Ann. Rev. Biochem. 1984. 53: 537–72

PRINCIPLES THAT DETERMINE THE STRUCTURE OF PROTEINS

Cyrus Chothia

Christopher Ingold Laboratory, University College London, 20 Gordon Street, London WC1H 0AJ, England; and Medical Research Council, Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, England

CONTENTS

0066-4154/84/0701-0537$02.00

PERSPECTIVES AND SUMMARY

The principles governing the structure of globular proteins were not revealed, as it was hoped they would be, by the first atomic descriptions of protein molecules. The early protein structures did confirm certain central ideas already present in the 1950s (1), but significant progress beyond that has occurred only in recent years. This review outlines these recent advances. Certain aspects of this work have been reviewed previously (2–6).

Here I deal with the static structure of soluble proteins built from α helices and β-pleated sheets. The intrinsic properties of polypeptides mean that many of the principles reviewed also apply to structural and membrane proteins, and to the very small number of proteins that do not contain α helices or β sheets.

The principles described in this review are concerned with six different levels or aspects of protein structure:

1. The local folding of the polypeptide chain—it was established in the 1950s and 1960s that the principle underlying the structure of helices, sheets, and turns is the simultaneous formation of hydrogen bonds by buried peptide groups and the retention of conformations close to those of minimum energy. Recent work has given precise information on the preferred conformations of secondary structures and side chains. It has also shown how variations within allowed conformational regions permit β sheets to twist, coil, bend, and bulge, and α helices to kink.
2. The association of secondary structures—α helices and β sheets in proteins close pack. The shape of their surfaces is determined, to a first approximation, by the conformation of the main chain. As a consequence, α helices and β sheets usually pack in one of a small number of relative orientations.
3. The links between secondary structures tend to be right-handed and short, and do not form knots. These features probably arise from the kinetics of the folding process.
4. A consequence of the principles outlined in 1–3 is that protein chains usually fold to give secondary structures arranged in one of a few common patterns. Thus, there are families or classes of proteins that have a similar tertiary structure but no evolutionary or functional relationship.
5. The stability of protein structures arises from the reduction in the surface accessible to solvent that occurs on folding and the formation of intramolecular hydrogen bonds. The chemical nature of the surfaces buried by residues and by secondary structures shows clear regularities. The total surface buried within proteins is a function of their molecular

weight. The principles outlined so far arise from the intrinsic chemical and physical properties of polypeptides, the thermodynamics of their structure, and the kinetics of the folding process. Quite different principles are involved in the following.

6. Functional determinants of protein structure—of the protein structures that are possible on thermodynamic and kinetic grounds, only some contain crevices appropriate for the formation of active sites.

CONFORMATION OF HELICES, PLEATED SHEETS, AND SIDE CHAINS

The refinement of protein structures using X-ray data of very high resolution has resulted in accurate descriptions of the conformations of helices, β sheets, and side chains. These confirm the dominating influence on conformation of the steric limits on torsion angles and of the need for buried polar groups to be hydrogen bonded (7–9). There are only rare instances of torsion angles outside the allowed regions or of buried polar groups without hydrogen bonds (10–15).

Hydrogen Bonds

Surveys of hydrogen bonds in parallel and antiparallel β sheets and in α helices show little or no significant difference in the principal features of their geometry. For all three secondary structures the mean bond lengths and angles are close to N...O, 2.9 Å; NH...O, 2.0 Å; NH...Ô=C, 155°; and N-Ĥ...O, 160°. Hydrogen bonds in 3_{10} helices and reverse turns are ~0.2 Å longer and a little more bent (12, 16, 17). The standard deviation of the NH...Ô=C angle is lower in α helices, ~6°, than it is in β sheets, ~12°. This is because the peptides in α helices tilt so that the carbonyl points ~15° away from the helix axis (18, 19), while in β sheets the peptides tilt both above and below the plane of the β sheets (J. Janin, unpublished information).

The constraints that hydrogen bonds place on protein structure principally arise from the need to keep bond lengths close to the optimum: Their standard deviations are ~0.15 Å. The constraints due to angular geometry are much less stringent.

α Helices

In α helices the mean values of main-chain torsion angles are very close to $\langle\phi\rangle = -65°$, $\langle\psi\rangle = -41°$ and $\langle\omega\rangle = 178°$ (12, 13, 16, 17, 19–22). The standard deviations of $\langle\phi\rangle$ and $\langle\psi\rangle$ are usually ~6°.

In two structures, determined at 1.2 Å and 0.98 Å respectively, it was noted that hydrogen bonds on the buried side of an α helix were about 0.2 Å

shorter than those on the side accessible to solvent (22, 23). The shortness of the interior hydrogen bonds is related to their peptide groups being less tilted away from the helix axis than those on the surface, whose carbonyl oxygens form additional hydrogen bonds to the solvent (18, 22, 24).

The residues at the ends of α helices are usually irregular in conformation. This commonly involves one or two residues at the N termini and two to four at the C termini. At the C termini a turn of 3_{10} or α_{II} helix is often found and occasionally a turn of π helix (12, 16, 17, 26).

The bending of α helices has been analyzed. Of the helices with 15 residues or more, two fifths had bends of 10°–30°. These bends are produced by small cumulative changes in a few $\phi\psi$ values in one region of the helix: Helices kink rather than bend smoothly. Bends of more than 20° involve the loss of one or two hydrogen bonds (A. M. Lesk, C. Chothia, unpublished).

Left-Handed α Helices, Polyproline Helices, and Turns

A single turn of a left-handed α helix is formed by residues 226–229 in thermolysin (27). This turn has the sequence –Asp–Asn–Gly–Gly– and its mean $\phi\psi$ values are 64°(15), 42°(9). This conformation is also found for single residues at the end of some α helices (28) and for a significant fraction of Asn residues (28a).

Helices with a conformation close to that of polyproline are observed in four proteins (13, 20, 22, 28b). The regions involved include residues 7–9 in pancreatic trypsin inhibitor, Gln–Pro–Pro; 59–63 in cytochrome c_{551}, –Ile–Pro–Met–Pro–Pro–; and 2–8 in avian pancreatic polypeptide, –Pro–Ser–Gln–Pro–Thr–Tyr–Pro– (13, 20, 22). Their mean $\phi\psi$ values are $-69°(6), 150°(6)$; $-78°(18), 140°(15)$; and $-72°(8), 140°(14)$ respectively; these $\phi\psi$ values are close to those of the polyproline II (9) helix, $-78°, 149°$.

The steric limits on main-chain torsion angles restricts the number of conformations possible for the turns in the protein chain that are formed by three or four residues (29–33). As noted in two detailed reviews (34, 35) the observed conformations are close to those expected. Turns nearly always occur on the protein surface as they contain several polar groups that do not form intramolecular hydrogen bonds (36, 37). The rare turns that are buried within proteins have trapped water molecules hydrogen bonded to these polar atoms (38).

β Sheets: Twisting and Coiling

The $\phi\psi$ values observed for β sheets are not localized as they are for helices but spread over most of the region normally allowed for extended polypeptide chains. This spread of $\phi\psi$ values gives the β sheets a twist that is right handed when they are viewed along their strands (39).

The twisted conformation of β sheets arises from a combination of factors

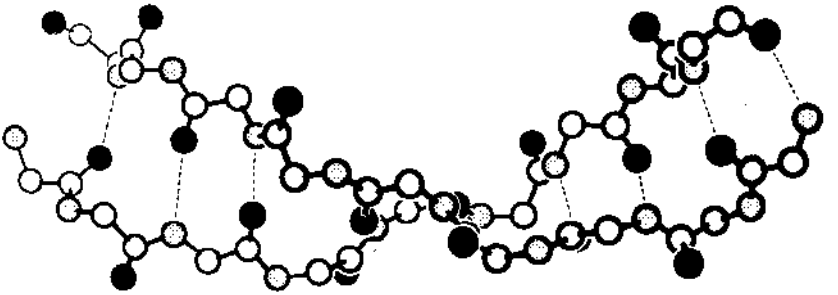


Figure 1 A two-stranded strongly twisted β sheet in which the strands are coiled in right-handed direction. Only main-chain atoms are shown. The structure illustrated here is formed by residues 4–12 and 16–24 in thermolysis (27). [Reprinted with permission from (51).]

acting at different levels. Model energy calculations, of differing degrees of sophistication, have shown that a conformation with a right-hand twist is favored for individual peptides (9, 39), by intrastrand forces (for polyalanine and polyvaline but not for polyisoleucine) and by interstrand forces (41–43). The twist is also favored at the tertiary structure level. β Sheets with a right-hand twist have surfaces complementary to those of α helices (see below). For β sheets packed together, a mean $\phi\psi$ value near the center of the normally allowed region permits the greatest degree of local accommodation. These factors act to differing extents in different protein structures and the degree of the twist in β sheets varies. The dihedral angle between adjacent strands varies between $-50°$ and $+10°$ and has a mean of $\sim -20°$ (44–46).

