AN INTRODUCTION TO ENZYME CHEMISTRY

P. F. Leadlay

Chemical Society

NOGRAPHS FOR TEACHERS NUMBER 32

THE CHEMICAL SOCIETY
MONOGRAPHS FOR TEACHERS No. 32

An Introduction to Enzyme Chemistry

P. F. LEADLAY, MA, DPhil
Department of Biochemistry, Oxford University

LONDON: THE CHEMICAL SOCIETY

Monographs for Teachers

This is another publication in the series of Monographs for Teachers which was launched in 1959 by the Royal Institute of Chemistry. The initial aim of the series was to present concise and authoritative accounts of selected well-defined topics in chemistry for those who teach the subject at GCE Advanced level and above. This scope has now been widened to cover accounts of newer areas of chemistry or of interdisciplinary fields that make use of chemistry. Though intended primarily for teachers of chemistry, the monographs will doubtless be of value also to a wider readership, including students in further and higher education.

First published November 1978

Published by The Chemical Society, Burlington House, London W1V 0BN, and distributed by The Chemical Society, Distribution Centre, Blackhorse Road, Letchworth, Herts SG6 1HN

Printed by Adlard & Son Ltd, Bartholomew Press, Dorking

Preface

This Monograph is intended as an introduction to the remarkable chemistry of enzymes, which are involved in speeding up almost every essential reaction in living systems. An elementary background in chemistry is assumed, but special terms are defined as they are introduced.

The central problem in enzyme chemistry is how to relate the structure of an enzyme to its ability to catalyse a specific chemical reaction. Here, the aim is to emphasise that a reasonable approach to this problem requires information from various complementary experimental methods. Two enzymes are described in detail, to give an idea of the progress which has been made in favourable instances.

The discussion throughout relates to purified enzymes studied in isolation, and in the case of intracellular enzymes their behaviour in the intact cell may well be different. Biologically important interactions between enzymes and other cell components are of great current interest, but, like an extended discussion of the varied chemistry of enzyme cofactors, this falls outside the present scope. Instead, attention is drawn to the growing importance of enzyme technology, that is, the production and use of enzymes for industrial, synthetic and analytical purposes.

I would like to express my thanks to all those who gave permission for reproduction of drawings or figures, and in particular Professors D. Blow, J. Rupley and C.-I. Brändén, who kindly sent photographs or unpublished material. I would also like to thank my wife for typing the manuscript, and several colleagues who read all or part of the manuscript and offered useful criticism.

CONTENTS

1. The Enzyme Molecule

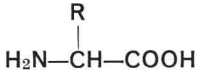
Introduction

In 1926 the American biochemist James Sumner obtained a crystalline substance from jack beans which appeared to consist wholly of protein, and which catalysed the rapid breakdown of urea. This was the enzyme urease, and Sumner's work was direct evidence for the idea that urease, and indeed other enzymes, are proteins. At the time his claims met with considerable scepticism, but it is now firmly established that all enzymes are proteins, although in many cases additional components are essential for catalytic activity. These are usually metal ions or small organic molecules.

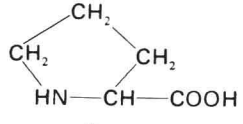
Proteins are large (macro-) molecules, with a relative molecular mass typically in the range 10 000 to 1 million. They consist of a large number of *amino acid* units or *residues*, covalently linked together to form long, unbranched chains. Formally, proteins are linear polymers (*polypeptides*) with the amino acids as the monomer units, but they differ sharply from their counterparts in synthetic polymer chemistry, in having the different monomer units in a defined order. This chapter aims to show how the number of each type of amino acid in the polypeptide chain, and the sequence in which they are connected together, give to each enzyme protein a well-defined three-dimensional structure and a unique chemical identity.

