The Arrangement of Amino Acids in Proteins

By F. SANGER

Sir William Dunn Institute of Biochemistry, University of Cambridge, Cambridge, England

CONTENTS

		Page
I.	Introduction	2
II.	Nomenclature	3
	1. Polypeptide Chains	3
	a. Open Polypeptide Chain	3
	b. Cyclic Chain	3
	c. Branched Chain	4
	2. Terminal Residues and Peptides	4
	3. Abbreviations	5
III.	Determination of the Position of Individual Residues in Proteins	5
	1. The Dinitrophenyl (DNP) Method	5
	a. Fractionation of DNP Derivatives	6
	b. *N*-Terminal Residues of Proteins	6
	c. *N*-Terminal Peptides	7
	2. The Phenylthiocarbamyl Method	8
	3. The Thiocarbamate Method	9
	4. The Use of *S*-Methylisothiourea	9
	5. *C*-Terminal Residues	9
	a. Carboxypeptidase	9
	b. Reduction to β-Amino Alcohols	10
IV.	Methods for the Degradation of Proteins	11
	1. Introduction	11
	a. Non-Specific Degradation	11
	b. Specific Degradation	11
	2. General Aspects of Partial Hydrolysis	12
	a. Yield of Peptides	12
	b. Complexity of Partial Hydrolyzates	13
	c. The Question of Rearrangement of Peptide Sequences during Hydrolysis	15
	3. Hydrolysis in Concentrated Acid	18
	a. Electrostatic Effects	18
	b. Steric Effects	19
	c. Bonds Involving the Amino Groups of the Hydroxy Amino Acids	21
	4. Hydrolysis in Dilute Acid	22
	5. Hydrolysis in Alkali	23
	6. Hydrolysis with Proteolytic Enzymes	25
	7. Rates of Hydrolysis of Proteins	26
	8. Non-Hydrolytic Methods of Degradation	27

Page

a. Splitting the Disulfide Bridges ... 27
b. Splitting by Radiation ... 29
c. Other Methods ... 29
V. Fractionation of Peptides ... 29
1. Ionophoretic Methods ... 30
2. Ion Exchange Methods ... 31
3. Adsorption Chromatography ... 34
4. Partition Chromatography ... 35
a. Paper Chromatography ... 35
b. Starch Chromatography ... 37
c. Other Partition Chromatography ... 37
5. Detection of Peptides from Columns ... 38
6. Countercurrent Distribution ... 38
7. Lysine Peptides ... 39
8. Cystine Peptides ... 40
9. Other Methods of Fractionation ... 40
10. Conclusions ... 40
VI. Determination of Peptide Structure ... 41
1. Identification and Estimation of Amino Acids ... 41
2. Amino Acid Sequence ... 42
3. Estimation of Peptides ... 43
The Pipsyl Method ... 43
VII. Results of Investigations on Various Proteins ... 44
1. Silk Fibroin (*Bombyx Mori*) ... 44
2. Protamines ... 47
3. Wool Keratin ... 48
4. Insulin ... 50
a. *N*-Terminal Peptides ... 52
b. Amino Acid Sequence in the Phenylalanyl Chains ... 54
5. Ovalbumin ... 57
6. γ-Globulin ... 58
7. Hemoglobin ... 59
8. Gelatin ... 59
VIII. General Conclusions ... 59
References ... 62

I. Introduction

A comprehensive review of the earlier literature on the partial hydrolysis products of proteins was given by Synge in 1943. Up to that time only a few simple peptides had been clearly identified from proteins by the classical and rather laborious methods of organic chemistry and Synge concluded that "the main obstacle to progress in the study of protein structure by the methods of organic chemistry is inadequacy of technique!" Probably the greatest advance that has been made recently in this field was the development by Martin and Synge (1941) of the entirely new technique of partition chromatography. The great problem in peptide chemistry has always been to find methods of frac-

tionating the extremely complex mixtures produced by the partial degradation of a protein. Older methods of fractional crystallization and precipitation with various reagents were as a rule inadequate to deal with these mixtures, and countercurrent methods of high resolving power, which could fractionate non-volatile, water-soluble substances, were needed. Partition chromatography, especially in the form of paper chromatography (Consden *et al.*, 1944), is such a method, so that it has already been possible to identify as breakdown products of proteins more peptides using this technique than had previously been identified by the classical methods of organic chemistry. During the last few years, work in this field has centered largely on the development of methods, so that this review will be more a consideration of techniques and their uses than a discussion of results, which are still rather few.

As an initial working hypothesis it will be assumed that the peptide theory is valid, in other words, that a protein molecule is built up only of chains of α-amino (and α-imino) acids bound together by peptide bonds between their α-amino and α-carboxyl groups. While this peptide theory is almost certainly valid (see Vickery and Osborne, 1928; Pauling and Niemann, 1939; Synge, 1943), it should be remembered that it is still a hypothesis and has not been definitely proved. Probably the best evidence in support of it is that since its enunciation in 1902 no facts have been found to contradict it. It is to be expected that investigations of the types described in the present article will throw further light on the accuracy of the peptide theory and on the possible existence of non-peptide bonds in proteins.

II. Nomenclature

Some of the terms to be used are new and require definition.

1. *Polypeptide Chains*

Three types of polypeptide chains are possible, open, cyclic and branched.

a. Open Polypeptide Chain. A chain of amino acids joined together by peptide bonds between the α-amino and α-carboxyl groups with a free amino group at one end of the chain and a free carboxyl group at the other end is the most usually considered type of polypeptide chain. The number of these chains in a protein may be estimated by the number of α-amino or α-carboxyl groups.

b. Cyclic Chain. A cyclic chain may be derived from an open polypeptide chain by peptide bond formation between the two terminal residues, and contains no free α-amino or α-carboxyl groups. The anti-

biotics "gramicidin S" (Sanger, 1946) and tyrocidine (Christensen, 1945) have been shown to possess cyclic structures and the absence of any free α-amino groups in ovalbumin and certain muscle proteins suggests that they too are built up from cyclic chains.

c. Branched Chain. The presence of two carboxyl groups in glutamic and aspartic acids and of two amino groups in lysine suggests the possibility that branched chains may occur in proteins. For instance, such a branched system could be formed by the formation of a peptide bond between the free α-amino group of one open chain and a γ-carboxyl group of a glutamic acid residue in another chain.

For all proteins that have been studied by the dinitrophenyl (DNP) method (see below) it was found that all the ϵ-amino groups of the lysine residues were free, indicating that these groups are not involved in any bond formation. Thus it is unlikely that there is very much branching of chains from the lysine residues. If the branching points are very few, however, it is just possible that some of the ϵ-amino groups thus masked have escaped detection. No evidence is available as to whether branching can occur from the ω-carboxyl groups of aspartic or glutamic acids. The presence of the γ-peptide linkage of glutamic acid in glutathione, glutamine and the capsular substances from *B. Anthracis* (Hanby and Rydon, 1946) suggests that it may also occur in proteins.

Some proteins are built up of two or more polypeptide chains held together by a stable bond other than the peptide link. Such a bond will be referred to as a "cross linkage." The only one that is definitely known to exist is the —S—S— bridge of cystine. Here two cysteine residues are joined together through their side chains.

2. *Terminal Residues and Peptides*

There are two types of terminal residues, those with a free amino group and those with a free carboxyl group. In previous publications (Sanger, 1945, 1949b) the expression "terminal residue" has been used to denote only that residue which carries a free amino group. However, it seems that a distinction should be made. Fox (1945) has suggested that those residues containing a free amino group and a bound carboxyl group be referred to as terminal *amino* acids and those with a free carboxyl group as terminal amino *acids*. Although clear on paper, this distinction is rather difficult to make in conversation. Following a suggestion by Dr. K. Bailey, it is proposed to use the term *N*-terminal residue for the residue having a free amino group and *C*-terminal residue for that having a free carboxyl group. The same nomenclature will apply to the terminal peptides.

3. *Abbreviations*

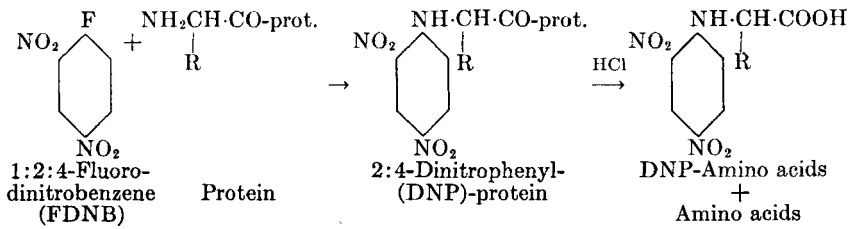
Throughout this review the abbreviations for the amino acid residues suggested by Brand and Edsall (1947) are used. Cysteic acid is abbreviated $CySO_3H$. Their method of writing empirical formulae of proteins and peptides is also adopted, *e.g.*, $Gly_{29}Ala_{11}Val_{21}$ . . . etc., the subscripts representing the number of residues of the amino acid per protein molecule. In this type of formula the order in which the amino acids are given has no significance. Where the order of residues is known, as in the description of a peptide, the symbols are joined by a period. Thus, whereas Gly,Ala signifies a dipeptide containing glycine and alanine, Gly.Ala represents glycylalanine. Gly.(Ala,Leu) indicates a peptide containing glycine, alanine and leucine with glycine as the *N*-terminal residue, and the order of the alanine and leucine is unknown. As is customary, the first residue is the *N*-terminal residue and the last the *C*-terminal.

III. Determination of the Position of Individual Residues in Proteins

The terminal residues of proteins differ from other residues in the chain, since they contain free amino or free carboxyl groups, and this fact may be used to identify them. Fox (1945) has reviewed the earlier literature on the study of terminal residues. At that time the position of only one amino acid in one protein was known. This was the presence of phenylalanine as an *N*-terminal residue in insulin. It was identified by Jensen and Evans (1935) who isolated the phenylhydantoin of phenylalanine from a hydrolyzate of insulin that had been treated with phenylisocyanate. More recently a general method has been worked out for the study of *N*-terminal residues, and preliminary investigations have been carried out on three methods of stepwise degradation which promise to be of great use in the future.

1. *The Dinitrophenyl (DNP) Method*

The principle of this general method (Sanger, 1945; Porter and Sanger, 1948) for the identification and estimation of the *N*-terminal residues of proteins may be summarized by the following formulae:



The FDNB reacts with the free amino groups of the protein under mild (slightly alkaline) conditions where the peptide bond is quite stable. On hydrolysis of the protein the *N*-terminal residues are liberated in the form of DNP amino acids. These are bright yellow compounds that can be extracted with an organic solvent, fractionated chromatographically, and estimated colorimetrically. The accuracy varies somewhat with the particular amino acids involved, due to differences in the stability of the DNP derivatives. In most cases the *N*-terminal residues of proteins and peptides may be estimated to within 10–15%.

a. Fractionation of DNP Derivatives. Several methods of fractionation have been suggested, all depending essentially on partition chromatography. Originally (Sanger, 1945; Porter and Sanger, 1948; Porter, 1950c) a scheme was worked out for separating the DNP derivatives of all the known amino acids using silica gel saturated with water as the stationary phase and various organic solvents as the mobile phase. This method was found to give satisfactory and reproducible separations in the author's laboratory; other workers (Consden *et al.*, 1947b; Blackburn, 1949), however, have found difficulty in obtaining suitable gels. These may be prepared from any type of sodium silicate by appropriate modification of the method of preparation (Desnuelle *et al.*, 1950), and should be rather strongly adsorbent in order to hold the relatively insoluble DNP derivatives in the stationary phase. Other methods of fractionation have been suggested in which the stationary phase is a buffer, adsorbed on kieselguhr (Bell *et al.*, 1949) or silica (Blackburn, 1949; Middlebrook, 1949). The DNP-derivatives are then partially ionized and are thus rendered more soluble in the aqueous phase. Such systems may prove of more general use, though *R* values are still not reproducible on different batches of silica (K. Bailey, unpublished observation). The use of paper chromatograms using buffer solutions has also been suggested (Blackburn and Lowther, 1950; Monnier and Penasse, 1950) though details have not been reported. Recently Partridge and Swain (1950) have obtained excellent separations using butanol adsorbed on rubber as the stationary phase and buffer solution as the mobile phase. Such a system with an aqueous moving phase would be expected to be more satisfactory for compounds which are more soluble in organic solvents than in water.

b. N-Terminal Residues of Proteins. In Table I are listed the *N*-terminal residues of a number of proteins as determined by the DNP-technique.

Besides reacting with the α-amino groups of proteins, FDNB also reacts with the ϵ-amino groups of the lysine residues, and an estimation of the ϵ-DNP-lysine in the hydrolysate indicates how many of these amino groups are free in the intact protein. For all proteins studied reasonable

agreement is found between the ε-DNP-lysine and the total lysine content of the protein, if the protein is first denatured. It thus seems unlikely that any branching of peptide chains occurs through the ε-amino

TABLE I

N-Terminal Residues of Proteins

Protein	Assumed mol. wt.	*N*-Terminal residue	Number per molecule
Insulin[a] (Ox, pig, sheep)	12,000	Phenylalanine	2
		Glycine	2
Hemoglobin[b] (horse, donkey)	66,000	Valine	6
Hemoglobin[b] (ox, sheep, goat)	66,000	Valine	2
		Methionine	2
Hemoglobin[b] (human adult)	66,000	Valine	5
Hemoglobin[b] (human fetal)	66,000	Valine	2-3
Myoglobin[b] (horse)	17,000	Glycine	1
Myoglobin[c] (whale)	17,000	Valine	1
β-Lactoglobulin[d]	40,000	Leucine	3
Ovalbumin[e,f]	—	None	—
γ-Globulin[f] (rabbit)	160,000	Alanine	1
Edestin[g]	300,000	Glycine	6
		Leucine	1
Salmine[b]	—	Proline	?
Clupein[h]	—	Proline	?
Myosin[i]	—	None	—
Tropomyosin[i]	—	None	—
Lysozyme[j]	14,000	Lysine	< 1
Pancreatic Trypsin Inhibitor[k]	9,000	Arginine	1
Serum albumin[l] (human, horse, ox)	69,000	Aspartic acid	1

[a] Sanger (1945).
[b] Porter and Sanger (1948).
[c] Schmid (1949).
[d] Porter (1948).
[e] Desnuelle and Casal (1948).
[f] Porter (1950a).
[g] Sanger (1949d).
[h] Felix *et al.* (1950).
[i] Bailey (1951).
[j] Green and Schroeder (1951).
[k] Green and Work (1951).
[l] van Vunakis and Brand (1951).

groups of the lysine residues. In certain native proteins (*e.g.*, β-lactoglobulin) some of the ε-amino groups do not react, indicating that they are in some way masked in the native molecule (Porter, 1948).

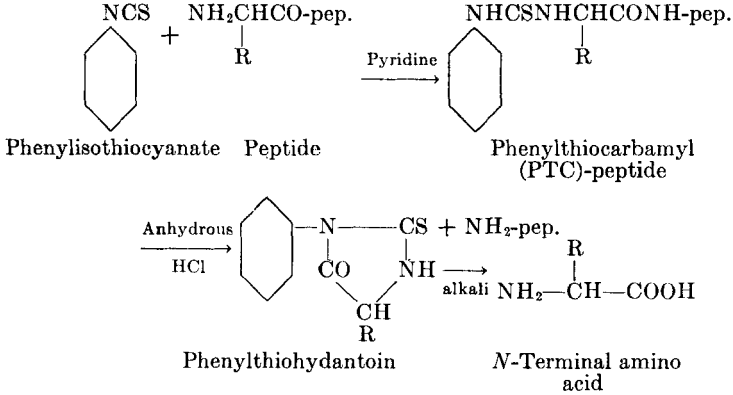
c. N-Terminal Peptides. The DNP technique may also be used to identify and estimate *N*-terminal peptides (Sanger, 1949b). Thus if a DNP-protein is only partially hydrolyzed, one obtains DNP-peptides

which may be extracted into an organic solvent and fractionated on suitable chromatograms. Complete hydrolysis of these purified DNP-peptides then reveals the nature of the amino acids and the *N*-terminal residue present, and amino acid arrangement may be determined by a second partial hydrolysis. In this way it is possible to identify the residues that occupy positions in the polypeptide chains near to the *N*-terminal residues. So far it has been possible to identify peptides up to about four residues long. Longer DNP-peptides are more difficult to separate from the other unsubstituted peptides and to fractionate, though the possibilities have not yet been fully explored.

For an illustration of the use of these methods the reader is referred to the section on insulin (p. 50).

2. *The Phenylthiocarbamyl Method*

Recently Edman (1950) described a method for determining the sequence of amino acids in a polypeptide chain, by splitting off one residue at a time starting from the *N*-terminal residue. The principle of the method is illustrated by the following equations:

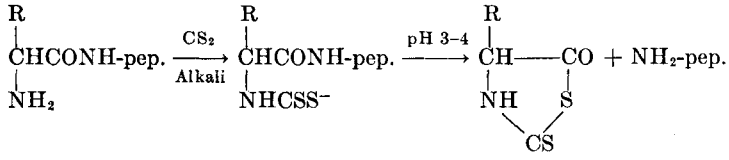


The formation of the phenylthiohydantoin does not require the presence of water, as does the hydrolysis of peptide bonds. Thus by heating the PTC-peptide with anhydrous HCl in nitromethane it is possible to break off the *N*-terminal residue as a phenylthiohydantoin without splitting other bonds in the peptide. The hydantoin dissolves in the nitromethane and is then separated and hydrolyzed to the amino acid which may be identified by paper chromatography. The rest of the peptide with the *N*-terminal residue removed is insoluble in nitromethane and the process may be repeated. The second residue is thus split off and identified. This method has given excellent results with synthetic peptides and it will be interesting to see how far it may be applied to a

protein. Theoretically, it should be possible to determine the complete structure of a pure single chain polypeptide or protein. At least the method should be extremely valuable for working out the structure of smaller peptides.

3. *The Thiocarbamate Method*

Another method for the stepwise degradation of a polypeptide chain has been suggested by Levy (1950). Here the *N*-terminal residue is split off as a 2-thiothiazolid-5-one derivative as follows:



Peptide Thiocarbamate 2-Thiothiazolid-5-one

Here again the process may be repeated on the peptide chain containing one residue less. This method has not yet been worked out on a small scale, but was found to give satisfactory results with synthetic peptides.

The reaction of CS_2 with amino groups has also been used by Léonis (1948) as the basis for a titration method for estimating the total α-amino groups as well as ϵ-amino, imino, and thiol groups. CS_2 reacts about ten times as rapidly with α-amino groups as with ϵ-amino groups, so that the two may be distinguished.

4. *The Use of S-Methylisothiourea*

Christensen (1945) has used *S*-methylisothiourea to study the free amino groups of tyrocidine. The amino groups are converted to guanidine groups which may be estimated by the Sakaguchi reaction. This method is especially useful for detecting the free amino group of ornithine, which is converted to arginine and thus may be estimated using arginase. In the case of tyrocidine the Sakaguchi reaction was negative after the action of arginase, indicating that the only free amino groups were the δ-amino groups of ornithine, and that tyrocidine was thus a cyclopeptide. It is doubtful if this method could be used to identify the *N*-terminal residue of more complex proteins.

5. *C-Terminal Residues*

a. Carboxypeptidase. Specificity studies (reviewed in Neurath and Schwert, 1950) indicate that carboxypeptidase attacks only those peptide bonds that are adjacent to a free α-carboxyl group. It appears that all

such bonds with the possible exception of those involving glycine are attacked to some extent although the rate of hydrolysis varies greatly with the nature of the residue involved. Thus if a protein or peptide is treated with carboxypeptidase the first amino acid to be liberated in the free form is the *C*-terminal residue. Lens (1949) has applied the method to insulin. Samples of the digest were removed at various intervals, ultrafiltered and the ultrafiltrate analyzed by paper chromatography. Free alanine was clearly liberated first and was therefore present in insulin as a *C*-terminal residue.