In strongly twisted β sheets the strands must be coiled as well as twisted otherwise they would splay apart (47). Such strands give coiled-twisted β sheets (Figure 1). The properties of model β sheets built from straight and coiled strands have been investigated (48–50). These models suggest the extent of twist is limited by hydrogen-bond geometries and that antiparallel β sheets can have greater conformational diversity than parallel. The fitting of models to observed structures showed that the strands of strongly twisted β sheets are indeed coiled. For strands to coil in the right-handed direction appropriate for the formation of strongly twisted β sheets, their $\phi\psi$ values must have the following relations (51):

$$\psi_i \simeq -\phi_{i+1}, \psi_{i+1} > -\phi_{i+2}, \psi_{i+2} \simeq -\phi_{i+3}, \psi_{i+3} > -\phi_{i+4}, \cdots$$

β Sheets: β Bends and β Bulges

β Sheets not only twist and coil; they can also fold upon themselves so the strand direction changes by $\sim 90°$. This occurs, for example, in orthogonal

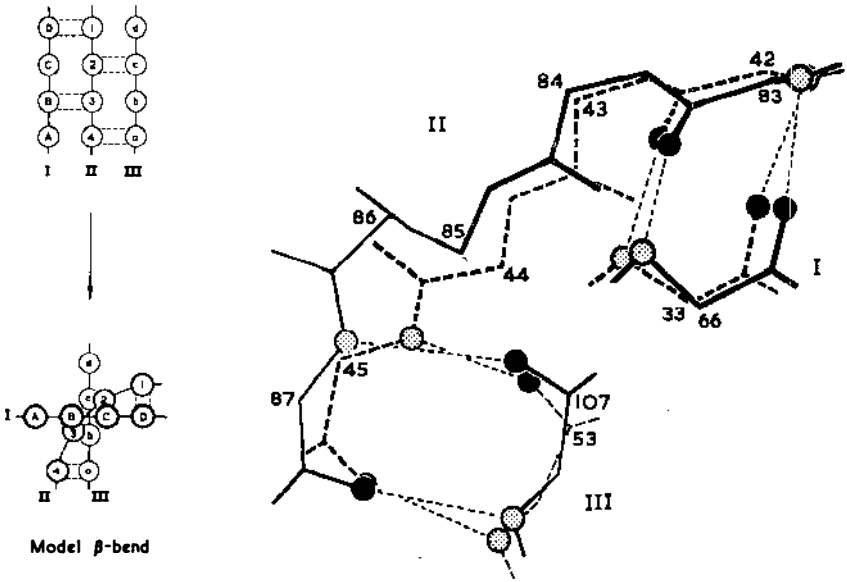


Figure 2 β bends and β bulges. The model on the left shows a flat β sheet (*top*) that twists and folds on itself so strands I and II are at right angles to each other and strand II makes a right-handed bend of $\sim 90°$. Hydrogen bonds are maintained. The bend in the central strand can be made by a residue in a polyproline conformation producing a local coiling of the strand [a β-bend (46)] or by the insertion of a residue in an α-conformation [a β-bulge (6, 51)]. On the right a β bend and a β bulge from elastase (11) are shown superposed. Main-chain atoms equivalent to model strand II and model residues ID and IIIa are shown. The superposition demonstrates the similar effect of β bends and β bulges on strands of β sheet. [Reprinted with permission from (46).]

β-sheet packings (see below and Figure 10). This local bending is usually produced in one of two ways: by β bends (46) or β bulges (6, 52). In β bends a residue with a conformation near that of polyproline produces a local coiling of the strand. In β bulges the insertion of a residue in an α conformation makes the change in strand direction. The net effect of both of these is to produce in the β sheets a right-handed bend of $\sim 90°$ (Figure 2).

Side Chains

Energy calculations have been used to calculate the preferred conformations of side chains and the results have been compared with the conformations observed in protein structures (53–62). There is very good agreement between the expected and observed conformations (17, 59, 60).

To a large extent, difference in side-chain conformation involves rotations around single bonds. Steric hindrance between substituents on the first and fourth atoms and in the case of Ser and Thr hydrogen bonding to main-chain polar groups mean that there are a small number of preferred

Table 1 Principal side-chain configurations observed in proteins[a]

Side Chain	Configurations[b] (%)				
Met, Glu, Gln, Lys, Arg	g^+t33,	tt28,	g^+g^+14,	tg^-8,	g^-t7
Leu	g^+t38,	tg^-19			
Ile	g^+t46,	g^+g^+16,	g^-t14,	tt10	
Val	g^+66,	g^-21,	t13		
Thr	e^+48,	g^-39,	t13		
Cys	g^+57,	t27,	g^-16		
Ser	g^-38,	g^+34,	t28		
Asn, Asp	g^+51,	t34,	g^-15		
Trp, Tyr, Phe	g^+57,	t31,	g^-13		
His	g^+45,	t45,	g^-10		

[a] Data taken from (59).
[b] The g^+, g^-, and t conformations are defined in Figure 3.

conformations for each side chain. These are listed in Table 1 (59). Inspection of five very accurate protein structures shows that the range of torsion angles within these different conformations is small (17).

The roles of the peptide backbone and the protein environment in side-chain conformation have been determined (59, 60). For a dipeptide the different preferred conformations (Table 1) have only small differences in energy. There is a weak relationship between secondary structure and side-chain conformation. The χ_1 values are on average 54% g^+, 35% t, and 11% g^- (59). (See Figure 3 for definition of t, g^+, and g^- conformations.) In

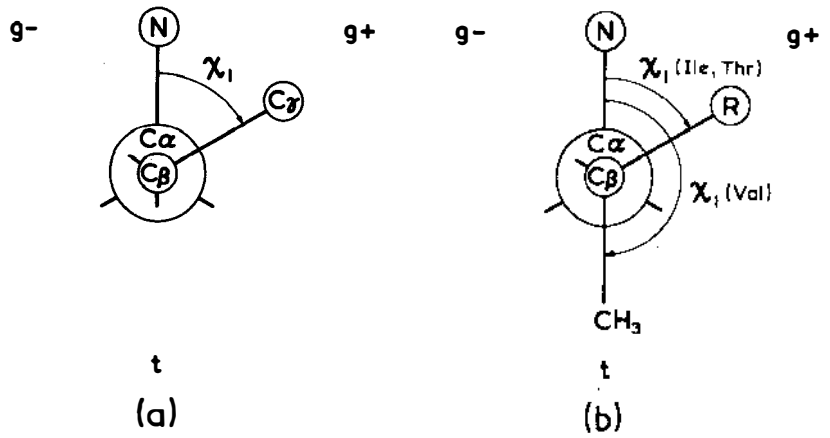


Figure 3 Torsion angle definitions for side chains. (*a*) For side chains unbranched at Cβ the ideal g^+, t, and g^- conformations are equivalent to χ_1 values of $-60°$, 180°, and $+60°$. (*b*) For Ile and Thr the ideal g^+, t and g^- conformations are also equivalent to χ_1 values of about $-60°$, 180°, and $+60°$. For Val, however, the ideal g^+, t, and g^- conformations analogous to those of Ile and Thr are given by χ_1 values of 180°, $+60°$, and $-60°$.

helices the t conformation restricts the $\phi\psi$ helical region and three-quarters of side chains have χ_1 in the g^+ conformation (59). Interaction of the side chain with the rest of the protein selects one preferred conformation and sharpens the potential around it (60).

CLOSE PACKING IN PROTEINS

The efficiency of residue packing in proteins—their packing density—has important implications for their structural, thermodynamic, and mechanical properties. The packing density of a molecule or residue is the ratio of the volume enclosed by its van der Waals envelope to the volume it actually occupies in a crystal, liquid, or protein interior. For organic solids, except those formed from molecules of unusual shape, this ratio has values in the range 0.68–0.80. For liquids the packing density is about 15% lower.

Protein Interiors

The controversy over whether the protein interior is more accurately described as an oil drop (63) or a crystalline molecule (64) has been decided in favor of the latter.

The comparison, for six proteins, of the sum of van der Waals volumes of their constituent amino acids with their partial specific volume gives packing densities of 0.72–0.77 (65). These values are typical of organic solids. The use of partial specific volumes in this calculation is not quite correct (see below), but the general conclusion has been confirmed by all latter work.

The introduction of the Voronoi polyhedra procedure to calculate, from atomic coordinates, the volume occupied by protein atoms gave the first quantitative description of packing in proteins (66). The application of this method and how detailed results vary with different parameters has been discussed by several authors (3, 66–69). For residues buried in the protein interior the results are unambiguous and accurate within the limits of the atomic coordinates used.

The initial use of the Voronoi procedure showed that the packing density in the interior of ribonuclease and lysozyme is the same as that of organic solids (66, 67). The mean volume occupied by residues in the interior of nine proteins was calculated (70). The results are given in Table 2, and show that the mean volume of a residue in the interior proteins is the same as that in crystals of its amino acid (Figure 4). The standard deviation of the mean residue volumes is $\sim 6\%$. A significant but unknown part of this is due to the inaccuracies of the atomic coordinates used in the calculations.

Within the protein interior the main-chain regions that form hydrogen bonds have a higher density than the side-chain regions that are packed by

Table 2 The mean volumes of residues buried in proteins[a]

Residue	Volume (Å^3)	Residue	Volume (Å^3)	Residue	Volume (Å^3)
Gly	66	Ser	99	His	167
Ala	92	Thr	122	Asn	135
Pro	129	Cys	106	Gln	161
Val	142	Cyh	118	Asp	125
Ile	169	Met	171	Glu	155
Leu	168	Tyr	204	Lys	171
Phe	203	Trp	238	Arg	202

[a] Data from (70) except for the Arg value that comes from (156).

