CONFORMATION OF POLYPEPTIDES AND PROTEINS*†

By G. N. RAMACHANDRAN

Centre of Advanced Study in Biophysics
University of Madras, Madras, India
and
Department of Biophysics, University of Chicago
Chicago, Illinois

and

V. SASISEKHARAN

Centre of Advanced Study in Biophysics
University of Madras, Madras, India

* The literature survey for this review was completed in September 1967, with the journals which were then available in Madras and the preprints which the authors had received.

† By the authors' request, the publishers have left certain matters of usage and spelling in the form in which they wrote them.

I. Introduction

This review deals with the recent developments regarding the description and nature of the conformation of proteins and polypeptides, with special reference to the stereochemical aspects of the problem. We shall not consider here, except insofar as its affects the main argument, the methods by which the conformation is determined, such as X-ray diffraction, optical rotary dispersion, and so on. We shall restrict our discussion to the methods that are available for an adequate description of the conformation and its variations and of the various theoretical methods that have been developed to work out these aspects.

We shall therefore consider first the parameters that are required for an adequate description of a polypeptide chain. In this we shall focus our attention on what may be called "internal parameters," i.e., those which could be defined in terms of the relationships between atoms or units which form the building blocks of the polypeptide chains. We shall then give an account of the mathematical method of utilising these parameters for calculating the coordinates of all the atoms in a suitable frame of reference, so that all the interatomic distances, bond angles, etc., can be calculated and their consequences worked out.

Although in general a polypeptide chain can have a highly coiled conformation, it is not completely random and there are conditions restricting the allowed conformations in the neighbourhood of a residue along the chain. The laws governing such restrictions will first be discussed and then the theories which can be used for calculating the most favoured conformation will be discussed. The former can be given in terms of what are known as contact criteria, which do not allow the distances between pairs of atoms to be less than specified minima. The latter is done by making use of potential energy functions, which give the variation of the energy of interaction of a pair of nonbonded atoms in terms of the distance between them and also of the energies involved in various distortions and so on.

Finally, the observed conformations in amino acids, peptides, polypeptides, and proteins will be briefly discussed and related to the theoretical approaches outlined earlier. This discussion is necessarily made rather brief and restricted to those aspects which have relevance to the conformation in the immediate neighbourhood of residues in a chain. In other words, this review is not intended to contain an account of all aspects of conformation. Particularly in the case of globular proteins, the tertiary structure and the way in which the molecules pack up and interact with the surrounding medium are not discussed at all.

A few excellent reviews have been published on the subject of protein structure. Particular reference may be made to those by Dickerson (1964), Davies (1965, 1967), and Harrington *et al.* (1966). Some of the earlier reviews dealing specifically with the application of X-ray diffraction to protein structure are those by Crick and Kendrew (1957), Kendrew and Perutz (1957), Rich and Green (1961), and Kraut (1965).

II. Description of Polypeptide Conformation

A. The Peptide Unit

From the chemists' point of view, a protein chain is composed of amino acid residues linked together linearly in the form —NH—CHR—CO— and it may be described by a sequence of the type (NH_2—CHR_1—CO)—(NH—CHR_2—CO)—. . . .—(NH—CHR_n—COOH). It is a well-known fact that the polypeptide chains occurring in proteins are always linear and are not branched.

The above unit of the organic chemists, namely the amino acid residue, is, however, not the most useful for understanding the stereochemistry or conformation of this chain. This is because the groups NH and CO interact with each other in such a manner as to produce a partial double bond character to the bond C—N. The two resonance structures which lead to this partial double bond character are shown

in Fig. 1 (Pauling, 1960, p. 281). As a consequence, the atoms starting from one α-carbon atom and going to the next α-carbon atom, namely the set of atoms C^{α}—C'O—NH—C^{α}, form a planar group which has considerable rigidity. The best values for the bond lengths and bond angles of this group, which may be called the peptide unit, are given in

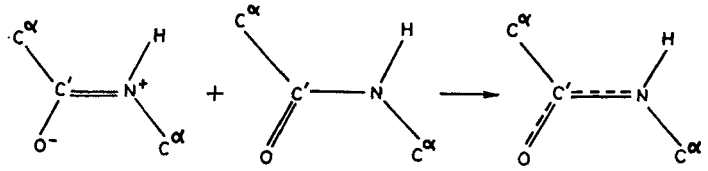


FIG. 1. The two resonant structures of the peptide unit which produce a partial double bond character for the peptide bond C'—N.

Fig. 2 and these are taken from a paper by Corey and Pauling (1953). These workers arrived at these data by correlating the available information at that time from X-ray structure analyses of various amino acids, peptides, and related compounds and also by taking into account theoretical considerations. It has since been found that the Pauling-Corey values for the dimensions of the peptide unit agree with the most

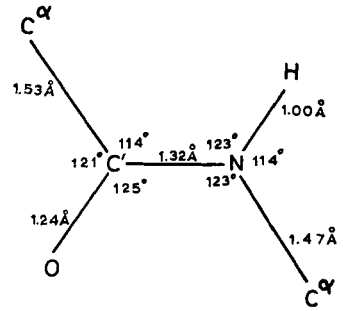


FIG. 2. The standard dimensions of the peptide unit, according to Pauling and Corey.

probable average values of all the later data (Hahn, 1957; Sasisekharan, 1962; Davies, 1965; Marsh and Donohue, 1967; and an extensive analysis recently made in the authors' laboratory). We may therefore take this unit as the building block of a polypeptide chain.

Cis and Trans Peptide Units

As mentioned above, the double bond character of the C—N group, due to the resonance between the two structures in Fig. 1, demands a planar configuration for the entire peptide group. However, for such a planar configuration, there are two possibilities, namely the cis and

the trans forms, as shown in Fig. 3. The cis form occurs, for instance, in the cyclic peptide diketopiperazine (Fig. 4A). However, it has now been proved that the trans conformation is the one that occurs uniformly in all open polypeptide chains (see Section VIII). Good evidence for this is provided by infrared spectra (Mizushima and Shimanouchi, 1961; see also Sasisekharan, 1962). In the early days, however, both cis and trans peptide units were often used in proposing structures for fibrous proteins. The greater stability of the trans conformation in relation to the cis may be judged by the fact that, in the cis conformation, the two large C^{α} atoms, at either end, occur as third neighbours in close contact,

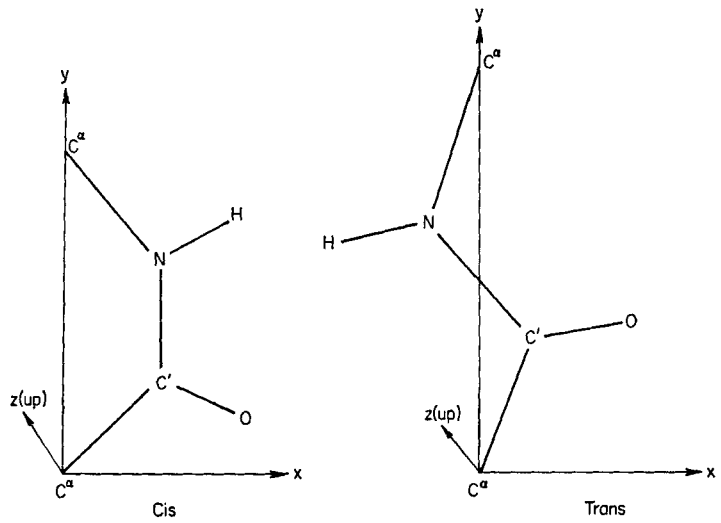


FIG. 3. Cis and trans modifications of the peptide unit. The choice of standard coordinate axes are also shown. The z-axis is up in both cases.

while in the trans conformation, only an oxygen occurs in this way as the third neighbour of C^{α}. It can be shown that there is an energy difference greater than 2 kcal/mole between the two conformations. This can be deduced from a nuclear magnetic resonance study (LaPlanche and Rogers, 1964) on the model compound *N*-methylacetamide (Fig. 5A), in which the cis form could not be observed at room temperature. This is mainly due to the large repulsion of the two methyl groups in the cis form, which makes it of much higher energy. In view of this, although the cis peptide unit might occur in small cyclic peptides like diketopiperazine, it is unlikely to be observed in an open peptide or polypeptide chain.

However, if the peptide unit has for its side group a five-membered ring, as in proline or hydroxyproline (Fig. 5B), it will be noticed that

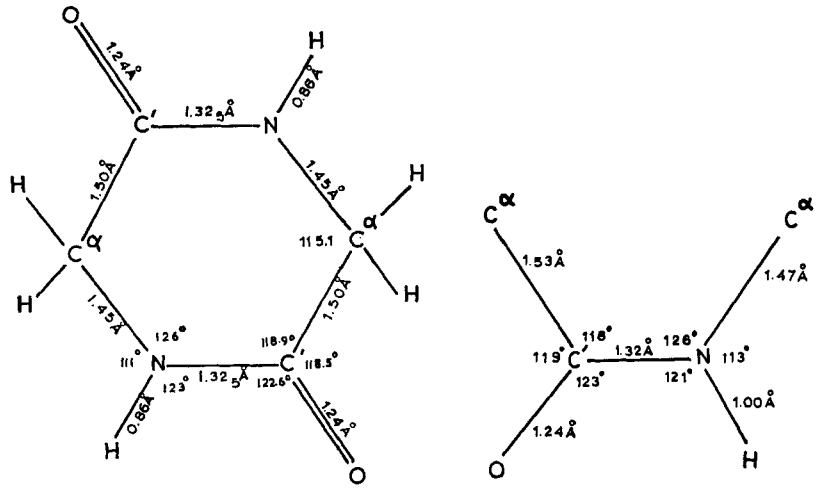


FIG. 4. A. Bond lengths and bond angles in diketopiperazine. The N—H bond length given is that determined by X-rays and is inaccurate. It is likely to be close to 1.0 Å. B. Suggested standard dimensions of the cis peptide unit.

the third neighbours of oxygen and of C^{α} are very similar, whether the peptide group is in the cis or trans conformation. Consequently, it would be expected that, when prolyl (or hydroxyprolyl) residues occur, it might be possible to form a cis peptide chain with good facility. This result is of interest in relation to the conformation of poly-L-proline (see Section VIII,E,3).

The coordinates of the atoms in cis and trans peptide units in a suitably chosen coordinate system are given in Tables I and II. In both cases, it is convenient to take the y-axis along the line joining one C^{α}-atom with the next one and the x-axis to be perpendicular to it in the plane of the peptide unit, as shown in Fig. 3. The z-axis would obviously be normal to the peptide plane upwards, so that x, y, z form

TABLE I

Rectangular Coordinates of the Atoms in the Pauling-Corey trans Peptide Unit

Atom	x(Å)	y(Å)
C_1^{α}	0.000	0.000
C_1'	0.578	1.417
O_1	1.804	1.607
N_1	−0.335	2.370
H_1	−1.317	2.181
C_2^{α}	0.000	3.801

a right-handed system of coordinates. The coordinates in Table I correspond to the Pauling-Corey dimensions mentioned earlier for the trans peptide unit. In the case of the cis peptide unit, two sets of values are given, one [Table II,(a)] corresponding to a rotation of the atoms beyond N in the standard trans peptide unit by 180°, and the second

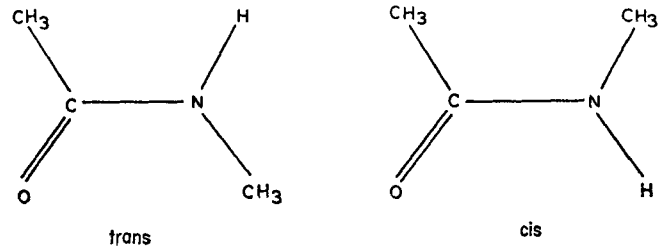


FIG. 5A. Structure of the cis and trans forms of *N*-methylacetamide.

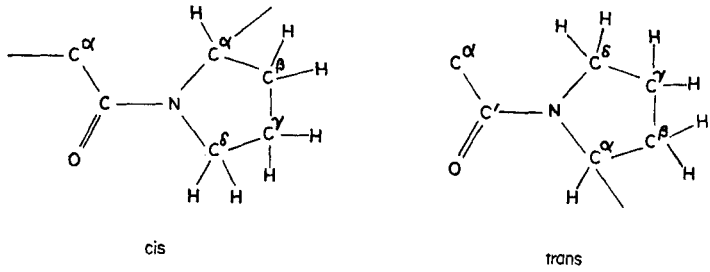


FIG. 5B. Cis and trans peptide units having a proline side group.

[Table II,(b)] corresponding to the dimensions as actually found in glycyldiketopiperazine (Degeilh and Marsh, 1959). Diketopiperazine is a closed ring system with resonances occurring in it and the dimensions of the cis peptide unit found in it are probably not the best for an open cis peptide unit. The best dimensions suggested for the latter are shown in Fig. 4B and the coordinates corresponding to this planar unit are given in Table II,(c). This will be called the standard cis peptide unit. (See Section VII,H for a justification of these dimensions.)

B. Two-Linked Peptide Units

As mentioned earlier, a protein chain consists of a number of peptide units linked to one another at α-carbon atoms. The two bonds by which neighbouring peptide units are joined at an α-carbon atom are N—C^{α} and C^{α}—C′. From the lengths of these two bonds in the peptide unit, it appears that they have practically single bond character and therefore one would expect a relatively free rotation about these two

TABLE II

Rectangular Coordinates of the Atoms in the cis *Peptide Unit*

(a) Planar peptide unit obtained by a rotation of 180° about the C—N bond from the trans peptide unit			(b) As observed in glycyl diketopiperazine[a]			(c) For the suggested standard conformation		
Atom	x(A)	y(A)	Atom	x(A)	y(A)	Atom	x(A)	y(A)
C_1^α	0.000	0.000	C_1^α	0.000	0.000	C_1^α	0.000	0.000
C_1'	1.358	0.705	C_1'	1.276	0.787	C_1'	1.309	0.792
O_1	2.415	0.056	O_1	2.351	0.170	O_1	2.385	0.176
N_1	1.279	2.023	N_1	1.212	2.110	N_1	1.235	2.110
H_1	2.083	2.617	H_1	2.023	2.695	H_1	2.063	2.672
C_2^α	0.000	2.748	C_2^α	0.000	2.905	C_2^α	0.000	2.907

[a] The hydrogen position as determined by X-rays is not very accurate. In this table, the bond length NH has been made equal to 1.00 Å.

bonds. Consequently, a variety of conformations can exist for a pair of linked peptide units because of the range of values that may be assumed for the angles of rotation about the two bonds by which they are linked at the α-carbon atoms. A particular conformation for a pair of peptide units may therefore be specified by stating the orientation of each of the two linked units about the bonds by which they are linked. For making such a specification, however, it is necessary to have a standard reference conformation with respect to which the others can be described. Although this way of describing the relative conformation of a pair of peptide units by two dihedral angles of rotation looks quite obvious, its consequences and the allowed ranges of these angles were worked out for the first time only in 1962–1963 (Ramachandran, 1962; Sasisekharan, 1962; Ramachandran *et al.*, 1963 a,b). These Madras workers called the two angles ϕ and ϕ' and used a certain way of defining the standard reference plane. At about the same time, a few others (for example, Schellman and Schellman, 1964; DeSantis *et al.*, 1965) also attempted to work out the limitation in the relative conformation of a pair of peptide units and it so turned out that the standard reference planes adopted by these persons were different. Some ideas of these dihedral angles had been adopted even earlier (Mizushima and Shimanouchi, 1961) and their conventions were still different. It was thus found that each of these sets of authors had adopted different conventions, both in defining the standard reference plane as well as the sense of positive rotation for the dihedral angles. In view of this, it was felt necessary to arrive at some standard conventions and this was done at a small workshop meeting of some of the leading workers in the field held in Bethesda in 1965 and later at the Gordon Conference on Proteins in the same year. As a consequence of this, certain standard notations and recommendations were published by Edsall *et al.* (1966a,b,c). Although the present authors have found it necessary to enlarge these (see Section II,E), they have tried as far as possible to conform to these conventions.

C. The Dihedral Angles Phi and Psi

The standard conformation from which the dihedral angles of rotation are measured is best described by saying that it corresponds to a fully extended form of the two peptide units. In this configuration (which is shown in Fig. 6), both the units are coplanar with the plane of the three atoms N, C^{α}, and C′, and the distance between the first C^{α} and the third C^{α} atoms is 7.2 Å, corresponding to a value of 110° for the angle N—C^{α}—C′. A general conformation may be obtained from this standard one by making rotations (clockwise) of angles ϕ and ψ about

the two bonds N—C^α and C^α—C′ respectively, as shown in Figs. 6A and 6B. The two rotations are best visualized by keeping in one's hand a pair of wire models of the peptide units (the so-called Kendrew skeletal models) linked at the common α-carbon atom. The two units are first kept in the same plane as each other and fully stretched as shown in Fig. 6A. (The C=O groups of the two units will then point roughly in opposite directions.) The screw of the second bond joining C^α to C′ is now tightened. Holding the first residue rigidly in the left hand, the second unit as a whole is given a rotation by the right hand through an angle ϕ about the bond N—C^α, in a clockwise sense looking from N towards C^α. This bond N—C^α is then tightened and the other bond

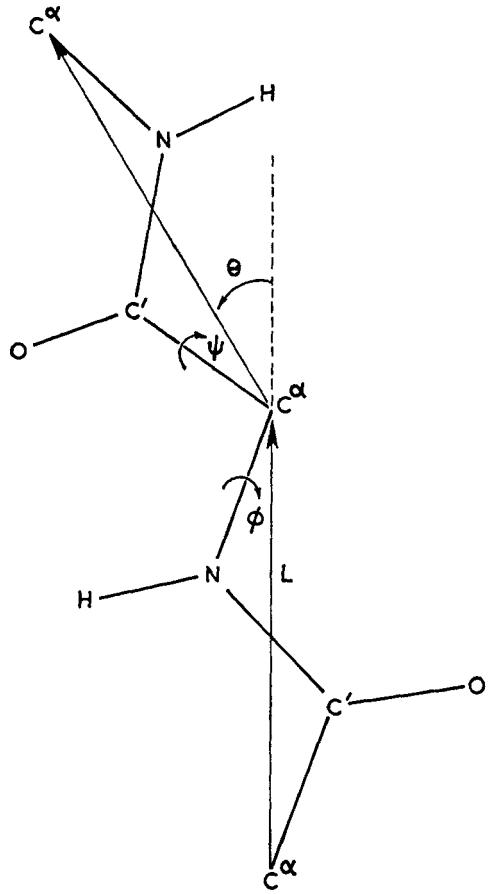


FIG. 6A. The standard conformation ($\phi=0$, $\psi=0$) of two linked peptide units. The dihedral angles of rotation, ϕ and ψ are also marked. The length L and the angle θ are referred to in Section II,A.

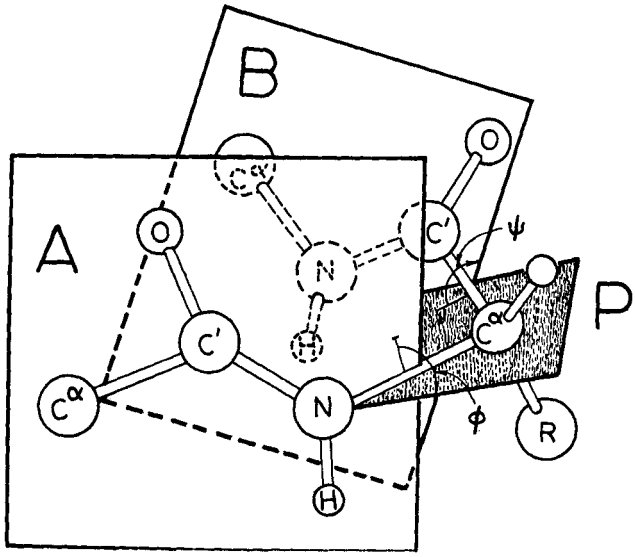


FIG. 6B. Diagram showing a general conformation (ϕ, ψ) about an α-carbon atom C^α. The shaded plane (P) contains the atoms N, C^α, C′ and the bonds N—C^α and C^α—C′ about which the rotations ϕ and ψ are made. The other two planes A and B represent the planes of the two linked peptide units. A clockwise rotation ϕ about N—C^α brings the plane A to the plane P and a clockwise rotation ψ about C^α—C′ brings the plane P to the plane B.

C^α—C′ is made loose. Holding the first unit firmly, the second unit is then rotated through the angle ψ about the second bond C^α—C′ once again in a clockwise sense, looking from C^α towards C′. Now if the second bond is also tightened, one has a conformation of the two linked peptide units corresponding to the pair of dihedral angles (ϕ, ψ) (Fig. 6B).

It will be noticed that, as soon as the ϕ-rotation is made, the orientation of the side group C^α—C^β bond becomes fixed and that this is unaffected by the subsequent ψ-rotation (see Section II,F,3 for building up of the side chain beyond the atom C^β).

Set of Dihedral Angles for a Peptide Chain

As will be noticed from the detailed description given above for obtaining a relative conformation (ϕ, ψ) for a pair of peptide units, it is a process which goes forward, as one proceeds along the chain. That is to say, one first makes a rotation about the bond N—C^α and then, from the resulting conformation, one makes a rotation about the bond C^α—C′. It is therefore obvious that if there are three linked peptide units, it may be described in full as $—C_1^\alpha—C_1'O_1—N_2H_2—C_2^\alpha—C_2'O_2—N_3H_3—C_3^\alpha—$

C'_3O_3—N_4H_4—C^α_4— (in this method of nomenclature, which involves giving the same ordinal number to all atoms in the same amino acid residue, we follow the recommendations of Edsall *et al.*, 1966a,b,c), and a conformation of the three linked peptide units may be specified by two sets of dihedral angles, namely (ϕ_2, ψ_2) and (ϕ_3, ψ_3), assuming of course that the peptide units themselves are planar and undistorted. Following the description given above, this conformation can readily be obtained, starting from a fully stretched conformation of the three units, by making the two rotations ϕ_2 and ψ_2 at the α-carbon atom 2 and then making the two rotations ϕ_3 and ψ_3 at the α-carbon atom 3. At each stage, all the atoms previous to the bond about which the rotation is made are held fixed, while all the atoms further ahead are rotated through the appropriate dihedral angle. This procedure indicates a general method of describing the conformation of a complete polypeptide chain containing N amino acid residues by first keeping them fully stretched and then carrying out stepwise the series of rotations (ϕ_1, ψ_1), (ϕ_2, ψ_2) . . . up to (ϕ_N, ψ_N) at each succeeding α-carbon atom. The set of N pairs of dihedral angles therefore contains, in a coded form, all the information that is necessary to describe the conformation of the backbone of the entire polypeptide chain.

We assume in this that each peptide unit is undistorted and has the standard Pauling-Corey dimensions and that at every α-carbon atom the angle is the same, equal to some constant value, e.g., 110°. This will never be true in practice and therefore in order to describe these variations, it is necessary to have further parameters to specify the conformation; but it is important to realise that, although these additional parameters are necessary to precisely delineate the exact conformation, the most important of the parameters are really ϕ and ψ and a good knowledge of the conformation of the peptide chain may be obtained from a set of such pairs of dihedral angles.

D. The Dihedral Angle Omega

Although the ideal Pauling-Corey peptide unit is an exact plane, in practice the atoms may not all lie in a plane. Insofar as the progress of the backbone of the polypeptide chain is concerned, this aspect of nonplanarity may be specified by having an extra dihedral angle ω in addition to the two angles ϕ and ψ. This angle ω represents a rotation about the peptide bond C′—N and may be defined as the angle between the planes C^α—C′—N and C′—N—C^α. More specifically, ω_i is the rotation about the bond C'_i—N_{i+1} linking the ith and $(i+1)$th amino acid residues. Following Edsall *et al.* (1966a,b,c), the angle ω may be taken to be equal to zero for a trans peptide unit and is taken to be

positive for a clockwise rotation looking along the peptide bond from C′ towards N. The values of ω occurring in a few crystal structures are shown in Table III, from which it will be seen that when the peptide unit is not involved in a strained situation, as in cyclic peptides, the value of ω rarely goes beyond a magnitude of about 10°. In cyclic hexapeptides, it is found to increase up to about 15°. It may be worthwhile mentioning at this point that a destablising energy of the form $K\omega^2$ is involved in forming a nonplanar peptide unit and that the value of K is of the order of from 15 to 30 kcal/mole (see Section VI,F,3,b for fuller details).

Obviously, a planar cis peptide unit would be described by a value of 180° for the parameter ω which is likely to be the most probable value for the angle in that region. It might be expected that the destabilising energy for a given ω for a cis peptide would be slightly smaller than that for a trans peptide.

Thus, the course of the backbone of a polypeptide chain is completely specified by a set of parameters $(\phi_i, \psi_i, \omega_i)$. The zero values of these parameters are specified by conditions 1-3 as follows and all the angles are measured clockwise looking along the direction of progress of the chain.

(1) $\phi_i = 0$ when the atoms C'_{i-1} and C'_i are trans about the bond $N_i - C^a_i$,

(2) $\psi_i = 0$ when the atoms N_i and N_{i+1} are trans about the bond $C^a_i - C'_i$, and

(3) $\omega_i = 0$ when the atoms C^a_i and C^a_{i+1} are trans about the bond $C'_i - N_{i+1}$,

Apart from the ω-rotation, nonplanarity of the peptide unit could also arise from the atoms O and H going out of plane and some methods must be available for describing this also. We shall not consider this now, but shall return to this aspect towards the end of this section (Section II,F) where a number of such subsidiary facts, and the methods for describing these, are considered together.

E. Description of the Conformation of the Side Group

We shall now consider the groups R_1, R_2, etc., which are attached to the α-carbon atoms of the main polypeptide chains. There are about twenty different types of these and their chemical formulae are well known. However, a precise definition of their stereochemical configuration requires a number of parameters. We shall follow for this the system adopted by Edsall *et al.* (1966a,b,c), with suitable additional notations for describing the hydrogen atoms. In their description of the

TABLE III
Values of ω and υ (in Degrees) Found in Simple Peptides[a]

Structure	ω	υ[d]
(a) Simple Peptides		
(i) *Asymmetric*[b]		
Gly-Asp	0.2	358.7
Cys-Gly-NaI	45.5	59.5
Glutathione	5.0	6.6
	15.5	10.8
NN'-Digly-Cys S	9.1	356.5
β-Ala-His	10.7	0.6
Leu-Gly-HBr	358.9	10.7
Leu-Pro-Gly	355.2	359.5
Thr-Phe nitrobenzyl ester HBr	0.2	357.7
Gly-Phe-Gly	4.6	359.4
	359.7	357.4
(ii) *Glycyl (nonasymmetric)*[c]		
β-Gly-Gly	1.3	355.6
Gly-Gly-Gly CuCl	1.5	5.6
	8.5	0.1
Na Gly-Gly-Gly Cu	6.9	7.9
	0.3	1.5
2Na Gly-Gly-Gly-Gly Cu	2.0	0.7
	0.4	1.1
	2.3	1.1
(b) Cyclic Peptides		
(i) *Asymmetric*[b]		
Ferrichrome A	3.4	355.1
	356.3	355.0
	10.6	10.9
	6.3	5.6
	355.9	358.0
	359.8	358.5
(ii) *Glycyl (nonasymmetric)*[c]		
Cyclohexaglycyl hemihydrate	3.2	6.2
	1.8	3.0
	9.2	10.8
	15.6	18.3
	1.3	0.9
	7.8	9.4
	6.3	3.0
	1.0	359.1
	2.6	12.7
	5.2	8.3
	0.4	356.7
	4.6	14.2
	6.1	1.2

TABLE III (Continued)

Structure	ω	v^d
Cyclohexaglycyl hemihydrate (Cont.)	7.8	1.9
	13.4	14.7
	5.5	2.5
	10.9	12.4
	5.8	354.2
	9.8	16.0
	6.8	357.6
	3.5	8.7
	6.1	2.8
	6.7	5.0
	2.5	7.6

[a] The data are reproduced from Lakshminarayanan (1968).

[b] The data for the asymmetric cases are for the L-configuration.

[c] In these cases, both the conformations with ω and v and $-\omega$ and $-v$ can occur.

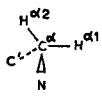
[d] The dihedral angle v is a measure of the nonplanarity of the carbonyl oxygen atom. (See Section II,F,1.)

side chain, Edsall *et al.* (1966a,b,c) considered only what might be called "heavy" atoms, that is nonhydrogen atoms, and neglected the hydrogen atoms. However, the hydrogens become important in studies of stereochemistry and conformational energy and therefore it is necessary to have a more comprehensive notation involving these also. This is described here.

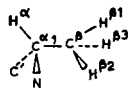
1. Notation Including Hydrogen Atoms

The symbols which are proposed for the bonds and atoms of the side chains of all the twenty commonly occurring amino acids and of hydroxyproline are listed in Fig. 7. The rules governing this nomenclature are described below.

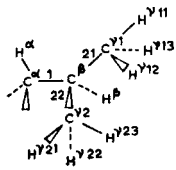
In this system, the heavy atoms going out from the α-carbon atom along the side chain are denoted sequentially by superscripts β, γ and so on, corresponding to the separation by a covalent bond at each stage. The bond from an alpha atom to a beta atom is denoted by a numerical symbol 1, that from a beta atom to a gamma atom by a symbol 2 and so on for all the bonds between heavy atoms. In branched side chains where more than one heavy atom occurs in the side chain beyond a particular atom, an additional symbol 1 or 2 is added to the notation both of the superscript symbol of the atom concerned and the bond symbol of the bond linking it to the earlier atom. For example, the two bonds from C^{β} to the two gamma carbon atoms $C^{\gamma 1}$ and $C^{\gamma 2}$ in valine will be denoted by symbols 21, 22, and similarly also in the case of the two



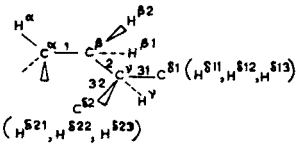
GLYCINE



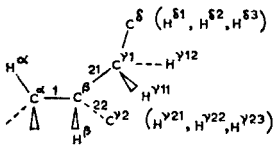
ALANINE



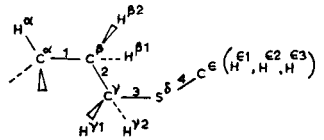
VALINE



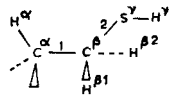
LEUCINE



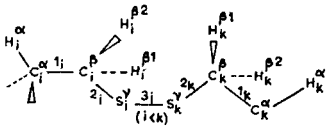
ISOLEUCINE



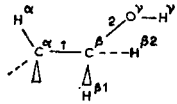
METHIONINE



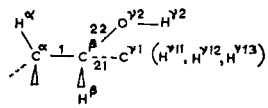
CYSTEINE



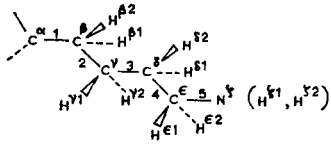
CYSTINE



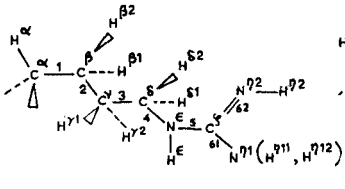
SERINE



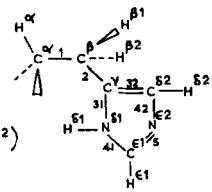
THREONINE



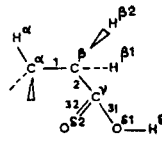
LYSINE



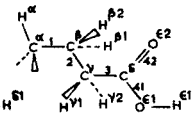
ARGININE



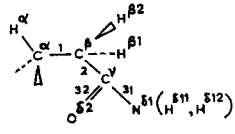
HISTIDINE



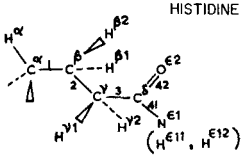
ASPARTIC ACID



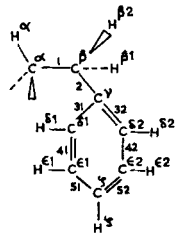
GLUTAMIC ACID



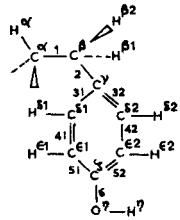
ASPARAGINE



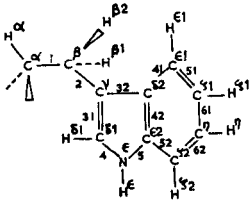
GLUTAMINE



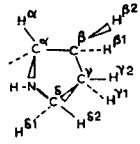
PHENYLALANINE



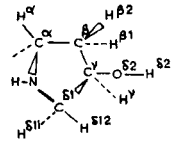
TYROSINE



TRYPTOPHAN



PROLINE



HYDROXYPROLINE

bonds C^{β}—O^{γ} and C^{β}—C^{γ} in threonine (Fig. 7). In such cases where there are multiple atoms attached to a preceding atom they will be denoted in increasing order according to the following rules:

1. If the atoms (groups) are dissimilar, the larger one will be given the symbol 1 and the other the symbol 2. In this case, by "larger" is meant a group with a larger number of atoms—thus CH_3 is larger than NH_2, which is larger than say CH, again larger than O. If both groups have the same number of atoms, the one which is heavier will be considered larger (this has not been found necessary in our examples). We shall call this the larger rule.

2. If the groups are alike, they will be numbered in increasing order in a clockwise sequence[1] looking down the previous bond, e.g., the two CH_3 groups of valine attached to the beta carbon atom. We shall call this the clockwise rule.

The hydrogen atoms attached to a heavy atom have the same superscript symbol as that of the corresponding heavy atom. If there is more than one hydrogen atom attached to a heavy atom, its superscript symbol will contain one more numerical symbol, which increases in a clockwise sense looking down the previous bond, by the clockwise rule mentioned above. For example, in the case of alanine, the three hydrogen atoms will be denoted by $H^{\beta 1}$, $H^{\beta 2}$, $H^{\beta 3}$ in clockwise order (Fig. 7); in the case of threonine, they will be numbered $H^{\gamma 11}$, $H^{\gamma 12}$, $H^{\gamma 13}$ and in the case of the two hydrogen atoms attached to $C^{\gamma 1}$ in isoleucine, $H^{\gamma 11}$ and $H^{\gamma 12}$, the latter being obtained from the former by a clockwise rotation about bond 21 (Fig. 7).

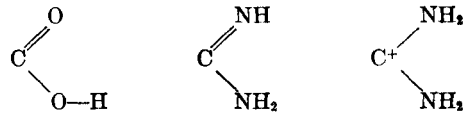
2. Ambiguous Cases

There are a few cases of terminal groups, such as (I) in aspartic and glutamic acids and (II) in arginine, in which the terminal heavy atoms are closely alike, and if the hydrogens have not been detected, or the determination of the bond lengths is not accurate enough to distinguish between the single and double bonds, the two atoms would be indis-

[1] This clockwise sense agrees with that generally used in polymer chemistry and in the description of organic molecules (see e.g., Klyne and Prelog, 1960). The dihedral angle (χ as defined below) specifying these bonds will increase with increasing ordinal number of the bond.

FIG. 7 (opposite). The details of the notation used for the description of all the atoms in the common amino acid chains, including the hydrogen atoms. The superscript symbols of the heavy atoms, the numerical symbols of the bonds connecting these, as well as the superscript symbols of the hydrogen atoms joined to the heavy atoms, are shown.

tinguishable. Particularly, in the case of arginine, the terminal group normally occurs in the charged form, (III), and the two NH_2 groups are



(I) (II) (III)

alike; or the carboxyl group may occur in the ionized form as COO^-, when the two oxygens are not different. In such cases, we must have an alternative definition of the branches 1 and 2—the branch 1 is taken to be the one which is closer to the cis configuration with respect to the previous heavy atoms. In aspartic acid, for example, the oxygen which is cis to C^α is denoted $O^{\delta 1}$ and in arginine, the nitrogen which is cis to C^δ is denoted as $N^{\eta 1}$. We shall denote this rule, which has to be applied when there is indefinite information, or when the two branches cannot be distinguished, as the cis rule. When distinction is possible, however, the larger rule will override the cis rule. Similarly, if, in an X-ray study, (OH) and (NH_2) cannot be distinguished, as in asparagine and glutamine, the symbols $(ON)^{\delta 1}$ or $(ON)^{\delta 2}$ or $(ON)^{\epsilon 1}$, $(ON)^{\epsilon 2}$ may be used, the 1 and 2 being obviously decided by the cis rule.

3. Side Groups Having Rings

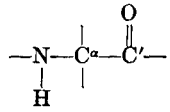
These rules suffice for all cases except those containing rings, such as phenylalanine, tyrosine, and tryptophan. In the former two cases, the two branches of the benzene ring are completely symmetrical, and it is immaterial which is called 1 and which 2. However, we may use the cis rule in this case. In practice (see Section VIII,B,2), neither branch is observed to be nearly cis, but the one closer to the cis configuration may be numbered 1. In the case of tryptophan, there are two fused rings, and there are multiple paths to the same atom from the α-carbon atom. Using the above rule, the numbering of the atoms may be given as in Fig. 7.[2] There is no difficulty in the case of the proline and hydroxyproline rings, where the hydrogens are automatically numbered by the clockwise rule.

4. Notation for Terminal Groups and Amino Acids

The symbols for the backbone atoms of an amino acid residue (IV) have already been given and these are all different from those of the

[2] For the heavy atoms, this differs slightly from the conventions adopted by Edsall *et al.* (1966a,b,c). In particular, the atoms $C^{\epsilon 1}$ and $C^{\epsilon 2}$ are interchanged, the former being called 1, since it is larger (CH) than the other (C). This necessitates the interchange of $C^{\delta 1}$ and $C^{\delta 2}$ which are identical (both CH).

hydrogen atom (H^{α}) and the side chain atoms attached to C^{α}. The subscript i may be added, where necessary, to all these symbols to indicate that they belong to the ith amino acid residue. We shall now consider the N-terminal group, say NH_2 or NH_3^+ and the C-terminal



(IV)

group, say COOH or COO^-, which occur at the two ends of the polypeptide chain. The rules already enunciated can be used in these cases.

1. For the NH_3^+ group, the three hydrogens are numbered H^1, H^2, H^3 in a clockwise sense, looking along the bond C—N (Fig. 8). If it is NH_2, the two hydrogens are numbered 1 and 2, the cis rule being operative here, as the two atoms are indistinguishable.

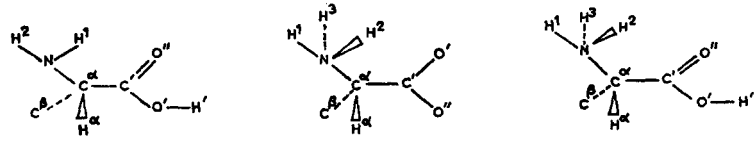


FIG. 8. The labelling of atoms in an amino acid, corresponding to the different modifications in which it is observed, e.g., in crystal structures. The notations for N-terminal and C-terminal groups in a polypeptide chain are consistent with these.

2. In the case of the COOH group, the two oxygens are distinguished as O′ and O″. If the group occurs in the un-ionized form, then the oxygen of the larger group OH is O′ and the other O″, and the hydrogen is denoted as H′. If the group is ionized as COO^- and the oxygen atoms are indistinguishable, the cis rule operates, and the one which is cis to N is called O′ and the other O″.[3] The corresponding ordinal number of the residue (i) can be added to these as subscript e.g., $i = 1$ for the N-terminal group and $i = N$ (the number of residues in the chain) for the C-terminal group.

The special case when both the N-terminal and C-terminal groups occur in the same "residue" ($i = 1 = N$) is an amino acid. The pos-

[3] It is not possible to call these atoms O^1 and O^2 (as in the side chain COOH) as this will lead to the atom H in the OH group being labelled H^1 to be consistent with our definitions; but this clashes with the symbol H^1 for one of the N-terminal hydrogens. Also, the use of O′ and O″ is in conformity with the accepted practice in a well-known laboratory working on X-ray protein structure studies.

sibilities which occur in this case, and the way in which the atoms may be labelled, are shown in Fig. 8.

5. Simplified Symbols Without Hydrogen Atoms

We have so far considered the notation that is necessary for describing the conformation of all the atoms in the side chains. However, if, as in describing the results of X-ray studies, it is necessary only to refer to the heavy atoms, all the details given in Fig. 7 are not necessary. The

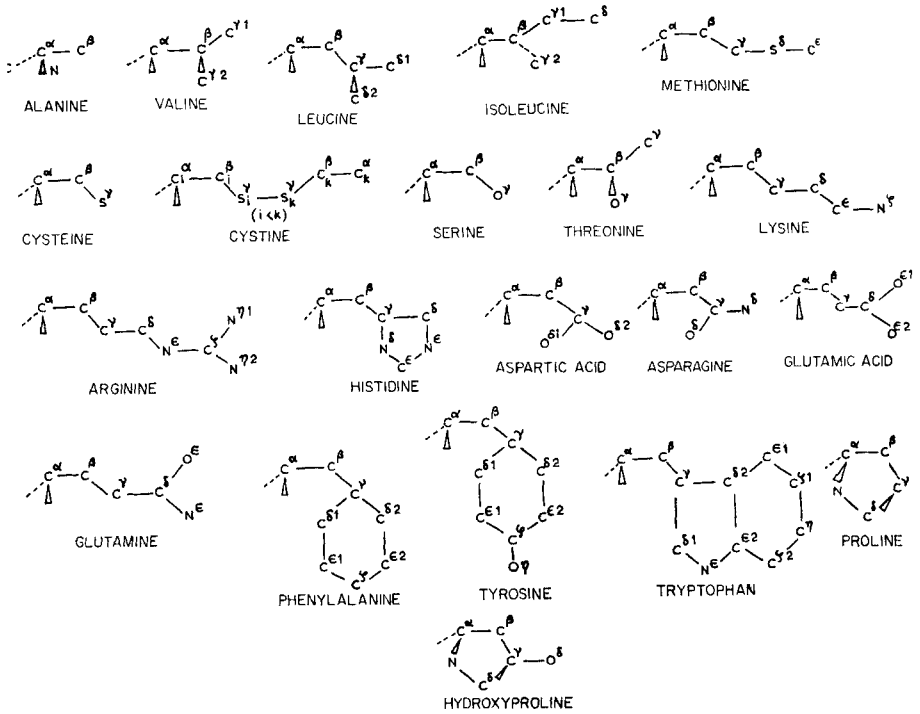
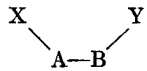


FIG. 9. Simplified notation for the heavy atoms alone in the common amino acid chains. This is a version of Fig. 7 in which the hydrogen atoms are removed and the consequently unnecessary symbols are removed.

minimum that are necessary to avoid ambiguity and which are concordant with the definitions mentioned above are shown in Fig. 9. The numbering in this figure is very closely analogous to that given by Edsall *et al.* (1966a,b,c), in almost all cases, except for minor deviations which have been necessitated by conforming to a consistent set of more generalized definitions adopted here.

6. Side Chain Dihedral Angles

The conformation of the atoms in the side chain may be denoted by a series of dihedral angles χ^1, χ^2, etc., these being rotations about the bonds 1 (C^α—C^β), 2 (C^β—C^γ), etc., from a standard conformation.[4] The standard conformation for any of the χ's to be equal to zero is taken to be that in which the two atoms of the bond, the heavy atom before and the heavy atom ahead are all in a plane and the last two are cis with respect to each other. (In the case of χ^1, the rotation about C^α—C^β, the heavy atom which is taken for reference is N of the same residue, i.e., $\chi^1 = 0$ corresponds to N and C^γ being cis to each other.) As in the case of ϕ, ψ, and ω, the direction of rotation of χ is taken to be positive for a clockwise rotation looking in the direction of progress of a side chain. These conventions, namely that dihedral angles are taken as positive for clockwise rotations and that the cis conformation is taken as corresponding to zero angle, agree with current usage in organic chemistry. [Figures 1 and 2 of the paper by Klyne and Prelog (1960) may be referred to for a clear picture of the angles—Fig. 1 in particular shows that the sense of the dihedral angle is independent of the direction from which system (V) is viewed, whether from A to B or from B to A.]



(V)

We may consider that in the case of the backbone alone, we make an exception to this cis rule for the definition of ϕ, ψ, and ω, because we take the trans peptide unit as the standard and the fully stretched conformation as the standard one corresponding to $\phi = \psi = 0$. These particular conventions are very convenient.

In those cases where more than one bond exists at the same level (e.g., C^β—C^γ in valine), an additional numerical symbol is added following the same conventions for the numbering of the bonds as mentioned above. It will be noticed that, according to our earlier definitions, the value of χ^{jk} will increase with k in the case of identical atoms. Values of χ^{jk} observed in several amino acids are listed later in this review (Section VIII,B,2).

7. Description of D*-Amino Acids and Residues*

It is possible to take over *in toto* all the rules that have been enunciated in Section II,E,6 for D-amino acids. It will then be found that all

[4] We have used superscripts for the different χ's, instead of subscripts as used by Edsall *et al.* (1966a,b,c). This makes the subscript available for indicating the number of the residue to which the side chain belongs.

χ^j and χ^{jk} will be the reverse of the corresponding values for L-amino acids, i.e.,

$$\chi_D{}^{jk} = -\chi_L{}^{jk}$$

Such a simple rule does not hold with the definitions of Edsall *et al.* (1966a,b,c) in the case of amino acids with two asymmetric carbon atoms, such as isoleucine and threonine and their isomers. The branches 21 and 22 will be interchanged in their system for the D-isomer with respect to the L-isomer, while they are unchanged in the present system.

8. Summary

Summarizing what has been discussed in this section, we may point out that the nomenclature of all the atoms in the side chain follows from the following rules.

We have (i) the increasing rule, according to which successive heavy atoms proceeding from C^α are labelled C^β, C^γ, etc. and the successive bonds 1, 2, 3, etc.

When there is branching, we have a second numerical label for the bond and the heavy atom it ends in, and this label increases according to (ii) the clockwise rule if the branches are alike (either two out of three, or all three).

If the branches are dissimilar, we have (iii) the larger rule, according to which the second label is 1 for the one having the largest group, 2 for the next smaller one and 3 for the smallest.

If there are only two branches and they are dissimilar, the above larger rule holds for them, but if they are alike, we have (iv) the cis rule, which makes the succeeding group nearer to the preceding group removed by three bonds from it have the label 1 and the other the label 2.

The hydrogens attached to a heavy atom have the same symbol or symbols as the heavy atom and in addition one more symbol, decided according to the same rules as for heavy atoms—namely the clockwise rule if there are two or three attached to a tetrahedrally bonded atom and the cis rule if there are two in a plane.

It will be seen from the above that the rules are very few and do not have any exceptions.

F. Complete Description of the Backbone and Side Group

We have seen above that the set of dihedral angles (ϕ_i, ψ_i, ω_i) of the backbone and the series of angles $\chi_i{}^{jk}$ for the side groups define the configuration of a polypeptide chain fairly completely. We shall now

consider what further data are necessary for a complete specification of the conformation.

1. Backbone

The peptide unit itself may be fully specified if we are able to give the positions of all the atoms in this unit, in a suitable coordinate system. On the other hand, our attempt has been to describe everything in terms of internal coordinates, i.e., bond lengths and bond angles and dihedral angles in the system. Therefore, it follows that, in this method of specification, one must have in addition to ω, which denotes the nonplanarity of the atoms C_i^{α}—C'_i—N_{i+1}—C_{i+1}^{α}, two other angles which will indicate the deviations from planarity of the atoms O and H. It appears that a self-contained definition of this can be given in terms of two angles, which may be called v^{O} and v^{H} which indicate the dihedral angles of rotation about the peptide bond C'_i—N_{i+1}, describing the nonplanarity of the set of atoms C_{i+1}^{α}—N_{i+1}—C'_i—O_i and C_i^{α}—C'_i—N_{i+1}—H_{i+1} respectively. These angles v^{O} and v^{H} which are defined to be zero for a planar peptide unit (and thus agree with the cis rule) are expected to be small and their values as observed in actual structures are also given in Table III. (The use of the Greek symbol v for these angles is suggested by its similarity to ϕ, ψ, ω, χ at the end of the Greek alphabet.)

Of course, the dimensions of the various bonds and the angles between the bonds have also to be specified if they do not correspond to the standard Pauling-Corey parameters. A suitable notation for the specification of the angles is the use of the symbol τ (as suggested by Edsall *et al.*, 1966a,b,c) e.g., τ(N—C′—O). The bond lengths may be denoted by the symbol l e.g., $l(C^{\alpha}—C')$, $l(C^{\beta}—C^{\gamma})$, etc. In many problems, the angle $\tau(N—C^{\alpha}—C')$ at the α-carbon atom which links two peptide units plays an important part. It is then simply referred to by the symbol τ. So also, when there is no ambiguity, the angle may be specified using only the label of the atom at which it occurs, e.g., $\tau(C^{\alpha})$ or simply as τ^{α}.

2. Side Group

In the case of the side groups, the dihedral angles that are involved in the specification of their conformation have been discussed above. It is only necessary therefore to specify in addition all the values of the bond lengths (l) and bond angles (τ) that are involved in describing the particular side group. This would make the description complete.

3. Attachment of the Side Group to the Backbone

In an ideal case, namely, when the angle N—C^{α}—C′ is equal to the tetrahedral angle 109°28′ and all the four bonds attached to the α-carbon

atoms are tetrahedrally oriented, the values of the dihedral angle of rotation ϕ about N—C^α corresponding to the atoms C′, H^α and C^β will be ϕ, $\phi + 120°$, and $\phi + 240°$ for an L-amino acid residue. We shall denote the value of the dihedral angle specifying H^α and C^β as ϕ^α and ϕ^β respectively (measured from the trans configuration). When, however, the angle at C^α has a value different from the tetrahedral angle, it is necessary to specify, in addition to all the data mentioned earlier for the backbone and all the data mentioned for the side chain, the following data for specification of the atoms C^β and H^α, in order to indicate the relationship of the side chain to the backbone.