With a single pure polypeptide chain it should theoretically be possible to determine the complete sequence of residues by following the rate of liberation of different amino acids under the action of carboxypeptidase. No such experiments have been described, but it would be interesting to know how far this method could be applied in practice. Clearly it would be impossible to draw any conclusions beyond the *C*-terminal residues if more than one peptide chain were present, as in the case of insulin.

b. Reduction to β-Amino Alcohols. Fromageot *et al.* (1950) and Chibnall and Rees (1951) have independently worked out techniques for the identification of *C*-terminal residues by reduction to β-amino alcohols. In the former method lithium aluminum hydride is used to reduce the protein. After hydrolysis, the β-amino alcohols are extracted into ether, identified by paper chromatography and estimated by reaction with periodate. Some reduction of peptide bonds was observed, and the yields were rather low. In the method of Chibnall and Rees, the carboxyl groups are first esterified with diazomethane and then reduced with the less violent reagent, lithium borohydride. The amino alcohols are separated from the amino acids in the hydrolyzate by electrodialysis, and treated with periodate, which decomposes them according to the equation:

$$HIO_4 + R{\cdot}CHNH_2{\cdot}CH_2OH \rightarrow HIO_3 + R{\cdot}CHO + HCHO + NH_3$$

A determination of the extra formaldehyde and ammonia produced in this reaction gives an estimation of the number of free α-carboxyl groups, and the residue on which they are located is identified from the nature of the amino alcohol and of the aldehyde R·CHO. Chibnall and Rees have also used this technique to determine the distribution of the protein amide groups between asparagine and glutamine. Residues of aspartic or glutamic acid which have a free ω-carboxyl group are destroyed by the above treatment, while those in amide form remain intact and can be estimated after hydrolysis of the protein.

IV. Methods for the Degradation of Proteins

1. *Introduction*

At present the only residues to which definite positions in the protein chain can be assigned, are those at or near the terminal positions, and these residues constitute only a small part of the molecule. The position of other residues in the chain can only be determined in the first place relative to one another by the identification of products of partial breakdown. The absolute location of any one residue can be decided only when the relative positions of all the residues are known and when the complete structure of the protein is thus worked out.

Thus in determining the amino acid sequence of a protein or polypeptide the main task would seem to be to identify as many degradation products as possible. There are essentially two ways of approaching this problem, the first requiring non-specific methods of degradation and the second specific methods.

a. Non-Specific Degradation. One method of determining amino acid sequence is to degrade the polypeptide or protein directly to a mixture of small peptides whose structure can be determined and to work out a unique sequence by fitting together the degradation products. It is possible to determine directly the structure of di- and possibly tripeptides. To obtain a unique solution for the amino acid sequence of a polypeptide it is necessary to obtain as many different small peptides as possible. This may best be achieved by using unspecific methods of degradation to give a mixture of maximum complexity and using reagents having different specificities. If we consider a hypothetical polypeptide whose structure may be written as A.B.C.D.E.F.G.H. where the letters represent different amino acid residues, degradation to the dipeptides AB,BC,CD,DE,EF,FG,GH would give a unique solution. If only AB,DE,FG and GH are obtained, various structures are possible, but if a different type of hydrolysis yields EF and the tripeptide BCD then the structure is determined.

This method will generally be applicable to simpler peptides. In the case of larger peptides, the complexity of the mixture will render fractionation difficult and a proportion of larger peptides will be required to work out an unambiguous result.

b. Specific Degradation. Most proteins contain at least 100 residues. It seems unlikely that a complete solution of their amino acid sequence will ever be obtained solely by the non-specific method and it will probably be necessary to approach the problem by the second method, involving a gradual specific breakdown into a small number of large fragments which can be purified and again broken down to smaller

products. Thus if the peptide A.B.C.D.E.F.G.H. could be split into two products A.B.C.D. and E.F.G.H. which could be separated, their structure could be determined much more readily than that of the original peptide.

In the choice of a suitable method of degradation three main factors are to be considered:

1. The reagent should cause a minimum of side reactions such as destruction of the constituent amino acids, or if such reactions do occur they should occur quantitatively to produce known products.

2. The reagent should exhibit the desired specificity. Specific methods are required for the initial breakdown and several relatively non-specific methods are needed for the final elucidation of sequence.

3. There should be a minimum of synthesis or rearrangement of peptide bonds under the influence of the reagent.

The first two points will be considered in connection with the different methods of hydrolysis, and the last will be discussed separately on page 15.

2. *General Aspects of Partial Hydrolysis*

By far the most important method of breaking down proteins is by hydrolysis of the peptide bonds. All the peptide bonds in a protein are susceptible to hydrolysis but there are great differences in the stability of the different bonds depending on the nature of the residues involved.

a. Yield of Peptides. If we assume that 1 mole of a polypeptide A.B.C.D.E.F is hydrolyzed to such an extent that the mole fractions of the bonds A—B, B—C, C—D etc. that are split are b,c,d etc. respectively, the yield of the amino acid C will be cd moles, since only those molecules of the polypeptide in which the bonds B—C and C—D are broken will give rise to C. To obtain the dipeptide CD the bonds BC,DE must be broken and CD unbroken, so that the yield will be c(1 − d)e. In general the yield F of a peptide $A_1.A_2.A._3 \ldots A_n.$ is given by:

$$F = \alpha_1\alpha_{n+1} \prod_{\alpha_2}^{\alpha_n} (1 - \alpha) \qquad (1)$$

where $\alpha_1,\alpha_2 \ldots \alpha_n$ are the mole fractions of the bonds involving the amino groups of A_1, A_2, $\ldots$ A_n, respectively that are split, and F is expressed in moles of peptide. This treatment assumes that the rate of hydrolysis remains constant as the reaction proceeds, which is not entirely true in all cases.

Equation 1 indicates that the yield of a peptide depends not only on the lability of the bonds involved in its terminal residues but also on the stability of the bonds within the molecules. It is also evident that in

general the yield will be greater if the peptide is a terminal one since α_1 or α_{n+1} is already a maximum ($= 1$).

Various workers (Kuhn, 1930; Montroll and Simha, 1940; Warner, 1942b; Myrbäck, 1949) have considered the mathematical treatment of the breakdown of high molecular chains assuming that all bonds were broken at the same rate, *i.e.*, that $\alpha_1 = \alpha_2 = \alpha_n$ etc. Equation 1 then becomes:

$$F = \alpha^2(1 - \alpha)^{n-1} \tag{2}$$

Assuming the chain to be of infinite length or cyclic, the fraction of the original chain appearing as peptides containing n residues is given by:

$$F_n = n\alpha^2(1 - \alpha)^{n-1} \tag{3}$$

It is difficult to know how far this type of treatment can really be applied to proteins where there is such great variation in the susceptibility of various bonds. Presumably in a large molecule where the intrinsic rates of hydrolysis are fairly evenly distributed about the mean, equation (3) would apply.

It is interesting to note that for any value of α in equation (2), F is a maximum if $n = 1$ and decreases as n increases, in other words the molar yield of smaller peptides is always greater than the yield of larger peptides. On the average this is also true for equation (1), which expresses more closely the situation present in a protein, since F is never greater than $\alpha_1\alpha_{n+1}$; however, as each α is different it will not apply to every case and the yield of certain higher peptides may be greater than the yield of certain smaller ones.

Though the values for α_1,α_2, etc. in equation (1) can of course be expressed in terms of the hydrolysis constants for the separate bonds, the treatment becomes extremely complicated as a different constant will be required for each polypeptide in which the particular bond occurs. In other words, the constant will vary as the reaction proceeds. Thus Kuhn *et al.* (1932) found that a three line formula was required to express the rate of hydrolysis of tetraglycine. Thus it does not seem that any rigid mathematical approach can be given at present, and we shall have to be content with a few generalizations derived largely from experimental observations.

b. Complexity of Partial Hydrolyzates. When a protein is partially hydrolyzed, a very complex mixture of peptides is produced, the exact complexity of which is difficult to assess.

If we consider an open polypeptide chain consisting of N residues, complete hydrolysis will give rise to N amino acids, many of which may

be identical. If it is partially hydrolyzed, the number of possible dipeptides is $N - 1$, of tripeptides $N - 2$ and of n-peptides (*i.e.*, peptides with n residues) $N - n + 1$, very few of which are likely to be identical. The total number of possible peptides is $N(N + 1)/2$. Clearly the shorter the time of hydrolysis the greater will be the number of higher peptides present in significant amounts. From equation (1) (p. 12) it was inferred that the yield of the higher peptides is always less than that of the smaller ones, so that the shorter the time of hydrolysis, the greater will be the complexity of the mixture. When a protein is hydrolyzed, there is only one molecular species to start with. The number of species present in significant amounts then increases rapidly to a maximum and gradually falls off. The initial rise in complexity will depend on what one considers to be a significant amount. Some splitting of each bond starts immediately with the instantaneous production of small amounts of all the $N(N + 1)/2$ different peptides. If one defines the "significant amount," then for a while the only species will be the original protein, and others will gradually be added to it. It is clear however that there is really no phase during the hydrolysis when the complexity of the mixture is increasing, and that it is impossible to make use of this apparently simple composition during the initial phase of hydrolysis to obtain a mixture in which only a few polypeptides are present "in significant amount," as there will be too many others present in insignificant amounts.

As an example we may consider a protein such as ovalbumin which is probably a cyclic polypeptide of 400 residues. If the hydrolysis were such that all bonds were split at the same rate, then during the initial phase of hydrolysis there will be 160,000 different chemical species produced, and this number will gradually decrease to the 20 free amino acids. In the presence of any agent, such as acid or alkali, which splits all bonds to some extent, all the 160,000 will be produced, but some in negligible amounts depending on the specificity of the reagent. Bull and Hahn (1948) have suggested that when ovalbumin is hydrolyzed by strong acid about 50 bonds are readily broken and that the rest are broken more slowly, say at one tenth of the rate. Consider first only the initial phase in which the 50 labile bonds are broken. When they are all broken they should give rise to 50 peptides of average length 8 residues. At the beginning of hydrolysis $50^2 = 2500$ different combinations of these peptides will be produced the number gradually falling to the 50 octapeptides. At the same time, however, all the other bonds within these octapeptides have been subject to hydrolysis at one tenth of the rate, so that $\frac{8}{10}$ of these octapeptides are broken down and each will give rise to $\frac{8}{2}(8 + 1) = 36$ different degradation products. The

hydrolysis will thus have the following composition. One-fifth will be in the form of 50 octapeptides and the remaining four-fifths in the form of $50 \times 36 = 1800$ smaller peptides. In reality the situation is even more complex since there are not two types of bonds but bonds with every type of stability. However, the above example does make it clear that at no stage in the hydrolysis, except near the end, will the composition of the hydrolyzate be sufficiently simple to justify investigation.

In the case of enzymic hydrolysis, a large proportion of bonds are unattacked and here again the simplest mixture and the most profitable one to investigate is the complete hydrolyzate. At earlier stages in the hydrolysis the mixture will be increasingly complex. The only other means of obtaining a relatively simple hydrolyzate would arise if there was a very sharp break in the hydrolysis curve, that is to say, if there are two types of bonds with very different labilities. For instance, if the 50 labile bonds in ovalbumin were hydrolyzed 100 times as rapidly as the other bonds, one would at a certain stage obtain a mixture in which the 50 octapeptides are present to the extent of 90%. However, it is unlikely that any one of the commonly-used enzymes could bring about such a sharp differentiation. Since methods of fractionation at present available may be capable of separating 100 short peptides but not 1000 long peptides, it would seem advisable to confine our attention to the later stages of hydrolysis by any agent. Special emphasis is laid on this point as it appears to be quite a common practice for protein chemists to attempt to determine the nature of proteins by splitting them to a few large peptides by partial hydrolysis, on the assumption that by starting with one compound and ending with 20, at some stage there must be only two or three compounds. This is true only when the first molecule in the solution is split. The chance that the second molecule will split in the same place is rather remote.

c. The Question of Rearrangement of Peptide Sequences during Hydrolysis. It is clear that if any synthesis or rearrangement of peptide bonds takes place during the course of hydrolysis of a protein, the amino acid sequences identified in products of partial hydrolysis may not be the actual sequences that were present in the original protein, so that this approach to the problem would be useless. The hydrolysis of the peptide bond is reversible, so that the theoretical possibility exists that any peptide bond may be synthesized. The free energy of formation of a peptide bond is probably about 3000–4000 calories per mole for a dipeptide (Huffman, 1942) and 2000 calories per mole for a peptide bond within a protein (Haugaard and Roberts, 1942) so that the equilibrium will be very much on the side of hydrolysis for most peptide bonds, and a direct reversal of hydrolysis would seem rather unlikely.

Synthetic reactions are more likely to take place if conditions are such that the product of the reaction is rapidly removed. Bergmann and his colleagues have shown that in the presence of certain proteolytic enzymes such conditions may be obtained either when the synthetic product is insoluble and is removed from the solution by crystallization (Bergmann and Fraenkel-Conrat, 1937, 1938; Bergmann and Behrens, 1938; Bergmann and Fruton, 1938) or else when it is removed by rapid hydrolysis to other products (Behrens and Bergmann, 1939). The former possibility may be avoided by keeping the reaction mixture always in solution. However the latter possibility, is rather more difficult to eliminate. Such a reaction may be formulated as follows:

$$A.B + C.D \leftrightharpoons A.B.C.D \rightarrow A.B.C + D$$

If the reaction $A.B.C.D \rightarrow A.B.C + D$ is very rapid it will shift the equilibrium of the other reaction to the right and thus bring about the synthesis of the bond B—C. Such a reaction will only occur to an appreciable extent if the hydrolytic reagent has a much greater affinity for the bond C—D in the peptide A.B.C.D than in C.D, that is to say if its specificity is determined not only by the residues involved in the susceptible bond, but by residues further removed from it. This would seem to be possible in the case of hydrolysis by proteolytic enzymes but less likely when acid or alkali are used (Sanger, 1949c).

Fruton (1950) has recently pointed out the possibility that "transpeptidation" reactions may occur in the presence of proteolytic enzymes. Such reactions may be represented as follows:

$$RCO{-}NH\underset{\displaystyle R'}{\underset{|}{C}}HCO{-}NHR'' + NH_2R''' \leftrightharpoons RCO{-}NH\underset{\displaystyle R'}{\underset{|}{C}}HCO{-}NHR''' + NH_2R''$$

It was originally shown by Bergmann and Fraenkel-Conrat (1937) that benzoylglycylanilide may be synthesized more rapidly by papain from benzoylglycinamide than from benzoylglycine, thus indicating that a direct transformation of the amide to the anilide occurs without intermediate formation of the acid. Following up this work Johnston *et al.* (1950) showed that if papain is allowed to act on benzoylglycinamide in the presence of NH_3 containing N^{15}, a small amount of the isotope is introduced into the amide. Clearly, if this type of reaction which involves very little change in free energy occurs to any great extent during a partial hydrolysis of a protein it may lead to rearrangement of peptide sequences.

Another type of rearrangement that might occur is through the amino acid anhydrides or diketopiperazines. The formation of these stable six-membered rings presumably involves a smaller free energy of forma-

tion than the synthesis of most other peptide bonds and takes place readily under various conditions. If such conditions are present the following equilibrium will exist in the hydrolyzing solution

$$\text{A.B} \leftrightharpoons \left[\begin{matrix}\text{A}\\ \text{B}\end{matrix}\right] \leftrightharpoons \text{B.A}$$

where A.B represents a dipeptide and $\left[\begin{matrix}\text{A}\\ \text{B}\end{matrix}\right]$ the corresponding anhydride. The relative rates of the different reactions may be such as to cause the partial or complete inversion of the order of the residues in the dipeptide, even if the anhydride intermediate never appears in the solution. Anhydrides are very much more labile both to alkali (Levene *et al.*, 1930, 1932; Kuhn *et al.*, 1932) and to acid (Abderhalden and Mahn, 1927, 1928), than the corresponding dipeptides, presumably since they contain no free charged group. This property tends to accelerate any possible inversion of a dipeptide. The formation of diketopiperazines seems to be associated with high temperatures. Brigl (1923) and Abderhalden and Komm (1924a, b) were able to convert dipeptides to anhydrides in quite high yield by heating in water or dilute acids at temperatures of 150–250°. No conversion could be demonstrated when the strength of acid was greater than 1 *M*. Small amounts of anhydrides were formed when the dipeptides were refluxed in water for several days. No reaction could be demonstrated under the action of 70% H_2SO_4 or concentrated HCl at room temperature. This does not necessarily prove that no anhydrides were formed, as they may have been broken down as fast as they were formed. However these results do at least indicate the possibility of anhydride formation and inversion under certain conditions.

Clearly the question of whether rearrangements do actually take place in partial hydrolysis experiments can only be solved by experience. If any such reactions occur it would be impossible to interpret results in terms of a unique sequence of amino acids unless the syntheses are completely specific and quantitative, which is most unlikely. In fact, using concentrated acid at low temperatures it has been possible to work out a unique sequence for "gramicidin S" (Consden *et al.*, 1947b from the peptides identified, and there was no evidence of any peptide that did not fit this sequence. Similarly a unique structure could be determined for the phenylalanyl chains of insulin (p. 54). It thus seems unlikely that any rearrangement occurs under the action of this type of reagent. On the other hand, syntheses have been definitely shown to occur in the presence of proteolytic enzymes. The formation of plastein by the action of pepsin or trypsin on concentrated peptide mixtures has clearly been shown in certain cases to be accompanied by a decrease in amino

nitrogen, which can only be interpreted as a net synthesis of peptide bonds (reviewed by Wasteneys and Borsook, 1930; Virtanen *et al.*, 1950) and the results with synthetic substrates are even more clear-cut, although in no case have naturally-occurring peptides been used as substrates.

In conclusion it may be said that while the theoretical possibility exists that synthetic reactions may occur during the partial hydrolysis of proteins, it is the opinion of the reviewer that future research will show that such syntheses are insignificant and will not interfere with the interpretation of data derived from partial hydrolysis experiments in terms of protein structure.

3. *Hydrolysis in Concentrated Acid*

Acid hydrolysis is the most generally used method of degrading proteins, and it is almost universally employed when amino acids are to be isolated or estimated, since it leads to complete hydrolysis with a minimum of destruction. The only amino acid that is extensively destroyed is tryptophan, the destruction of which may be largely due to the presence of traces of heavy metals during hydrolysis and may be reduced by using very pure HCl and quartz vessels (Jacobsen, 1949; Monnier and Jutisz, 1950).

Slight destruction of serine and threonine (Rees, 1946) also takes place, but in partial hydrolyzates this would be almost negligible. No synthetic reactions or rearrangements have been shown to take place under the action of strong acids.

The relative rate of hydrolysis in acid of any peptide bond and hence the yield of a given peptide is determined mainly by the number of hydrogen ions that can approach the bond. While the rate probably depends on a number of different factors, we may consider two which probably play a major role, namely, electrostatic effects and steric effects.

a. Electrostatic Effects. The presence of any charged groups in the neighborhood of a peptide bond will clearly affect the approach of hydrogen ions. In strong acid all the carboxyl groups on the proteins will be uncharged but all the basic groups (amino, imidazole, and guanidyl) will be fully charged and will oppose the approach of the similarly charged hydrogen ions.

From their studies of the course of hydrolysis of various proteins by concentrated acid Gordon *et al.* (1941) were able to calculate the ratio of free basic amino acids in the hydrolyzate to total basic residues (free and in peptide combination). This ratio gives an average estimate of the stability of basic peptides. In general, the values found were slightly lower than the corresponding figures for the neutral residues, indicating

that the basic peptides were slightly, though not very much more stable, than the average peptide. Thus the ratio obtained for wool at a certain stage of hydrolysis was 0.33 and the corresponding figure for the neutral residues 0.39.