The amino acids

There are 20 amino acids commonly encountered in enzyme proteins. With one exception, they are carboxylic acids where the adjacent (α) carbon atom is also attached to a primary amino group, a hydrogen atom, and a *side-chain* that characterises the particular amino acid. Their general structure may be written:



where R represents the side-chain. The exception is proline, which is strictly speaking an α-imino acid in which the side-chain is looped onto the α-nitrogen atom to form a ring:



proline

1

Acid–base properties

The presence of both an amino and a carboxyl group confers interesting acid–base properties on α-amino acids. An acid is defined, following Brønsted and Lowry, as a proton donor, while a base is a proton acceptor. For the dissociation of a Brønsted acid HA in water

$$HA + H_2O \rightleftharpoons A^- + H_3O^+$$

it is useful to define the quantity pK_a

$$pK_a = -\log_{10} K_a = -\log_{10} \frac{[H_3O^+][A^-]}{[HA]} \qquad 1$$

where K_a is the equilibrium constant for the dissociation. Comparison of 1 with the usual definition of pH

$$pH = -\log_{10}[H_3O^+] \qquad 2$$

gives, after rearrangement,

$$pK_a = pH + \log_{10} \frac{[HA]}{[A^-]} \qquad 3$$

This equation allows the degree of ionisation of the acid to be calculated at any pH, if the pK_a is known.

When the pH is equal to pK_a, the acid is exactly half-dissociated. Thus the pK_a value of an acid expresses its strength directly on the pH scale of acidity, a lower pK_a characterising a stronger acid. The strength of an organic base B is expressed by the pK_a for the dissociation of its conjugate acid, BH^+:

$$BH^+ + H_2O \rightleftharpoons B + H_3O^+$$

Clearly a stronger base is characterised by a higher pK_a for BH^+.

An α-amino acid with no ionisable groups in the side-chain shows two acid dissociations in the pH range 1–13, one near pH 3 and a second near pH 10. There is considerable evidence that such α-amino acids exist, in water at neutral pH, almost entirely in the dipolar form **I** rather than the undissociated form **II**:

I **II**

The dipolar or zwitterionic structure **I** explains the relatively high solubility of α-amino acids in water, their relatively low solubility in non-polar solvents, and their large dipole moments. It is confirmed by detailed study of their infrared and Raman spectra in solution.

The two acid dissociation processes for, say, glycine are, therefore, best represented:

$$H_3N^+—CH_2—COOH + H_2O \rightleftharpoons H_3N^+—CH_2—COO^- + H_3O^+ \quad pK_a\ 2.3$$

and

$$H_3N^+—CH_2—COO^- + H_2O \rightleftharpoons H_2N—CH_2—COO^- + H_3O^+ \quad pK_a\ 9.8$$

In proteins, all except the terminal α-amino and α-carboxyl functions are masked, and the ionisable groups on the side-chains of certain amino acids largely determine the acid–base properties. The chief importance of the acid–base behaviour of individual amino acids is that it provides the basis for several important separation methods. These include electrophoresis, where amino acids migrate under the influence of an applied electric field at a rate proportional to their net charge; and various combinations of electrophoresis with thin layer chromatography, in which amino acids partition between a polar stationary phase and a non-polar moving phase. One method, ion-exchange chromatography, merits a detailed description because of the key role it has played in structural studies on enzymes, but it will be helpful to examine first the properties of individual amino acid side-chains.

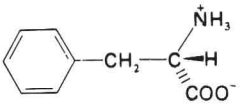
Side-chain properties

The structures, names (all of which are unashamedly non-systematic) and usual abbreviations for the common amino acids are given in *Fig 1*. The form shown is in each case the predominant one at neutral pH in water. Inspection of this list shows that the variation in structure of the side-chain is extremely wide. Two proteins differing from one another in amino acid composition, or in the sequence in which the amino acids are linked, may be anticipated to show quite different chemical properties.

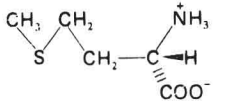
The amino acids may be usefully grouped, on the basis of the polarity of their side-chain, into three classes: non-polar, neutral polar, and charged polar. Those regarded as non-polar include glycine and the four amino acids with aliphatic hydrocarbon side-chains, alanine, valine, leucine and isoleucine. Proline, phenylalanine and methionine are also in this class.

The larger the non-polar side-chain, the more hydrophobic (water-hating) character the amino acid acquires. This may be verified experimentally by measuring, for each amino acid, the free energy change on transfer from water into a less polar solvent such as ethanol. As expected, the free energy change is least favourable for glycine, and most favourable for phenylalanine, tyrosine and tryptophan.