One method is to specify all the τ-values at C^α which will immediately fix C^β and H^α when the lengths of the bonds C^α—C^β and C^α—H^α are known. Another method is to specify ϕ^α and ϕ^β and the angles τ(N—C^α—H^α) and τ(N—C^α—C^β). The latter seems to be the one which is preferable and in line with the philosophy of approach to the problem adopted above.

In problems of theoretical computation of conformations when $\tau \neq 109°28'$, it is often necessary to fix a side group to the backbone in order to work out the consequences. The various workers in the field have not published the exact way in which they have done this, and personal discussion has shown that there is a considerable variation in the method adopted. Therefore, the method adopted in our laboratory will be stated here.

The plane containing the α-carbon atom and the side group atoms C^β and H^α is taken to bisect the angle N—C^α—C′, and the bonds C^α—C^β and C^α—H^α are taken to be equally above and below the plane of the

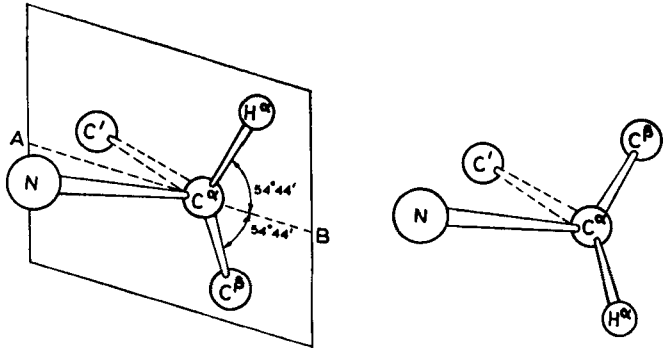


FIG. 10 (left). Diagram showing the attachment of the atoms C^β and H^α to the backbone for L-amino acid residues. The plane containing C^α,C^β, and H^α is shown and it passes through the line AB which bisects the angle N—C^α—C′ and is at right angles to the plane containing N, C^α and C′. (right) Corresponding diagram for D-residues. Note that the two configurations about C^α are related by reflection about the plane N—C^α—C′.

backbone atoms N—C^α—C′ at angles of $\pm 54°44'$. This is indicated in Fig. 10. A calculation shows that, given the angle τ(N—C^α—C′), this procedure closely minimizes the sum of the deviations of all the six angles at C^α from their ideal tetrahedral values. This aspect is mentioned in particular detail because it is one that is generally glossed over in most publications in the field and we felt that we should at least make a positive statement and a probable suggestion, so that, if necessary, this might lead to the suggestion of better methods of obtaining the position of the side group with relation to the backbone.

Figure 10 has been drawn both for an L- and a D-amino acid residue. The C^β atom will be down and the H^α atom up for an L-residue side chain and vice versa for a D-residue. This diagram seems to be a simple way to remember the absolute configuration of L- and D-amino acids. Another method is to note that ϕ, ϕ^α, ϕ^β are in increasing sequence for an L-residue and in decreasing sequence for a D-residue.

4. Description in Terms of Peptide Units

In what has been discussed above, we have used the amino acid residue —NH—CHR—CO— as the unit and denoted it by a subscript i, following Edsall *et al.* (1966a,b,c). However, in most of the theoretical (mathematical) developments, it is found necessary to have a symbol for the peptide unit (see for example Section III). The symbol (i) within brackets is suggested for this. Thus the atoms in the ith peptide unit are

$$—C^\alpha(i)—C'(i)O(i)—N(i)H(i)—C^\alpha(i+1)—$$

where $C^\alpha(i) = C_i^\alpha$. The last condition correlates the numbering in terms of peptide units and in terms of amino acid residues. Thus, ϕ_i, ψ_i are associated with rotations about bonds meeting at the atom $C^\alpha(i)$ and may therefore be labelled $\phi(i)$ and $\psi(i)$. The ith side group is affixed to this α-carbon atom of label (i) and ω_i is an internal rotation within the ith peptide unit, which makes the peptide nonplanar. Therefore, the angles χ_i^{jk} in the notation using residues as links in the chain will be labelled $\chi^{jk}(i)$ when using peptide units for this purpose and the angle $\omega_i = \omega(i)$. Thus, all the dihedral angles ϕ, ψ, ω, and χ could be denoted with labels i either as subscripts or within brackets and they will denote the same angles in either case.

G. Conclusion

We have discussed above the complete set of factors that have to be taken into account for a geometrical description of a polypeptide chain conformation. However, in practice, one need not consider the variations of all of them in order to understand which conformations are

allowed and which are disallowed. The physical criteria which decide whether a conformation is allowed or not are discussed in Section IV. When testing whether or not a particular conformation is allowed by these criteria, it may be possible to restrict appreciably the variables which are to be considered. For instance, planar peptide units are a good enough approximation in most cases involving open polypeptide chains (but not in cyclic peptides). This eliminates the angles ω and υ. Again, as will be seen from later sections, the dihedral angles in the side chain have only a limited range of variation, unlike ϕ and ψ which have much larger ranges of variation. This means that, as far as the side groups are concerned, one can, so to say, choose a finite number of conformations and consider only those conformations which are close to these.

So also, for the bond lengths and bond angles, one would normally take the average values found in various crystal structures. These are found to vary only by a few hundredths of an Ångstrom and by less than 2° or 3° from the mean values. The existence of such variations could be taken account of by suitably altering the conditions which are used for defining an allowed conformation, as shown in Section IV. We would like to mention that the reader need not be puzzled by the large number of parameters that have been specified above in this section. This has been done mainly with the purpose of making the definitions complete and capable of being put to use in a large-scale computer for a very detailed study, if it becomes necessary.

III. Analytical Calculation of Atomic Coordinates

A. Formulae for the Backbone Atoms of a General Polypeptide Chain

The theoretical methods for expressing the effect of the rotations (ϕ, ψ) in a standard coordinate system were first worked out for a pair of peptide units by Ramakrishnan (1964). This has been generalised for a chain of peptide units by later workers. Some details of the general formulae are available in papers by Némethy and Scheraga (1965), and Ramachandran *et al.* (1966b). They all depend on the well-known matrix expression (Jeffreys and Jeffreys, 1950) for effect of a rotation on the coordinates of a point by an angle θ about an axis through the origin having the direction cosines λ, μ, ν.

$$[M]^{\theta}_{\lambda,\mu,\nu} = \begin{bmatrix} \cos\theta + \lambda^2(1-\cos\theta) & \lambda\mu(1-\cos\theta) - \nu\sin\theta & \lambda\nu(1-\cos\theta) + \mu\sin\theta \\ \lambda\mu(1-\cos\theta) + \nu\sin\theta & \cos\theta + \mu^2(1-\cos\theta) & \mu\nu(1-\cos\theta) - \lambda\sin\theta \\ \lambda\nu(1-\cos\theta) - \mu\sin\theta & \mu\nu(1-\cos\theta) + \lambda\sin\theta & \cos\theta + \nu^2(1-\cos\theta) \end{bmatrix} \quad (1)$$

We shall now consider the problem of determining the coordinates o the atoms in a sequence of peptide units specified by (ϕ_i, ψ_i) which is an extension of this procedure (see e.g., Ramachandran *et al.*, 1966b). We shall first consider only the backbone atoms and assume that all $\omega_i = 0$. The method discussed here follows that of Ramachandran *et al.* (1966b), which is closely related to the general procedure suggested by Eyring (1932a) for a polymer chain. Suppose that every peptide unit (i) has a rectangular coordinate system $S(i)$ associated with it,[5] which may be the one shown in Fig. 3. These coordinate systems are then related to one another by affine transformation $T(i, j)$ consisting of both rotations and translations. Formally, if $\mathbf{r}_1(1)$ is the position vector of an atom in the first unit referred to its corresponding system $S(1)$, then the position vector $\mathbf{r}_1(i)$ of the corresponding atom in the ith unit, referred to $S(1)$, is given by

$$\mathbf{r}_1(i) = T(i,1)\mathbf{r}_1(1) \tag{2}$$

The transformation $T(i, 1)$ may also be inversely interpreted as relating the position vector $\mathbf{r}_i(i)$ of an atom in the ith unit referred to its own system $S(i)$ and its position vector $\mathbf{r}_1(i)$ referred to the standard system $S(1)$. The relation is

$$\mathbf{r}_1(i) = T(i,1)\mathbf{r}_i(i) \tag{3}$$

This follows from the fact that, referred to its own coordinate system, the coordinates of an atom in that unit are the same, irrespective of the system of reference, i.e.,

$$\mathbf{r}_1(1) = \mathbf{r}_i(i) \qquad \text{for all } i \tag{4}$$

Affine transformations, in general, have such a double interpretation (see e.g., Birkhoff and MacLane, 1963). Using these interpretations, it is possible to show that, for a transformation corresponding to rotation (ϕ, ψ) at the α-carbon atom C_2^{α},

$$\mathbf{r}_1(2) = T(2,1)\mathbf{r}_2(2) = \mathbf{L} + [\mathrm{R}(2,1)]\mathbf{r}_2(2) \tag{5}$$

where $[\mathrm{R}(2,1)]$ is the product of three matrices of the form

$$[\mathrm{R}(2,1)] = \begin{bmatrix} -\cos\theta & -\sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & -1 \end{bmatrix} [\mathrm{M}^{\phi}{}_{l,m,n}][\mathrm{M}^{\psi}{}_{l',m',n'}] \tag{6}$$

In this, θ is the angle between the positive directions of the y-axes of $S(1)$ and $S(2)$ when $\phi = \psi = 0$ (Fig. 6) and $\mathbf{L}$ is a vector $(0, L, 0)$, where L is the distance between two neighbouring α-carbon atoms in a

[5] We shall use the notation of the peptide unit here and elsewhere in this section, using brackets, because this brings to focus the idea that the dihedral angles of rotation of ϕ and ψ operate on all the atoms of a particular peptide unit.

peptide unit (Fig. 6). With the Pauling-Corey parameters (i.e., for the peptide unit of Table I), $\theta = 144° 36' - \tau(C^\alpha)$ and $L = 3.80$ Å. $[M^{\phi}_{l,m,n}]$ and $[M^{\psi}_{l',m',n'}]$ are matrices of the form given by Eq. (1), in which l, m, n and l', m', n' are respectively the direction cosines of N(1)—$C^\alpha(2)$ and $C^\alpha(2)$—C'(2) in the system $S(1)$ for the standard conformation $\phi = \psi = 0$. [Note that, more specifically, (ϕ, ψ) in Eq. (6) is (ϕ_2, ψ_2)].

More generally, Eq. (5) takes the form

$$\mathbf{r}_i(i+1) = \mathbf{L} + [R(i+1,i)]\mathbf{r}_{i+1}(i+1) \tag{7}$$

where $\mathbf{L}$ is the same as in Eq. (5), and $[R(i+1, i)]$ is of the same form as Eq. (6), but will depend on the parameters (ϕ_{i+1}, ψ_{i+1}). It is to be noted that all the quantities $\theta, L, l, m, n, l', m', n'$ are constants for a given peptide geometry and the only variables are the dihedral angles ϕ_i and ψ_i at each α-carbon atom. Equation (7) can be developed to give the transformation $T(N,1)$ for the coordinates of the atoms in the Nth unit referred to a standard system [namely $S(1)$], when all the (ϕ_i, ψ_i) for $i = 2$ to N are known. The formula is

$$T(N,1) = T(2,1)T(3,2) \ldots T(N-1,N) = \prod_{i=1}^{N-1} T(i+1,i) \tag{8}$$

and we have finally

$$\mathbf{r}_1(N) = T(N,1)\mathbf{r}_N(N) \tag{9a}$$

where $T(N,1)$ is of the form

$$T(N,1) = \mathbf{D}(N) + [P(N)] \tag{9b}$$

where

$$[P(N)] = \prod_{i=1}^{N-1} [R(i+1,i)] \tag{10}$$

and

$$\mathbf{D}(N) = \sum_{i=1}^{N-1} [P(i)]\mathbf{L} \tag{11}$$

Here, Eq. (10) is the rotational part, and Eq. (11), where $[P(1)]$ is a unit matrix, is the translational part of the transformation.

Note that $\mathbf{r}_N(N)$ in Eq. (9a) is equal to $\mathbf{r}_1$ (1) by Eq. (4) and is therefore specified by the peptide geometry which is fed into the calculation.

B. Application to the Side Chain and Effect of Omega Rotation

A slight modification is needed for computing the positions of the atom $H^\alpha(i)$ and the side group atoms starting from $C^\beta(i)$ attached to

the ith C^{α}. These atoms do not partake of the rotation ψ_i and hence the last $[R(i, i-1)]$ becomes $[R'(i, i-1)]$ which has the form of the right-hand side of Eq. (6), but without the last term $[M\psi_{l',m',n'}]$. Once having obtained C^{β}, the succeeding atoms in the side chain can be added and their coordinates determined, by making use of the following procedure.

We shall specifically consider the case of four atoms A, B, C, D linked in a chain A—B—C—D, and describe the procedure for finding the coordinates of the atom D corresponding to a dihedral angle of rotation χ about the bond B—C, given the positions of the atoms A, B, and C, the

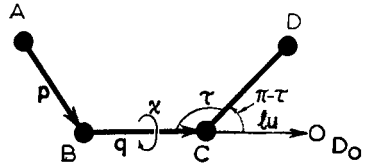


FIG. 11. The relevant parameters involved in the computation of the coordinates of the atom D, given the atoms A, B, and C and the rotation χ about the line BC. The unit vector **n** is normal to the plane of the paper and is pointing upwards.

angle τ(B—C—D) and the length l(C—D). In particular, A, B, C, D may be the atoms N, C^{α}, C^{β} and C^{γ} and this procedure would fix C^{γ} given χ^1. Figure 11 shows the relevant parameters used in the derivation.

Denote, in the coordinate system used, the position vectors of A, B, C, D by $\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3, \mathbf{r}_4$. Let

$$\mathbf{p} = \mathbf{r}_2 - \mathbf{r}_1, \qquad \mathbf{q} = \mathbf{r}_3 - \mathbf{r}_2 \tag{12}$$

Let **n** be a unit vector normal to the plane A—B—C, pointing upwards. Then

$$\mathbf{n} = (\mathbf{p} \times \mathbf{q})/\,\mathbf{p} \times \mathbf{q} \tag{13}$$

Denote unit vector along **q** by **u**, i.e.,

$$\mathbf{u} = \mathbf{q}/\,\mathbf{q} \tag{14}$$

Then, the final position of D is obtained by initially putting it in the position D_0 in the direction BC, such that $CO_0 = l$, i.e.,

$$\overrightarrow{CD_0} = l\mathbf{u} = \mathbf{v} \tag{15}$$

and then (a) rotating it about the vector **n** by an angle $\pi - \tau$ and (b) then rotating it about the direction B—C (i.e., about the vector **u**) by an angle χ. Both the rotations $(\pi - \tau)$ and χ are clockwise looking down the relevant vectors. Following the formulae discussed earlier, this gives Eq. (16).

$$\mathbf{r}_4 = \mathbf{r}_3 + [\mathrm{M}^{\chi_u}][\mathrm{M}^{\tau-\tau_n}]\mathbf{v} \tag{16}$$

In this, we have used the notation $[\mathrm{M}^{\theta}_{\mathbf{a}}]$ for the matrix $[\mathrm{M}^{\theta}_{\lambda,\mu,\nu}]$ of Eq. (1) with **a** as a unit vector in the direction λ, μ, ν.

This procedure can be readily extended to building up further atoms in the side chain, given the relevant l's, τ's, and χ's. In fact, the above formulation is readily extended to the construction of a general polymer chain. A compact formulation of this type has been given by Sugeta and Miyazawa (1967).

So also, the rotations ω_i about the peptide bonds can be introduced if necessary. If ω_i is the same in all the peptide units, it may be convenient to redefine a coordinate system for the distorted peptide unit and determine the coordinates of the backbone atoms in this system. The procedure discussed above can then be carried over *in toto*. (This is convenient, for example, for helical chains with nonplanar peptide units.) If the ω_i's are different, transformations of the type of Eq. (16) will have to be written for every bond.

C. Helical Structures

It is obvious that, in a helical structure, the relationship of one peptide unit to the next will be the same at all α-carbon atoms. This would mean that (ϕ_i, ψ_i) has the same value for all i. The converse of this can also be proved (C. M. Venkatachalam, unpublished), namely that if ϕ_i, ψ_i is same for all i, the resultant structure will take up a helical conformation about some axis in space. About this axis, the helix can be specified by two parameters, the number of units per turn (n) and the resolved height of a unit along the axis of the helix (h). The quantity $1/n = t$ or $360°/n = t°$ gives the angular twist about the axis per unit. The quantities h and t are denoted by the terms unit height and unit twist respectively (Ramachandran, 1960; Edsall *et al.*, 1966a,b,c). Conventionally, h is always taken to be positive. Then a positive value of n (or t) denotes a right-handed helix (i.e., one which twists clockwise looking along the direction of its progress) and a negative value of n (or t) represents a left-handed helix. Following Cahn *et al.* (1966), we shall call these helices "plus" (symbol P) and "minus" (symbol M) respectively. The description of the helices by P and M will agree with the assignment of the sign of n of the helices.

The relation between the pair of helical parameters n and h and the conformational parameters ϕ and ψ can be worked out geometrically for planar peptide units if the backbone angle at the α-carbon atom (τ) is specified. A method of calculating this was given by Ramakrishnan (1964).

In terms of our notation, this can be done as follows. As shown in Section III,A [Eqs. (5) and (6)], the matrix [R(2,1)] which gives the rotation by which unit (2) is obtained from unit (1), can be calculated if we are given ϕ and ψ and the geometry of the peptide unit. Denote the elements of this matrix by $R_{kl}(k, l = 1, 2, 3)$. This matrix must be of the form $[M^t{}_{\lambda,\mu,\nu}]$ where λ, μ, ν are the direction cosines of the helical axis and t is the unit twist of the helix. From Eq. (1), Eq. (17) follows.

$$\begin{aligned} \lambda \sin t &= \tfrac{1}{2}(R_{32} - R_{23}) \\ \mu \sin t &= \tfrac{1}{2}(R_{13} - R_{31}) \\ \nu \sin t &= \tfrac{1}{2}(R_{21} - R_{12}) \end{aligned} \tag{17}$$

Since $\lambda^2 + \mu^2 + \nu^2 = 1$, we get Eq. (18) immediately.

$$\sin t = \pm\tfrac{1}{2}[\Sigma(R_{kl} - R_{lk})^2]^{\frac{1}{2}} \tag{18}$$

Knowing sin t, λ, μ, and ν are determined from Eq. (17). There is an ambiguity of a plus-minus sign in all of the quantities t, λ, μ, ν. (This is equivalent to choosing the helical axis to go in either direction, namely up or down.) Since, by definition, the unit height h is taken to be positive, and taking **L** of Eq. (5) to represent the direction of progress of the helix we have Eq. (19).

$$h = \lambda L_1 + \mu L_2 + \nu L_3 = \text{positive} \tag{19}$$

This fixes the signs of λ, μ, ν and hence of the unit twist t, and of $n = 1/t$.

Thus, the unit height h, unit twist t, and the direction of the helical axis, defined by λ, μ, ν, can all be determined from Eqs. (6) and (5). Having done this, it is possible to go over into cylindrical polar coordinates (r, ϕ, z) (with the helical axis as the z-axis) or to rectangular coordinates, if necessary, referred to the helical axis and calculate the coordinates of the atoms in the chain referred to the appropriate coordinate system.

A diagram showing the variation of helical parameters n and h with ϕ and ψ is given in Ramachandran *et al.* (1963b). Table IV gives summarized data at intervals of 10° for ϕ and ψ for a Pauling-Corey peptide unit and for $\tau = 110°$. Extensive tables of these helical parameters have been prepared (Ramakrishnan, 1965; Ramachandran and Venkatachalam, 1969). A summarized version of the variation of these parameters is given diagrammatically in Fig. 12B (Section V,A,2). Similar data have been calculated in this laboratory for nonplanar peptide units (R. Balasubramanian, unpublished) and for cis peptide units (C. M. Venkatachalam, unpublished).

TABLE IV

Values of Number of Units per Turn (n) and Unit Height (h in Å, Shown in Italics) of Helical Polypeptide Chains Corresponding to Different Values of (ϕ,ψ)[a]

ϕ \ ψ	0	10	20	30	40	50	60	70	80	90	100	110
0	2.00	−2.09	−2.19	−2.29	−2.41	−2.54	−2.68	−2.83	−2.99	−3.17	−3.36	−3.56
	3.63	*3.63*	*3.61*	*3.58*	*3.54*	*3.48*	*3.41*	*3.32*	*3.20*	*3.07*	*2.90*	*2.69*
10	−2.10	−2.20	−2.31	−2.43	−2.56	−2.70	−2.85	−3.01	−3.19	−3.38	−3.58	−3.78
	3.63	*3.62*	*3.60*	*3.58*	*3.53*	*3.47*	*3.40*	*3.30*	*3.17*	*3.01*	*2.82*	*2.58*
20	−2.21	−2.32	−2.44	−2.57	−2.71	−2.87	−3.04	−3.22	−3.41	−3.61	−3.82	−4.03
	3.62	*3.61*	*3.59*	*3.56*	*3.52*	*3.45*	*3.37*	*3.26*	*3.11*	*2.94*	*2.72*	*2.45*
30	−2.33	−2.46	−2.59	−2.73	−2.89	−3.06	−3.25	−3.44	−3.65	−3.86	−4.07	−4.26
	3.60	*3.59*	*3.57*	*3.54*	*3.49*	*3.42*	*3.32*	*3.20*	*3.03*	*2.83*	*2.57*	*2.26*
40	−2.47	−2.60	−2.75	−2.91	−3.09	−3.28	−3.47	−3.69	−3.90	−4.11	−4.31	−4.49
	3.57	*3.56*	*3.54*	*3.51*	*3.45*	*3.37*	*3.26*	*3.11*	*2.92*	*2.68*	*2.39*	*2.04*
50	−2.62	−2.77	−2.93	−3.11	−3.31	−3.51	−3.73	−3.95	−4.17	−4.37	−4.55	−4.69
	3.53	*3.52*	*3.50*	*3.46*	*3.39*	*3.30*	*3.17*	*2.99*	*2.77*	*2.49*	*2.15*	*1.75*
60	−2.79	−2.96	−3.14	−3.34	−3.55	−3.77	−4.00	−4.23	−4.44	−4.63	−4.76	−4.83
	3.48	*3.47*	*3.44*	*3.39*	*3.31*	*3.19*	*3.04*	*2.84*	*2.57*	*2.24*	*1.86*	*1.42*
70	−2.98	−3.16	−3.37	−3.59	−3.82	−4.06	−4.29	−4.56	−4.71	−4.85	−4.92	−4.91
	3.42	*3.40*	*3.36*	*3.29*	*3.20*	*3.06*	*2.87*	*2.63*	*2.31*	*1.94*	*1.51*	*1.05*
80	−3.19	−3.40	−3.62	−3.86	−4.11	−4.36	−4.59	−4.80	−4.94	−5.07	−5.01	−4.92
	3.34	*3.31*	*3.25*	*3.17*	*3.05*	*2.87*	*2.64*	*2.35*	*1.99*	*1.57*	*1.11*	*0.64*
90	−3.43	−3.66	−3.91	−4.17	−4.43	−4.67	−4.89	−5.04	−5.12	−5.11	−5.01	−4.86
	3.23	*3.19*	*3.12*	*3.01*	*2.85*	*2.63*	*2.35*	*2.00*	*1.60*	*1.15*	*0.69*	*0.23*
100	−3.69	−3.94	−4.21	−4.49	−4.75	−4.98	−5.14	−5.23	−5.22	−5.12	−4.95	4.71
	3.10	*3.04*	*2.94*	*2.79*	*2.59*	*2.32*	*1.99*	*1.59*	*1.16*	*0.70*	*0.25*	*0.18*
110	−3.98	−4.26	−4.54	−4.82	−5.07	−5.25	−5.34	−5.34	−5.23	−5.03	4.78	4.52
	2.93	*2.84*	*2.70*	*2.51*	*2.26*	*1.94*	*1.55*	*1.13*	*0.68*	*0.24*	*0.17*	*0.55*
120	−4.30	−4.59	−4.88	−5.14	−5.35	−5.50	−5.50	−5.33	−5.14	4.89	4.59	4.29
	2.72	*2.59*	*2.40*	*2.16*	*1.85*	*1.47*	*1.07*	*0.62*	*0.19*	*0.21*	*0.57*	*0.89*

130	−4.64	−4.94	−5.22	−5.44	−5.57	−5.56	−5.46	−5.24	4.96	4.67	4.36	4.06
	2.45	*2.27*	*2.03*	*1.72*	*1.36*	*0.95*	*0.52*	*0.11*	*0.28*	*0.63*	*0.93*	*1.19*
140	−4.98	−5.26	−5.49	−5.63	−5.67	−5.56	5.34	5.06	4.75	4.42	4.11	3.82
	2.10	*1.87*	*1.56*	*1.21*	*0.81*	*0.39*	*0.01*	*0.38*	*0.71*	*1.00*	*1.24*	*1.44*
150	−5.30	−5.56	−5.70	−5.72	−5.63	5.42	5.14	4.82	4.49	4.16	3.86	3.59
	1.68	*1.39*	*1.03*	*0.64*	*0.23*	*0.16*	*0.52*	*0.83*	*1.10*	*1.33*	*1.51*	*1.67*
160	−5.61	−5.72	−5.77	−5.71	5.50	5.22	4.88	4.55	4.21	3.91	3.62	3.37
	1.18	*0.83*	*0.44*	*0.04*	*0.34*	*0.68*	*0.98*	*1.23*	*1.44*	*1.61*	*1.75*	*1.86*
170	−5.78	−5.84	5.80	5.55	5.25	4.93	4.60	4.26	3.95	3.66	3.40	3.17
	0.61	*0.23*	*0.16*	*0.54*	*0.86*	*1.15*	*1.38*	*1.57*	*1.72*	*1.84*	*1.94*	*2.02*
180	5.85	5.82	5.62	5.32	4.97	4.64	4.30	3.98	3.65	3.43	3.19	2.98
	0.00	*0.39*	*0.75*	*1.06*	*1.33*	*1.55*	*1.72*	*1.86*	*1.97*	*2.05*	*2.11*	*2.17*
190	5.78	5.60	5.33	5.01	4.66	4.32	4.01	3.71	3.45	3.21	3.00	2.81
	0.61	*0.97*	*1.28*	*1.53*	*1.73*	*1.89*	*2.01*	*2.10*	*2.17*	*2.22*	*2.26*	*2.29*
200	5.61	5.35	5.02	4.67	4.34	4.02	3.73	3.46	3.22	3.01	2.82	2.65
	1.18	*1.49*	*1.73*	*1.92*	*2.06*	*2.17*	*2.25*	*2.31*	*2.34*	*2.37*	*2.39*	*2.39*
210	5.30	5.00	4.68	4.35	4.03	3.74	3.48	3.24	3.02	2.83	2.66	2.51
	1.68	*1.92*	*2.11*	*2.24*	*2.34*	*2.40*	*2.45*	*2.48*	*2.49*	*2.50*	*2.49*	*2.48*
220	4.98	4.66	4.34	4.03	3.74	3.48	3.25	3.03	2.84	2.67	2.52	2.38
	2.10	*2.28*	*2.41*	*2.50*	*2.56*	*2.60*	*2.62*	*2.62*	*2.62*	*2.61*	*2.59*	*2.56*
230	4.64	4.33	4.02	3.73	3.48	3.25	3.04	2.85	2.68	2.52	2.38	2.26
	2.45	*2.57*	*2.66*	*2.72*	*2.75*	*2.76*	*2.76*	*2.75*	*2.73*	*2.70*	*2.67*	*2.63*
240	4.30	4.00	3.73	3.48	3.24	3.04	2.85	2.68	2.53	2.39	2.26	2.15
	2.72	*2.80*	*2.86*	*2.89*	*2.90*	*2.91*	*2.88*	*2.85*	*2.82*	*2.78*	*2.74*	*2.69*
250	3.98	3.72	3.47	3.24	3.03	2.85	2.67	2.53	2.39	2.27	2.16	2.05
	2.93	*2.99*	*3.02*	*3.03*	*3.03*	*3.01*	*2.98*	*2.95*	*2.90*	*2.85*	*2.80*	*2.74*
260	3.69	3.45	3.22	3.02	2.84	2.68	2.53	2.39	2.27	2.16	2.06	−2.04
	3.10	*3.14*	*3.15*	*3.15*	*3.13*	*3.10*	*3.07*	*3.02*	*2.97*	*2.91*	*2.85*	*2.78*

[a] The data correspond to the dimensions of Fig. 2 and Table I for the peptide unit and a value of 110° for the angle τ(N—C—C′). (From Ramakrishnan, 1965.)

TABLE IV (Continued)

Values of Number of Units per Turn (n) and Unit Height (h in Å, Shown in Italics) of Helical Polypeptide Chains Corresponding to Different Values of (φ,ψ)[a]

φ\ψ	0	10	20	30	40	50	60	70	80	90	100	110
270	3.43	3.21	3.01	2.83	2.67	2.52	2.39	2.27	2.16	2.06	−2.04	−2.14
	3.23	*3.25*	*3.25*	*3.24*	*3.21*	*3.18*	*3.14*	*3.09*	*3.03*	*2.97*	*2.90*	*2.82*
280	3.19	3.00	2.82	2.66	2.52	2.38	2.26	2.16	2.06	−2.04	−2.14	−2.24
	3.34	*3.34*	*3.34*	*3.32*	*3.28*	*3.24*	*3.20*	*3.14*	*3.08*	*3.01*	*2.93*	*2.85*
290	2.98	2.81	2.65	2.51	2.38	2.26	2.15	2.05	−2.04	−2.13	−2.24	−2.36
	3.42	*3.42*	*3.41*	*3.38*	*3.34*	*3.30*	*3.26*	*3.20*	*3.12*	*3.05*	*2.96*	*2.87*
300	2.79	2.64	2.50	2.37	2.25	2.14	2.05	−2.04	−2.14	−2.24	−2.36	−2.49
	3.48	*3.48*	*3.46*	*3.44*	*3.40*	*3.35*	*3.30*	*3.24*	*3.17*	*3.08*	*2.99*	*2.88*
310	2.62	2.48	2.36	2.24	2.14	2.04	−2.05	−2.14	−2.25	−2.36	−2.49	−2.63
	3.53	*3.52*	*3.51*	*3.48*	*3.43*	*3.39*	*3.34*	*3.27*	*3.19*	*3.10*	*3.01*	*2.89*
320	2.47	2.35	2.23	2.13	2.04	−2.05	−2.15	−2.26	−2.37	−2.50	−2.63	−2.78
	3.57	*3.56*	*3.55*	*3.52*	*3.48*	*3.43*	*3.37*	*3.30*	*3.22*	*3.12*	*3.01*	*2.88*
330	2.33	2.22	2.12	2.03	−2.06	−2.16	−2.26	−2.38	−2.50	−2.64	−2.79	−2.95
	3.60	*3.59*	*3.57*	*3.54*	*3.50*	*3.45*	*3.39*	*3.32*	*3.23*	*3.13*	*3.00*	*2.86*
340	2.21	2.11	2.02	−2.07	−2.17	−2.27	−2.39	−2.51	−2.65	−2.80	−2.96	−3.14
	3.62	*3.61*	*3.59*	*3.56*	*3.52*	*3.47*	*3.40*	*3.33*	*3.24*	*3.12*	*2.98*	*2.81*
350	2.10	2.02	−2.08	−2.18	−2.28	−2.40	−2.53	−2.66	−2.81	−2.98	−3.15	−3.34
	3.63	*3.62*	*3.60*	*3.57*	*3.53*	*3.48*	*3.41*	*3.33*	*3.23*	*3.10*	*2.95*	*2.77*

φ\ψ	120	130	140	150	160	170	180	190	200	210	220	230
0	−3.76	−3.96	−4.16	−4.33	−4.47	−4.55	4.59	4.55	4.47	4.33	4.16	3.96
	2.45	*2.15*	*1.81*	*1.42*	*0.97*	*0.50*	*0.00*	*0.50*	*0.97*	*1.42*	*1.81*	*2.15*
10	−3.99	−4.19	−4.39	−4.49	−4.55	−4.59	4.56	4.46	4.32	4.14	3.95	3.74
	2.31	*1.97*	*1.59*	*1.15*	*0.68*	*0.18*	*0.32*	*0.80*	*1.24*	*1.64*	*1.99*	*2.30*
20	−4.22	−4.38	−4.52	−4.59	−4.60	4.57	4.47	4.32	4.13	3.93	3.73	3.52
	2.13	*1.75*	*1.32*	*0.85*	*0.36*	*0.14*	*0.62*	*1.07*	*1.47*	*1.83*	*2.14*	*2.41*

30	−4.43	−4.56	−4.63	−4.65	−4.60	4.45	4.33	4.14	3.94	3.73	3.52	3.32
	1.90	*1.48*	*1.02*	*0.53*	*0.03*	*0.45*	*0.90*	*1.31*	*1.67*	*1.99*	*2.26*	*2.49*
40	−4.61	−4.68	−4.70	−4.63	4.52	4.35	4.16	3.95	3.73	3.52	3.32	3.12
	1.63	*1.17*	*0.69*	*0.20*	*0.29*	*0.74*	*1.15*	*1.51*	*1.83*	*2.11*	*2.35*	*2.55*
50	−4.75	−4.75	−4.69	4.56	4.39	4.19	3.97	3.74	3.52	3.32	3.12	2.94
	1.31	*0.83*	*0.34*	*0.14*	*0.59*	*1.00*	*1.37*	*1.69*	*1.96*	*2.21*	*2.41*	*2.59*
60	−4.83	−4.75	4.62	4.44	4.22	3.99	3.76	3.54	3.33	3.13	2.95	2.78
	0.95	*0.47*	*0.00*	*0.46*	*0.87*	*1.23*	*1.58*	*1.83*	*2.07*	*2.28*	*2.46*	*2.62*
70	−4.83	−4.69	4.48	4.26	4.02	3.79	3.56	3.34	3.14	2.94	2.78	2.63
	0.57	*0.10*	*0.35*	*0.75*	*1.11*	*1.43*	*1.71*	*1.95*	*2.16*	*2.34*	*2.50*	*2.64*
80	−4.77	4.55	4.31	4.07	3.82	3.58	3.36	3.15	2.96	2.79	2.63	2.49
	0.18	*0.26*	*0.66*	*1.02*	*1.33*	*1.60*	*1.84*	*2.04*	*2.22*	*2.38*	*2.52*	*2.64*
90	4.64	4.38	4.11	3.86	3.61	3.38	3.17	2.98	2.80	2.64	2.50	2.36
	0.21	*0.59*	*0.94*	*1.25*	*1.51*	*1.74*	*1.94*	*2.12*	*2.27*	*2.41*	*2.53*	*2.64*
100	4.45	4.17	3.90	3.65	3.41	3.19	2.99	2.81	2.65	2.50	2.37	2.25
	0.55	*0.89*	*1.19*	*1.45*	*1.67*	*1.86*	*2.03*	*2.18*	*2.31*	*2.43*	*2.53*	*2.62*
110	4.23	3.95	3.69	3.44	3.22	3.01	2.83	2.66	2.51	2.38	2.26	2.14
	0.88	*1.16*	*1.41*	*1.62*	*1.80*	*1.96*	*2.11*	*2.23*	*2.34*	*2.44*	*2.52*	*2.60*
120	4.00	3.73	3.48	3.24	3.04	2.85	2.68	2.53	2.39	2.26	2.15	2.05
	1.16	*1.39*	*1.59*	*1.77*	*1.92*	*2.05*	*2.16*	*2.26*	*2.35*	*2.44*	*2.51*	*2.57*
130	3.77	3.51	3.27	3.06	2.87	2.70	2.54	2.40	2.27	2.16	2.05	−2.04
	1.40	*1.59*	*1.75*	*1.89*	*2.01*	*2.11*	*2.21*	*2.29*	*2.36*	*2.43*	*2.49*	*2.54*
140	3.55	3.31	3.09	2.89	2.71	2.55	2.41	2.28	2.17	2.06	−2.04	−2.14
	1.62	*1.76*	*1.89*	*2.00*	*2.09*	*2.17*	*2.24*	*2.31*	*2.36*	*2.41*	*2.46*	*2.50*
150	3.34	3.11	2.91	2.73	2.57	2.43	2.30	2.18	2.07	−2.03	−2.13	−2.24
	1.80	*1.91*	*2.01*	*2.09*	*2.16*	*2.22*	*2.27*	*2.32*	*2.36*	*2.39*	*2.42*	*2.45*
160	3.14	2.93	2.75	2.59	2.44	2.31	2.19	2.08	−2.02	−2.10	−2.23	−2.36
	1.96	*2.04*	*2.10*	*2.16*	*2.21*	*2.25*	*2.29*	*2.32*	*2.34*	*2.36*	*2.37*	*2.39*
170	2.96	2.77	2.60	2.45	2.32	2.20	2.09	−2.01	−2.11	−2.22	−2.35	−2.48
	2.09	*2.15*	*2.19*	*2.23*	*2.26*	*2.28*	*2.30*	*2.31*	*2.32*	*2.32*	*2.32*	*2.31*
180	2.79	2.62	2.47	2.33	2.21	2.10	−2.00	−2.10	−2.21	−2.33	−2.47	−2.62
	2.21	*2.24*	*2.26*	*2.28*	*2.29*	*2.29*	*2.30*	*2.29*	*2.29*	*2.28*	*2.26*	*2.24*

TABLE IV (Continued)

Values of Number of Units per Turn (n) and Unit Height (h in Å, Shown in Italics) of Helical Polypeptide Chains Corresponding to Different Values of (φ,ψ)[a]

φ\ψ	120	130	140	150	160	170	180	190	200	210	220	230
190	2.64	2.48	2.35	2.22	2.11	2.01	−2.09	−2.20	−2.32	−2.45	−2.60	−2.77
	2.31	*2.31*	*2.32*	*2.32*	*2.32*	*2.31*	*2.30*	*2.28*	*2.26*	*2.23*	*2.19*	*2.15*
200	2.50	2.36	2.23	2.10	2.02	−2.08	−2.19	−2.31	−2.44	−2.59	−2.75	−2.93
	2.39	*2.39*	*2.37*	*2.36*	*2.34*	*2.32*	*2.29*	*2.25*	*2.21*	*2.16*	*2.10*	*2.04*
210	2.37	2.24	2.13	2.03	−2.07	−2.18	−2.30	−2.43	−2.57	−2.73	−2.91	−3.11
	2.47	*2.45*	*2.42*	*2.39*	*2.36*	*2.32*	*2.27*	*2.22*	*2.16*	*2.09*	*2.01*	*1.91*
220	2.25	2.14	2.04	−2.06	−2.17	−2.28	−2.41	−2.55	−2.71	−2.89	−3.09	−3.31
	2.53	*2.50*	*2.46*	*2.41*	*2.36*	*2.31*	*2.24*	*2.17*	*2.09*	*2.00*	*1.89*	*1.76*
230	2.15	2.04	−2.05	−2.16	−2.27	−2.40	−2.54	−2.70	−2.87	−3.06	−3.27	−3.51
	2.59	*2.54*	*2.49*	*2.43*	*2.36*	*2.29*	*2.21*	*2.11*	*2.01*	*1.89*	*1.75*	*1.59*
240	2.05	−2.05	−2.15	−2.26	−2.39	−2.53	−2.68	−2.85	−3.04	−3.24	−3.48	−3.73
	2.63	*2.57*	*2.51*	*2.44*	*2.35*	*2.26*	*2.16*	*2.05*	*1.92*	*1.77*	*1.59*	*1.39*
250	−2.04	−2.14	−2.26	−2.38	−2.51	−2.66	−2.83	−3.01	−3.22	−3.44	−3.69	−3.95
	2.68	*2.60*	*2.52*	*2.44*	*2.34*	*2.23*	*2.11*	*1.96*	*1.80*	*1.62*	*1.41*	*1.16*
260	−2.14	−2.25	−2.37	−2.50	−2.65	−2.81	−2.99	−3.19	−3.41	−3.65	−3.90	−4.17
	2.71	*2.62*	*2.53*	*2.43*	*2.31*	*2.18*	*2.03*	*1.86*	*1.67*	*1.45*	*1.19*	*0.89*
270	−2.25	−2.36	−2.50	−2.64	−2.80	−2.98	−3.17	−3.38	−3.61	−3.86	−4.11	−4.38
	2.73	*2.64*	*2.53*	*2.41*	*2.27*	*2.12*	*1.94*	*1.74*	*1.51*	*1.25*	*0.94*	*0.59*
280	−2.36	−2.49	−2.63	−2.79	−2.96	−3.15	−3.36	−3.58	−3.82	−4.07	−4.31	−4.55
	2.75	*2.64*	*2.52*	*2.38*	*2.22*	*2.04*	*1.84*	*1.60*	*1.33*	*1.02*	*0.66*	*0.26*
290	−2.49	−2.63	−2.78	−2.94	−3.14	−3.34	−3.56	−3.79	−4.02	−4.26	−4.48	4.69
	2.76	*2.64*	*2.50*	*2.34*	*2.16*	*1.95*	*1.71*	*1.43*	*1.11*	*0.75*	*0.35*	*0.10*
300	−2.63	−2.78	−2.95	−3.13	−3.33	−3.54	−3.76	−3.99	−4.22	−4.44	−4.62	4.75
	2.76	*2.62*	*2.46*	*2.28*	*2.07*	*1.83*	*1.58*	*1.23*	*0.87*	*0.46*	*0.00*	*0.47*
310	−2.78	−2.94	−3.12	−3.32	−3.52	−3.74	−3.97	−4.19	−4.39	−4.56	4.69	4.75
	2.75	*2.59*	*2.41*	*2.21*	*1.96*	*1.69*	*1.37*	*1.00*	*0.59*	*0.14*	*0.34*	*0.83*

320	−2.95	−3.12	−3.32	−3.52	−3.73	−3.95	−4.16	−4.35	−4.52	4.63	4.70	4.68
	2.72	*2.55*	*2.35*	*2.11*	*1.83*	*1.51*	*1.15*	*0.74*	*0.29*	*0.20*	*0.69*	*1.17*
330	−3.13	−3.32	−3.52	−3.73	−3.94	−4.14	−4.33	−4.45	4.60	4.65	4.63	4.56
	2.69	*2.49*	*2.26*	*1.99*	*1.67*	*1.31*	*0.90*	*0.45*	*0.03*	*0.53*	*1.02*	*1.48*
340	−3.32	−3.52	−3.73	−3.93	−4.13	−4.32	−4.47	−4.57	4.60	4.59	4.52	4.38
	2.63	*2.41*	*2.14*	*1.83*	*1.47*	*1.07*	*0.62*	*0.14*	*0.36*	*0.85*	*1.32*	*1.75*
350	−3.53	−3.74	−3.95	−4.14	−4.32	−4.46	−4.56	4.59	4.55	4.49	4.39	4.19
	2.56	*2.30*	*1.99*	*1.64*	*1.24*	*0.80*	*0.32*	*0.18*	*0.68*	*1.15*	*1.59*	*1.97*

$\phi \backslash \psi$	240	250	260	270	280	290	300	310	320	330	340	350
0	3.76	3.56	3.36	3.17	2.99	2.83	2.68	2.54	2.41	2.29	2.19	2.09
	2.45	*2.69*	*2.90*	*3.07*	*3.20*	*3.32*	*3.41*	*3.48*	*3.54*	*3.58*	*3.61*	*3.63*
10	3.53	3.34	3.15	2.98	2.81	2.66	2.53	2.40	2.28	2.18	2.08	−2.02
	2.56	*2.77*	*2.95*	*3.10*	*3.23*	*3.33*	*3.41*	*3.48*	*3.53*	*3.57*	*3.60*	*3.62*
20	3.32	3.14	2.96	2.80	2.65	2.51	2.39	2.27	2.17	2.07	−2.02	−2.11
	2.63	*2.81*	*2.98*	*3.12*	*3.24*	*3.33*	*3.40*	*3.47*	*3.52*	*3.56*	*3.59*	*3.61*
30	3.13	2.95	2.79	2.64	2.50	2.38	2.26	2.16	2.06	−2.03	−2.12	−2.22
	2.69	*2.86*	*3.00*	*3.13*	*3.23*	*3.32*	*3.39*	*3.45*	*3.50*	*3.54*	*3.57*	*3.59*
40	2.95	2.78	2.63	2.50	2.37	2.26	2.15	2.05	−2.04	−2.13	−2.23	−2.35
	2.72	*2.88*	*3.01*	*3.12*	*3.22*	*3.30*	*3.37*	*3.43*	*3.48*	*3.52*	*3.55*	*3.56*
50	2.78	2.63	2.49	2.36	2.25	2.14	2.05	−2.04	−2.14	−2.24	−2.36	−2.48
	2.75	*2.89*	*3.01*	*3.10*	*3.19*	*3.27*	*3.34*	*3.39*	*3.43*	*3.48*	*3.51*	*3.52*
60	2.63	2.49	2.36	2.24	2.14	2.04	−2.05	−2.14	−2.25	−2.37	−2.50	−2.64
	2.76	*2.88*	*2.99*	*3.08*	*3.17*	*3.24*	*3.30*	*3.35*	*3.40*	*3.44*	*3.46*	*3.48*
70	2.49	2.36	2.24	2.13	2.04	−2.05	−2.15	−2.26	−2.38	−2.51	−2.65	−2.81
	2.76	*2.87*	*2.96*	*3.05*	*3.12*	*3.20*	*3.26*	*3.30*	*3.34*	*3.38*	*3.41*	*3.42*
80	2.36	2.24	2.14	2.04	−2.06	−2.16	−2.26	−2.38	−2.52	−2.66	−2.82	−3.00
	2.75	*2.85*	*2.93*	*3.01*	*3.08*	*3.14*	*3.20*	*3.24*	*3.28*	*3.32*	*3.34*	*3.34*
90	2.25	2.14	2.04	−2.06	−2.16	−2.27	−2.39	−2.52	−2.67	−2.83	−3.01	−3.21
	2.73	*2.82*	*2.90*	*2.97*	*3.03*	*3.09*	*3.14*	*3.18*	*3.21*	*3.24*	*3.25*	*3.25*
100	2.14	2.04	−2.06	−2.16	−2.27	−2.39	−2.53	−2.68	−2.84	−3.02	−3.22	−3.45
	2.71	*2.78*	*2.85*	*2.91*	*2.97*	*3.02*	*3.07*	*3.10*	*3.13*	*3.15*	*3.15*	*3.14*

TABLE IV (Continued)

Values of Number of Units per Turn (n) and Unit Height (h in Å, Shown in Italics) of Helical Polypeptide Chains Corresponding to Different Values of (ϕ,ψ)[a]

ϕ\ψ	240	250	260	270	280	290	300	310	320	330	340	350
110	2.04	−2.05	−2.16	−2.27	−2.39	−2.53	−2.67	−2.85	−3.03	−3.24	−3.47	−3.72
	2.68	*2.74*	*2.80*	*2.85*	*2.90*	*2.95*	*2.98*	*3.01*	*3.03*	*3.03*	*3.02*	*2.99*
120	−2.05	−2.15	−2.26	−2.39	−2.53	−2.68	−2.85	−3.04	−3.24	−3.48	−3.73	−4.00
	2.63	*2.69*	*2.74*	*2.78*	*2.82*	*2.85*	*2.88*	*2.91*	*2.90*	*2.89*	*2.86*	*2.80*
130	−2.15	−2.26	−2.38	−2.52	−2.68	−2.85	−3.04	−3.25	−3.48	−3.73	−4.02	−4.33
	2.59	*2.63*	*2.67*	*2.70*	*2.73*	*2.75*	*2.76*	*2.76*	*2.75*	*2.72*	*2.66*	*2.57*
140	−2.25	−2.38	−2.52	−2.67	−2.84	−3.03	−3.25	−3.48	−3.74	−4.03	−4.34	−4.66
	2.53	*2.56*	*2.59*	*2.61*	*2.62*	*2.62*	*2.62*	*2.60*	*2.56*	*2.50*	*2.41*	*2.28*
150	−2.37	−2.51	−2.66	−2.83	−3.02	−3.24	−3.48	−3.74	−4.03	−4.35	−4.68	−5.00
	2.47	*2.48*	*2.49*	*2.50*	*2.49*	*2.48*	*2.45*	*2.40*	*2.34*	*2.24*	*2.11*	*1.92*
160	−2.50	−2.65	−2.82	−3.01	−3.22	−3.46	−3.73	−4.02	−4.34	−4.67	−5.02	−5.35
	2.39	*2.39*	*2.39*	*2.37*	*2.34*	*2.31*	*2.25*	*2.17*	*2.06*	*1.92*	*1.73*	*1.49*
170	−2.64	−2.81	−3.00	−3.21	−3.45	−3.71	−4.01	−4.32	−4.66	−5.01	−5.33	−5.60
	2.31	*2.29*	*2.26*	*2.22*	*2.17*	*2.10*	*2.01*	*1.89*	*1.73*	*1.53*	*1.28*	*0.97*
180	−2.79	−2.98	−3.19	−3.43	−3.65	−3.98	−4.30	−4.64	−4.97	−5.32	−5.62	−5.82
	2.21	*2.17*	*2.11*	*2.05*	*1.97*	*1.86*	*1.72*	*1.55*	*1.33*	*1.06*	*0.75*	*0.39*
190	−2.96	−3.17	−3.40	−3.66	−3.95	−4.26	−4.60	−4.93	−5.25	−5.55	−5.80	5.84
	2.09	*2.02*	*1.94*	*1.84*	*1.72*	*1.57*	*1.38*	*1.15*	*0.86*	*0.54*	*0.16*	*0.23*
200	−3.14	−3.37	−3.62	−3.91	−4.21	−4.55	−4.88	−5.22	−5.50	5.71	5.77	5.72
	1.96	*1.86*	*1.75*	*1.61*	*1.44*	*1.23*	*0.98*	*0.68*	*0.34*	*0.04*	*0.44*	*0.83*
210	−3.34	−3.59	−3.86	−4.16	−4.49	−4.82	−5.14	−5.42	5.63	5.72	5.70	5.56
	1.80	*1.67*	*1.51*	*1.33*	*1.10*	*0.83*	*0.52*	*0.16*	*0.23*	*0.64*	*1.03*	*1.39*

220	−3.55	−3.82	−4.11	−4.42	−4.75	−5.06	−5.34	5.56	5.67	5.63	5.49	5.26
	1.62	*1.44*	*1.24*	*1.00*	*0.71*	*0.38*	*0.01*	*0.39*	*0.81*	*1.21*	*1.56*	*1.87*
230	−3.77	−4.06	−4.36	−4.67	−4.96	5.24	5.46	5.56	5.57	5.44	5.22	4.94
	1.40	*1.19*	*0.93*	*0.63*	*0.28*	*0.11*	*0.52*	*0.95*	*1.36*	*1.72*	*2.03*	*2.27*
240	−4.00	−4.29	−4.59	−4.89	5.14	5.33	5.50	5.50	5.35	5.14	4.88	4.59
	1.16	*0.89*	*0.57*	*0.21*	*0.19*	*0.62*	*1.07*	*1.47*	*1.85*	*2.16*	*2.40*	*2.59*
250	−4.23	−4.52	−4.78	5.03	5.23	5.34	5.34	5.25	5.07	4.82	4.54	4.26
	0.88	*0.55*	*0.17*	*0.24*	*0.68*	*1.13*	*1.55*	*1.94*	*2.26*	*2.51*	*2.70*	*2.84*
260	−4.45	−4.71	4.95	5.12	5.22	5.23	5.14	4.98	4.75	4.49	4.21	3.94
	0.55	*0.18*	*0.25*	*0.70*	*1.16*	*1.59*	*1.99*	*2.32*	*2.59*	*2.79*	*2.94*	*3.04*
270	−4.64	4.86	5.01	5.11	5.12	5.04	4.89	4.67	4.43	4.17	3.91	3.66
	0.21	*0.23*	*0.69*	*1.15*	*1.60*	*2.00*	*2.35*	*2.63*	*2.85*	*3.01*	*3.12*	*3.19*
280	4.77	4.92	5.01	5.07	4.94	4.80	4.59	4.36	4.11	3.86	3.62	3.40
	0.18	*0.64*	*1.11*	*1.57*	*1.99*	*2.35*	*2.64*	*2.87*	*3.05*	*3.17*	*3.25*	*3.31*
290	4.83	4.91	4.92	4.85	4.71	4.56	4.29	4.06	3.82	3.59	3.37	3.16
	0.57	*1.05*	*1.51*	*1.94*	*2.31*	*2.63*	*2.87*	*3.06*	*3.20*	*3.29*	*3.36*	*3.40*
300	4.83	4.83	4.76	4.63	4.44	4.23	4.00	3.77	3.55	3.34	3.14	2.96
	0.95	*1.42*	*1.86*	*2.24*	*2.57*	*2.84*	*3.04*	*3.19*	*3.31*	*3.39*	*3.44*	*3.47*
310	4.75	4.69	4.55	4.37	4.17	3.95	3.73	3.51	3.31	3.11	2.93	2.77
	1.31	*1.75*	*2.15*	*2.49*	*2.77*	*2.99*	*3.17*	*3.30*	*3.39*	*3.46*	*3.50*	*3.52*
320	4.61	4.49	4.31	4.11	3.90	3.69	3.47	3.28	3.09	2.91	2.75	2.60
	1.63	*2.04*	*2.39*	*2.68*	*2.92*	*3.11*	*3.26*	*3.37*	*3.45*	*3.51*	*3.54*	*3.56*
330	4.43	4.26	4.07	3.86	3.65	3.44	3.25	3.06	2.89	2.73	2.59	2.46
	1.90	*2.26*	*2.57*	*2.83*	*3.03*	*3.20*	*3.32*	*3.42*	*3.49*	*3.54*	*3.57*	*3.59*
340	4.22	4.03	3.82	3.61	3.41	3.22	3.04	2.87	2.71	2.57	2.44	2.32
	2.13	*2.45*	*2.72*	*2.94*	*3.11*	*3.26*	*3.37*	*3.45*	*3.52*	*3.56*	*3.59*	*3.61*
350	3.99	3.78	3.58	3.38	3.19	3.01	2.85	2.70	2.56	2.43	2.31	2.20
	2.31	*2.58*	*2.82*	*3.01*	*3.17*	*3.30*	*3.40*	*3.47*	*3.53*	*3.58*	*3.60*	*3.62*

It is well known that a number of helical structures occur in proteins and in polypeptides (see e.g., reviews by Dickerson, 1964; Schellman and Schellman, 1964). A brief account of this is given in Section IV.

