary to obtain 10-mg. to 50-mg. quantities of the larger and more complicated peptides in order to permit further structural work to be done. As a first step in this structural work, quantitative amino acid analyses were performed by ion exchange chromatography on an

aliquot of the material in each peak. The results were particularly informative in deciding whether or not a single peptide had been obtained, for a reasonably well purified peptide should contain only a limited number of amino acids in quantities that give integral molar ratios. The amino acid composition of each peptide obtained in relatively pure form is shown in Fig. 6. The sequence

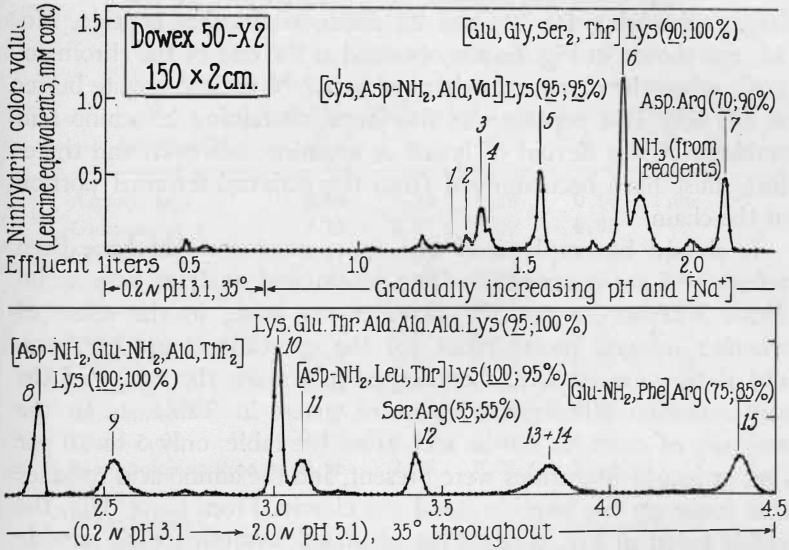


FIG. 6. Separation of the peptides present in a 20-hour tryptic hydrolyzate of 200 mg. of oxidized ribonuclease.[22] A 2 × 150-cm. column of Dowex 50-X2 was employed. An additional peptide (No. 16) was eluted after a change at 4.5 liters to 2 *N* sodium acetate buffer at pH 6.6. The abbreviation Cys represents a cysteic acid residue in the peptides from the oxidized protein.

of the residues in the brackets is not known. The numbers in parenthesis give the yield of each peptide obtained after 3 and after 20 hours of tryptic hydrolysis. On the basic of the specificity of trypsin, the basic amino acid, lysine or arginine, is assigned to the carboxyl-terminal position. It will be noted that peptide No. 10 contains two lysine residues. Inasmuch as DNP end-group analyses of this peptide placed one lysine at the amino-terminal position, it seems extremely probable that this fragment represents the amino-terminal sequence. Since the first four amino acids of

this sequence were shown by Anfinsen *et al.*[23] to be Lys.Glu.Thr.-Ala—, the finding of two more alanine residues and a lysine makes it possible to extend this sequence to seven amino acids.

The amino acid analyses indicated that mixtures of peptides were present in the zones marked 2, 3, 4, and 13+14, and upon rechromatography under slightly different conditions, four additional purified peptides were obtained. Three of them were large, containing 19, 21, and 22 residues. Another peptide, No. 16, not shown in Fig. 6, was obtained at the end of the chromatogram when the eluent was changed to a 2 *N* sodium acetate buffer of pH 6.6. This peptide was also large, containing 20 amino acid residues. It was devoid of lysine or arginine, however, and therefore must have been derived from the carboxyl-terminal portion of the chain.

It should be emphasized that the procedures that have been referred to are quantitative. The amino acid analyses help to indicate whether or not the peptides are pure, in the sense of whether integral molar ratios for the constituents are obtained, and if they are impure, the analyses may show the extent of the contamination. Illustrative data are given in Table 2. In the majority of cases, as can be seen from the table, only 5 to 10 per cent or less of impurities were present, and the amino acid residues that make up the peptide stand out clearly. From these data, the yields given in Fig. 6 could be calculated, assuming each peptide to occur only once in the parent protein.

As a result of the action of trypsin, thirteen peptides have been obtained in yields of 50 to 100 per cent, and these account for all of the 124 amino acid residues in oxidized ribonuclease. The compositions of these peptides, and the high yields in which they were isolated, indicate that transpeptidation by the enzyme is not occurring, and that such rearrangements will not complicate the interpretation of the results.