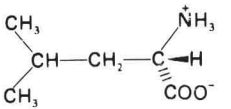
Non-polar



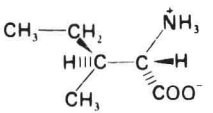
Phenylalanine (Phe)



Methionine (Met)

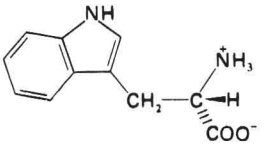


Leucine (Leu)

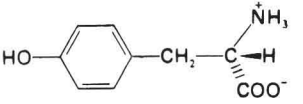


Isoleucine (Ile)

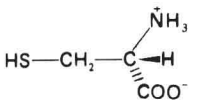
Neutral polar



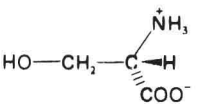
Tryptophan (Trp)



Tyrosine (Tyr)

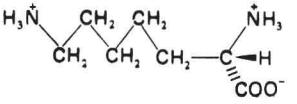


Cysteine (Cys)

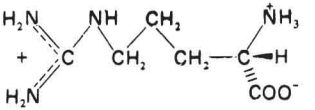


Serine (Ser)

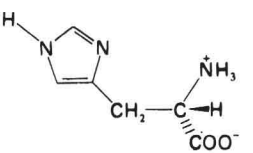
Charged polar



Lysine (Lys)



Arginine (Arg)



Histidine (His)

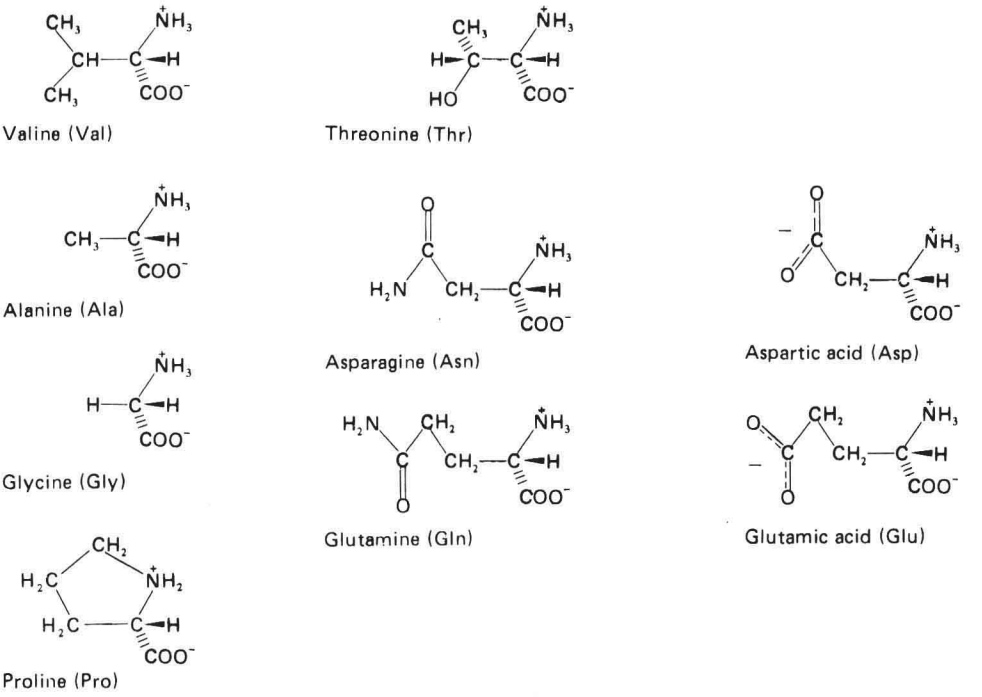


FIG. 1. The 20 amino acids commonly found in enzyme proteins. Dashed bonds are directed below and thickened bonds above the plane of the paper.

The neutral polar class includes the thiol amino acid cysteine, the hydroxylic amino acids serine and threonine, the aromatic amino acids tyrosine and tryptophan (which also have a great deal of non-polar character) and glutamine and asparagine, which have amide side-chains. All these side-chains can engage in strong hydrogen bonding through oxygen, nitrogen or sulphur, to other polar residues or to water molecules in the solvent. Cysteine plays a very important role in maintaining protein structure because of its ability to dimerise, through mild oxidation, to give cystine:

Such a cystine or disulphide –S–S– bridge may form between two cysteine residues in the same polypeptide chain, creating a loop, or may cross-link cysteine residues in two separate chains. Note that the –S–S– bridge is much more hydrophobic than the parent cysteine residues.