Probably the charged groups that are most effective in stabilizing peptide bonds are the α-amino groups, which are closer to a peptide bond than the ϵ-amino, imidazole and guanidyl groups. A bond involving an *N*-terminal residue should thus be relatively stable to acid. An important result of this is the stability of dipeptides which leads to their accumulation at a certain stage of hydrolysis (Gordon *et al.*, 1941). This stability is evident from the results of Stein *et al.* (1944), who followed the course of hydrolysis of silk fibroin in concentrated HCl at 40° using the van Slyke nitrous acid method for estimating the rate of liberation of free amino groups and the ninhydrin method (van Slyke *et al.*, 1941) for free amino acids. From these results it is possible to calculate the average length of the peptides excluding in the average the free amino acids. After 43 hours hydrolysis 60% of the peptide bonds were split, and the hydrolyzate contained 25% of its nitrogen in the form of free amino acids and 75% in the form of peptides whose average length was 2.05 residues; in other words almost the whole of the N in peptide linkage was assignable to dipeptides. From equation (3) (p. 13) it may be calculated that if a completely random splitting had occurred the yield of amino acids would have been 36%, of dipeptides 29%, of tripeptides 17%, etc.

Similar conclusions may be drawn from the rates of hydrolysis of gramicidin (Synge, 1945; Christensen and Hegsted, 1945) the effect being more marked at 37° than at the boiling temperature. On the contrary, however, it was found that the yield of free amino acids during the hydrolysis of ovalbumin with 1 *N* HCl was almost exactly theoretical (Warner, 1942b).

b. Steric Effects. Apart from the effect of positively charged groups, probably the most important factor influencing the rate of hydrolysis of a peptide bond is the effective size of the amino acid side chains on either side the bond, preventing the approach of hydrogen ions by steric means. The effect will be expected to depend on the actual size of the side chain and its position relative to the bond in question.

Synge (1945) has made a kinetic study of the hydrolysis of a number of simple peptides by a mixture of equal volumes of 10 *N* HCl and glacial acetic acid at 37°. His results are shown in Table II.

The most stable peptides appear to be those containing valine. The bulky CH_3CHCH_3 group is close to the main peptide chain and effectively prevents the approach of hydrogen ions from a fairly wide angle.

In leucine the CH_3CHCH_3 is slightly further from the peptide bond, so that leucyl peptides are less stable than the corresponding valyl peptides, and peptides containing alanine and glycine are still more labile. A side chain seems to be less effective in stabilizing the peptide bond if it is on the residue whose amino group forms part of the bond (Levene *et al.*, 1932). Val.Gly for instance, is more stable than Gly.Val, though it should be noted that these generalizations are derived only from a study of peptides of glycine.

TABLE II

Hydrolysis of Dipeptides in Strong Acid (Synge, 1945)

Peptide	Relative velocity of hydrolysis (Gly.Gly = 1)
Gly.Gly	1
Gly.Ala	0.62
Ala.Gly	0.62
Gly.Leu	0.40
Gly.Try	0.35
Gly.Val	0.31
Leu.Gly	0.23
Leu.Leu	0.048
Leu.Try	0.041
Val.Gly	0.015

The marked stability of valine peptides has frequently been noted. Thus Christensen (1943) was able to isolate Val.Val in 1.5% yield from gramicidin after boiling for 24 hours with 16% HCl, and in 5–6% yield after 6 hours hydrolysis. Synge (1944) could find no free valine in partial hydrolyzates of gramicidin that had been treated with 5 *N* HCl for 10 days. The stability of valyl peptides was also apparent from the work on the partial hydrolysis of insulin (Sanger and Tuppy, 1951a). Thus tripeptides containing valine as the central residue were present in higher concentrations than other tripeptides. A lability of bonds involving the carboxyl groups of glycine was also noted. Peptides containing proline also appear to be unusually stable presumably due to steric factors (Consden *et al.*, 1947b).

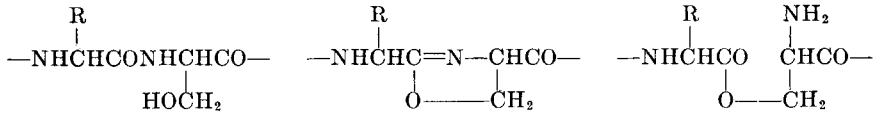
In this connection the unusual stability of cyclic structures should be mentioned. In the case of carbohydrates this was clearly demonstrated by Swanson and Cori (1948) and by Myrbäck (1949), who showed that the cyclic Schardinger dextrins are considerably more stable than corresponding open chain polysaccharides. Similarly Consden *et al.* (1947b) found the cyclopeptide "gramicidin S" to be unusually

resistant to acid hydrolysis. It has been suggested by Synge (personal communication) that immobilization of the peptide bond from rotation may be an important factor in inhibiting hydrolysis. Such an effect might also partly account for the relative stability of native proteins to enzymic hydrolysis (p. 26) and possibly for the stability of proline peptides.

c. Bonds Involving the Amino Groups of the Hydroxy Amino Acids. While the electrostatic and steric effects mentioned above probably play an important part in determining the stability of a given peptide bond, they are certainly not the only factors concerned and the course of hydrolysis of a protein cannot wholly be explained in this simple manner. In fact it has been observed that the first bonds in a protein to be split by strong acid are those involving the amino groups of the serine and threonine residues. Abderhalden and Bahn (1935) made use of this lability to prepare seryl peptides from the corresponding anhydrides. Under suitable conditions only one of the bonds in the anhydrides split.

Gordon *et al.* (1941) hydrolyzed a number of proteins with strong acid and followed the rate of liberation of the free amino groups of serine and threonine residues by means of the periodate reaction. Only those residues with a free hydroxyl and amino group react with periodate to give ammonia. The serine and threonine residues could be differentiated by estimating the acetaldehyde produced by periodate from threonine. It was found that the free amino groups of these hydroxy amino acids were liberated much more rapidly than the average amino group (estimated by the van Slyke method). Similar results were obtained by Christensen and Hegsted (1945).

More recently Desnuelle and Casal (1948) followed the relative rates of liberation of the amino groups of the different amino acids using the DNP method. During the initial stages of hydrolysis with 10 *N* HCl at 30° there appeared to be a very specific breakdown of the bonds involving the amino groups of serine and threonine. Thus after one hour about 20–30% of these bonds were split whereas very few other amino groups were liberated. There was no difference between serine and threonine. In order to account for this effect Desnuelle and Casal have suggested that under the action of concentrated HCl an intermediate oxazoline ring is formed which breaks at the amino group:



In this way the peptide chain is considered to migrate from the amino to the hydroxy group. The ester bond then formed will be fairly

rapidly broken down. Such reactions have been shown to take place with simple compounds such as benzoyl serine (Bergmann and Miekeley, 1924) under the action of chlorinating agents, *e.g.*, thionyl chloride. In support of the above mechanism it was found that the rate of liberation of the free amino groups of the hydroxy amino acids was greater when estimated by the DNP method than when estimated by the periodate method, thus suggesting that the hydroxy group was being simultaneously masked.

To sum up, when a protein is hydrolyzed with strong acid we may expect to find an initial rather specific hydrolysis liberating the free amino groups of the hydroxy amino acids followed by a more random breakdown at different bonds, the relative rates depending largely on the nature of the residues involved, and there is likely to be a slowing up of hydrolysis and accumulation of dipeptides towards the end of the reaction. In general, the specificity of hydrolysis will be greater at lower temperatures, since the activation energies for hydrolysis of the various bonds will be expected to show greater differences relative to the mean thermal energy of the molecules. This is evident in the results of Christensen and Hegsted (1945), who found a more random splitting at higher temperature. Desnuelle and Casal (1948) also found that the liberation of hydroxy amino residues was much more specific at lower temperatures.

4. *Hydrolysis in Dilute Acid*

While concentrated acid is usually preferred as a hydrolytic agent, it may be advantageous in certain cases to use dilute acid, which seems to exercise a rather different kind of specificity. Thus for instance very poor yields of the *N*-terminal glycyl peptides of insulin were obtained after hydrolysis in concentrated HCl, whereas much higher yields were obtained by boiling in 0.1 *M* HCl (Sanger, 1949b). This difference in specificity may to some extent be ascribed to differences in the charged groups. In sufficiently dilute acid the acidic groups may become negatively charged, so that they will attract hydrogen ions and labilize any nearby peptide bonds. An attempt to utilize this effect to obtain a specific hydrolysis near the cysteic acid residues of oxidized insulin gave no clear-cut results (Sanger, 1949b). A very specific hydrolysis has, however, been demonstrated by Partridge and Davis (1950). They found that when a protein was boiled with acetic or oxalic acid and the hydrolyzate investigated by paper chromatography, free aspartic acid was liberated much more rapidly than any other amino acid. For most proteins during the first 8 hours of hydrolysis the only ninhydrin positive spot was that due to aspartic acid. Free glutamic acid was the next

amino acid to appear, but at a very much slower rate. While it is possible that other bonds besides those next to the aspartic acid residues are also split, this type of hydrolysis seems to exhibit a marked specificity and may prove a useful method for obtaining larger polypeptides from proteins. Partridge and Davis suggest further that the presence of a free carboxyl group should labilize the *C*-terminal residues of a peptide chain and it may thus be possible to effect a stepwise liberation of residues from the *C*-terminal group.

The main disadvantage of using dilute acid for hydrolysis is the possibility of anhydride formation, already referred to (p. 16).

Another type of hydrolysis that may exhibit a different specificity is that catalyzed by long-chain anionic detergents. Steinhardt and Fugitt (1942) showed that the rate of hydrolysis of proteins by acid was dependent, not only on the concentration of hydrogen ion, but also on the nature of the anion present. For a homogeneous series of long chain sulfonic acids the rate of hydrolysis was proportional to the chain length of the anion. Thus the hydrolysis proceeded about a hundred times as rapidly in dodecylsulfonic acid as in HCl. It seems that the detergent is adsorbed on the protein and the presence of the $—SO_3^-$ group labilizes the peptide bonds. At low concentrations of detergent the amide groups are especially labile to this type of hydrolysis. Being slightly stronger bases than the peptide bonds, they preferentially bind the anions (Steinhardt, 1941). Thus it was possible to remove amide groups with very little destruction of peptide bonds. This suggests a different type of specificity from other methods of hydrolysis. The lability of amide groups, however, seems to be a characteristic of low temperature hydrolysis in dilute acids, since Virtanen and Hamberg (1947) hydrolyzed zein at pH 1.5–1.8 and 37° for 4 months and found that 50% of the amide groups were split off while very little amino N was liberated.

An obvious advantage of using this method of catalyzed hydrolysis is that there will be a minimum of destruction of the amino acids or of any bonds other than the peptide bonds. A disadvantage of the method is the practical difficulty of working with solutions of these strongly surface-active reagents.

5. *Hydrolysis in Alkali*

The main disadvantage of alkaline hydrolysis is that much more destruction of amino acids occurs than with acid. The chief amino acids to be affected are cysteine, serine, threonine and arginine. Wieland and Wirth (1949) have used paper chromatography to study the effect of strong alkali on certain amino acids. They found that serine broke down to give appreciable quantities of glycine and alanine, threonine

gave glycine, alanine and α-amino butyric acid and cysteine gave alanine. While relatively violent conditions were required for these reactions, they were shown to take place more readily if the amino acids were in peptide form. Hellerman and Stock (1938) and Warner (1942a) showed that arginine is broken down by strong alkali to give ornithine and citrulline.

Sanger and Tuppy (1951a) studied an alkaline hydrolyzate of an oxidized insulin fraction. Although the conditions used would not have been expected to cause any destruction of free serine or threonine, it was evident that most of these residues in the protein had been broken down. Thus, for instance, Gly.Pro was found in the alkaline hydrolyzate whereas Thr.Pro was present in the acid hydrolyzate. The arginine residues were also converted to ornithine (or citrulline).

Extensive racemization of the amino acids also occurs in the presence of alkali, which may complicate the results of a partial hydrolysis experiment (Levene and Bass, 1928, 1929).

In spite of these obvious disadvantages it may, in certain cases, be worth while to employ alkaline hydrolysis, which may exhibit somewhat different specificities from those found for acid hydrolysis. Synge (1945) compared the relative rates of hydrolysis in acid and in alkali of a number of simple dipeptides of the monoamino acids, and found them to be similar in both reagents. In these experiments the relative stabilities of the peptide bonds were probably determined largely by the steric effects produced by the side chains of the residues, which inhibit the approach of hydroxyl ions, and these effects will be similar for both acid and alkali hydrolysis. However, where the effects of charged groups are concerned one may expect to find a somewhat different specificity. In strongly alkaline solution the basic groups are uncharged, whereas the carboxyl groups will be negatively charged and will stabilize bonds in their neighborhood. Similarly, dipeptides would be stable in alkaline solution due to the presence of the charged α-carboxyl group.

It is also probable that the peptide bonds involving the amino groups of the hydroxyamino acids are more stable in alkali than in acid. Abderhalden and Bahn (1935) showed that whereas anhydrides containing serine were split by acid at the bond involving the amino group of the serine residue, they were split at the other bond by alkali. It was also found by the DNP method (Sanger, unpublished) that the relative rate of liberation of the amino group of serine from fraction A of oxidized insulin was slower in alkali than in strong acid.

Alkaline hydrolysis offers a possible advantage for the investigation of tryptophan peptides since tryptophan itself is more stable in alkali than in acid (Lugg, 1938; Brand and Kassell, 1939).

The possibility that long chain bases, such as hexyl trimethyl ammo-

nium, will show catalytic effects similar to those found for long chain anions is suggested by the work of Steinhardt and Zaiser (1950), who showed that they are also bound to protein and cause anomalous titration effects. It may be that the mild conditions necessary for such an hydrolysis may make it possible to avoid the undesirable side reactions that occur during alkaline hydrolysis.

6. *Hydrolysis with Proteolytic Enzymes*

The action of proteolytic enzymes on proteins and synthetic peptides has been extensively studied and reviewed (see Bergmann and Fruton, 1941; Bergmann, 1942; Neurath and Schwert, 1950; Linderstrøm-Lang, 1949) so that the subject will not be dealt with in great detail here.

One obvious advantage of using proteolytic enzymes is that they are very unlikely to cause any destruction of amino acid residues, since they act under very mild conditions. Another important advantage is their specificity. Not only do they exhibit an entirely different specificity from that shown by acid and alkali but a specificity which is much more limited and exacting. Enzymes should therefore be useful for the initial specific splitting of proteins into large peptides. The finding of synthetic substrates of known structure that are hydrolyzed by the endopeptidases has made it possible to predict what bonds will be susceptible to particular enzymes. This subject has been reviewed in detail by Bergmann and Fruton (1941) and by Neurath and Schwert (1950). In brief it may be said that trypsin splits those bonds in which the carboxyl groups of arginine or lysine are involved, chymotrypsin those involving the carboxyl groups of tyrosine, phenylalanine, tryptophan or methionine, and pepsin those involving the amino groups of the aromatic amino acids. It seems that the rate at which synthetic substrates are split is considerably slower than the rate at which peptide bonds in proteins are split (Northrop *et al.*, 1948). The question therefore arises as to whether bonds of the type present in the synthetic peptides are the most labile in a protein or whether there are other types of bonds that are broken down. In the case of pepsin, there is evidence that other types of bonds are also split. Thus Harington and Pitt-Rivers (1944) found that pepsin would hydrolyze the peptide Tyr.CySH, in which the bond involves the carboxyl group and not the amino group of tyrosine. Recently Desnuelle *et al.* (1950) have studied the action of pepsin on ovalbumin and on horse globin using the DNP technique. If the above specificity were the only one, the only free amino groups liberated should be those of the aromatic residues. In fact this was not the case. For ovalbumin no specificity could be detected, and free amino groups of almost all amino acids were liberated simultaneously. With globin the bonds split first were those

involving the amino groups of alanine, phenylalanine, leucine, and serine; during a second slower phase of hydrolysis no specificity was apparent.

In a study of the action of proteolytic enzymes on an oxidation product (fraction B, p. 54) of insulin, Sanger and Tuppy (1951b) found that other bonds besides those adjacent to aromatic residues were split by pepsin, including those of Leu-Val, Ala-Leu, and Glu(—NH_2)-His. It thus seems that in the case of pepsin at least there is much to be learnt about its specificity when proteins act as substrate. Trypsin and chymotrypsin were found to split oxidized insulin with the same specificity as was found for synthetic peptides, and it seems probable that this specificity may be shown in their action on other proteins. Clearly a knowledge of the exact mode of action of these enzymes would greatly help in the elucidation of protein structure just as advances in our knowledge of protein structure must throw light on the behavior of the endopeptidases.

The possibility of rearrangement of sequences of amino acids under the action of proteolytic enzymes has already been mentioned (p. 15) but this danger would not seem sufficiently great to offset the advantages to be gained by using them. Nevertheless, the results must be interpreted with caution.

It has already been emphasized (p. 14) that the best stage of hydrolysis at which to attempt the fractionation of peptides is at the point where enzyme action will proceed no farther or when there is a very sharp break in the hydrolysis curve. The disadvantage of a long incubation period is of course that the danger of rearrangements increases (see Linderstrøm-Lang and Ottesen, 1949).

It may be that proteolytic enzymes act more specifically on native than on denatured proteins. The work of Linderstrøm-Lang and coworkers (reviewed by Linderstrøm-Lang, 1949) has indicated that at least in the case of the action of trypsin on β-lactoglobulin, the initial step in proteolysis is some type of denaturation. Nevertheless, certain hydrolyses proceed without denaturation of the substrate. Examples are the formation of plakalbumin from ovalbumin (p. 57) and the splitting of globulin molecules by papain (Petermann and Pappenheimer, 1941; Petermann, 1942), which appear to be rather specific reactions. Presumably only a few susceptible sites are available to the enzyme on the native protein, and if these can be split without denaturing the protein and exposing the other susceptible sites, a specific type of hydrolysis may take place.

7. *Rates of Hydrolysis of Proteins*

In experimental work it is very often desirable to know how far a protein will be hydrolyzed under a given set of conditions. The course

of hydrolysis of a number of proteins in various concentrations of acid and alkali have been recorded in the literature, though it is often difficult to find the particular and relevant data one is interested in. In Table III are listed some references where such data may be found. Since protein hydrolysis follows no known kinetic laws it is impossible to define the rate in terms of any constant, so that it is best to refer to the original work. In the last column are listed the approximate half-lives of the proteins in the various reagents. This is defined as the time at which 50% of the peptide bonds in the protein are broken.