IV. Model Building of Polypeptide Structures

Integral and Nonintegral Helices

Although the formal theories of the type mentioned in the last section connecting the structural properties of a polypeptide helix with the conformation for a pair of peptide units were developed only recently, a number of structures had been predicted to occur in proteins and polypeptides, beginning in the 1940's. A good review of these has been given in several articles published earlier (see e.g., the one by Dickerson, 1964). We shall not consider these in detail except briefly to mention the properties of a few of these, which will be useful in connection with the discussion of the more exhaustive studies made recently.

Starting from the early investigations of Astbury (1940), Huggins (1943), and Bragg *et al.* (1950), who restricted themselves to helices with integral number of residues per turn, a real breakthrough in this field may be said to have come with the proposal of the now famous α-helix (Pauling *et al.*, 1951). Making use of the standard (Pauling-Corey) dimensions of the peptide unit, Pauling and co-workers studied the possible helical polypeptide structures in which there exist a maximum number of hydrogen bonds between different units in the chain. If the stereochemistry is maintained rigorously, they found that they could get a stable conformation in which there are about 3.6 residues per turn and the NH of a residue was hydrogen bonded to the carbonyl oxygen of the residue five residues behind it. This postulation of a nonintegral number of 3.6 residues per turn was in contrast with the proposal of earlier workers who assumed that n should be 2, 3, 4 or 6. A similar analysis also led to the suggestion of another helix, namely the γ-helix, in which there is a hydrogen bond between an NH and a CO of a residue five residues *ahead*. The structural parameters as well as the dihedral angles of these helices are given in Table V which also contains similar data on a number of other postulated structures which we shall briefly mention below. Following Bragg *et al.* (1950), these helices may be denoted by a symbol n_m, where n is the number of residues per turn and m the number of atoms contained in the ring of atoms joined by the hydrogen bond. The α-helix will thus have the name 3.6_{13}-helix.

Although Pauling and Corey believed that only the α- and γ-helices are possible, other types were found to be possible stereochemically by other workers, e.g., the so-called π-helix was described by Low and

TABLE V
Values of Parameters Connected with Some Standard Helical Structures Which Have Been Proposed[a]

Structure	τ (in degrees)	ϕ (in degrees)	ψ (in degrees)	n	h(Å)	Reference[c]
			trans Peptide unit ($\omega = 0°$)			
α-Helix	109.5	133.0	122.8	3.615	1.495	(1)
α-Helix	110.0	122.2	133.0	3.600	1.50	[a]
3_{10}-Helix	111.6	130.7	154.3	3.00	2.00	(3)
γ-Helix	110.1	263.7	258.0	5.14	0.98	(2)
π-Helix	114.9	122.9	110.3	4.40	1.15	(4)
2.2_7-Helix	111.6	101.9	239.2	2.17	2.75	(3)
2_7-Helix	111.3	105.1	249.5	2.00	2.80	(1)
4.3_{14}-Helix	100.5	268.1	271.7	4.34	1.20	(3)
Polyglycine II	109.1	102.0	325.8	−3.00	3.10	(5)
		258.0	34.2	+3.00	3.10	
Polyglycine II	110	100	330	−3.00	3.10	(10)
		260	30	+3.00	3.10	
Poly-L-proline II	110.0	102.8	325.9	−3.00	3.12	(6)
Poly-L-proline II	108.8	103.7	325.1	−3.00	3.12	(8)
Poly-L-hydroxyproline A:	105.5	103.1	327.6	−3.00	3.05	(7)
Collagen	109	104	340	−3.27	2.91	(12)
	107	112	336			
	107	113	348			
Silk	110	40	315	2.00	3.45	(11)
ω-Helix[b]	109.9	244.4	235.4	−4.00	1.325	(9)
			cis Peptide unit ($\omega = 180°$)			
Poly-L-proline I	114	97	338	3.33	1.9	(13)

[a] Calculated to correspond exactly to $n = 3.6$, $h = 1.5$ Å and $\tau = 110°$.
[b] ω-form of poly-β-benzyl-L-aspartate. The dihedral angle $\omega = 5°$ in this structure.
[c] References
(1) Bamford *et al.*, (1956); (2) Pauling and Corey (1951); (3) Donohue (1953); (4) Low and Grenville-Wells (1953); (5) Crick and Rich (1955); (6) Sasisekharan (1959a); (7) Sasisekharan (1959b); (8) Cowan and McGavin (1955); (9) Bradbury *et al.*, (1962); (10) Ramachandran *et al.*, (1966a); (11) Marsh *et al.*, (1955a,b); (12) Ramachandran (1967); (13) Traub and Shmueli (1963).

Baybutt (1952) and considered in more detail by Low and Grenville-Wells (1953). This helix, however, requires $\tau \sim 115°$. Donohue (1953) described some other helices of which the only new one which is expected to occur is the so-called ribbon structure (2.2_7-helix mentioned in Table V). Donohue also concluded that, of all the different types of helices predicted, the α-helix is the one most stable and, as is well known, that is found to occur in a number of synthetic polypeptides, as well as in proteins. The α-helix invariably occurs in its right-handed form both in

polypeptides and in proteins. But more recently, it has been found that poly-β-benzyl-L-aspartate has a left-handed screw sense of α-helix (Bradbury *et al.*, 1960). If films prepared from this polymer are heated to 110°C and cooled, they give X-ray patterns which correspond to another structure called the ω-helix (Bradbury *et al.*, 1959). This has four residues per turn and may be termed the 4_{13}-helix. The relative stabilities of right- and left-handed helices and of the various other helices are discussed in more detail in Section VII,B.

Expanding their methods to conformations with interchain hydrogen bonds, Pauling and Corey (1951) arrived at the so-called extended beta conformation, which can occur both in parallel and in antiparallel array.

An entirely different fold of the polypeptide chain was deduced from the chemistry and X-ray pattern of the protein collagen by Ramachandran and Kartha (1954, 1955a,b). This is characterized by the conformation of a triple helix held together by interchain hydrogen bonds between the helices, each of which has approximately three residues per turn. A conformation very similar to this occurs in the chains of one of the modifications of polyglycine (Crick and Rich, 1955; Ramachandran *et al.*, 1966a), poly L-proline (Cowan and McGavin, 1955), and poly-L-hydroxyproline (Sasisekharan, 1959b). These three amino acids are the ones which occur in largest amount in collagen. In more recent years, copolymers of these have been found to exhibit a conformation very similar to collagen itself. This is discussed later in Section VIII,D,4.

Thus, the application of simple stereochemical reasoning and model building has led to the discovery of several important types of structures for proteins, which have also been observed for polypeptides. There are really only three basic types of these conformations, namely the α-helix type (highly coiled and internally hydrogen bonded), the extended beta form (which forms sheets by interchain hydrogen bonds) and the triple helix (which also has interchain hydrogen bonds between the three helices). Each of these structures can aggregate to form a fibre, the first and the last by close packing of the approximately cylindrical protofibrils and in the beta form by stacking of sheets. These types of aggregations are particularly relevant to the fibrous proteins. In the case of crystalline proteins, one has chains of finite length and these coil up to form a mass of approximately spherical shape externally. The prediction of the folding of such a chain is a formidable task. However even in the early days, attempts had been made to work out the conformation in a few cases and mention may be made especially of the study of insulin (Lindley and Rollett, 1955) and of ribonuclease by Scheraga (1960). In the absence of more information, and led by the

success of the α-helix in describing several polypeptide structures, the model building procedure in these studies was made by assuming maximum α-helical content in the chain. The best way of defining a polypeptide chain in terms of dihedral angles ϕ and ψ had not then been developed. However, Mizushima and Shimanouchi (1961) attempted to work out some of the restrictions on these dihedral angles by the classical method of assuming staggered conformations. However, now that we have all the methods described in Sections II and III for the description of a polypeptide conformation, it will be possible to work out the ranges of allowed conformations in terms of these parameters. This is discussed in the succeeding Section V. The method is then extended to calculate the energies of the conformations defined by these parameters, and these calculations are discussed in Sections VI and VII.

V. Allowed Conformations of Polypeptide Chains

A. A Pair of Linked Units

1. Conditions Restricting the Allowed Conformations

We have seen in the last section how the relative conformation of a pair of planar peptide units can be defined by two dihedral angles (ϕ, ψ), in which both ϕ and ψ can take all the values from 0 to 360°. The question now arises whether all the geometrically possible values are realizable physically. In fact, it is not so, for the allowed conformations are limited by restrictions on the allowed contact distances between different atoms. Thus, the conformation $\phi = 180°$, $\psi = 180°$ is impossible, for the expected positions of the atoms O(1) and H(2) almost coincide in this case (the theoretical distance is 0.35 Å, which is impossibly short). It is therefore necessary to have a list of allowed interatomic contacts for working out the ranges of ϕ and ψ which are allowed, i.e., which are expected to occur in actual systems. Such ideas of nonbonded contacts between atoms have been considered in the literature and there are data on the so-called van der Waals radii of atoms. (See e.g., Pauling, 1960, for a table of such radii. See also Section VI,C,2 and Table X.) However, the sum of the van der Waals radii of two atoms would represent only the equilibrium distance between those two atoms; and what we require is not this, but the limiting distance to which they can be brought together before the conformation becomes impossible. The latter would obviously be definitely lower than the sum of the van der Waals radii. No clear data were available in the literature regarding such limiting contact distances, but by examining a number of exam-

ples of crystal structures, a set of permitted contact distances between different types of atoms were arrived at in this laboratory (Sasisekharan, 1962; Ramachandran *et al.*, 1963a,b). This list has been very slightly modified in the light of later experience and the latest set which has been commonly used in this laboratory, is listed in Table VI. The table contains two sets of values for each distance r_{AB} between two atoms A and B. These are called "normally allowed limits" and "extreme limits."[6] The former is such that if the interatomic distance is larger,

TABLE VI

Values of Limiting Distances in Ångstroms for Various Interatomic Contacts

Type of contact	Normal limit	Extreme limit
H · · · H	2.0	1.9
H · · · O	2.4	2.2
H · · · N	2.4	2.2
H · · · C	2.4	2.2
O · · · O	2.7	2.6
O · · · N	2.7	2.6
O · · · C	2.8	2.7
N · · · N	2.7	2.6
N · · · C	2.9	2.8
C · · · C	3.0	2.9
C · · · C(H)[a]	3.2	3.0
C(H) · · · C(H)[a]	3.2	3.0

[a] C(H) stands for a CH_2 or CH_3 group in which the hydrogens have not been definitely located.

there are no restrictions on the conformation, and a conformation in which all the interatomic distances are larger than the fully allowed distances is very likely to occur. It is called a "fully allowed" or "normally allowed" conformation. However it is found that, if there is a compensating feature in the crystal structure, such as hydrogen bonds or other attractive effects in the neighbourhood, then contact distances as low as the extreme limits are observed. Therefore, if in a conformation some of the contact distances lie between the normal and extreme limits, then they are still possible, but would be less stable than the fully allowed conformations. Such conformations are called "partially allowed" conformations. If, on the other hand, some of the contacts are less than the extreme limits, then such a conformation is very unlikely to occur and is called a "disallowed" conformation.

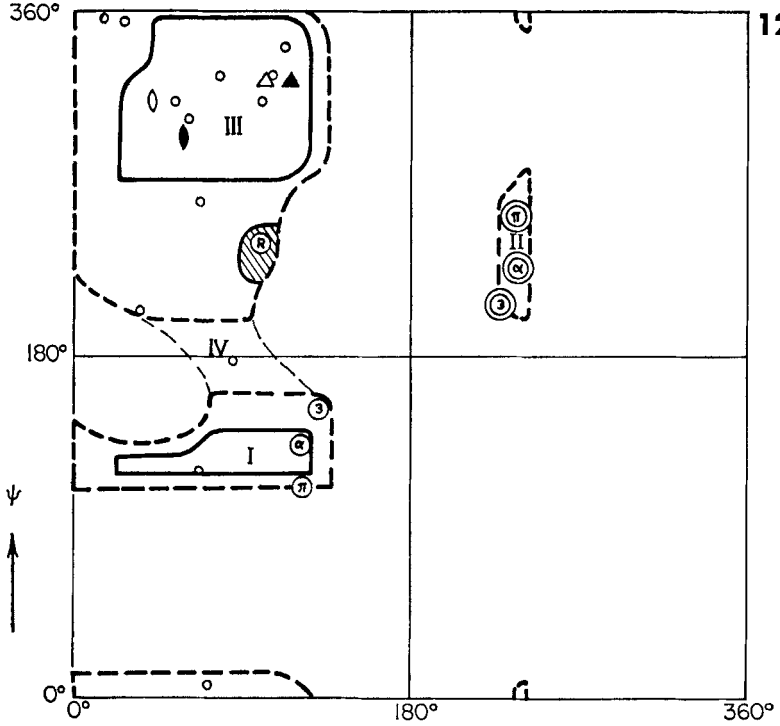
[6] We have here used the term "extreme limit" instead of "outer limit" used earlier in our studies in this laboratory. The word "extreme" conveys the desired meaning better than the word "outer."

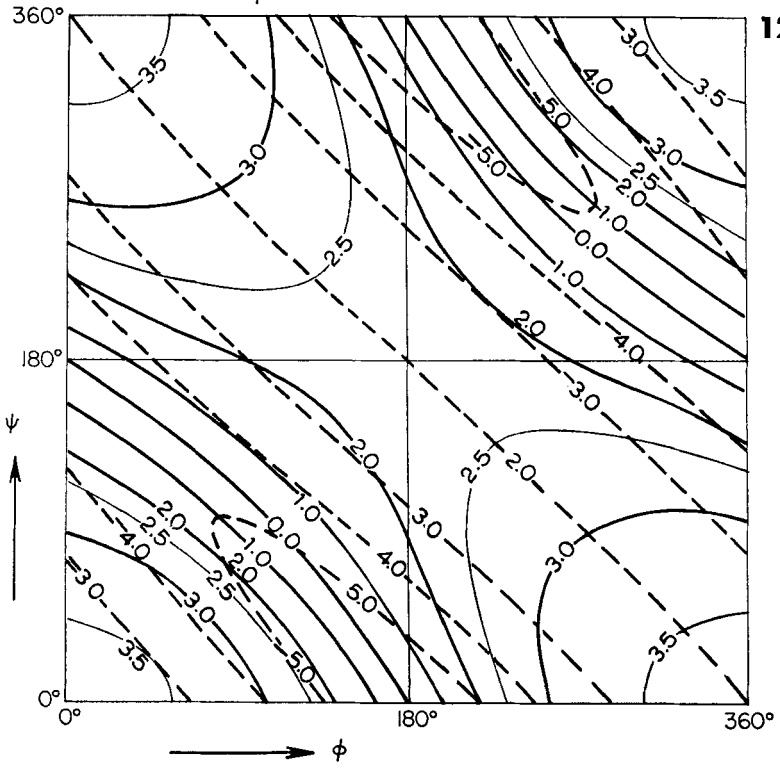
An inspection shows that, for any r_{AB}, the normal limit is 0.3 to 0.5 Å shorter than the sum of the van der Waals radii ($r_A + r_B$) listed in Table X.

2. *Conformational Map—Diagram Showing the Range of Allowed Conformations*

a. Conformational Map When C^β Is Present. Using the above ideas, a map showing the allowed conformations in the ϕ—ψ plane was worked out by Ramachandran *et al.* (1963a,b), and the latest version of this, which is very close to the one first reported, is shown in Fig. 12A. This diagram also contains the expected conformations of some of the helical structures mentioned in Table V as well as the conformations observed in a few simple peptides which have been studied crystallographically. The map corresponds to a value of $\tau = 110°$ and is for a pair of peptide units with a C^β attached to the C^α atom in the L-configuration. In the calculation of contacts, atoms which are third neighbours, as well as those which are separated by three or more atoms, were considered. This practice has been continued in all the later work in the authors' laboratory. Scheraga and co-workers, whose studies are discussed in Sections V, C and V, D, have apparently omitted third-neighbour interactions, as stated by Scott and Scheraga (1966c). These authors also state that if third-neighbour interactions were included, "far too many conformations would have been classified as 'not allowed'." This is not so, for third-neighbour interactions have been included by the Madras group. In fact, our own calculations lead to the conclusion that far too many conformations are *not* found to be "not allowed."

It will be noticed that there are three regions of allowed conformations, which are labelled I, II, and III. Region I contains the α-, 3_{10}- and π-helices, which are all highly coiled structures, all in the right-handed P-form. Region II is only partially allowed and contains the M-form of the helices contained in I. Region III is the largest of all, and a good portion of it is fully allowed; in this range are observed the beta structures and the collagen structure. In addition to these, a portion IV is also marked, bounded by thin dashed lines. This region was marked as being disallowed by Ramachandran *et al.* (1963a,b). However, only a small number of contacts are slightly smaller than the extreme limits, the greatest deviations being $N_1 \ldots N_2 = 2.58$ Å (limit = 2.60 Å) and $N_1 \ldots H_2 = 2.13$ Å (limit = 2.20 Å). Since the violation of the extreme limiting contact distances is only small, it is likely that conformations will be possible in this range also. In fact, as will be seen below, if the angle τ is increased, for example to 115°, the two regions I and II completely connect up and there is no gap in this region IV.


12A
360°
180°
0°
0°
180°
360°
ψ
φ
III
II
IV
I
R
3
α
π


12B
360°
180°
0°
0°
180°
360°
ψ
φ
3.5
3.0
2.5
2.0
1.0
0.0
1.0
2.0
3.0
4.0
5.0

In Fig. 12A the fraction of the total area that is fully allowed is 7.5% while the partially allowed region, including IV, makes up 22.5%.

The corresponding map with a β-carbon atom for $\tau = 115°$ is shown in Fig. 13B. It will be seen that the allowed regions in general expand. On the other hand, for $\tau = 105°$ [Fig. 13A], these regions contract and there is no link between the regions I and III (see e.g., Ramakrishnan and Ramachandran, 1965).

b. Conformational Map with a Glycyl α-Carbon Atom. When there is no C^β attached to the α-carbon atom, the restrictions on the allowed conformations are much less severe than those shown in Figs. 12 and 13A. The allowed conformations in this case are shown in Fig. 14 for $\tau = 110°$. Here, although the map shows four allowed regions in the four quadrants, there is really only one large area of allowed conformations which is completely connected up. In addition, there is a small bridge of slightly disallowed regions across $\psi = 180°$ exactly as in the case of Fig. 12A. The contacts are only very slightly smaller than the extreme limiting distances within this bridge. The observed conformations for glycyl residues in simple peptides and cyclic peptides are also plotted in Fig. 14. For $\tau = 115°$, the bridges across $\psi = 180°$ become partially allowed, as in Fig. 13B. It is interesting to note that all the observed conformations occurring in these bridges belong to cyclic peptides, and they actually have $\tau \geqslant 112°$.

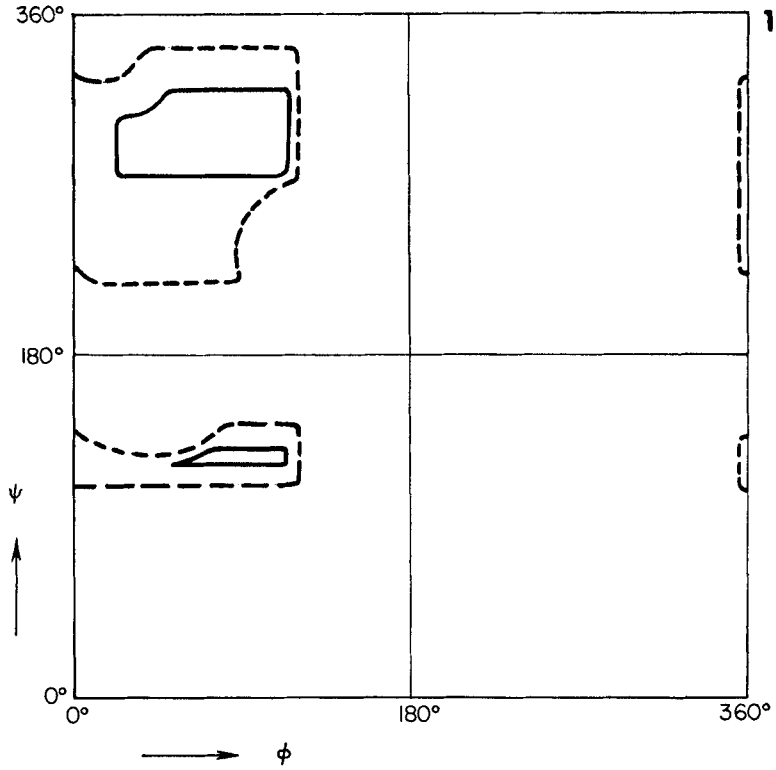
Comparing Figs. 12A and 14, it is seen that Fig. 12A is asymmetric about the points (0°, 0°) and (180°, 180°) but that Fig. 14 is centrosymmetric about these points. The latter result occurs because glycine has two H^α's which are equivalent, and therefore the α-carbon atom is not asymmetric. In the case of a chain composed of D-amino acids, the map corresponding to Fig. 12A would be obtained by inverting it about the point (180°, 180°).

It is interesting to note that the proportion of allowed conformations

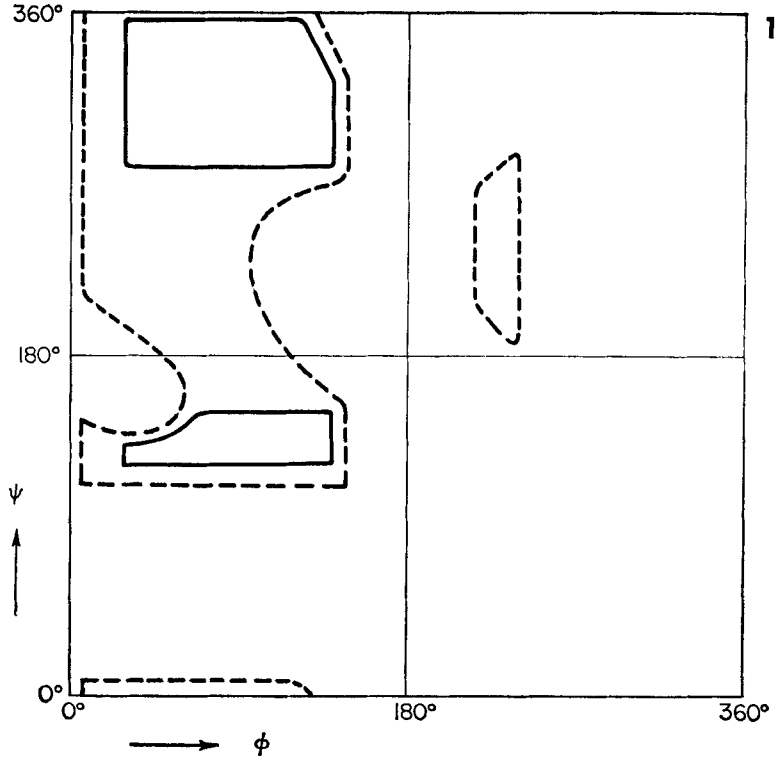
---

FIG. 12A. Conformational map for the case when there is a C^β atom ($\tau = 110°$). Fully allowed conformations are outlined by a solid line and partially allowed conformations by a thick broken line (see text for the explanation of the thin broken lines, region IV near $\psi = 180°$). The conformations plotted are (α) = α_P, (α) = α_M, (π) = π_P, (π) = π_M, (3) = $3_{10,P}$, (3) = $3_{10,M}$, ◆ = β-structure, ◇ = silk, (R) = Ribbon structure, 2.2$_7$-helix, ▲ = collagen, △ = poly-Gly, -Pro, -Hypro, o = observed in simple peptides. The shaded region corresponds to the occurrence of a hydrogen bond of the type N(2)H(2) . . . 0(1) (see Section V,D).

FIG. 12B. Diagram showing lines of constant n and h (in Ångstroms) in the conformational map. The data corresponds to the standard planar trans peptide unit and $\tau = 110°$. (Data from Ramakrishnan, 1965.) ---- Curves of constant n, ——— curves of constant h.



13A



13B

is much larger for a glycyl than for an alanyl side chain (i.e., when a C^β atom is present). In the glycyl case, 45% of the total area is fully allowed, while 61% comes within the extreme limits, as compared with 7.5% and 22.5% respectively for the alanyl case, as already mentioned.

B. Comparison with Observations on Globular Proteins

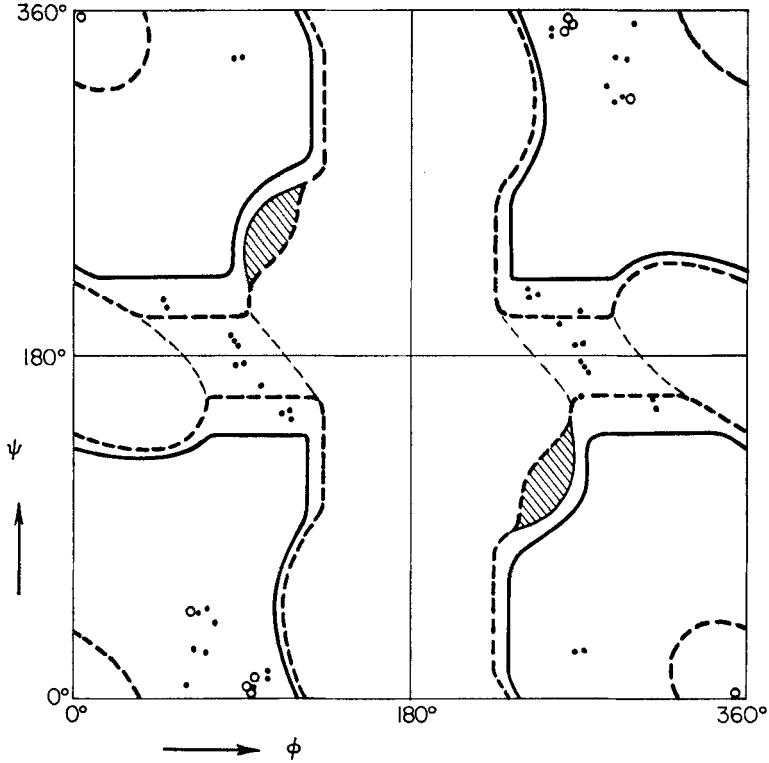
The recent structural determinations of the protein myoglobin (Kendrew *et al.*, 1961; Kendrew, 1963) and of lysozyme (Blake *et al.*, 1967a,b) have yielded information about the conformation of the polypeptide chains as they occur in a protein molecule. It would therefore be of interest to compare the observations in these cases with the predictions of the simple contact theory developed in the preceding section.

It is found that, in myoglobin, a good amount of the chain is in the α-helical conformation. There are in fact several segments of α-helices connected by nonhelical portions. The α-helical conformation is an allowed one according to Fig. 12A and therefore does not require any special comment. The values of ϕ and ψ observed in the nonhelical regions are plotted in Fig. 15. It will be seen that a good number or most of the conformations occur in the allowed regions. A fairly large number occur in the bridging across $\psi = 180°$ and this seems to suggest that, at these α-carbon atoms, τ may be slightly larger than 110°. Although a few are slightly outside the allowed regions in the neighbourhood of the α-helix, they border on these and can be considered to be satisfactory. There are, however, three open circles plotted on these maps, which are completely outside the conformations that are allowed when there is a C^β atom. They must therefore correspond only to glycyl residues and this is in fact actually found to be the case (see also Section VII,C,1).

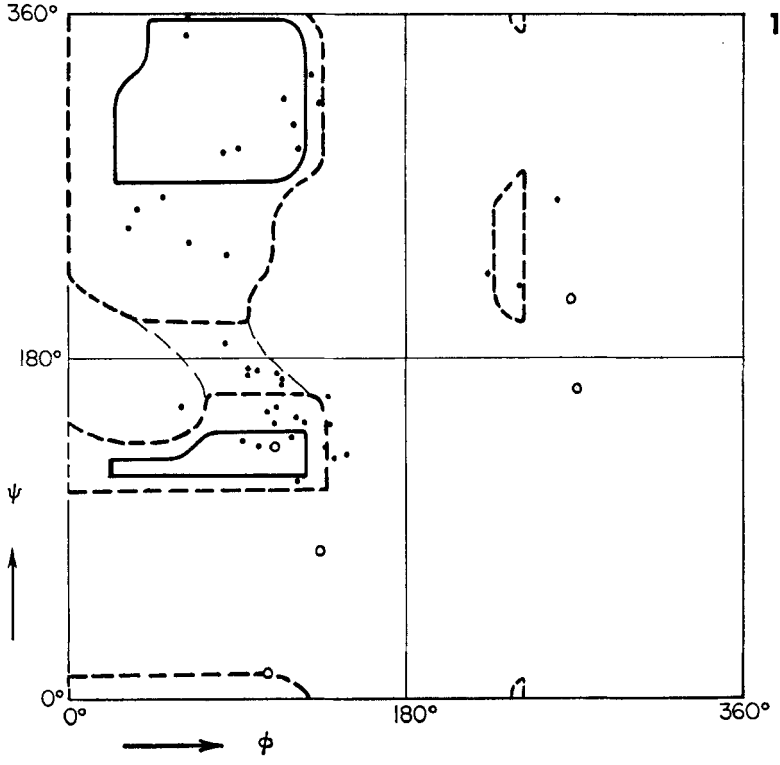
A similar remarkable agreement between the allowed ranges of the predicted conformations and those actually observed has again been found in the structure of lysozyme. Figure 16 shows this and in this case, the individual conformations in the helical regions are also plotted. It will be noted that in addition to a concentration of points in the region of the α-helix and the 3_{10}-helix, and a good number in the bridge across $\psi = 180°$, there is a fairly large number of conformations occurring in the region I of this map, particularly in the fully allowed section of this. It is interesting that very few conformations occur near that of the left-handed α-helix, very similar to what has been observed in myoglobin. In this case, there are several glycyl residues, but it is interesting to note

---

FIG. 13A. Conformational map similar to Fig. 12A, but for an angle $\tau = 105°$.
FIG. 13B. Corresponding map for $\tau = 115°$.



14



15

that, although a few occur in the allowed regions for an alanyl α-carbon atom, most of the glycyl conformations occur in the regions of the map which are disallowed for an α-carbon atom with a side chain. A table of these conformations is given later in Section VIII,C,2 dealing with observational data. It may, however, be mentioned here that two of the conformations, viz. (185°, 291°) and (292°, 153°) occur so far away from the allowed regions that they require explanation. They are discussed further in Section VIII,C,2.

C. Effect of Longer Side Groups

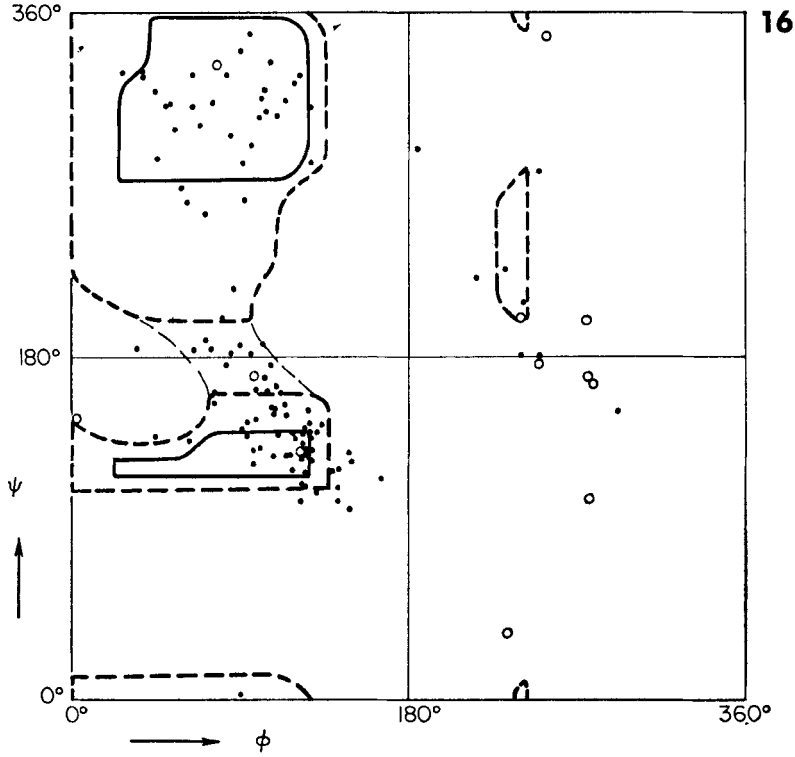
The broad agreement between the observed data and the predicted conformations as indicated above has led to more detailed studies on the allowed ranges of conformations when the effects of atoms beyond C^β are included. One such study has been published by Ramachandran *et al.* (1965). They have studied the resulting conformational map when a C^γ atom is also present in addition to the C^β atom in the side group. It is found that C^γ occurs only for values of χ^1 close to 60° (position I), 180° (position II), and 300° (position III). (This is discussed further in Section VIII,B.) The C^γ atom was assumed to occupy the three positions one by one, and the effect on the conformational map was worked out, which is shown in Fig. 17. It will be seen from this figure that the effect is mainly to remove regions which are not highly populated in the maps we have already discussed. Only the case of position I requires some comment. For this case, the conformation of the left-handed α-helix is close to the allowed limit while that of the right-handed α-helix is fully allowed.

Such studies have been extended by other workers, notably by Leach *et al.* (1966a). They considered the effect of the variations in the groups occurring in the side chains. In addition, they also studied the effects of the variation in the geometry of the planar amide backbone and the variation in limiting contact distances on the conformational map. The results have been reported in the form of stepwise diagrams. In the

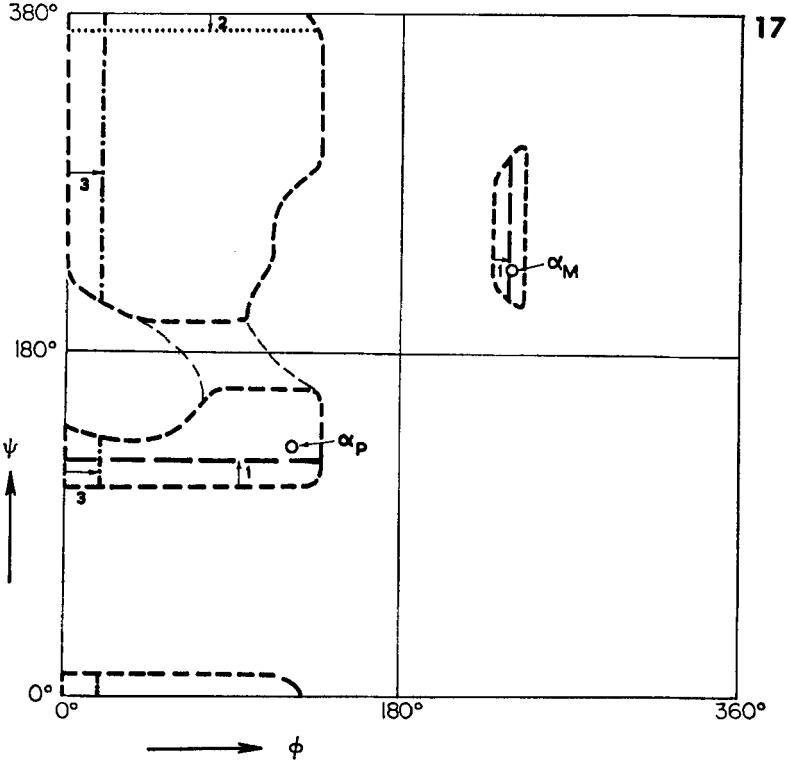
---

FIG. 14. Conformational map for glycyl α-carbon atom ($\tau = 110°$). Note the large increase in allowed conformations relative to Fig. 12A. The open circles are the observed data for simple peptides containing glycyl residues and the closed circles for cyclic peptides.

FIG. 15. A conformational map for $\tau = 110°$ with the observed conformations in the nonhelical regions of myoglobin plotted on it. The latest data were kindly supplied by Dr. H. C. Watson of Cambridge. (An earlier version of this figure has appeared already in Ramachandran *et al.*, 1966b, and in Davies, 1965.) The conformations (ϕ, ψ) at glycyl α-carbon atoms are denoted by open circles and all others by black circles.



16



17

case of glycyl and alanyl α-carbon atoms as the linking atom, the maps are nearly the same as those reported by Ramachandran *et al.* (1965). However, Scheraga *et al.* (1967) found that the inclusion of atoms beyond C^γ imposed significant restrictions on ϕ and ψ. An example of such a conformational map is shown in Fig. 18 (Némethy *et al.*, 1966; Scheraga *et al.*, 1967).

D. Helical Chains

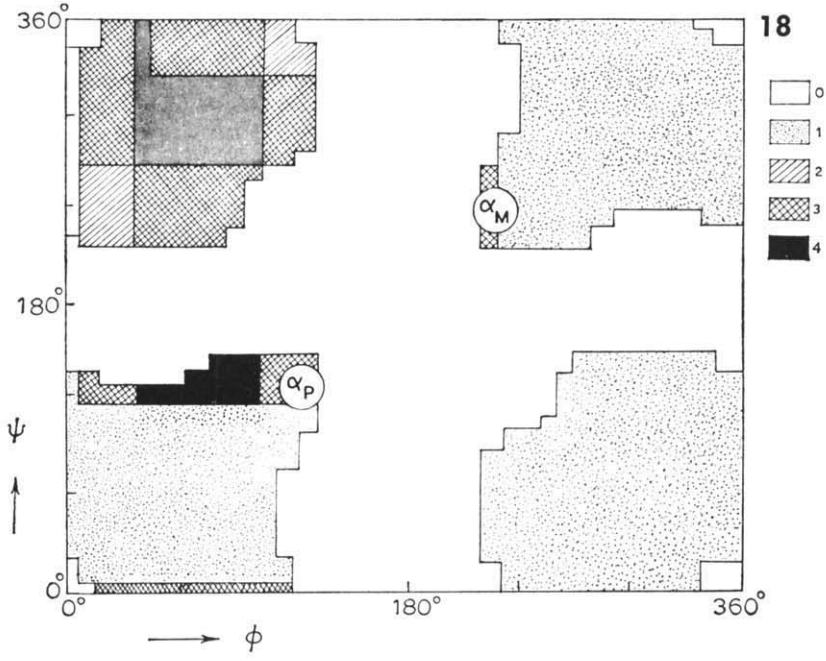
The extension of the above ideas for a pair of peptide units to a study of allowed conformations of helical structures, which are characterised by the same (ϕ, ψ) at all linking α-carbon atoms, was made by Ramachandran *et al.* (1966b) and by Leach *et al.* (1966b). Obviously, more restrictions will be encountered in the allowed ranges of (ϕ, ψ) as one goes up from dipeptidelike to tripeptidelike and higher helical structures. Such additional steric restrictions were investigated for helical pentapeptidelike and hexapeptidelike structures by Leach *et al.* (1966b). They introduced, in addition, the criteria for the formation of hydrogen bonds and examined the influence of the peptide group geometry on these properties.

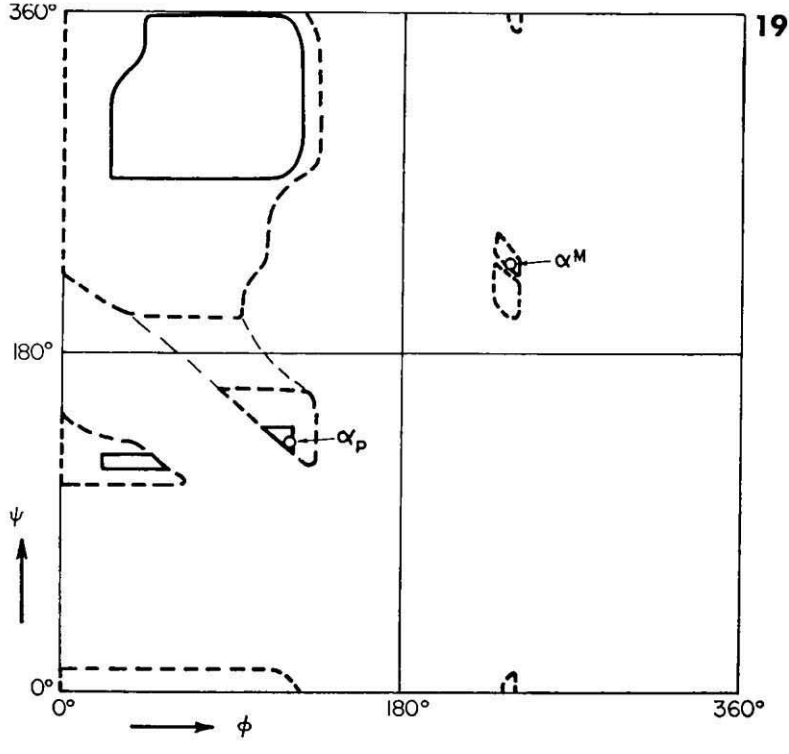
The restriction on the allowed regions of (ϕ, ψ) in the conformational map (with a β-carbon atom) when one goes up to a helical chain of poly-L-alanine is shown in Fig. 19 (taken from Venkatachalam and Ramachandran, 1967). The positions of the α-helices α_P and α_M which come within the allowed ranges are also shown in the figure, showing that contacts between a peptide unit and one higher up removed from it by four or five units do not interfere with this conformation. On the other hand, conformations close to α_P and α_M are stabilised by hydrogen bonds. The conformation of α_P is in the fully allowed region of the contact map, while that of α_M is only in the partially allowed region. This is for a value of $\tau = 110°$. The actual value of (ϕ, ψ) is highly sensitive to τ (Table V) and we shall discuss this further in Section VIII,C,1. The plotted points for all the helices (in particular α-helices) in this review correspond throughout to a value of $\tau = 110°$. Such hydrogen bonds have been investigated in detail by Ramachandran *et al.* (1966b). Typical regions of the conformational map showing the for-

---

FIG. 16. Conformational map for $\tau = 110°$ (with C^β) with the observed conformations in lysozyme plotted on it. Glycyl and nonglycyl α-carbon atoms are distinguished as in Fig. 15. (Data for lysozyme conformations from Table XXV.)

FIG. 17. Effect of the C^γ-atom at the three positions I, II, and III on the conformational map of two linked peptide units ($\tau = 110°$). (From Ramachandran *et al.*, 1965.) --- position I, ... position II, -·-·- position III. The positions of α_P and α_M are also marked.


18
360°
180°
0°
0°
180°
360°
ψ
φ
α_M
α_P
0
1
2
3
4


19
360°
180°
0°
0°
180°
360°
ψ
φ
α^M
α_P

mation of 3—1, 4—1, and 5—1 hydrogen bonds for a value of $\tau = 110°$ are shown in Fig. 20A–C. Making a slight departure from the conventions of Edsall *et al.* (1966a,b,c), we shall call the hydrogen bond from a NH group of a residue i to the CO group of a residue j as i—j, the direction of the bond being from the donor to the acceptor atom. Figures 20A–C show clearly that it is possible to have hydrogen-bonded helices topologically similar to the α-helix, 3_{10}-helix, and 2.2_7-helix for this value of τ and that these are the only intrahelical types of hydrogen bonds that can occur in a single helical polypeptide chain. No hydrogen bonds are found to be possible between a NH of a preceding residue and a CO of a succeeding residue, i.e., those of the type 1—j, $j > 1$, if there is a side group.

The π-helix does not come up in this map and therefore Ramachandran *et al.* (1966b) studied the variations of hydrogen bonding with the value of τ varying from 105° to 115°. At 105°, no hydrogen-bonded helical conformation of any type is possible with planar peptide units. The 4—1 and 5—1 types of bonds become possible only for a range from 108° to 112°. However, if τ is increased above 112°, a 6—1 hydrogen bond becomes possible and therefore both the right- and left-handed conformations of π-helix, namely π_P and π_M are possible for τ greater than 112°. However, the region in the (ϕ—ψ) map in which they occur is extremely limited, and in a survey made at intervals of 2° for $\tau = 115°$, the data observed are shown in Table VII. This, combined with the fact that the π-helix occurs only near the edge of the partially allowed region, may be the reason why it has not been observed so far.

The effect of the γ-carbon atom on the formation of α-helices has also been investigated (C. M. Venkatachalam, unpublished). It is found that position I is not possible for this atom for an α- or 3_{10}-helix, while positions II and III are possible. This would mean that if valine occurs in the α-helical portion of a chain, then the two γ-carbon atoms can occur only in one of the three possible combinations, namely positions II and III.

However, it is found that an oxygen atom, as in serine, can occur in

---

FIG. 18. Allowed areas in the (1—2) conformational map corresponding to different side groups. Conformations in areas 0 are disallowed, 1 to 4 allowed for glycyl-α-carbon atom, 2 to 4 for alanine side group, 3 to 4 for higher straight chain homologues, while only 4 is allowed for L-valine and L-isoleucine side groups at the linking α-carbon atom. (From Scheraga *et al.*, 1967.)