The charged polar amino acids are those which contain an ionising group, either acidic or basic, in their side-chain. The acidic amino acids, aspartic acid and glutamic acid, carry an extra negative charge at neutral pH, because the pK_a values of their side-chain carboxyl groups are 3.9 and 4.3 respectively. That the pK_a values are different illustrates how sensitively this function reflects small structural differences between acids. The thiol group in cysteine (pK_a 8.3) will be largely in the uncharged form at pH 7, and the phenolic side-chain of tyrosine (pK_a 10.1) almost completely so. The strongly basic amino group at C–6 (the ϵ-carbon) in lysine (pK_a 10.5) and the guanidino group in arginine (pK_a 12.5) will both be completely protonated at neutral pH, giving these amino acids an extra positive charge, while the imidazole ring in histidine (pK_a 6.0) will remain largely (90 per cent) in the uncharged form.

These polar charged side-chains may be expected to exert great influence on protein structure, because of their strong interaction with water and their sensitivity to pH changes. At the same time, the protein environment may profoundly affect the ionisation of individual groups and hence their chemical reactivity. A polar side-chain, at pH values above its pK_a, will be capable of acting either as a Brønsted base, by donating a pair of electrons to a proton, or as a nucleophile, by donating the electron pair to (for example) a suitably electrophilic carbon atom. At pH values lower than its pK_a, the conjugate acid form predominates, and this may function as a Brønsted acid. Catalysis by Brønsted acids and bases is discussed in detail in Chapter 2.

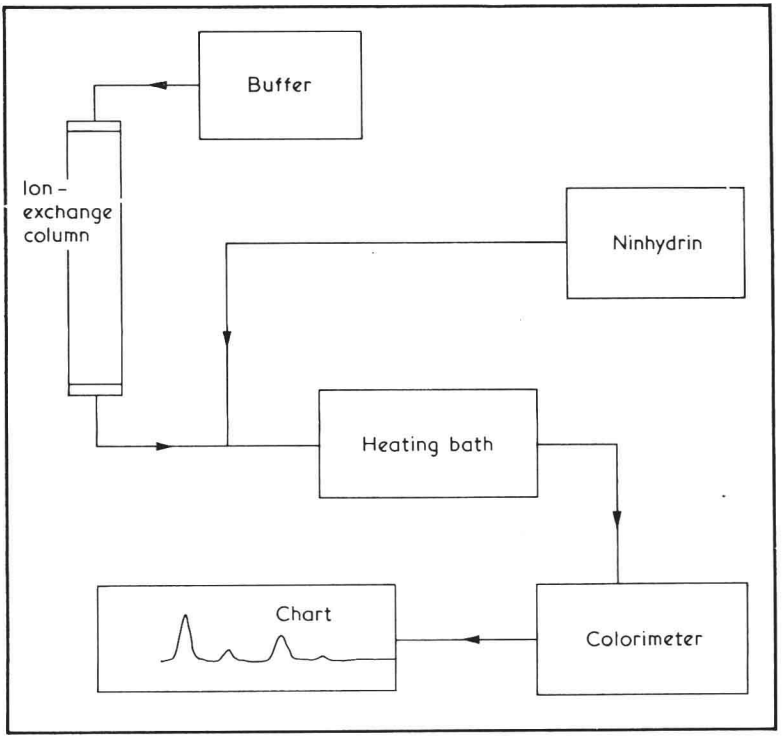
Amino acid analysis

All 20 common amino acids may be cleanly separated from one

another by column ion-exchange chromatography. An ion-exchange resin is used whose matrix contains sulphonic acid groups, which interact strongly with positively charged ions. A column of this material is prepared, and the mixture of amino acids is applied, at a pH (less than 3) where all the amino acids are positively charged and bind strongly to the resin. The ion-exchanger is then washed with a buffer of slightly higher pH, so that the acidic amino acids glutamic, aspartic and cysteic acids (the latter an oxidation product of cysteine), which are least firmly held, are gradually eluted from the column. By steadily increasing the pH and concentration of the buffer, all the other amino acids are eluted in turn. Among the neutral amino acids, the less polar tend to be eluted later, because of favourable interactions with the hydrophobic matrix of the ion-exchange resin. The basic amino acids lysine and arginine are eluted last of all.