8. *Non-Hydrolytic Methods of Degradation*

Critics of the peptide theory have claimed that too much emphasis has been laid on studies in which hydrolysis is used for degrading proteins, and that other methods of degradation such as oxidation and reduction should be employed. In the earlier years of protein chemistry many attempts were made to study proteins in this way but very few recognizable products could be obtained, and with those that could be identified it was difficult to know what was their relationship to the original protein. This was undoubtedly due to the great complexity of the problem, the different side chains and bonds in the proteins each reacting to give a variety of products of unknown origin. Clearly, in order to obtain any recognizable breakdown products, one must use reagents that attack only one or a very few types of bond or residue. Hydrolytic reagents, which attack almost exclusively the peptide bond, are the obvious first choice. In the search for other reagents for degrading proteins it would seem more profitable to consider the nature of proteins first and to choose a reagent for a particular purpose rather than to take any reagent off the shelf and see what it does to the protein. There is obviously a great use for reagents that will attack a protein in a specific manner. If it were possible to destroy one type of residue exclusively and split it out of a protein chain, it would be a valuable step in the degradation of proteins.

a. Splitting the Disulfide Bridges. Probably the only covalent linkage that occurs in proteins other than the peptide bond between the amino acid residues, is the disulfide bridge of cystine; this may be split by oxidation to sulfonic acid groups. Toennies and Homiller (1942) studied the action of performic acid on a large number of amino acids and found that the only acids attacked were cystine, methionine, and tryptophan. Cystine was quantitatively converted to cysteic acid, and methionine to the corresponding sulfone. In the case of tryptophan the products were not identified. Sanger (1949a) used performic acid to split the disulfide bridges of insulin, which contains no tryptophan or methionine (see p. 51). This method could probably be used as an

TABLE III

Rates of Hydrolysis of Proteins[a]

Reference	Proteins studied	Reagents used	Temp. (°C.)	Approx. half-life of protein (hours)
Abderhalden and Mahn, 1928	Gelatin	1 *N* HCl	15, 38, 50, 70	
Acher *et al.*, 1950	Lysozyme	10 *N* HCl	37	96
Bull and Hahn, 1948	Ovalbumin	7.95 *N* HCl	30	195
		7.95 *N* HCl	45	42
		7.95 *N* HCl	60	9
Desnuelle and Casal, 1948	Casein	2.5 *N* HCl	Boiling	2.3
		10 *N* HCl	37	115
		10 *N* HCl	30	192
	Silk fibroin	10 *N* HCl	30	144
Dunn, 1925	Casein	3.6 *N* H_2SO_4	Boiling	5
Gordon *et al.*, 1941	Wool	10 *N* HCl	37	95
	Edestin	10 *N* HCl	37	120
	Gelatin	10 *N* HCl	37	35
Levene and Bass, 1928	Casein	5 *N* HCl	125	
		0.5, 1.0, 5 *N* NaOH	25	
Pittom, 1914	Casein	5.7 *N* HCl	Boiling	
	Ovalbumin	5.7 *N* HCl	Boiling	
Stein *et al.*, 1944[b]	Silk fibroin	12 *N* HCl	40	17
Vickery, 1922	Gliadin	0.1, 0.2 *N* HCl	93–94	
		0.5 *N* HCl	93–94	35
		1.0 *N* HCl	94–95	17
		2.0 *N* HCl	94–95	7
		4.0 *N* HCl	98–104	1.8
		20% HCl	102–110	< 1
		0.2 *N* H_2SO_4	93–94	
		4.0 *N* H_2SO_4	96–98	6
		0.2 *N* NaOH	93–94	
		1.0 *N* NaOH	93–94	16
		0.2 *N* $Ba(OH)_2$	93–94	36
Warner, 1942b	Ovalbumin	1 *N* HCl	Boiling	4.5
		20% H_2SO_4	100	
		1.4, 4.3 *N* NaOH	100, 68, 35	
		0.43 *N* NaOH	100	
		0.2 *N* NaOH	68	
		2.3, 3.7 *N* $Ba(OH)_2$	100	

[a] For other earlier references see Vickery, 1922.

[b] In these experiments 1 g. protein was hydrolyzed with only 2 ml. HCl.

initial specific method of degrading other proteins that contain disulfide bridges, though the presence of tryptophan might lead to side reactions. Disulfide bridges can also be split specifically by mild reduction, but the thiol groups formed tend to reoxidize and cause polymerization of the products (Miller and Andersson, 1942).

b. Splitting by Radiation. If a protein is irradiated with ultraviolet light of a suitable wavelength, the only residues that absorb energy are those of the aromatic amino acids phenylalanine, tyrosine, and tryptophan, and it has been suggested that the photochemical energy may be sufficient to split the peptide bonds adjoining these residues (Carpenter, 1940; McLaren, 1949; Mandl *et al.*, 1950). Rideal and Mitchell (1937) showed that substances such as stearic anilide could be split. Here the aromatic residue is directly adjoining the peptide bond. However it was also shown that cleavage could take place when the peptide bond and aromatic residue were separated by a number of $—CH_2—$ groups, since substances such as stearyl benzylamine were also split (Carpenter, 1940). Propionylphenylalanine and phenylpropionylalanine, which resemble more closely a natural peptide were both split with equal efficiency (Mandl *et al.*, 1950). In each case the main nitrogenous end product was ammonia rather than the amino acid, indicating that deamination had also occurred. No liberation of free amino acids from Tyr.Leu or Leu.Tyr could be detected but this was probably due to deamination. However it has been shown (Carpenter, 1941; Kaplan *et al.*, 1950) that when insulin is irradiated free tyrosine is liberated into the solution. Clearly the effects of light on a protein may be expected to be rather complex (McLaren, 1949) but it would be extremely valuable if conditions could be found for a specific photolysis near the aromatic residues.

c. Other Methods. Fodor and coworkers (see Fodor, Fodor and Kuk-meiri, 1947) used anhydrous glycerol at 130–140° to break down proteins. The products formed, which were termed "acropeptides," appeared to be large cyclic polypeptides. Using a similar method Uchino (1934) and Tazawa (1949) reported the formation of large amounts of diketopiperazines from ovalbumin.

Troensgaard (1947) has used reduction with sodium and amyl alcohol to split proteins and has isolated a number of piperazines and pyrrole derivatives. The relationship of these products to structures in the original protein is not clear at present.

V. Fractionation of Peptides

It has already been emphasized that a partial hydrolyzate of a protein is a complex mixture of closely related compounds, so that very sensitive

methods of fractionation are required to separate pure peptides. In recent years countercurrent methods, especially those depending on partition effects have been found to be most useful for this purpose. Before applying such methods it is often desirable to carry out preliminary group separations of the hydrolyzate into fractions containing fewer peptides than the original mixture. For instance, not more than about twenty peptides can be satisfactorily fractionated by paper chromatography. Thus if a partial hydrolyzate of a protein were applied directly to such a chromatogram there would probably be considerable overlapping of the spots and interpretation of the results would be difficult. Separation into a number of relatively simple groups would make fractionation much easier. We may thus distinguish between two classes of methods of fractionation, those for preliminary group separations and general methods involving the countercurrent principle. The distinction is somewhat arbitrary as most methods may be used to some extent for both purposes. The most useful methods for group separation are those which give the most clear-cut fractionations with a minimum of overlapping, the ideal being that each peptide should be present in only one fraction. Such an ideal is rarely achieved but it is sometimes possible, as in the case of the cystine peptides, to make use of a unique property of one amino acid, to separate out all peptides in which it is involved. It is clearly an advantage to use for group separations methods which depend on different properties of the peptides than those on which the final general fractionation depends. Thus, whereas the methods of general fractionation usually depend on differences in the partition coefficients of the peptides, methods of group separation more often depend on differences in ionophoretic mobility or in the adsorption coefficients.

1. *Ionophoretic Methods*

Since amino acids and peptides contain several differently charged groups, they can be fractionated by methods which make use of differences in isoelectric point or electrophoretic mobility. The various techniques available for such separations have been comprehensively reviewed by Svensson (1948).

A simple compartment type of apparatus may be used to separate a peptide mixture into basic, neutral and acidic fractions (Gordon *et al.*, 1941, 1943; Sanger and Tuppy, 1951a). Since the simplification of the mixture is usually more important than the yield of peptides, it is often advisable to repeat the ionophoresis on each fraction, and in this way clear cut fractionations may be obtained with very little "overlapping." This method is especially useful for separating the basic peptides. By

choosing suitable conditions of pH etc. it is also possible to carry out more extensive group separations using this type of apparatus; thus, by working at an alkaline pH for instance it is possible to obtain a fraction containing only arginine peptides (Sanger and Tuppy, 1951a).

For more extensive separations it is best to use methods that depend on differences in mobility. The method that has been used most is that of Consden, Gordon and Martin (1946) for ionophoresis in silica jelly. It was successfully used to separate the acidic peptides from a partial hydrolyzate of wool into fractions of different mobilities (Consden *et al.*, 1949; Consden and Gordon, 1950). This type of apparatus is less suitable for peptides containing aromatic or basic residues, due to adsorption on the silica, which causes "tailing" and difficulties in eluting the peptides. The use of agar-agar instead of silica may prove more suitable for such peptides (Gordon *et al.*, 1949).

Svensson and Brattsten (1949) and Grassmann (1950) have described a method of ionophoresis in which the separations are carried out in a box of glass powder. The solution is allowed to flow down the box while a voltage is applied horizontally, so that the direction of flow of each solute depends on its mobility; fractions are collected from outlet tubes at the bottom of the box. This apparatus should prove especially useful for longer peptides and proteins, since adsorption on the stabilizer (glass powder) is unlikely. A similar continuous method using filter paper was described by Grassmann and Hannig (1950).

A number of other workers have devised methods for ionophoresis on filter paper. Haugaard and Kroner (1948) have combined ionophoresis with paper chromatography by threading electrodes down the side of a paper chromatogram, which was first soaked in buffer solution. A potential was applied during the development of the chromatogram with phenol. Wieland and Fischer (1948) have used an apparatus in which a potential is applied to the ends of a strip of filter paper which is soaked in buffer and held in a small glass chamber. The ends of the strip are outside the chamber and dip into the electrode vessels. Good separations with amino acids and proteins (Turba and Enenkel, 1950) were reported. A similar principle is employed in the methods reported by Biserte (1950), Durrum (1950) and Cremer and Tiselius (1950). The method of the latter authors was designed specially for the fractionation of proteins. The whole paper is immersed in a bath of chlorobenzene, which prevents heating and loss of solvent by evaporation.

2. *Ion Exchange Methods*

Another method of fractionating peptides according to their charge is by the use of ion exchange materials. These have frequently been

used for the group separation of amino acids (reviewed by Turba, 1948; Block, 1949; Martin and Synge, 1945). In most work the amino acids have merely been separated into two groups in each experiment, those which are adsorbed, and those that are not adsorbed, the latter being eluted with a different solvent. In this way it is possible to obtain quite sharp separations by choosing suitable adsorbents and conditions of pH etc. Thus, for instance, basic amino acids may be adsorbed on "basic Al_2O_3" (Wieland, 1942), silica gel (Schramm and Primosigh, 1944), permutite (Felix and Lang, 1929) or suitable synthetic resins (Block, 1942; Wieland, 1944). Acidic amino acids may be adsorbed on "acid Al_2O_3" (Wieland, 1942; Jutisz and Lederer, 1947; Turba and Richter, 1942) and synthetic anion exchange resins (Cannan, 1944; Tiselius *et al.*, 1947). The synthetic resins are probably to be generally preferred because of their greater capacity and because their properties are more reproducible from batch to batch. There are also a large number of different ones now commercially available, which extends their use. It is probable that simple group separations of this type may also be carried out with peptide mixtures (*cf.* Waldschmidt-Leitz and Turba, 1941) though in such mixtures of substances with a much more scattered range of isoelectric points, considerable overlapping may be expected. Also large peptides and those containing aromatic residues will probably be held on the exchangers by ordinary adsorption which may cause difficulties.

Jutisz and Lederer (1947) and Lederer and Tchen (1947) have devised a method for the group separation of neutral amino acids and peptides by making use of differences in the apparent pK of the amino groups in the presence of formaldehyde. Neutral amino acids and peptides are not adsorbed on "acid Al_2O_3" from water but in the presence of formaldehyde they acquire acidic properties and those with the lower pKa′ values are adsorbed. Table IV shows the approximate pKa′ values for the three groups in 10% formaldehyde, and their behavior on columns of "acid Al_2O_3" in the presence of 10% and 1% formaldehyde, respectively. In this way three clearly separated groups may be obtained. Presumably seryl, threonyl, and cysteinyl peptides will behave as the glycyl peptides. Final elution from the columns is effected with hot water which dissociates the formaldehyde complexes. This method, which gives rather a different group separation from other techniques may prove useful for the initial simplification of peptide mixtures.

Under suitable conditions it is possible to obtain separations on an ion exchange column by elution analysis. The pH of the developing solution should be such that the solutes are distributed between the resin and solvent so that they move down the column as definite bands.

In this way Consden *et al.* (1948) and Drake (1947) separated glutamic and aspartic acid on columns of Amberlite IR-4 (polyamine anion exchanger) maintained at pH 2.5, and this type of column has been used to separate acidic peptides from a wool hydrolyzate (Consden *et al.*, 1949). More recently Stein and Moore (1949) have obtained excellent fractionations of amino acids on columns of Dowex-50 (cation exchanger, sulfonic acid groups) developed with 1.5–4 *N* HCl. The separations obtained are almost certainly due to differences in adsorption affinity for the resin, as well as to charge effects. At the pH used the only partially

TABLE IV

Group Separation of Amino Acids and Peptides according to Jutisz and Lederer (1947)

Group	Approx. pKa′ in 10% H·CHO	Adsorption on "acid Al_2O_3" in 10% H·CHO	in 1% H·CHO
1. Monoamino acids other than those in group 2	7.0	−	−
2. Glycine Serine Threonine Cysteine Simple dipeptides other than those in group 3	5.8	+	−
3. Glycyl peptides	4.2	+	+

ionized groups will be the sulfonic acid groups of the resin, the amino acids having either one or two positive charges. This technique appears to be one of the most efficient methods of fractionation and it will be interesting to see how far it may be used for the separation of peptides. In contrast to most other methods depending on adsorption (see next section) the bands exhibit no tailing, but give elution curves showing sharp well-defined peaks.

Separations may also be carried out on ion exchangers by the principle of displacement chromatography. If the ionizing groups of a resin are completely saturated with an amino acid, and a second amino acid, having a greater affinity for the resin is introduced onto the column, the first amino acid will be displaced and will move down the column. In this way columns may be obtained from which the amino acids are displaced in the order of their isoelectric points. Partridge (1949a, b, c; Partridge and Westall, 1949) has developed this principle using the sulfonic acid resin "Zeo-Karb 215," for the separation of amino acids. Elution is carried out with ammonia which has a greater affinity for the resin than most of the amino acids. This method, which is best carried

out on a relatively large scale, may prove useful for the group separation of peptides, but no such results have yet been reported.

3. *Adsorption Chromatography*

The possibilities of fractionating amino acids, peptides and proteins by adsorption chromatography have been fairly extensively explored, especially by Tiselius and coworkers (reviewed in Tiselius, 1947; Turba, 1948; Martin and Synge, 1945). Active carbon is probably the most effective adsorbent for amino acids, being one of the few on which they are retained. Aromatic amino acids are strongly adsorbed and may be separated from the other amino acids (Schramm and Primosigh 1943; Jutisz and Lederer, 1947; Partridge, 1949b) by adsorption from acetic acid solution. They can be eluted then with phenol or ethyl acetate. A similar type of group separation may also be used to separate aromatic peptides from a partial hydrolyzate (Synge and Tiselius, 1949; Sanger and Tuppy, 1951a) though larger peptides and peptides containing basic amino acids, are rather strongly adsorbed, and may appear in the "aromatic" fraction. Considerable losses are often involved in this type of fractionation, as it is difficult to elute the substances completely.

Attempts to use adsorption on charcoal for a more detailed chromatographic fractionation of peptides have not met with great success. Adsorption isotherms are usually of the Langmuir type and the only method that is generally applicable is the technique of frontal analysis. This may be used as an analytical method (Moring-Claesson, 1948; Synge and Tiselius, 1947, 1949), but does not give appreciable fractionation. Usually the bands tail too much to make separations possible by elution analysis. The most satisfactory method for separating compounds having this type of adsorption isotherm is the displacement method, though unfortunately it is not a general method and has only been used in a few special cases. One of the disadvantages of displacement analysis is that there is no intermediate zone between two consecutive bands of adsorbed substances, so that a clear separation is almost impossible. This difficulty has been very neatly overcome by Tiselius and Hagdahl (1950) by the addition of a volatile substance having an adsorption affinity intermediate between those of the two substances to be separated. For instance when a mixture of methionine, Leu.Gly.Gly and *n*-butanol was subjected to displacement chromatography, the bands were eluted in the following order: methionine, *n*-butanol, Leu.Gly.Gly. By cutting in the middle of the *n*-butanol fraction it was possible to obtain complete separation of the methionine and Leu.Gly.Gly.

It is likely that adsorption methods may be useful for fractionating larger peptides, which cannot readily be separated by other methods.

Thus Synge and Tiselius (1947) were able to fractionate the components of tyrocidin, both by elution and displacement methods. In a study of the partial hydrolysis of ovalbumin Moring-Claesson (1948) was able to separate by adsorption the unchanged protein from the breakdown products. The former was adsorbed much more strongly on alumina and less strongly on carbon than the smaller peptides and amino acids.

On most of the more commonly used adsorbents, peptides and proteins are only weakly adsorbed from aqueous solution. Tiselius (1948; Shepard and Tiselius, 1949) has shown that in the presence of a high concentration of salt, substances are much more strongly adsorbed, thus making it possible to use adsorbents such as paper or silica for the chromatography of proteins (see also Mitchell *et al.*, 1949). This technique which is known as "salting out adsorption" may prove useful for the fractionation of larger peptides.

Hamoir (1945) was able to fractionate amino acids into four groups by adsorption on silver sulfide. In general those amino acids that form the least soluble silver salts were the most strongly adsorbed. The order in which the amino acids are adsorbed on the column, is rather different from the order on other columns, so that a different type of group separation may be obtained.

4. *Partition Chromatography*

Partition chromatography was originally introduced as a method for the fractionation of acetamido acids (Martin and Synge, 1941) and was early applied to the separation of acetamido peptides from a partial hydrolyzate of gelatin (Gordon, Martin and Synge, 1943). In this method the fractionation was carried out on columns of silica gel and the bands were located by incorporating an indicator in the aqueous phase of the chromatogram. The mixture studied was rather too complex to give really satisfactory resolution of the peptides and since much better results are obtained by direct fractionation of peptides on paper or starch chromatograms, the technique has not been used further.

a. Paper Chromatography. Undoubtedly the most satisfactory procedure for fractionating amino acids and smaller peptides is paper chromatography (Consden *et al.*, 1944) which is now familiar to most workers. The method has been the subject of a number of reviews (*e.g.* Consden, 1948; Martin, 1950; Jones, 1949) and its application to peptides has been described by Consden *et al.* (1947a). In this method samples containing about 1 mg. of the peptide mixtures to be analyzed are fractionated on two-dimensional chromatograms. The position of the peptide spots is determined either by spraying with ninhydrin, which causes only a small amount of destruction of the peptide (see Woiwod, 1949) or from the

fluorescence which is produced if the paper is heated (Phillips, 1948). The exact nature of this latter reaction is still obscure but it seems to depend both on the peptide and on the paper (Patton *et al.*, 1949). It is extremely valuable for locating spots with a minimum of concomitant destruction. The peptides may then be cut out and eluted by running a small amount of water through the "cut." Microtechniques for identifying the constituent amino acids and the *N*-terminal residues of these peptides have been described by Consden *et al.* (1947a) (see p. 48).

The choice of suitable solvents for running the chromatograms will depend on the nature of the particular mixture to be analyzed. Probably the most generally useful ones are phenol, collidine and butanol-acetic acid mixtures (Partridge, 1948). The latter are especially useful for the larger peptides that "tail" badly on other solvents (Jones, 1948, 1949). Thus Phillips (1949a) found that a mixture of peptides from insulin which could not be satisfactorily fractionated on phenol or collidine could be separated into well-defined spots using butanol-acetic acid. By varying the acetic acid content, mixtures with different properties can be prepared.