Clearly, the next step is to determine the order in which these thirteen peptides are linked together, and for this purpose, it is necessary to study the nature of the products formed when oxidized ribonuclease is cleaved at other sites by enzymes possessing different specificities. To effect this cleavage, chymotrypsin and pepsin have been used.

Chymotrypsin does not possess as attractively high a degree of specifiy as does trypsin. A number of types of bond are split at measurable rates, but there is a considerable preference for linkages involving the carboxyl groups of phenylalanine, tyrosine,

TABLE 2

AMINO ACID COMPOSITION OF PEPTIDE FRACTIONS OBTAINED FROM TRYPTIC HYDROLYSIS OF OXIDIZED RIBONUCLEASE[22]

	Peptide				
Amino Acid	O-Tryp 2	O-Tryp 4	O-Tryp 5	O-Tryp 6	O-Tryp 7
Aspartic acid	**3.06**	**3.15**	**1.10**	0.14	**1.06**
Glutamic acid	**2.10**	**2.07**	0.09	**0.95**	
Glycine	**0.99**	0.03	0.03	**1.00**	
Alanine	0.02	**2.09**	**1.01**		
Valine			**0.95**		0.07
Leucine		0.01	0.04	0.06	0.07
Isoleucine	**0.95**	0.08	0.08	0.03	0.03
Serine	**2.84**	**6.31**	0.24	**2.23**	0.17
Threonine	**2.85**	**0.96**	0.11	**0.91**	0.05
Cysteic acid	**2.01**	**1.13**	**1.00**	0.06	
Methionine sulfone	**0.95**	**2.98**	0.04	0.04	
Proline					
Phenylalanine					0.03
Tyrosine	**1.97**	**0.96**			
Histidine		**1.14**			
Lysine	0.06	**0.99**	**0.95**	**1.05**	0.07
Arginine	**0.97**	0.07	0.06		**0.94**
Ammonia (amide)	(**4.2**)	(**3.3**)	(**1.2**)	−0.1	0.0
Total number of residues*	19	21	5	6	2

* Totals of boldface figures given in round numbers.

methionine (and methionine sulfone), and leucine. As might be expected, therefore, and as is evident from Fig. 7, the peptide mixture obtained after 24 hours of hydrolysis by chymotrypsin[27] is more complex than that obtained with trypsin. The yields are also usually lower. Amino acid analyses indicated the presence of mixtures of peptides in the regions of the curve marked 4+5,

11 + 12, and 14 + 15 + 16. These were separated by rechromatography under different conditions of pH and temperature, as was done before. A total of eighteen purified peptides, containing from 3 to 27 amino residues each, were isolated from the chymotryptic hydrolyzate in yields of from 6 to 80 per cent. And finally, additional although less complete data were secured through the use of a third enzyme, pepsin.[28]

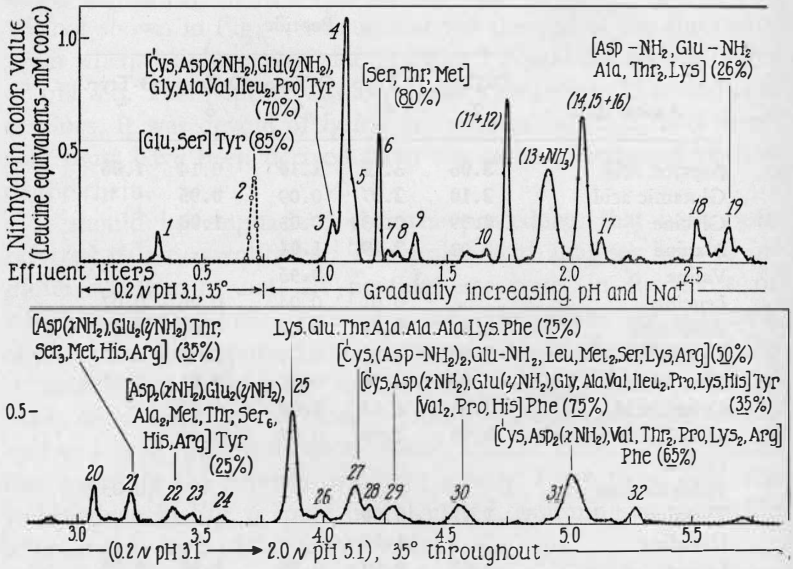


FIG. 7. Separation of the peptides in a 24-hour chymotryptic hydrolyzate of oxidized ribonuclease.[27]