FIG. 19. Conformational map for a perfect helix of poly-L-alanine, ——— fully allowed, --- extreme limit. (From Venkatachalam and Ramachandran, 1967.) The conformations of α_P and α_M are also shown.

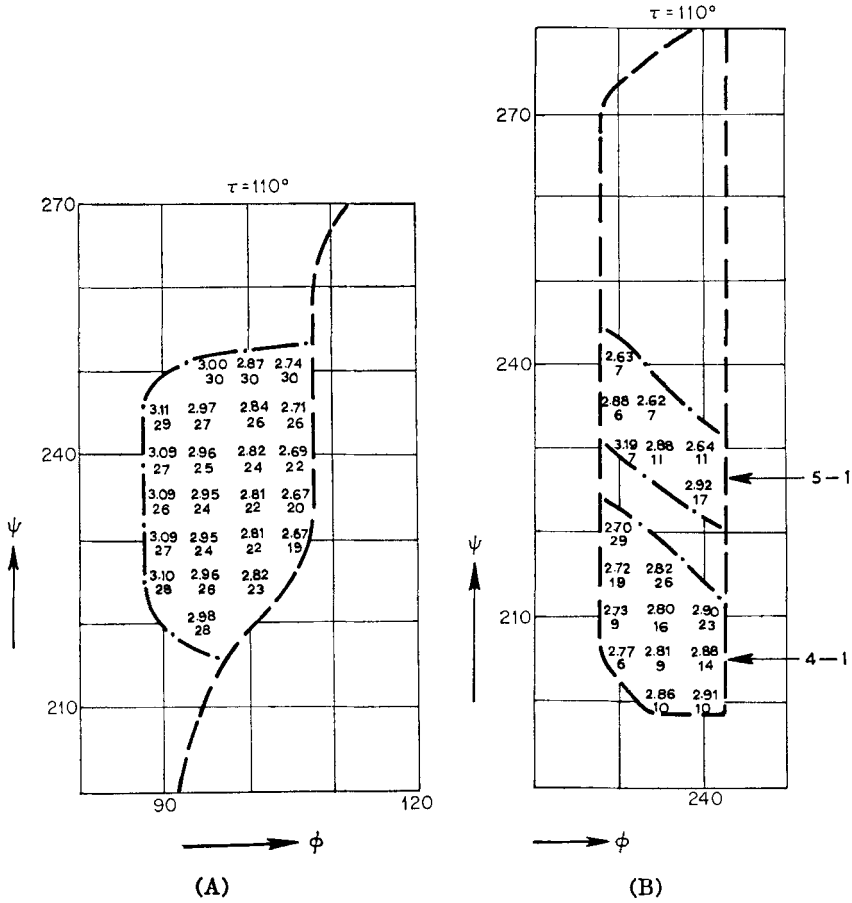
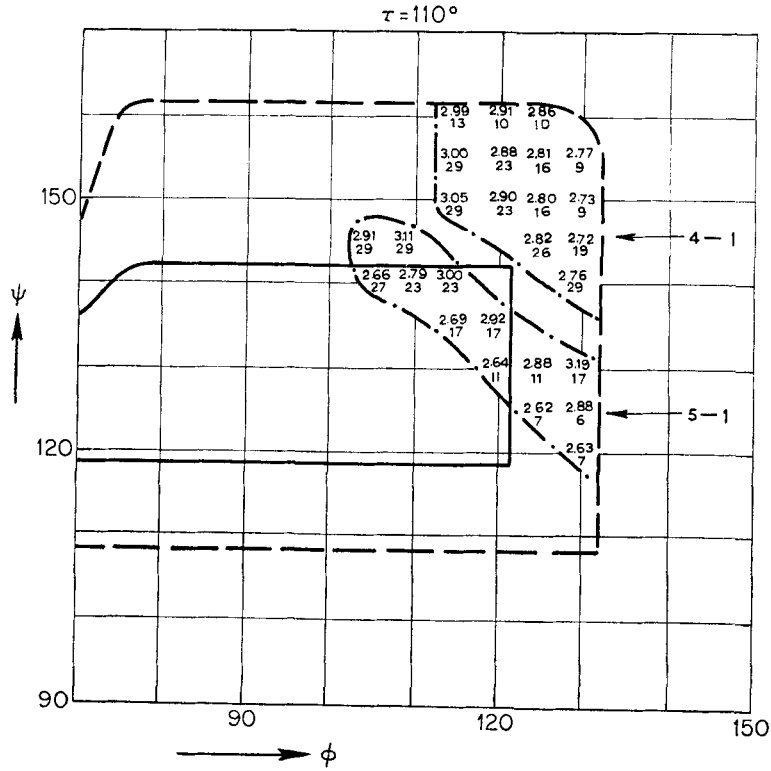


FIG. 20. Regions of the conformational map in which hydrogen bonds occur. A. 3—1 hydrogen bonded conformations (P-type helices). B. 4—1 and 5—1 hydrogen bonds with M-type helices. C. 4—1 and 5—1 hydrogen bonds with P-type helices. Only regions allowed by the helical contact map of Fig. 19 are shown. The values of the hydrogen bond length (in Ångstroms) and the H—N—O angle (in degrees) are shown at grid points at intervals of 5°. (From Ramachandran *et al.*, 1966b.)

position I of the γ-atom (as it is appreciably smaller than a CH_3 group) for a 3_{10}-helical conformation, but not for a right-handed α-helix. In fact, in this position, for $\chi^1 \simeq 60°$, $\phi \simeq 125°$, $\psi \simeq 155°$, an OH . . . O bond can occur between the serine side group and the same carbonyl oxygen O to which the NH is hydrogen bonded (i.e., a 4—1 bond) in a right-handed 3_{10}-helical arrangement (Sarathy and Ramachandran, 1968), Sarathy finds that both OH . . . O and NH . . . O bonds (4—1) can



(C)

FIG. 20. (Continued.)

occur together also for $\chi^1 \simeq 300°$ for the $3_{10,P}$-helix. In addition, it is also possible to have both OH . . . O and NH . . . O bonds for a left-handed α-helical conformation (Sarathy and Ramachandran, 1968). In this case, the NH . . . O is of the type 5—1, while OH . . . O is 2—1, and $\chi^1 \simeq 65°$, $\phi = 227°$, $\psi = 238°$. Thus, a series of hydrogen-bonded conformations, leading either to a right-handed, or a left-handed helix, are possible for poly-L-serine, and this may be the reason why ORD studies on this polymer in solvents promoting α-helix formation do not show the existence of this conformation (the right- and left-handed forms giving opposite effects).

E. Extension to Three Linked Peptide Units

The case of three linked peptide units (1), (2) and (3), whose conformation can be defined by the parameters (ϕ_2, ψ_2) and (ϕ_3, ψ_3), has been worked out by Venkatachalam (Venkatachalam, 1968a). It is

TABLE VII

Data at Intervals of 2° for ϕ and ψ for Hydrogen Bonds of Allowed Helices with the 6—1 Type Bond (π-Helix) for τ = 115°

ϕ	ψ	Length A	Angle NH ∧ NO	Angle CO ∧ ON	n	h (Å)
(a) *Right-handed helix*						
122	112	2.88	12	14	4.20	1.00
(b) *Left-handed helices*						
218	272	3.11	30	22	−4.66	1.18
220	268	3.19	24	20	−4.57	1.22
	270	3.00	28	20	−4.65	1.16
222	266	3.09	22	18	−4.56	1.20
	268	2.90	25	17	−4.63	1.15
224	264	3.01	19	16	−4.54	1.18
	266	2.81	22	15	−4.61	1.13
226	260	3.15	16	17	−4.46	1.22
	262	2.93	17	14	−4.53	1.17
228	258	3.08	14	15	−4.44	1.21
	260	2.86	14	12	−4.51	1.16
230	256	3.02	12	14	−4.43	1.20
	258	2.79	11	10	−4.50	1.15
232	256	2.74	9	8	−4.48	1.14

found that there are practically no short contacts between atoms in the units (1) and (3), when the conformation is allowed by (1) and (2) and also allowed by (2) and (3). There are only two small regions of disallowed conformations of the three linked peptide units, around points designated by the parameters (ϕ_2, ψ_2; ϕ_3, ψ_3) = (120°, 180°; 60°, 160°) and (120°, 270°; 300°, 180°).

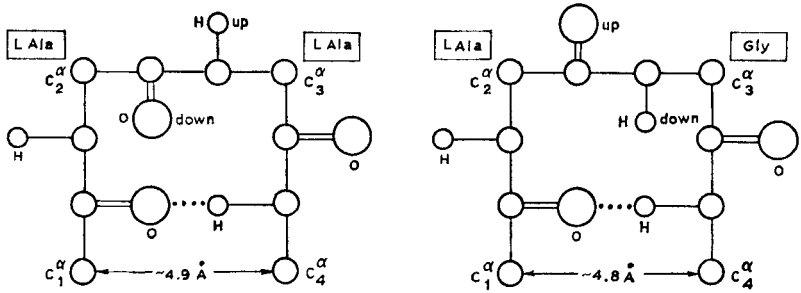


FIG. 21. The two possible conformations of three peptide units which lead to chain reversal and have hydrogen bonds between units (1) and (3). (From Venkatachalam, 1968a.) (A) around (120°, 150°; 90°, 180°), (B) around (120°, 300°; 260°, 180°). (See text for comparison with observation.)

TABLE VIII

Calculated and Observed Conformation of Chain Reversals Involving One Intermediate Peptide Unit in Three Units

	Type A		Type B	
	Can accommodate C^{β}_2 and C^{β}_3		Can accommodate C^{β}_2 and only H^{α}_3	
	Calculated[a]	Observed	Calculated[a]	Observed
(ϕ_2, ψ_2)	(120°, 150°)	(112°, 151°)	(120°, 310°)	(123°, 312°)
(ϕ_3, ψ_3)	(90°, 180°)	(86°, 189°)	(260°, 180°)	(262°, 179°)
$\omega_1, \omega_2, \omega_3$	0°, 0°, 0°	3°, −2°, 8°	0°, 0°, 0°	6°, −1°, −4°
$N_1H_1 \cdots O_3$	2.72A	3.00A	2.84A	2.98A
NH ∧ NO	21°	[b]	24°	[b]
N ··· O ∧ OC	70°	66°	48°	34°
$C^{\alpha}_1 \cdots C^{\alpha}_4$	4.64A	5.16A	5.04A	5.61A
Sequence	—	—Gly-Gly— in cyclohexaglycyl	—	—L-Ser-Gly— in Ferrichrome A

[a] The "calculated" values are for a conformation nearest in the tabulated list. All the ω's were taken to be zero (corresponding to planar peptide units).

[b] Not available, since the hydrogen atom was not located.

On the other hand, there are several favoured conformations stabilised by hydrogen bonds of the type 4—1, i.e., between N—H of unit 3 and C=O of unit 1. This feature occurs around three distinct regions, characterised by specific foldings of the polypeptide chain. These are (A) around (120°, 150°; 90°, 180°), (B) around (120°, 300°; 260°, 180°) and (C) around (120°, 150°; 120°, 150°). While (C) is a type of folding which could form part of a 3_{10}-helix, the other two are clearly non-helical. In both of them, the peptide chain turns around, reversing its direction of progress on either side of the middle unit. The distance between them is of the order of 4.8 Å. They could both therefore be extended in either direction to form an antiparallel pleated sheet and could form the basis of the folding back and forth of the chain in a cross-beta structure. They are shown diagrammatically in Fig. 21. Examples of both these conformations have been observed in cyclic peptides, that of (A) in cyclohexaglycyl (Karle and Karle, 1963) and that of (B) in ferrichrome A (Zalkin *et al.*, 1966). The nearest predicted conformation in the calculated list and the observed one in each case are given in Table VIII.

F. Extension to Cyclic Peptides

1. Dipeptides

The case of diketopiperazine, which is a dipeptide with two peptide units in the cis modification, has been mentioned in Section II,A,1. Diketopiperazines of amino acids other than glycine are also known. In all these, the cis peptide unit is expected to occur, although it is energetically less favourable than the trans peptide unit, because ring closure is not possible for the trans form. The dimensions of the peptide unit in glycyl diketopiperazine are not typical, probably because of resonance in the ring. However, the easy formation of this cyclic dipeptide indicates that it is possible for the peptide unit to take up the cis conformation, if it is stabilised by the formation of a covalent bond which is required to close the ring.

2. Tripeptides

In this case also, simple model building shows that it is impossible to get a ring structure either with three trans peptides, or with a combination of cis and trans peptides. On the other hand, three cis peptide units can readily join together to form a symmetric tripeptide with a three-fold axis of symmetry. However, the formation of such a compound requires the initial occurrence of the cis conformation not only near the region

where ring closure occurs, but also in the intermediate peptide unit, and hence it is best realised in the case of the tripeptide of an amino acid like proline or hydroxyproline, for which the cis and the trans conformations are very nearly of the same energy (cf. Section II,A,1).

In fact, cyclotriprolyl has been synthesized (Rothe *et al.*, 1965). The problem now is to find its conformation. As mentioned above, this can be done by having all the three peptide units in the cis conformation; but it is found that if planar peptide units are used, the side chain five-membered ring cannot be readily closed and further there are bad contacts between the hydrogens H^{α} of neighbouring units, of the order of 1.4 Å. The former of these defects can be removed by making the peptide units nonplanar and this also reduces the distortion needed to make the $H^{\alpha} \ldots H^{\alpha}$ contact reasonable. The best conditions have been worked out by Venkatachalam (1968b) and it is found that under the conditions given below, a satisfactory structure free of bad contacts is possible. The cis peptide unit was taken to have the standard dimensions given in Table II(c). With these and assuming ω to be approximately $-20°$, the conformation is satisfactory and corresponds to $\phi = 97°$ and $\psi = 287°$ for the tripeptide of L-proline. The value of $\phi = 97°$ is close to that found in poly-L-proline I, an open chain having cis peptide units with proline side chains (cf. Table V and Section V,G,2). A slight distortion in the position of the atoms H^{α} and C^{β} is necessary, but the resultant six angles at the α-carbon atom do not differ from the tetrahedral angle by more than 3°. It may be mentioned that distortions of the tetrahedral angle at the α-carbon atom of this order and deviations from planarity of the peptide units with ω going up to 15° are observed even in the X-ray structure of cyclohexaglycyl (Karle and Karle, 1963), which is a much less strained cyclic peptide.

3. Tetrapeptides and Pentapeptides

Practically no cyclic tetrapeptides have been reported in the literature. However, Balasubramanian and Wetlaufer (1967), have studied the optical rotatory dispersion of cyclotetraalanine, which they found was the real nature of a commercial sample of L-alanyl-L-alanine diketopiperazine. This was found to have an optical rotatory dispersion similar to that of an α-helix. An investigation of the conformation of this cyclic peptide made by Ramakrishnan and Sarathy (1968) showed that it is possible to have a cyclic symmetrical structure with a distortion of the peptide plane corresponding to a value of ω of about $-10°$ and having $\tau = 109.5°$, using only trans peptide units. Thus, a helix with $n = 4$ and $h = 0$ is possible for a distorted peptide unit. The corresponding values of ϕ and ψ are 77° and 137°. (All these are for L-

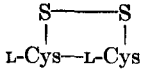
residues; see Section VII,G for the effect of introducing D-residues.) Cyclotetraglycyl also can obviously have a similar structure, for the contact restrictions are less severe in this case.

The case of the pentapeptides is not simple, because it is possible that the chain may not have a symmetric configuration with a five-fold axis of symmetry, but may fold up irregularly. However, the symmetric conformation has been investigated by C. Ramakrishnan and K. P. Sarathy (unpublished) and they find that it is possible to have this with very little distortion of the planar peptide unit, for a chain having all L-residues. Only an ω-distortion of the order of 6° is required to give a cyclic structure with $n = 5$, $h = 0$, $\tau = 111°$. This corresponds to $\phi = 70°$, $\psi = 128°$. Cyclopentaglycyl also can have a similar structure.

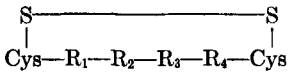
The extension to cyclic peptides with $n > 5$ is not easy, for the number of possible conformations are large, and contact criteria are only able to show that they are possible, but cannot predict which are the likely ones to occur. It is necessary to calculate the energies of the different conformations to find out which are most likely. Attempts made in this direction are discussed in Section VII,G.

4. Cyclic Peptides with S—S Bridges

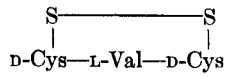
Several cyclic peptides with six or more residues linked by a —Cys—S—S—Cys— bridge are known (Schröder and Lübke, 1966). Of these, those with six in the ring of type (VI) are the most common, e.g., in oxytocin, vasopressin, and insulin. However, rings with a smaller number of residues are known, particularly when the R's are Gly's. In particular,



has been synthesised and a three-residue ring of the type (VII) with mixed D and L residues occurs as part of an antibiotic peptide, malformin (Schröder and Lübke, 1966).

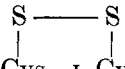


(VI)

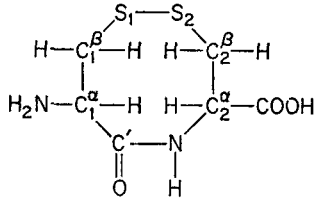


(VII)

The cases with two and three residues linked by an S—S bridge have been considered at Madras. For this purpose, the allowed contact distances between sulphur and other atoms were suitably worked out and used in the calculations.



a. Dipeptide L-Cys—L-Cys. In this case (Chandrasekaran, 1968), it is found that ring closure by an S—S bridge is possible only if the only peptide unit that is involved in the chain is in the cis conformation. Structure (VIII) shows the compound schematically.



(VIII)

The relevant dihedral angles are, starting from S_2 and going clockwise along the ring: χ_2^2 (about C_2^β—S_2), χ_2^1 (about C_2^α—C_2^β), ϕ_2 (about N—C_2^α), ω (about C′—N) ψ_1 (about C_1^α—C′), χ_1^1 (about C_1^α—C_1^β), χ_1^2 (about C_1^β—S_1) and finally the torsional angle of rotation about the S_1—S_2 bond, which in our notation will be χ_1^3 (we may denote it for convenience as χ^S).

The various allowed ranges of these angles were first explored by means of a model, with the criterion that τ_1^α and τ_2^α were made equal to 110°. The dimensions used by Traub and Shmueli (1963) for the cis peptide unit in connection with their studies on poly-L-proline I were adopted for this unit. These were taken from the structure of Leu-Pro-Gly and are very close to our standard dimensions (see Section VIII,H). τ_1^β and τ_2^β were made equal to 115°, a value close to that found in various Cysteine and cystine derivatives. The conformations which lead to a distance between the sulphur atoms close to 2.05 Å and the angles $\tau(S_1)$ and $\tau(S_2)$ less than 120° were explored. It was found that none were possible for $\omega = 0$, but when the cis peptide plane was distorted, allowed conformations occurred both for $\omega \sim 20°$ to 30° and for −20° to −30°. In the former case, the dihedral angle χ^S was close to +90° and in the latter to −90°, both of which are commonly found (see Sections VIII,B,2 and VIII,C,2). A typical example (which is nearly the best conformation) has the following parameters.

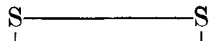
$\omega = +30°$, $\phi = 30°$, $\psi = 310°$, $\chi_1^1 = 190°$, $\chi_1^2 = 298°$,
$\chi_2^1 = 310°$, $\chi_2^2 = 308°$, $\chi^S = 92°$

and

$$l(S_1—S_2) = 2.05\ \text{A}, \quad \tau(S_1) = 112°, \quad \tau(S_2) = 104°$$

Thus this compound can definitely occur with a cis peptide unit.

Later energy calculations (R. Chandrasekaran, unpublished) indicate that the minimum energy conformation corresponds to a value of $\omega \simeq -10°$, if the angles at the α-carbon atoms are allowed to change from 110° by a small amount.



b. Tripeptide L-Cys—L-Cys—L-Cys. This has been found to be possible to occur with two trans peptide units. So far, only the planar case ($\omega = 0$) has been explored, but even with this, using $\tau(C^\alpha) = 110°$ and $\tau(C^\beta) = 115°$, an allowed conformation occurs with $\tau(S_1) = 120°$, $\tau(S_3) = 120°$ and $\chi^S = \chi(S_1—S_3) = 256°\,(-104°)$ and $l(S_1—S_3) = 2.03$ Å. The two angles at the sulphur atom are a little too large, but they are likely to come down when nonplanar ω-distortion is introduced (as was found to be so in the dipeptide case). The other relevant dihedral angles are (R. Chandrasekaran, unpublished):

$$\psi_1 = 0°, \quad \phi_2 = 90°, \quad \psi_2 = 220°, \quad \phi_3 = 100°$$
$$\chi_1^1 = 50°, \quad \chi_1^2 = 114°, \quad \chi_3^1 = 305°, \quad \chi_3^2 = 96°$$

It is believed that the introduction of finite ω-distortion would lead to even better structures. The middle L-Cys can be replaced by L-Ala or Gly and in all cases, both the peptide units are trans. The case when it is a D-residue is being investigated by Chandrasekaran. As mentioned above, such a D-residue actually occurs in a peptide.

The case of larger ring structures with S—S bridges has been studied using energy minimization methods and is discussed in Section VII,F.

G. Chains with cis Peptide Units

Most of the discussion so far has been based on trans peptide units, except for a few instances of occurrence of cis peptide units in small cyclic peptides. In these cases, the closure of the ring provides the stabilising energy necessary to overcome the loss of energy attendant on assuming the cis conformation.

As already mentioned, a chain composed solely of cis peptide units is observed to occur in poly-L-proline I (see Section VIII,E,3 for details). It is therefore of interest to discuss the (ϕ, ψ) conformational map for cis peptide units. Preliminary studies of this type have been made in the authors' laboratory (Ramachandran and Venkatachalam, 1968) and they may be briefly summarized here.

Analogous to the trans case, the contact criteria given in Table VI were used to calculate the conformational map for a pair of cis peptide units of the standard type given in Table II(c). The map is shown in

Fig. 22 for $\tau = 110°$, when there is a C^β atom in the L-configuration. The map has two allowed regions, bridged over a small range of ψ-values in which just one contact $C^\beta \ldots O$ falls down to 2.67 Å, only 0.03 Å less than the extreme limit. Within this allowed region, there are four blocks of fully allowed conformations.

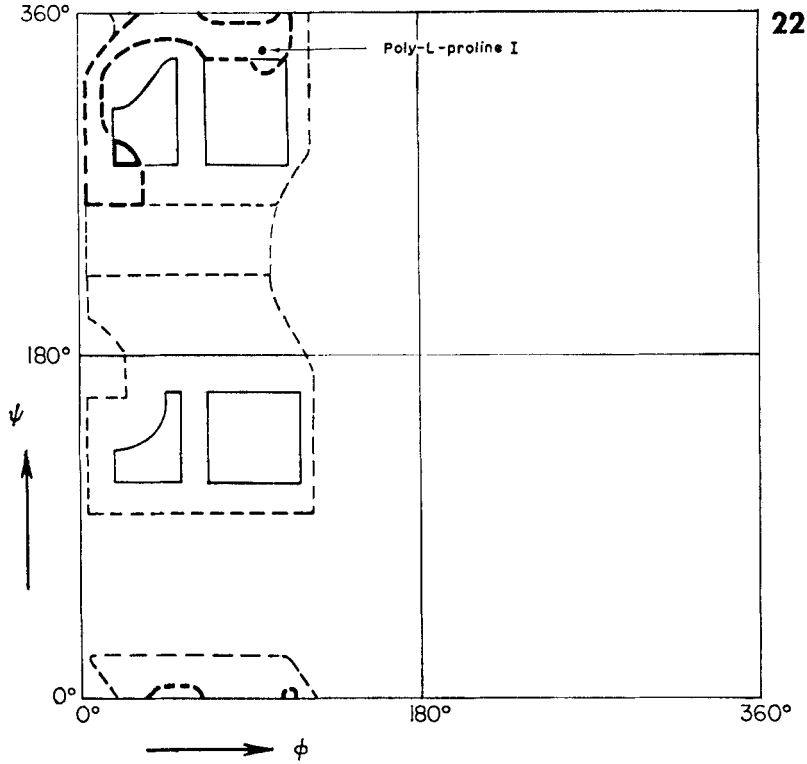
When a helix of cis peptide units of poly-L-alanine is considered, it is found that only the curved strip marked by thick dashed lines in Fig. 22 is allowed. It is interesting that this strip contains the conformation (ϕ, ψ) of the poly-L-proline I helix at (100°, 340°). When the proline side chain occurs linking the C^α atom with the nitrogen, ϕ can only have a value in the neighbourhood of 100°, and for such a value of ϕ, the only range of ψ that is allowed for a helix is from about 330° to 355°, which is just what is observed in poly-L-proline. This probably explains the ready occurrence of the poly-L-proline I structure with cis peptide units.

VI. Calculation of Energies Associated with Conformations

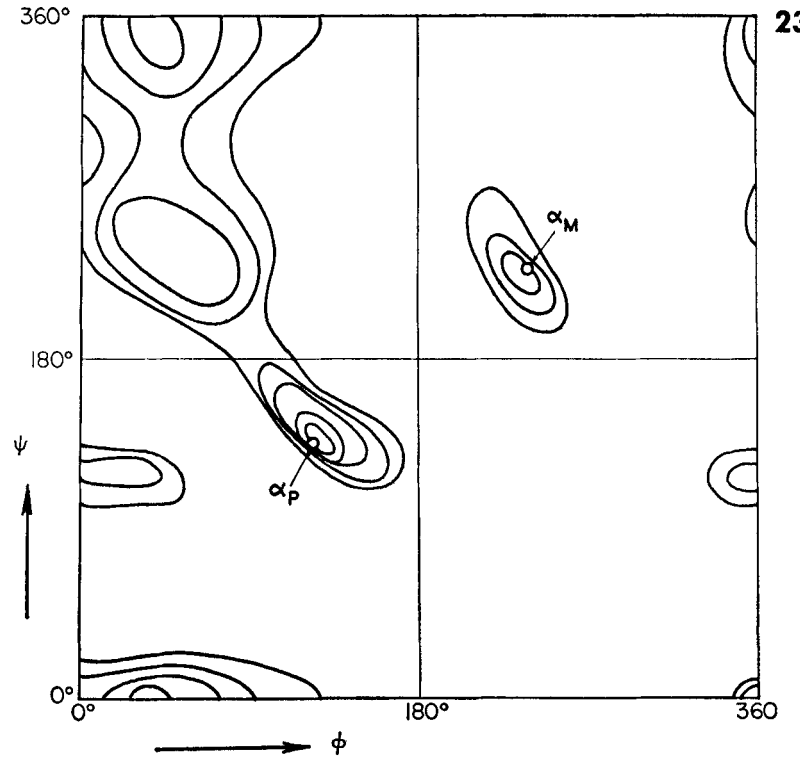
A. Introduction

The use of contact criteria which were considered in Section V, gives significant information, but it has its limitations. The use of two sets of limits, namely those giving the fully allowed and partially allowed regions, gives some idea of how the allowed regions merge into disallowed ones; but this information is not quite enough. For instance, the question may arise as to how many contacts are bad in a disallowed region and how bad they are—for example, we saw that contacts in region IV of Fig. 12A, which is given as disallowed by the computer, are really only very slightly worse than the extreme limit of the partially allowed region. On the contrary, there may be portions of a partially allowed region in which a number of contacts are rather bad, although none of them is shorter than the extreme limits, and the conformation may therefore be an unfavourable one. It is therefore desirable to work out a method which would give the potential energy of a conformation based on the various interatomic distances that occur in it and which would give a much better idea of the stability or otherwise of a particular conformation.

However, such a procedure is not easy and straightforward, because the theory of nonbonded interactions is still not fully developed and much of it is somewhat empirical. For instance, different types of formulae have been proposed for such interactions and, even for a particular type, varying sets of values of the parameters occurring in it have been proposed by different workers. Therefore, we shall only


360°
180°
0°
ψ
Poly-L-proline I
0°
180°
360°
φ

22


360°
180°
0°
ψ
α_M
α_P
0°
180°
360
φ

23

attempt here to give an account of the present status of the subject and briefly indicate what appears to be the most reasonable set of assumptions.

The first attempt at an application of such potential functions to calculate the energies of polypeptide conformations was that of Liquori and co-workers (DeSantis *et al.*, 1965), who did it for a helical chain of poly-L-alanine. Their map, in our notation (ϕ, ψ), is shown in Fig. 23. As will be seen from this, their calculations indicate a much lower minimum of energy in the region of the right-handed α-helix, α_P, than in the region of α_M. Although there are some details of this map which are somewhat doubtful (see Section VII,C), this first attempt paved the way for a systematic application of potential functions to the theory of polypeptide conformations. We shall consider below the nature of the nonbonded interactions and the available theoretical methods of describing these in some detail, because these interactions play a vital part in the formation of particular conformations. In addition, other types of interatomic interactions, as well as the effects of intramolecular distortions will also have to be taken into account. A brief description of the theory of these is also considered in the subsections that follow. The application of these ideas to the various cases will be discussed in Section VII. A very detailed review of potential energy calculations of conformations of polypeptides by Scheraga (1968) is now in press.

B. Total Potential Energy of a Conformation

In general, the nonbonded forces operating between bonded atoms may be classified into two groups—one which is operative essentially for short separations and the other for large separations between the atoms. We assume, as a first approximation, that the covalent bonds in the molecules which interact are not affected and as a consequence the interatomic distances and angles occurring in them are fairly well defined. The variations in these produced by the intermolecular interactions will also

---

FIG. 22. Conformational map of allowed regions for cis peptide units with C^β atoms in the L-configuration. ———, fully allowed for a pair of units, ---- partially allowed by extreme limits for a pair, ▬▬ fully allowed for a helix, - - - partially allowed for a helix. The allowed region for the helix contains the conformations observed in poly-L-proline I. (From Ramachandran and Venkatachalam, 1968.)

FIG. 23. Potential energy contours for the nonbonded energy of a helical chain of poly-L-alanine, calculated by DeSantis *et al.* (1965). The map has been redrawn to conform to our notation of ϕ and ψ. Note that the energy is lower in the region of α_P than that of α_M.

lead to opposing forces and a corresponding increase in energy, and these will also have to be considered in a complete theory.

The long-range nonbonded force between two atoms is essentially attractive, namely of the van der Waals interaction type between two neutral atoms, to which will be added the electrostatic interactions between the partial charges on the atoms, arising from covalent bond formation. The short-range forces, which are mainly repulsive in nature, occur due to the overlapping of the electron shells of the two atoms which come into close contact.

Although we thus divide the interatomic interaction into different types of forces, the distinction between these is arbitrary and, in intermediate situations, one may gradually go over from one into the other. When one attempts to obtain the total potential energy as the sum of the various contributions of this type, it is necessary to be careful that the same effect is not counted twice over. As a very good approximation, the total energy of a conformation V can be expressed as a sum of the energies due to various interactions in the form of Eq. (20).

$$V = V_a + V_r + V_{es} + V_{hb} + V_l + V_\tau + V_\theta \tag{20}$$

Here V_a is the van der Waals attraction term (also called the London term), V_r is the repulsive term, V_{es} is the electrostatic term, V_{hb} is the energy of hydrogen bond formation, V_l is the strain energy associated with deformation of bond length, V_τ is the corresponding energy for bond angle deformation, and V_θ is the energy of dihedral angle deformation (torsional distortion). In addition, there is also a possibility of hydrophobic effects when adjacent nonpolar groups are present in an aqueous medium.

In the sections that follow, we shall describe each of these terms in more detail and present them in a formulation that is amenable for calculation of the conformational energy.

C. Nonbonded Interaction—Attractive and Repulsive Forces

As already mentioned, there is an attractive force between neutral atoms or molecules at distances where the overlap between the electronic wave functions of the two is not appreciable. At shorter distances, the closed shells of electrons produce a net tendency for the interacting groups to separate to larger distances, thereby reducing the overlap (Born and Mayer, 1932).

1. Attractive van der Waals Forces

A theory of the attractive forces based on quantum mechanics was given by London (1937) who obtained an expression for the interaction

energy as a function of the separation of two atoms. These forces are usually called the van der Waals attraction, from the term arising from them which was first introduced into the gas equation by van der Waals. They are also called London dispersion forces, and they are responsible for bringing together inert gas atoms to a liquid state. In a classical way, the theory may be given as follows. The nuclei and the electrons of an atom will be in continuous relative oscillatory motion with respect to each other, which gives rise to a separation of the centres of the positive and negative charges leading to the formation of transient dipole moments. These transient dipoles interact with one another, and the form of the energy of interaction is given by $V_a = -A/r^6$, where the constant A is connected with atomic properties, in particular the polarizabilities of the two atoms which interact. The exact expression for A was derived by Slater and Kirkwood (1931; see Pitzer, 1959) and is given by

$$A = \frac{\frac{3}{2}e(\hbar/\sqrt{m})\alpha_1\alpha_2}{\sqrt{\alpha_1/N_1} + \sqrt{\alpha_2/N_2}} \tag{21}$$

where α_1 and α_2 are the atomic polarisabilities of the interacting pairs of atoms and N_1 and N_2 are the effective number of polarizable electrons on

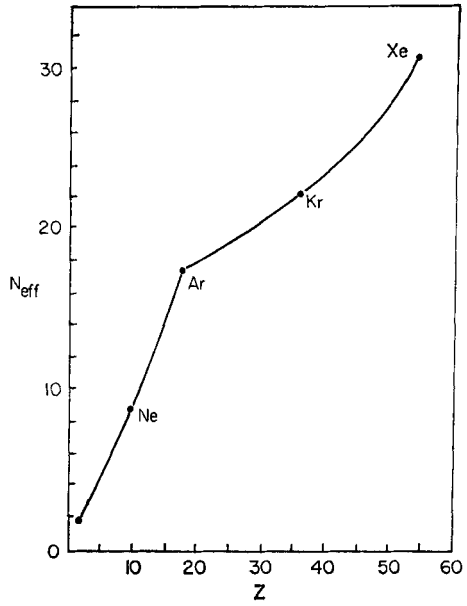


FIG. 24. Graph showing the variation of the effective number of polarizable electrons (N_{eff}) with the atomic number (Z). (From Scott and Scheraga, 1965.)

these atoms and e, m, and $\hbar$ $(= h/2\pi)$ are the usual fundamental constants. The values of N are available from Pitzer (1959) and have been given in a graph of N versus the atomic number Z by Scott and Scheraga (1965), which is shown in Fig. 24. The polarizabilities are obtained from measurements of refractivities and are adopted from Ketelaar (1958). The data relevant to polypeptides are given in Table IX.

TABLE IX
Polarizability Values α for Various Atomic Species

Atom	$\alpha \times 10^{24}$
H	0.42
O(peptide)	0.84
N(peptide)	1.15
C′(peptide)	1.30
CH_2	1.77

They are very much dependent on the chemical nature of the atoms—for instance the occurrence of a double bond increases the value by 0.58. Hence, in an actual application, one has to use the corrected value of the polarizability to take into account the magnitude of the bond orders (Brant and Flory, 1965a,b). The entries in Table IX are given after taking such considerations into account.

2. Repulsive Forces and Different Forms of Potential Functions

a. The Buckingham 6-exp Potentials. The question of choosing a suitable term to represent the repulsion of two neutral atoms has led to different forms of functions. Two of these have been particularly used, one an exponential dependence and the other an inverse power dependence. The Buckingham function uses the former in the form $V_r = B \exp(-\mu r)$, so that the nonbonded interaction takes the form

$$V_a + V_r = -A/r^6 + B \exp(-\mu r) \tag{22}$$

In this, the constant μ can be evaluated from collision experiments for rare gases and similar atoms and is taken to have a similar value for other bonded atoms having closed shells of electrons. The value of B is adjusted to give a potential minimum at $r = R$, the sum of the van der Waals radii of the interacting pair. A typical set of values of the van der Waals radii (Bondi, 1964) are given in Table X.

This type of function has been extensively used in the study of polypeptide conformation, namely by Scott and Scheraga (1965) and by Brant and Flory (1965a,b). Although Brant and Flory did not give

TABLE X
Van der Waals Radii of Atoms that Normally Occur in Polypeptide Chains

Atom	Van der Waals radii
H	1.20 Å
O	1.52
N	1.55
C	1.70
CH_2[a]	1.80

[a] CH_2 denotes a carbon atom with hydrogens, of the type CH_2 or CH_3, when the hydrogens are not located.

such data, a set of values of the constants A and B using their methods has been calculated by Ramachandran *et al.* (1966b), and is shown in Table XI. In this, following Brant and Flory, a constant value of $\mu = 4.6$ is used for all types of contacts.

The Buckingham type of function has a defect in that, although it rises to large positive values for short contacts, it becomes negative for still smaller interatomic distances and may occasionally lead to spurious results. Also for computational purposes, it requires the evaluation of

TABLE XI
Typical Constants A and B in the Buckingham Type of Potential[a]

Interaction[b]	A	B
H · · · H	46.8	0.829
C · · · H	165.8	7.79
N · · · H	156.0	5.34
O · · · H	124.1	3.83
CH_2 · · · H	226.9	14.9
C · · · C	599.9	92.4
N · · · C	571.2	60.5
O · · · C	461.6	43.3
CH_2 · · · C	822.8	187.0
N · · · N	546.9	40.4
O · · · N	446.1	29.4
CH_2 · · · N	783.6	121.0
O · · · O	368.9	21.7
CH_2 · · · O	633.9	86.0
CH_2 · · · CH_2	1128.0	382.0

[a] For use in Eq. (22) when r is measured in Ångstroms and V is in kcal/mole. μ is taken to be 4.6 for all interactions.

[b] CH_2 denotes a carbon atom with hydrogens, of the type CH_2 or CH_3, when the hydrogens have not been located.

an exponential term which takes time. In view of these, another form of the repulsive potential, namely one of the form B/r^{12}, is found to be more convenient.

b. The Lennard-Jones Potential. This form of the potential function makes use of a term proportional to r^{-12} for the repulsive part, so that $V_r = B/r^{12}$. Since the attractive part will have the form $-A/r^6$ as already mentioned, the total interaction, which is the sum of the two and is called the Lennard-Jones 6-12 potential, takes the form

$$V_a + V_r = -A/r^6 + B/r^{12} \tag{23}$$

The constant A can be calculated exactly as mentioned earlier from the polarisabilities of the interacting pair of atoms. The constant B can be calculated by requiring that the minimum value of the potential occurs when the separation between the pair of atoms is equal to the sum of the van der Waals radii, which may be denoted by R. This yields the relation $B = \frac{1}{2}AR^6$, so that

$$V = -\frac{A}{r^6}\left[1 - \frac{R^6}{2r^6}\right] \tag{24}$$

The minimum value of the potential V_m, which occurs at $r = R$ is seen from Eq. (24) to be equal to

$$V_m = -\frac{1}{2}\frac{A}{R^6} = -\frac{B}{R^{12}} \tag{25}$$

Thus, in terms of the minimum energy and the distance corresponding to it, the Lennard-Jones 6-12 function may be rewritten in the form

$$V_a + V_r = -2V_m\frac{R^6}{r^6}\left[1 - \frac{R^6}{2r^6}\right] \tag{26}$$

The form $V = -A/r^6 + B/r^{12}$ for the nonbonded interaction has been used recently by Scott and Scheraga (1966c) and a set of values of A and B as used by them for various interacting pairs is given in Table XII. (The values of A in Tables XI and XII should agree, but they differ essentially because of the different values of α used by the different authors for the various atoms.) The same form of the potential function has also been used by Brant *et al.* (1967).

In this formulation [Eq. (26)], the two parameters V_m and R are dependent on the species of atoms which interact. However, one may postulate some law governing the variation of V_m with the species of atoms or even consider it to be a constant. In the latter case, the curve showing the variation of V with distance will be the same for all pairs

of atoms, but shifted along the axis of r depending on the value of R. An idea of this type was adopted by Kitaigorodsky (1961, 1965), to obtain what he called a universal function and is discussed below.

c. The Universal Function of Kitaigorodsky. Kitaigorodsky (1961, 1965) developed a form of the Buckingham type of function with the idea of theoretically obtaining the lattice constants and crystal structures of simple organic crystals. For this purpose he simplified our

TABLE XII
Typical Constants A and B in the Lennard-Jones Type of Potential[a,b]

Interacting pair	R	A	$B \times 10^{-4}$
C · · · C	3.40 Å	370	28.6
C · · · N	3.25	366	21.6
C · · · O	3.22	367	20.5
C · · · H	2.90	128	3.8
N · · · N	3.10	363	16.1
N · · · O	3.07	365	15.3
N · · · H	2.75	125	2.7
O · · · O	3.04	367	14.5
O · · · H	2.72	124	2.5
H · · · H	2.40	46.7	0.45

[a] From Scott and Scheraga (1966c).
[b] For use in Eq. (23), to give V in kcal/mole where r is measured in Ångstroms.

general equation to be of the form $V = V_a + V_r + V_{es}$. Thus, the interatomic force between two atoms could be represented by a potential produced by the nonbonded interaction of a pair of "universal neutral atoms" and characteristic residual charges Xe at the atomic centres, where X would in general be different from Z, the atomic number. For the universal neutral atom, Kitaigorodsky used the 6-exp function as the basis and putting $z = r/R$ and $\alpha = \mu R$ in Eq. (22), he obtained $V_a + V_r$ to be of the form

$$V = V_{2/3}\left[\frac{1}{z^6} - \frac{6}{\alpha}e^{\alpha}e^{-\alpha z}\right]\left[11.4 - \frac{6}{\alpha}e^{\alpha/3}\right] \tag{27}$$

where $V_{2/3}$ is the value of V at $r = \frac{2}{3}R$. For the types of interactions C . . . C, C . . . H and H . . . H, Kitaigorodsky assumed that $V_{2/3} =$ 3.5 kcal/mol and $\alpha = 13$ and with these values we have the simple relation (Venkatachalam and Ramachandran, 1967) as follows:

$$V = V_a + V_r = 3.5(8600e^{-13z} - 0.04/z^6) \tag{28}$$

Equation (28) is different from that of Brant and Flory in that it leads to a value of μ different for the different types of interactions, unlike a

constant value of $\mu = 4.6$ used by them. The equation has the merit that there is only one parameter that requires to be specified for the calculation of the total nonbonded interaction, namely, the equilibrium distance R. Kitaigorodsky used certain values for these distances for the three types of interactions that he used. It was found by Dr. V. S. R. Rao in this laboratory (private communication) that as a working rule, Kitaigorodsky's values of R can be fitted to the condition that the potential energy is zero for the sum of the van der Waals radii mentioned in Table X. Extending this rule to the interactions also of oxygen and nitrogen with various atoms, and using the van der Waals radii of these atoms also, Venkatachalam and Ramachandran (1967) have given a list of values of R that could be used in the Kitaigorodsky equation for different pairs of atoms. It may be mentioned that Kitaigorodsky's equation with these parameters has been found to explain very well the conformation of various sugar residues and polysaccharides by Rao *et al.* (1967). A variant of the Kitaigorodsky functions with a smaller set of values for R has also been suggested by Venkatachalam and Ramachandran (1967) (see Section VII,C). In this case, R is made equal to the sum of the van der Waals radii of Table X.

D. Electrostatic Energy

As mentioned earlier, there are partial charges on atoms even in their covalently bound state and therefore the interaction of molecules must take into account the electrostatic forces between these charges, which are easily described by a coulombic law. The calculations may be made in one of two ways—either they may be expressed as charge-charge interactions or as dipole-dipole interactions. If the former method is used, it is possible to calculate the net charges on the various atoms of a molecule from the tables of bond moments that are available (e.g., Smyth, 1955). The electrostatic energy V_{es} is then

$$V_{es} = -\frac{e_i e_j}{\epsilon r_{ij}} \tag{29}$$

where e_i and e_j are the charges on the atoms i and j in neighbouring molecules or groups and ϵ is the effective dielectric constant of the medium. The bond moments of the various bonds found in the peptide group are listed in Table XIIIA as given by several authors. Using these and making suitable assumptions, two or three workers have listed the electronic charges on the atoms of the peptide group which are given in Table XIIIB. The last column of the table gives a set of approximate values, which can be considered to be as good as any that can be estimated at the present time.

TABLE XIIIA
Bond Moments of the Bonds Occurring in the Backbone of the Peptide Unit[a]

Bond	References[b] (1)	(2)	(3)	(4)
C=O	2.30	2.48	2.35	—
C′—N	0.22	0.21	—	0.47–0.62
N—H	1.31	1.31	1.35	1.08–1.21
C^{α}—N	0.22	—	—	0.47–0.62

[a] Values are in Debyes.
[b] References: (1) Smyth (1955); (2) Scheraga *et al.* (1967); (3) Brant *et al.* (1967) (4) Del Re *et al.* (1963).

The dipole moment of the peptide group is known to be 3.7 D (Brant and Flory, 1965a,b). Though the direction of the moment is not exactly certain, in calculating the energy of a long polypeptide chain, one may as an approximation replace the monopole charges by an effective dipole moment located at the centre of the peptide group, as was done by Brant and Flory (1965a,b). However, in evaluating the interactions between atoms at distances of a few Ångstroms, the monopole method is likely to yield more accurate values of the potential than one using the dipole moment.

There is considerable doubt as to the value of the dielectric constant ϵ which must be used in Eq. (29). Brant and Flory (1965a,b) used a value of $\epsilon = 3.5$ in calculating the potential energy map of a pair of peptide units. Scott and Scheraga (1966c) used a value of 4.0 for this parameter. A theoretical study of the effective dielectric constant has been made recently by Ramachandran and Srinivasan (1968), using an idealised model of a spherical cavity containing the interacting atoms.

TABLE XIIIB
Charges Associated with the Bonded Atoms in the Peptide Unit[a]

Atom	References[b] (1)	(2)	(3)	Suggested value
C′	0.425	0.449	0.394	0.4
O	−0.381	−0.416	−0.394	−0.4
N	−0.300	−0.305	−0.281	−0.3
H	0.260	0.272	0.281	0.3

[a] The numbers are in fractions of the electronic charge.
[b] References: (1) J. A. Schellman (personal communication); (2) Scheraga *et al.* (1967); (3) Brant *et al.* (1967).

It is then found that a value of ϵ in the range 3 to 5 is quite reasonable for the interaction of static charges even though the medium outside the cavity may have a large dielectric constant of the order of 20 to 100. A value of $\epsilon \simeq 1$ is reasonable for the van der Waals and similar interactions, which involve the dielectric constant at optical frequencies.

E. Hydrogen Bond Energy

This is the energy of formation of the so-called hydrogen bond between two electronegative elements X and Y, of the type

$$\text{X—H} \cdots \text{Y—Z},$$

where X and Y can be N, O, Cl, etc. Although the nature of the hydrogen bond has not been clearly understood, it has been pictured by many as being essentially an electrostatic attraction of the partly polarised hydrogen atom covalently bonded to the atom X towards the negative charge density around Y. This attraction energy is of the order of 3 to 8 kcal/mole, which is significantly larger than the van der Waals attraction and hence should be explicitly included in the energy calculations.

1. Theoretical

The first attempt to include such energies in the potential function for the calculation of polypeptide and protein conformations was made by DeSantis *et al.* (1965), who used the well-known Stockmayer relation (Stockmayer, 1941) for the interaction between polar molecules. This function is represented by

$$V = -2V_m \frac{R^6}{r^6}\left(1 - \frac{R^6}{r^6}\right) - \frac{\mu_1\mu_2}{r^3} g(\theta_1,\theta_2,\phi_1 - \phi_2) \qquad (30)$$

where the first term represents the nonbonded interaction between the two dipolar groups [identical with Eq. (26)], and the second term the electrostatic interaction between the dipoles μ_1 and μ_2 attached to the groups X—H and Y—Z. The directional character of the dipole interaction is taken care of by the angular dependence of g, given by

$$g(\theta_1,\theta_2,\phi_1 - \phi_2) = 2\cos\theta_1\cos\theta_2 - \sin\theta_1\sin\theta_2\cos(\phi_1 - \phi_2) \qquad (31)$$

where θ_1 and θ_2 are the angles which μ_1 and μ_2 make with the line joining them and ϕ_1 and ϕ_2 are the azimuthal orientations of the two dipoles. DeSantis *et al.* (1965) have assumed that the two dipoles N—H and O=C are centred on the H and O atoms respectively. In such a case, the first term of the Stockmayer equation represents the nonbonded interaction of the hydrogen atom of the N—H group with the oxygen

atom of the C=O group. Correspondingly θ_1 and θ_2 would describe the angles which N—H and C=O make with the line joining H and O and $(\phi_1 - \phi_2)$ the angle between the plane containing the atoms N, H, and O, which take part in the hydrogen bond and the plane containing the atom N and the acceptor C=O bond respectively. The parameters V_m, R, μ_1, μ_2 were obtained by them by approximately minimizing the first term of the equation at the van der Waals distance between the atoms O and H, and the whole function at a distance of 2.85 Å for N . . . O, with the C=O and N—H bonds being colinear.

A more accurate description of the hydrogen bond energies is available from the work of Lippincott and Schroeder (1955) and by Schroeder and Lippincott (1957) for systems of the type O—H . . . O, N—H . . . O, N—H . . . N, etc. This has been adopted by Scott and Scheraga (1966a) in their calculations on the conformation of polypeptides.

