THE STRUCTURAL CHEMISTRY OF PROTEINS

★

SPRINGALL

THE STRUCTURAL CHEMISTRY OF PROTEINS

H. D. SPRINGALL, M.A., D.Phil., F.R.I.C.
Professor of Chemistry
University College of North Staffordshire

LONDON
BUTTERWORTHS SCIENTIFIC PUBLICATIONS
1954

FOREWORD

For many years chemists and biologists have been interested in proteins and a very large effort has been expended in studying the problems of their composition, properties, and structure. As a result, a great amount of information about them has been gathered, and progress has been especially rapid during the last decade or two.

So many techniques of investigation have been applied to the proteins that it is difficult to find an author who is able to discuss all of them in a thoroughly competent way. Professor Springall is such a man : he is an organic chemist who for many years has been interested also in physical chemistry and modern structural chemistry, and who has himself carried out investigations not only in the field of organic chemistry, but also in the field of the determination of molecular structures by diffraction methods.

In this book he has successfully attacked the problem of presenting a succinct, reasonably detailed, and up-to-date account of all of the work that has been done on proteins, including their composition, chemical properties, physico-chemical properties, and molecular structure. I expect that the book will be found valuable not only to chemists and biologists who are working with proteins, but also to others who have an interest in the subject.

LINUS PAULING

Pasadena, California

PREFACE

THIS book has grown from a course of lectures which I used to give to undergraduates in the final year of the Honours School of Chemistry, and to post-graduate research workers, in the University of Manchester. Its origins go further back than that, to 1937, when, as a Commonwealth Fund Fellow, it was my good fortune to be present at memorable discussions between Professor L. Pauling, Professor W. T. Astbury, Dr. K. Landsteiner and Professor H. Neurath; and to realize the fascination of protein chemistry.

I hope the book will be found useful by those following an Honours course in Chemistry and by research workers, academic and industrial, and others interested in the chemistry of biological systems.

Each chapter is self-contained in the matter of references to the literature. I have endeavoured to make these references reasonably full, tending to stress certain review articles which provide ready access to a more extensive survey. [The valuable and monumental work, *The Proteins*, edited by H. Neurath and K. Bailey, New York, began to appear late in 1953 when this book was already in the press. That is why it is not mentioned in the references.] The abbreviations of titles of journals are those of the *World List*. I am not happy about some of them (notably *Rep. Progr. Chem.* and *Proc. roy. Soc.* for publications I always think of as *Ann. Rep.* and *Proc. Roy. Soc.* but the overriding consideration of facilitating library work by using standardized abbreviations seems to me of paramount importance.

In common with others working in a large and active field, I am very conscious that the subject has not stood still for the benefit of the recorder and that this has led to a certain jerkiness in the record.

It is a pleasant duty to acknowledge my indebtedness to those friends who have fashioned my interest in structural organic chemistry, and in whose laboratories I have been privileged to work: firstly, to my old tutor, Professor N. V. Sidgwick, to Sir Robert Robinson, Professor L. Pauling, Professor E. L. Hirst and Professor E. R. H. Jones. To them is very largely due such milk as may be in the coconut. If some of it has curdled with keeping, the fault is mine, not theirs.

In the actual preparation of the book I am indebted, above all, to my wife for her unfailing encouragement and help in proof-reading and preparing illustrations and indexes: to my colleagues,

Drs. S. C. Nyburg and G. F. Smith, for criticism of the sections on x-ray crystallography and pure organic chemistry, respectively. to my secretary, Miss E. E. Perkins, who has cheerfully typed the manuscript, much of it several times: to those who have generously permitted, and assisted in, the reproduction of illustrations: and to the publishers for their helpfulness and consideration in many ways.

H. D. S.

Keele, Staffordshire.

CONTENTS

1

INTRODUCTION

BIOLOGICAL IMPORTANCE OF THE PROTEINS

In recent years it has become more and more clearly recognized that very large molecules, macromolecules of molecular weight > 10,000, are of critical importance in biological processes, having functions of great diversity and often of extreme specificity. The structural material of plants and, to some extent, of animals is composed of such molecules, and modern investigation presents a view of the life processes within the biological cell as involving intimate chemical and physical interactions between such molecules. These chemical interactions are of a special type in that they proceed very gently, without violent liberation or absorption of energy, but they will only proceed if the reactant molecules are arranged in a particularly favourable pattern—they have low energies, but high entropies, of activation—characteristics which have been discussed, for example, by C. N. Hinshelwood[1,2] (1946 and 1947).