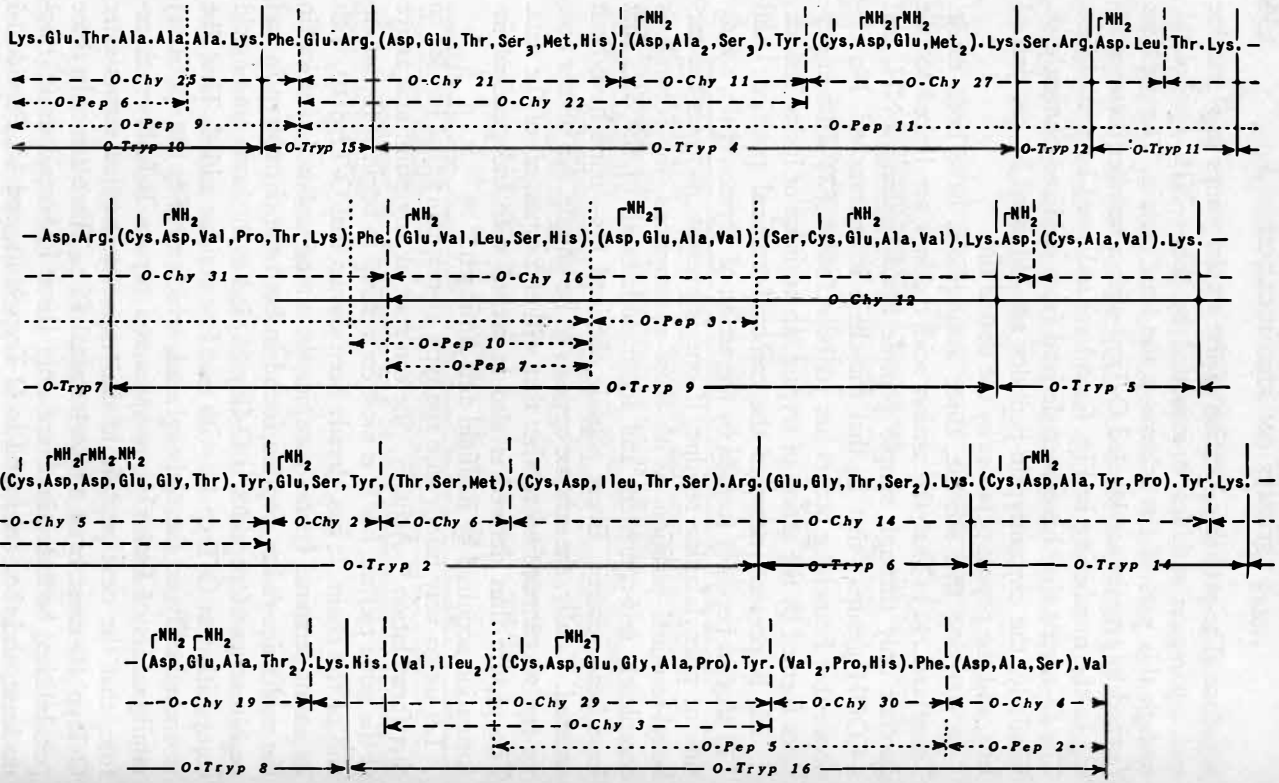
In all, the amino acid compositions of forty or more peptides formed by the action of trypsin, chymotrypsin, and pepsin on oxidized ribonuclease have been established. What this very considerable body of data can tell us about the structure of the molecule is shown in Fig. 8. Here is given a partial structural formula for oxidized ribonuclease that shows how the thirteen segments obtained by tryptic hydrolysis are linked together.

In Fig. 8 the chain has been divided into segments, which should be imagined as being joined end to end. The residues enclosed within the parentheses in this formula are of undetermined

sequence. The probable location of the amide groups in glutamine and asparagine residues is indicated by the —NH_2 symbol, although this part of the scheme is the least certain. The peptides formed by trypsin are labeled O-Tryp with a number corresponding to the numbering in Fig. 6, and are indicated by a solid line. The O- means they have been derived from oxidized ribonuclease. Similarly, the chymotrypsin-peptides are indicated by the dashed line, and the pepsin-peptides by the dotted line.