Lippincott and Schroeder have approximated the hydrogen bond to a system of two diatomic molecules of the form X—H and H . . . Y and have employed features such as the X—H frequency shift, bonded X—H distance, van der Waals repulsion term and an electrostatic term. These are described by the usual diatomic potentials of the form

$$V = D\{1 - \exp[-n(r - r_0)^2/2r]\} \tag{32}$$

where D is the dissociation energy, r is the actual bond length, r_0 is the equilibrium bond length and n is a parameter. Two such terms, one for X—H and the other for H . . . Y are included. A van der Waals repulsion term of the type $Be^{-\mu d}$ between the atoms X and Y and an electrostatic term $-A/d^m$ are also included (d is the distance between the atoms X and Y). Thus the final form of the function is

$$\begin{aligned} V_{hb} = {} & D\{1 - \exp[-n(r - r_0)^2/2r]\} \\ & - D^*\{1 - \exp[-n^*(d - r - r_0^*)^2/2(d - r)]\} + Be^{-\mu d} - A/d^m \end{aligned} \tag{33}$$

Here D is the strength of the X—H bond, D^* of the H . . . Y bond, r is the distance X—H and r_0 its equilibrium value and d is the X . . . Y distance, r_0^* is the equilibrium value for the distance H . . . Y and n and n^* are related to the ionization potential. The parameter m determines the power law for the electrostatic term. The values of these parameters for the commonly occurring hydrogen bonded systems given by Lippincott and Schroeder are reproduced in Table XIV. This function has been modified by Moulton and Kromhout (1956) to include terms which depend on the angle θ_1 between bond X—H and line X . . . Y and θ_2 between bond H . . . Y and the line X . . . Y. Actually, the modification suggested is that the factor $\exp[-n(r - r_0)^2/2r]$

TABLE XIV

Parameters for Calculating the Energy of Hydrogen Bond Systems of the Type X—H · · · Y[a]

	O—H · · · O	N—H · · · O	O—H · · · N	N—H · · · N	O—H · · · Cl	N—H · · · Cl
D(kcal/mole)	118	104	118	104	118	104
D^*(kcal/mole)	81	74	79	72	82	74
n(A^{-1})	9.18	9.30	9.07	9.30	9.07	9.30
n^*(A^{-1})	13.32	13.15	13.49	13.49	13.07	13.07
r_0(A)	0.97	1.014	0.97	1.014	0.97	1.014
r_0^*(A)	0.97	0.97	1.014	1.014	1.275	1.275
u(A^{-1})	4.80	4.80	4.80	4.80	4.80	4.80

[a] The data are adapted from Schroeder and Lippincott (1957).

be multiplied by $\cos^2\theta_1$ and the factor $\exp[-n^*(d - r - r_0^*)^2/2(d - r)]$ by $\cos^2\theta_2$. Scott and Scheraga (1966a) used the parameters given in Table XIV with $m = 6$ in their calculations. However, as the N—H bond length was held fixed in the peptide, the first term of the expression was omitted and the Moulton-Kromhout modification was included in the second term. For the latter purpose, they assumed that the normal bonding direction lies along the direction of the lone pair electron of the oxygen of the C=O bond. Two such terms were included, one for each lone pair.

While including the energy term for the hydrogen bond in the evaluation of the total energy, care must be exercised in that the nonbonded terms for the interactions X . . . Y and H . . . Y should not be included in the van der Waals energy term also. The term $-A/d^m$ in Eq. (33) involves an electrostatic interaction and therefore no additional electrostatic term V_{es} is necessary for this interaction.

2. Observational

An enormous amount of literature has grown up about the hydrogen bond and it will not even be possible here to briefly summarize the state of the present knowledge. An excellent book on the subject is that of Pimentel and McClellan (1960) and a recent review is one by Hamilton and Ibers (1968). Although there is general information from infrared data that the A—H frequency is shifted to lower and lower values with decrease of the distance A . . . B, it is not obvious that the shorter bond is always of lower energy. Equations (30) and (33) predict a minimum at some intermediate distance. In fact, for a particular type of bond, e.g., NH . . . O, the length is distributed around a most likely value, in this case about 2.9 Å (Pimentel and McClellan, 1960, pp. 286–287; Ramakrishnan and Ramachandran, 1967). The mean distance is even larger if we consider only bonds between NH (as distinct from N^+H) groups and C=O (as distinct from COO^-) groups. The latter authors also found in addition that the angle H—N . . . O is in general small, more than 70% of the examples having a value less than 20°. On the other hand, the angle between the directions C—O and O . . . N is not in general small, and there is a pronounced maximum in the distribution at a value of about 50° to 60°, corresponding to the lone pair orbital of the oxygen, as mentioned above.

It would be desirable to explore further the theory of hydrogen bonding and obtain at least semiempirical formulae for its energy which fit observations like these and which contain a small number of constants. It should go over into the nonbonded interaction value at large enough distances.

F. Strain Energy of Bond Length, Bond Angle, and Dihedral Angle Variation

In trying to fit in with certain conformations, the geometry of a molecule may itself be slightly altered. Such effects are particularly present in closely packed structures and cyclically linked structures, for example, cyclic peptides and small peptides closed by S—S bonds. In such structures, the bond lengths (l), bond angles (τ), and dihedral angles (e.g. ω) would change and would lead to increase in potential energy. The energy associated both with bending and stretching of bonds can be obtained to a good approximation from the force constants of the molecules associated with the relevant quantities, which are often available from data of infrared spectroscopy (Westheimer, 1956; Herzberg, 1945). If K_i is the force constant associated with a deformation Δq_i (i.e., $F_i = -K_i \Delta q_i$, where Δq_i may be the change in length (l) of a bond or the change in valence angle (τ)), then the strain energy V_i is given by

$$V_i = \tfrac{1}{2} K_i (\Delta q_i)^2 \tag{34}$$

Values of K_i are available in the literature on infrared spectroscopy, but not in a direct way except for a few examples and we shall consider only typical values relevant for peptide units.

1. Bond Length Variation

Miyazawa *et al.* (1958), have considered the infrared spectrum of *N*-methylacetamide and worked out the force constants for the stretching of the various bonds. We shall indicate the corresponding force constants by K_l and express them in kcal/mole/Å^2. The increase in energy for a deformation Δl (in Ångstrom) is then given by

$$V_l = \tfrac{1}{2} K_l (\Delta l)^2 \tag{35}$$

Values of K_l are also available from a study of the normal vibrations of polyglycine I by Fukushima *et al.* (1963). The averages of these two

TABLE XV
Stretching Force Constants K_l (in kcal/mole/A^2) for the Bonds Occurring in the Peptide Unit

C^α—C′	C′=O	C′—N	N—H	N—C^α	References[a]
500	1200	800	800	500	(1)
600	1400	1100	900	600	(2)

[a] References: (1) Mean of Miyazawa *et al.* (1958), and Fukushima *et al.* (1963). (2) T. Miyazawa (personal communication, 1967).

(rounded off) are given in the first row of Table XV. Very recently, T. Miyazawa (personal communication) has made a more thorough study of the infrared spectrum of *N*-methylacetamide, *N*-methylformamide, and their isotope-substituted derivatives and obtained the data shown in the second row of Table XV.

It is seen that all the values of K_l are of the order of 500 to 1200 kcal/mole/Å^2, so that for a change of 0.1 Å, the increase in energy is of the order of 5 kcal/mole. In view of this, bond strains greater than 0.05 Å are not expected to occur and such a deformation can be neglected in all calculations of the conformational energy.

2. Bond Angle Variation

Both Miyazawa *et al.* (1958), and Fukushima *et al.* (1963) have given values of the bending force constants K_τ which give the energy for a deformation $\Delta\tau$ to be equal to

$$V_\tau = \tfrac{1}{2}K_\tau(\Delta\tau)^2 \tag{36}$$

K_τ is expressed in units of kcal/mole ($\Delta\tau$ is to be measured in radians). The values for the different angles in the backbone of the peptide unit vary by about 30% about the mean value, and since they are not very accurate, a value of 80 kcal/mole (which is the mean) may be used for K_τ for all the angles. Incidentally, this agrees very well also with the value of K_τ for the angle H—C—H in methane (as calculated using the data in Herzberg, 1945, p. 182).

Thus, a change of 5° in bond angle will produce an increase in energy of about 0.3 kcal/mole, so that deformations of this order are to be expected to occur, if a conformation cannot be otherwise accommodated because of nonbonded repulsive interactions.

3. Torsional Potential

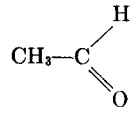
The torsional potential refers to the preferential relative orientation of the different groups attached to either end of a single bond. In the early days, it was supposed that rotations about single bonds were free to occur; but Kemp and Pitzer (1936) found that they could explain the observed entropy of the ethane molecule only by postulating that not all states of rotation about the central C—C bond are equally probable. This discovery of the phenomenon of hindered rotation in ethane has led more generally to the postulation of torsional potentials for rotations about single bonds. It is necessary to mention that this potential is something characteristic of the bond itself and arises from the interaction between the orbitals of the bonded atoms. Thus, the contributions of nonbonded repulsion and the electrostatic interaction

alone between the CH—CH bonds in ethane are not sufficient to explain the energy barrier of close to 2.7 kcal/mole (Wilson, 1959). In the case of double bonds, the torsional potential would be much larger, because the atoms attached to the two bonds linked by double bonds would be planar and the distortion of this planar configuration would require much larger energies. The term V_θ is then quite important.

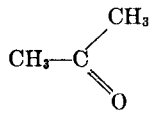
a. Single Bonds. The usual form of the torsional potential for a single bond of a tetrahedrally bonded atom is taken to be of the form

$$V_\theta = \tfrac{1}{2}V_1(1 - \cos 3\theta) + \tfrac{1}{2}V_2(1 - \cos 6\theta) + \cdots \qquad (37)$$

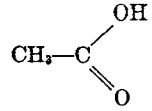
where the barriers V_1, V_2, etc. have a three-fold, six-fold, etc. symmetry. The barriers of a six-fold potential are comparatively small, e.g., the calculated value is 0.006 kcal/mole in nitromethane (Millen, 1962). Hence, for the rotations ϕ and ψ in a polypeptide, only a three-fold barrier seems to be relevant (Scott and Scheraga, 1966c; Brant and Flory, 1965a,b). Direct evidence showing the presence of barriers about the N—C^α and C^α—C′ bonds in the peptide chain is not available, though information about this factor in a series of analogous model compounds is known. The barrier about the C^α—C′ bond may thus be esetimated by comparison with the values for the methyl rotation in carboxyl compounds, all of which have a three-fold potential (Dale, 1966). If the two atoms in the planar group attached to the atom concerned are dissimilar, as in this case of ϕ-rotation, e.g.,



(barrier 1.16 kcal/mole) or



(barrier 0.78 kcal/mole) a value of 1.0 kcal/mole seems to be reasonable for $V_{\phi 1}$. When they are similar (N and O), as in the ψ-rotation, a typical example is acetic acid



for which the barrier is 0.48 kcal/mole, and a reasonable value for $V_{\psi 1}$ is thus 0.5 kcal/mole. The minima for V_ψ would occur at $\psi = 0°$,

120°, 240° and for V_ϕ at $\phi = 60°$, 180°, 300°. However, for rotations around the side chain C—C bonds, the barrier would be expected to be nearly 3 kcal/mole, as in ethane. It may be mentioned that Brant and Flory (1965a,b) used values of 1.5 and 1.0 kcal/mole for $V_{\phi 1}$ and $V_{\psi 1}$ respectively, while Scott and Scheraga (1966c) used values of 0.6 and 0.2 kcal/mole respectively. (See below for correction for nonbonded effects to the above recommended values of 1.0 and 0.5 kcal/mole.)

b. Double Bonds. In the case of a regular double bond like C=C, the barrier is very high and possible rotation about such bonds may be neglected. However, the ω-rotation about the partial double bond C—N in the peptide unit is important and should be considered. In this case, the torsional potential has a variation of the form $\frac{1}{2}V_{\omega 1}(1 - \cos 2\omega)$. For small values of $\Delta\omega$, not far from the minimum, this gives the increase in energy

$$\Delta V_\omega = V_{\omega 1}(\Delta\omega)^2 \tag{38}$$

The value of $V_{\omega 1}$ has been estimated from different methods and some of the available data are summarized in Table XVI. They deal mostly with model compounds of amides. The barrier itself has been deter-

TABLE XVI
Energy Barrier to Internal Rotation About the Peptide Bond C—N

Substance	Free energy of activation ΔF^* (kcal/mole)	Technique used	Reference[c]
N,N-Dimethylformamide	21.0	nmr	(1)
N,N-Dimethylformamide	22.0	nmr	(2)
N,N-Dimethylformamide	19.6	nmr	(3)
N,N-Dimethylacetamide	17.4	nmr	(1)
N,N-Dimethylacetamide	19.0	nmr	(2)
N,N-Diethylacetamide	16.9	nmr	(3)
Poly-L-proline[a]	24.4	Optical rotation kinetics	(4)
Poly-*O*-acetyl-L-hydroxyproline[a]	22.5[b]	Optical rotation kinetics	(5)
N-Methylformamide and *N*-Methylacetamide:	17.5	Infrared	(6)
Theory (~1 ev)	22.6	Quantum mechanics	(7)

[a] Calculated from the rate constants using absolute theory of reaction rates.

[b] Mean value.

[c] References: (1) Rogers and Woodbrey (1962); (2) Gutowsky and Holm (1956); (3) Hammaker and Gugler (1965); (4) Steinberg *et al.* (1960); (5) Downie and Randall (1962); (6) T. Miyazawa (personal communication); (7) Pauling and Sherman (1933).

mined by most of the methods, but the variation in energy of the form of Eq. (37) for small deviations from planarity is best obtained from infrared data. The latest value of this type (T. Miyazawa, personal communication) is 17.5 kcal/mole. In view of the uncertainties in the available knowledge, a rounded off value of 20 kcal/mole is recommended for $V_{\omega 1}$. Thus, a nonplanar distortion ω of 10° would lead only to an increase in energy of 0.65 kcal/mole and is quite likely to occur.

It should be remembered that in using the torsional potentials (which contain also the nonbonded term), the latter should not be used again twice over. The steric factor in ethane is about 0.4 kcal/mole (Eyring, 1932b) and hence the true value of V_1 in Eq. (37) is only 2.5 kcal/mole. If such effects are substracted for V_ϕ and V_ψ, the true values of $V_{\phi 1}$ and $V_{\psi 1}$ would be close to those used by Scott and Scheraga (1966c) mentioned above.

c. Out-of-plane Distortions of O and H. T. Miyazawa (personal communication) states that it is very difficult to estimate from infrared data the force constants for the individual out-of-plane distortions of O in the C=O and H in the N—H groups, apart from that associated with ω. We shall not therefore attempt to give any constants associated with the distortions υ^O and υ^H but merely point out their existence.

G. Hydrophobic Effects

Another type of nonbonded interaction which may play a role in stabilizing the configuration of a polypeptide or protein chain is known as the hydrophobic effect. This effect is of importance when the polypeptide chain is considered in an aqueous medium and refers to the interaction between the side-chain groups of the chain. Nearly all proteins have about 20–30% content of nonpolar side chains like valine, leucine, isoleucine, and phenylalanine. These groups have low affinity to water and would tend to eliminate contact with water molecules (hence the name hydrophobic) and to cluster together so as to stabilize the chain configuration. Such an effect is known as a hydrophobic bond (Kauzmann, 1959)—a better term would be hydrophobic interaction. It is believed that both van der Waals interaction and change in the ordering of water molecules near the surface of the hydrophobic groups (Frank and Evans, 1945) are responsible for this effect. The thermodynamics of such a situation have been worked out by various persons (Kauzmann, 1959; Némethy and Scheraga, 1962a) using hydrocarbons in aqueous solution as model systems. Recently, Scheraga *et al.* (1967), report that they have worked out potential energy expressions to take into account the effect of water in hydrophobic bonding between nonpolar groups. Further details are given by Gibson and Scheraga (1967). In general,

the introduction of water molecules would tend to drive the nonpolar groups towards the interior of the protein molecule and the polar groups towards the periphery (as noticed for example in myoglobin and lysozyme—see Section VII,C). A recent review on hydrophobic interactions is one by Némethy (1967).

H. Summary

Summarizing this section, we may say that the potential energy of intermolecular interaction, or that arising from the interaction of a peptide unit with its neighbours in a polypeptide chain, can be expressed by Eq. (20). In this, the first two terms, namely $V_a + V_r = V_{nb}$, may be calculated using either the 6-exp form (Eq. 22) or the 6-12 form (Eq. 23). As will be seen from Section VII,C, the different forms of this expression and the different values of the parameters used in them lead only to comparatively minor differences in the results. The electrostatic interaction V_{es} is best calculated by using the monopole approximation and Eq. (29). The theory of the hydrogen bond energy is the one that is most uncertain, and the formulae proposed have first to be tested with the available data before they are adopted for predicting conformations. Unfortunately, the experimental data also are not too precise as regards the stabilising energy of the bond and its variation with bond distance and angle.

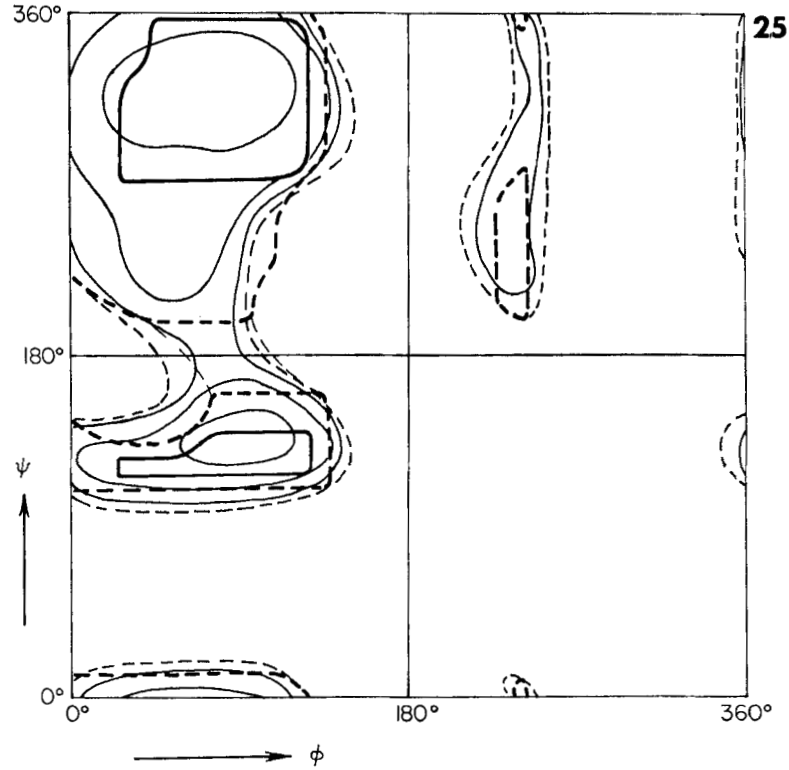
In the same way, the distortional energies are also known only roughly from theory. However, they are important, since appreciable distortions can occur only with slight increase in energy. As will be seen from Section VII,E, the dihedral angle ω could vary from the value 0°, for a planar peptide unit, by as much as 20° or even more. Similarly, the out-of-plane distortions v^O and v^H (as defined in Section II,F,1) are also likely to occur with fair facility, but in general, they would be expected to go hand-in-hand with the ω-distortion.

Hydrophobic effects are expected to play an important role in the conformation of proteins in solution, and what is lacking here is a set of quantitative formulae, although the qualitative theory seems to be clear enough.

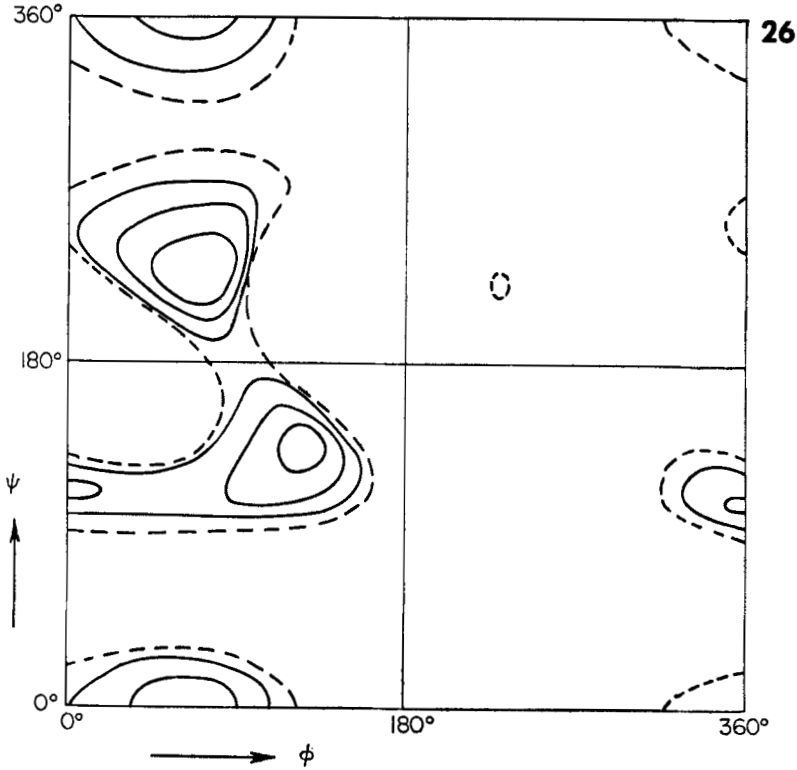
In spite of these lacunae in the theory, we have enough theoretical basis to predict the conformations to be expected in simple examples and the results of such studies are described in Section VII.

VII. Tests of the Different Potential Functions

In the previous section, we discussed in detail the various terms that may contribute to the total conformational energy of a polypeptide chain. In this section we shall discuss the application of the different potential



25



26

functions to polypeptide conformation. We shall also discuss the effect of including the various energy terms, in steps, in the calculation of the energy of a pair of linked peptide units and of a polypeptide chain.

A. Conformational Energy of a Pair of Linked Peptide Units

The first attempt to evaluate the conformational energy of a pair of linked peptides was made by Brant and Flory (1965a,b) and later by Ramachandran *et al.* (1966b), and by Scott and Scheraga (1966c). Brant and Flory and Scott and Scheraga used nonbonded energy terms, torsional potential and electrostatic potential terms to evaluate the total energy of the dipeptide. Ramachandran *et al.*, on the other hand, computed only the nonbonded interactions and compared these with the observations. It would therefore be of interest to discuss first the nonbonded energy of a pair of peptides and see how well the results compare with those of the hard sphere approach. Then we shall consider how the other energy terms influence the total conformational energy.

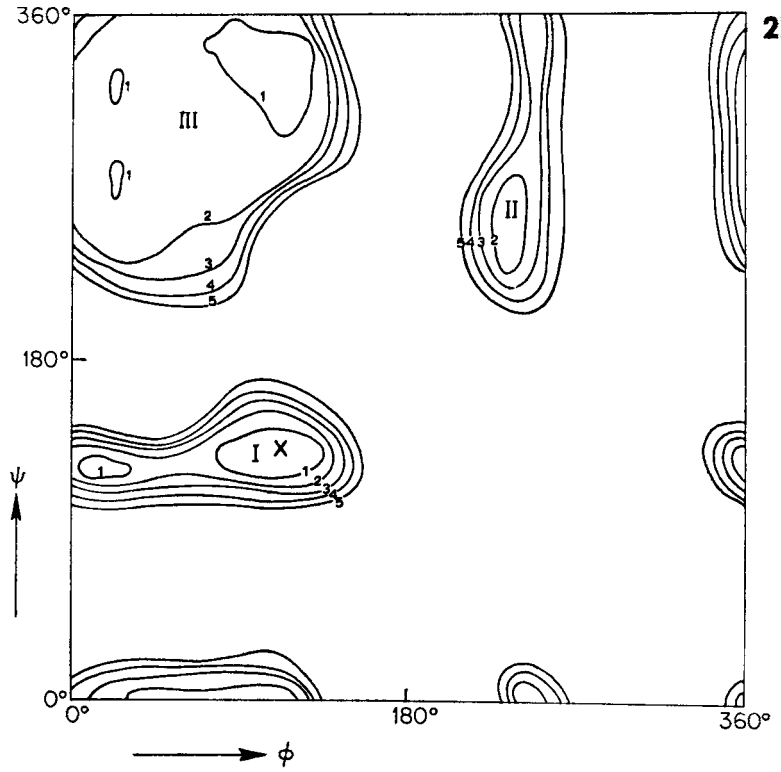
1. Effect of Nonbonded Interactions Alone

We will first discuss the case of a pair of peptide units when a C^{β} atom is present, as has been done by Ramachandran *et al.* (1966b). The parameters used by them were discussed in Section VI,C and shown in Table XI. With these parameters, the nonbonded energy of interaction, V_{nb} between the atoms in the pair were calculated over the whole range of ϕ and ψ (0° to 360°) at intervals of 10°. All nonbonded interactions between atoms separated by more than two covalent bonds were included. This practice has been followed in all the investigations on potential energies of calculations made in the authors' laboratory, which are briefly referred to below.

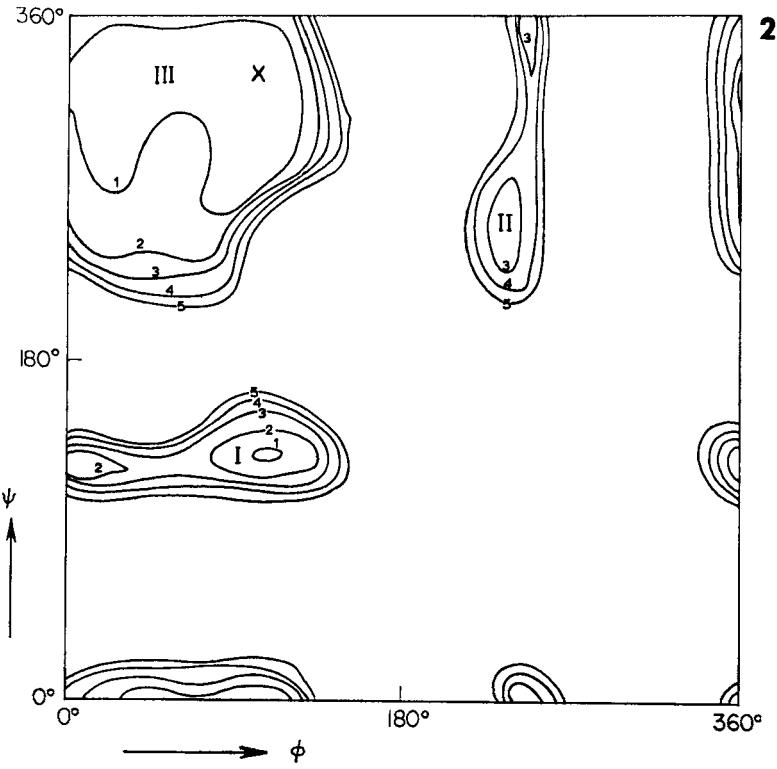
The contours of constant V_{nb} are reproduced in Fig. 25, in which the contact map is also superimposed. It will be noticed that the contour for $V_{nb} = 0$ nicely fits in with the outer limit contour of the contact map.

FIG. 25. Contours of constant nonbonded energy (V_{nb}) for a pair of peptide units linked at an alanyl α-carbon atom using the Scheraga-Flory potentials. (From Ramachandran *et al.*, 1966b.) The dashed contour corresponds to $V_{nb} = 0$ and the contours inside are at intervals of 1 kcal/mole, going downwards. The contact map is also shown superposed on the contours. Note the very good agreement between the extreme limit outline of the contact map and the zero contour of V_{nb}.

FIG. 26. Contour map of V_{nb} using the Liquori potentials (DeSantis *et al.*, 1965), similar to those in Fig. 25, reproduced from Ramachandran *et al.* (1966b). This potential map is in poor agreement with the contact map. In particular, it shows the region near (100°, 300°) to be unfavorable (having high energy).


360°
180°
0°
0°
180°
360°
ψ
φ
III
II
I X
1
2
3
4
5

27A


360°
180°
0°
0°
180°
360°
ψ
φ
III
X
II
I
1
2
3
4
5

27B

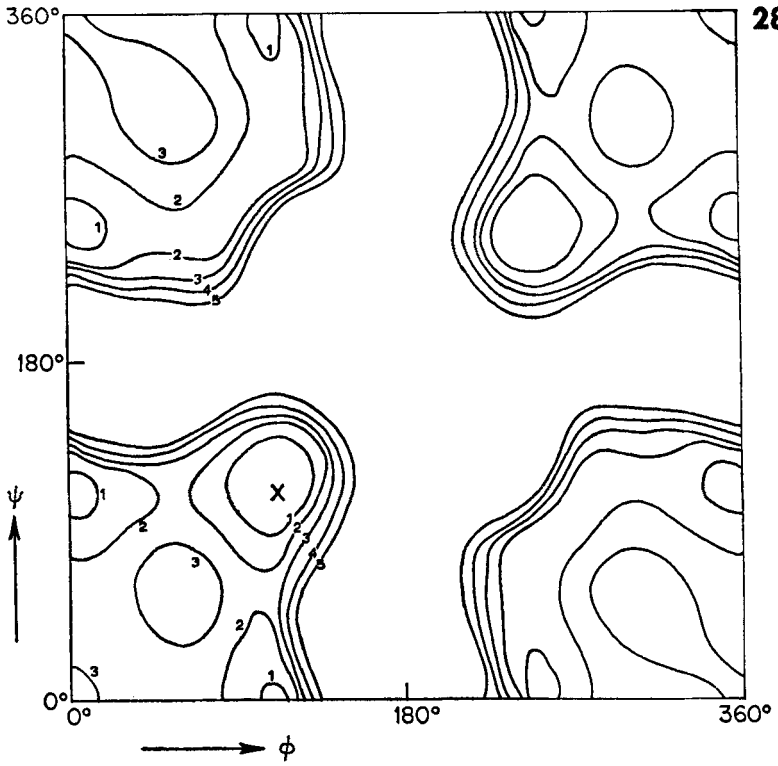
In particular, the bridging across $\psi = 180°$ in the left side of the figure is quite evident. The extension of the contour on the right-hand side up from about $\psi = 240°$ to about $\psi = 360°$ should also be noticed. The nice fitting of the contact map with the energy contour map shows the dominant role played by the repulsive terms of the nonbonded interactions in the total energy of a conformation.

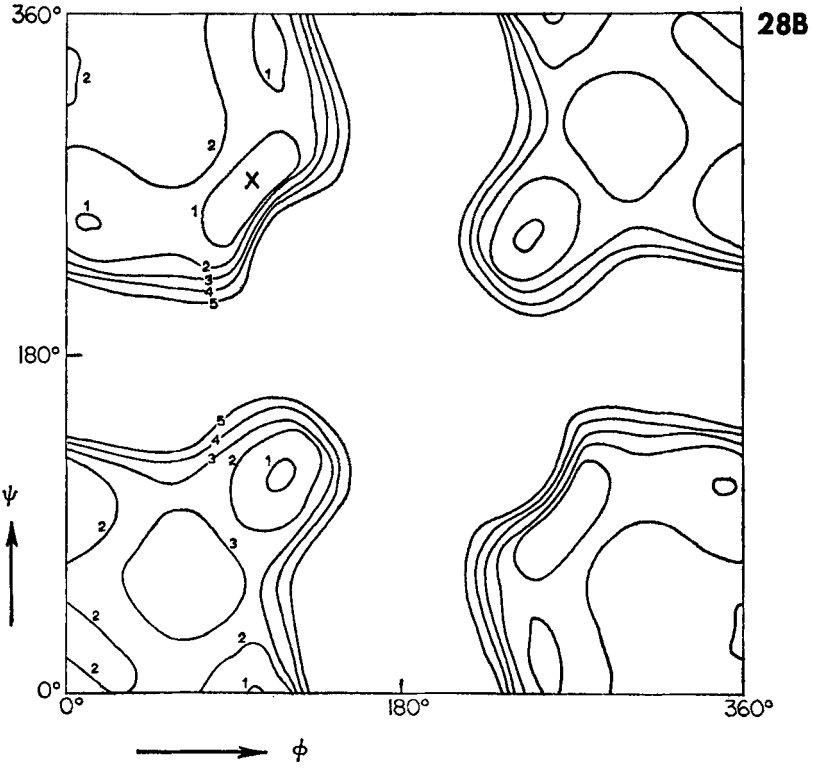
When the (ϕ, ψ) data of myoglobin and lysozyme are plotted, they are found to lie mostly within the zero contour. In lysozyme, in particular, the two deepest minima (regions I and III) are the regions which are most populated. Such a good agreement was not obtained for the contours calculated using the potential energy functions (nonbonded interactions only) adopted from DeSantis *et al.* (1965) and shown in Fig. 26. The difference in shape between the contours obtained by Ramachandran *et al.* (1966b), using the parameters employed by Scheraga and Flory on the one hand and by Liquori on the other, would appear to require further study.

2. Effect of Including Other Interactions to the Nonbonded Energy

a. Inclusion of Torsional and Electrostatic Potentials. The comparison of the contact map with that of the nonbonded energy map for a pair of linked peptides has already indicated the overriding importance of the repulsive forces. We will now see how far the torsional and electrostatic energy terms affect the energy constants in the ϕ-ψ plane. Such a comparison of the effect of including torsional and electrostatic energy terms in the total energy has been made possible by the results of Brant *et al.* (1967) and Flory (1967), who have published energy contour maps with and without electrostatic energy added to the nonbonded and torsional energies. The energy contour maps obtained by them for a pair of linked peptides with the C^β atom present are shown in Figs. 27A and B. Figure 27A shows the energy contour when the electrostatic interactions are not included but only the nonbonded and torsional energies are considered. The parameters used in these calculations have already been discussed in Sections VI,C,D,F. The minimum energy point in each case has been marked with an $\times$ and the contours have been drawn relative to the minimum at $\times$. The regions of low energy are marked as I, II, and III in both Figs. 27A and B.

FIG. 27. Contour maps of potential for a pair of peptide units linked at an alanyl α-carbon atom, according to Brant *et al.* (1967) and Flory (1967)—(A) using only nonbonded and torsional potentials; and (B) including also electrostatic interactions. Contours are at intervals of 1 kcal/mole, starting from the lowest one, and $\times$ indicates the point of lowest energy.


28A
360°
180°
0°
0°
180°
360°
ψ
φ


28B
360°
180°
0°
0°
180°
360°
ψ
φ

The general features of the two maps are essentially the same and are similar to Fig. 25. The main difference between Fig. 25 and Figs. 27A and B is that, in the former case, the two regions in the left-hand side are bridged across $\psi = 180°$, whereas in the latter cases they are not. This is further discussed below when the results of Scott and Scheraga (1966c) are compared with those of Ramachandran *et al.* (1966b).

However, the values of the actual minima in the allowed regions in the two maps are different. In Fig. 27A, the lowest minimum is observed in the region I near the right-handed α-helix, whereas in Fig. 27B it occurs in the region III of $\phi \simeq 100°$ and $\psi \simeq 330°$ near the collagen helix. This difference has been attributed by Brant *et al.* (1967) to the orientation of the peptide dipoles and their electrostatic interaction. The dipole moments are antiparallel for $\phi = \psi = 0$ and all along the diagonal $\phi + \psi = 0$. This negative energy along the line $\phi + \psi = 0$ causes the saddle point at about 90°, 300° in Fig. 27A to be replaced by a minimum in Fig. 27B at a slightly shifted position. But at the minimum of Fig. 27A corresponding to a conformation near the α_P-helix, the dipole moments are almost parallel and are repulsive and this raises the value to be as shown in Fig. 27B. Except for the changes in the position of the deepest minimum, the two maps are very similar.

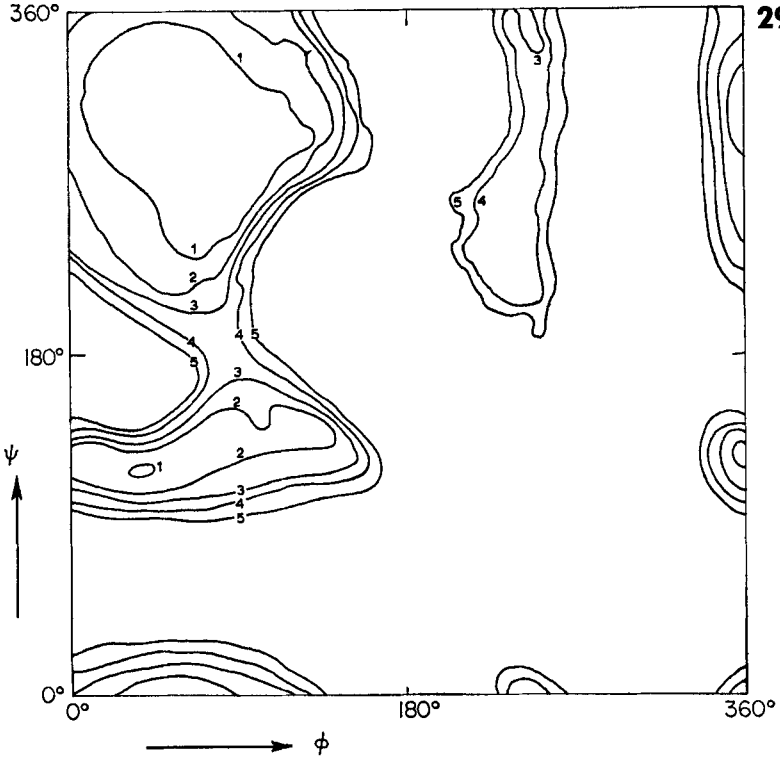
A comparison of Fig. 25 and Fig. 27A shows that small local minima are observed along the margin (near $\phi = 30°$) in Fig. 27A nearly at $\psi \simeq 120°$, 270°, and 330°. They are in part due to the torsional potential contributions. It is apparent that the actual positions of the minima would depend on both the contributions from the torsional potential and nonbonded interactions.

Maps of the conformational energy for a glycyl α-carbon atom calculated by Brant *et al.* (1967), are shown in Figs. 28A and B. Figure 28A shows the contour map without electrostatic interactions, while in Fig. 28B this contribution is included, as has been done for the case of the alanyl α-carbon atom. Comparison of these with Fig. 14 obtained by Ramakrishnan and Ramachandran (1965) from contact criteria shows that the general features are essentially the same, again reflecting the important role of repulsive forces in the conformational energy. As in the case of an alanyl α-carbon atom, the effect of the inclusion of the electrostatic energy is to shift the minimum. Small minima observed along the margins ($\phi \simeq 0°$ and $\psi \simeq 0°$) may again be attributed in part to the torsional potential.

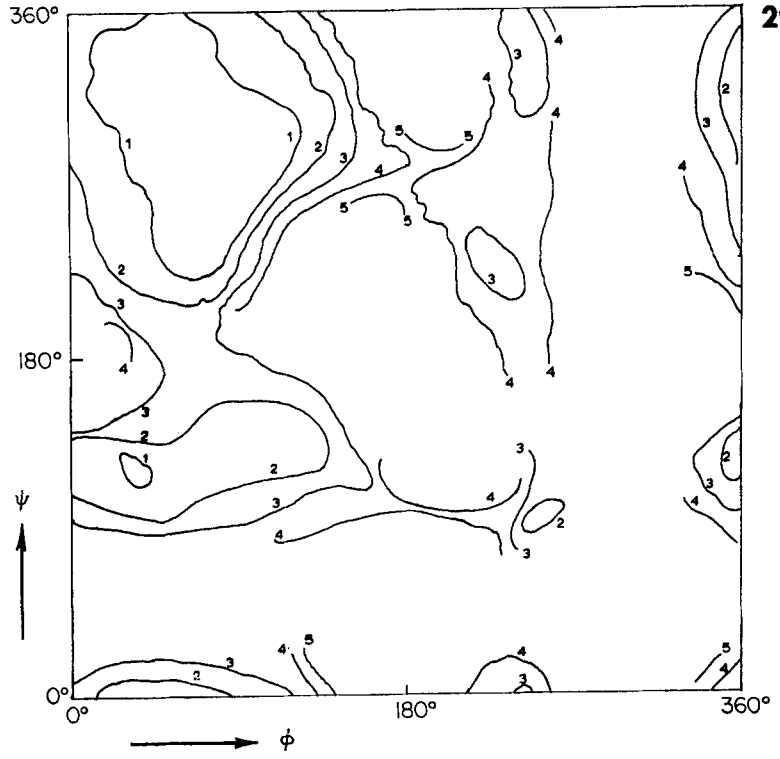
Similar calculations of conformational energy have been carried out by

---

FIG. 28. Same as Fig. 27, but for a glycyl α-carbon atom—(A) with only V_{nb} and torsional potential, and (B) including also V_{es}.



29A



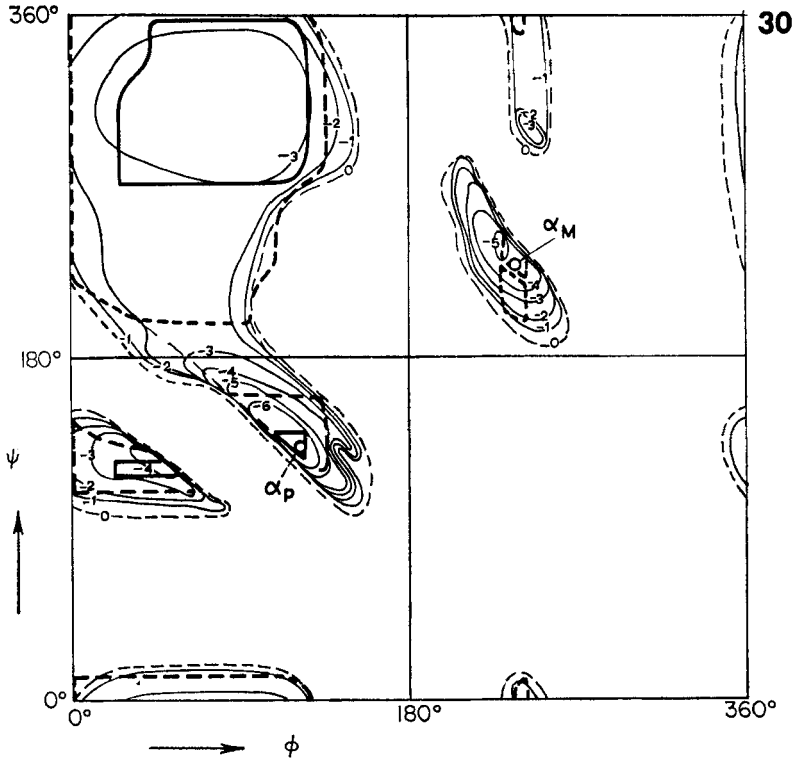
29B

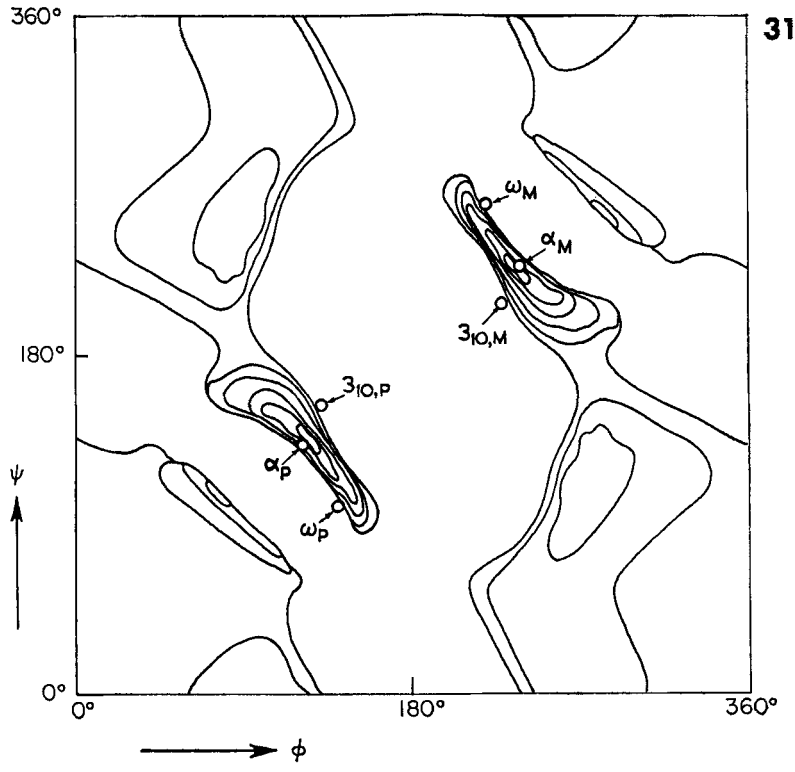
Scott and Scheraga (1966c) for a pair of linked peptides both when a C^{β} atom is present and when it is not present, corresponding to the above two cases of Brant *et al.* (1967). They have included in their calculations terms due to nonbonded, torsional, and electrostatic energies. The parameters of the terms have already been discussed in the last section. Their calculations agree well with those of Brant *et al.* (1967), except that they have obtained somewhat lower energies throughout the ϕ-ψ plane and that the two low-energy regions in the left-hand side are bridged across $\psi = 180°$, as shown in Fig. 29A. In this respect, the calculations of Scott and Scheraga (1966c) resemble more closely the calculations of Ramachandran *et al.* (1966b) (compare Fig. 25 and Fig. 29A). These minor differences between the contour maps of Brant *et al.* (1967), and Scott and Scheraga (1966c) are due to the fact that each group of workers has employed parameters for the nonbonded, torsional, and electrostatic energies slightly different from the other. Therefore, calculations of this nature, being empirical in basis and approximate in numerical data, cannot be accepted as being accurate in detail. Only those features which persist, when the parameters are varied and different methods are compared, should be regarded as being reliable.

b. Inclusion of Strain Energy of Bond Angle Variations. As mentioned above, the inclusion of torsional and electrostatic potentials or energies, in the calculation of nonbonded energies, does not appreciably alter the features of the conformational energy map of a pair of linked units both with alanyl and glycyl side chains. Gibson and Scheraga (1966) have varied the bond angles in a suitable manner and have found that the regions of low energy expand somewhat. This can be seen by comparing Fig. 29A for a rigid model, in which the bond angles were kept constant, with Fig. 29B for a flexible model, in which the bond angles were allowed to vary slightly. Increasing the flexibility of the peptide group, however, has not altered the positions of the energy minima on the map for the rigid peptide units, except in regions near $\phi \simeq 260°$ and $\psi \simeq 100°$ which become considerably deeper. The energies in the regions between the minima become very much lower and the contours spread out as to include much larger areas within the regions of inter-

---

FIG. 29. Potential energy contours at intervals of 1 kcal/mole, using nonbonded, torsional and electrostatic energies, obtained by Scheraga and co-workers, for an alanyl linking α-carbon atom—(A) for planar peptide units with Pauling-Corey parameters; (B) minimum values obtained by distortion of bond angles in the plane. (From Scheraga *et al.*, 1967.)


30
360°
180°
0°
0°
180°
360°
ψ
φ
α_M
α_P


31
360°
180°
0°
0°
180°
360°
ψ
φ
ω_M
α_M
$3_{10,M}$
$3_{10,P}$
α_P
ω_P

mediate energy. Nonplanar distortions and their effects are under study in the authors' laboratory.

B. Conformational Energy Map of a Helical Polypeptide Chain

Ramachandran *et al.* (1966b) and Scott and Scheraga (1966c) have extended their computation to the cases of polyglycine and poly-L-alanine in a helical form. In working out the conformations available for the helical polypeptide, which is characterised by the value of ϕ and ψ being the same in every residue, nonbonded interactions, torsional, electrostatic and hydrogen bond energies were all taken into effect by Scott and Scheraga, while Ramachandran *et al.* have computed only the nonbonded interactions making use of Flory-Scheraga potential functions (Table XI).

The energy contour map due to Ramachandran *et al.* (1966b) for helical conformations of poly-L-alanine over the complete range of ϕ and ψ for $\tau = 110°$, including only the nonbonded interactions, is shown in Fig. 30, in which is superimposed the contact map for a perfect helix of poly-L-alanine. Again the striking feature is the close resemblance between the contact map and the energy contours, emphasizing once more the dominant role of the repulsive forces of the nonbonded interactions in deciding the allowed conformations. The unique positions of the right and left-handed α-helices may be noted. Regions around these are domains of large negative values going down to —7.0 kcal/mole and —5.0 kcal/mole respectively. According to these calculations, the right-handed α_P-helix is stabler than the left-handed α_M-helix by about 2 kcal/mole. The nonbonded energies corresponding to values of (ϕ, ψ) for different types of helices obtained by Ramachandran *et al.* are listed in Table XVII. It will be noticed that the right-handed α-helix is the stablest even with regard to nonbonded energy.

The results of Scott and Scheraga (1966c) do not differ very much from those of Ramachandran *et al.* (1966b). Again, as in the case of two

---

FIG. 30. Nonbonded potential energy contour map for helical chains of poly-L-alanine. Note the deep minima near the conformations of α_P and α_M, the former being lower. The contact map is also shown superposed and there is good agreement with this. (Data from Ramachandran *et al.*, 1966b, and Venkatachalam and Ramachandran, 1967.)

FIG. 31. Potential energy contour map for polyglycine helical chains, including hydrogen bond energy. (From Scott and Scheraga, 1966c.) Note the deep minima at the α-helical regions, and slightly higher energies for the ω- and 3_{10}-helices. The position of the 3_{10}-helix shown is different from that in the original reference, but agrees with the (ϕ, ψ) values in Table V.

linked peptides, the inclusion of the various energy terms, such as torsional, electrostatic, and hydrogen-bonded energies, has only a marginal effect on the energy map of helical poly-L-alanine due to nonbonded interactions alone. However, the actual values of minima and their corresponding values of (ϕ, ψ) are not the same in the two cases. For example, Scott and Scheraga report that by carrying out energy minimization calculations, both in the right-handed and left-handed regions, the right-handed α-helix is stabler than the left-handed one by a few tenths of a kcal/mole per residue and not by as much as 2 kcal/mole per residue as reported by Ramachandran *et al.* (1966b).