The macromolecules responsible for both the structural and the cell transformation functions belong to three chemical classes: the proteins, the polysaccharides and the nucleic acids, of which we shall consider the first. This class is of extreme importance, thus:

(*1*) Many animal structural materials, *e.g.* hair, horn, feathers, skin, cartilage and muscle, are composed of proteins.

(*2*) Some of the hormones, *e.g.* insulin, are proteins.

(*3*) The enzymes are all either proteins or complexes made up of a protein–non-protein prosthetic group, the prosthetic groups in such systems being called co-enzymes.

(*4*) The substances responsible for oxygen transport in respiration processes, *e.g.* haemoglobin, are also complexes each made up of a protein–non-protein prosthetic group, the prosthetic groups being organo-metallic pigments.

(*5*) The material of the cell nucleus, which is of critical, if only partly understood, significance in reproduction processes, is largely composed of complexes made up of protein–non-protein (nucleic acid) ' nucleoproteins '. Complexes of protein–non-protein (lipoid) ' lipoproteins ', and of protein–non-protein (polysaccharide) ' mucoproteins ', are also found in the cell nucleus, and widely distributed elsewhere in living systems.

(*6*) Many viruses are either proteins or nucleoproteins, characterized by extremely high molecular weight ($\sim 10^7$).

(*7*) The bacterial toxins are proteins and, amongst them, that from *Clostridium botulinum* is the most powerful poison yet known.

(*8*) Protein–protein, and protein–polysaccharide interactions underly the complex chemistry of the important fields of immunity and allergy, and protein–carotenoid interactions are of great importance in visual processes.

PROPERTIES AND CLASSIFICATION OF PROTEINS

Their importance in life processes, has led to intensive organic and physico-chemical investigation of the proteins. Their properties are, however, such as to render the work extremely difficult; for example, proteins cannot be distilled, and are insoluble in organic solvents. Many of them are very sensitive to heat and even to the mildest reagents. Any given protein is liable to occur as a component of a complex mixture of closely similar proteins. Moreover, the ordinary organic chemical criteria of purity, such as analysis and melting point, are of little or no use in the protein field and special criteria, often involving elaborate techniques, such as electrophoresis, sedimentation, or diffusion, have to be used. Thus, despite great advances, which undoubtedly give the broad outlines of the solution of the problem of protein structure, there are still many obscurities challenging the investigator.

There are two main classes of proteins, fibrous and globular. Most fibrous proteins are insoluble in ordinary aqueous media. Many of them appear to be of indefinitely high molecular weight. These fibres are composed of long, approximately linear, molecules arranged roughly parallel to the fibre axis. They are amorphous. Most are capable of being stretched and the stretched forms, when released, contract again. The proteins involved in biological structural material are of this fibrous class.

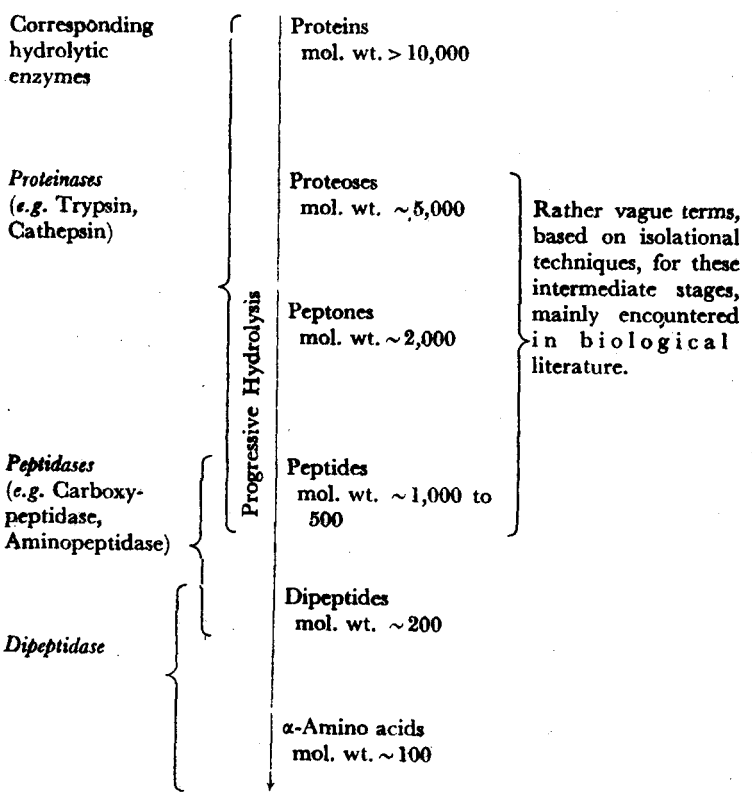
The globular proteins are soluble in aqueous media, either in water or in aqueous solutions of salts, acids, bases or alcohol. They have definite molecular weights falling in the range ten thousand to some millions*, the molecules in many cases being approximately spherical in shape. A large number have been crystallized. All such proteins undergo the characteristic process of ' denaturation ' —a molecular disorganization of far-reaching effect on physical and physiological properties. The proteins involved in physiological chemical reactions are of this globular class.