In deriving this scheme, three assumptions have been made. These are: (1) That the amino acid residues are linked to one another only through simple peptide bonds joining α-NH_2 and α-COOH groups—that is, that the chain is unbranched. No evidence that branching occurs has yet been found. (2) That in peptides formed by the action of trypsin, an arginine or lysine residue, if one is present, occupies the carboxyl-terminal position. And (3) that in peptides formed by the action of chymotrypsin, a tyrosine or phenylalanine residue, if one is present, occupies the carboxyl-terminal position. With these assumptions, the formula is completely self-consistent and accommodates all of the data that have been secured. Every peptide obtained fits—none have been omitted. Finally, the scheme appears to be unique. No matter how we try, we cannot find another that will encompass all of the data. The picture thus derived is also in accord with independent information acquired by Redfield and Anfinsen.[30]

Let us go very briefly into the type of reasoning that has led to this formulation.[27] Actually, one can start at any point and arrive at the same picture, but it is most convenient to begin at the amino end of the chain. It has already been shown that O-Tryp 10, with its amino-terminal lysine, occupies the amino-terminal position in the molecule. A key peptide found in both the chymotrypsin and pepsin series of peptides is O-Chy 25. It has the same amino acid composition as O-Tryp 10 with one Phe residue added. This Phe is undoubtedly on the carboxyl end, which satisfies the specificity requirements of both chymotrypsin and pepsin. It follows, therefore, that the next peptide in the trypsin series that comes after O-Tryp 10 must have amino-terminal Phe. There are only three possibilities, because there are only three Phe residues in ribonuclease, and they are found in O-Tryp 9, 15, and 16. Peptide 16


Lys.Glu.Thr.Ala.Ala.Ala.Lys.Phe.Glu.Arg.(Asp,Glu,Thr,Ser3,Met,His).(Asp,Ala2,Ser3).Tyr.(Cys,Asp,Glu,Met2).Lys.Ser.Arg.Asp.Leu.Thr.Lys.
O-Chy 25
O-Chy 21
O-Chy 11
O-Chy 27
O-Pep 6
O-Chy 22
O-Pep 9
O-Pep 11
O-Tryp 10
O-Tryp 15
O-Tryp 4
O-Tryp 12
O-Tryp 11
Asp.Arg.(Cys,Asp,Val,Pro,Thr,Lys).Phe.(Glu,Val,Leu,Ser,His).(Asp,Glu,Ala,Val).(Ser,Cys,Glu,Ala,Val).Lys.Asp.(Cys,Ala,Val).Lys.
O-Chy 31
O-Chy 16
O-Chy 12
O-Pep 3
O-Pep 10
O-Pep 7
O-Tryp 7
O-Tryp 9
O-Tryp 5
(Cys,Asp,Asp,Glu,Gly,Thr).Tyr.Glu,Ser,Tyr.(Thr,Ser,Met).(Cys,Asp,Ileu,Thr,Ser).Arg.(Glu,Gly,Thr,Ser2).Lys.(Cys,Asp,Ala,Tyr,Pro).Tyr.Lys.
O-Chy 5
O-Chy 2
O-Chy 6
O-Chy 14
O-Tryp 2
O-Tryp 6
O-Tryp 14
(Asp,Glu,Ala,Thr2).Lys.His.(Val,Ileu2).(Cys,Asp,Glu,Gly,Ala,Pro).Tyr.(Val2,Pro,His).Phe.(Asp,Ala,Ser).Val
O-Chy 19
O-Chy 29
O-Chy 30
O-Chy 4
O-Chy 3
O-Pep 5
O-Pep 2
O-Tryp 8
O-Tryp 16

is excluded because, as we have seen, it is devoid of basic amino acids and comes from the carboxyl end of the chain. O-Tryp 9 can be eliminated because it was found, by the DNP technique, not to contain amino-terminal Phe. This, by elimination, leaves O-Tryp 15 as the peptide following O-Tryp 10. Since O-Tryp 15 is a tripeptide, and Phe is at one end and Arg at the other, its structure is established and our accounting from the amino end has been increased to ten residues.

We now turn back to the chymotrypsin series of peptides. The placing of O-Tryp 15 requires the next chymotrypsin peptide to begin with the sequence Glu-Arg—. There are only four O-Chy peptides that contain glutamic acid and arginine, namely, O-Chy 14, 21, 22, and 27. O-Chy 14 and O-Chy 27 can be eliminated because if they were placed in this position, their amino acid composition would make it impossible for any remaining peptide of the trypsin series to follow O-Tryp 15. This leaves only O-Chy 21 and 22, which have so similar and unique an amino acid composition—each contains much Ser, one Met, and one His—that it can be decided that they come from the same part of the molecule. The two are placed in the manner indicated in Fig. 8. This uniquely defines the beginning of the trypsin peptide that follows O-Tryp 15. O-Tryp 4 is the only one that contains Met, His, and six serines and hence must be the next segment in the chain.