This procedure has been refined to the point where it can be incorporated into an automatic system for the quantitative analysis of amino acid mixtures. The principle is shown in *Fig. 2*. As the

FIG. 2. An automatic system for the separation and quantitative determination of amino acids.



separated amino acids emerge from the column, they are mixed with ninhydrin reagent, which on heating with any amino acid gives—quantitatively—a strongly coloured derivative. The optical density of each peak of colour is measured, and traced out on a chart. Each amino acid emerges in a characteristic position on the chart, and the relative peak areas give a quantitative measure of the original sample composition.

Optical properties of the natural amino acids

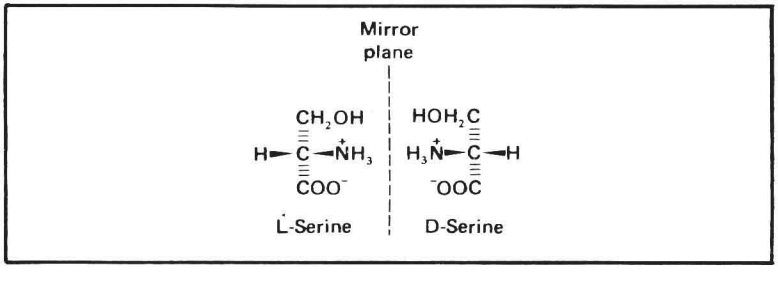
The optical activity of any compound, that is the ability to interact with plane-polarised light and rotate its plane of polarisation, is a function of its *symmetry* properties. Two alternative representations of the amino acid serine, conventionally labelled L and D respectively, are shown in *Fig. 3*. Dashed bonds are directed below, and thickened bonds above the plane of the paper.

These two compounds differ only in the spatial disposition of the groups and atoms about the central carbon, and are termed *stereoisomers*. Although they are related to one another just as an object to its mirror image, they are not identical. Using three-dimensional models, it can readily be shown that no manipulation makes one structure superimposable on the other. Such stereoisomers, which are non-superimposable mirror images of one another, are termed *enantiomers*. With the exception of glycine, all the α-amino acids can exist in either of two enantiomeric forms.

Compounds with this stereochemical property are said to be *chiral*, and to possess chirality (=handedness, from the Greek χειρ, hand). The higher symmetry of glycine renders it *achiral*: it cannot exist in enantiomeric forms. Chirality in α-amino acids is associated with the presence in the molecule of a carbon atom to which four different atoms or groups are attached. Such 'asymmetric' carbon atoms are often described as chiral centres, although strictly speaking chirality is a property of the whole molecule.

Chiral compounds are always optically active, but since one

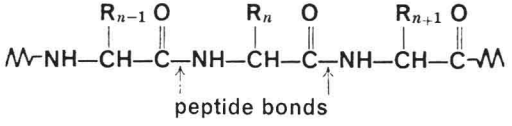
FIG. 3. The stereoisomeric forms of serine.



enantiomer will rotate the plane of plane-polarised light in a clockwise sense, and the other by the same amount in the opposite sense, the optical activity will only be measurable if one form is present in excess. Chemical synthesis of an amino acid such as serine will generally provide material which is optically inactive, because it contains equal amounts of both enantiomers. The amino acids obtained by hydrolysis of natural proteins are, however, optically active, because only the L-enantiomer of each chiral amino acid is used in protein biosynthesis.

The peptide bond

The covalent links between adjacent amino acid residues in a polypeptide chain are formed by the condensation of the α-carboxyl group of one residue with the α-amino group of the next:



The newly-formed amide bonds are called *peptide* bonds, and the condensation product is termed a *peptide* if the number of amino acid units is small, but a polypeptide or protein if the relative molecular mass exceeds about 10 000. A polypeptide generally contains a free α-amino group at one end, which is usually called the *N-terminus*, and one free α-carboxyl group at the other end (the *C-terminus*). Peptides and polypeptide chains are conventionally drawn with the *N*-terminus on the left. Additional peptide bonds could conceivably be formed involving side-chain amino and carboxyl groups, which would lead to branched or cross-linked chains, but such linkages are extremely rare.