* The vaccinia virus body, a complex of protein, nucleic acids and lipids and probably other components, has a particle weight of $\sim 5 \times 10^9$.

HYDROLYTIC DEGRADATION OF PROTEINS

Proteins of both classes are hydrolysed, rather slowly by (*a*) strong acids, (*b*) strong alkalis, and (*c*) certain groups of specific enzymes. Alkali tends to racemize the optically active products of hydrolysis, and so is usually avoided. Enzymes act extremely slowly on fibrous proteins, and, in any case, being proteins themselves, they may contaminate the products (see Chapter 5). For normal organic chemical investigation the acid conditions are usually adopted, the technique being to reflux the protein for some hours with either ~35 per cent sulphuric acid (the excess can be conveniently removed, after hydrolysis, by barium carbonate) or ~20 per cent hydrochloric acid.

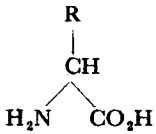
In the hydrolytic process the proteins are progressively broken down into smaller and smaller molecular fragments :



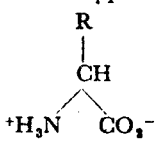
NATURAL α-AMINO ACIDS

On complete hydrolysis, proteins give almost exclusively (> 99 per cent) mixtures of a limited number of α-amino acids*. Many proteins, even after repeated crystallization, are found, on hydrolysis, to yield traces of compounds other than α-amino acids. For instance, many contain carbohydrates, *e.g.* egg-albumin, ~ 1 per cent mannose + glucosamine ; this carbohydrate is probably combined chemically, in a way not yet fully understood, with the protein, and its presence may well be of critical importance in determining the immunological properties of the protein. Some proteins contain phosphorus, *e.g.* myosin, ~ 0·05 per cent, while crystalline insulin contains zinc, ~ 0·3 per cent. There can be no doubt, however, that the general structure of proteins is dictated by their α-amino acid contents.

The α-amino acids are often assigned the general formula

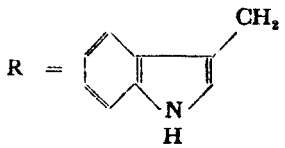


but their most characteristic properties depend on their existence as internal salts, zwitterions or dipolar ions, of general formula



Only about twenty-five different α-amino acids are known to occur in proteins : of these, nineteen occur frequently. A single protein usually contains between ten and twenty of these different α-amino acids.

The acids range in complexity from glycine, R=H, to tryptophan



In a given acid, the side chain, R, may be a chemically-inert aliphatic alkyl group, or may contain reactive groupings such as $\cdot CO_2H$, $\cdot NH_2$, $\cdot SH$, $\cdot OH$.

* No single enzyme will effect the complete hydrolysis from intact protein to individual α-amino acids in good yield. A series of enzymes, each specific for a given range in the progressive hydrolysis, must be used.

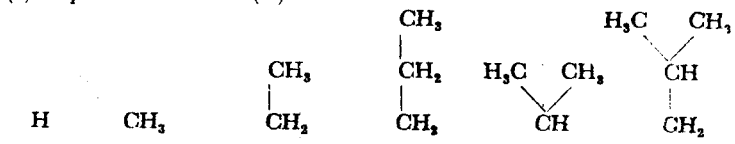
The α-amino acids with non-reactive side chains are not physiologically inert or interchangeable. This is demonstrated most clearly by the nutritional studies of W. C. ROSE[3] (1935) on rats. Thus, valine, $R=CH(Me)_2$, is essential to the diet for health; in its absence a characteristic and serious nervous disorder sets in, and this cannot be cured by supplying the closely related aliphatic side-chain acids leucine $R=CH_2{\cdot}CHMe_2$ or *iso*leucine $R=CHMeEt$, though it responds rapidly to the introduction of valine to the diet.