Without further pursuing this crossword puzzle type of occupation here, it can be stated that by alternating between the peptides of the trypsin and the chymotrypsin series, and occasionally calling on the pepsin experiments when necessary, it is possible to assemble the complete scheme shown in Fig. 8.

What amounts to almost clinching evidence in favor of this structure has come from experiments in which the larger trypsin

---

FIG. 8. A partial structural formula for oxidized ribonuclease.[27] The amino acids are considered to be arranged in a single chain with the segments in the figure joined end to end. The formula is based upon the study of the peptides liberated by tryptic, chymotryptic, and peptic hydrolysis of the oxidized protein. The tentative allocation of amide groups (—NH_2) has been based upon the most recent results, which have permitted definite placement of some of the asparaginyl and glutaminyl residues.[29] The residues in parentheses are of undetermined sequence.

peptides, Nos. 4, 9, 2, and 16, have been isolated and then hydrolyzed with chymotrypsin. In each case, peptides of the expected amino acid composition have been isolated from the hydrolyzate.

This partial structural formula of ribonuclease serves several useful purposes. In the first place, although it is only a partial formula, it still reveals much about the way the enzyme is put together. None of the unknown sequences are longer than eight amino acid residues, and many are shorter. In the second place, the formula serves as a guide for further work. It permits three more or less independent investigations connected with the detailed structure of the enzyme to proceed simultaneously. These are: the determination of the complete sequence of the amino acid residues in the oxidized protein; the determination of the positions of the four disulfide bonds in the native protein, or in other words, finding out which pairs of half-cystine residues were linked together before oxidation split them apart; and finally, the study of whether any particular portion of the entire large molecule is responsible for its enzymatic activity—briefly that is, whether ribonuclease has a so-called "active center," and if it does, where this "active center" is located. Let us take up these questions in order.

Because the manner is already known in which the forty or more peptides formed by enzymatic hydrolysis are linked, detailed sequence studies of all of these peptides are not required. Only those are selected for sequence work that do not overlap one another and that are of a suitable size. The thirteen peptides of the trypsin series were the starting products. The nine smallest could be studied directly, but the four larger ones, O-Tryp 4, 9, 2, and 16, were first hydrolyzed with chymotrypsin to give fourteen smaller fragments.

In determining the sequence of amino acids in all of these peptides, the phenylthiocarbamyl technique of Edman has been the principal procedure employed by Dr. Hirs. The reaction permits residues to be removed one at a time from the amino end of a peptide. The cleavage frequently is not quantitative, however, and carefully performed amino acid analyses at each stage of the degradation are required to permit clear-cut interpretation of the results.

In addition to the Edman procedure, Dr. Hirs has been using two enzymatic methods of degradation. These employ carboxypeptidase, which splits amino acids from the carboxyl end of peptides, and leucine aminopeptidase, prepared according to Spackman, Smith, and Brown,[31] which acts in the same way at the opposite, or amino, end. By performing quantitative determinations of the free amino acids released after varying periods of enzymatic action, the sequence of amino acids at either the carboxyl or amino end of the peptide may frequently be inferred. Since the basic amino acid residues are split off only with difficulty by carboxypeptidase, the hydrazinolysis method of Akabori and associates has been used to confirm the presence of carboxyl-terminal lysine. The enzymatic methods have the advantage that asparagine and glutamine residues are liberated as such, without hydrolysis of the sensitive β- or γ-amide bond.

Sequence work of this kind is slow. Each peptide must be prepared in adequate amounts, and its purity in terms of amino acid composition established before further work can proceed. Nevertheless, by a combination of techniques, Dr. Hirs has thus far established or confirmed the position of about half of the residues in oxidized ribonuclease. The results to date (Hirs[29]) are shown in Fig. 9. The sequence of the first ten residues has been confirmed. The tetrapeptide sequence at O-Tryp 11, originally determined by Redfield and Anfinsen[30] has also been confirmed, and the order of the eight amino acids preceding it has been established. Another sequence around the fourth and fifth half-cystine residues has been worked out, and near the carboxyl end of the chain a sequence of twenty-one amino acid residues has been determined. The sequence of the four amino acids at the carboxyl end has been established by Anfinsen and associates.[23,30]

While this sequence work has been proceeding, Dr. D. H. Spackman in our laboratory, and Drs. A. P. Ryle and C. B. Anfinsen have been attempting to determine which pairs of half-cystine residues (numbered 1 to 8 in Fig. 9) are connected by disulfide bonds. The problem is basically similar to the one already solved by Sanger and his colleagues in their work on insulin. Since ribonuclease in which the disulfide bonds are intact is not

hydrolyzed in simple aqueous solution by trypsin or chymotrypsin, Dr. Spackman has allowed enzymatic action to proceed in 2 *M* guanidinium chloride.