Figure 31 shows the results of the calculations of Scott and Scheraga for regular helical polyglycine structures. There is complete symmetry in the figure in the sense that each kind of right-handed helix has a left-handed counterpart of the *same* energy. The regions of lowest energy again very closely correspond to the right- and left-handed α-helices.

TABLE XVII
Nonbonded Potential Energies of Various Standard Helical Polypeptide Chain Conformations

Type of hydrogen bond	Description of helix	ϕ (in degrees)	ψ (in degrees)	Energy (kcal/mole)
3—1	2.2_7-Helix ($n = 2.17$; $h = 2.75$ Å)			
	Right-handed	100	240	−2.0
	Left-handed	260	120	Large
4—1	3_{10}-Helix ($n = 3.00$; $h = 1.80$ Å)			
	Right-handed	122	158	−4.0
	Left-handed	238	202	−3.2
5—1	α-Helix: 3.6_{13} ($n = 3.60$; $h = 1.50$ Å)			
	Right-handed	122	134	−7.1
	Left-handed	238	226	−5.1

Also shown in the figure are the positions of the ω-helix and the 3_{10}-helix, which lie in relatively high energy regions. In the upper part of the figure, in the positions corresponding to both polyglycine I and polyglycine II, the energy value is between 3 and 4 kcal/mole above that of the α-helix. However, we know that these two structures are stabilized by interchain hydrogen bonds (which are not included in their calculations), which can more than make up for the difference in energy of 4

kcal/mole between it in a single chain state and the hydrogen-bonded α-helix. (See Section VII,F for a brief discussion of this.)

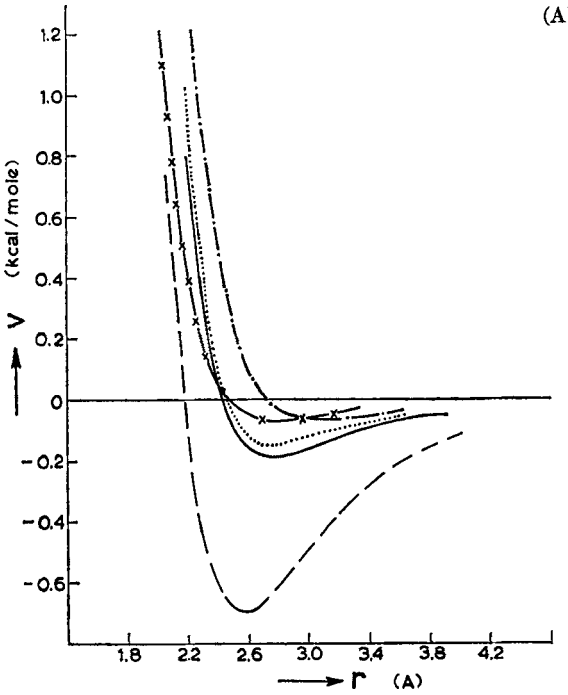
C. Comparison of the Different Nonbonded Potential Functions

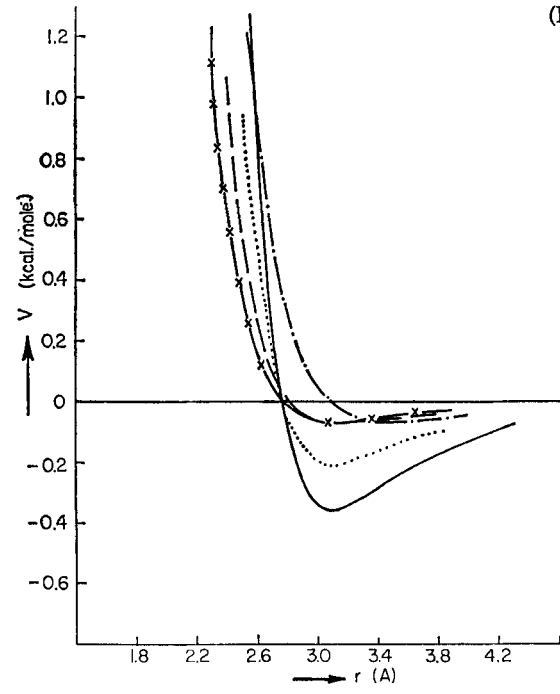
It would appear from the above discussion that the potential energy calculations would be very useful in the prediction of polypeptide conformations. However, all the above examples have only emphasized the important role of the repulsive term of the nonbonded interactions in determining the regions of low energy. Therefore, it would be interesting to compare the nonbonded energy terms proposed by different workers. This has actually been done by Venkatachalam and Ramachandran (1967), who have made a comparative study on the various nonbonded potential functions that have been proposed.

There are many different types of potential functions with different parameters proposed in the literature. If graphs of the variation of $V_{nb}(r)$ with interatomic distance r are plotted for these, then the different curves differ appreciably from one another. This is shown in Fig. 32 for two typical cases of N . . . N and N . . . H interactions. It will be noticed that the position and value of the minimum differ from one function to the other. A comparative study of the differences that these would make in predicting polypeptide conformations is therefore of great interest.

The functions used by Venkatachalam and Ramachandran were those proposed by DeSantis *et al.* (1965), Brant and Flory (1965b), Scott and Scheraga (1966c), and Kitaigorodsky (1961). These are denoted by the symbols L, F, S, and K_1, respectively. They have in addition used a variant of the Kitaigorodsky function, referred to as K_2, by requiring that the potential be a minimum at a distance of separation equal to the sum of the van der Waals radii. In the K_1-function, the minimum occurs at a distance always larger than the sum of the van der Waals radii by about 10%.

The energy maps for a pair of peptides linked at an alanyl α-carbon atom and also for poly-L-alanine have been reported by these authors in a numerical form, with the value of the potential function rounded off to the nearest integer. It was observed that there was general agreement between the maps using different functions and all agreed well with the contact map. For example, in the maps of two linked peptides, they all showed minimum energy values in three regions corresponding to the three allowed regions I, II, and III in the contact map. One occurred near the right-handed α-helical conformation, the second near the left-handed α-helical conformation and the third was a broad minimum


(A)
1.2
1.0
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
V (kcal/mole)
1.8
2.2
2.6
3.0
3.4
3.8
4.2
r (A)


(B)
1.2
1.0
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
V (kcal/mole)
1.8
2.2
2.6
3.0
3.4
3.8
4.2
r (A)

around the collagen conformation. However, the depth of the minimum within the three allowed regions was different for the different functions.[7]

Again, the energy maps of the different potential functions for poly-L-alanine resembled one another and also the contact map for helical poly-L-alanine. The two regions containing the right-handed and left-handed α-helical conformations respectively of the alanyl map were both split into regions omitting conformations corresponding to helices of small unit height, in all the functions. All the maps showed an optimal conformation near α_P and a less-preferred one near α_M. Numerical values of the energies have been given by Venkatachalam and Ramachandran (1967) using the five different potential functions for a few specific helical conformations. It was noticed by them that the energy difference shown by the various functions between the right- and left-handed forms of the α-helix were different. For the functions L and F, α_P is more stable than α_M by a difference of the order of 2 kcal/mole whereas it is only of the order of 0.2 to 0.4 kcal/mole for the functions S, K_1 and K_2. However, all the functions showed that the 3_{10}-helix is much less stable than the α-helix (both in the P- and the M-forms) by about 2 to 3 kcal/mole. The 2.2_7-helix has a distinctly higher nonbonded energy.[8]

The energy maps obtained using the L, F, and S functions have been

[7] There was a difference between the results obtained from the functions L and all the others put together. The map with the L-functions did not have a minimum near the collagen conformation, which is densely populated, for example in the lysozyme ϕ—ψ map, and the shape of its contours also did not follow the outline of the contact map. Apart from this, there are some errors of statement in the papers by Liquori and co-workers on the subject (e.g., DeSantis *et al.*, 1965; Liquori, 1966). For instance, in Table IV of Liquori (1966), a conformation (45°, 60°), which corresponds very nearly to $n = +3$, is said to be the same as that of "the right-handed three-fold helix of polyglycine II," which is incorrect, as the latter is really near (260°, 35°) (see Table V), as first worked out by Crick and Rich (1955).

[8] C. M. Venkatachalam (private communication) reports that there are small numerical errors in the data reported in the paper by Venkatachalam and Ramachandran (1967), owing to misbehaviour of the computer. However, the main conclusions are unaffected.

FIG. 32 (opposite). Variation of nonbonded potential energy with distance for the interaction (A) N . . . H and (B) N . . . N for the various functions. ——— (F, Brant and Flory, 1965a, as adapted by Ramachandran *et al.*, 1966b), - - - - (L, DeSantis *et al.*, 1965), ---- (K_1, Kitaigorodsky, 1961, 1965), -x-x- (K_2, modification of the Kitaigorodsky function by Venkatachalam and Ramachandran, 1967), ... (S, Scott and Scheraga 1966c). Note the variations between the different curves. In particular, the L-function for the N . . . H interaction is four times deeper at the minimum than all the others.

given in the preceding sections. That obtained with the K_2 functions, as well as a map published by Dunnill (1965), using functions whose parameters have not been tabulated, are shown in Figs. 38B and C in Section VIII,C,2. These functions show a surprisingly good agreement with the data on lysozyme, particularly with a few (ϕ, ψ) values which occur well outside the allowed limits of the contact map.

The important conclusion arrived at by these studies is that the potential maps, whatever form of function is used, are in conformity with the contact map, as far as regions of low energy are concerned, and these are decided mostly by the repulsive interactions V_r. There are significant differences, however, in the prediction of relative stabilities of different conformations between the various proposed functions, and this points out the need for obtaining better data from different sources, which could be used to obtain a definitive set of potential functions.

D. Energy Minimization Methods

1. Introduction

For a pair of linked trans peptide units (at a glycyl or alanyl α-carbon atom) the problem of deciding the conformation of lowest energy is rather simple. This is because we require only two parameters ϕ and ψ to characterize the conformation and the total energy $V(\phi, \psi)$ is a function of these two variables only. Thus, the problem of locating the stable conformation is merely one of searching a two-dimensional map for the minimum. As described in the last section, one needs only to calculate the energy at every point in the plane at suitable intervals and select the minimum by inspection. The exact location of the minimum can be made by interpolation between the intervals of calculation. A similar procedure can be used for helical structures, as described above in Section VII,B. The same procedure is more cumbersome when, in addition to the pair of parameters ϕ and ψ, a third rotation χ is also included. The problem is then a three-dimensional one, as in the case of a valine side chain attached at the linking atom of a pair of peptide units. For this case of valine, Liquori (1966) has adopted the simple procedure of considering this as a three-dimensional map with ϕ, ψ, χ as axes and the (ϕ, ψ) contours at suitable sections of χ and stacking the contours drawn on transparent sheets. The minimum values of energy in each section are then located and the profile of minima drawn. The deepest minimum is directly obtained and the exact location can then be calculated by employing interpolation methods.

Recently, in the authors' laboratory, the energy minima of a few cyclic peptides, such as cyclotetraalanine, have been determined by this method

of evaluating the energy over a range of values of the parameters involved and locating the minimum (Ramakrishnan and Sarathy, 1968). This is briefly discussed in Section VII,G.

When more parameters are involved, several simplifying assumptions could be incorporated in the search for the stablest conformation. For instance, Ooi *et al.* (1967) have assumed that the possible minima for the χ-rotations occur near the theoretically expected positions characterized by $\chi = 60°$, 180°, and 300°. Such an assumption enormously reduces the computational steps, though it might not lead to the real minimum. There is every chance that the minimum lies at a point not very near the sample points.[9] However, recently, Ooi *et al.* (1967) have reported a reasonable amount of success of this method in working out the conformation of gramicidin-S, a cyclic decapeptide (see Section VII,G).

An entirely different approach for finding out a configuration of minimum energy by model-building through computers has been developed by Levinthal (1966). In this method, a particular conformation characterised by a set of parameters is displayed by a computer on a screen. For this configuration, the total energy of interaction between various atoms and the direction of decreasing energy are calculated. Using these data, fresh instructions are issued to the computer and an altered conformation is studied. The search of the energy minimum can then proceed in this fashion.

From the above discussion it will be seen that for a pair of linked peptides with a fairly big side chain involving a number of χ-parameters, the evaluation of the minimum is a complicated procedure, although still amenable to the method of calculating the energy for varying values of the parameters involved and finding the minimum. The problem is of the same magnitude for a homopolypeptide helix for which the ϕ, ψ, and χ values are all the same along the length of the helix. However, for a segment of a nonhelical chain, the number of parameters that affect the total energy is so large that the direct method of evaluating the energy for all possible sets of values of the parameters and finding the minimum is almost impossible. In view of this, some mathematical procedure of energy minimization is necessary to get at the optimal conformation. This is described in the next subsection.

2. The Energy Surface

The principle behind the energy minimization procedure is the following: The n parameters characterising the conformation may be denoted

[9] However, in the case of poly-L-tyrosine, Ooi *et al.* (1967) actually verified that the minimum occurred close to the assumed values of χ.

by $a_1, a_2, \ldots, a_n$. These may be, for example, ϕ, ψ, χ^1, χ^2, etc. in the polypeptide case. These constitute the axes of an n-dimensional space. The energy V is a function of $a_1 \ldots, a_n$. A point in the n-dimensional space represents a conformation and has associated with it a value of V. All such values of V make up a multidimensional V-surface. This may be called the energy surface. The basic approach in the energy minimization procedure is essentially the construction of a course of steepest energy gradients from a randomly selected point, and iterating the procedure towards the real local minimum in its vicinity.

The application of minimization procedures to obtain the coordinates (in n dimensions) of the minimum has been used for the determination of crystal structures of simple molecules (Kitaigorodsky, 1965). The validity of the approach has been demonstrated by constructing the V-surface in the vicinity of the point representing the actually observed structure of naphthalene, for which the surface is seven-dimensional, and by evaluating the elastic coefficients from the shape (or differential coefficients) of the V-surface near the minimum and showing them to be in agreement with observation.

3. Gradient Methods

The methods usually employed to attain the minima on the V-surface, starting from selected points, are known as gradient methods. These employ the relation that, for a function to have a minimum, its derivatives should all be zero:

$$\partial V/\partial a_i = 0 \qquad \text{for } i = 1, 2, \ldots, n \tag{39}$$

Thus, a method of proceeding towards the minimum would be to calculate the rate of variation of V with respect to the parameters involved at the sample point and to move the conformation to another point along the direction of the gradient in which the energy decreases, i.e., along the vector—grad V and then repeat the procedure. This method is called the method of steepest descent.

A difficulty in the method is in the choice of the magnitude of the step (say l) to be taken along the direction of the negative gradient. In case the expected value of the minimum energy (or the quantity that is minimised) is known, the value of l may be estimated from the deviation of the local value from the minimum. Otherwise, l has to be chosen arbitrarily and adjusted to give the quickest convergence to the minimum—if it is taken too large, there would be fluctuations, while if it is taken too small, the convergence would be very slow.

A second difficulty is that the convergence would take place only towards the minimum point of the local valley, and this may not be the

absolute minimum, or the lowest point of the V-surface. Several methods of getting over this difficulty have been suggested, and we shall mention only two here. The method of Gel'fand, which is first discussed, proceeds along what may be described as the course that a river would take in going down from the top of a hill along valleys and gorges and is superior to the method of steepest descent. But even here, it may wind up in a lake and not in the ocean, particularly if the lake is in a large plateau. So, it will be a good idea to start from two or three entirely different points and move along the path of decreasing energy using one of the methods described below to verify that one has reached the absolute rock bottom minimum. This is the real problem in the energy minimization procedures.

A method of searching for the location of the minimum of a multiparameter function has been suggested by Gel'fand *et al.* (1963). The method is described as a nonlocal search in which local structures of a function at various points are compared and "regularly organized." Exploration of these regions of low values in a prescribed fashion is expected to lead to the overall minimum. The search is begun at a random point X_0 (Fig. 33) and a local minimization by the gradient method is applied to reach A_0. This is repeated from a nearby point X_1 to get to A_1. The third search point X_2 is taken along the line A_0A_1 at a step l from A_1. The gradient search starts from X_2 to lead to A_2 and X_3 is taken along A_1A_2 at a distance l from A_2 and so on. Once the choice of the starting point and l are made, the convergence is claimed to be quick. This procedure was applied to the solution of the crystal structure of L-proline (Kayushina and Vainshtein, 1966). The positions of the atoms in the structure were obtained by minimizing the disagreement factor R for the intensities, which is a function of the coordinates of the atoms and is a minimum for the correct structure. It should be mentioned that, in this problem, the expected value of the minimum was known and the function was not expected to have too many false minima, or large maxima in between minima in the path.

Though the above procedure sets out to find the absolute minimum of the function, it is very difficult to apply such a method for the potential function V, since it has abrupt changes due to the energy barriers present. Hence, local minimization procedures are more useful along with a set of predetermined starting points of expected minima. Such a procedure is described by Scott and Scheraga (1966b). The method is briefly described as follows: A starting conformation P_1 is chosen (Fig. 34) where the minimum is expected to occur (such as the staggered positions for χ^1 etc.). From the gradient of V at P_1, the direction in which V decreases is obtained. The values of V for various points along

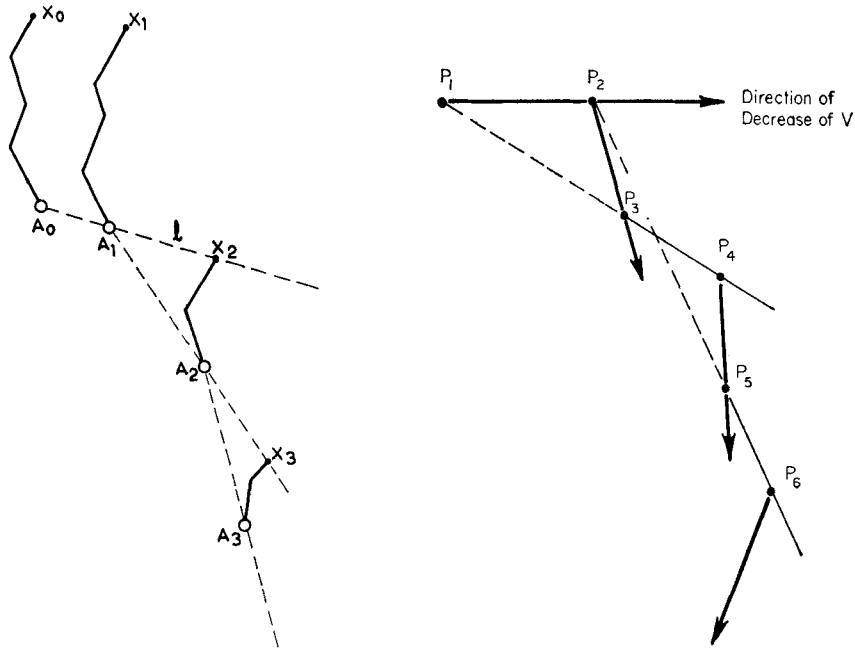


FIG. 33. (left). Method of minimization of a function as suggested by Gel'fand.
FIG. 34. (right). Energy minimization procedure adopted by Scott and Scheraga (1966b).

this line are calculated and the point P_2 of minimum energy is found out. From P_2 again, the direction of decrease of V is computed and the minimum along this line is located at P_3. The next search point P_4 is chosen as the minimum point on the extension of the line P_1P_3. At P_4 again, the direction of decrease of V is calculated and the minimum is located at P_5. The minimum is then computed on the line P_2P_5 and so on alternately, until the lowest minimum is reached. Such a procedure was first applied by these workers for finding out the stable conformations of *n*-alkanes.

Scheraga *et al.* (1967), have reported energy minimization calculations on poly-β-methyl-L-aspartate and poly-γ-methyl-L-glutamate. In the former case, four parameters were varied and in the latter five parameters were changed. They have also indicated that several energy minimization techniques are being tried on gramicidin-S and homopolymer helices. The success of this procedure, or the nonlocal search of Gel'fand *et al.* (1963), in polypeptide conformations can be judged only after several applications have been made. However, in the next subsection, we shall discuss the results of the application of the energy

minimization method to homopolymer helices as discussed by Scheraga *et al.* (1967) and Ooi *et al.* (1967).

E. Conformational Energies of Homopolymer Polypeptide Helices

Using the energy minimization procedure discussed above, the minimum conformational energies of several homopolar helices such as poly-L-valine, poly-β-methyl-L-aspartate, poly-γ-methyl-L-glutamate and poly-L-tyrosine have been determined by Scheraga *et al.* (1967). In these structures, the side-chain interactions were also taken into account. So also, contributions due to nonbonded, torsional, electrostatic and hydrogen bond energies have been included. The following are some of the interesting results reported by them. A word of caution must be offered here. As discussed in Section VII,B and C, these predictions should be accepted with reservation, because the potential functions that have been used are not definitive. Because of this, if two positions of minima are found, which differ say by less than 0.5 kcal/mole, one set of potential functions may lead to the lowest energy near one, while the other may yield the absolute minimum at the other or a nearby conformation. Thus, the attempt should be to find the regions of low energies, rather than the exact point of absolute minimum.

Poly-L-valine was shown by Scheraga and co-workers (1967) and Ooi *et al.* (1967), to have a favourable conformation as a right-handed α-helix, when the energy was minimized with respect to the dihedral angles of the backbone and the side chains. This is in contradiction with the predictions of Liquori (1966). Incidentally, this agrees with the prediction of the contact criterion (see end of Section V,D), which shows that positions II and III of the γ-carbon atom are allowed for the helices α_P and α_M, and the valine γ-carbon atoms could take up these positions, but α_P is distinctly superior. Optical rotatory dispersion (ORD) measurements on a block copolymer of the type (D,L-Lys)$_x$—(L-Val)$_x$—(D,L-Lys)$_x$ with $x \approx 40$, have indicated that the L-valine portion exists in the form of an α_P-helix in 90% aqueous methanol at room temperature (Scheraga *et al.*, 1967).

In the case of poly-β-methyl-L-aspartate and poly-γ-methyl-L-glutamate, conformational energy calculations have been carried out by Ooi *et al.* (1967), near the right- and left-handed α-helical regions. Energy minimization techniques led to the prediction of α_M for the aspartate polymer and α_P for the glutamate polymer. This difference in the screw sense has been attributed by these authors to the interaction of the dipole of the ester group with that of the backbone. This interaction accordingly stabilizes the right-handed form in the glutamate polymer, but

destabilizes the right-handed form in the aspartate polymer. In order to understand the role of the dipole-dipole interactions in the free energy calculations, Ooi *et al.* (1967) have repeated the energy calculations for different nonbonded potentials and for several values of the dielectric constant. It was always found that the aspartate polymer had a lower energy in the left-handed α-helical form.

In the case of poly-L-tyrosine, it was found that α_P was more stable than α_M by about 1.8 kcal/mole, the main stabilizing contribution being the nonbonded energy. However, the dipole-dipole interaction between the side chain and the backbone is not sufficient to reverse the screw sense, unlike the situation found in the aspartate polymer, where the reversal takes place.

F. Conformational Energies of Polypeptide Structures

In Section VII,B, we had considered the energies of single helical chains of polypeptides. The energy of actual crystal structures in which these chains are put together has been calculated by Venkatachalam (1968c). He considered the cases of the parallel and antiparallel pleated sheets and worked out the nonbonded energy per unit for different distances of separation of the chains in the sheet, using the functions F mentioned in Section VII,C. The results are shown in Table XVIIIA, and it will be seen that, in both cases, the energy is a minimum at about 4.7 Å, which is the observed separation.

In addition, Venkatachalam has calculated the nonbonded energies of a three-dimensional structure containing the pleated sheet β-structure (polyglycine I) and three-fold helices (polyglycine II) and these are given in Table XVIIIB. The energy in both cases was in the region of -13 to -15 kcal/mole per residue and was nearly equal for both, showing the possibility of easy conversion of one into the other. When α-helices of polyglycine were packed together, the total energy of the structure was distinctly higher (per residue) than the above value, showing that this structure was less stable than the β- and the triple-helical structures. However, this particular aspect requires more detailed study.

G. Energy Calculations for Cyclic Peptides

Studies on cyclic tetra- and pentapeptides based on contact criteria were mentioned in Section V,F,3. Calculation of energy minima of their conformation have been made by Ramakrishnan and Sarathy (1968). A value of $K_\tau = 80$ kcal/mole was used for the distortion of the tetrahedral angle $NC^\alpha C'$ and a value of $K_\omega = 30$ kcal/mole was used for the ω-distortion (based on the value given by Donohue, 1953). Otherwise,

TABLE XVIIIA

Variation of Nonbonded Energy with Distance of Separation Between a Pair of Polypeptide Chains in the Parallel and Antiparallel Pleated Sheet Arrangements

Parallel pleated sheet (h = 6.50 Å)			Antiparallel pleated sheet (h = 7.00 Å)		
Separation (Å)	Energy (kcal/mole/residue)		Separation (Å)	Energy (kcal/mole/residue)	
	Polyglycine	Poly-L-alanine		Polyglycine	Poly-L-alanine
4.35	−3.77	−5.12	4.25	−0.56	−2.47
4.45	−5.26	−6.52	4.35	−3.37	−5.21
4.55	−6.06	−7.23	4.45	−4.91	−6.66
4.65	−6.41	−7.50	4.55	−5.67	−7.34
4.75	−6.49	−7.50	4.65	−5.97	−7.56
4.85	−6.40	−7.34	4.75	−6.00	−7.52
4.95	−6.23	−7.11	4.85	−5.89	−7.33
5.05	−6.01	−6.83	4.95	−5.70	−7.08
5.15	−5.78	−6.54	5.05	−5.48	−6.79
5.25	−5.54	−6.26	5.15	−5.24	−6.50
5.35	−5.32	−5.99	5.25	−5.01	−6.23

TABLE XVIIIB
Nonbonded Energy of Polyglycine Structures

Structure	Reference[a]	Energy (kcal/mole/residue)
Polyglycine I Monoclinic[b] $a = 4.75$ Å, $b = 7.00$ Å, $d(100) = 3.45$ A, $\gamma = 67°$	(1)	−15.5
Polyglycine II Hexagonal	(2)	−13.2
$a = 4.8$ Å, $c = 9.3$ Å	(3)	−12.7

[a] The reference is to the article where the structure is described. (1) Bamford *et al.* (1956); (2) Crick and Rich (1955); (3) Ramachandran *et al.* (1966a).
[b] The diagram of this structure is given in Ref. (1), p. 285.

the Pauling-Corey dimensions were used for the bond lengths and bond angles and the potentials F (Section VII,C) were used.

With these parameters, if the cyclic tetrapeptide was assumed to have a four-fold axis of symmetry and all the peptide units were taken to be equivalent, cyclotetra-L-alanine had a deep minimum energy of −6.7 kcal/mole per residue for $\omega \simeq 10°$ and $\tau = 105°$, $\phi = 83°$, $\psi = 129°$. If all the residues are D, the inverse conformation $\phi = 277°$, $\psi = 231°$ was stable. Thus, cyclotetra-L-alanine or D-alanine is a stable structure, and as already mentioned, the former has been found to occur. Its (ϕ, ψ) value is near that of the right-handed α-helix.

It was found that if even one of the four L-residues in cyclotetra-L-alanine was replaced by a D-residue, the minimum energy shot up by about 7 kcal/mole per residue and such a mixed peptide is therefore unlikely to have a cyclic structure. Cyclotetraglycine has also a low energy of less than −7 kcal/mole per residue with $\omega \simeq 10$–15°.

The study also showed that cyclopenta-L-alanine can have a very stable structure with an energy of less than −7 kcal/mole per residue for $\omega \sim 5°$, $\tau \sim 110°$ and $\phi \sim 65°$, $\psi \sim 135°$. The exact minimum depends on the parameters chosen for the potential function. Here also, if there is even one D-residue in the midst of all other L-residues, the energy rises by about 6 kcal/mole per residue. However, mixed pentapeptides of the type

└Gly—D-Leu—Gly—L-Leu—Gly┘ and └Gly—D-Leu—L-Leu—Gly-Gly┘

have been synthesised (Kenner *et al.*, 1958; Hardy *et al.*, 1963) and their conformations are under study.

Energy minimization methods have been used by Vanderkooi *et al.* (1966), and Scheraga and co-workers (1967), for the determination of the conformation of the cyclic decapeptide gramicidin-S. A possible low-energy structure has been obtained. Liquori *et al.* (1966) have also obtained a minimum energy structure using their L-type potential functions. The two structures are very different. The structure of this decapeptide has not yet been worked out by X-ray methods and it would be of interest to see how well it agrees with the theoretical predictions, when the X-ray determination is made.

H. Stable Conformation of the cis Peptide Unit

The suggested standard dimensions of the cis peptide unit were mentioned in Section II,A,1, Fig. 4B and Table II(c). The expected conformation of minimum energy for the cis peptide unit could be obtained by taking as the starting point the one obtained from the Pauling-Corey trans peptide by giving an ω-rotation of 180° (Ramachandran and Venkatachalam, 1968). This starting unit has $\tau(C^{\alpha}C'N) = \tau_1 = 114°$ and $\tau(C'NC^{\alpha}) = \tau_2 = 123°$. However, the C_1^{α} . . . C_2^{α} distance is short in this and this short contact could be relieved by an increase of τ_1 and τ_2 by δ_1 and δ_2 say. Accompanying the change $+\delta_1$ of τ_1, the other two angles at C′ were each changed by $-\delta_1/2$. The resultant change in energy ΔV due to the components V_{nb}, V_{τ} (for all three angles at C′ and N) and V_{es} ($\epsilon = 1$) were computed and the sum of all the effects is shown in Table XIXA. It will be seen that there is a broad minimum

TABLE XIXA

Variation of Potential Energy (in kcal/mole) with Distortion of the cis Peptide Unit[a]

$\delta_1 \backslash \delta_2$	1°	2°	3°	4°	5°
1°	−0.382	−0.499	−0.558	−0.562	—
2°	−0.499	−0.495	−0.634	−0.622	−0.558
3°	−0.558	−0.634	−0.657	−0.628	−0.552
4°	−0.559	−0.617	−0.626	−0.584	−0.493
5°	—	−0.549	−0.543	−0.489	−0.369

[a] The starting point is at $C^{\alpha}C'N = 114°$ and $C'NC^{\alpha} = 123°$. The distortion of these two angles are respectively δ_1 and δ_2. (See text for other details.)

at about $\delta_1 = +3°$, $\delta_2 = +3°$. (The contribution V_{es} varied negligibly for a change of $\pm 2°$ about the minimum even for $\epsilon = 1$ and so the value of ϵ used is immaterial.)

Table XIXB gives the values of the angles occurring in the cis peptide unit, corresponding to this minimum, in diketopiperazine (Degeilh and

TABLE XIXB
Values (in Degrees) of the Angles Involved in the Planar cis Peptide Unit

Structure\Angle	$C^{\alpha}C'N$	$C^{\alpha}C'O$	$OC'N$	$C'NC^{\alpha}$	$C'NH$	HNC^{α}
Diketopiperazine	119	118.5	122.5	126	123	111
Leu-Pro-Gly	119	119	122	126	121	113
Minimum energy (calculated)	117	119.5	123.5	126	121.5	112.5
Recommended standard	**118**	**119**	**123**	**126**	**121**	**113**

Marsh, 1959), the relevant unit picked out of the structure of Leu-Pro-Gly (Leung and Marsh, 1958) and the recommended standard values.

On calculating the variation of energy with ω, it was found that it was a minimum for $\omega = 180°$, showing that nonbonded interactions are not expected to lead to nonplanarity of the cis peptide unit. This agrees with the observations in diketopiperazine and Leu-Pro-Gly. In the latter, all the torsional angles are within 5° of that for a planar conformation, and the former is highly planar.

VIII. CONFORMATIONS OBSERVED FOR AMINO ACIDS, PEPTIDES, POLYPEPTIDES, AND PROTEINS

A. Introduction

In the previous sections we have described the various methods of approach available for the theoretical prediction of the possible conformations of polypeptide chains. We have also seen that, in general, the predictions, both of the contact criteria as well as of potential energy calculations, agree reasonably well with observation in a good number of examples. In this section, we shall discuss the observed conformations in proteins and polypeptides and in simpler compounds like amino acids and small peptides, which have a bearing on polypeptide conformation. We shall only consider those aspects of conformation which have a relevance at the molecular level, i.e., we shall only consider the primary and secondary structure of these materials, and leave out the tertiary and higher order structures which occur in biological systems. Thus, in the case of globular proteins, we will discuss only the distribution of the various dihedral angles ϕ, ψ, χ and the types of structures like α-helix, extended chain, and so on which occur in these structures. We shall also give a brief discussion of the effect of different residues on helix formation.

In the case of fibrous proteins and polypeptides, the discussion will be restricted to some of the recent studies, as earlier work has been adequately reviewed by various workers. We may refer in particular

to a recent review by Davies (1965). However, some recent studies on polypeptides which form model compounds of typical fibrous proteins will be discussed in some detail.

B. Observations on Amino Acids and Peptides

Since amino acids are the building blocks of polypeptide chains, a study of the conformations in which they occur, particularly with reference to the conformation of the side group, is very interesting. In fact, in the case of simple amino acids or their derivatives like hydrohalides and of small peptides containing up to six residues, X-ray crystallographic methods provide very accurate data concerning their conformation. A preliminary analysis of such data was published by Sasisekharan (1962) and a more detailed study was made by Ramakrishnan and Ramachandran (1965). A good review of the available data on bond lengths and bond angles in amino acids has been published in Volume 22 of this series by Marsh and Donohue (1967). An extensive study of the conformation of the side groups as they exist in amino acids and peptides has been reported by Lakshminarayanan *et al.* (1967). A still more detailed monograph on amino acids is under preparation by Lakshminarayanan (1968).

The conformational parameters characterising the amino acid, peptide, C- and N-terminal residues, and the side groups have been discussed in Section II. In particular, in an amino acid, the two oxygen atoms O′ and O″ of the carboxyl group would require two parameters ψ' and ψ'' for their specification, while the angle ϕ which refers to the hydrogens at N is not relevant, since the hydrogens have not been located in most structures. The angle ϕ is again not necessary for the N-terminal end of a peptide, which, however, has one ψ-value (Fig. 8). For a C-terminal peptide, as can be seen from Fig. 8, the relevant parameters for the backbone are one value of ϕ and two values of ψ. For the description of a peptide unit in the middle of a chain, the two values that are necessary are ϕ and ψ. It will also be recalled (see Section II) that the side group conformation is described by a set of χ-values. In the following subsections, we shall describe briefly the general features regarding these dihedral angles observed in the backbone and side group conformations of amino acids and peptides, following Lakshminarayanan *et al.* (1967). A more extensive review, including more recent data, is under preparation by Lakshminarayanan (1968).

1. Backbone Conformation

Table XX lists the values of the relevant dihedral angles ϕ and ψ observed at glycyl α-carbon atoms in various structures. While the values of ϕ and $\phi'(=\psi-180°)$ reported by Ramakrishnan and Rama-

TABLE XX

The Values of ϕ and ψ (in Degrees) for Glycine and Glycyl Residues[a]

Structure		ϕ	$\psi 1$	$\psi 2$
		Glycine		
α-Glycine		—	198.3	19.2
β-Glycine		—	203.7	27.3
γ-Glycine		—	191.7	15.0
Diglycine HBr-1		—	200.8	22.0
Diglycine HBr-2		—	182.4	355.8
Diglycine HCl-1		—	192.1	16.4
Diglycine HCl-2		—	182.5	359.4
Bisglycino-Cu-1		—	184.2	4.5
Bisglycino-Cu-2		—	173.9	351.7
N-Acetylglycine		—	183.0	3.2
		C-Terminal residues		
β-Gly-Gly		357.7	184.7	0.2
Leu-Gly HBr		95.3	176.1	354.6
Gly (in glutathione)		95.6	190.0	10.9
Gly-Phe-Gly		264.3	179.3	355.4
Leu-Pro-Gly		3.2	180.5	358.1
Cys-Gly NaI		207.6	213.1	32.4
Gly-Gly-Gly Cu Cl $1\frac{1}{2}H_2O$		264.6	175.3	353.2
Na-Gly-Gly-Gly cuprate		311.2	160.9	342.8
2Na Gly-Gly-Gly-Gly cuprate		98.6	179.7	0.4
		N-Terminal and middle residues		
β-Gly-Gly		—	—	330.8
Gly-Asp		—	—	8.0
Gly-Trp $2H_2O$		—	—	344.1
Gly-Tyr HCl		—	—	352.3
NN′ diglycyl cystine		—	—	24.7
Gly-Phe-Gly		—	—	312.6
Gly-Gly-Gly Cu Cl $1\frac{1}{2}H_2O$		—	—	342.8
Gly-*Gly*-Gly Cu Cl $1\frac{1}{2}H_2O$		294.2	—	311.0
2Na *Gly*-Gly-Gly-Gly cuprate		—	188.6	—
2Na Gly-*Gly*-Gly-Gly cuprate		359.6	184.3	—
2Na Gly-Gly-*Gly*-Gly cuprate		5.5	179.6	—
		Cyclic peptides		
Gly (ferrichrome A)		262.2	178.2	—
Cyclohexaglycyl hemihydrate	1	86.0	188.2	—
	2	110.7	151.3	—
	3	85.3	186.7	—
	4	111.4	150.1	—
	5	87.1	188.3	—
	6	112.2	149.0	—
	7	87.9	184.0	—
	8	111.2	150.4	—

TABLE XX (Continued)

Structure		ϕ	$\psi 1$	$\psi 2$
Cyclohexaglycyl hemihydrate	9	50.1	209.4	—
	10	84.6	183.7	—
	11	48.3	209.5	—
	12	82.5	182.6	—
	13	261.6	186.4	—
	14	90.4	—	336.2
	15	261.8	188.6	—
	16	94.7	—	331.3
	17	289.3	—	312.4
	18	283.8	—	317.7
	19	285.0	—	317.6
	20	290.0	—	312.7
	21	59.0	—	10.0
	22	59.4	—	10.0
	23	65.3	—	2.1
	24	65.7	—	5.4

[a] Modified from Lakshminarayanan *et al.* (1967). References to the original literature may be obtained from this paper.

chandran (1965) were calculated as angles between suitably defined least squares planes, the values given here have been obtained by Lakshminarayanan *et al.* (1967), as simple torsional angles as defined in Section II.

It is interesting to note that the absence of L- and D-isomers for glycine results, in general, in the duplication of every (ϕ, ψ), leading to another one defined by ($-\phi$,$-\psi$). For instance, the value listed in the table for the centrosymmetric structure of cyclohexaglycyl hemihydrate would correspond only to one of the pair (ϕ, ψ) and ($-\phi$,$-\psi$) related by an inversion symmetry. The values in Table XX are thus only representative and each (ϕ, ψ) would also have another ($-\phi$,$-\psi$) associated with it.

It will be observed that the values of ψ are usually around 0° and 180° (with a deviation of less than 25°). Occasionally, deviations from these up to 40° are observed, as in the case of cyclohexaglycyl hemihydrate, where it is necessitated by closure of the ring.

In Table XXI are reproduced the values of ϕ and ψ observed in non-glycyl residues. It has been found that, unlike glycine, the tilt of the carboxyl group characterised by ψ is asymmetric in this case, and it is mostly negative for L-amino acids (although in general small), with only rare exceptions. In a few cases, the tilt is opposite with ψ being positive, but small. The positive tilt is large ($>15°$) only very rarely, e.g., Asp in aspartic acid HCl and Gly-Asp and Orn 2 in the cyclic peptide ferrichrome A.

TABLE XXI

Values of ϕ and ψ (in Degrees) for Amino Acids and Peptides with Side Chains[a]

Structure	ϕ	$\psi 1$	$\psi 2$
Alanine	—	164.7	342.9
Alanine	—	161.5	340.7
Arginine $2H_2O$	—	167.9	349.3
Arginine 2HI	—	165.5	336.5
Arginine HBr H_2O (molecule 1)	—	177.2	353.6
Arginine HBr H_2O (molecule 2)	—	156.1	331.2
Arginine HCl H_2O (molecule 1)	—	175.0	352.9
Arginine HCl H_2O (molecule 2)	—	155.3	332.9
Arginine HCl (molecule 1)	—	134.3	309.1
Arginine HCl (molecule 2)	—	138.3	319.2
Asparagine H_2O	—	186.2	11.4
Aspartic acid HCl	—	222.4	42.3
Cysteine HCl	—	174.9	349.1
S-Methylcysteine sulphoxide	—	170.2	356.6
Cysteine ethyl ester urea	—	179.8	359.0
Cystine	—	167.7	345.8
Cystine 2HBr	—	180.8	0.7
Cystine 2HCl	—	192.6	9.7
Glutamic acid	—	147.5	314.2
Glutamic acid HCl	—	162.0	342.4
Glutamine	—	168.7	340.5
Histidine HCl H_2O	—	179.5	0.4
Di-(histidino)Zn $2H_2O$	—	170.9	348.2
Di-(histidino)Zn $2H_2O$	—	176.1	356.4
Hydroxyproline	—	178.0	356.9
Isoleucine HBr	—	159.4	345.7
Isoleucine HCl	—	165.7	6.4
Leucine HBr	—	170.2	342.5
Lysine HCl H_2O	—	162.1	340.1
α-Methionine	—	149.6	326.3
β-Methionine	—	148.4	330.6
Norleucine	—	155.6	324.7
Phenylalanine HCl	—	177.5	358.1
Cu proline $2H_2O$	—	191.0	11.0
Serine	—	181.3	3.8
Threonine	—	156.1	333.9
Tryptophan HBr	—	201.4	355.8
Tyrosine HBr	—	154.7	322.2
Tyrosine HCl	—	149.6	323.9
Valine HBr	—	165.3	348.4
Valine HCl	—	171.2	352.2
Valine HCl·H_2O	—	174.2	356.3
	C-Terminal residues		
Gly-Asp	69.2	249.2	63.3
Cys (in glutathione)	89.2	174.9	358.7

TABLE XXI (Continued)

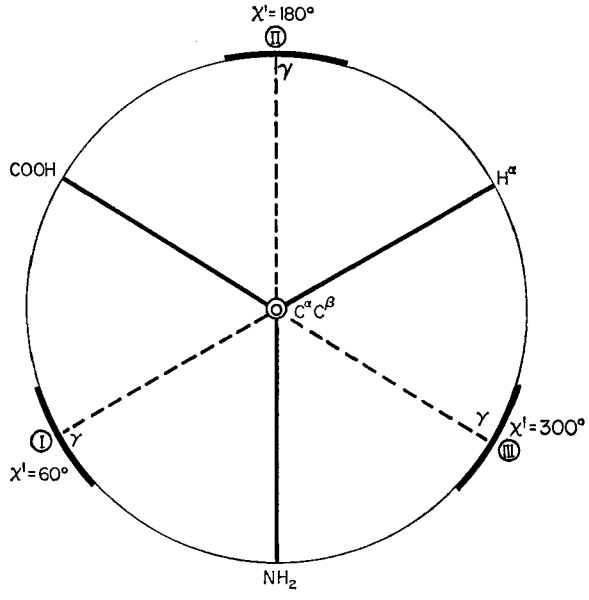
Structure	ϕ	$\psi 1$	$\psi 2$
	C-Terminal residues		
N,N'-Diglycylcystine $2H_2O$	20.4	174.9	356.4
β-Ala-His Cu $2H_2O$	27.4	174.3	353.6
p-Tosyl-Pro-Hypro	137.8	136.4	316.5
Thr-Phe-nitrobenzyl ester HBr	66.9	117.9	304.5
Gly-Trp $2H_2O$	107.1	156.4	326.7
Gly-Tyr HCl	93.2	149.1	326.6
	N-Terminal and middle residues		
Cys-Gly NaI	—	—	349.0
γ-Glutamyl (in glutathione)	—	167.5	350.4
Leu-Gly HBr	—	—	319.7
Leu-Pro-Gly	—	—	329.3
p-Tosyl-Pro-Hypro	—	—	353.2
Leu-Pro-Gly	111.8	161.9	—
Orn2 (ferrichrome A)	35.6	—	27.2
Orn3 (ferrichrome A)	103.5	131.0	—
Orn4 (ferrichrome A)	75.4	183.6	—
Gly-Phe-Gly	53.7	—	314.3
Ser5 (ferrichrome A)	17.2	—	354.5
Ser6 (ferrichrome A)	123.0	—	313.3
Thr-Phe-nitrobenzyl ester HBr	—	—	323.1

[a] Taken from Lakshminarayanan *et al.* (1967). References to the original literature may be obtained from this paper.

In most of the cases of peptides, Ramakrishnan and Ramachandran (1965) found that the observed (ϕ, ψ) values were within the allowed regions of the conformational map as shown in Fig. 12. However, a few cases of (ϕ, ψ) were found to occur in region IV, but these have a τ-value at the α-carbon larger than the tetrahedral angle, and would correspond to allowed conformations in a map for $\tau > 110°$.

2. Side-Chain Conformation

The conformational features of the side group have been discussed by Lakshminarayanan *et al.* (1967) with reference to the various rotations about the single bonds in the side chains. For example, the γ-atom in a side group would be described by a value χ^1 as shown in Fig. 35A. The figure shows how χ^1 is measured looking down the C^α—C^β bond. It is observed that χ^1 is distributed sharply around the values 60°, 180°, and 300°, corresponding to the staggered positions of the γ-atom. Similarly, the projection down the C^β—C^γ atom is shown in Fig. 35B. It shows how χ^2 defines the positions of the δ-atom. It is



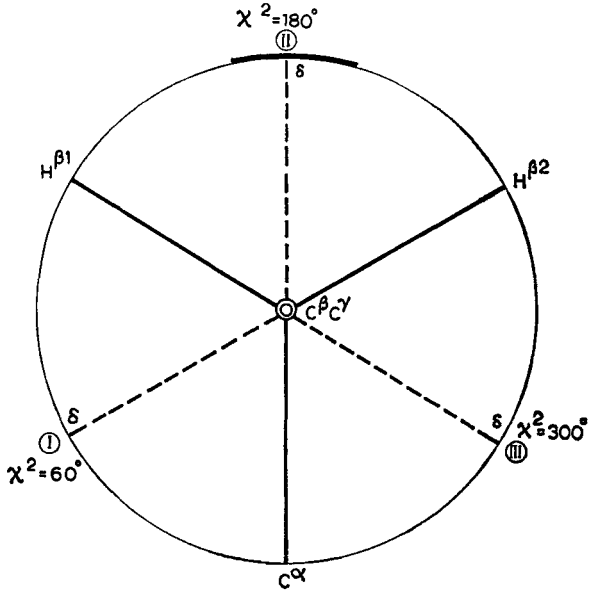
(a)

FIG. 35(a). Positions of the γ-atom looking down the bond C^{α}—C^{β}. The observed range of values of χ^1 are shown by thick lines on either side of the three positions I, II, III.

found that χ^2 occurs often near 180°, corresponding to the conformation in which C^{δ} is trans to C^{α} about the bond C^{β}—C^{γ}. The values of χ observed in various structures are reproduced in Table XXII.

The most important features of the conformation of side groups are as follows: Wherever the local arrangement is ethanelike, the three staggered configurations characterised by $\chi = 60°, 180°, 300°$ are preferred. For example, in arginine, the γ-carbon atom is found to occur at all the three positions in different crystals (Table XXII). For the δ- and ε-atoms, the position with $\chi \sim 180°$ is observed to occur in most of the cases of unbranched chains.

Cysteine, cystine, and serine show a striking behaviour. For these side groups, the γ-atom, which may either be an oxygen or a bulky sulphur atom, occupies the position with $\chi^1 \simeq 60°$ (with rare exceptions, when it is $\simeq 300°$). In this conformation it can be seen from Fig. 35A that S or O lies in between N and C′. Also for cystine, χ^2 is near about $\pm 90°$. The conformation about the S—S bond in cystine is also interesting. This angle, which may be denoted by χ^3, or χ^S, has a value near about $\pm 90°$. This means that the two bonds coming out of either sulphur atom occur in planes passing through the S—S bond which are



(b)

FIG. 35(b). The three positions of the δ-atom looking down the bond $C^{\beta}-C^{\gamma}$. The position II with $\chi^3 = 180°$ is the most common one.

approximately at right angles. This type of relationship is known to occur also for hydrogen peroxide, which has the structure H—O—O—H, and is explained by the interaction of the nonbonding electrons on the two oxygen atoms, or sulphur atoms, as is the case in cystine. Both a right-handed twist as well as a left-handed twist (i.e., $\chi^S = +90°$ and $-90°$) are observed for the L-configuration of cystine in different crystals. (Note a similar observation in lysozyme, discussed in Section VIII,C,2.)

Another interesting observation relates to the planar terminal groups in arginine, aspartic, and glutamic acids and the rings in histidine, tyrosine, etc. This plane is found to be either coplanar with, or perpendicular to, the plane defined by the previous three atoms. For example, in arginine HCl,H_2O and HBr,H_2O, the molecules A and B are characterised by $\chi^4 \sim -90°$ and $+90°$, while in other crystals containing arginine, it is near 0° or 180°. In aspartic acid, the end carboxyl group has χ^{21} and $\chi^{22} \simeq 0°$ and 180°, while in glutamic acid χ^{31} and χ^{32} are 0° and 180° ($\pm 20°$) and a similar observation holds for the terminal amide group of asparagine and glutamine. An exception is the γ-glutamyl side chain in glutathione, for which the plane is at right angles (χ^{31}, $\chi^{32} \approx \pm 90°$).