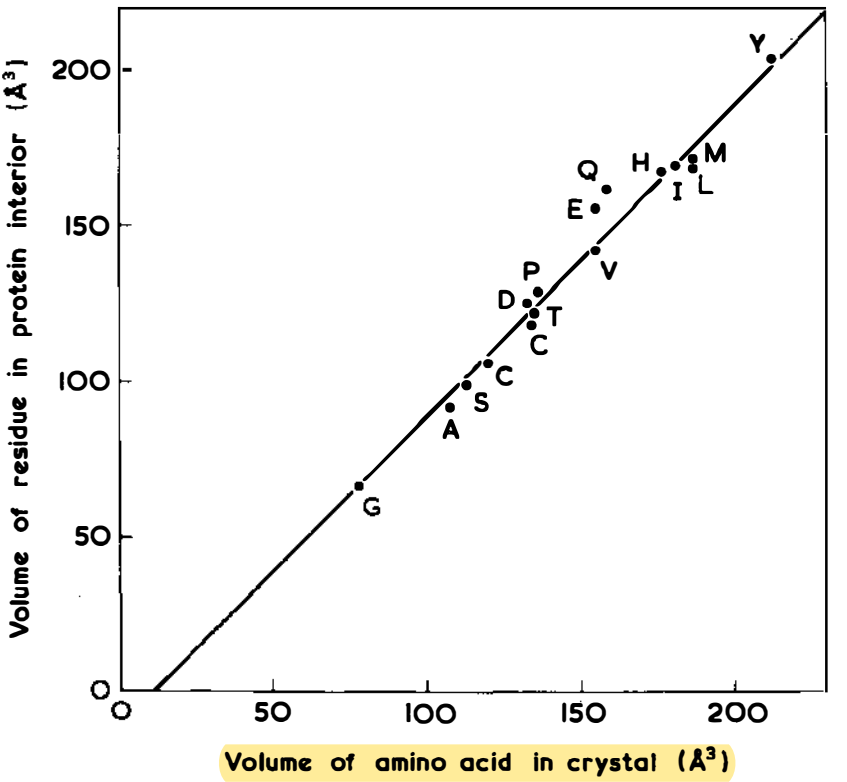


Figure 4 The mean volume of residues buried in protein interiors plotted against the volume of their amino acid in crystals. The line has a slope of 1 and an intercept of 11 Å^2—the volume lost by an amino acid on forming a residue. Drawn from data in (70).

van der Waals forces (71). Most of the contacts formed by residues are to others close in the sequence or, if it is in a secondary structure, to those in the same α helix or β sheet (72).

Even with perfect close packing, $\sim 25\%$ of the total volume is not occupied by protein atoms. Nearly all this space is in the form of very small cavities. Cavities the size of water molecules or larger do occasionally occur (73, 74) but form only a small percentage of the total protein volume. The size of the cavities varies with thermal fluctuations of their surrounding atoms. Such fluctuations, or mobile defects, facilitate in proteins dynamic behavior such as the rotation of side chains (74–76).

Protein Surfaces

The Voronoi procedure can be used to estimate contact areas, surface roughness, and solvent accessibility (68). For residues on the surface of proteins the calculation of atomic volumes depends upon the structure of the solvent shell. This has not been determined for any protein used in these calculations and the use of artificial solvent shells gives results of uncertain accuracy.

If the total volume of a protein is calculated from its residue composition using the volumes in Table 2, a value is obtained that is about 10% greater than the experimentally determined partial specific volume. The reason for this discrepancy is not known. A plausible explanation is that the rough protein surface occupies interstices in the open water structure. Semiquantitative calculations support this idea (77).

PACKING OF α HELICES AND β SHEETS

Two factors have dominating influence on the packing of α helices and β sheets in proteins. First, residues in the protein interior close pack; second, packed secondary structures have a conformation close to the minimum free-energy conformation of the isolated secondary structures (see previous sections). These factors imply that the manner in which α helices and β sheets pack depends upon the shape of their surfaces (78).

The geometrical principles governing the packing of secondary structures have been embodied in a set of simple models. These models assume that, to a first approximation, surface shape arises from the main-chain conformation of α helices and β sheets. Size, shape, and conformation of side chains determine which particular class of packing occurs and modulates its exact geometry. Comparison of the geometrical predictions of the models with the features observed in real packing shows that this assumption is usually correct.

Here I outline only the different packing models. Details of the

comparison between these models and packings observed in known protein structures can be found in the references cited. I do not discuss the attempts to predict the three-dimensional structure using packing and chain topology rules [this work has been recently reviewed (79)].

Packing of α Helices

The problem of how α helices pack together was first discussed in 1953 when a model colloquially known as "knobs into holes" was developed (80). With the availability of atomic coordinates of a large number of globular proteins it became apparent that helices do not generally pack with their axes inclined at $+20°$ and $-70°$ as expected from the knobs into holes model. The observed distribution of angles covers the whole range of possible values though there is a sharp peak at $-50°$ with a shoulder at $+20°$ (81; see Figure 5). Several analyses of helix packing have been published (78, 80–85) and a comparison of their results is given in (81).

The "ridges into grooves" model was developed to explain the observed distribution of interaxial angles and the patterns formed by residues that make interhelix contacts (78, 81; Figure 6). In this model, residues on the surfaces of helices form ridges separated by grooves. Helices pack together by the ridges of one helix packing into the grooves of the other and vice versa. The ridges on the helix surface are usually formed by the side chains of residues four separate in the sequence: i, $i+4$, $i+8$,... and $i+1$, $i+5$, $i+9$,... (the $\pm 4n$ ridges). Occasionally they are formed by residues whose separation is three ($\pm 3n$ ridges), and, rarely, by adjacent residues (the $\pm 1n$

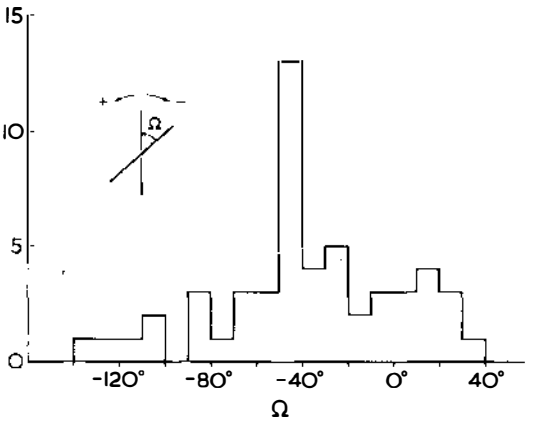
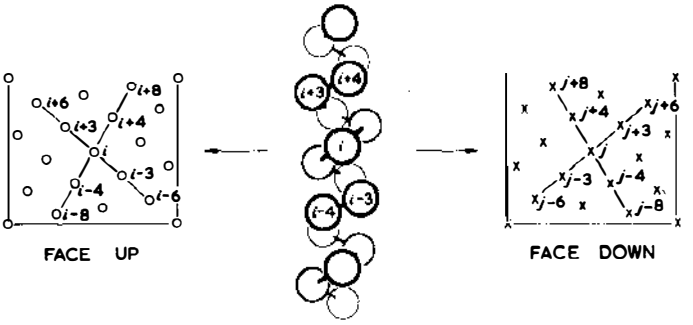
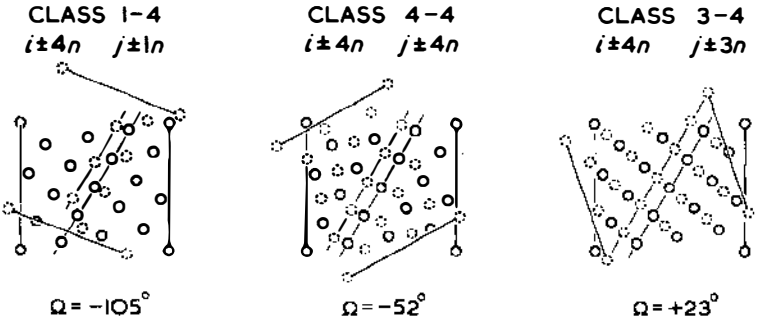
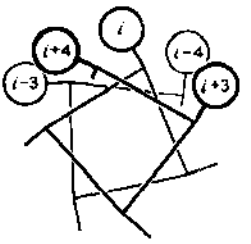


Figure 5 The relative orientations observed for packed α helices in proteins. The angle Ω describes the relative orientation of packed secondary structures. It is defined as the angle between the strands of β sheet and/or helix axes when projected onto their plane of contact. Data for this figure is taken from (81) and supplemented by my unpublished calculations.







ridges). Intercalculation of different ridges and grooves by packed helices inclines the axis at different characteristic angles: $\sim -50°$ when both helices use $\pm 4n$ ridges (class 4-4); $\sim +20°$ when one helix uses $\pm 4n$ ridges and the other $\pm 3n$ ridges (class 3-4), and so on (78, 81; Figure 6). These different packing classes give different patterns for the contacts between residues at the interface.

The use of particular ridges ($\pm 4n$, $\pm 3n$ or $\pm 1n$) by a helix depends upon the size and conformation of side chains, and upon the surface of the other helix. The $\pm 4n$ ridges are more commonly involved in packing because side chains with this separation splay apart less than do side chains in $\pm 3n$ ridges or $\pm 1n$ ridges (Figure 6).

The general validity of this packing model can be demonstrated by direct inspection of helix interfaces observed in protein structures (81). The actual Ω values depart somewhat from ideal values because of heterogeneity in the size of side chains and because of variations in helix geometry. Larger departures occasionally occur when, in a ridge, a very small residue is next to a large one, as the gap produced can allow ridges to cross each other. Thus, the wide distribution of observed Ω values arises from the different packing classes and the spread of values within particular classes (Figure 7). The dominance of the $\pm 4n$ ridges explains the peak at $\Omega \simeq -50°$ as this is the orientation when helices pack with both using the $\pm 4n$ ridges.

In helix packings the distance between the helix axes varies between 6.8 Å and 12.0 Å. This variation is principally a function of the size of the side chains at the center of the interface (81). The mean interaxial distance for packed helices is 9.4 Å (81). The mean interpretation of atoms at the interface is 2.3 Å (84). Thus, the contacts between packed helices mainly involve the ends of side chains. This observation is important in considering how helices can move relative to each other (86).