The characterizing side chains of the natural α-amino acids are classified and listed in *Table I.*

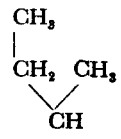
Table I

1. *Neutral α-Amino Acids*

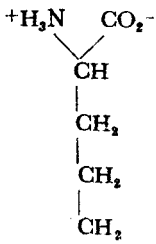
(*a*) *Aliphatic Side Chains* (R)



Glycine* Alanine α-Aminobutyric acid Norvaline Valine Leucine

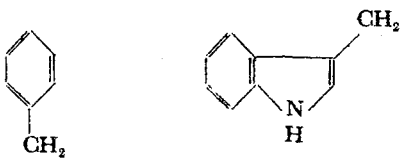


*iso*Leucine



α: α-Diaminopimelic acid.

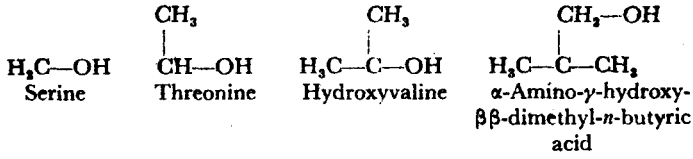
(*b*) *Side Chains* (R) *Containing Unsubstituted Aromatic Systems*

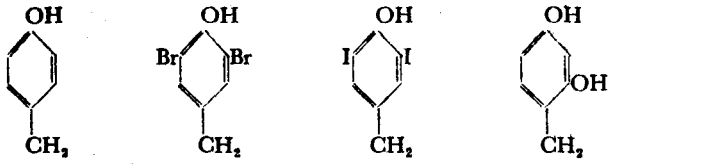


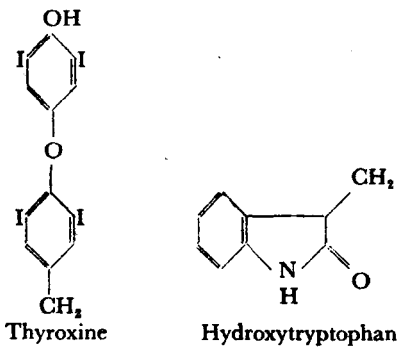
Phenylalanine Tryptophan

* N-Methylglycine, sarcosine, $Me^+NH_2{\cdot}CH_2{\cdot}CO_2^-$, long known as a decomposition product of creatine, has recently been isolated from the antibiotic peptide actinomycin[4] (C. E. DALGLEISH, A. W. JOHNSON, A. R. TODD and L. C. VINING, 1950), and from ground nut protein[5] (R. D. HAWORTH, R. MACGILLIVRAY and D. H. PEACOCK, 1951).

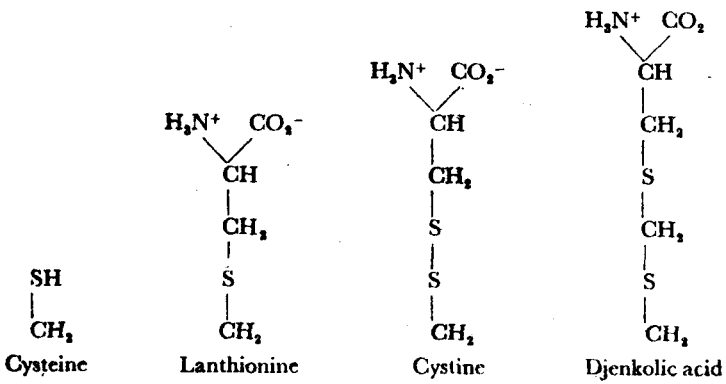
(*c*) *Side Chains* (R) *Containing Hydroxy Groups*



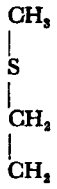


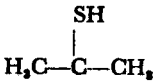


(*d*) *Side Chain* (R) *Containing Sulphur*

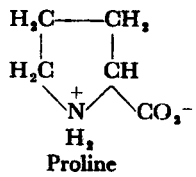


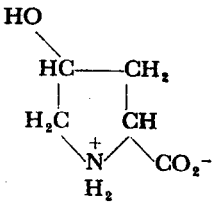

Homocysteine


Methionine


Dimethylcysteine

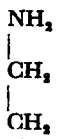
(*e*) *Side Chain* (R) *Cyclized : Imino Acids*