TABLE XXII

Values of χ (in Degrees) Observed for Amino Acid Side Chains[a,b]

Structure	χ^1	χ^2	χ^3	χ^4	χ^{51}	χ^{52}
			Straight chains			
Arginine 2HI	60.4	188.4	184.2	191.0	3.6	179.7
Arginine $2H_2O$	62.2	151.1	175.2	162.0	351.9	171.5
Arginine HCl (Molecule A)	170.8	187.4	172.5	188.3	352.9	175.0
Arginine HCl (Molecule B)	168.5	166.1	174.8	170.0	5.6	181.4
Arginine HCl·H_2O (Molecule A)	299.7	195.7	179.5	276.3	11.9	191.0
Arginine HCl·H_2O (Molecule B)	308.6	172.8	182.2	98.8	344.3	164.8
Arginine HBr·H_2O (Molecule A)	298.8	197.0	180.4	274.4	9.3	182.9
Arginine HBr·H_2O (Molecule B)	304.6	168.1	186.3	100.9	342.2	165.3
Cysteine HCl	64.6	—	—	—	—	—
S-Methylcysteine sulphoxide	294.9	—	—	—	—	—
Cysteinylglycine NaI (N-terminal)	64.9	—	—	—	—	—
Cysteinyl (in glutathione)	71.8	—	—	—	—	—
Cysteine ethyl ester HCl-urea	74.9	—	—	—	—	—
Cystine	55.3	81.5	73.8	—	—	—
Cystine 2HBr	70.6	271.1	278.7	—	—	—
Cystine 2HCl	69.1	271.3	280.9	—	—	—
N,N'-Diglycyl cystine (C-terminal)	64.0	263.1	281.0	—	—	—
Lysine HCl·H_2O	304.2	184.2	188.9	179.0	—	—
α-Methionine	299.9	176.9	80.5	—	—	—
β-Methionine	299.0	183.6	190.4	—	—	—
Norleucine	303.1	182.2	195.2	—	—	—
Ornithine 2 (ferrichrome A)	55.4	177.1	303.7	—	—	—
Ornithine 3 (ferrichrome A)	300.7	76.2	45.4	—	—	—
Ornithine 4 (ferrichrome A)	299.7	149.0	303.8	—	—	—
Serine	69.2	—	—	—	—	—
Serine 5 (ferrichrome A)	59.5	—	—	—	—	—
Serine 6 (ferrichrome A)	177.2 58.2	—	—	—	—	—

TABLE XXII (Continued)

Structure	χ^{11}	χ^{12}	χ^{21}
		Branched at C^{β}	
Isoleucine HBr	191.7	66.5	68.1
Isoleucine HCl	48.4	287.5	172.1
Threonine	185.5	305.2	—
Threonylphenyl-alaninenitrobenzyl ester HBr (N-terminal)	173.3	295.6	—
Valine HBr	74.0	195.5	—
Valine HCl	64.4	194.3	—
Valine HCl·H_2O	289.0	53.9	—
	χ^{1}	χ^{21}	χ^{22}
		Branched at C^{γ}	
Aspartic acid HCl	296.0	174.1	352.5
Aspargine H_2O	72.3	183.0	2.7
Glycylasparagine (C-terminal)	296.6	170.7	352.6
Histidine HCl·H_2O	71.5	239.5	61.1
Di-(Histidino)Zn·$5H_2O$	73.2	317.9	139.8
Di-(Histidino)Zn·$2H_2O$	70.9	314.8	134.4
β-Alanylhistidino Cu $2H_2O$	53.1	244.7	69.6
Leucine HBr	187.5	58.4	182.2
Leucylglycine HBr (N-terminal)	292.8	155.2	265.6
Leucylprolylglycine (N-terminal)	279.4	170.3	291.5
Phenylalanine HCl	62.1	83.6	262.4
Threonylphenyl-alaninenitrobenzyl ester HBr (C-terminal)	172.0	83.6	266.9
Glycylphenylalanyl-glycine (middle)	185.3	102.5	278.5
Tryptophan HBr	65.9	80.7	253.3
Glycyltryptophan $2H_2O$ (C-terminal)	294.1	60.6	237.7
Tyrosine HBr	187.3	64.7	250.3
Tyrosine HCl	185.2	64.3	243.7
Glycyltyrosine HCl (C-terminal)	189.8	65.3	241.6

TABLE XXII (Continued)

Structure	χ^1	χ^2	χ^{31}	χ^{32}
		Branched at C^δ		
Glutamic acid	309.9	282.9	204.6	25.0
Glutamic acid HCl	290.4	188.2	195.9	17.6
Glutamine	70.6	176.1	164.8	341.6
γ-Glutamyl (in glutathione)	289.0	291.0	103.3	287.9

[a] Taken from Lakshminarayanan *et al.* (1967).

[b] The χ^{ij} in this table are according to the definitions adopted in this review.

The distribution of the χ-values for the γ-, δ- and ϵ-atoms has been worked out by Lakshminarayanan *et al.* (1967) in terms of the number of instances in which a particular range of χ-values is observed. Their data are reproduced in Table XXIII. It is seen from this that the positions $\chi^1 \sim 300°$, corresponding to C^γ trans to COOH, and χ^2, $\chi^3 \sim 180°$, in which C^δ is trans to C^β and C^ϵ is trans to C^γ, are more populated than others. Use of such information is often helpful in working out protein conformation. Actually, Ooi *et al.* (1967) and Scheraga *et al.* (1967) have used assumptions of this kind for working out the conformations of homopolypeptides (discussed in Section VIII,E). In the

TABLE XXIII
Distribution of χ-values

Values around	Number of cases		Number of cases with values	
	Without O or S	With O or S	>Column 1	<Column 1
γ-Position				
60	13	11	17	7
180	10	2	8	4
300	21	—	7	14
δ-Position				
60	3	1	3	1
180	20	—	9	11
300	2	—	—	2
90	—	1	—	1
270	—	3	2	1
ϵ-Position (Excluding structures having a planar group at δ)				
60	—	—	—	—
180	10	3	8	5
300	—	—	—	—

absence of detailed torsional potential functions, the empirical observations described above would be highly valuable for such studies.

C. Conformations of Globular Proteins

In recent years, X-ray structure determinations of many crystalline proteins have been attempted. Brief accounts of the various attempts are given in the review by Dickerson (1964) and Davies (1965, 1967). The more accurately known structures are myoglobin, lysozyme, ribonuclease, carboxypeptidase A and tosyl-α-chymotrypsin, which have been solved by using calculated electron density distributions at 2 Å resolutions. At this level of resolution, the course of the polypeptide chains of these proteins has been traced in the unit cell and the various side groups could be recognized, although with some difficulty. It is therefore possible to study the conformation of the chains in such instances, with the reservation that, while the gross conformational features could be ascertained with reasonable certainty, the actual values of ϕ, ψ, and χ may be in error considerably.

Actually, a brief account of the conformations of myoglobin and lysozyme was given in Section V,B, where it was shown that the plots of the (ϕ, ψ) values for both these proteins lie mostly within the allowed regions of the contact map. In this section, we shall deal with these aspects in more detail, including the χ-values (made available to us by the kind courtesy of Dr. J. C. Kendrew) in the myoglobin molecule. However, for lysozyme, only the (ϕ, ψ) values are available (Blake *et al.*, 1967a,b) and in other cases only the gross conformational features have been ascertained at present. For example, bovine pancreatic ribonuclease A (Kartha *et al.*, 1967) has been reported to have a low content of helix, of only about two turns near the amino end and a couple more in the middle. The chain is cross-linked by four S—S bridges, and in between these the chains run roughly in an antiparallel array. Although a 2 Å resolution map has been obtained, no details of dihedral angles are available. Again, in the structure of bovine tosyl-α-chymotrypsin, obtained at a low resolution, Mathews *et al.* (1967) report that the polypeptide chain is almost entirely in the extended conformation with the exception of eight residues at the C-terminal end which form a short section of an α-helix, but further data are not available. The structure of the enzyme carboxypeptidase A was reported at a resolution of 6 Å (Lipscomb *et al.*, 1966) and very recently this has been refined to 2 Å resolution (Lipscomb *et al.*, 1967).

The structures of a few other globular proteins have been obtained at comparatively low resolutions, e.g., that of haemoglobin at 5.5 Å resolution (Muirhead *et al.*, 1967), papain at a resolution of 4.4 Å

(Drenth *et al.*, 1967) and ribonuclease S (Wyckoff *et al.*, 1967) at 3.5 Å resolution.

The importance of studies of this type lies in the fact that they help in finding out the general scheme that is involved in the building up of proteins, apart, of course, from the knowledge that the detailed structure provides to the chemist for his interpretation of the reactivity of the particular protein.

1. Myoglobin

The X-ray structure determination of myoglobin (Kendrew *et al.*, 1960, 1961) has revealed that the molecule is convoluted into a compact shape. The interior side chains of the molecule are in close van der Waals interaction and are mostly hydrophobic. Almost all the polar side chains are found in the outside of the molecule. These studies have also

TABLE XXIVA
Parameters of the α-Helical Regions in Myoglobin[a]

Helix	Number of residues	n[a]	h(Å)[a]	ϕ(°) Mean[b]	ψ(°) Mean[b]
A	16	3.63	1.50	123	134
B	16	3.83	1.47	128	126
C	7	[c]	[c]	123	142
D	7	3.63	1.45	125	129
E1	10	3.61	1.52	127	128
E2	10	3.67	1.40		
F	9	3.70	1.46	124	127
G	19	3.59	1.53	127	129
H	24	3.63	1.49	126	129

[a] From Kendrew (1962).
[b] Calculated from data supplied by Drs. Kendrew and Watson.
[c] Not given in Kendrew (1962).

confirmed the presence of several segments of α-helix along the chain. Typical helical parameters in sperm-whale myoglobin are given in Table XXIVA (Kendrew, 1962). This structure provided the first direct evidence for the existence of the α-helix as a stable conformation in a globular protein, and proved its right-handed form to be the one that occurs most readily.

The values of ϕ, ψ and some of the χ's for all residues have been made available to the authors recently (J. C. Kendrew and H. C. Watson, private communication). Dr. Kendrew has informed the authors that these data are not very accurate, because of the inherent difficulties in interpreting the electron density map at this resolution. A rough idea

of the accuracy to be expected is that ϕ, ψ and χ^1 may have errors of the order of 20° and the other χ's more than 30°. The values are therefore subject to considerable revision during further refinement of the data, which is now proceeding in Dr. Kendrew's laboratory. In view of this fact we are not presenting here the table of data provided by Drs. Kendrew and Watson but instead will offer a general discussion of the data and a plot of the distribution of the angles ϕ and ψ.

The (ϕ, ψ) plot of the nonhelical segments has already been discussed in Section V,B. It may be recalled that most of the conformations lie in the allowed regions of the contact map. Of particular interest are the two groups of conformations, one clustering around the α-helical conformation, some of which lie on the side of bad contacts, and the other in the bridge region IV. These conformations are actually found in the region of low potential energy in the potential map. Only one point Gln (260°, 260°) appears to lie in the disallowed region, but it is not highly disallowed, as the potential energy map extends in this direction (see Section VII,E for the potential maps).

Another interesting aspect found in the myoglobin plot which justifies the theoretical calculations is the occurrence of three glycyl conformations well outside the allowed region on the alanyl contact map and in a portion where the β-carbon has severe short contacts. As such, the contact map predicts that these residues should be glycyl and this is actually the case (Ramakrishnan and Ramachandran, 1965).

The (ϕ, ψ) values for the helical regions alone of the molecule are shown plotted in Fig. 36. For comparison with theory, the part of the allowed conformations of the poly-L-alanine map in that region is also shown superposed. It will be seen that there is appreciable scatter in this region, whereas a strictly regular helix would be represented by a single point on the (ϕ, ψ) diagram. However, it can be seen that most of the plotted points occur in or just outside the allowed regions. The expected positions of the conformation of an α-helix in this map, corresponding to $n = 3.6$ and $h = 1.5$ Å are shown by A and B in the diagram. These two points are for values of the angle τ at the α-carbon atom = 109.5° and 110° respectively (see Table V). It is found that although the individual points are widely scattered, the mean of these for the different helices are not so different. The mean values of (ϕ, ψ) are given in Table XXIVA and it will be noticed from this that these are closely corresponding to one another for all the helices except helix C. The grand average value for (ϕ, ψ) for all the points plotted in Fig. 36 is (126°, 129°). According to Ramachandran *et al.* (1966b), the conformation (125°, 130°) corresponds to $n = 3.62$, $h = 1.50$ Å and a hydrogen bond distance of 2.88 Å and angle of 11°. Also, this mean conformation

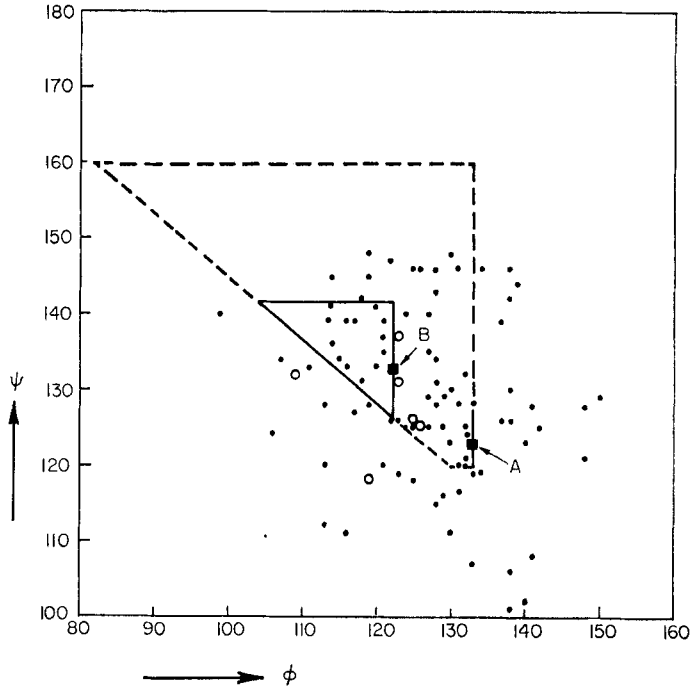


FIG. 36. Distribution of the values of (ϕ, ψ) in the helical regions of myoglobin. The large spread is not significant as the data have an accuracy of only about 20°. The (ϕ, ψ) position corresponding to α_P with $n = 3.6$, $h = 1.5$ Å are marked for $\tau = 109.5°$ (A) and $\tau = 110°$ (B). The lines marked enclose the allowed regions for an alanyl helix, ——— fully allowed, ----- extreme limit.

is close to the one calculated for poly-L-alanine by Arnott and Wonacott (1966), which is referred to in Section VIII,E,1,a. It is interesting that the mean conformation corresponds to one in which the peptide plane is not vertical (i.e., parallel to the axis of the helix) but slightly inclined, so that the N—H group points inwards, as originally suggested by Sasisekharan (1962). In fact, this tendency is even more predominant for the residues in helix C, which take up a conformation midway between the α-helix and the 3_{10}-helix, on the average. This feature is also found with the helix C of lysozyme (Table XXIVB), as may be seen by comparing the mean (ϕ, ψ) values of the corresponding helices in the two proteins. Further, the last peptide unit in the helix in almost all the helices of myoglobin tend to take up this type of conformation, namely, tending towards the 3_{10}-helix.

It has been suggested by Némethy *et al.* (1967) that this conformation, which is deviated from that of the α-helix, does not really go

towards that of an exact 3_{10}-helix (which is energetically unfavourable as discussed in an earlier section) but moves diagonally upwards and towards the left in the conformational map. It will be seen from Fig. 30 that there is an extended valley in the potential map running in this direction. Also, as pointed out by Némethy *et al.* (1967), there is a second solution for $n = 3.6$ and $h = 1.5$ Å which is different from the α-helix and which is shifted in this direction. Therefore, without disturbing the arrangement of the α-carbon atoms as in the standard α-helix, the peptide unit alone can rotate, so that it takes up the second conformation. If this happens at any one or more of the last three peptide units at the C-terminal end of an α-helix, then the corresponding N—H groups point approximately midway between the C=O groups of the fourth and the fifth preceding residues, thus forming a bifurcated hydrogen bond of the types 5—1 and 4—1. It is interesting that a similar feature exists at the C-terminal end of a number of helices in myoglobin before it becomes nonhelical. This second conformation has been termed the α_{II}-helix by Némethy *et al.* (1967). This however, does not seem to be a suitable term because a complete helix with all the residues having this particular conformation is not stabilised by hydrogen bonds and is probably unlikely to occur. (Leach *et al.* 1966b; Némethy *et al.* 1967.) A better notation might be "α'-conformation" denoting that this is an alternative conformation which can occur at isolated places in the α-helix, which is stabilised in the usual way by 5—1 hydrogen bonds.

The distribution of the position of the γ-atoms in the helical sections is given in Fig. 37 (the first three residues and the last residue in each helix are omitted, following Kendrew, 1967, private communication). At

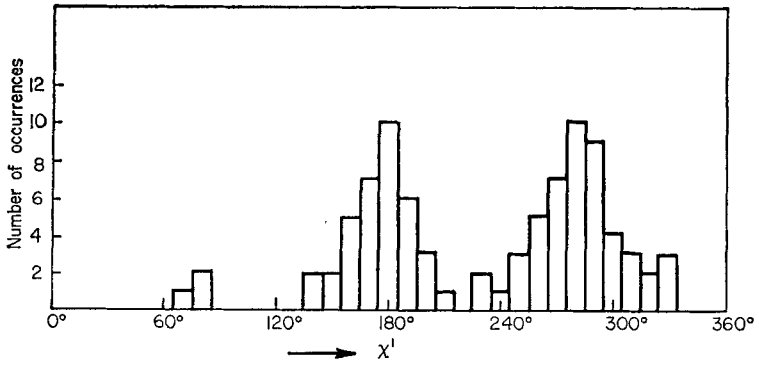


FIG. 37. Distribution of the values of χ^1 in the helical regions of myoglobin. Note the preponderance of values near position III (300°) and of only two examples in position I(60°). The latter are a serine and threonine residue (see text for explanation).

the current stage of refinement, the scatter of the values of χ^1 is believed to be really not significant, although the preponderance of position III is worthy of note, as also the fact that only three values of χ^1 occur in the 60° region, one for a seryl γ-oxygen in the helical section F, the other for a threonyl γ-oxygen in the helical section C and the third for an isoleucyl γ-carbon atom in the helical section G. Actually, as was pointed out in Section V,D, a γ-carbon atom in position I ($\chi^1 \simeq 60°$) is unlikely in an α-helix, because it has bad contacts with the backbone. But, on the other hand, a γ-oxygen atom at this position is favourably disposed to form a hydrogen bond with one of the backbone oxygens (as mentioned in Section V,D). The occurrence of a γ-carbon atom of the isoleucyl residue in position I requires further study. It is interesting that, even in the nonhelical regions, the position III of the γ-carbon atom, namely $\chi^1 \simeq 300°$, is the one that is most common—18 out of 25 examples occur in position III.

2. Lysozyme

The structure of hen egg-white lysozyme molecule has been studied by Blake *et al.* (1967a) and a brief discussion of the chain conformation was given in Section V,B. We shall now consider its conformation in more detail. The lysozyme molecule is a chain of 128 amino acid

TABLE XXIVB
Parameters of Helical Regions in Lysozyme[a]

Helix	Number of residues	n	h(A)	ϕ(°) Mean[b]	ψ(°) Mean[b]
A	11	3.59	1.51	127	128
B	11	3.66	1.48	120	133
C	6	3.38	1.66	119	150
D	9	3.72	1.45	128	128

[a] From Blake *et al.* (1967a).
[b] Calculated from data in Table XXV.

residues folded into a compact structure and held by four S—S bridges. Here also, most of the polar side groups are located on the periphery of the molecule and the majority of hydrophobic groups are in the interior. In contrast to myoglobin, which consists of long stretches of α-helix, lysozyme is found to have only a small proportion of helix. The helical parameters of these sections are given in Table XXIVB (from Blake *et al.*, 1967a). It can be seen that there is considerable distortion from the "classical" α-helix for helix C and even in the other helices, a

number of conformations are displaced towards the 3_{10}-helix region (i.e., to larger values of ψ). The observations made in the case of the last peptide unit making a bifurcated hydrogen bond hold here also in a majority of cases.

Table XXV gives the observed data of (ϕ, ψ) in the various residues. These (ϕ, ψ) values were shown plotted on the contact map in Fig. 16. Here, the same data are shown plotted in Fig. 38A and B along with the potential energy contour diagrams for an alanyl α-carbon atom corresponding to the functions F and K_2 of Section VII,C. Figure 38C is a corresponding map reproduced from a paper by Dunnill (1965). The parameters of the functions used for this map are not available. The agreement with these maps, particularly with the second and third ones, is remarkable. In fact, most of the nonglycyl residues are enclosed within the zero energy contours in all the cases. The two aspartic acid residues a little below the α_M-conformation may refine at a later stage and move into a region of lower potential. So also, the serine residue occurring at about (180°, 290°) is found on a ridge in Fig. 38A, but is at the boundary of a bridge in Figs. 38B and 38C. If slight variations in the geometry of the peptide unit are made, it is very likely to have a lower energy. The only doubtful point is the phenylalanine residue plotted near (300°, 150°). However, the contours in Figs. 38B and 38C extend towards this, showing that it is also not very unlikely. Dr. A. C. T. North (personal communication) informs the authors that these conformations, which are a little outside the low energy regions, occur in portions of the lysozyme structure which are not very clearly resolved and their exact (ϕ, ψ) values are subject to revision. Also, the same comment as was made with myoglobin is true here also, namely that all ϕ and ψ are likely to have errors of the order of 20°.

Another interesting feature is observed in the region of residues 41 to 54, where the chain doubles back on itself to form an arrangement similar to the antiparallel pleated sheet, with hydrogen bonds between the two lengths of the chain perpendicular to the direction of the progress of the chain. This type of beta-structure similar to that in silk fibroin (see Section VIII,F) has been observed so far only in this globular protein.

A third interesting feature is the nature of the disulphide bond dihedral angles. All four of these have a magnitude close to 100°, which is typical of what is found in cystine structures, as mentioned in Section VIII,B. However, it is interesting that two of these are right-handed and two left-handed. In fact, such an occurrence of either a positive or a negative value of χ^S has been observed in cystine compounds also, all in the L-configuration.

TABLE XXV

Conformational Parameters ϕ and ψ (in Degrees) Observed in the Lysozyme Molecule[a]

Residue		ϕ	ψ
1	LYS	—[b]	—[c]
2	VAL	56	301
3	PHE	107	338
4	GLY	79	336
Helical section—A[d]			
5	ARG	133	115
6	CYS	143	116
7	GLU	127	153
8	LEU	119	121
9	ALA	126	130
10	ALA	115	140
11	ALA	141	121
12	MET	109	153
13	LYS	123	114
14	ARG	142	105
15	HIS	117	177
16	GLY	277	172
17	LEU	128	146
18	ASP	105	330
19	ASN	241	182
20	TYR	103	319
21	ARG	242	220
22	GLY	240	202
23	TYR	85	298
Helical section—B			
24	SER	97	352
25	LEU	149	131
26	GLY	123	132
27	ASN	128	126
28	TRP	120	138
29	VAL	102	132
30	CYS	99	147
31	ALA	125	112
32	ALA	115	150
33	LYS	115	129
34	PHE	108	151
35	GLU	98	131
36	SER	77	156
37	ASN	251	181
38	PHE	292	153
39	ASN	96	292
40	THR	110	165
41	GLN	94	139
42	ALA	143	330
43	THR	36	330
44	ASN	46	285
45	ARG	76	315
46	ASN	90	3
47	THR	111	161
48	ASP	87	217
49	GLY	277	107
50	SER	185	291
51	THR	36	332
52	ASP	69	303
53	TYR	65	314
54	GLY	255	352
55	ILE	134	145
56	LEU	72	190
57	GLN	233	228
58	ILE	115	318
59	ASN	92	343
60	SER	75	185
61	ARG	99	163
62	TRP	45	139
63	TRP	94	146
64	CYS	27	331
65	ASP	83	329
66	ASN	35	185
67	GLY	251	180
68	ARG	46	200
69	THR	60	270
70	PRO	132	141
71	GLY	99	171
72	SER	128	313
73	ARG	51	314
74	ASN	61	262
75	LEU	124	135

[a] From Blake *et al.* (1967a).

[b] Not relevant.

[c] Not stated in the reference.

[d] The residues included in this and the other helical sections are those included within the brackets.

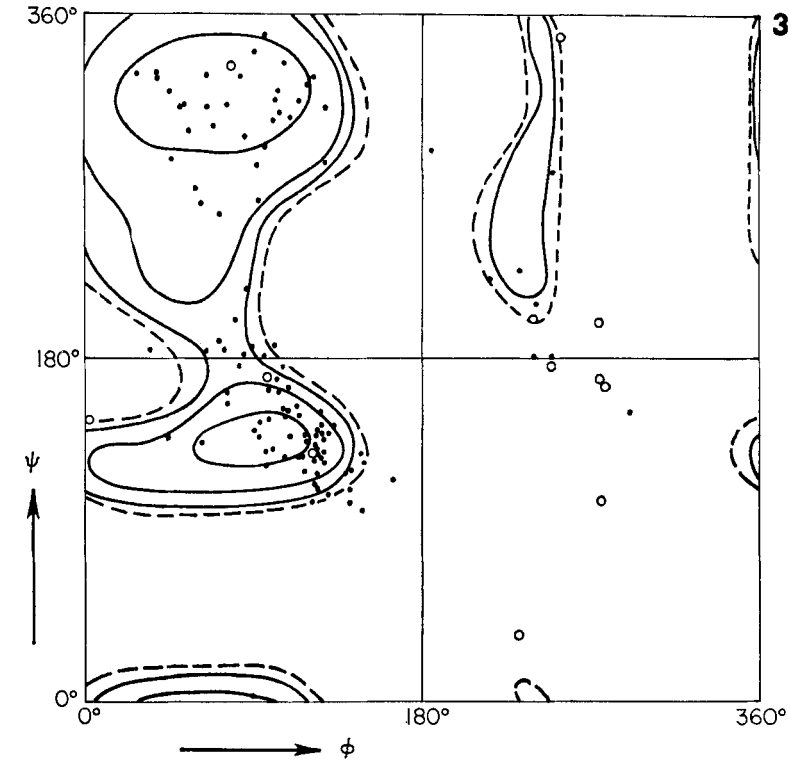
TABLE XXV (Continued)

Residue		ϕ	ψ	Residue		ϕ	ψ
76	CYS	81	201	101	ASP	98	124
77	ASN	218	223	102	GLY	4	149
78	ILE	45	321	103	ASP	85	181
79	PRO	102	309	104	GLY	234	37
Helical section—C				105	MET	103	188
80	CYS	166	117	106	ASN	108	154
81	SER	124	141	107	ALA	113	161
82	ALA	128	138	108	TRP	72	258
83	LEU	128	141	109	VAL	150	126
84	LEU	96	183	110	ALA	118	141
85	SER	120	326	111	TRP	124	127
86	SER	91	187	112	ARG	123	113
87	ASP	52	312	113	ASN	123	135
Helical section—D				114	ARG	77	163
88	ILE	111	177	115	CYS	64	136
89	THR	144	122	116	LYS	127	285
90	ALA	149	100	117	GLY	276	202
91	SER	126	120	118	THR	65	350
92	VAL	125	134	119	ASP	93	263
93	ASN	114	150	120	VAL	116	157
94	CYS	109	129	121	GLN	125	142
95	ALA	128	141	122	ALA	121	161
96	LYS	117	129	123	TRP	116	142
97	LYS	105	162	124	ILE	83	176
98	ILE	91	141	125	ARG	110	309
99	VAL	103	169	126	GLY	279	168
100	SER	66	184	127	CYS	104	321
				128	ARG	92	283

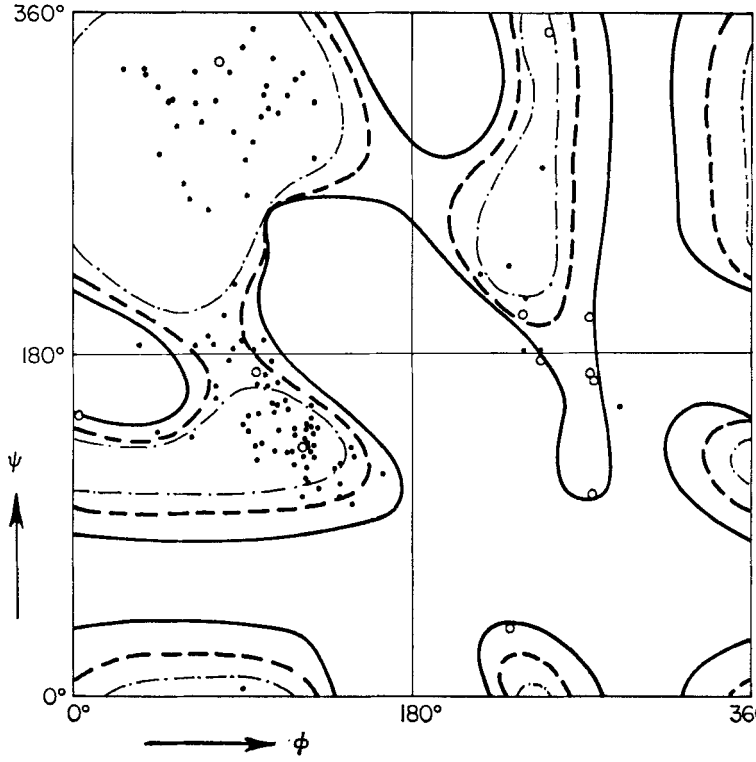
D. Conformations of Fibrous Proteins

1. Introduction

The observed conformations of peptides and globular proteins were described above in terms of the dihedral angles ϕ, ψ, and χ. This was possible because fairly precise atomic coordinates were available from single crystal structure analysis of these compounds, by X-ray diffraction methods. However, in the case of fibrous proteins, as the name implies, one is dealing with filamentous bundles of polypeptide chains rather than convoluted finite molecules and this makes the X-ray analysis of fibrous proteins different from that of the globular proteins. It is not possible to derive their structure from X-ray data; on the other hand, satisfactory models of conformations for the fibrous proteins are



38A



38B

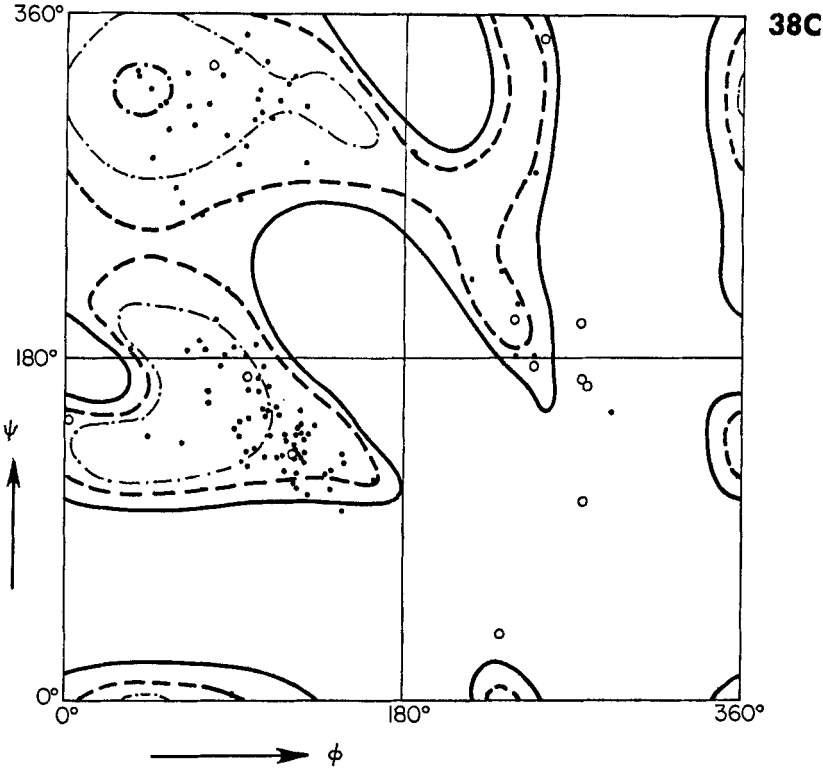


FIG. 38. Plot of (ϕ, ψ) values of lysozyme along with energy contours. A. Using potential functions F. ----, zero contour; ———, contours at intervals of 1 kcal/mole going down to −3 kcal/mole. B. Using the potential functions K_2. ---, zero contour; ———, +1 kcal/mole -·-·- (thin), −0.5 kcal/mole; C. As given by Dunnill (1965). Contours as in B. In addition -·-·- (thick) corresponds to −1.0 kcal/mole.

built up, which could then be tested against the X-ray diffraction pattern. In this trial-and-error method, use is made of auxiliary information, such as infrared dichroism, known crystal structure data from related simple molecules, electron microscopic studies and so on. Table XXVI, adapted from a paper by Ramachandran (1962), gives the prominent features of the X-ray diffraction patterns and infrared spectra of the commonly occurring forms of fibrous proteins.

The common mode of analysis is to compare the actually observed diffraction pattern with the calculated diffraction pattern corresponding to the proposed helical structure. The inverse procedure, namely using Fourier methods with observed diffraction data to obtain electron density maps, so commonly used in single crystal work, has not been done for the

TABLE XXVI

Main Features of X-ray Diffraction and Infrared Spectra of Fibrous Proteins

Protein structure	X-ray Reflections			Infrared dichroism
	Equatorial	Meridional	Other features	
α-Form (Normal form of proteins of the keratin-myosin family)	9.5 Å	5.17 Å 1.5 Å	The 5.17 Å is a wide arc. The true meridional reflection is at 1.5 Å.	Both NH stretching (3300 cm^{-1}) and CO stretching (1650 cm^{-1}) have parallel dichroism
β-Form	9.7 Å 4.8 Å	3.3 Å	Layer at 6.6 Å	Both bands have perpendicular dichroism
(i) of keratin, etc.				
(ii) of silk	9.7 Å 4.7 Å	3.5 Å	Layers at 7.0, 3.5, 2.3 and 1.75 Å.	Both bands have perpendicular dichroism
Cross β-form (Supercontracted form of keratin, etc.)	10 Å	4.7 Å	If resolved, as in *Chrysopa* egg stalk silk,[a] equatorial spots near 5 Å and layer line at 9.4 Å.	Both bands have parallel dichroism
Collagen	11 Å 4.5 Å (diffuse)	2.9 Å	Layers at 10 Å and 4 Å.	Both bands have perpendicular dichroism

[a] See e.g., Dickerson (1964, p. 674).

fibrous material. This is mainly due to the poor quality of the fibre diffraction photographs and the limited number of reflections that are observed, which is due to the inherent disorder present and other similar factors.

Similarity of wide-angle X-ray pattern has been used, as was first done by Astbury (1940), to imply similarity of structure at the atomic level in fibrous proteins. For example, α-keratin and muscle myosin have been classified as having the same α-helical structure, although the building up of these α-helical units into macromolecular aggregates might be quite different.

In this and the subsequent subsections we shall therefore discuss the conformations of these fibrous materials in a general way, emphasising the results of the X-ray diffraction methods to correlate the proposed structure for these substances. Excellent reviews dealing mainly with the X-ray diffraction of protein and polypeptide structures are available (Dickerson, 1964; Davies, 1965). The details presented here are by no means exhaustive. They are restricted to the extent to which the details are relevant to the conformation.

2. α-Helical Proteins

To this class belongs, for example, the α-form of keratin. All the α-forms of fibrous proteins give a characteristic wide-angle X-ray diffraction pattern, as mentioned in Table XXVI. This X-ray pattern could not be accounted for simply by parallel α-helices. The structures giving rise to this kind of pattern are now generally assumed to consist of some form of coiled-coil structures (Crick, 1952, 1953; Pauling and Corey, 1953a), in which the α-helical protofibrils are twisted as in a rope. The question as to how many α-helices form the primary coiled-coil has not been answered definitely as yet. Fraser and his colleagues (Fraser and MacRae, 1961a,b; Fraser *et al.*, 1962a) had indicated in their earlier studies on α-keratin that the agreement was best between the calculated and observed intensities for a three-stranded coiled-coil. However, a detailed investigation carried out by Cohen and Holmes (1963) on paramyosin showed that the agreement was better with a two-stranded α-helical coiled-coil.

More recently, Fraser *et al.* (1965d) have compared the intensity distribution in the X-ray patterns of paramyosin, myosin, α-keratin, and tropomyosin in the dry state. From their calculations they have concluded that myosin, α-keratin, and tropomyosin should also have a two-stranded coil like the paramyosin molecule. Recently, Atkins (1967) has shown evidence for the presence of a four-stranded α-helical coiled-coil in some types of silk which occur in the α-form.

3. Beta Structures in Proteins

The X-ray patterns provided by stretched keratin, by silk fibroin, and also by certain synthetic polypeptides are distinctly different from that of the α-form. The prominent spacings of these so-called β-forms of proteins are listed in Table XXVI.

Silk contains a high proportion of glycine and alanine and a small amount of serine. The structure of silk was successfully tackled by Marsh *et al.* (1955a,b), who concluded that the polypeptide chains in silk are more or less completely extended, having a two-fold axis along the chain axis, and that adjacent chains run antiparallel, such that N—H groups of one chain could easily form hydrogen bonds with the C=O groups of neighbouring chains in a sheetlike manner. The structure of silk from different sources has been reviewed by Rudall (1962, 1963).

In the case of β-keratin, the polypeptide chain is not fully extended as in silk, and a parallel sheet structure has been proposed by Pauling and Corey (1953b). In this structure, the polypeptide chains run parallel and are again hydrogen-bonded in the form of sheets. However, a more recent investigation on the β-keratin of stretched hair of horse-tail carried out by Fraser and MacRae (1962b) showed that the agreement between the observed pattern and the calculated one was best only for a heterogeneous structure containing both parallel and antiparallel sheets. In either case, the NH and CO bonds are nearly perpendicular to the fibre axis and they both should exhibit perpendicular infrared dichroism for their stretching frequencies, as actually observed (Table XXVI).

A variant of the extended β-form is the so-called "cross β-form" which is exhibited by supercontracted keratin. In this, the chains run at right angles to the fibre axis in a zig-zag manner and they are then hydrogen-bonded to one another as in the antiparallel pleated sheet, except near the turning points. The 4.8 Å reflection on the equator in the extended β-form therefore occurs on the meridian in the cross β-form and the infrared dichroism is parallel (Table XXVI) in this structure. The best example of the cross β-form giving a well-resolved X-ray pattern is that of *Chrysopa* egg stalk silk (Parker and Rudall, 1957).

The exact nature of the conformation near the turning points in the cross β-structure is not clear. Astbury *et al.* (1959) used three units which are not hydrogen bonded in this region. However, the studies reported in Section V,E indicate that one such unit might be sufficient.

4. Triple-Helical Conformation of Collagen

Fibrous proteins of the collagen group give a wide-angle diffraction pattern which is distinctly different from that shown by the above two forms of proteins. The most important spacings are shown in Table XXVI. The structure of collagen was for many years the subject of intensive investigation by many workers. Several structures were proposed, but none correctly until 1955, when the correct clue to the structure was given by Ramachandran and Kartha (1955a,b) in terms of a triple-chain coiled-coil. It was shown that agreement between the observed pattern and the calculated one was best for a coiled-coil structure consisting of three polypeptide chains shown schematically in Fig. 39. Each chain has very nearly 3.3 units per turn and has an average unit height of 2.9 Å. The three chains (A,B,C of Fig. 39) occur side by side and they are all wound about a common central axis by a twist of 30° for every three units. The chains B and C are derived from chain A by successive screw operations of a rotation of 110° and a

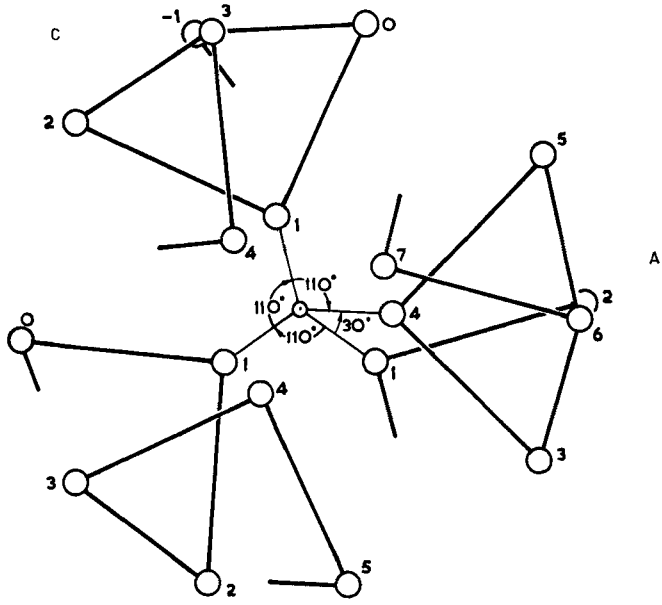


FIG. 39. Schematic diagram of the triple chain coiled-coil structure of collagen, showing the α-carbon atoms as circles and the peptide units as lines joining them. The three chains are marked A, B, C and the atoms with the same numerical symbols in the three chains are equivalent, as regards the helical coiling about the central axis.

translation parallel to the helix axis by 2.9 Å. The chains are held together by a system of hydrogen bonds between them, so that there are two bonds formed for every three residues and the contacts between atoms are such that every third residue along each chain could only be a glycine. Studies of the amino acid composition of collagen invariably show that glycine makes up one-third of the total number of residues, and approximately another third consists of proline and hydroxyproline. The NH and CO bonds are nearly perpendicular to the fibre axis as required by infrared dichroism. A few modifications of the above structure were proposed to incorporate the amino acid sequence Gly-Pro-Hypro (Rich and Crick, 1955, 1961; Cowan *et al.*, 1955; Ramachandran *et al.*, 1962). When this sequence occurs, the structure can have only one hydrogen bond occurring for every three residues. Recently, after a thorough investigation, the structure of collagen was refined by Ramachandran and Sasisekharan (1965). A projection of the molecular structure is shown in Fig. 40. It contains, in addition to two NH . . . O hydrogen bonds, one CH . . . O bond for every three residues in each chain.

Collagen occurs in all connective tissue and the typical collagen X-ray

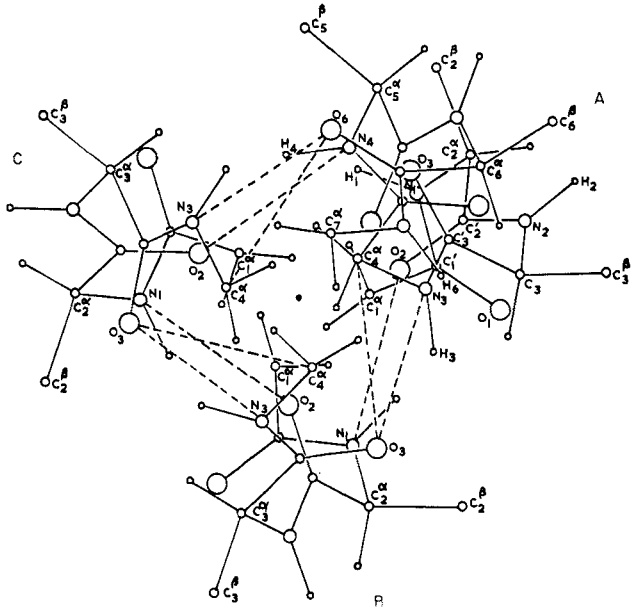


FIG. 40. Projection of the detailed molecular arrangement in the triple helical structure of collagen, projected down the helical axis on a plane perpendicular to the axis.

pattern indicating the triple helical structure is shown by all these materials. A recent review of the molecular structure of collagen has been given by Ramachandran (1967). Apart from material that is classified as collagen by the biologists, some silks which have a high proline content also exhibit the typical collagen X-ray pattern (Lucas and Rudall, 1967; Rudall, 1967).

5. Other Types of Structures

Apart from the three standard types of structures discussed above, other types of X-ray patterns have also been observed. Feather keratin exhibits a particularly rich pattern, but its structure is not definitely known. Entirely different structures have been proposed by Schor and Krimm (1961a,b) Ramachandran and Dweltz (1962), and Fraser and MacRae (1962a, 1963).

Recently, E. D. T. Atkins (personal communication) has observed a very rich X-ray diffraction pattern given by a silk, which is different from the pattern given by any other fibrous protein known so far. It is being analyzed. So, it is still worthwhile examining fibrous protein materials from various biological systems not examined so far, in order to find out if newer types of organization of protein chains exist in nature.

E. Conformations of Synthetic Polypeptides

The synthetic polypeptides, in general, give much better X-ray diffraction patterns than the fibrous proteins. They can be obtained in the form of well-oriented fibres of films, both for X-ray diffraction and infrared studies. Consequently, a good deal of attention has been paid to the various forms of synthetic polypeptides in recent years. We shall describe below the details of the more recent investigations on these polypeptides.

1. Synthetic Polypeptides of the Alpha and Beta Forms

a. The Alpha Form. Poly-γ-benzyl-L-glutamate was the first one to be obtained in a highly oriented form and studied by X-ray diffraction and infrared methods (Bamford *et al.*, 1951, 1952; Ambrose and Elliott, 1951; Perutz, 1951). X-ray patterns and infrared data suggested the polymer to be in the α-helical form. It may be said that the confirmation of the α-helix structure proposed by Pauling, Corey, and Branson (1951) came first from the above studies on poly-γ-benzyl-L-glutamate. Soon, it was shown that the diffraction patterns of poly-γ-methyl-L-glutamate and poly-L-alanine were very similar to that of

poly-γ-benzyl-L-glutamate. They may also, therefore, be considered to exist in the solid state as α-helices.

Most polypeptides can be obtained in this α-form from dichloroacetic acid solutions. A detailed study of poly-L-alanine was made by Brown and Trotter (1956) who concluded that the agreement between the observed data and the calculated intensities was better for the left-handed α-helix than for the right-handed form. However, Elliott and Malcolm (1959), in a careful study, showed that the agreement was good for the right-handed helix and not so for the left-handed one. They also pointed out that the observed data could be easily explained on the basis of packing of helices with random arrangement of chain directions, going either up or down. Very recently, a refinement of the structure of poly-L-alanine has been made by Arnott and Wonacott (1966), making use of accurately measured intensities. They have given atomic coordinates for the α-helix occurring in this polymer based on their refinement. Energy values for the packing of the helices have also been computed using nonbonded interactions. The conformational parameters of the α-helix present in this polymer have been reported to be $\phi = 113.1°$ and $\psi = 136.3°$ with $\omega = -1°$. The hydrogen bond length is 2.87 Å, but the bond is not linear, the deviation from linearity being 12°. Sasisekharan (1962) had shown that such a tilt of the peptide planes from the vertical, which makes the NH groups point slightly inwards and the CO groups point outwards, making the NH . . . O bond nonlinear by 10°–15°, relieves the short contact C′(2) . . . O(1), and is therefore expected to occur. Arnott and Wonacott in fact found this contact to be 2.90 Å in their refined model, which is well above the normal limit value, unlike the earlier model of Elliott and Malcolm (1959), where this contact was short.

Tsuboi *et al.* (1961) and Tsuboi (1964) also investigated the racemic mixture of L- and D-forms of poly-γ-benzyl glutamate. They observed that the racemic mixture was also α-helical, just as the enantiomorph. However, the X-ray pattern and the infrared data were found to be slightly different for the two cases. Elliott *et al.* (1965), who also investigated these polymers, attributed the difference in the patterns and the infrared data to distortion in the α-helix and to disorder arising out of interactions between the side chains. Very recently, Mitsui *et al.* (1967) reinvestigated the problem and concluded that, in the enantiomorphous form, the side chains are not as well ordered as they are in the racemic mixture and that the α-helical conformation is present in both the forms.

Investigations on the copolymers of the synthetic polypeptides have also been carried out. Both X-ray work and solution studies have

indicated the presence of α-helical conformation for the copolymers. Mention may be made of the recent work by Vollmer and Spach (1967) who have studied the copolymers of *O*-carbobenzoxy-L-tyrosine and benzyl-L(D)-glutamate and benzyl-L-aspartate respectively, using optical rotatory dispersion and X-ray diffraction methods. They all exhibit the α-helical structure.

Fraser *et al.* (1967) have investigated the effect of glycyl residues on the stability of the α-helix by studying ordered sequences of glycyl and γ-ethyl-L-glutamate residues along the polypeptide chain, using X-ray diffraction, infrared, and optical rotatory dispersion techniques. They have concluded that glycyl residues appreciably reduce the stability of the γ-ethyl-L-glutamate helix. This is further discussed in Section VIII,F.

b. The Beta Form. The synthetic polypeptides can also be obtained in the β-form. As in the case of fibrous proteins, it is possible to convert the α-form to the β-form and vice versa by physical and chemical methods. For example, a film of the α-form of poly-L-alanine can be converted to the β-form by stretching in steam. A number of polypeptides can also be obtained in the β-form by casting from solutions in formic acid.

Recently, the β-form has been observed to occur in solution from studies on infrared and ORD. The earliest evidence was obtained by Yang and Doty (1957) with oligomers of γ-benzyl-L-glutamate. Later, Rosenheck and Doty (1961) observed an α-β transition for poly-L-lysine. Yang (1967) believes the β-form of poly-L-lysine to be of the antiparallel type, as in silk fibroin, but Sarkar and Doty (1966) believed it to be of the cross-β type. Davidson *et al.* (1966) consider that the β-conformation receives additional stabilization energy from hydrophobic side-chain interactions. Poly-L-lysine hydrochloride, which has a β-conformation in solution under suitable conditions, has been obtained in a dry form from solution and shown to have a β-pleated sheet structure (Traub *et al.*, 1967).