α-Helix–β-Sheet Packing

Analysis of observed protein structures shows that there is a strong tendency for the helices packed onto sheets to have their axes nearly parallel

Figure 6 The ridges into grooves model for the packing of α helices. Helical nets can be used to describe the relative position and the packing of residues. The nets are made by marking the position of residues on a piece of paper wrapped around a helix. Placing a face-down net over a face-up net gives a representation of the residue arrangements that can occur at helix interfaces. The central part of the figure shows how packing the $j \pm 1n$, $j \pm 4n$, and $j \pm 3n$ rows of one helix between the $i \pm 4n$ rows of a second helix inclines their axis at ideal angles of $-105°$, $-52°$, and $+23°$. The $i \pm 4n$ rows are more common in helix packings because residues in these rows splay apart less than those in other rows (see lower part of figure). On going from $i-4n$ to $i+4n$ we rotate 80° around the helix axis; on going from $i-3n$ to $i+3n$ we rotate 120°. [This figure is a modified version of two figures in (81).]

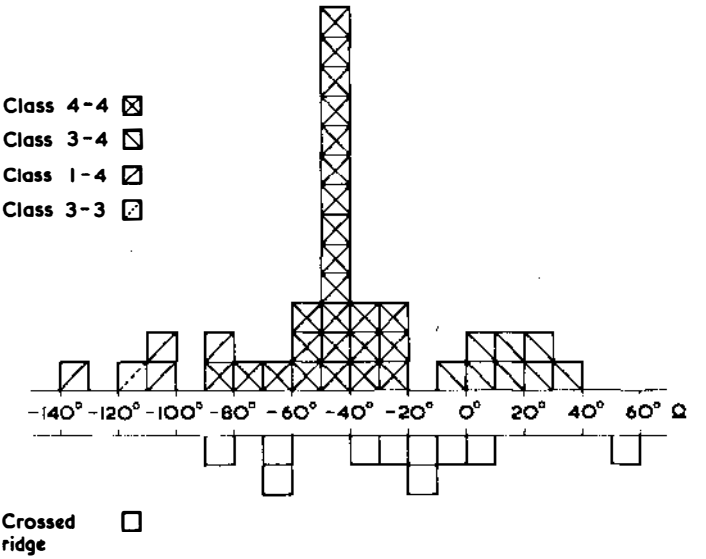


Figure 7 The distribution of observed Ω values for the different helix-helix packing classes. The letter Ω is defined in the caption of Figure 5 and the packing classes in Figure 6 and text. [Reprinted with permission from (81).]

to the sheet strands (44, 78, 87; Figure 8). A model for the packing of α helices onto β sheets (44, 78) is illustrated in Figure 9. I call this the complementary twist model for helix-sheet packing.

The β sheets observed in globular proteins have a right-handed twist

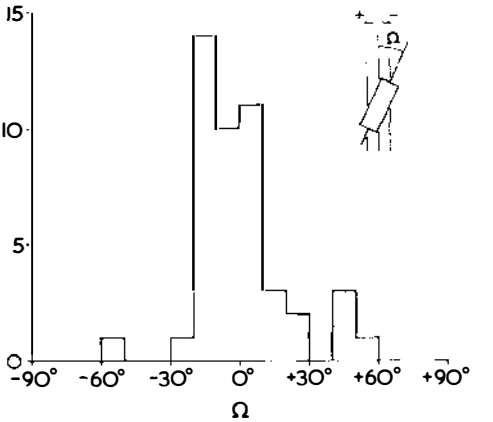


Figure 8 The relative orientations observed for α helices packed onto β sheets. The letter Ω is the angle between the helix axes and β strands (see caption to Figure 5). Data taken from (44, 87).

when viewed in a direction parallel to the polypeptide chain. Now in an α helix the adjacent two rows of residues $i, i+4, i+8, \ldots$ and $i+1, i+5, i+9, \ldots$ form a surface that also has a right-handed twist (Figure 9). The essential feature of the packing model is that α helices pack onto β sheets with their axes parallel to the β-sheet strand because in this orientation these two rows

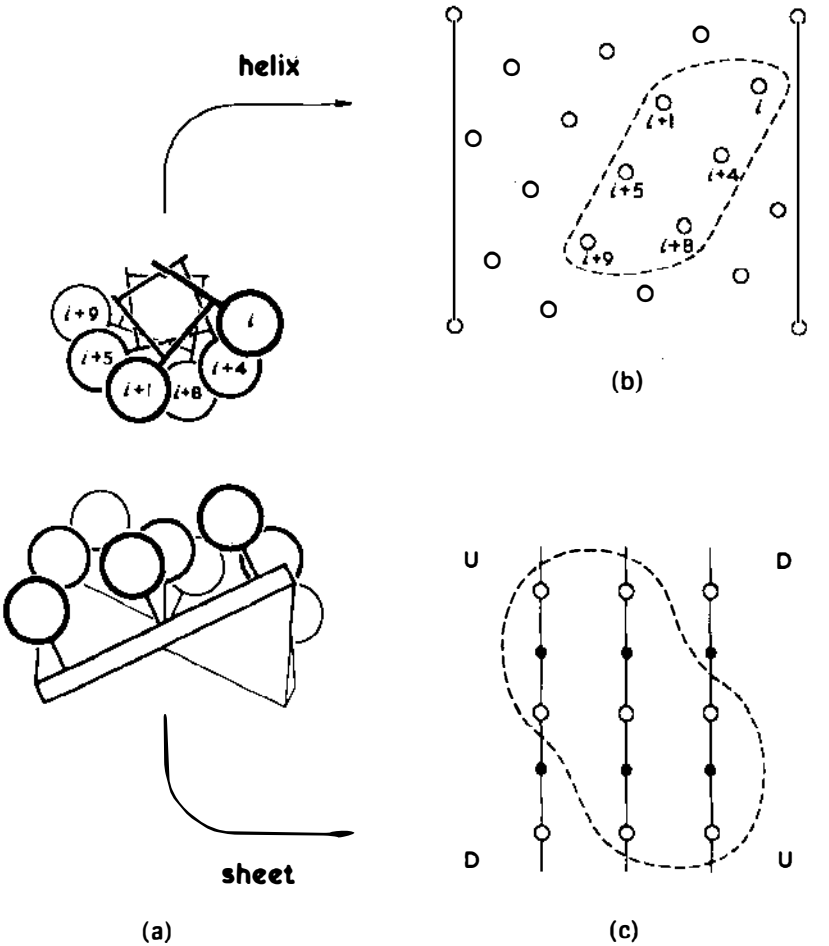


Figure 9 The complementary twist model for α-helix to β-sheet packing. Side chains on one side of an α helix and a twisted β sheet are shown as open circles (*a*). When the helix axis is parallel to the strand direction, the helix residues $i, i+1, i+4, i+5, i+8, i+9, \ldots$ form a twisted surface that is complementary to that of the twisted β sheet. The patterns formed by the residues in contact at the interface are shown on flattened projections of the α helix (*b*) and β sheet (*c*). U and D mark the corners of the β sheet that move up and down from the plane when the β sheet is twisted. Note that the β-sheet contact residues cluster about a line joining U to U, the concave diagonal of the β sheet. [Reprinted with permission from (44).]

of residues have a surface complementary to that of the twisted β sheet (Figure 9).

Two detailed analyses of helix-sheet packings (44, 87) have demonstrated the validity of the complementary twist model. The model predicts that the helix residues forming the interface with the β sheet will have the sequential relationship of the type $i, i+1, i+4, i+5, i+8, \ldots$ (Figure 9). For observed packing $\sim 90\%$ of the helix interface residues do have relationships of this type (44, 87). A subset of these, $i+1$, $i+4$, $i+5$, and $i+8$, forms the larger part of the contact (87). On the β-sheet side of the interface residues are expected to cluster about a line parallel to the concave diagonal of the β sheet (Figure 9), and this is indeed observed. Note that these oblique patterns for the interface residues occur even though the α-helix axis and β-sheet strands are parallel.

If two α helices pack side by side on the face of a β sheet its twist will rotate one helix relative to the other. From the mean values of β-sheet twist ($-19°$ between strands ~ 4.5 Å apart) and of the distance between helix axis (9.4 Å), we would expect such helices to be inclined at $\sim -40°$ to each other. This implies that the interface between helices must be formed by the $\pm 4n$ ridges (class 4-4), as use of other ridges would produce a quite different relative orientation (see previous section). Inspection of the helices that pack onto β sheets and against each other shows that four fifths pack as expected (44). The other fifth are in other packing classes; they have one end of the helix in contact with the β sheet and then splay away.

There is one issue over which the two analyses of helix-sheet packing have led to different conclusions. This concerns the details of the side-chain packing. As the periodicity of residues in α helices is different to that in β sheets, there cannot be the regular overall pattern of intercalation found in helix-helix packing. One group (44) argues that the helix and sheet surfaces can be regarded as essentially smooth with only occasional irregular intercalation where particularly large or small residues occur. The other group (87) argues that regular local intercalation occurs at the center of the contact with 4 helix residues, $i+1$, $i+4$, $i+5$, and $i+8$, surrounding one β-sheet residue. Further investigation is required.

Helices are sometimes associated with orthogonal β-sheet packings (see below). No analysis or discussion of this minor class of helix-sheet packing has been published.