The flexibility of the polypeptide chain is strongly influenced by geometrical restrictions on the peptide links. All the atoms of a peptide bond, including the α-carbon atoms of adjacent amino acid residues, are constrained to lie in a single plane (*Fig. 4*). Ample evidence for the planarity of the peptide bond is provided by extensive x-ray spectroscopic studies on the crystal structure of simple amides and peptides, from which it also emerges that the amide C–N bond is shorter and stronger than a normal C–N single bond. This extra stability is associated with the planar geometry, and there is a considerable energy barrier to any twisting of the C_α–C–N–C_α grouping out of the plane.

A peptide bond could, in principle, possess either of the two planar configurations shown in *Fig. 4*. However, the *cis* configuration is energetically less favourable (because carbon has a larger van der Waals radius than oxygen, and the two C_α groups get in each other's

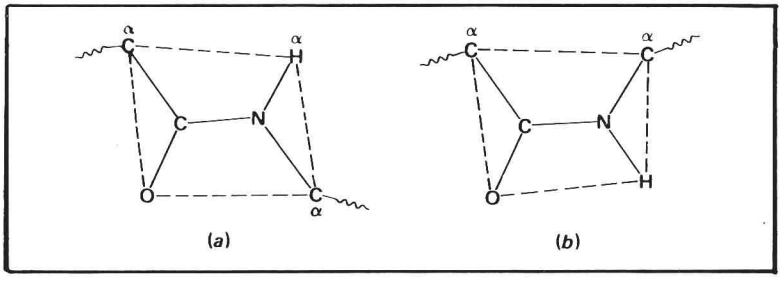
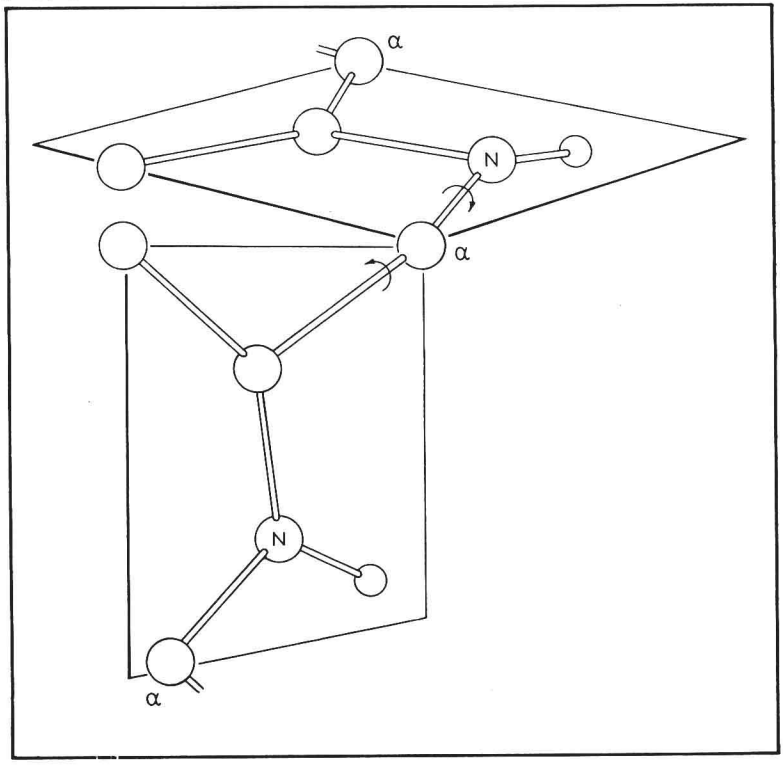


FIG. 4. The two alternative configurations of the planar peptide bond (*a*) *trans* and (*b*) *cis*.

way in *Fig. 4b*) and the *trans* configuration is almost invariably adopted in natural proteins.