It is not possible at present to predict the exact position of a peptide spot on a chromatogram from a knowledge of its composition, but an approximate determination of its R_F values may be obtained as follows. Martin (1949a) has shown theoretically that the partition coefficient of a dipeptide divided by the product of the partition coefficients of the constituent amino acids is a constant for any given phase pair, *i.e.*, that for a peptide A.B

$$\frac{\alpha_{A.B}}{\alpha_A \alpha_B} = \text{constant}$$

where α_A, α_B and $\alpha_{A.B}$ are the respective partition coefficients. Consden *et al.* (1944) showed that the R_F values are related to the partition coefficients by the equation

$$\alpha = \frac{A_L}{A_S}\left(\frac{1}{R_F} - 1\right)$$

$\frac{A_L}{A_S}$ is a constant for any solvent system so that

$$\frac{\left(\frac{1}{R_F^{A.B}} - 1\right)}{\left(\frac{1}{R_F^{A}} - 1\right)\left(\frac{1}{R_F^{B}} - 1\right)} = \text{constant} = K$$

$$R_F^{A.B} = \frac{R_F^A R_F^B}{K(1 - R_F^A)(1 - R_F^B) + R_F^A R_F^B}$$

Thus the R_F value of a dipeptide may be calculated in terms of the R_F values of the constituent amino acids and of a constant K, which is best determined experimentally for each solvent system. The relationship is not absolutely accurate, since, for instance, peptides containing the same amino acids in different order may frequently be separated. It has, however, been found to apply satisfactorily in most cases and to be a useful check on the identity of a peptide.

b. Starch Chromatography. Starch chromatography, originally introduced by Elsden and Synge (1944; Synge, 1944) has been developed into a very accurate method of amino acid analysis by Moore and Stein (1948, 1949; Stein and Moore, 1948). Amino acids and simple peptides move as sharp well defined bands and excellent resolutions may be obtained. Synge (1944, 1949) fractionated a partial hydrolyzate of gramicidin on starch and was able to identify a number of peptides. Amino acids and peptides on starch columns behave much as they do on paper chromatograms; but the use of starch makes possible separations on a larger scale. The efficiency of separations is probably rather less on starch and it is not possible of course to use the two-dimensional technique. Starch chromatograms run extremely slowly and it is necessary to run each column for several days to obtain satisfactory fractionations.

c. Other Partition Chromatography. A number of other types of partition chromatograms have been suggested and may be useful in the separation of peptides.

Paper chromatography is only applicable on a micro scale and several attempts have been made to extend it to a larger scale. Mitchell and Haskins (1949) have described a "chromatopile" for such fractionations. This consists of a pile of filterpapers which are used as the column. Jones (1949) suggests the use of thick paper which may best be run by ascending chromatography. Recently there has become available preparations of powdered cellulose which are suitable for use in columns.

Diatomaceous earth (Kieselguhr) may be used as an inert support for the aqueous phase of a partition chromatogram (Martin, 1949a; Bell *et al.*, 1949). Being only a very weak adsorbent, it is likely to be useful for larger peptides which may tail badly on other chromatograms due to excessive adsorption.

Certain hydrophobic peptides are difficult to fractionate as they all tend to run fast on the usual chromatograms in which water is the stationary phase (Synge, 1949). Recently systems have been described using adsorbents such as rubber, in which an organic solvent is held as the stationary phase and the column is developed with water or a buffer solution (Boldingh, 1948; Howard and Martin, 1950).

5. *Detection of Peptides from Columns*

In any of the chromatographic methods described above it is necessary to have some method for locating and if possible, for estimating the colorless peptides on the column or as they are eluted from it. In paper chromatography they are located on the paper but where a column arrangement is used it is usually simpler to identify the bands as they are eluted. This is most usually done by collecting the effluent in a large number of small fractions. A number of automatic fraction collectors have been described that may be used for this purpose (Stein and Moore, 1948; Randall and Martin, 1949; Phillips, 1949b) and some are commercially available. A suitable test may then be applied to each fraction. For this purpose the ninhydrin reaction is most generally used (Moore and Stein, 1948). Drake (1947) has described an automatic arrangement for spotting aliquot drops of the effluent on to filter paper, which can be developed by a suitable reagent. This could probably be used in conjunction with an automatic fraction collector.

Several methods have been described for continuously recording the concentration of solute in the effluent from a column. Tiselius and Claesson (1942) have observed changes in the refractive index of the solution using a special interferometer. This method has proved very useful where adsorption chromatography has been used, but is not suitable for use with partition chromatograms as any slight changes in the composition of the solvent will cause changes of refractive index and will interfere with the recording.

A small conductivity cell attached to the bottom of a column may also be used to locate the bands of amino acids and peptides, which will cause changes in the conductivity of the solution (Randall and Martin, 1949).

Recently Drake (1950) has described a polarographic method for following the chromatography of proteins. Only substances containing cystine or cysteine will be detected but it may be useful where large peptides are being studied. The method is unaffected by changes in the solvent or salt concentration.

6. *Countercurrent Distribution*

The development of the method of countercurrent distribution is largely due to the work of Craig and his associates (see Craig *et al.*, 1949). The method is applicable to the separation of any substances that can be reversibly distributed between two immiscible solvents, and these, of course, include peptides. Several different types of apparatus

have been devised for this purpose (Craig, 1944; Kies and Davis, 1950; Rometsch, 1950), the most efficient probably being the all-glass apparatus recently designed by Craig and Post (1949).

Compared with chromatographic methods, this countercurrent distribution method is probably less efficient, the apparatus required is very much more complicated, and the labor involved is considerably greater; nevertheless, it possesses several definite advantages over chromatographic methods. The chief of these is the absence of any solid phase, which may act as an adsorbent for the solutes. In partition chromatographic methods this adsorption may often lead to distortion of the bands and render fractionation very inefficient. The behavior of a solute on a countercurrent distribution depends only on its partition coefficient, which in most cases is constant, so that it is possible to calculate the exact theoretical distribution curve and this may be used as a very sensitive test for purity.

The method is especially useful for the fractionation of larger polypeptides, which cannot be fractionated easily by other techniques. Thus, for instance, Gregory and Craig (1948) studied crystalline gramicidin by this method, and found it to consist of at least three components, although it had previously been thought to be pure. Other naturally occurring polypeptides have similarly been purified (Craig *et al.*, 1949; Barry *et al.*, 1948; Livermore and du Vigneaud, 1949).

7. *Lysine Peptides*

The DNP method has been used for the separation and identification of peptides containing lysine (Sanger, 1949b). When a DNP-protein is partially hydrolyzed, the only colored products present are the DNP derivatives of the *N*-terminal peptides and of those peptides which contain lysine residues. The DNP-terminal peptides are extracted by an organic solvent as already described (p. 7). This procedure can be regarded as a very specific type of group separation in which only a few special peptides are separated out. The peptides containing ϵ-DNP-lysine, which remain in the aqueous hydrolyzate solution mixed with other unsubstituted peptides, can be separated by adsorption on talc from acid solution, since the DNP group is held strongly on this adsorbent. Elution can be effected with acid ethanol and the peptides subsequently fractionated on suitable partition chromatograms. FDNB also reacts with tyrosine and histidine residues to give colorless products, which are likely to be retained by the talc. In the case of insulin no such interference was observed, but it is doubtful if the method could be applied successfully to other more complicated proteins.

8. *Cystine Peptides*

Consden and Gordon (1950) have described an elegant method for investigating peptides involving cystine residues. After removal of acidic peptides from a partial hydrolyzate using an ion exchange column (Amberlite IR-4), the remaining peptides are oxidized with bromine, which converts the cystine residues to cysteic acid residues. The acidic peptides so formed can be separated from the remaining neutral peptides on another ion exchange column. Only peptides which originally contained cystine or cysteine linked with neutral amino acids appear in this fraction as cysteic acid peptides. This simplified mixture may then be fractionated by ionophoresis and paper chromatography.

9. *Other Methods of Fractionation*

So far we have considered only the newer countercurrent methods for fractionating peptides. These will probably play a predominant role in the future, though the classical methods of fractional crystallization and precipitation should not be forgotten. They are still the most effective methods of fractionating proteins and probably larger polypeptides, such as the oxidation products of insulin (Sanger, 1949a).

The aromatic sulfonic acids, which were developed by Bergmann for the specific precipitation of amino acids, have also been used for the separation of peptides from a partial hydrolyzate of silk (Stein *et al.*, 1944).

Synge and Tiselius (1950) have recently described a method for fractionating substances according to their molecular weight by electrokinetic ultrafiltration which may be applicable to the group separation of peptides. A number of ingenious methods of fractionation have recently been suggested by Martin (1949b).

10. *Conclusions*

Many of the methods considered above have not been extensively used for the fractionation of peptides, so that it is impossible to know how far they may be applied and which are the most effective methods. Also it is to be expected that considerable improvements in these techniques will take place and that other new methods will be devised in the near future.

The properties of amino acids and small peptides render them suitable to fractionation by methods employing partition chromatography. Paper chromatography is especially to be preferred for the final fractionation of a simplified peptide mixture because of the good resolutions obtained, the ease and rapidity of technique, and the possibility of using

the two-dimensional method. The chief disadvantage is that only a very small amount of each peptide can be obtained from a chromatogram, though it is usually possible to obtain sufficient material to determine the structure of di- and tripeptides. When larger amounts of peptides are required, several methods are available; chromatography on starch, cellulose powder or ion exchange resins, or countercurrent distribution. Probably none of these gives such good resolution as paper chromatography and the work involved is greater.

For the preliminary separation of a complex protein hydrolyzate into simpler peptide mixtures, ionophoretic methods are probably the most generally useful. Aromatic peptides may be separated by adsorption on charcoal and cystine peptides by oxidation.

One of the main requirements at present is for methods that will efficiently fractionate large polypeptides with molecular weights between say 1,000 and 10,000. At present the best method for dealing with them is by countercurrent distribution and possibly by certain ionophoretic methods. While techniques for separating this class of substances would be extremely useful, it seems that the chief difficulty is to find suitable degradation procedures for producing a simple mixture of large polypeptides from a protein.

VI. Determination of Peptide Structure

1. *Identification and Estimation of Amino Acids*

The first stage in the determination of the structure of a peptide is to identify the amino acid residues present in it, and the obvious technique for this is paper chromatography. By no other method is it possible to identify completely the amino acids present in a mixture so simply and so rapidly. Unfortunately no solvent has yet been found on which it is possible to separate all the naturally occurring amino acids and it is necessary to run each hydrolyzate on a two-dimensional chromatogram or else to run aliquots on two separate one-dimensional chromatograms. Where a large number of peptides is to be analyzed, the latter technique is usually preferable (Consden *et al.*, 1949).

It is often desirable to carry out an approximate amino acid analysis to determine whether there are one or two residues of a particular amino acid in a peptide. This may be done with sufficient accuracy by carrying out the paper chromatography in a semi-quantitative manner (Polson, 1948; Consden *et al.*, 1949). Several modifications of greater accuracy have been described (Martin and Mittelmann, 1948; Wieland and Fischer, 1948; Woiwod, 1949; Fowden and Penney, 1950; Boissonnas, 1950). The question as to how accurately it is possible to determine an amino

acid on a paper chromatogram has been frequently discussed (Jones, 1949; Gordon, 1949; Martin, 1950) and is probably still unsettled. If careful control experiments are carried out the above methods should give results to within 5–10%. Other methods (starch chromatography, microbiological methods, or the pipsyl method) may be used to obtain an accurate analysis but are considerably more laborious.

An elegant method for determining the optical configuration of amino acids on paper chromatograms using the enzyme D-amino acid oxidase has been described by Synge (1949).

2. *Amino Acid Sequence*

The determination of the amino acid sequence in peptides is essentially a question of identifying terminal residues. Thus if the *N*-terminal residue of a dipeptide is known its structure is determined. The sequence in a tripeptide would be determined by identification of the *N*- and *C*-terminal residues. The structure of larger peptides may be worked out from their terminal residues and by degrading them to di and tripeptides. Thus, for instance, if a tetrapeptide has residue A as an *N*-terminal residue and on hydrolysis gives the dipeptides A.B and C.D, its structure must be A.B.C.D. An alternate method of determining the amino acid sequence is by a method of step-wise degradation from the terminal residues (p. 8). The method of Edman (1950), which has already been applied to synthetic peptides would seem especially suitable for this purpose. Consden *et al.* (1947a) have developed a rapid micro method for the deamination of the *N*-terminal residues of peptides eluted from paper chromatograms. A sample of the peptide is completely hydrolyzed and the amino acids present are identified. Another sample is then deaminated with nitrosyl chloride which destroys the *N*-terminal residue. The amino acids remaining are then identified after hydrolysis of the deaminated peptide. Although the reaction rarely goes completely smoothly and quantitatively, and a certain amount of destruction of the non-terminal residues usually takes place, it is nevertheless possible in most cases to identify the *N*-terminal residue.

The DNP method (p. 5) may also be used for the identification of the *N*-terminal residues in peptides. It is rather more laborious than the above deamination procedure, but the results are usually more clear-cut. Two methods are available. Either the terminal DNP amino acid is identified by chromatography or the remaining unsubstituted amino acids are identified by paper chromatography. The former method has the advantage that it may be carried out quantitatively but requires rather more material than is usually available from the elution of one paper chromatogram.

Bowman (1950) has shown that amino acids and peptides may be quantitatively converted to their dimethyl derivatives by reductive alkylation in the presence of formaldehyde, and has used the method to detect the *N*-terminal residues of peptides. A micro-modification was described by Ingram (1950) who identified the amino acid which is absent from a hydrolyzate of the dimethyl peptide.

3. *Estimation of Peptides*

It is clear that the estimation of the amount of a particular peptide in a partial hydrolyzate may in certain cases yield considerably more information concerning the structure of the protein than its mere identification. Unfortunately it is impossible to determine directly the total amount of a particular peptide sequence in a protein molecule. Only a minimal value can be obtained from the composition of a partial hydrolyzate. In the case of the *N*-terminal peptides of insulin it was possible to estimate certain sequences in the protein from the yields of the peptides by allowing for the rate of breakdown of the bonds involved (p. 52). Such an approach may clearly be useful in the future where the yield of a peptide can be determined.

Usually only small amounts of peptide are available after chromatographic fractionation of a partial hydrolyzate, so that micro methods of estimation are required. Consden *et al.* (1949) obtained an approximate estimate of peptides from a paper chromatogram by the color intensities of the amino acid spots produced on hydrolysis. Clearly careful control analyses with synthetic substances should accompany any estimaitons by this type of technique to allow for losses during fractionation.

The use of isotopic methods for the estimation of peptides appears to offer considerable advantages of accuracy and general applicability. In the original "isotope dilution" technique of Rittenberg and Foster (1940) it was necessary to isolate a pure sample of the compound to be estimated, from a mixture of comparable amounts of other similar substances. In the case of peptides this would clearly be a formidable task, and has never been attempted.

The Pipsyl Method. Using radioactive isotopes Keston, Udenfriend and Cannan (1946, 1949; Keston and Udenfriend, 1949) have developed an accurate micro method, which theoretically could be applied to the estimation of any amino acid or peptide in a mixture. Two different techniques, the carrier technique and the indicator technique, have been worked out. In the former the peptide mixture to be analyzed is treated with *p*-iodophenylsulfonyl chloride (pipsyl chloride) containing radioactive I^{131}, which reacts quantitatively with the amino groups giving pipsyl derivatives. To this is added a great excess of the non-isotopic pipsyl

derivative of the peptide to be estimated, so diluting the corresponding isotopic derivative in the hydrolyzate. A pure sample of this pipsyl peptide is then isolated from the mixture, and from its isotope content, the amount of peptide in the original mixture may be calculated. The chief difficulty lies in the isolation of the pipsyl derivative in an absolutely pure state free from any other radioactive pipsyl derivatives of closely related compounds which tend to form mixed crystals. In many cases a large number of recrystallizations are necessary before a constant isotope concentration is obtained.

To overcome the above difficulties the indicator technique was developed for the estimation of amino acids (Keston *et al.*, 1947, 1950). The amino acid mixture is again treated with pipsyl chloride containing I^{131}. This mixture can then be fractionated by countercurrent distribution and paper chromatography and if all the derivatives are completely separated, the amount of amino acid can be determined from the isotope content of the spot. By adding an indicator to the mixture before fractionation it is possible to obtain an estimate even if the spots are not completely resolved. For instance, if alanine is to be estimated the indicator added would be pipsyl alanine containing S^{35}, which may be estimated independently of the I^{131}. The ratio of I^{131} to S^{35} in a part of the pipsyl-alanine spot may be used to estimate the original alanine content. This is only true for those parts of the alanine spot which contain no other pipsylderivatives (containing I^{131}). In such parts the ratio of I^{131} to S^{35} is constant, and this is used as the test of purity. This method has not yet been applied to peptides, but there seems no reason why it should not be, though rather complex fractionation procedure might be necessary to obtain a part of the peptide spot in a pure form.

VII. Results of Investigations on Various Proteins

In this section we shall consider the results obtained since 1943, which provide information concerning the arrangement of amino acids in proteins. Similar studies on the naturally occurring polypeptides will not be dealt with here.

1. *Silk Fibroin* (Bombyx Mori)

Assuming a molecular weight of 30,000 for silk fibroin (Coleman and Howitt, 1946) its composition is given by the following formula:

$$Gly_{169}Ala_{103}Ser_{40}Tyr_{19}Asp_6Glu_4Leu_6Val_9Pro_2Phe_3Thr_4Arg_2Lys_1His_1$$

(Levy and Slobodiansky, 1949; Tristram, 1949)

It may be noticed that this formula tends approximately but not exactly to the formula $(Gly_3Ala_2X_2)_n$ where X represents any residues

other than glycine or alanine, and this unusually simple composition suggests that the amino acid arrangement may also be simple.

By the classical methods of fractionation the peptides Gly.Ala (Fischer and Abderhalden, 1907) Ala.Gly (Abderhalden, 1909a) and Gly.Tyr (Abderhalden, 1909b) were isolated in considerable yield from the silk fibroin of *Bombyx Mori* and unambiguously characterized. Several longer peptides were also isolated in small yield and fairly convincing evidence for their structure was presented, though they were not finally identified by synthesis. Thus Abderhalden and Bahn (1932) benzoylated a fraction from a hydrolyzate of fibroin which had been obtained by the action of 1 *N* NaOH at 37°, and obtained several benzoyl peptides. One of these contained the amino acids glycine, serine, tyrosine, and proline in the ratio 1:1:1:2 and on partial hydrolysis with 10% H_2SO_4 gave hippuric acid, Ser.Pro and Tyr.Pro. In an experiment with synthetic peptides it was found that benzoyl Gly.Ser was completely broken down under the conditions of hydrolysis in 10% H_2SO_4 whereas benzoyl Gly.Tyr was rather stable. From this it was concluded that the only possible structure for the pentapeptide was Gly.Ser.Pro.Tyr.Pro. A second fraction was similarly identified as Ser.Pro.Tyr.Pro. A third peptide (Abderhalden and Bahn, 1933) containing the amino acids tyrosine, serine, and proline in the ratio 2:1:1 was isolated from a similar hydrolyzate. On treatment with trypsin (presumably a crude preparation) this gave tyrosine, Ser.Pro and a tripeptide containing one molecule of the three amino acids. This tripeptide was then treated with benzylamine and phenylisocyanate, according to the method of Abderhalden and Brockman (1930) and from the hydrolyzate the phenylhydantoin of tyrosine and Ser.Pro benzylamide were isolated. This established the structure of the tripeptide as Tyr.Ser.Pro, and it was concluded that the tetrapeptide was Tyr.Ser.Pro.Tyr. The presence of tyrosine as a *N*-terminal amino acid was assumed since the tetrapeptide was precipitable with mercuric sulfate, as were tyrosine and other dipeptides containing a free tyrosine carboxyl group, whereas the tripeptide was not precipitable. While the evidence for the structure of these peptides is fairly good, it is possible that they may have been mixtures of peptides having slightly different structures. At least they do show that there are tetrapeptide sequences in fibroin that contain no glycine or alanine.

Stein *et al.* (1944) have made use of their method of specific precipitation with aryl sulfonic acids to isolate peptides from an acid hydrolyzate of silk fibroin. Gly.Ala was isolated in 5.5% yield using 2:5-dibromobenzene sulfonic acid and Ala.Gly in 6.0% yield using 2,6-diiodophenol-4-sulfonic acid.