Packing of β Sheets

β Sheets, in those proteins formed by their association, are of two types. In one, the β sheets are just twisted. In the other, they are both twisted and folded upon themselves (78; Figure 10). This particular folding of β sheets is produced by a local coiling, or a β bulge, bending the strands by $\sim 90°$. The

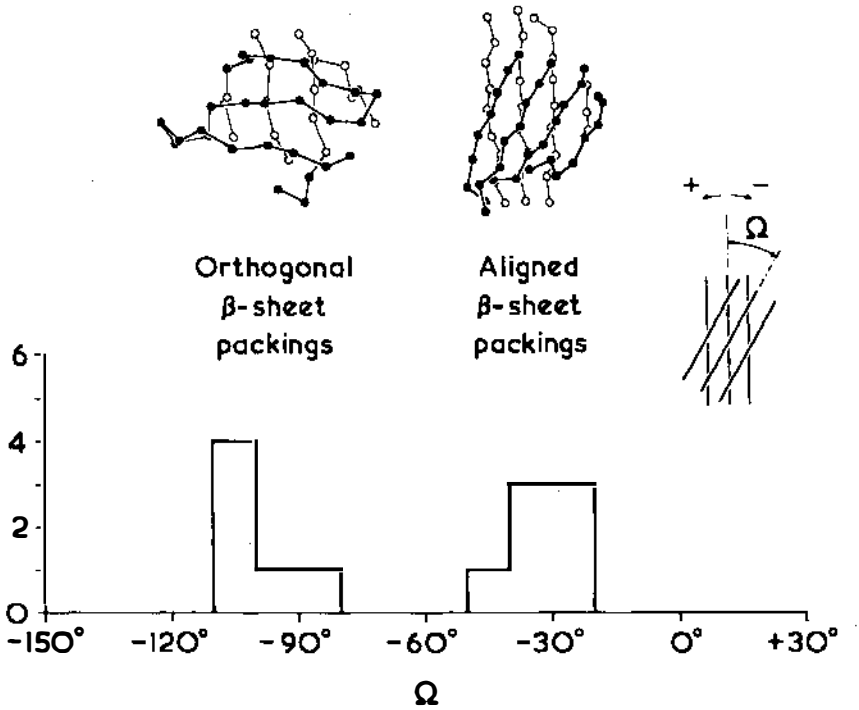


Figure 10 The relative orientation of packed β sheets in proteins in the aligned and orthogonal classes. The letter Ω is defined in the caption to Figure 5. The upper part of the figure shows an example of each class. Circles, representing the Cα atoms of residues in β sheets, are closed for the β sheet above the plane of the page and are open for the β sheet below the plane. Note how in the orthogonal packing strands move from one β sheet to the next without interruption. Data in this figure were taken from (45, 46).

two types of β sheet pack differently giving *aligned β-sheet packings* and *orthogonal β-sheet packings* respectively (Figure 10; 46).

ALIGNED β-SHEET PACKINGS Structures in this class are formed by the face-to-face packing of β sheets that can be regarded as essentially independent: In nearly all cases the links between the β sheets involve several residues in a non β-sheet conformation (45). In this class the main-chain direction in one β sheet is at an angle of $\sim -30°$ to the main-chain direction in the other β sheet (78; Figure 10).

Models for, and analyses of, aligned β-sheet packings have been published (45, 88). The models relate the relative orientation of the packed β sheets to their right-handed twist. Figure 11 shows this relation. On the left of this figure I show two hypothetical untwisted β sheets packed with the rows of side chains at their interface aligned. On the right I show what

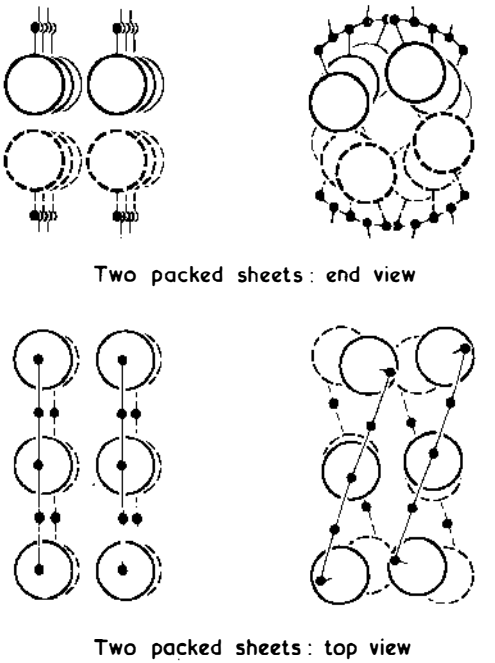


Figure 11 A model for the aligned packing of β sheets. The main chain is represented by small closed circles and the side chain at the interface by large open circles. The figure shows views of packed β sheets with two strands each. On the left the β sheets are flat and residues at the interface aligned. On the right we show what happens when the β sheets twist about an axis in the plane of their interface. The ends of the side chains maintain their alignment but the main chains now point in different directions.

happens when the β sheets are given a right-handed twist about an axis parallel to the strands and in the plane of their interface. Though the ends of the side chains that form the contact across the interface maintain their alignment, the main chains of the two β sheets do not. They point in different directions. If one looks from on top, the main chains make an angle (Ω) that is negative. If the main chains are 10 Å apart and the twist along the chain 4°/Å (typical values), Ω is $-35°$. The observed values of Ω are in the range $-20°$ to $-50°$, departure from the model value arising from variation in the extent of the twist and from the imprecise alignment of side chains at the interface.

The residues that form the β-sheet–β-sheet interface have an anti-complementary pattern and this has also been related to their twist (88).

ORTHOGONAL β-SHEET PACKINGS The packings in this class are formed by β sheets folded on themselves. As in the aligned class, the β sheets are twisted and pack face to face. They differ in that the direction of the main

chains in the different layers are at ~90° and in that the strands at one corner, or two diagonally opposite corners, pass from one layer to the next without interruption (46; Figure 10).

A model for orthogonal β-sheet packings (46) is shown in Figure 12. The principal features arise from the right-handed twist of the β sheets and from the manner in which they fold upon themselves. The folding is produced by the strands at the corners that go directly from one layer to the next by right-hand 90° bends. These bends are usually produced by a local coiling of the strands, a β bend, or by a β bulge (46, 52). The right-handed geometry of these conformations arises from the values normally allowed for $\phi\psi$ (see a previous section).

The effect of the β sheets in the two layers being twisted and inclined at 90° is that they are only in contact along one diagonal. Along the other diagonal they splay apart (Figure 12). The space between the splayed corners is usually filled by helices or large side chains in loops (46).

The β-sheet hydrogen bonding in these structures is often shown on

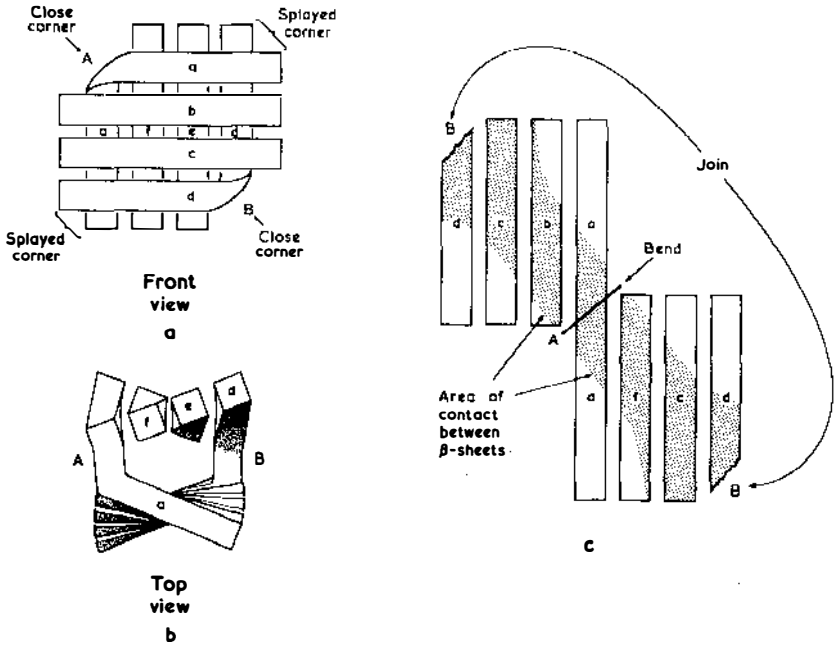


Figure 12 A model for the orthogonal packing of β sheets (*a*, *b*). The strands are represented by rods with square cross-section. The β sheets have the normal right-handed twist so that two corners come close together (*A* and *B*) and two splay apart. At the close corners, strands pass from one β sheet to the next through β bends or β bulges (see Figure 2 and text). If the packing is opened out by straightening bend *A* and breaking the chain at *B* we get a single β sheet with staggered strands (*c*). [Reprinted with permission from (46).]

two-dimensional plans obtained by the equivalent of straightening and breaking the chains that go between layers (Figure 12). These plans show the structure as a single β sheet with staggered strands and hydrogen bonds going right round. This latter feature, and the appearance of these structures in certain projections, has led to their being referred to as barrels or cylinders. Such descriptions are inaccurate and possibly misleading (46).

Residue Composition and Packing in β Sheets

The β-sheet surfaces that form interfaces to other secondary structures are predominantly formed by five residues—Val, Leu, Ile, Phe, and Ala. In Table 3 we give the proportion of these residues in β sheets, in the interface regions of β sheets, and for buried residues in proteins. For β sheets, the five residues form about two thirds of the interface regions with just three—Val, Leu, and Ile—forming half (44-46).