The hydrolyzate obtained by tryptic plus chymotryptic cleavage of unoxidized ribonuclease was chromatographed on Dowex 50-X2, and from two of the cystine-containing zones from the chromatogram, two pairs of cysteic acid-containing peptides were isolated.[32] One pair was ($CySO_3H$, Asp—NH_2, Glu—NH_2, $MetSO_2$) and ($CySO_3H$, Asp, Ileu, Thr, Ser, Arg), and the other was (Asp—NH_2, $CySO_3H$, Ala, Val, Lys) and ($CySO_3H$, Asp—NH_2, Asp—NH_2, Glu—NH_2, Gly, Thr, Tyr). From the

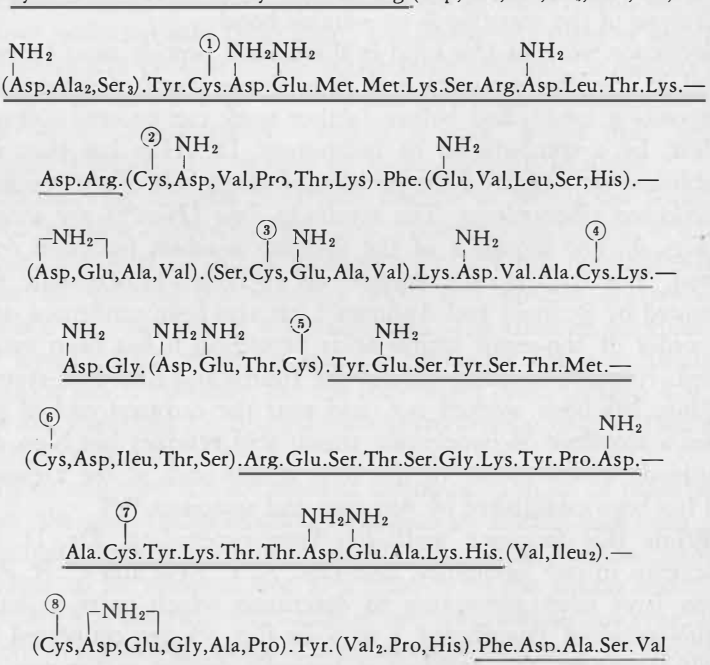


FIG. 9. Arrangement of the amino acid residues in oxidized ribonuclease (Hirs).[20] The sequences determined to date are underlined. The residues in parenthesis are of undetermined order. The cysteic acid residues have been numbered consecutively, 1 to 8, starting from the amino end of the chain.

amino acid compositions of these peptides, without the performance of any sequence studies, it could be concluded that one disulfide bond links half-cystine residues 1 and 6, and another unites residues 4 and 5. Ryle and Anfinsen, proceeding in a different manner, have also obtained evidence in favor of a 4—5 —S—S— bond. There are still two more to be located, and very recently, Dr. Anfinsen has kindly informed us that they have secured evidence[33] indicating a linkage between half-cystines 2 and 8, and 3 and 7. As a result of these experiments, all four of the

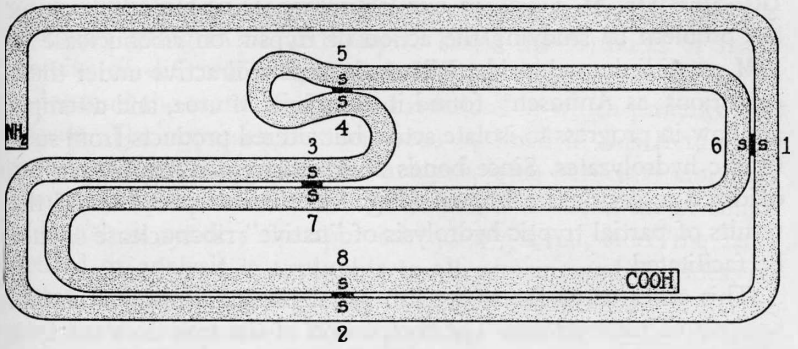


FIG. 10. A two-dimensional diagram indicating the positions of the four disulfide bonds in native ribonuclease (based upon the results of Spackman *et al.*[32] and Ryle and Anfinsen[33]).

disulfide bonds would appear to be accounted for, and from their arrangement, it is clear that the peptide chain in ribonuclease is tightly tied together. This (through Dr. Anfinsen's courtesy in letting us include his data) is shown schematically in two dimensions in Fig. 10. One feature of the molecule is readily apparent from Figs. 9 and 10. Although the sequence around the 4—5 disulfide link has not yet been completely worked out, there is no doubt that the small disulfide ring in this part of the molecule is about equivalent in size to the one already found in insulin and also in the peptide hormones, vasopressin and oxytocin.