A more precise estimation of these two peptides present at various

stages of hydrolysis was obtained by Levy and Slobodiansky (1949) with the pipsyl carrier technique. Samples of fibroin that had been hydrolyzed with 12 *N* HCl at 39° for 16, 24, and 48 hours respectively were analysed for glycine, alanine, Ala.Gly, Gly.Ala and Gly.Gly. The results of one such experiment are shown in Table V.

It was calculated that if the arrangement were completely random, the maximum possible yield of Gly.Gly would be 18.2% and of Ala.Gly and Gly.Ala 12.2%. The above figures, especially those for Ala.Gly, which are over twice the theoretical value show that this is not the case. The relatively small yield of Gly.Gly suggests that this sequence does not occur to any great extent whereas the yield of Ala.Gly accounts for about half of the alanine present in the protein. Levy and Slobodiansky have pointed out that these results would be expected if the minimum

TABLE V

Analysis of Partial Hydrolyzate of Silk Fibroin (Levy and Slobodiansky, 1949)
(Yields expressed in terms of N as % of total N)

Time of hydrolysis in 12 *N* HCl at 39°	16 hr.	24 hr.	48 hr.
Glycine	4.4	7.1	12.9
Alanine	3.1	6.3	10.5
Ala.Gly	16.9	23.3	27.0
Gly.Ala	5.4	9.0	8.3
Gly.Gly	0.1	—	1.8

repeating peptide sequence were of the type: X.Ala.Gly.Ala.Gly.X.Gly. While this is probably the simplest explanation of the results, there are certainly many other more complicated ones. It does not account for the tri- and tetra-peptides isolated by Abderhalden and Bahn which contain no glycine or alanine or for the results of Drucker and Smith (see below). The results of Goldschmidt *et al.* (1933), and of Grant and Lewis (1935) referred to by Synge (1943) also suggest an uneven distribution of residues throughout the molecule. At least, it may be said that the results of Levy and Slobodiansky make it likely that the sequence X.Ala.Gly.Ala.Gly.X occurs frequently in silk fibroin. Silk fibroin may be dissolved in a solution of cupri-ethylenediamine (Coleman and Howitt, 1947). On dialysis part of the material remains in solution and is said to be "renatured." Drucker and Smith (1950) treated this material for a short period with trypsin and obtained a precipitate, which had an average molecular weight of about 7000 and contained only the amino acids glycine, alanine, and serine. The remaining amino acids were all left in solution. It was suggested on the basis of this evidence

that the silk fibroin molecule is built up of five parts, three of which have molecular weight 7000 and contain only the above three amino acids, the other two parts having molecular weight 5000 and a more complex composition.

These results are in agreement with the suggestion of Meyer *et al.* (1940), that silk fibroin is composed of two parts, an amorphous part of complex amino acid composition and a crystalline part built up simply of Gly.Ala units linked together. This theory was based on the observation that the X-ray data could be fitted by a unit cell containing four parallel Gly.Ala residues. It was considered that some of the alanine residues in the crystalline part could be replaced by serine since its molecular dimensions are similar. Clearly the idea of a long chain containing glycine and alanine alternately is untenable in view of the results of Levy and Slobodiansky since in such a case the yields of Ala.Gly and Gly.Ala should be equal. However such a sequence could be broken up to a certain extent by serine residues.

Abderhalden (1940, 1943) has recently renewed his investigations on the isolation of diketopiperazines from silk fibroin and has isolated considerable quantities of the anhydrides of Gly.Ala, Gly.Tyr, and Ala.Ser from partial hydrolyzates obtained by the use of strong acid or proteolytic enzymes. No evidence could be obtained of anhydrides containing only one amino acid, making it unlikely that such sequences occur to any great extent in fibroin. Abderhalden believes that these diketopiperazines are derived from some cyclic structure in the protein. It is possible however that they may have been formed from peptides during the isolation procedure since certain quite mild conditions are known to catalyze their formation (Ågren, 1940; Huang and Niemann, 1950). Hot ethanol was used to extract them, and this might have brought about ring formation. Control experiments were carried out to show that dipeptides did not undergo ring closure under the conditions of hydrolysis, but the other obvious control, boiling a dipeptide with ethanol, was not reported.

2. *Protamines*

Little progress has been made concerning the chemistry of the protamines since Synge reviewed the subject in 1943. However it would seem that much could be learnt from the application of the newer methods of peptide chemistry to these proteins, which from their amino acid composition and molecular weight appear to have a rather simple structure (see Tristram, 1949).

Porter and Sanger (1948) showed the presence of proline as a *N*-terminal residue in salmine, but were unable to estimate it due to the instability

of DNP-proline. Felix and Mager (1937) had already suggested that proline was the *N*-terminal residue in clupein on the basis of titration data. This was confirmed using the DNP method, and it was suggested there was also a small amount of *N*-terminal serine (Felix *et al.*, 1950).

The presence of arginine as a *C*-terminal residue was indicated by Dirr and Felix (1932). If dibenzoyl arginine is treated with acetic anhydride, an acetyl derivative is formed which on hydrolysis gives benzoyl acetyl urea and β-benzoylamino-α-piperidone. This is readily split with acid to give ornithine. It was found that ornithine was produced after benzoylation, acetylation, and hydrolysis of clupein. The above series of reactions could only take place if the α-carboxyl group was not involved in peptide linkage so that it must have originated from a *C*-terminal arginine residue.

Since two-thirds of the residues in clupein are arginine it follows that the sequence Arg.Arg must occur frequently in the molecule, and the dipeptide has often been isolated. Felix and Schuberth (1942) have described a method for the preparation of Arg.Arg and obtained 12.5 g. of the pure diflavianate from 50 g. clupein.

The fractionation of partial hydrolyzates of clupein by adsorption chromatography on filtrol-neutrol (Waldschmidt-Leitz and Turba, 1940, 1941) and on the cation exchange resin Wofatite C (Rauen and Felix, 1948) have been studied. Considerable fractionations could be obtained, but no definite peptides were identified.

3. *Wool Keratin*

Considerable doubt exists as to whether wool can be regarded as a single protein or whether it is a mixture of different insoluble proteins, so that it is perhaps not one of the most suitable proteins to study by partial hydrolysis methods. However, because of its practical importance the Leeds workers have applied their newer methods to it, and a number of peptides have been identified.

Consden *et al.* (1949) studied an acidic fraction of a partial hydrolyzate of wool. Preliminary group separation was carried out on an ion exchange column (Amberlite IR-4) to obtain an acidic fraction, which was separated into ten fractions by ionophoresis in silica jelly. Each fraction was then subjected to two dimensional paper chromatography using phenol and collidine. One of these ionophoretic fractions contained only one peptide which was identified as Glu.Glu. This was present in much greater amount than any other acidic peptide and accounted for about 10% of the total glutamic acid of wool (estimated from the total N of the ionophoretic fraction). In Table VI are listed the peptides that were considered to be probably present. About an

equal number of peptides were given as being possibly present. Interpretation of the results was rendered difficult by the extreme complexity of the mixture, so that only a few of the spots appeared to contain a single peptide. Where no difference in composition was found before and after deamination of a dipeptide spot, it was assumed that both possible peptides were present. It is doubtful if this is entirely justified, since some dipeptides are extremely stable to deamination. Clear evidence was obtained for the presence of Glu.Glu, Ala.Glu, Glu.Ala

TABLE VI

Acidic Peptides Identified from Wool (Consden et al., 1949)

Peptide	Yield*
Glu.Asp	4
Glu.Glu	600
Glu.Gly	5
Glu.Ala	29
Glu.Tyr	29
Glu.Leu	21
Glu.CyS	4
Asp.Glu	4
Ser.Glu	17
Gly.Glu	8
Ala.Glu	27
Tyr.Glu	27
Val.Glu	19
Leu.Glu	51
CyS.Glu	4
Asp.Val	11
Asp.Leu	11
Ser.Asp	5
Leu.Asp	19

* mg. N of peptide per 100 g. N of wool.

and Asp.Leu, and the peptides listed in Table VI are probably present also. The approximate yields of the peptides were estimated from the strength of ninhydrin color of the amino acids produced on hydrolysis. These yields are listed in Table VI.

Consden and Gordon (1950) have studied the cysteic acid peptides derived from the cystine peptides of wool as described on p. 40. After a preliminary group separation of the cysteic acid peptides by ionophoresis, each fraction was fractionated on paper chromatograms with phenol and collidine. In Table VII are listed the peptides considered to be probably present and their approximate yield. The results were more clear-cut in this case than with the aspartic and glutamic acid

peptides, largely due to the fact that it was possible to separate by ionophoresis or chromatographically any two dipeptides containing the same amino acids in different order. This separation was made possible by the low pK of the amino group in cysteic acid peptides.

It is clear from the results that the structure of wool is extremely complex. Almost all of the monoamino acids occur linked to both sides of glutamic and cysteic acids. The small number of aspartic acid peptides identified is probably due to the lower content of aspartic acid in wool so that the peptide spots would probably be faint and would not show on the chromatograms.

TABLE VII

Cysteic Acid Peptides from Wool (Consden and Gordon, 1950)

Peptide	Yield*
Asp.$CySO_3H$	2
Glu.$CySO_3H$	2
Ser.$CySO_3H$	60
Gly.$CySO_3H$	24
Thr.$CySO_3H$	20
Ala.$CySO_3H$	30
Leu.$CySO_3H$	8
$CySO_3H$.Gly	10
$CySO_3H$.Thr	20
$CySO_3H$.Ala	14
$CySO_3H$.Val	20
$CySO_3H$.Leu	10
$CySO_3H$.Phe	4

* mg. N of peptide per 100 g. N of wool.

Middlebrook (1949) investigated the *N*-terminal residues by the DNP method and found rather a complex mixture, which would seem to be a further indication of the heterogeneity of wool.

4. *Insulin*

Insulin has been studied in rather more detail than most other proteins due to the interest arising from its physiological properties and to the fact that it is one of the few proteins that can be obtained in a reasonably pure form. It possesses a relatively simple structure, being built up of fairly short open polypeptide chains. For the purposes of the following discussion a value of 12,000 (Gutfreund, 1948) will be assumed for the molecular weight.

The most recent analytical figures (see Tristram, 1949) indicate the following composition:

$Gly_7Ala_6Val_3Leu_{12}Ileu_3Pro_3Phe_6(CyS\text{-})_{12}Arg_2His_4Lys_2Asp_2Glu_6$
$(Asp\text{-}NH_2)_4(Glu\text{-}NH_2)_8Ser_6Thr_2Tyr_8$

The DNP method showed the presence of four *N*-terminal residues two of which were glycine and two phenylalanine (Sanger, 1945). From a study of the action of carboxypeptidase on insulin Lens (1949) found that free alanine was liberated before any other free amino acids, indicating the presence of alanine in the *C*-terminal position. Using their respective methods of reduction of the free carboxyl groups to alcohol

TABLE VIII

Properties of the Fractions of Oxidized Insulin (Sanger, 1949a)

	A	B
Yield from insulin	30–40%	25%
N-terminal residue	Glycine	Phenylalanine
Amino acids absent	Lysine, arginine, histidine, phenylalanine, threonine, proline	Isoleucine
Mol. wt.		
1. From estimation of *N*-terminal residues	2900	3800
2. By ultracentrifugation (Gutfreund and Ogston, 1949)	2900	7000

groups (p. 10) Fromageot *et al.* (1950) and Chibnall and Rees (1951) both identified alanine and glycine as the *C*-terminal amino acids.

These results indicate that insulin is built up of four open polypeptide chains, which seem to be held together by —S—S— bridges of cystine. These are present to the extent of six residues per molecule. By oxidation with performic acid (p. 27) it was possible to split these bridges and thus to liberate the separate peptide chains (Sanger, 1949a). The oxidized insulin was fractionated by precipitation methods and two main fractions were obtained, which appeared to represent the whole of the oxidized insulin. The properties of these two fractions, designated A (acidic) and B (basic), are summarized in Table VIII. The most probable composition of fraction A is,

$$Gly_2Ala_1Val_2Leu_2Ileu_1(CyS)_4Asp_2Glu_4Ser_2Tyr_2(—NH_2)_4$$

and of fraction B,

$Gly_3Ala_2Val_3Leu_4Pro_1Phe_3(CyS)_2Arg_1His_2Lys_1Asp_1Glu_3Ser_1Thr_1Tyr_2$
$(—NH_2)_2$

Butler and coworkers have studied the action of various proteolytic enzymes on insulin. From a chymotryptic hydrolyzate they obtained

two fractions, one precipitable by trichloracetic acid and one soluble (Butler *et al.*, 1948). The latter fraction consisted of relatively small peptides with an average chain length of seven residues. These could be separated satisfactorily on paper chromatograms using butanol-acetic acid (Phillips, 1949a) but they have not yet been investigated further. The trichloracetic acid precipitate, which was referred to as the "core" had a mean molecular weight of about 5000 and a high proportion of glycyl (approximately 3 equivalents) and a smaller amount of valyl

TABLE IX

DNP Derivatives from Fraction B of Oxidized Insulin (Sanger, 1949b)

Derivative	Products of complete hydrolysis	Products of partial hydrolysis	Yield* from DNP-insulin	Yield* from B4
B1	DNP-phenylalanine	—	13	13
B2	DNP-phenylalanine Valine	—	16	14
B3	DNP-phenylalanine Valine Aspartic acid	B1, B2	13	12
B4	DNP-phenylalanine Valine Aspartic acid Glutamic acid	B1, B2, B3	30	55
B5	—	B1, B2, B3, B4	20	
Total			92	94

* Moles of peptide as % of the total *N*-terminal phenylalanyl residues.

and other *N*-terminal residues (Butler *et al.*, 1950). On oxidation with performic acid the core was mostly converted to a fraction corresponding in properties to the fraction A of the oxidized insulin.

a. N-Terminal Peptides. The *N*-terminal peptides of the two fractions from the oxidized insulin were determined by subjecting their DNP derivatives to partial hydrolysis (Sanger, 1949b). In the case of fraction B four main bands (B1–B4) were identified in an ethyl acetate extract of a partial hydrolyzate. Their properties are summarized in Table IX. It is clear that all these derivatives are derived from the one *N*-terminal sequence Phe.Val.Asp.Glu. Other fainter bands (B5) were also present; on partial hydrolysis these gave rise to B3 and B4 and were therefore higher peptides from the same peptide sequence. The aspartic and glutamic acid residues are probably in the form of asparagine and glutamine residues in the intact protein, since other bands containing the

same amino acids as B3, B4 and probably therefore amides, were also present when a shorter time of hydrolysis was used. The yields of the different peptides from a partial hydrolyzate of DNP-insulin are shown in column 4 of Table IX. It can be seen that virtually all the *N*-terminal phenylalanyl residues may be accounted for in terms of this one sequence. This is confirmed by comparing the yields of the DNP peptides from DNP-insulin with the yields from a sample of peptide B4 which had been hydrolyzed under similar conditions (column 5, Table IX). Agree-

TABLE X

Peptides Containing ε-DNP-lysine from DNP-insulin (Sanger, 1949b)

Peptide	Amino acids present	*N*-terminal residue	Products of partial hydrolysis	Structure	Yield* from DNP-insulin	Yield* from L4
L1	ε-DNP-Lysine	—	—	ε-DNP-Lys	14	14
L2	ε-DNP-Lysine Alanine	ε-DNP-Lysine	—	ε-DNP-Lys.Ala	19	23
L3	Threonine Proline ε-DNP-Lysine	Threonine	L1	Thr.Pro.(ε-DNP)Lys	32	32
L4	Threonine Proline ε-DNP-Lysine Alanine	Threonine	L1, L2, L3	Thr.Pro.(ε-DNP) Lys. Ala	23	21
Other unidentified bands					6	
				Total	94	

* Moles of peptide as % of the total lysine.

ment between the two sets of figures makes it clear that both the *N*-terminal phenylalanyl residues of insulin are present in the form of this one sequence, Phe.Val.Asp.Glu, and it was therefore concluded that the two phenylalanyl chains of insulin, which contain the same *N*-terminal tetrapeptides are in fact identical.

It was also possible to separate and identify the peptides containing ε-DNP-lysine from a partial hydrolyzate of the DNP derivative of fraction B. Here again four main colored bands were present and their properties are summarized in Table X. These all fit into the sequence Thr.Pro.Lys.Ala and the yields from DNP-insulin and from peptide L4 make it clear that both the lysyl residues of insulin are present in this single tetrapeptide sequence.

When the DNP-derivative of fraction A of the oxidized insulin was partially hydrolyzed, four DNP derivatives were produced, all of which fitted into the sequence DNP-Gly.Ileu.Val.Glu. When strong HCl was

used for hydrolysis the peptides were produced in very small yields. Because of the great lability of the bond involving the carboxyl group of the glycyl residue most of the DNP-glycyl residues were present as DNP-glycine itself and as larger peptides which could not readily be fractionated. However when dilute acid was used for hydrolysis the yield of these peptides was raised and another band which appeared to be Gly.Ileu.Val.Glu.Glu was obtained. The yields of these peptides indicated that both the *N*-terminal glycyl residues are combined in the one amino acid sequence and it was concluded that the two glycyl chains were also identical.

b. Amino Acid Sequence in the Phenylalanyl Chains. The results with the *N*-terminal peptides showed for the first time that the fractions A and B of the oxidized insulin were each an essentially homogeneous preparation of a polypeptide chain of 20 and 30 residues respectively. It was therefore considered worth while to investigate the peptides present in their partial hydrolysates by the methods of Consden *et al.* (1947b). Fraction B was subjected to hydrolysis in conc. HCl and the hydrolyzate separated into a number of fractions by ionophoresis, charcoal adsorption, and adsorption on an ion exchange resin (Sanger and Tuppy, 1951a). The resulting peptide mixtures were then fractionated on two-dimensional paper chromatograms and their structures investigated.

To illustrate the methods used a fraction containing peptides of cysteic acid will be considered. This was obtained by adsorption on an ion exchange resin (Amberlite IR-4B at pH 2.6; Consden *et al.*, 1948). A chromatogram of this fraction on phenol/butanol-acetic acid revealed eight peptide spots, whose compositions are given in Table XI. The *N*-terminal residues were determined by hydrolysis of the DNP derivatives. The structures of the peptides, as far as they can be deduced from the data in the table are given in the last column. Three dipeptides were identified: $CySO_3$.H.Gly, Val.$CySO_3$H, and Leu.$CySO_3$H. Fraction B contains only two cysteic acid residues so that any peptide containing cysteic acid which is not the *N*-terminal residue must either contain the sequence Val.$CySO_3$H or Leu.$CySO_3$H. Peptide 3 must therefore be Val.$CySO_3$H.Gly and peptide 5 Leu.$CySO_3$H.Gly. Thus both cysteic acid residues are joined through their carboxyl groups to a glycine residue. Peptide 8 which contains no glycine but has leucine as the *N*-terminal residue can only be Leu.Val.$CySO_3$H and peptide 7 Leu.Val.$CySO_3$H.Gly. These peptides establish therefore the presence of the sequences Leu.Val.-$CySO_3$H and Leu.Val.$CySO_3$H.Gly in fraction B.