The hydrophobicity of the inner strands of β sheets and the high proportion of Val and Ile in β sheets, particularly in parallel β sheets, has been noted by several authors (89–93). It has been suggested that this arises from the intrinsic nature of interaction between strands in the β sheets. However, the high Val, Leu, and Ile content is only a characteristic of the interface regions. Parallel β sheets usually have both faces covered by α helices with two thirds of the residues being part of an interface (44, 87). Antiparallel β sheets are usually packed against just one other β sheet and only one third of their residues are part of an interface. The more prominent

Table 3 Common nonpolar residues in β sheets, α helices, and their contact regions

	Parallel β sheets[a]		Antiparallel β sheets[b]		α Helices[c]	
Residue	Total β sheet (%)	Contact region[d] (%)	Total β sheet (%)	Contact regions[d] (%)	Total α helices (%)	Contact regions[d] (%)
Val	18	24	14	22	7	12
Leu	9	10	8	16	7	12
Ile	11	14	8	15	6	9
Phe	4	6	4	11	3	6
Ala	8	7	8	8	9	12
Total	50	61	42	72	32	51

[a] Data from (44). Similar values for total composition are given in (92, 93).
[b] Data from (45, 46). Similar values for total composition are given in (92, 93).
[c] Data from (81).
[d] Contact regions are those parts of the β sheets or α helices that make contacts with other β sheets or α helices.

part played by interface residues in parallel β sheets accounts for their high overall Val and Ile content (45, 46).

The common occurrence of Val, Ile, and Leu in interface regions probably arises from their packing requirements. The spacing of residues along the strands ($\sim$7.0 Å) is different to that between strands ($\sim$4.5 Å), so in orthogonal packings and in any but the most ideal aligned packing there could not be a regular pattern of intercalation. This also applies to helix-sheet interfaces (see above). Well-packed surfaces favor Val, Ile, and Leu residues that make good interstrand contacts. Intercalation for particularly large or small residues is occasional and irregular (44–46).

CHAIN TOPOLOGY IN PROTEINS

The examination of the topology of protein chains (37, 94–107) has shown clear regularities. These regularities have been described by different authors in different ways. A comparison of their results suggests that the topological regularities so far observed can be expressed as three rules.

1. *Pieces of secondary structure that are adjacent in the sequence are also often in contact in three dimensions* (37, 94, 97, 102–104, 107). Three possible structures for sequentially linked pieces of secondary structure are shown in Figure 13: $\alpha\alpha$, two antiparallel packed α helices; $\beta\beta$, two antiparallel β-sheet strands; and $\beta\alpha\beta$, a helix packed against two adjacent parallel β-sheet strands. An examination of 31 proteins showed that two thirds of the pieces of secondary structure were part of an $\alpha\alpha$, $\beta\beta$, or $\beta\alpha\beta$ unit (37). These substructures were named folding units. The number of folding units in the 31 proteins is *close* to the maximum possible. An examination of more recent protein structure shows that $\alpha\beta\beta$ and $\beta\beta\alpha$ folding units (a helix packed against two adjacent antiparallel strands) occur to a lesser but still significant extent. This close connectivity of strands and helices is very significantly different to that which would be expected if connections were made randomly (37).

   The common occurrence of $\beta\alpha\beta$ units had previously been noted and called super secondary structures (96). An extensive analysis of the connections within β sheets has been made (102–104, 107).
2. *The connections in β-X-β units* (where the βs are parallel strands in the same β sheet, though not necessarily adjacent, and X is an α helix, a strand in a different β sheet, or an extended piece of polypeptide) *are right-handed* (Figure 13). This observation, made for certain $\beta\alpha\beta$ units (96), was later shown to be generally true for proteins (98–101). There are over a hundred cases that satisfy this rule and only three exceptions (6).
3. *The connections between secondary structures neither cross each other nor make knots in chain.* This rule is partly or totally stated in somewhat

different ways in a number of papers (102–105a). The only clear exception at present known to this rule is the knot found in the chain of carbonic anhydrase. It is a rather trivial exception in that the knot is formed by the N terminus of the chain looping round the C terminus (102). However, loop penetration is occasionally observed in protein structures (105a; D. F. Dykes, quoted in 105b).

Topological rules in addition to these three have been put forward. In all cases so far they can be shown to be the same as, or arise from a combination of, the three given here. This is the case with the left-handed

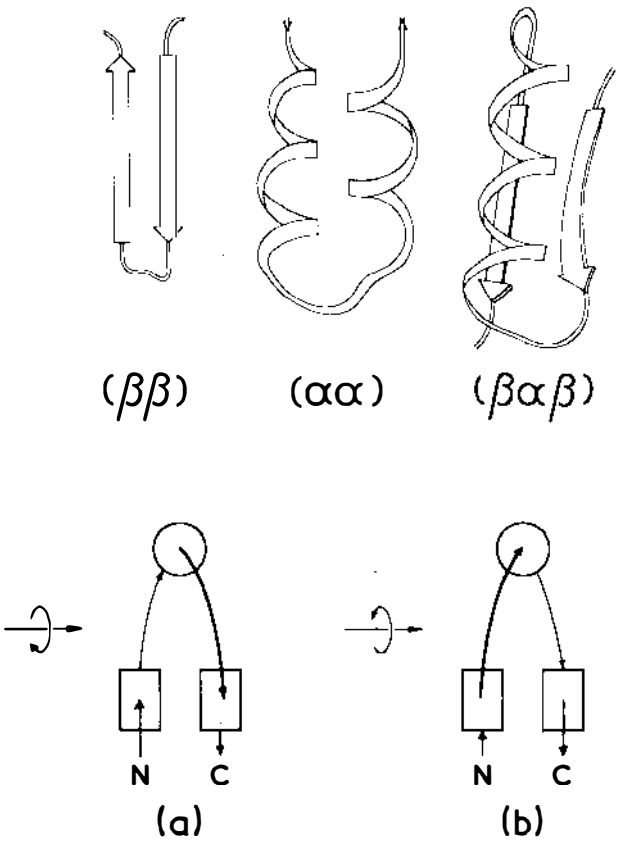


Figure 13 Topological feature of proteins: αα, ββ and βαβ are folding units, or super-secondary structures, frequently found in protein (32). βαβ Units are represented in (*a*) and (*b*) in an end-on view with circles representing an α helix and rectangles the β-sheet strands (see 37). This unit can be right-handed (*a*) or left-handed (*b*). In nearly all cases they are observed to be right-handed (6, 96, 98–101). This is also true when the α helix is replaced by a β-sheet strand or an extended chain.

"Greek Key" and "Jelly Roll" topologies (6), which can be shown to arise from a combination of the rules given above (C. Chothia, unpublished).

The origin of the topological regularities in proteins is uncertain. There are no apparent thermodynamic origins and explanations have centered around the kinetics of protein folding. It has been suggested that the limits on the rates of diffusion of chain segments favor the association of those close in the sequence (rule 1). Estimates of the actual rate of diffusion (108), however, make this explanation unlikely. Another explanation is that, during folding, intermediates made up from segments close in the sequence are used because they have a lower entropic cost than those formed by segments distant in the sequence.

CLASSIFICATION OF PROTEIN STRUCTURES

The packing models and topology rules embody the basic principle determining the arrangement of secondary structures in nearly all proteins. Actual structures are built by a particular set of variations on, and combinations of, these models and rules. (They also have sufficient stability and a function—see later sections.) The small number of common packing classes and the limitations on chain topologies mean that certain arrangements of secondary structures occur frequently. Thus, proteins quite unrelated by evolution or function may have similar structures (37, 98–102). This also means that a system for the structural classification of proteins or protein domains can be made on the basis of the arrangement of their secondary structures (4–6, 37).

The original system (37) divided proteins into four classes:

1. All-α proteins are composed mainly of α helices. One common motif is the "four α-helical bundle" (109)—four are nearly parallel but with a small left-hand twist because class 3-4 helix packings put their axes at $\sim +20°$ to each other (81). Proteins built by other classes of helix packing tend to be larger and more heterogeneous, presumably because it is difficult to make stable proteins with a few α helices inclined at large angles.
2. All-β proteins are composed mainly of β sheets. The large majority of proteins in this class contain aligned or orthogonal β-sheet packings with connection between the strands limited by the topology rules.
3. α/β Proteins have α helices and β-sheet strands that approximately alternate along the chain. These structures usually contain a central β sheet with α helices packed on both sides. The right-handed nature of the $\beta\alpha\beta$ units (see above) is important for the function as well as the structure of this class of protein (see a later section).
4. $\alpha+\beta$ Proteins contain α helices and β-sheet strands that tend to

segregate rather than alternate along the chain. The amount of α helix or β sheet in these proteins varies, making their structure heterogeneous. If the amount of either is appreciable the two secondary structures pack as in the all-α or all-β classes. The contact between α helices and β sheets may then involve packings of the kind found in α/β or in the splayed corners of orthogonal β-sheet packings (46).

The original four classes have been divided by some authors (5, 6), who give lists of the proteins they consider as belonging to each subclass. These subclasses are based on function, possible evolutionary relationships (5), or the details of chain topology (6). This more elaborate approach can produce inconsistencies. For example, the system that has a strong emphasis on chain topology puts in the same subclass proteins that would be in separate subclasses if their packing was emphasized.

The great variation possible within normal packing and topology rules, together with the rare exceptions to these rules, means that any classification systems that cover the large majority of protein structures can only be approximate. It can be a reasonable but not exact guide to the secondary structures present in the protein, the nature of their packings, and the topology of the chain.