As mentioned previously, the partial structural formula of ribonuclease (Fig. 8) is extremely useful as a guide in studies designed to learn more about the so-called "active center" of this

enzyme. The general problem of "active sites" is one of the most important in protein chemistry, because, of course, one of the main reasons for determining the structure of an enzyme is the hope that, in this manner, some insight will be gained into the basis for its catalytic properties. In research directed toward this end, Anfinsen[34] has found that even minimal hydrolysis of ribonuclease by pepsin abolishes enzymatic activity. Dr. F. M. Richards, now at Yale University, has observed that, after mild hydrolysis by the bacterial proteinase subtilisin, a new active peak appears upon chromatography of the hydrolyzate on columns of IRC-50.[35] Dr. M. Uziel, in our laboratory, has been approaching the problem by studying the action of trypsin on ribonuclease in 2 *M* guanidinium chloride. Ribonuclease is still active under these conditions, as Anfinsen[36] found it to be in 8 *M* urea, and attempts are now in progress to isolate active but altered products from such tryptic hydrolyzates. Since bonds at which trypsin may cleave the oxidized molecule are known (Fig. 8), the interpretation of the results of partial tryptic hydrolysis of "native" ribonuclease should be facilitated.

This takes us as far as we can go at present in our structural analysis of ribonuclease. The completion of the task does not look as if it would present major difficulties, but it should be emphasized that one must always be on the lookout for surprises. Although all that has been learned about ribonuclease so far is in accord with the working hypothesis that the molecule contains a single unbranched peptide chain, an experimenter must be prepared to encounter unexpected linkages in proteins—linkages of the type that Craig and his colleagues encountered in bacitracin, for example.

It is fair to ask, at this point, whether we have learned anything, in general, about the structure of proteins. We certainly do not yet know why ribonuclease is an enzyme. Moreover, examination of the structure of insulin, ACTH, glucagon, and ribonuclease has not yet revealed any common plan for the arrangement of the amino acid residues. Nor can we deduce anything about the nature of the metabolic machinery that synthesizes these proteins, except that it is amazingly precise and has not yet been found to make a mistake. A valine has never been found where a leucine

usually goes, for example. Each molecule of bovine ribonuclease appears to be like each other molecule, just as is the case for insulin from a given species.

Proteins are so complex in structure, so numerous, and exhibit such diverse activities, that we cannot expect the many problems they raise to be solved by the incomplete information we now possess. Current knowledge of proteins is still relatively small; learning how to learn has been a major objective of the research thus far. It is hoped that the method of approach initiated by Sanger and expanded in the work on ribonuclease will also be generally applicable to larger and more complex molecules, for it is obviously going to be necessary to know the formulas of many proteins of different origins and different functions if we are to understand the many biochemical reactions in which proteins participate. The approach that has been outlined in this lecture depends heavily upon amino acid analysis, with particular emphasis upon the acquisition of quantitative data. The liberal use of amino acid analysis renders the approach quite general, since the same method of analysis is applicable to all proteins and peptides, no matter how different they may be. The quantitativeness of the procedures is important. When working with molecules as large and complex as the proteins, it is, in our experience, hard to overestimate the extent to which progress is expedited and error avoided by the availability of dependable quantitative data.

It is obvious, however, that this scheme is only feasible if a relatively quick, simple, and accurate method of amino acid analysis is at hand. In order to fulfill this need, automatic recording equipment has recently been devised with the skillful collaboration of Dr. Spackman—equipment that permits a complete quantitative analysis to be turned out every 24 hours with a minimum of labor.[37] This new technique has been of great value in the later phases of the work on ribonuclease, and it is hoped that it will catalyze further advances in our general knowledge of the structure of proteins. This, of course, is only one new development in a field that is expanding so rapidly that we feel sure that future Harvey Lecturers will be able to give us much more definitive information about Nature's most intricate molecular design—the proteins.