In the upper part of Table XII are listed the various peptides identified in this work and the sequences which were deduced from them as

being present in fraction B. Hydrolyzates obtained by the use of dilute acid and of alkali were also studied but only a few new peptides were found. These are also included in Table XII. From the results it was possible to deduce five sequences as being present in fraction B, accounting for all the amino acids present except for one leucine, one tyrosine and

TABLE XI

Cysteic Acid Peptides from Fraction B of Oxidized Insulin (Sanger and Tuppy, 1951a)

Peptide	Amino acids present	Strength of amino acid after hydrolysis of: Peptide	Strength of amino acid after hydrolysis of: DNP-Peptide	Structure
1	Cysteic acid	xxxx	—	$CySO_3H$.Gly
	Glycine	xxxx	xxx	
2	Aspartic acid	x		(Asp, Glu)
	Glutamic acid	x		
3	Cysteic acid	xx	x	Val.($CySO_3H$, Gly)
	Glycine	x	x	
	Valine	xx	—	
4	Cysteic acid	xxx	xx	Val.$CySO_3H$
	Valine	xxx	—	
5	Cysteic acid	x	x	Leu.($CySO_3H$, Gly)
	Glycine	x	?	
	Leucine	x	—	
6	Cysteic acid	xxx	xxx	Leu.$CySO_3H$
	Leucine	xxx	—	
7	Cysteic acid	xx	xx	Leu.($CySO_3H$, Gly, Val)
	Glycine	xx	xx	
	Valine	xx	xx	
	Leucine	xx	—	
8	Cysteic acid	xxxx	xxx	Leu.($CySO_3H$, Val)
	Valine	xxxx	xxx	
	Leucine	xxxx	—	

two phenylalanine residues. It was not, however, possible to fit these five sequences together into a single unique structure. This was partly due to the lability of certain bonds to hydrolysis, especially those involving the amino groups of the serine and threonine residues. Thus no peptide was identified which contained these bonds intact. Another difficulty was experienced in separating the less polar peptides containing phenylalanine, leucine, valine, and tyrosine, since these residues are grouped together in the polypeptide chain and give rise to a large number of similar peptides.

The action of the proteolytic enzymes pepsin, trypsin, and chymotrypsin on this fraction was next investigated (Sanger and Tuppy,

Table XII

Peptides Identified in Hydrolyzates of Fraction B of Oxidized Insulin (Sanger and Tuppy, 1951a, b)

Dipeptides from acid and alkaline hydrolyzates	Phe.Val Glu.His CySO$_3$H.Gly His.Leu Glu.Ala CySO$_3$H.Gly Arg.Gly Lys.Ala Val.Asp His.Leu Leu.Val. Ala.Leu Leu.Val. Gly.Glu Gly.Phe Thr.Pro Asp.Glu Leu.CySO$_3$H Ser.His Val.Glu Val.CySO$_3$H Glu.Arg
Tripeptides from acid and alkaline hydrolyzates	Phe.Val.Asp Leu.CySO$_3$H.Gly Val.Glu.Ala Tyr.Leu.Val Gly.Glu.Arg Pro.Lys.Ala Glu.His.Leu Ser.His.Leu Val.CySO$_3$H.Gly Val.Asp.Glu Leu.Val.Glu Leu.Val.CySO$_3$H His.Leu.CySO$_3$H Ala.Leu.Tyr
Higher peptides from acid and alkaline hydrolyzates	Phe.Val.Asp.Glu Ser.His.Leu.Val Tyr.Leu.Val.CySO$_3$H Thr.Pro.Lys.Ala His.Leu.CySO$_3$H.Gly Leu.Val.Glu.Ala Leu.Val.CySO$_3$H.Gly Phe.Val.Asp.Glu.His Ser.His.Leu.Val.Glu Glu.His.Leu.CySO$_3$H His.Leu.Val.Glu Ser.His.Leu.Val.Glu.Ala
Sequences deduced from above peptides	Phe.Val.Asp.Glu.His.Leu.CySO$_3$H.Gly Tyr.Leu.Val.CySO$_3$H.Gly Thr.Pro.Lys.Ala Ser.His.Leu.Val.Glu.Ala Gly.Glu.Arg.Gly
Peptides identified in peptic hydrolyzate	Phe.Val.Asp.Glu.His.Leu.CySO$_3$H.Gly.Ser.His.Leu Leu.Val.CySO$_3$H.Gly.Glu.Arg.Gly.Phe Tyr.Thr.Pro.Lys.Ala Val.Glu.Ala.Leu His.Leu.CySO$_3$H.Gly.Ser.His.Leu
Peptides identified in chymotryptic hydrolyzate	Phe.Val.Asp.Glu.His.Leu.CySO$_3$H.Gly.Ser.His.Leu.Val.Glu.Ala.Leu.Tyr Tyr.Thr.Pro.Lys.Ala Leu.Val.CySO$_3$H.Gly.Glu.Arg.Gly.Phe.Phe
Peptides identified in tryptic hydrolyzate	Gly.Phe.Phe.Tyr.Thr.Pro.Lys.Ala
Structure of the phenylalanyl chain of insulin	Phe.Val.Asp.Glu.His.Leu.(CyS—).Gly.Ser.His.Leu.Val.Glu.Ala.Leu.Tyr.Leu.Val.(CyS—).Gly.Glu.Arg.Gly.Phe.Phe.Tyr.Thr.Pro.Lys.Ala

1951b). The larger peptides present, which contained up to about fifteen residues, could be satisfactorily separated on paper chromatograms and with the knowledge obtained from the lower peptides it was possible to deduce the structure of many of them from their amino acid composition and *N*-terminal residues and to work out a unique sequence for fraction B. These results are also shown in Table XII, where the structure of those peptides which played a major part in the elucidation of the sequence are given. The structure of fraction B was worked out as the only possible sequence which would fit all the experimental results, assuming that it was a single polypeptide chain of about thirty residues. The fact that it was possible to find a unique sequence was regarded as proof that such an assumption was correct, and that there is only one type of phenylalanyl chain in insulin.

5. *Ovalbumin*

Using the DNP method it has been shown that ovalbumin has no *N*-terminal residue (Desnuelle and Casal, 1948; Porter, 1950a). Either the amino groups at the end of the chains are masked by the carbohydrate moiety or the protein contains one or more cyclopeptide units.

A very specific type of hydrolysis occurs when ovalbumin is incubated with a proteolytic enzyme prepared from *B. subtilis*. A new crystalline protein, plakalbumin, is produced (Linderstrøm-Lang and Ottesen, 1949), which differs from ovalbumin in crystalline form and in solubility and has a somewhat lower molecular weight (Güntelberg and Linderstrøm-Lang, 1949). At pH values below 7.0 it has a slightly different electrophoretic mobility which is concordant with the loss of two acidic groups per molecule (Perlmann, 1949). The conversion is accompanied by the liberation of six atoms of non-protein nitrogen per molecule of ovalbumin (Eeg-Larsen *et al.*, 1948). This fraction contains two free amino groups, 3.6 free carboxyl groups and 4 atoms of peptide bond N per six nitrogen atoms; it contains no free amino acids but on hydrolysis yields alanine, glycine, valine and aspartic acid (Villee *et al.*, 1950). The results suggest that about two small peptides are present. Since the ratio of the rate of formation of plakalbumin to the rate of formation of nonprotein N was not constant, it was evident that the reaction takes place in at least two stages, probably the rupture of two peptide bonds. On further incubation with the enzyme, the plakalbumin is broken down slowly to a mixture of products. Clearly this is a case where it is possible to obtain characteristic degradation products at the early stages of an hydrolysis. It would seem that about two peptide bonds are split about ten times as rapidly as any others.

6. *γ-Globulin*

Porter (1950a) found rabbit γ-globulin to have one *N*-terminal alanine residue per molecule of molecular weight 160,000, which suggests that it is probably a single long chain of over a thousand residues. He also determined the *N*-terminal peptide sequence. All the DNP-peptides found fitted into the one sequence DNP-Ala.Leu.Val.Asp and the fifth residue was probably glutamic acid. There were no other DNP peptides present and the whole of the terminal alanine residues could be quantitatively accounted for in peptides fitting the above sequence. In contrast to the results of physicochemical studies, these findings suggest that the protein is chemically homogeneous, since it would seem unlikely that two different proteins would have the same terminal tetrapeptide sequence. It is difficult to be absolutely certain about this question, since nothing is known about the principles that govern the order of amino acid residues in proteins, but it may be pointed out that on a purely random basis there are 19^4 possible terminal tetrapeptide sequences. The fact that various different residues have been found to occupy the *N*-terminal positions in proteins indicates that there is no general principle that defines closely the amino acid that occupies a particular position in a protein, so that the presence of a single *N*-terminal peptide sequence in a preparation can best be explained on the basis of chemical homogeneity.

Porter also compared the *N*-terminal peptides of normal γ-globulin and of purified antivoalbumin, which was studied in the form of a specific precipitate with ovalbumin. No difference could be found, again suggesting the chemical similarity between the antibody and the normal γ-globulin, from which it is formed. An alternative explanation is that only a small part of the polypeptide chain, "the active center," has a different amino acid sequence in the antibody. In an attempt to identify the "active center" of the antiovalbumin molecule, Porter (1950b) studied the action of proteolytic enzymes on it. By the action of papain a molecule about a quarter the size of the original γ-globulin was split off, which acted as a specific inhibitor in the antibody reaction. Since it had a *N*-terminal alanine group it appeared to come from the *N*-terminal quarter of the γ-globulin molecule. On further hydrolysis all activity was lost and it was not possible to obtain an active molecule small enough for chemical investigation.

These results illustrate the possibilities of applying chemical methods to biologically active proteins, most of which have been studied so far only by physicochemical techniques.

7. *Hemoglobin*

Porter and Sanger (1948) determined the *N*-terminal residues of hemoglobins from a number of different animal species. These results are summarized in Table I. Considerable species differences were evident both in the nature of the *N*-terminal residues and in the number of open polypeptide chains present. In the case of horse hemoglobin which was studied in more detail, there are six valyl *N*-terminal residues and therefore six open polypeptide chains. It is interesting to note that the cystine content is not more than three residues, so that some of the chains must be held together by another type of cross-linkage. An investigation of the *N*-terminal peptides by the DNP-technique showed the presence of the following *N*-terminal sequences:

Val.Leu, Val.Glu(—NH_2).Leu

Thus there are at least two different types of polypeptide chains present in the molecule (Sanger, 1948).

8. *Gelatin*

Gordon *et al.* (1943) attempted to fractionate the acetyl derivatives of peptides from gelatin on silica gel chromatograms. The hydrolyzate was separated into neutral and basic fractions by ionophoresis in a three compartment cell. The neutral peptides were then fractionated on a silica chromatogram using ethyl acetate, and each fraction was refractionated using butanol-chloroform. In general the fractionation was not sufficient to give clear-cut results though the presence of the following peptides was suggested: Leu.Gly, Gly.Leu, Gly.Pro, a dipeptide containing proline and alanine, and at tripeptide containing proline, alanine and glycine. These investigations are largely of historical interest, since they were carried out before the introduction of paper chromatography. The amino acids present in the peptides were identified and estimated as their acetyl derivatives on silica chromatograms.

VIII. General Conclusions

It is clear from the above that considerable progress has been made in recent years in the development of methods for investigating the arrangement of amino acid residues in proteins. These methods have not yet been extensively applied and a vast amount of work is still required in this field. Such work may be expected to be rather unrewarding at first. The separation of a few peptides from a protein is not likely to

make possible the formulation of any general theory of protein structure or to explain the physiological action of a protein. Only by the accumulation of a large amount of experimental evidence can such objectives be attained. It does not appear, however, at present that there is any easy short cut to the solution of the problem of protein structure and action. Probably only when more is known about the exact chemical structure of proteins will it be possible to understand the unique part played by proteins in the living organism. Every peptide identified will contribute towards this ideal, even though it may appear to have no significance in itself.

By the application of the methods described in this review the structure of a pure polypeptide containing thirty residues has been determined (p. 54) and there seems no reason why it should not be possible to work out the complete amino acid sequence in proteins which are as simple as insulin. How far it will be possible to apply these techniques to more complex proteins is difficult to say. The larger the polypeptide chains in a protein, the greater the necessity of isolating larger peptide breakdown products. Probably the chief need in this field is for techniques for the specific breakdown of proteins into larger peptides and for the fractionation of such peptides. Most of the more commonly studied proteins contain more than 300 residues but it is possible that some of them, when studied in greater detail may be found to have a simpler structure than is at present believed. The relative simplicity of insulin may be merely apparent as insulin has been studied in more detail than have other proteins.

It has frequently been suggested that proteins may not be pure chemical entities but may consist of mixtures of closely related substances with no absolute unique structure. The chemical results so far obtained suggest that this is not the case and that a protein is really a single chemical substance, each molecule of one protein being identical with every other molecule of the same pure protein. Thus it was possible to assign a unique structure to the phenylalanyl chains of insulin. Each position in the chain was occupied by only one amino acid and there was no evidence that any of them could be occupied by a different residue. Whether this is true for other proteins is not certain but it seems probable that it is. The *N*-terminal residues of several pure proteins have been determined (Table I) and this position is always found to be occupied by a single unique amino acid. These results would imply an absolute specificity for the mechanisms responsible for protein synthesis and this should be taken into account when considering such mechanisms.

It is impossible with the small amount of experimental evidence at present available to form any general theory of protein structure or to

formulate any principles that govern the arrangement of amino acids in proteins, though several such theories have been put forward on very much less evidence. Certainly there is no simple periodic arrangement of residues along the chains of the type suggested by Bergmann and Niemann (1936). The presence of peptides such as Glu.Glu in a hydrolyzate of wool (Consden *et al.*, 1949) precluded this possibility and no periodicity whatsoever was evident in the structure of the phenylalanyl chains of insulin (Table XI).

Although they are not immediately apparent, it still seems probable that there may be certain principles which determine amino acid sequences. The mechanisms of protein synthesis, about which almost nothing is known, would be expected to have their limitations so that one might at least expect to find certain sequences that occur more frequently than others. The results at present available do suggest that this may be the case. Thus, for instance, in the phenylalanyl chain of insulin there are three dipeptide sequences (His.Leu, Leu.Val, CyS.Gly) that occur twice in the chain of thirty residues. One of these (CyS.Gly) also occurs in glutathione. Other dipeptide sequences in this chain were Thr.Pro, which was also detected in the antibiotic actinomycin (Dalgliesh *et al.*, 1950) and Ala.Leu, which was found in gramicidin (Synge, 1949). The sequence Glu.Glu occurs both in wool (Consden *et al.*, 1949) and in insulin (Sanger, 1949b) and is probably also present in gliadin, since Nakashima (1927) obtained a peptide fraction that contained tyrosine and glutamic acid in the ratio 1:3. It may be that this sequence has a special significance since glutamic acid and glutamine residues frequently occur linked together in natural products, such as folic acid, the capsular substance from *B. Anthracis* (Hanby and Rydon, 1946) and in the derivative of triglutamine isolated by Dekker *et al.* (1949) from a marine alga. In this connection Woolley (1949) has suggested that the same amino acid sequence occurs both in insulin and in trypsinogen, since two similar fractions were obtained from tryptic digests of the two proteins. How far these results do reflect a general principle of protein structure will only be known when considerably more experimental evidence is available. The results obtained with wool (p. 48) on the contrary suggest that almost every possible dipeptide containing glutamic acid or cystine is produced on hydrolysis of this protein and that there are therefore no obvious limitations to the type of sequences that can occur. It is certain that proteins are extremely complex molecules but they are no longer completely beyond the reach of the chemist, so that we may expect to see in the near future considerable advances in our knowledge of the chemistry of these substances which are the essence of living matter.

ACKNOWLEDGMENTS

I wish to express my thanks to Dr. A. C. Chibnall, Dr. R. L. M. Synge, and Dr. H. Tuppy for valuable discussions and criticisms which have greatly helped in the preparation of the present review.

REFERENCES

Abderhalden, E. (1909a). *Z. physiol. Chem.* **62,** 315.
Abderhalden, E. (1909b). *Z. physiol. Chem.* **63,** 401.
Abderhalden, E. (1940). *Z. physiol. Chem.* **265,** 23.
Abderhalden, E. (1943). *Z. physiol. Chem.* **277,** 248.
Abderhalden, E., and Bahn, A. (1932). *Z. physiol. Chem.* **210,** 246.
Abderhalden, E., and Bahn, A. (1933). *Z. physiol. Chem.* **219,** 72.
Abderhalden, E., and Bahn, A. (1935). *Z. physiol. Chem.* **234,** 181.
Abderhalden, E., and Brockmann, H. (1930). *Biochem. Z.* **225,** 386.
Abderhalden, E., and Komm, E. (1924a). *Z. physiol. Chem.* **134,** 121.
Abderhalden, E., and Komm, E. (1924b). *Z. physiol. Chem.* **139,** 147.
Abderhalden, E., and Mahn, H. (1927). *Z. physiol. Chem.* **169,** 196.
Abderhalden, E., and Mahn, H. (1928). *Z. physiol. Chem.* **174,** 47.
Acher, R., Jutisz, M., and Fromageot, C. (1950). *Biochim. et Biophys. Acta* **5,** 493.
Agren, G. (1940). *Arkiv. Kemi Mineral. Geol.* **14B,** No. 21.
Bailey, K. (1951). *Biochem. J.* **49,** 23.
Barry, G. T., Gregory, J. D., and Craig, L. C. (1948). *J. Biol. Chem.* **175,** 485.
Behrens, O. K., and Bergmann, M. (1939). *J. Biol. Chem.* **129,** 587.
Bell, P. H., Bone, J. F., English, J. P., Fellows, C. E., Howard, K. S., Rogers, M. M., Shephard, R. G., and Winterbottom, R. (1949). *Ann. N. Y. Acad. Sci.* **51,** 897.
Bergmann, M. (1942). *Advances in Enzymol.* **2,** 49.
Bergmann, M., and Behrens, O. K. (1938). *J. Biol. Chem.* **124,** 7.
Bergmann, M., and Fraenkel-Conrat, H. (1937). *J. Biol. Chem.* **119,** 707.
Bergmann, M., and Fraenkel-Conrat, H. (1938). *J. Biol. Chem.* **124,** 1.
Bergmann, M., and Fruton, J. S. (1938). *J. Biol. Chem.* **124,** 321.
Bergmann, M., and Fruton, J. S. (1941). *Advances in Enzymol.* **1,** 63.
Bergmann, M., and Miekeley, A. (1924). *Z. physiol. Chem.* **140,** 128.
Bergmann, M., and Niemann, C. (1936). *J. Biol. Chem.* **115,** 77.
Biserte, G. (1950). *Biochim. et Biophys. Acta* **4,** 416.
Blackburn, S. (1949). *Biochem. J.* **45,** 579.
Blackburn, S., and Lowther, A. G. (1950). *Biochem. J.* **46,** xxvii.
Block, R. J. (1942). *Proc. Soc. Exptl. Biol. Med.* **51,** 252.
Block, R. J. (1949). *In* Ion Exchange, edited by Nachod, F. C. Academic Press, New York, p. 295.
Boissonnas, R. A. (1950). *Helv. Chim. Acta* **33,** 1975.
Boldingh, J. (1948). *Experientia* **4,** 270.
Bowman, R. E. (**1950**). *J. Chem. Soc.* 1349.
Brand, E., and Edsall, J. T. (1947). *Ann. Rev. Biochem.* **16,** 224.
Brand, E., and Kassell, B. (1939). *J. Biol. Chem.* **131,** 489.
Brigl, P. (1923). *Ber.* **56,** 1887.
Bull, H. B., and Hahn, J. W. (1948). *J. Am. Chem. Soc.* **70,** 2128.
Butler, J. A. V., Dodds, E. C., Phillips, D. M. P., and Stephen, J. M. L. (1948). *Biochem. J.* **42,** 116.