PROTEIN STABILITY: HYDROGEN BONDS, HYDROPHOBICITY, AND SURFACE AREAS

The previous sections have discussed the principles governing the conformation and packing of secondary structures and the topology of the links between them. To be viable, however, a protein must also have overall stability. The net energy gained from hydrophobicity, hydrogen bonds, van der Waals forces, and electrostatic interactions must compensate for the loss of conformational entropy and any conformational strain.

Although the earlier emphasis on hydrogen bonding as the major source of protein stability (7) later shifted to hydrophobicity (1, 63), it is clear that both forces have very important effects on the structure of proteins. Recently the contribution of electrostatics has been discussed. Though electrostatics is very important for protein function, the magnitude of its contribution to protein stability and its influence on actual structure is not clear at the moment. As we have already seen, van der Waals forces (packing) determine the manner in which secondary structures associate. They may also contribute to stability (110, 111).

Hydrogen Bonds

Estimates of the energy of hydrogen bonds between protein atoms relative to those between protein atoms and water were made by measuring the

association in water of small molecules containing groups similar to those found in proteins, e.g. urea and N-methylacetamide (112–116). These experiments gave very small energy differences, suggesting that hydrogen bonds made only a marginal contribution to protein stability. Recently it has been pointed out that, for entropic reasons, intramolecular interactions can have substantially greater free energies than intermolecular, so the contribution of hydrogen bonds can be substantial (117). This also accounts for the lower enthalpy of the folded state (110, 111, 117).

An early survey of protein structures showed that the proportion of polar groups that form intramolecular hydrogen bonds is essentially constant and close to 0.5. As proteins have similar amino acid compositions, this implies that the contribution made by hydrogen bonds to protein stability is proportional to molecular weight (70).

Hydrophobicity and Surface Areas

The hydrophobic contribution to protein stability arises of course from the folded protein making fewer contacts with water than the unfolded. With the introduction of algorithms to calculate the surface areas of protein atoms (73, 118–120), it became possible to calculate changes in the areas of protein-solvent and protein-protein contact that occur in folding or association. Three different measures of the protein surface have been used: accessible surface area, accessibility, and contact surface area. These are defined and the relations between them discussed in (3). For the same structure, the value for the accessible surface area is *approximately* 3.5 times as large as that for the contact surface area.

For amino acids there is a good correlation between their surface area and the energy of transfer from water to organic solvents (121, 122). [A very similar correlation is found for other molecules (123–128).] The correlation suggests that the changes in surface area that occur in any process can be used to calculate the contribution of hydrophobic free energy.

Four important features of the relation between surface area and hydrophobic free energy include the following:

1. For any particular set of transfer experiments the correlation is very high, >0.99 (121, 122).
2. The magnitude of the unitary free energy derived from the correlation is dependent upon the nature of the organic solvent involved in the transfer process (129). The value obtained from amino acid data varies from 86 (Ethanol) to 131 (Carbon Tetrachloride) cal/Å^2 of contact surface with a mean 100 cal/Å^2 (122). For accessible surface areas a mean value of 25 cal/Å^2 has been used (70).
3. The relationship is also found for the transfer of amino acids from water to 8 M urea and 6 M guanidine hydrochlorite (130) and shows that the

hydrophobic interactions are reduced relative to water by 7.1 and 8.3 cal/Å^2 of accessible surface. This can explain how urea and guanidine hydrochloride unfold protein and the greater effectiveness of the latter (130).

4. The different correlations show small threshold values that vary between $\sim$10 and $\sim$40 Å^2 of accessible surface area [(128, 130) and inspection of data in (121, 122)].

Changes in surface area in proteins have been widely used to calculate hydrophobic free energies. If these results are used without some estimate of the other nonhydrophobic energy changes, they can be misleading.

Objections have been made to the application to proteins of the values obtained from amino acid transfer experiments. The values are only a half or a third of that derived from macroscopic concepts of surface chemistry (131). Surface area calculations do not account for solvent effects on the dimerization of methane nor on the gauche/trans equilibrium of butane (132).

Residue Surfaces and Environments

Accessible surface areas and accessibilities have been published for residues in proteins, model tripeptides, (Gly–X–Gly and Ala–X–Ala) and dipeptides (73, 118, 133, 134). Surface areas calculated for residues and side chains in different conformations show small variations (118, 134). The extent to which residue types are buried in proteins has been determined (133, 135) as has the composition of the inside and surface of proteins (135). Some of these results are given in Table 4. They show that about half of the nonpolar residues are buried within the protein, about a quarter of the residues with one polar atom, and about one tenth of the residues with two polar atoms or a charged atom. The residues in this last group are, when buried, usually involved in the coordination of metals, part of domain interfaces, or part of the active site.

The partition, between the inside and outside of proteins, of these *groups* of residues correlates well with the hydration energy of their side chains (136), i.e. their affinity for water. Within the nonpolar group hydrophobicity differentiates the lower extent to which Gly and Ala are buried but it does not differentiate between Val, Ile, Leu, and Phe.

The distribution of residues observed in protein structures has been used to develop "empirical hydrophobicity scales." In one of these, a residue's hydrophobicity is related to its average distance from the protein center of mass and the average orientation of its side chain (137–140). In another, the average surroundings of residue types are used to characterize their hydrophobicity (141–143).

A "hydropathy scale" has been developed (143a) in which the value for

Table 4 Residue surface areas and hydrophobicities and the composition of buried and accessible surfaces in proteins

			Hydrophobicity[c] (kcal · mol^{-1}) relative to Gly		Molar fraction (%)		
Residue	Accessible surface area (Å^2)[a]	Proportion buried in proteins[b] (%)	Transfer free energy[d]	Hydration free energy[e]	Average protein[f]	Accessible residues[g]	Buried residues[g]
Gly	75	36	0.00	0.00	8.4	6.7	11.8
Ala	115	38	0.5	0.45	8.6	6.2	11.2
Val	155	54	1.5	0.40	6.6	4.5	12.9
Ile	175	60	--	0.24	4.5	2.8	8.6
Leu	170	45	1.8	0.11	7.4	4.8	11.7
Phe	210	50	2.5	3.15	3.6	2.4	5.1
Pro	145	18			5.2	4.8	2.7
Ser	115	22	−0.3	7.45	7.0	9.4	8.0
Thr	140	23	0.4	7.27	6.1	7.0	4.9
Cys	135	48		3.63	2.9	0.9	4.1
Met	185	40	1.3	3.87	1.7	1.0	1.9
Tyr	230	15	2.3	8.50	3.4	5.1	2.6
Trp	255	27	3.4	8.27	1.3	1.4	2.2
His	195	17	0.5	12.66	2.0	2.5	2.0
Asn	160	12	−0.2	12.08	4.3	6.7	2.9
Gln	180	7	−0.3	12.08	3.9	5.2	1.6
Asp	150	15	—	13.34	5.5	7.7	2.9
Glu	190	18	—	12.59	6.0	5.7	1.8
Lys	200	3		11.91	6.6	10.3	0.5
Arg	225	1	—	22.31	4.9	4.5	0.5

[a] Value for the residue R in the extended tripeptide Gly–R–Gly (133).
[b] Residues were taken to be buried if they had 5% or less of their potential surface accessible to solvent (133).
[c] Hydrophobic is a misnomer: There is no hydrophobia between water and hydrocarbons; there is just not enough hydrophilia to pry the water molecules apart (168).
[d] Measured for transfer from water to ethanol or dioxane (169).
[e] Measured for transfer from water to vapor (136).
[f] From (170).
[g] From (135). Here buried residues were those with an accessible surface area of less than 20 Å^2.

each residue is based upon both the average extent to which it is buried and on its energy of transfer from water to vapor. The mean hydropathy of proteins does not change with molecular weight; large proteins are not more or less hydrophobic than small.

The distribution of residues in protein sequences has been surveyed (144, 144a), as have the interactions between residues within proteins (145, 146). Certain distributions and interactions deviate from random expectations. The reason for this is not understood.

Secondary-Structure Surfaces and Residue Composition

One calculation has suggested that the observed probabilities of residues being found in α helices and β sheets is proportional to the surface that becomes buried when they are incorporated into these secondary structures (147). The calculation carried out is the following: Different residues, X, in an extended chain, Ala_3–X–Ala_3, were incorporated into an α helix, Ala_4–X–Ala_4, and into a β sheet, Ala_7/Ala_3–X–Ala_3/Ala_7. For each residue, X, the loss of contact surface that occurs in these two processes was determined. Residues were then ordered on the size of the area losses to give two lists, one for the chain to α-helix process and one for the chain to β-sheet process. The two lists were then compared to lists of the probability of the residues being found in α helices or β sheets (148). For β sheets the rank correlation coefficient between the two lists is 0.74 (Pro excluded); for α helices it is 0.65 (Pro, Tyr, and Glu excluded) (147).

All residues occur to some extent in α helices, β sheets, and turns (148, 148a). This means that prediction of amino acids based just on the residue composition of regions of polypeptide can only have limited success (144a). The best of such schemes predict correctly the secondary structure of not more than 56% of residues (148b). More accurate results are obtained with those schemes that include consideration of both local and long-range interactions (148c, 148d).

When an average residue goes from a chain in an extended conformation to an α helix or two stranded β sheets it buries about half its potential accessible surface ($\sim$80 Å^2). If the β sheet is extended by additional strands the inner residues bury another quarter (133). The effect on individual residue surfaces of secondary-structure formation and packing has been described in some detail (84, 87, 88).