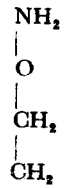

Proline

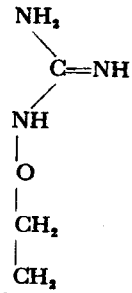

Hydroxyproline

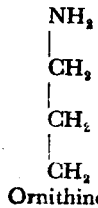
2. Basic α-Amino Acids

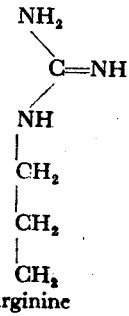
Side Chains (R)

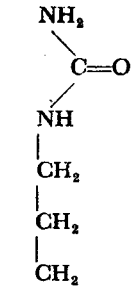

α : γ-Diaminobutyric acid

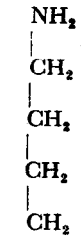

Canaline

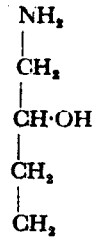

Canavanine

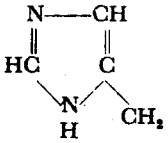

Ornithine

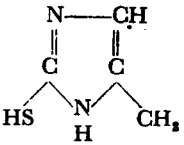

Arginine


Citrulline


Lysine


δ-Hydroxylysine

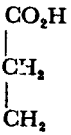

Histidine


Thiolhistidine

3. *Acidic α-Amino Acids*

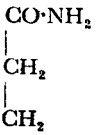
Side Chains (R)


Aspartic Acid


Glutamic Acid

These acids also occur as the corresponding amides—side chains (R)


Asparagine


Glutamine

H. B. VICKERY, in 1931[6] (with C. L. A. SCHMIDT) and 1941[7], discussed, authoritatively, the necessary steps in confirmation before a newly reported α-amino acid can be accepted as an established natural product: the subject has been further discussed by G. TOENNIES[8] (1943), R. J. BLOCK[9] (1946), and J. T. EDSALL[10] (1949). The main criteria are:

(*i*) the isolation must be repeated in another laboratory;
(*ii*) the constitution of the new acid must be confirmed by synthesis.

Not all the forty-one acids listed in *Table I* have complied with these conditions, nor have all of them been isolated from proteins. The position with regard to the actual constituents of the proteins may be stated as follows:

There are

(*a*) nineteen acids of general occurrence: glycine, alanine, valine, leucine, *iso*leucine, proline, phenylalanine, cysteine, cystine, methionine, tryptophan, arginine, histidine, lysine, aspartic acid (and asparagine), glutamic acid (and glutamine), serine, threonine and tyrosine*;

(*b*) five acids of definite occurrence but found infrequently: dibromotyrosine, diiodotyrosine, thyroxine, hydroxyproline and hydroxylysine;

(*c*) six acids of very rare occurrence which may be protein constituents: dihydroxyphenylalanine, lanthionine†, djenkolic acid, canavanine, citrulline and thiolhistidine.

Of the remaining eleven acids listed in *Table I*, two, norvaline and hydroxyvaline, have been claimed as protein constituents, but the claims have not yet been confirmed or disproved: seven, aminobutyric, diaminopimelic, and aminohydroxydimethylbutyric acids, dimethylcysteine, diaminobutyric acid, ornithine, and, probably, hydroxytryptophan, are definitely established constituents of natural products analogous to proteins but of lower molecular weight, and not yet identified in proteins: two, homocysteine and canaline, have, so far, been found only as intermediates in amino-acid metabolism.

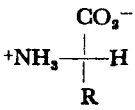
This statement is based on the analysis of H. B. Vickery[7] modified, as far as the protein constituents are concerned, by the inclusion of δ-hydroxylysine and the rejection of norleucine and hydroxyglutamic acid, following the studies of J. G. Heathcote[13] (1948), of R. Consden, A. H. Gordon, A. J. P. Martin, O. Rosenheim and R. L. M. Synge[14] (1945), and of H. D. Dakin[15] (1941), respectively.

All these α-amino acids, save glycine, have an asymmetric carbon atom (C_α), and are optically active when carefully isolated from proteins and allied natural sources. Almost all such optically active α-amino acids are of the same configuration, belonging to

* The common α-amino acids are here listed in the standard order suggested by E. Brand[11] (1946), and considered further in Chapter 5.

† Until 1951, lanthionine had only been isolated after alkali treatment and was suspected of being an artifact, but in that year G. Alderton and H. L. Fevold[12] isolated it unequivocally from the relatively low molecular weight antibiotic substance, subtilin, from *Bacillus subtilis*.