A polypeptide chain, therefore, has rotational freedom only around the single bonds to each α-carbon atom, as indicated in *Fig. 5*. Even so, the number of different structures, or *conformations*,

FIG. 5. Rotation can still take place around the bonds to each α-carbon in a polypeptide chain.



which the polypeptide backbone can adopt is very large. Not all conformations of the chain will have equal stability. For certain angles of alignment of successive peptide planes, there will be unfavourable van der Waals interactions as non-bonded groups are forced into contact. Conversely, other angles of alignment minimise such unfavourable interactions and, when adopted by a whole sequence of successive peptide units, the result is a strikingly regular folding of the chain. Particular stability is associated with folding schemes in which each peptide –NH– group is able to form a hydrogen bond to a peptide carbonyl group either within the same chain or in a separate chain.

In the right-handed α-helix, for example, the polypeptide chain is closely wrapped around the helical axis, with every peptide carbonyl group aligned parallel to this axis and within hydrogen bonding distance of the peptide –NH– group three residues further along the chain (*Fig. 6*). Since the amino acid side-chains project outwards they do not interfere, unless adjacent side-chains are identically-charged or highly-branched. Otherwise only a proline residue, with its special geometry, will disrupt the helical pattern.

One form of the so-called β-sheet structure is shown in *Fig. 7*. In contrast to the α-helix, each chain is fully extended, and forms strong hydrogen bonds to two closely neighbouring strands, one on either side, creating a corrugated sheet-like structure. Those parts of an enzyme molecule which show regular folding of the polypeptide chain, such as the α-helix or β-sheet, are referred to as *secondary structure*. Other aspects of the three-dimensional structure, including the disposition of side-chains, are usually referred to as *tertiary structure*. When an enzyme consists of two or more separately folded polypeptide chains that associate together by non-covalent bonds, it is said to possess *quaternary structure*.

In summary, the distinctive geometrical properties of the peptide bond place a fair amount of restriction on the three-dimensional structure of polypeptide chains. Certain conformations are particularly stabilised by the absence of unfavourable steric interactions and a correct geometry for the formation of strong hydrogen bonds. In enzymes, as in other globular proteins, regular folding patterns tend to play a supporting role, and an over-riding influence on the final structure is exerted by the amino acid side-chains.

The primary structure

The *in vivo* assembly of a polypeptide begins at the *N*-terminus, and the chain grows as amino acids are successively added. The information codified in the DNA sequence of the appropriate gene dictates which of the 20 common amino acids is inserted at each stage, and each copy of a given polypeptide therefore contains the

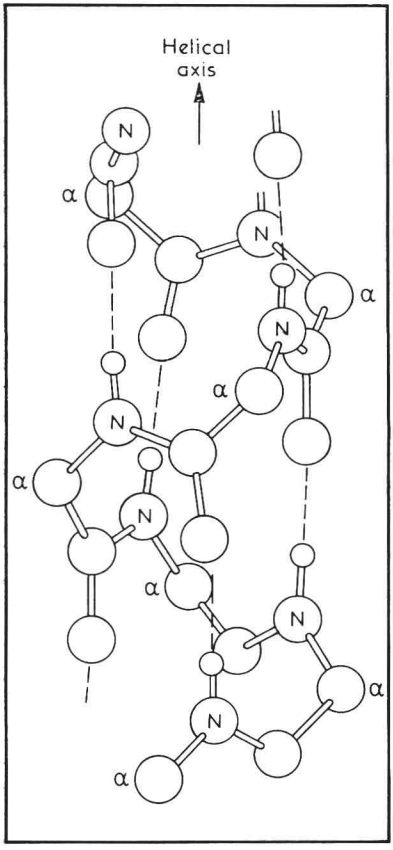


FIG. 6. The right-handed α-helix, in which hydrogen bonds are formed between peptide groups within the *same* chain. The side-chains are not shown. (From Linus Pauling, *The nature of the chemical bond.* New York: Cornell University Press, and Oxford: OUP, 1960.)

same unique sequence of amino acid residues. This unique sequence, referred to as the *primary structure*, confers a separate chemical identity on each enzyme protein.

The determination of primary structure

Because of its importance in determining the structure and reactivity of enzymes, there has been great interest in methods for determining primary structure. For a large number of enzymes this complex task has already been accomplished: chicken lysozyme, for example, an enzyme in egg-white which catalyses the degradation of bacterial cell walls, is now known to have 129 amino acid residues in a single polypeptide chain. The complete primary structure is shown in *Fig. 8.* The abbreviations used for the amino acids are those of *Fig. 1.*