Butler, J. A. V., Phillips, D. M. P., Stephen, J. M. L., and Creeth, J. M. (1950). *Biochem. J.* **46,** 74.
Cannan, R. K. (1944). *J. Biol. Chem.* **152,** 401.
Carpenter, D. C. (1940). *J. Am. Chem. Soc.* **62,** 289.
Carpenter, D. C. (1941). *J. Franklin Inst.* **232,** 76.
Chibnall, A. C., and Rees, M. W. (1951). *Biochem. J.* **48,** xlvii.
Christensen, H. N. (1943). *J. Biol. Chem.* **151,** 319.
Christensen, H. N. (1945). *J. Biol. Chem.* **160,** 75.
Christensen, H. N., and Hegsted, D. M. (1945). *J. Biol. Chem.* **158,** 593.
Coleman, D., and Howitt, F. O. (1946). Fibrous Proteins. Society of Dyers and Colourists, Bradford, p. 144.
Coleman, D., and Howitt, F. O. (1947). *Proc. Roy. Soc. London* **A190,** 145.
Consden, R. (1948). *Nature* **162,** 359.
Consden, R., and Gordon, A. H. (1950). *Biochem. J.* **46,** 8.
Consden, R., Gordon, A. H., and Martin, A. J. P. (1944). *Biochem. J.* **38,** 224.
Consden, R., Gordon, A. H., and Martin, A. J. P. (1946). *Biochem. J.* **40,** 33.
Consden, R., Gordon, A. H., and Martin, A. J. P. (1947a). *Biochem. J.* **41,** 590.
Consden, R., Gordon, A. H., and Martin, A. J. P. (1948). *Biochem. J.* **42,** 443.
Consden, R., Gordon, A. H., and Martin, A. J. P. (1949). *Biochem. J.* **44,** 548.
Consden, R., Gordon, A. H., Martin, A. J. P., and Synge, R. L. M. (1947b). *Biochem. J.* **41,** 596.
Craig, L. C. (1944). *J. Biol. Chem.* **155,** 519.
Craig, L. C., Gregory, J. D., and Barry, G. T. (1949). *Cold Spring Harbor Symposia Quant. Biol.* **14,** 24.
Craig, L. C., and Post, O. (1949). *Anal. Chem.* **21,** 500.
Cremer, H. D., and Tiselius, A. (1950). *Biochem. Z.* **320,** 273.
Dalgliesh, C. E., Johnson, A. W., Todd, A. R., and Vining, L. C. (1950). *J. Chem. Soc.* 2946.
Dekker, C. A., Stone, D., and Fruton, J. S. (1949). *J. Biol. Chem.* **181,** 719.
Desnuelle, P., and Casal A. (1948). *Biochim. et Biophys. Acta* **2,** 64.
Desnuelle, P., Rovery, M., and Bonjour, G. (1950). *Biochim. et Biophys. Acta* **5,** 116.
Dirr, K., and Felix, K. (1932). *Z. physiol. Chem.* **205,** 83.
Drake, B. (1947). *Nature* **160,** 602.
Drake, B. (1950). *Acta Chem. Scand.* **4,** 554.
Drucker, B., and Smith, S. G. (1950). *Nature* **165,** 196.
Dunn, M. S. (1925). *J. Am. Chem. Soc.* **47,** 2564.
Durrum, E. L. (1950). *J. Am. Chem. Soc.* **72,** 2943.
Edman, P. (1950). *Acta Chem. Scand.* **4,** 283.
Eeg-Larsen, N., Linderstrøm-Lang, K., and Ottesen, M. (1948). *Arch. Biochem.* **19,** 340.
Elsden, S. R., and Synge, R. L. M. (1944). *Biochem. J.* **38,** ix.
Felix, K., Fischer, H., Krekels, A., and Rauen, H. M. (1950). *Z. physiol. Chem.* **286,** 67.
Felix, K., and Lang, A. (1929). *Z. physiol. Chem.* **182,** 125.
Felix, K., and Mager, A. (1937). *Z. physiol. Chem.* **249,** 111.
Felix, K., and Schuberth, H. (1942). *Z. physiol. Chem.* **273,** 97.
Fischer, E., and Abderhalden, E. (1907). *Ber.* **40,** 3544.
Fodor, A., Fodor, P. J., and Kuk-meiri, S. (1947). *Enzymologia* **12,** 101.
Fowden, L., and Penney, J. R. (1950). *Nature* **165,** 846.

Fox, S. W. (1945). *Advances in Protein Chem.* **2,** 155.
Fromageot, C., Jutisz, M., Meyer, D., and Pénasse, L. (1950). *Biochim. et Biophys. Acta* **6,** 283.
Fruton, J. S. (1950). *Yale J. Biol. Med.* **22,** 263.
Goldschmidt, S., Freyss, G., and Strauss, K. (1933). *Ann.* **505,** 262.
Gordon, A. H. (1949). *Discussions of the Faraday Soc.* **7,** 128.
Gordon, A. H., Keil, B., and Sebesta, K. (1949). *Nature* **164,** 498.
Gordon, A. H., Martin, A. J. P., and Synge, R. L. M. (1941). *Biochem. J.* **35,** 1369.
Gordon, A. H., Martin, A. J. P., and Synge, R. L. M. (1943). *Biochem. J.* **37,** 92.
Grant, R. L., and Lewis, H. B. (1935). *J. Biol. Chem.* **108,** 667.
Grassmann, W. (1950). *Angew. Chem.* **62,** 170.
Grassmann, W., and Hannig, K. (1950). *Naturwissenschaften* **37,** 397.
Green, F. C., and Schroeder, W. A. (1951). *J. Am. Chem. Soc.* **73,** 1385.
Green, N. M., and Work, E. (1951). *Biochem. J.* **49,** xxxvii.
Gregory, J. D., and Craig, L. C. (1948). *J. Biol. Chem.* **172,** 839.
Güntelberg, A. V., and Linderstrøm-Lang, K. (1949). *Compt. rend. lab. trav. Carlsberg. Sér. chim.* **27,** 1.
Gutfreund, H. (1948). *Biochem. J.* **42,** 544.
Gutfreund, H., and Ogston, A. G. (1949). *Biochem. J.* **44,** 163.
Hamoir, G. C. M. (1945). *Biochem. J.* **39,** 485.
Hanby, W. E., and Rydon, H. N. (1946). *Biochem. J.* **40,** 297.
Harington, C. R., and Pitt-Rivers, R. (1944). *Nature* **154,** 301.
Haugaard, G., and Kroner, T. D. (1948). *J. Am. Chem. Soc.* **70,** 2135.
Haugaard, G., and Roberts, R. M. (1942). *J. Am. Chem. Soc.* **64,** 2664.
Hellerman, L., and Stock, C. C. (1938). *J. Biol. Chem.* **125,** 771.
Howard, G. A., and Martin, A. J. P. (1950). *Biochem. J.* **46,** 532.
Huang, H. T., and Niemann, C. (1950). *J. Am. Chem. Soc.* **72,** 921.
Huffman, H. M. (1942). *J. Phys. Chem.* **46,** 885.
Ingram, V. M. (1950). *Nature* **166,** 1038.
Jacobsen, C. F. (1949). *Compt. rend. lab. trav. Carlsberg. Sér. chim.* **26,** 455.
Jensen, H., and Evans, E. A. (1935). *J. Biol. Chem.* **108,** 1.
Johnston, R. B., Mycek, M. J., and Fruton, J. S. (1950). *J. Biol. Chem.* **185,** 629.
Jones, T. S. G. (1948). *Biochem. J.* **42,** xxxv.
Jones, T. S. G. (1949). *Discussions of the Faraday Soc.* **7,** 285.
Jutisz, M., and Lederer, E. (1947). *Nature* **159,** 445.
Kaplan, E. H., Campbell, E. D., and McLaren, A. D (1950). *Biochim. et Biophys. Acta* **4,** 493.
Keston, A. S., and Udenfriend, S. (1949). *Cold Spring Harbor Symposia Quant. Biol.* **14,** 92.
Keston, A. S., Udenfriend, S., and Cannan, R. K. (1946). *J. Am. Chem. Soc.* **68,** 1390.
Keston, A. S., Udenfriend, S., and Cannan, R. K. (1949). *J. Am. Chem. Soc.* **71,** 249.
Keston, A. S., Udenfriend, S., and Levy, M. (1947). *J. Am. Chem. Soc.* **69,** 3151.
Keston, A. S., Udenfriend, S., and Levy, M. (1950). *J. Am. Chem. Soc.* **72,** 748.
Kies, M. W., and Davis, P. L. (1950). *Federation Proc.* **9,** 189.
Kuhn, W. (1930). *Ber.* **63,** 1503.
Kuhn, W., Molster, C. C., and Freudenberg, K. (1932). *Ber.* **65,** 1179.
Lederer, E., and Tchen, P. K. (1947). *Biochim. et Biophys. Acta* **1,** 35.
Lens, J. (1949). *Biochim. et Biophys. Acta* **3,** 367.

Léonis, J. (1948). *Compt. rend. lab. trav. Carlsberg. Sér. chim.* **26,** 315.
Levene, P. A., and Bass, L. W. (1928). *J. Biol. Chem.* **78,** 145.
Levene, P. A., and Bass, L. W. (1929). *J. Biol. Chem.* **82,** 171.
Levene, P. A., Rothen, A., Steiger, R. E., and Osaki, M. (1930). *J. Biol. Chem.* **86,** 723.
Levene, P. A., Steiger, R. E., and Rothen, A. (1932). *J. Biol. Chem.* **97,** 717.
Levy, A. L. (**1950**). *J. Chem. Soc.* 404.
Levy, M., and Slobodiansky, E. (1929). *Cold Spring Harbor Symposia Quant. Biol.* **14,** 113.
Linderstrøm-Lang, K. (1949). *Cold Spring Harbor Symposia Quant. Biol.* **14,** 117.
Linderstrøm-Lang, K., and Ottesen, M. (1949). *Compt. rend. lab. trav. Carlsberg. Sér. chim.* **26,** 403.
Livermore, A. H., and du Vigneaud, V. (1949). *J. Biol. Chem.* **180,** 365.
Lugg, J. W. H. (1938). *Biochem. J.* **32,** 775.
McLaren, A. D. (1949). *Advances in Enzymol.* **9,** 75.
Mandl, L., Levy, B., and McLaren, A. D. (1950). *J. Am. Chem. Soc.* **72,** 1790.
Martin, A. J. P. (1949a). *Biochem. Soc. Symposia* **3,** 4.
Martin, A. J. P. (1949b). *Discussions of the Faraday Soc.* **7,** 332.
Martin, A. J. P. (1950). *Ann. Rev. Biochem.* **19,** 517.
Martin, A. J. P., and Mittelmann, R. (1948). *Biochem. J.* **43,** 353.
Martin, A. J. P., and Synge, R. L. M. (1941). *Biochem. J.* **35,** 1358.
Martin, A. J. P., and Synge, R. L. M. (1945). *Advances in Protein Chem.* **2,** 1.
Meyer, K. H., Fuld, M., and Klemm, O. (1940). *Helv. Chim. Acta* **23,** 1441.
Middlebrook, W. R. (1949). *Nature* **164,** 501.
Miller, G. L., and Andersson, K. J. I. (1942). *J. Biol. Chem.* **144,** 465.
Mitchell, H. K., Gordon, M., and Haskins, F. A. (1949). *J. Biol. Chem.* **180,** 1071.
Mitchell, H. K., and Haskins, F. A. (1949). *Science* **110,** 278.
Monnier, R., and Jutisz, M. (1950). *Bull. soc. chim. biol.* **32,** 228.
Monnier, R., and Penasse, L. (1950). *Compt. rend.* **230,** 1176.
Montroll, E. W., and Simha, R. (1940). *J. Chem. Phys.* **8,** 721.
Moore, S., and Stein, W. H. (1948). *J. Biol. Chem.* **176,** 367.
Moore, S., and Stein, W. H. (1949). *J. Biol. Chem.* **178,** 53.
Moring-Claesson, I. (1948). *Biochim. et Biophys. Acta* **2,** 389.
Myrbäck, K. (1949). *Arkiv. Kemi* **1,** 161.
Nakashima, R. (1927). *J. Biochem. (Japan).* **7,** 441.
Neurath, H., and Schwert, G. W. (1950). *Chem. Revs.* **46,** 69.
Northrop, J. H., Kunitz, M., and Herriott, R. M. (1948). Crystalline Enzymes, 2nd ed., New York, p. 73.
Partridge, S. M. (1948). *Biochem. J.* **42,** 238.
Partridge, S. M. (1949a). *Discussions of the Faraday Soc.* **7,** 296.
Partridge, S. M. (1949b). *Biochem. J.* **44,** 521.
Partridge, S. M. (1949c). *Biochem. J.* **45,** 459.
Partridge, S. M., and Davis, H. F. (1950). *Nature* **165,** 62.
Partridge, S. M., and Swain, T. (1950). *Nature* **166,** 272.
Partridge, S. M., and Westall, R. G. (1949). *Biochem. J.* **44,** 418.
Patton, A. R., Foreman, E. M., and Wilson, P. C. (1949). *Science* **110,** 593.
Pauling, L., and Niemann, C. (1939). *J. Am. Chem. Soc.* **61,** 1860.
Perlmann, G. E. (1949). *J. Am. Chem. Soc.* **71,** 1146.
Petermann, M. L. (1942). *J. Biol. Chem.* **144,** 607.
Petermann, M. L., and Pappenheimer, A. M. (1941). *J. Phys. Chem.* **45,** 1.

Phillips, D. M. P. (1948). *Nature* **161,** 53.
Phillips, D. M. P. (1949a). *Biochim. et. Biophys. Acta* **3,** 341.
Phillips, D. M. P. (1949b). *Nature* **164,** 545.
Pittom, W. W. P. (1914). *Biochem. J.* **8,** 157.
Polson, A. (1948). *Biochim. et Biophys. Acta* **2,** 575.
Porter, R. R. (1948). *Biochim. et Biophys. Acta* **2,** 105.
Porter, R. R. (1950a). *Biochem. J.* **46,** 473.
Porter, R. R. (1950b). *Biochem. J.* **46,** 479.
Porter, R. R. (1950c). *Methods in Medical Research* **3,** 256.
Porter, R. R., and Sanger, F. (1948). *Biochem. J.* **42,** 287.
Randall, S. S., and Martin, A. J. P. (1949). *Biochem. J.* **44,** ii.
Rauen, H. M., and Felix, K. (1948). *Z. physiol. Chem.* **283,** 139.
Rees, M. W. (1946). *Biochem. J.* **40,** 632.
Rideal, E. K., and Mitchell, J. S. (1937). *Proc. Roy. Soc. London* **A159,** 206.
Rittenberg, D., and Foster, G. L. (1940). *J. Biol. Chem.* **133,** 737.
Rometsch, R. (1950). *Helv. Chim. Acta* **33,** 184.
Sanger, F. (1945). *Biochem. J.* **39,** 507.
Sanger, F. (1946). *Biochem. J.* **40,** 261.
Sanger, F. (1948). *Nature* **162,** 491.
Sanger, F. (1949a). *Biochem. J.* **44,** 126.
Sanger, F. (1949b). *Biochem. J.* **45,** 563.
Sanger, F. (1949c). *Cold Spring Harbor Symposia Quant. Biol.* **14,** 153.
Sanger, F. (1949d). *Biochem. Soc. Symposia* **3,** 21.
Sanger, F., and Tuppy, H. (1951a). *Biochem. J.* **49,** 463.
Sanger, F., and Tuppy, H. (1951b). *Biochem. J.* **49,** 481.
Schmid, K. (1949). *Helv. Chim. Acta* **32,** 105.
Schramm, G., and Primosigh, J. (1943). *Ber.* **76,** 373.
Schramm, G., and Primosigh, J. (1944). *Ber.* **77,** 417.
Shepard, C. C., and Tiselius, A. (1949). *Discussions of the Faraday Soc.* **7,** 275.
Stein, W. H., and Moore, S. (1948). *J. Biol. Chem.* **176,** 337.
Stein, W. H., and Moore, S. (1949). *Cold Spring Harbor Symposia Quant. Biol.* **14,** 179.
Stein, W. H., Moore, S., and Bergmann, M. (1944). *J. Biol. Chem.* **154,** 191.
Steinhardt, J. (1941). *J. Biol. Chem.* **141,** 996.
Steinhardt, J., and Fugitt, C. H. (1942). *J. Research Natl. Bur. Standards* **29,** 315.
Steinhardt, J., and Zaiser, E. M. (1950). *J. Biol. Chem.* **183,** 789.
Svensson, H. (1948). *Advances in Protein Chem.* **4,** 251.
Svensson, H., and Brattsten, I. (1949). *Arkiv Kemi* **1,** 401.
Swanson, M. A., and Cori, C. F. (1948). *J. Biol. Chem.* **72,** 797.
Synge, R. L. M. (1943). *Chem. Revs.* **32,** 135.
Synge, R. L. M. (1944). *Biochem. J.* **38,** 285.
Synge, R. L. M. (1945). *Biochem. J.* **39,** 351.
Synge, R. L. M. (1949). *Biochem. J.* **44,** 542.
Synge, R. L. M., and Tiselius, A. (1947). *Acta Chem. Scand.* **1,** 749.
Synge, R. L. M., and Tiselius, A. (1949). *Acta Chem. Scand.* **3,** 231.
Synge, R. L. M., and Tiselius, A. (1950). *Biochem. J.* **46,** xli.
Tazawa, Y. (1949). *Acta Phytochim. (Japan)* **15,** 73.
Tiselius, A. (1947). *Advances in Protein Chem.* **3,** 67.
Tiselius, A. (1948). *Arkiv Kemi Mineral. Geol.* **26B,** No. 1.
Tiselius, A., and Claesson, S. (1942). *Arkiv Kemi Mineral. Geol.* **15B,** No. 18.

Tiselius, A., Drake, B., and Hagdahl, L. (1947). *Experientia* **3,** 651.
Tiselius, A., and Hagdahl, L. (1950). *Acta Chem. Scand.* **4,** 394.
Toennies, G., and Homiller, R. P. (1942). *J. Am. Chem. Soc.* **64,** 3054.
Tristram, G. R. (1949). *Advances in Protein Chem.* **5,** 83.
Troensegaard, N. (1947). *Acta Chem. Scand.* **1,** 672.
Turba, F. (1948). *Z. Vitamin-, Hormon- u. Fermentforsch.* **2,** 49.
Turba, F., and Enenkel, H. J. (1950). *Naturwissenschaften* **37,** 93.
Turba, F., and Richter, M. (1942). *Ber.* **75,** 340.
Uchino, T. (1934). *J. Biochem. (Japan)* **20,** 65.
van Slyke, D. D., Dillon, R. J., MacFadyen, D. A., and Hamilton, P. (1941). *J. Biol. Chem.* **141,** 627.
van Vunakis, H., and Brand, E. (1951). Abstracts, 119th Meeting, American Chemical Society, April, 1951, p. 28c.
Vickery, H. B. (1922). *J. Biol. Chem.* **53,** 495.
Vickery, H. B., and Osborne, T. B. (1928). *Physiol. Revs.* **8,** 393.
Villee, C. A., Linderstrøm-Lang, K., and Ottesen, M. (1950). *Federation Proc.* **9,** 241.
Virtanen, A. I., and Hamberg, U. (1947). *Acta Chem. Scand.* **1,** 847.
Virtanen, A. I., Kerkkonen, H., Hakala, M., and Laakonen, T. (1950). *Naturwissenschaften* **37,** 139.
Waldschmidt-Leitz, E., and Turba, F. (1940). *J. prakt. Chem.* **156,** 55.
Waldschmidt-Leitz, E., and Turba, F. (1941). *J. prakt. Chem.* **158,** 72.
Warner, R. C. (1942a). *J. Biol. Chem.* **142,** 705.
Warner, R. C. (1942b). *J. Biol. Chem.* **142,** 741.
Wasteneys, H., and Borsook, H. (1930). *Physiol. Revs.* **10,** 110.
Wieland, T. (1942). *Z. physiol. Chem.* **273,** 24.
Wieland, T. (1944). *Ber.* **77,** 539.
Wieland, T., and Fischer, E. (1948). *Naturwissenschaften* **35,** 29.
Wieland, T., and Wirth, L. (1949). *Ber.* **82,** 468.
Woiwod, A. J. (1949). *Biochem. J.* **45,** 412.
Woolley, D. W. (1949). *J. Biol. Chem.* **179,** 593.