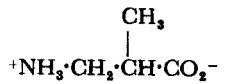
the L-series, and are represented by the conventional Fischer–Rosanoff projection formula



[see A. NEUBERGER[16] (1948) for a review]. The *sign* of rotation varies with R, thus natural alanine is L(+) while natural tyrosine is L(−).

Certain rare exceptions are known in which α-amino acids of the D-series occur in natural products, which though not proteins, are related to them, *e.g.* D(−)-glutamic acid in the capsular material of *Bacillus anthracis*, D(+)-phenylalanine in tyrocidin and gramicidin-S, D(+)-dimethylcysteine in penicillin, D(+)-proline in certain ergot alkaloids (see R. L. M. SYNGE[17] (1949) for a review).

Very rarely, β-amino acids are found in nature. β-Alanine was isolated from meat proteins by J. VON LIEBIG and was subsequently found to be a component of the essential growth-substance, pantothenic acid, a material widely distributed in nature but always occurring in minute amounts*. A new compound of this class, α-methyl-β-alanine,

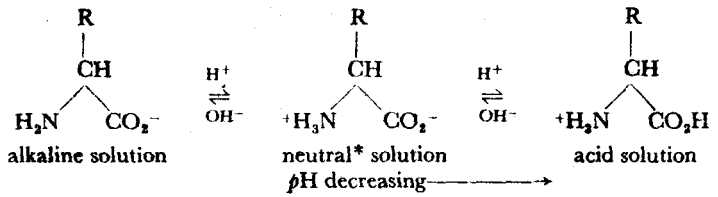


has been isolated (H. R. CRUMPLER, C. E. DENT, H. HARRIS, R. G. WESTALL, 1951)[19] from human urine. Taurine of ox bile is the sulphonic-acid analogue of β-alanine.

The α-amino acids are crystalline solids of high density (~1·5), melting indefinitely with decomposition at high temperatures (250°–300°C), soluble in water and some aqueous solutions, but insoluble in organic solvents.

These properties are characteristic of ionic substances and there is no doubt that the compounds exist in the crystalline state, and in neutral solution, as zwitterions (N. BJERRUM[20], 1923, see Part I of the monograph by E. J. COHN and J. T. EDSALL[21], 1943). The ionization effects operating on an amino acid in solution as the *p*H varies are as follows:

* The brilliant researches of R. J. WILLIAMS and his collaborators and of the Merck laboratories on pantothenic acid were reviewed by T. HENSHALL[18] in 1941.



The lack of volatility and the solubility relations of the α-amino acids make it difficult (*a*) to isolate them in a state of purity (since such an acid is usually prepared *either* mixed with other similar α-amino acids *or* mixed with inorganic salts, *e.g.* ammonium chloride, having similar solubilities); and (*b*) to investigate their organic chemical reactions. Nevertheless, versatile general methods have been devised which have resulted in the synthesis of almost all the naturally occurring acids. While, by combinations of inorganic and organic techniques, the isolational problem following synthesis has been largely solved (see H. B. VICKERY and C. L. A. SCHMIDT[6], M. S. DUNN[22] (1938–43), A. H. COOK[23] (1945), M. S. DUNN and L. B. ROCKLAND[24] (1947), and H. N. RYDON[25] (1950) for reviews).

The zwitterion structure can be broken down in two ways:

(*1*) The amino group can be acylated by a variety of reagents, when the product can be treated as a normal carboxylic acid;

(*2*) The carboxyl group can be esterified and the product treated as a rather unstable, but reasonably normal, amino compound.

The amino group of an α-amino acid can be eliminated, as nitrogen with nitrous acid, as ammonia with ninhydrin. The carboxyl group can be eliminated as carbon dioxide with ninhydrin. These reactions have been made quantitative (see Chapter 5).

The separation of a mixture of α-amino acids into its components and their individual estimation is a general problem of critical importance in the investigation of a protein and is considered in Chapter 5.

PEPTIDE BOND HYPOTHESIS FOR PROTEIN STRUCTURE

A consideration of the results of the hydrolysis of proteins leads, inescapably, to the idea that these compounds must be built up from α-amino acid units, and, since the average molecular weight

* For each amino-acid there is a definite *p*H, the isoelectric point, at which the compound does not migrate in an electric field, because the net charge on the molecule is zero. For most neutral amino-acids the isoelectric point is close to *p*H 6.