REFERENCES

1. Bergmann, M. 1937. *Harvey Lectures Ser.* **31,** 37.
2. du Vigneaud, V. 1956. *Harvey Lectures Ser.* **50,** 1.
3. Sanger, F. 1955. *Bull. soc. chim. biol.* **37,** 23.
4. Ryle, A. P., Sanger, F., Smith, L. F., and Kitai, R. 1955. *Biochem. J.* **60,** 541.
5. Shepherd, R. G., Willson, S. D., Howard, K. S., Bell, P. H., Davies, D. S., Davis, S. B., Eigner, E. A., and Shakespeare, N. E. 1956. *J. Am. Chem. Soc.* **78,** 5067.
6. Li, C. H., Geschwind, I. I., Cole, R. D., Raacke, I. D., Harris, J. I., and Dixon, J. S. 1955. *Nature* **176,** 687.
7. Li, C. H. 1956. *Advances in Protein Chem.* **11,** 101.
8. White, W. F., and Landmann, W. A. 1955. *J. Am. Chem. Soc.* **77,** 1711.
9. Bromer, W. W., Sinn, L. G., Staub, A., and Behrens, O. K. 1956. *J. Am. Chem. Soc.* **78,** 3858.
10. Harris, J. I., and Roos, P. 1956. *Nature* **178,** 90.
11. Geschwind, I. I., Li, C. H., and Barnafi, L. 1956. *J. Am. Chem. Soc.* **78,** 4494.
12. Jones, W. 1920. *Am. J. Physiol.* **52,** 203.
13. Dubos, R. J., and Thompson, R. H. S. 1938. *J. Biol. Chem.* **124,** 501.
14. Kunitz, M. 1940. *J. Gen. Physiol.* **24,** 15.
15. Brand, E. 1948. Quoted in "Crystalline Enzymes" by J. H. Northrop, M. Kunitz, and R. M. Herriott, 2nd ed., p. 26. Columbia Univ. Press, New York.
16. Brown, D. M., and Todd, A. R. 1955. *In* "The Nucleic Acids: Chemistry and Biology" (E. Chargaff and J. N. Davidson, eds.), Vol. I, p. 409. Academic Press, New York.
17. Martin, A. J. P., and Porter, R. R. 1951. *Biochem. J.* **49,** 215.
18. Hirs, C. H. W., Moore, S., and Stein, W. H. 1953. *J. Biol. Chem.* **200,** 493.
19. Hirs, C. H. W., Stein, W. H., and Moore, S. 1954. *J. Biol. Chem.* **211,** 941.
20. Moore, S., and Stein, W. H. 1954. *J. Biol. Chem.* **211,** 893.
21. Moore, S., and Stein, W. H. 1954. *J. Biol. Chem.* **211,** 907.
22. Hirs, C. H. W., Moore, S., and Stein, W. H. 1956. *J. Biol. Chem.* **219,** 623.
23. Anfinsen, C. B., Redfield, R. R., Choate, W. L., Page, J., and Carroll, W. R. 1954. *J. Biol. Chem.* **207,** 201.
24. Hirs, C. H. W. 1956. *J. Biol. Chem.* **219,** 611.
25. Sanger, F. 1952. *Advances in Protein Chem.* **7,** 1.
26. Bergmann, M., and Fruton, J. S. 1941. *Advances in Enzymol.* **1,** 63.
27. Hirs, C. H. W., Stein, W. H., and Moore, S. 1956. *J. Biol. Chem.* **221,** 151.
28. Bailey, J. L., Moore, S., and Stein, W. H. 1956. *J. Biol. Chem.* **221,** 143.
29. Hirs, C. H. W. 1957. *Federation Proc.* **16,** 196.
30. Redfield, R. R., and Anfinsen, C. B. 1956. *J. Biol. Chem.* **221,** 385.

31. Spackman, D. H., Smith, E. L., and Brown, D. M. 1955. *J. Biol. Chem.* **212,** 255.
32. Spackman, D. H., Moore, S., and Stein, W. H. 1957. *Federation Proc.* **16,** 252.
33. Ryle, A. P., and Anfinsen, C. B. 1957. *Biochim. et Biophys. Acta* **24,** 633.
34. Anfinsen, C. B. 1956. *J. Biol. Chem.* **221,** 405.
35. Richards, F. M. 1955. *Compt. rend. trav. lab. Carlsberg, Sér. chim.* **29,** 315.
36. Anfinsen, C. B. 1956. *Compt. rend. trav. lab. Carlsberg, Sér. chim.* **30,** 13.
37. Spackman, D. H., Stein, W. H., and Moore, S. 1956. *Federation Proc.* **15,** 358.