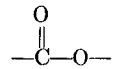
The β-form of poly-L-alanine is similar to that of silk. Actually, the X-ray patterns of tussah silk and poly-L-alanine are strikingly similar (Bamford *et al.*, 1954; 1956). Neither parallel nor antiparallel sheets nor any systematic alternations of these gave quantitative agreement with the observed X-ray pattern and the conclusion arrived at was that random variations of chain sense could occur (Brown and Trotter, 1956; Bamford *et al.*, 1956). β-Forms have also been obtained for quite a few other polymers and copolymers (Bamford *et al.*, 1956). Recently, Fraser *et al.* (1965c) have examined films of poly-L-alanylglycine, which forms a pleated sheet β-structure similar to that proposed by Marsh

et al. (1955a,b). However, the sheet separation between the glycine surfaces is greater (3.9 Å) than the value (3.5 Å) of Marsh *et al.*, with the alanine–alanine sheet distance being less (5.3 Å) than the earlier value of (5.7 Å).

2. The Omega Helix for Polypeptides

In solution, optical studies indicated that poly-β-benzyl-L-aspartate is in the left-handed α-helical form (Blout and Karlson, 1958; Karlson *et al.*, 1960; Bradbury *et al.*, 1960; for a review, see Blout, 1961). However, the α-form of the polymer gave only a poor X-ray pattern, but it was observed that oriented films of poly-β-benzyl-L-aspartate, heated to 160°C and cooled, gave a good X-ray diffraction pattern different from that of the α-form of the polymer (Bradbury *et al.*, 1959). The structure which gave rise to the rich X-ray pattern was designated as the ω-form. The ω-form of poly-β-benzyl-L-aspartate was thoroughly studied by X-ray diffraction and infrared methods by Bradbury *et al.* (1962). This form has been shown to be characterized by a helix having an exact four-fold screw axis with an axial translation of 1.325 Å. It was pointed out by Bradbury *et al.* (1959) that these dimensions are similar to those proposed by Bragg *et al.* (1950) for the 4_{13}-helix. In the ω-form, the hydrogen bonds are nonlinear and the peptide units are not planar. Bradbury *et al.* (1962) have concluded that steric conditions are favourable only for a left-handed ω-form with nonplanar peptide units.

The ω-form has also been observed for poly-*S*-benzylthio-L-cysteine (Fraser *et al.*, 1962b). This polymer is similar to the aspartate polymer discussed above but with the ester



group replaced by a disulfide group —S—S—. It was noted that the exact length of the side group may be a determining factor in stabilizing the ω-form, for poly-*S*-benzyl-cysteine with only one sulfur in place of the disulfide gave only a poor α-helix pattern under the same conditions of preparation.

3. Conformation of Polyimino Acids

Poly-L-proline in the solid state exhibits two modifications, which have been named poly-L-proline I and II. The X-ray diffraction patterns of the forms I and II are distinctly different from one another and also from the α- and β-forms of synthetic polypeptides discussed above. The structure of poly-L-proline II was first examined by X-ray diffraction by Cowan and McGavin (1955). They could interpret the pattern in terms of a left-handed helix with a three-fold axis, consisting

solely of trans peptide groups. The helix must necessarily be left-handed due to steric reasons, namely that $\phi \sim 100°$. The structure is topologically similar to one of the three chains of collagen and has in fact values of $n = 3$ and $h = 3.1$ Å, identical with the first triple-helical structure proposed for collagen by Ramachandran and Kartha (1954), before they proposed the coiled-coil structure. Sasisekharan (1959a) has reexamined the X-ray data and has given detailed co-ordinates for the atoms in the structure. This structure is stabilised by CH . . . O hydrogen bonds.

Traub and Shmueli (1963) investigated poly-L-proline I and have concluded that the X-ray data would fit best with a right-handed helix having cis peptide units with 3⅓ units per turn and a unit height of 1.90 Å. There are no hydrogen bonds stabilising the structure, but its stability is probably due to the extremely limited range of ψ that is possible, corresponding to $\phi \sim 100°$, which is necessary for proline (Fig. 22). Complexes of polyproline I with propionic acid have been shown to have the same chain structure as the polymer itself (Traub *et al.*, 1967).

Poly-L-hydroxyproline has been studied by Sasisekharan (1959b), who found its chain structure to be very similar to poly-L-proline II, but stabilised by OH . . . O hydrogen bonds. Another interesting structure similar to poly-L-proline II is polyglycine II whose structure was determined by Crick and Rich (1955). The chains in the structure have each a three-fold screw axis, but in addition, the chains are hydrogen-bonded in a hexagonal array, so that all the NH groups are hydrogen-bonded. In this case, both right-handed and left-handed helices are possible. Polyglycine I, on the other hand, exists as a β-form in the solid state (see Bamford *et al.*, 1956, for details).

A modification of the Rich-Crick structure of polyglycine II, which incorporates systematic CH . . . O hydrogen bonds in addition to NH . . . O hydrogen bonds, has been proposed by Ramachandran *et al.* (1966a). The existence of the CH . . . O bonds has been confirmed by recent infrared studies of Krimm *et al.* (1967), who find a splitting in the CH frequency corresponding to what are believed to be hydrogen-bonded and nonhydrogen-bonded α-carbon atoms, which exist when the chains occur randomly with either sense, going up or down (Ramachandran *et al.*, 1967). However, the helices in a single fibril will all be right-handed, or all left-handed, as shown by the latter workers.

4. Conformations of Synthetic Analogues of Proteins

In recent years, interest has turned to the study of possible polypeptide models of collagen and silk as an aid to the understanding of the conformation of these proteins. To this end, polypeptides with ordered

sequences of residues have been synthesized. We have briefly discussed poly-(Gly-L-Ala) in Section VIII,E,1 above. In order to understand the kind of amino acid triplets which would play a primary role in the formation of the triple helical structure of collagen, Andreeva and her colleagues in Moscow and Traub and co-workers in Israel have examined in great detail a number of polytripeptides as collagen models.

The first set of experiments was done with polymers having the sequence Gly-Pro-Hypro[10] by the Moscow group. Two structural forms were reported for this polytripeptide. A low molecular weight form gave an X-ray pattern which could be explained on the basis of three helical chains held together by sets of hydrogen bonds; but unlike the arrangement found in collagen, the helices run parallel and are not coiled about each other (Andreeva and Millionova, 1963, 1964). The other, a high molecular weight form, resembles collagen in X-ray pattern, infrared spectrum, and optical rotation (Andreeva *et al.*, 1963; Rogulenkova *et al.*, 1964).

Shibnev *et al.* (1965) have obtained an X-ray pattern for poly-Gly-Pro-Pro which was found to be similar to that of collagen. This polymer has been studied in detail by Traub and Yonath (1966) who also concluded that the pattern could be interpreted in terms of a collagenlike structure. Shibnev *et al.* (1966) have extended their studies to polytripeptides having the sequences Gly-Hypro-Hypro and Gly-Hypro-Pro. From these studies, Andreeva *et al.* (1967) conclude that all polytripeptides with sequences Gly-imino acid-imino acid could easily form the collagenlike triple helix.

Sequences with Gly-Pro-Ala and Gly-O-acetyl-Hypro-Pro have also been studied by Traub and Yonath (1966) and shown to have a collagenlike triple helical structures. Gly-Ala-Hypro also shows a similar structure (Andreeva *et al.*, 1967). However, it was found by Traub and Yonath that the polytripeptide having a sequence Gly-Pro-Gly forms a structure with an aggregation of parallel left-handed helices in sheets, in which, however, the individual chains have three residues per turn.

Andreeva *et al.* (1967) have also investigated polytripeptides with the sequence alanine-imino acid-imino acid, e.g., poly-Ala-Pro-Pro and poly-Ala-Hypro-Hypro. These polymers did not give the collagenlike X-ray pattern. These authors therefore concluded that the substitution of glycine by alanine in the first place prevents the formation of the collagen fold. In other words, for the formation of a triple helical protofibril as in collagen, no residue other than glycine can occur in this position in the polypeptide chain. This postulation of glycine as every

[10] Since the nonglycyl peptides discussed in this section are all of the L-configuration, the prefix L will be omitted here.

third residue in the chain of the collagen structure was the fundamental basis of the first proposal of the triple chain by Ramachandran and Kartha (1954, 1955a,b) and these recent studies amply support this assumption.

5. Conformation of Related Polypeptides

We have seen above that hydrogen bonds are not essential for a helical conformation, e.g. poly-L-proline I. It would be of interest to know whether helical conformations are at all possible when the hydrogen of the amide nitrogen is substituted with groups which prevent the formation of a hydrogen bond such as N—H . . . O=C. Very recently, experiments have actually been carried out on poly-*N*-methyl-L-alanine (this has a methyl group attached to the peptide nitrogen) by Goodman and Fried (1967). Nuclear magnetic resonance, optical rotatory dispersion, and circular dichroism studies showed that the above polymer could, in fact, have a helical conformation in solution. Conformational energies have been computed both by Mark and Goodman (1967) and by Liquori and DeSantis (1967) and it is found that four helical conformations are possible for poly-*N*-methyl-L-alanine. These conformations have local minimum energies. One corresponds nearly to the β-conformation, another lies near α_M, and two others are found in regions where no helical conformations have so far been reported in other polypeptides.

F. Effect of Side Groups on α-Helix Formation

In Section VIII,E,4, it was shown that the occurrence of certain types of sequences of side chains lead to the formation of the collagen fold. While in this case the conditions have become fairly clear, namely that every third residue must be glycine and there must be a good proportion of proline or hydroxyproline among the remaining residues, these sequences do not occur in globular proteins, which therefore cannot fold up into a triple-helical arrangement. The only regular helical conformations that are likely to occur in these proteins are one-chain helices with intrahelical hydrogen bonding, such as the α-helix, along with its variant, the 3_{10}-helix. These are internally hydrogen-bonded within the backbone of the chain and the question arises as to what influence the side group has on the formation of the helical conformation. (In the following, the term "helix" shall refer to the α-helix or its close analogues.)

It is well known that when a prolyl residue occurs in a chain, it has an inhibiting effect on the formation of α-helix. In fact, stereochemically, such a residue can occur only at the N-terminal end of a helix

within the first one or two residues. There cannot be a turn of α-helix preceding the proline residue, both because proline has no NH group for hydrogen bonding and also because the five-membered side chain ring makes bad contacts with the atoms in the peptide units three and four steps earlier in the main chain backbone. Thus, a proline residue is nonhelix-forming, in this sense, although the value of $\phi \sim 100°$ which it demands is close to that required for an α-helix. Therefore it may serve to initiate an α-helix, but it cannot occur afterwards in the helix.

Again, Kendrew (1962) observed that, in myoglobin, a break in an α-helix is often found in the vicinity of a seryl or a threonyl residue. In fact, these two residues, with γ-oxygen atoms, must be considered nonhelix-forming, for the side chain can hydrogen bond to the backbone in a variety of ways other than for a right-handed α-helix (Section V,D). This raises the question whether there are other such residues which inhibit helix formation.

The basic finding that initiated an analysis of the influence of side chains on the α-helix was that of Blout *et al.* (1960) and Bloom *et al.* (1962), who found that certain polyamino acids exist as helices in solution and certain others do not. In particular, polymers of valine, isoleucine, serine, threonine, cysteine, and proline do not form helices.[11] To work out theoretically whether these amino acids really inhibit the formation of α-helices, one has to take into account a large number of interactions, as described in Section VI, and this is not always an easy process. It would therefore appear to be profitable to formulate at least certain empirical rules governing the formation of helices and nonhelical stretches in a chain depending upon the sequence of amino acids.

Two methods of approach have been made in this connection. One is to analyse statistically the known amino acid sequence and helix content in proteins whose structures are known. The other approach is to study experimentally the role of sequential polypeptides and protein analogues, by analysing the conformation in solution by ORD techniques and in the solid state by infrared and X-ray studies.

1. Observation on Proteins

The first method, namely the statistical analysis, is based on the argument that, if a three-dimensional structure of a protein is a direct consequence of its amino acid sequence, then there should be a correlation between the structure (in particular its α-helix content) and the sequence. In some cases, instead of the sequence, the amino acid composition is

[11] However, as mentioned in Section VII,E, Scheraga and co-workers found that poly-L-valine can form a helix in solution under suitable circumstances.

made use of. Davies (1964) has examined the percentages of "non-helix-forming" amino acids (viz., valine, isoleucine, serine, threonine, cysteine, and proline) in proteins of known compositon. He has established that the best correlation exists between the total (Ser + Thr + Val + Ile + Cys) content versus percentage helix; the greater the total content of these, the less is the helix present.

Havsteen (1966) has made an exhaustive study similar to the above, by including observations on 40 proteins. A statistical correlation was made by studying the linear regression of the dependence of $1/b_0$ (where b_0 is the Moffitt parameter in the formula representing ORD, which depends on helix content) on the number of amino acids of a given type. According to him, the amino acids may be classified as follows on the basis of the correlation coefficient:

1. Helix-promoting:	Ala, Glu, Leu, Lys, Met, Tyr
2. Indifferent:	Gly
3. Nonhelix-forming:	(a) due to steric reasons: Val, Ile
	(b) due to other reasons: Ser, Thr, Pro, Hypro, etc.

It was found that the helix content was smaller, the larger the content of classes (3)a and 3(b). The correlation coefficients for class 3(a) were in general lower than those for class 3(b), but the trend was not highly significant. Similarly, the fit of (Ser + Thr + Pro) is slightly better than that of either (Ser + Thr + Pro + Cys) or of (Ser + Thr) alone.

Guzzo (1965), analysing the sequence in myoglobin and in the α- and β-chains of haemoglobin, has suggested a possible correlation between certain amino acids in the sequence and the location of the helical and nonhelical parts of a structure. He fails to find any apparent correlation with the presence of Val, Ile, Ser, and Thr. But the presence in the sequence of Pro, Asp, Glu, or His appears to lead to the breaking of the helix. He concludes, at least in the proteins he examined, that these amino acids might be *necessary* (and Pro sufficient) for the section to be nonhelical. He advances the experimental evidence that the substitution of Pro (which definitely breaks a helix) for Glu or Asp in a mutant still leaves the mutant active, indicating that Pro does not produce any new nonhelical region, since the structure remains presumably the same. However, the suggestion that Glu is helix-breaking is contrary to the results of Havsteen and the conclusion of Prothero and others discussed below.

Prothero (1966) has shown that the above criteria of Guzzo are not

valid in a number of proteins and has instead suggested that (a) any region of five residues will be α-helical if at least three include Ala, Val, Leu or Glu and (b) any segment of seven residues will be helical if at least three are Ala, Val, Leu or Glu and if one more is Ile, Thr, or Gln. [Of course, the absence of Pro is assumed; the inclusion of Val in the criterion (a) and of Ile and Thr in the criterion (b) is not in concordance with Havsteen's results.] Periti *et al.* (1967) have also employed statistical techniques by analysing pairs of amino acid residues in the myoglobin sequence and in α- and β-chains of haemoglobin which are adjacent and separated by one, two, . . . up to five residues and tabulating the results. These tables will be of value in deducing laws governing the formation of helical sequences.

Recently, Schiffer and Edmundson (1967) have employed a graphical method to study the amino acid sequences in the helical and nonhelical regions of myoglobin, haemoglobin, and lysozyme. The amino acid side chains in a helix are projected on a plane perpendicular to the helix axis. These are named "helical wheels," the perimeter of a wheel corresponding to the backbone and the amino acids being arranged sequentially on external spokes. For an α-helix, each spoke will be set at 100° from its predecessor. From an analysis using these helical wheels, they found that in the helical regions of myoglobin, the hydrophobic residues (Leu, Ile, Val, Met, Phe, Tyr, Trp, and Ala) form clusters, termed stabilizing arcs, and are believed to stabilise the structure by inter-helical interactions. Moreover, the hydrophobic residues tend to occupy positions i, $i+3$, $i+4$ on adjacent helical turns. These are believed to be the most favourable positions (Némethy and Scheraga, 1962a) resulting in stabilizing *intra*-helical interactions. Based on the criteria thus obtained, namely (a) clustering of hydrophobic residues on the wheel; and (b) arrangement of these residues in positions i, $i+3$, $i+4$, they have worked out regions of insulin, cytochrome c, ribonuclease A, and other proteins which are potentially helical. They have obtained from these studies approximate agreement between the percentage of helix observed in these proteins and the value expected from the method of wheels.

If we accept the argument of Schiffer and Edmundson, it is clear that what is relevant is not so much the amino acid composition, but the sequence. In this sense, their conclusions agree with those of Prothero, but go further. It should be mentioned that Havsteen's criterion 3(a), which assumes that Val and Ile are nonhelix-forming, must be carefully reexamined in the light of these conclusions. Theoretically, (as mentioned in Sections V,D and VII,E), these two amino acids, branched at the β-carbon atom, can form α-helices with the two γ-carbon atoms

going to positions II and III. This not only decreases the entropy because of the choice being restricted to only one of the three possible combinations of positions for the atoms $C^{\gamma 1}$ and $C^{\gamma 2}$, but also because position I (which is forbidden for an α-helix) is found to occur in the free amino acid in five examples (see Table XXII) and also in a recent sixth determination of DL-valine in the authors' laboratory (Mallikarjunan and Rao, 1968). Hence, in the case of Val and Ile, there are competing effects, namely the possible hydrophobic interactions when they are favourably disposed for helix formation and the need for a special (possibly unfavourable) side-chain conformation, which is necessary in the helix (see also the helix-inhibiting property of Val residues in polypeptides, discussed in the next section).

2. Observations on Synthetic Polypeptides

The influence of residue sequence on helix formation has been studied experimentally by Fraser and co-workers recently, by synthesising a series of polypeptides containing ordered sequences of helix-forming and supposedly helix-breaking residues and studying their helicity by ORD (b_o-values) in solution and by infrared and X-ray diffraction studies of films formed from these polymers. Fraser *et al.* (1965a) have studied sequences of L-valyl (V) and γ-methyl-L-glutamyl (G) residues.[12] ORD measurements of $(G_xVG)_n$, $x = 0, 1, 2, 3$ indicated that all these polymers, to some extent adopt a helical conformation. As the mole fraction of V is increased, the helicity decreased. Infrared spectra revealed the presence of α-helix in addition to a predominantly β-conformation in all cases, except for $(VG)_n$ and $(V)_n$ which were only in β-conformation. X-ray diffraction patterns yielded β-patterns for all polymers, except $(G_3)_n$ and $(G_xVG)_n$ where α-patterns superposed on β were obtained.

Similar studies on sequential polypeptides containing *S*-benzyl-L-cysteinyl (C) and γ-ethyl-L-glutamate (G) residues were made by Fraser *et al.* (1965b). Here again ORD studies indicated that all polymers studied adopt a certain amount of helical conformation in solution. Infrared spectra of films indicated the same results. However, $(G_3CG)_n$, $(GCG_2)_n$ gave an α-type of X-ray pattern and others only a β-pattern. These authors have attributed the decrease in the stability of the helix due to increasing percentage of valyl or cysteinyl residues to steric factors.

Recently, Fraser *et al.* (1967) have extended these studies to find out

[12] These and similar abbreviated symbols are used for convenience in this section and they do not correspond to the one-letter symbols for amino acid residues that have been suggested in the literature.

the effect of glycyl residues on the stability of the α-helix. It is well known that polyglycine itself does not exist in the α-conformation and that copolymers with glycine tend to have a β-conformation (apart of course from the special ones which form triple helices). Therefore, copolymers of γ-ethyl-L-glutamate (G) and glycine (g) were synthesised and their conformation studied in solution by ORD and in the solid state by infrared and X-ray diffraction methods. Results indicated that $(G_2)_n$ is highly α-helical, both in solution and solid, whereas $(Gg)_n$ is in the β-form; helicity increases with decreasing proportion of glycine in the order $(G_2g)_n$, $(G_3g)_n$ $(G_4g)_n$ and $(G_2gG_3)_n$. In the crystalline form, even the last one shows no α-helix in its X-ray pattern, though infrared data for the solid and ORD in solution suggest some amount of helix. Therefore, it can be inferred that the presence of glycyl residues, particularly in a regular sequence, results in a marked reduction in the stability of the α-helix. In glycine, the absence of a side chain β-carbon atom results in a large freedom of conformation—as could be seen from the large area available in the (ϕ, ψ) plane for a glycyl residue—and consequently a transition from the highly specialised helical conformation to a random one entails a large gain in entropy and this may perhaps explain the contribution to the instability of the α-helix by glycine residues. In fact, the lysozyme (ϕ, ψ) plot shows practically no glycyl conformations in the helical region, while they are distributed all over the map elsewhere (Figs. 16 and 38).

Acknowledgments

The authors would like to thank their colleagues in the Centre of Advanced Study in Biophysics at Madras, and in particular Mr. A. V. Lakshminarayanan and Dr. C. M. Venkatachalam, for their assistance in preparing the manuscript and for their comments and criticism. They would like to acknowledge the assistance of various persons who provided preprints of unpublished work, in particular Dr. J. C. Kendrew for the dihedral angle data on myoglobin. The secretarial work was throughout done by Mr. K. P. Ramachandran and all the figures were prepared by Mr. V. R. Sambandam, to both of whom the authors are most grateful. Thanks are also due to U.S. Public Health Service (Grant Nos. AM-10905 and AM-GM-11493) for partial financial assistance.

References

Ambrose, E. J., and Elliott, A. (1951). *Proc. Roy. Soc.* (*London*) **A205**, 47.

Andreeva, N. S., and Millionova, M. I. (1963). *Kristallografiya* **8**, 578.

Andreeva, N. S., and Millionova, M. I. (1964). *Soviet Phys. Cryst.* (*English Transl.*) **8**, 464.

Andreeva, N. S., Millionova, M. I., and Chirgadze, Yu. N. (1963). *In* "Aspects of Protein Structure" (G. N. Ramachandran, ed.), p. 137. Academic Press, New York.

Andreeva, N. S., Esipova, N. G., Millionova, M. I., Rogulenkova, V. N., and Shibnev, V. A. (1967). *In* "Conformation of Biopolymers" (G. N. Ramachandran, ed.), Vol. 1, p. 291. Academic Press, New York.

Arnott, S., and Wonacott, A. J. (1966). *J. Mol. Biol.* **21**, 371.

Astbury, W. T. (1940). *J. Soc. Leather Trades' Chemists* **24**, 69.

Astbury, W. T., Beighton, E., and Parker, K. D. (1959). *Biochim. Biophys. Acta* **16**, 1.

Atkins, E. D. T. (1967). *J. Mol. Biol.* **24**, 139.

Balasubramanian, D., and Wetlaufer, D. B. (1967). *In* "Conformation of Biopolymers" (G. N. Ramachandran, ed.), Vol. 1, p. 147. Academic Press, New York.

Bamford, C. H., Hanby, W. E., and Happey, F. (1951). *Proc. Roy. Soc.* (*London*) **A205**, 30.

Bamford, C. H., Brown, L., Elliott, A., Hanby, W. E., and Trotter, I. F. (1952). *Nature* **169**, 357.

Bamford, C. H., Brown, L., Elliott, A., Hanby, W. E., and Trotter, I. F. (1954). *Nature* **173**, 27.

Bamford, C. H., Elliott, A., and Hanby, W. E. (1956). "Synthetic Polypeptides." Academic Press, New York.

Birkhoff, G., and MacLane, S. (1963). "A Survey of Modern Algebra," p. 254. Macmillan, New York.

Blake, C. C. F., Mair, G. A., North, A. C. T., Phillips, D. C., and Sarma, V. R. (1967a). *Proc. Roy. Soc.* (*London*) **B167**, 365.

Blake, C. C. F., Johnson, L. N., Mair, G. A., North, A. C. T., Phillips, D. C., and Sarma, V. R. (1967b). *Proc. Roy. Soc.* (*London*) **B167**, 378.

Bloom, S. M., Fasman, G. D., deLozé, C., and Blout, E. R. (1962). *J. Am. Chem. Soc.* **84**, 458.

Blout, E. R. (1961). *Tetrahedron* **13**, 123.

Blout, E. R., deLozé, C., Bloom, S. M., and Fasman, G. C. (1960). *J. Am. Chem. Soc.* **82**, 3787.

Blout, E. R., and Karlson, R. H. (1958). *J. Am. Chem. Soc.* **80**, 1259.

Bondi, A. (1964). *J. Phys. Chem.* **68**, 44.

Born, M., and Mayer, J. E. (1932). *Z. Physik.* **75**, 1.

Bradbury, E. M., Brown, L., Downie, A. R., Elliott, A., Hanby, W. E., and McDonald, T. R. R. (1959). *Nature* **183**, 1736.

Bradbury, E. M., Downie, A. R., Elliott, A., and Hanby, W. E. (1960). *Proc. Roy. Soc.* (*London*) **A259**, 110.

Bradbury, E. M., Brown, L., Downie, A. R., Elliott, A., Fraser, R. D. B., and Hanby, W. E. (1962). *J. Mol. Biol.* **5**, 230.

Bragg, W. L., Kendrew, J. C., and Perutz, M. F. (1950). *Proc. Roy. Soc.* (*London*) **A203**, 321.

Brant, D. A., and Flory, P. J. (1965a). *J. Am. Chem. Soc.* **87**, 663.

Brant, D. A., and Flory, P. J. (1965b). *J. Am. Chem. Soc.* **87**, 2791.

Brant, D. A., Miller, W. G., and Flory, P. J. (1967). *J. Mol. Biol.* **23**, 47.

Brown, L., and Trotter, I. F. (1956). *Trans. Faraday Soc.* **52**, 537.

Cahn, R. S., Ingold, C., and Prelog, V. (1966). *Angew. Chem. Intern. Ed. English* **5**, 385.

Chandrasekaran, R. (1968). *Proc. Indian Acad. Sci.*, **A, 68**, 13.

Cohen, C., and Holmes, K. C. (1963). *J. Mol. Biol.* **6**, 423.

Corey, R. B., and Pauling, L. (1953). *Proc. Roy. Soc.* (*London*) **B141**, 10.

Cowan, P. M., and McGavin, S. (1955). *Nature* **176**, 501.
Cowan, P. M., McGavin, S., and North, A. C. T. (1955). *Nature* **176**, 1062.
Crick, F. H. C. (1952). *Nature* **170**, 882.
Crick, F. H. C. (1953). *Acta Cryst.* **6**, 689.
Crick, F. H. C., and Kendrew, J. C. (1957). *Advan. Protein Chem.* **12**, 134.
Crick, F. H. C., and Rich, A. (1955). *Nature* **176**, 780.
Dale, J. (1966). *Tetrahedron* **22**, 3373.
Davidson, B., Tooney, N., and Fasman, G. D. (1966). *Biochem. Biophys. Res. Commun.* **23**, 155.
Davies, D. R. (1964). *J. Mol. Biol.* **9**, 605.
Davies, D. R. (1965). *Progr. Biophys. Mol. Biol.* **15**, 189.
Davies, D. R. (1967). *Ann. Rev. Biochem.* **36**, 321.
Degeilh, R., and Marsh, R. E. (1959). *Acta Cryst.* **12**, 1007.
Del Re, G., Pullman, B., and Yonezawa, T. (1963). *Biochim. Biophys. Acta* **75**, 153.
DeSantis, P., Giglio, E., Liquori, A. M., and Ripamonti, A. (1965). *Nature* **206**, 456.
Dickerson, R. E. (1964). *In* "The Proteins" (H. Neurath, ed.), Vol. II, p. 603. Academic Press, New York.
Donohue, J. (1953). *Proc. Natl. Acad. Sci. U.S.* **39**, 470.
Downie, A. R., and Randall, A. A. (1962). *In* "Collagen" (N. Ramanathan, ed.), p. 209. Wiley (Interscience), New York.
Drenth, J., Jansonius, J. N., and Wolthers, B. G. (1967). *J. Mol. Biol.* **24**, 449.
Dunnill, P. (1965). *Sci. Progr.* **53**, 609.
Edsall, J. T., Flory, P. J., Kendrew, J. C., Liquori, A. M., Némethy, G., Ramachandran, G. N., and Scheraga, H. A. (1966a). *J. Mol. Biol.* **15**, 399.
Edsall, J. T., Flory, P. J., Kendrew, J. C., Liquori, A. M., Némethy, G., Ramachandran, G. N., and Scheraga, H. A. (1966b). *J. Biol. Chem.* **241**, 1004.
Edsall, J. T., Flory, P. J., Kendrew, J. C., Liquori, A. M., Némethy, G., Ramachandran, G. N., and Scheraga, H. A. (1966c). *Biopolymers* **4**, 121.
Elliott, A., and Malcolm, B. R. (1959). *Proc. Roy. Soc.* (*London*) **A249**, 30.
Elliott, A., Fraser, R. D. B., and MacRae, T. P. (1965). *J. Mol. Biol.* **11**, 821.
Eyring, H. (1932a). *Phys. Rev.* **39**, 746.
Eyring, H. (1932b). *J. Am. Chem. Soc.* **54**, 3191.
Flory, P. J. (1967). *In* "Conformation of Biopolymers" (G. N. Ramachandran, ed.), Vol. 1, p. 339. Academic Press, New York.
Frank, H. S., and Evans, M. W. (1945). *J. Chem. Phys.* **13**, 507.
Fraser, R. D. B., and MacRae, T. P. (1961a). *Nature* **189**, 572.
Fraser, R. D. B., and MacRae, T. P. (1961b). *J. Mol. Biol.* **3**, 640.
Fraser, R. D. B., and MacRae, T. P. (1962a). *Nature* **195**, 1167.
Fraser, R. D. B., and MacRae, T. P. (1962b). *J. Mol. Biol.* **5**, 457.
Fraser, R. D. B., and MacRae, T. P. (1963). *J. Mol. Biol.* **7**, 272.
Fraser, R. D. B., MacRae, T. P., and Rogers, G. E. (1962a). *Nature* **193**, 1052.
Fraser, R. D. B., MacRae, T. P., and Stapleton, I. W. (1962b). *Nature* **193**, 573.
Fraser, R. D. B., Harrap, B. S., MacRae, T. P., Stewart, F. H. C., and Suzuki, E. (1965a). *J. Mol. Biol.* **12**, 482.
Fraser, R. D. B., Harrap, B. S., MacRae, T. P., Stewart, F. H. C., and Suzuki, E. (1965b). *J. Mol. Biol.* **14**, 423.
Fraser, R. D. B., MacRae, T. P., Stewart, F. H. C., and Suzuki, E. (1965c). *J. Mol. Biol.* **11**, 706.

Fraser, R. D. B., MacRae, T. P., and Miller, A. (1965d). *J. Mol. Biol.* **14**, 432.
Fraser, R. D. B., Harrap, B. S., MacRae, T. P., Stewart, F. H. C., and Suzuki, E. (1967). *Biopolymers* **5**, 251.
Fukushima, K., Ideguchi, Y., and Miyazawa, T. (1963). *Bull. Chem. Soc. Japan* **36**, 1301.
Gel'fand, I. M., Pyatetsky-Shapiro, I. F., and Federov, Yu. G. (1963). *Soviet Math.* **4**, 1487.
Gibson, K. D., and Scheraga, H. A. (1966). *Biopolymers* **4**, 709.
Gibson, K. D., and Scheraga, H. A. (1967). *Proc. Natl. Acad. Sci. U.S.* **58**, 420.
Goodman, M., and Fried, M. (1967). *J. Am. Chem. Soc.* **89**, 1264.
Gutowsky, H. S., and Holm, C. H. (1956). *J. Chem. Phys.* **25**, 1228.
Guzzo, A. V. (1965). *Biophys. J.* **5**, 809.
Hahn, T. (1957). *Z. Krist.* **109**, 438.
Hamilton, W., and Ibers, J. (1968). "Hydrogen Bonding in Solids." W. A. Benjamin, New York.
Hammaker, R. M., and Gugler, B. A. (1965). *J. Mol. Spectr.* **17**, 356.
Hardy, P. M., Kenner, G. W., and Sheppard, R. C. (1963). *Tetrahedron* **19**, 95.
Harrington, W. F., Josephs, R., and Segal, D. M. (1966). *Ann. Rev. Biochem.* **35**, 599.
Havsteen, B. H. (1966). *J. Theoret. Biol.* **10**, 1.
Herzberg, G. (1945). "Molecular Spectra and Molecular Structure. II. Infrared and Raman Spectra of Polyatomic Molecules." Van Nostrand, New York.
Huggins, M. L. (1943). *Chem. Rev.* **32**, 195.
Jeffreys, H., and Jeffreys, B. S. (1950). "Methods of Mathematical Physics," p. 122. Cambridge Univ. Press, London and New York.
Karle, I. L., and Karle, J. (1963). *Acta Cryst.* **16**, 969.
Karlson, R. H., Norland, K. S., Fasman, G. D., and Blout, E. R. (1960). *J. Am. Chem. Soc.* **82**, 2268.
Kartha, G., Bello, J., and Harker, D. (1967). *Nature* **213**, 863.
Kauzmann, W. (1959). *Advan. Protein Chem.* **14**, 1.
Kayushina, R. L., and Vainshtein, B. K. (1966). *Soviet Phys.-Cryst. (English Transl.)* **10**, 698.
Kemp, J. D., and Pitzer, K. S. (1936). *J. Chem. Phys.* **4**, 749.
Kendrew, J. C. (1962). *Brookhaven Symp. Biol.* **15**, 216.
Kendrew, J. C. (1963). *Science* **139**, 1259.
Kendrew, J. C., and Perutz, M. F. (1957). *Ann. Rev. Biochem.* **26**, 327.
Kendrew, J. C., Dickerson, R. E., Strandberg, B. E., Hart, R. G., and Davies, D. R. (1960). *Nature* **185**, 422.
Kendrew, J. C., Watson, H. C., Strandberg, B. E., Dickerson, R. E., Phillips, D. C., and Shore, V. C. (1961). *Nature* **190**, 666.
Kenner, G. W., Thompson, P. J., and Turner, J. M. (1958). *J. Chem. Soc.* p. 4148.
Ketelaar, J. (1958). "Chemical Constitution," p. 91. Elsevier, Amsterdam.
Kitaigorodsky, A. I. (1961). *Tetrahedron* **14**, 230.
Kitaigorodsky, A. I. (1965). *Acta Cryst.* **18**, 585.
Klyne, W., and Prelog, V. (1960). *Experientia* **16**, 521.
Kraut, J. (1965). *Ann. Rev. Biochem.* **34**, 247.
Krimm, S., Kuroiwa, K., and Rebane, T. (1967). *In* "Conformation of Biopolymers" (G. N. Ramachandran, ed.), Vol. 2, p. 439. Academic Press, New York.

Lakshminarayanan, A. V. (1968). *In* "Conformation of Biological Macromolecules. Vol. I: Conformation of Amino Acids" (G. N. Ramachandran, ed.). In preparation.

Lakshminarayanan, A. V., Sasisekharan, V., and Ramachandran, G. N. (1967). *In* "Conformation of Biopolymers" (G. N. Ramachandran, ed.), Vol. 1, p. 61. Academic Press, New York.

LaPlanche, L. A., and Rogers, M. T. (1964). *J. Am. Chem. Soc.* **86**, 337.

Leach, S. J., Némethy, G., and Scheraga, H. A. (1966a). *Biopolymers* **4**, 369.

Leach, S. J., Némethy, G., and Scheraga, H. A. (1966b). *Biopolymers* **4**, 887.

Leung, Y. C., and Marsh, R. E. (1958). *Acta Cryst.* **11**, 17.

Levinthal, C. (1966). *Sci. Am.* **214**, 42.

Lindley, H., and Rollett, J. S. (1955). *Biochim. Biophys. Acta* **18**, 183.

Lippincott, E. R., and Schroeder, R. (1955). *J. Chem. Phys.* **23**, 1099.

Lipscomb, W. N., Coppola, J. C., Hartsuck, J. A., Ludwig, M. L., Muirhead, H., Searl, J., and Steitz, T. A. (1966). *J. Mol. Biol.* **19**, 423.

Lipscomb, W. N., Hartsuck, J. A., Ludwig, M. L., Quiocho, F. A., Geeke, G. N., and Steitz, T. A. (1967). *Am. Cryst. Assoc. Summer Meeting, August.* Abstr. G4.

Liquori, A. M. (1966). *Ciba Found. Symp. Principles Biomol. Organ.* p. 40.

Liquori, A. M., and DeSantis, P. (1967). Private communication.

Liquori, A. M., DeSantis, P., Kovacs, A. L., and Mazzarella, L. (1966). *Nature* **211**, 1039.

London, F. (1937). *Trans. Faraday Soc.* **8**, 33.

Low, B. W., and Baybutt, R. B. (1952). *J. Am. Chem. Soc.* **74**, 5806.

Low, B. W., and Grenville-Wells, H. J. (1953). *Proc. Natl. Acad. Sci. U.S.* **39**, 785.

Lucas, F., and Rudall, K. M. (1967). *In* "Comprehensive Biochemistry" (M. Florkin and E. H. Stotz, eds.), Vol. 26B, p. 475. Elsevier, Amsterdam.

Mallikarjunan, M., and Rao, S. T. (1968). *Acta Cryst.* (in press).

Mark, J. E., and Goodman, M. (1967). *J. Am. Chem. Soc.* **89**, 1267.

Marsh, R. E., and Donohue, J. (1967). *Advan. Protein Chem.* **22**, 234.

Marsh, R. E., Corey, R. B., and Pauling, L. (1955a). *Biochim. Biophys. Acta* **16**, 1.

Marsh, R. E., Corey, R. B., and Pauling, L. (1955b). *Acta Cryst.* **8**, 710.

Mathews, B. W., Sigler, P. B., Henderson, R., and Blow, D. M. (1967). *Nature* **214**, 652.

Millen, D. J. (1962). *Progr. Stereochem.* **3**, 138.

Mitsui, Y., Iitaka, Y., and Tsuboi, M. (1967). *J. Mol. Biol.* **24**, 15.

Miyazawa, T., Shimanouchi, T., and Mizushima, S. (1958). *J. Chem. Phys.* **29**, 611.

Mizushima, S., and Shimanouchi, T. (1961). *Advan. Enzymol.* **23**, 1.

Moulton, W. G., and Kromhout, R. A. (1956). *J. Chem. Phys.* **25**, 34.

Muirhead, H., Cox, J. M., Mazzarella, L., and Perutz, M. F. (1967). *J. Mol. Biol.* **28**, 117.

Némethy, G. (1967). *Angew. Chem. Intern. Ed. Engl.* **6**, 195.

Némethy, G., and Scheraga, H. A. (1962a). *J. Phys. Chem.* **66**, 1773.

Némethy, G., and Scheraga, H. A. (1962b). *J. Chem. Phys.* **36**, 3382.

Némethy, G., and Scheraga, H. A. (1965). *Biopolymers* **3**, 155.

Némethy, G., Leach, S. J., and Scheraga, H. A. (1966). *J. Phys. Chem.* **70**, 998.

Némethy, G., Phillips, D. C., Leach, S. J., and Scheraga, H. A. (1967). *Nature* **214**, 363.
Ooi, T., Scott, R. A., Vanderkooi, G., and Scheraga, H. A. (1967). *J. Chem. Phys.* **46**, 4410.
Parker, K. D., and Rudall, K. M. (1957). *Nature* **179**, 905.
Pauling, L. (1960). "The Nature of the Chemical Bond," 3rd Ed. Cornell Univ. Press, Ithaca, New York.
Pauling, L., and Corey, R. B. (1951). *Proc. Natl. Acad. Sci. U.S.* **37**, 235.
Pauling, L., and Corey, R. B. (1953a). *Nature* **171**, 59.
Pauling, L., and Corey, R. B. (1953b). *Proc. Natl. Acad. Sci. U.S.* **39**, 253.
Pauling, L., and Sherman, J. (1933). *J. Chem. Phys.* **1**, 606.
Pauling, L., Corey, R. B., and Branson, H. R. (1951). *Proc. Natl. Acad. Sci. U.S.* **37**, 205.
Periti, P. F., Quagliarotti, G., and Liquori, A. M. (1967). *J. Mol. Biol.* **24**, 313.
Perutz, M. F. (1951). *Nature* **167**, 1053.
Pimentel, G. C., and McClellan, A. L. (1960). "The Hydrogen Bond," p. 286. Freeman, San Francisco, California.
Pitzer, K. S. (1959). *Advan. Chem. Phys.* **2**, 59.
Prothero, J. W. (1966). *Biophys. J.* **6**, 367.
Ramachandran, G. N. (1960). *Proc. Indian Acad. Sci.* **A52**, 240.
Ramachandran, G. N. (1962). *In* "Collagen" (N. Ramanathan, ed.), p. 3. Wiley (Interscience), New York.
Ramachandran, G. N. (1967). *In* "Treatise on Collagen" (G. N. Ramachandran, ed.), Vol. 1, p. 103. Academic Press, New York.
Ramachandran, G. N., and Dweltz, N. E. (1962). *In* "Collagen" (N. Ramanathan, ed.), p. 147. Wiley (Interscience), New York.
Ramachandran, G. N., and Kartha, G. (1954). *Nature* **174**, 269.
Ramachandran, G. N., and Kartha, G. (1955a). *Nature* **176**, 593.
Ramachandran, G. N., and Kartha, G. (1955b). *Proc. Indian Acad. Sci.* **A45**, 124.
Ramachandran, G. N., and Sasisekharan, V. (1965). *Biochim. Biophys. Acta* **109**, 314.
Ramachandran, G. N., Sasisekharan, V., and Thathachari, Y. T. (1962). *In* "Collagen" (N. Ramanathan, ed.), p. 81. Wiley (Interscience), New York.
Ramachandran, G. N., Ramakrishnan, C., and Sasisekharan, V. (1963a). *J. Mol. Biol.* **7**, 95.
Ramachandran, G. N., Ramakrishnan, C., and Sasisekharan, V. (1963b). *In* "Aspects of Protein Structure" (G. N. Ramachandran, ed.), p. 121. Academic Press, New York.
Ramachandran, G. N., Ramakrishnan, C., and Venkatachalam, C. M. (1965). *Biopolymers* **3**, 591.
Ramachandran, G. N., Sasisekharan, V., and Ramakrishnan, C. (1966a). *Biochim. Biophys. Acta* **112**, 168.
Ramachandran, G. N., and Srinivasan, R. (1968). In preparation.
Ramachandran, G. N., and Venkatachalam, C. M. (1968). *Biopolymers* (in press).
Ramachandran, G. N., and Venkatachalam, C. M. (1969). *In* "Conformation of Biological Macromolecules. Vol. II: Conformation of Peptides and Polypeptides" (G. N. Ramachandran, ed.). In preparation.
Ramachandran, G. N., Venkatachalam, C. M., and Krimm, S. (1966b). *Biophys. J.* **6**, 849.
Ramachandran, G. N., Ramakrishnan, C., and Venkatachalam, C. M. (1967). *In*

"Conformation of Biopolymers" (G. N. Ramachandran, ed.), Vol. 2, p. 429. Academic Press, New York.
Ramakrishnan, C. (1964). *Proc. Indian Acad. Sci.* **A59**, 327.
Ramakrishnan, C. (1965). Ph.D. Thesis, Univ. of Madras.
Ramakrishnan, C., and Ramachandran, G. N. (1965). *Biophys. J.* **5**, 909.
Ramakrishnan, C., and Ramachandran, G. N. (1967). *Proc. Symp. Fibrous Proteins, August.* Australian Acad. Sci., Canberra.
Ramakrishnan, C., and Sarathy, K. P. (1968). *Biochim. Biophys. Acta* (in press).
Rao, V. S. R., Sundararajan, P. R., Ramakrishnan, C., and Ramachandran, G. N. (1967). *In* "Conformation of Biopolymers" (G. N. Ramachandran, ed.), Vol. 2, p. 721. Academic Press, New York.
Rich, A., and Crick, F. H. C. (1955). *Nature* **176**, 915.
Rich, A., and Crick, F. H. C. (1961). *J. Mol. Biol.* **3**, 483.
Rich, A., and Green, D. W. (1961). *Ann. Rev. Biochem.* **30**, 193.
Rogers, M. T., and Woodbrey, J. C. (1962). *J. Phys. Chem.* **66**, 540.
Rogulenkova, V. N., Millionova, M. I., and Andreeva, N. S. (1964). *J. Mol. Biol.* **9**, 253.
Rosenheck, K., and Doty, P. (1961). *Proc. Natl. Acad. Sci. U.S.* **47**, 1775.
Rothe, M., Steffen, K. D., and Rothe, I. (1965). *Angew. Chem. Intern. Ed. English* **4**, 356.
Rudall, K. M. (1962). *In* "Comparative Biochemistry" (M. Florkin and H. S. Mason, eds.), Vol. IV, p. 397. Academic Press, New York.
Rudall, K. M. (1963). *Advan. Insect Physiol.* **1**, 257.
Rudall, K. M. (1967). *In* "Treatise on Collagen" (G. N. Ramachandran and B. S. Gould, eds.), Vol. 2, p. 83. Academic Press, New York.
Sarathy, K. P., and Ramachandran, G. N. (1968). *Biopolymers* **6**, 461.
Sarkar, P. K., and Doty, P. (1966). *Proc. Natl. Acad. Sci. U.S.* **55**, 981.
Sasisekharan, V. (1959a). *Acta Cryst.* **12**, 897.
Sasisekharan, V. (1959b). *Acta Cryst.* **12**, 903.
Sasisekharan, V. (1962). *In* "Collagen" (N. Ramanathan, ed.), p. 39. Wiley (Interscience), New York.
Schellman, J. A., and Schellman, C. (1964). *In* "The Proteins" (H. Neurath, ed.), Vol. 2, p. 1. Academic Press, New York.
Scheraga, H. A. (1960). *J. Am. Chem. Soc.* **82**, 3847.
Scheraga, H. A. (1968). *Advan. Phys. Org. Chem.* **6**, (in press).
Scheraga, H. A., Scott, R. A., Vanderkooi, G., Leach, S. J., Gibson, K. D., Ooi, T., and Némethy, G. (1967). *In* "Conformation of Biopolymers" (G. N. Ramachandran, ed.), Vol. 1, p. 43. Academic Press, New York.
Schiffer, M., and Edmundson, A. B. (1967). *Biophys. J.* **7**, 121.
Schor, R., and Krimm, S. (1961a). *Biophys. J.* **1**, 467.
Schor, R., and Krimm, S. (1961b). *Biophys. J.* **1**, 489.
Schröder, E., and Lübke, K. (1966). "The Peptides," Vol. 2, Chapt. 2. Academic Press, New York.
Schroeder, R., and Lippincott, E. R. (1957). *J. Phys. Chem.* **61**, 921.
Scott, R. A., and Scheraga, H. A. (1965). *J. Chem. Phys.* **42**, 2209.
Scott, R. A., and Scheraga, H. A. (1966a). *J. Chem. Phys.* **44**, 3054.
Scott, R. A., and Scheraga, H. A. (1966b). *Biopolymers* **4**, 237.
Scott, R. A., and Scheraga, H. A. (1966c). *J. Chem. Phys.* **45**, 2091.
Shibnev, V. A., Rogulenkova, V. N., and Andreeva, N. S. (1965). *Biofizika* **10**, 164.

Shibnev, V. A., Lisovenko, A. V., Esipova, N. G., Rogulenkova, V. N., and Millionova, M. I. (1966). *Biofizika* **11**, 123.

Slater, J. C., and Kirkwood, J. G. (1931). *Phys. Rev.* **37**, 682.

Smyth, C. P. (1955). "Dielectric Behaviour and Structure," pp. 244–245. McGraw-Hill, New York.

Steinberg, I. Z., Harrington, W. F., Berger, A., Sela, M., and Katchalski, E. (1960). *J. Am. Chem. Soc.* **82**, 5263.

Stockmayer, W. H. (1941). *J. Chem. Phys.* **9**, 398.

Sugeta, H., and Miyazawa, T. (1967). *Biopolymers* **5**, 673.

Traub, W., and Shmueli, U. (1963). *In* "Aspects of Protein Structure" (G. N. Ramachandran, ed.), p. 81. Academic Press, New York.

Traub, W., and Yonath, A. (1966). *J. Mol. Biol.* **16**, 404.

Traub, W., Shmueli, U., Suwalski, M., and Yonath, A. (1967). *In* "Conformation of Biopolymers" (G. N. Ramachandran, ed.), Vol. 2, p. 449. Academic Press, New York.

Tsuboi, M. (1964). *Biopolymers Symp.* **1**, 527.

Tsuboi, M., Wada, A., and Nagashima, N. (1961). *J. Mol. Biol.* **3**, 707.

Vanderkooi, G., Leach, S. J., Némethy, G., Scott, R. A., and Scheraga, H. A. (1966). *Biochemistry* **5**, 2991.

Venkatachalam, C. M. (1968a). *Biopolymers* (in press).

Venkatachalam, C. M. (1968b). *Biochim. Biophys. Acta* (in press).

Venkatachalam, C. M. (1968c). *Biochim. Biophys. Acta* (in press).

Venkatachalam, C. M., and Ramachandran, G. N. (1969). *In* "Conformation of Biopolymers" (G. N. Ramachandran, ed.), Vol. 1, p. 83. Academic Press, New York.

Vollmer, J. P., and Spach, G. (1967). *Biopolymers* **5**, 337.

Westheimer, F. H. (1956). *In* "Steric Effects in Organic Chemistry" (M. S. Newman, ed.), Chapt. 12, p. 523. Wiley (Interscience), New York.

Wilson, E. B. (1959). *Advan. Chem. Phys.* **2**, 367.

Wyckoff, H. W., Hardman, K. D., Allewell, N. M., Inagami, T., Johnson, L. N., and Richards, F. M. (1967). *J. Biol. Chem.* **242**, 3984.

Yang, J. T. (1967). *In* "Conformation of Biopolymers" (G. N. Ramachandran, ed.), Vol. 1, p. 157. Academic Press, New York.

Yang, J. T., and Doty, P. (1957). *J. Am. Chem. Soc.* **79**, 761.

Zalkin, A., Forrester, J. D., and Templeton, D. H. (1966). *J. Am. Chem. Soc.* **88**, 1810.