Chemical Character of Buried and Accessible Surfaces

Of the total surface buried within proteins, about three fifths is buried within secondary structures and two fifths between them (88, 133, 149). The formation of hydrogen bonds between peptides makes half the surface buried *within* secondary structures polar (133). The surfaces buried *between*

secondary structures are very hydrophobic (1, 73, 133). Quantitative expressions of this are given by the calculation of their hydropathy (143a) or hydrophobic moment (149a). Two thirds of the surface buried between secondary are formed by nonpolar atoms and more than half of the polar part is formed by groups that hydrogen bond within their own secondary structure (133). Thus, the net general effect is that, on folding, both polar and nonpolar surfaces are reduced by similar amounts, approximately three quarters (73, 118, 133). On a more detailed level, this general statement has one important qualification: As proteins increase in size they bury an increasing proportion of their surface (see next section). This relative increase in buried surface occurs at the expense of just the protein's nonpolar surface (133).

The polar surface buried in proteins is very largely formed by peptide groups while the nonpolar part is mainly formed by side chains (73, 118, 133). However, a large proportion of nonpolar side chains remains on the surface (Table 4). The role of a particular nonpolar side chain in protein folding depends upon which other side chains are in its vicinity on the formation of the secondary structures. If the neighbors are also nonpolar, the residue is usually part of the surfaces buried within the protein; if they are polar it is usually on the surface, whatever its intrinsic hydrophobicity.

Extent of the Accessible and Buried Surfaces

The accessible surface area, A_s, of small and medium size monomeric protein (50–320 residues) is a function of their molecular weight, M_r (70). The relationship can be expressed by Equation 1 (150, 151):

$$A_s = 11.1\, M_r^{2/3}. \qquad 1.$$

For all but a few cases this equation gives A_s values to within a few per cent of those observed. No data for monomeric proteins with more than 320 residues have been published.

The implications of Equation 1 for the shape of those proteins to which it applies have been discussed (152). The ratio of the protein's actual surface area, A_s, to the surface area of a smooth ellipsoid of the same volume increases with molecular weight. For small proteins ($M_r \simeq 6{,}000$) the ratio is 1.4 : 1; for larger proteins ($M_r \simeq 35{,}000$) it is 1.65 : 1. The greater surface area of real proteins is, of course, due to the textured nature of their surface. The actual ratios have two interesting implications (152): First, they imply that larger proteins must be more highly textured or more aspherical than smaller proteins, as had been predicted much earlier (153). Second, the ratios are large enough to allow a great variety of protein shapes. Spheres, spheres with large clefts, and ellipsoids or rods or discs with axial ratios of 1 to 2, when they have the same volume, differ in surface area by less than

10%. This analysis (152) disproves previous suggestions that proteins have surface area/volume ratios that approach minimum values characteristic of a sphere.

For oligomeric proteins the surface area of isolated monomers with 50–250 residues is given quite accurately by Equation 1 (70, 150, 151). For oligomers whose monomers have 330–840 residues, this is not the case. They have surface areas that are approximately proportional to molecular weight, and are 20%–50% greater than the values given by equation 1 (70, 135, 154).

The accessible surface area of protein chains in an extended conformation, A_t, is quite accurately given by Equation 2 (70, 151):

$$A_t = 1.45\, M_r. \qquad 2.$$

Thus, for those proteins whose accessible surface area is given by Equation 1, the potential surface buried by the protein's folding, A_B, will be approximately given by Equation 3:

$$A_B = 1.45\, M_r - 11.1\, M_r^{2/3}. \qquad 3.$$

This means that small proteins, ~50 residues, will bury ~90 Å^2 for each residue and that large proteins, ~320 residues, will bury ~120 Å^2. As just noted, Equation 1 does not apply to large oligomeric proteins but the values so far published for their A_s values (70, 120, 154) show that they bury a constant proportion of their surface that is also equivalent to ~120 Å^2 per residue.

The extent of the buried (and therefore accessible) surfaces in proteins has been rationalized on the basis of the following argument: The area buried on the folding of a protein is directly related to the hydrophobic energy that helps compensate for the loss of conformational entropy (70, 155). The loss of entropy by the main chain will be proportional to its length, i.e. its molecular weight. The loss by the side chains will be proportional to the number buried. For monomeric and small oligomeric proteins, this increases with molecular weight. The number of buried residues, n_B, is related to the number in a protein, n, by the equation $n_B = (n^{1/3} - k)^3$ (135). For large oligomeric proteins, the proteins, the proportion is constant (154) and the same as that found in monomers of 300 residues.

Surfaces Buried Between Protein Subunits

For oligomeric proteins the accessible surface area that becomes buried, when two monomers are brought together, varies between 1000 Å^2 and 5500 Å^2 (120, 154–158). The lower values, 1000 Å^2 to 2000 Å^2, are found in oligomers where the structure of the isolated and associated monomers are very similar, and it has been argued that the size of the buried surfaces can

account for the stability of the associations (156–158). The larger values are found where the monomers have their structure stabilized or changed by the association. In some cases the interface is formed by loops or chain termini that would be flexible in the isolated monomer (158). In other cases the active structure of the monomers appears only on association (159).

FUNCTIONAL DETERMINANTS OF STRUCTURE: DOMAINS IN PROTEINS

The proteins found in biological systems are, of course, present because they perform some function. Many specific examples of the relation between structure and function have been described (see any recent text book). Here I briefly mention just two general features of protein structure that are selected by function: They relate to the topology of α/β proteins and the domain structure found in many proteins.

In α/β proteins a clear relationship has been demonstrated between the general topology of the structure and the position and geometry of the active sites (160). α/β Proteins have a β sheet with its strands connected by α helices that pack onto the faces of the β sheet (37). In all these proteins, substrates and cofactors bind in crevices formed by the connections two adjacent β-sheet strands make to α helices on *opposite* sides of the β sheet (Figure 14). The occurrence and position of such crevices are determined by

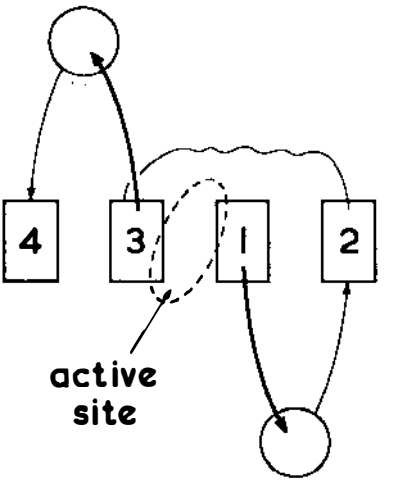


Figure 14 The active site in α/β proteins is formed at a position where the direction of the strand order changes (160): between strands 1 and 3 in the example shown here. The twist of the β sheet, not shown here, increases the crevice nature of the space between the loops that come from strands 1 and 3.

the right-hand nature of the $\beta\alpha\beta$ units and the strand order of the β sheet. They can occur only where the direction of the strand order changes: between strands 3 and 1 in the β sheet with strand order 4312 in Figure 14. Thus, the requirement to form an active site discriminates between the arangements of $\beta\alpha\beta$ units that are possible on thermodynamic and kinetic grounds (160).

Many proteins have a domain structure—globular units separated by a cleft. [The definitions of domains and the methods for their location have recently been reviewed (161).] It has been suggested that domains have a structural role in proteins in facilitating folding (94). In large proteins, at least, experimental evidence had previously shown the independent refolding of fragments [(162); more recent work is reviewed in (163)].

Domains are clearly of central importance to protein function: Many active sites occur at domain interfaces; in enzymes with more than one substrate the different binding sites usually occur on different domains; in proteins with more than one function different domains usually perform the different functions (161). In certain proteins the domain structure seems to have been created by gene duplication or gene fusion. Gene duplication is indicated by a protein's domains being similar in structure. This gives the molecule pseudo-symmetry (164, 165). Gene fusion is indicated by domains with the same function and similar structure being found in different proteins in association with other domains of quite different structure (166).

CONCLUSION

The seeming complexity of the first protein structures created shock of the kind expressed in the contemporary description of myoglobin (171) as "almost nothing but a complicated set of rods [of polypeptide] sometimes going straight for a distance then turning a corner and going off in a new direction . . . much more complicated and irregular than most of the early theories of the structure of proteins had suggested." The work described here has shown that the apparently complex structure of proteins is in fact governed by a set of relatively simple principles. Individual proteins arise from particular combinations of and variations on these principles. An analogous situation is found in linguistics, where a set of simple grammatical rules govern the generation of different, and sometimes complex, sentences.

Although the principles described here do not, in most cases, predict exactly what will occur for a particular protein, they do strongly predict the sorts of things that are to be expected and suggest how exceptions may arise. As more protein structures become known, the principles will be refined and extended.

Much remains to be done. The detailed energetics of protein structure are

not clearly understood. The relation between the folding process and the final structure is, at this moment, mostly speculative. The recent achievements, however, together with those of the 1950s and the 1960s, have laid the foundations of a comprehensive theory for protein structure.

ACKNOWLEDGMENTS

I thank John Cresswell for drawing the figures and Arthur Lesk and Joël Janin for their comments on the manuscript. The author's work described in this review has been supported by the Royal Society, National Institute of General Medical Sciences, Science Research Council, and Medical Research Council.

Literature Cited

1. Perutz, M. F., Kendrew, J. C., Watson, H. C. 1965. *J. Mol. Biol.* 13:669–78
2. Schulz, G. E., Schirmer, R. H. 1979. *Principles of Protein Structure.* New York: Springer-Verlag. 314 pp.
3. Richards, F. M. 1977. *Ann. Rev. Biophys. Bioeng.* 6:151–76
4. Janin, J. 1979. *Bull. Inst. Pasteur* 77:337–73
5. Rossmann, M. G., Argos, P. 1981. *Ann. Rev. Biochem.* 50:497–532
6. Richardson, J. S. 1981. *Adv. Prot. Chem.* 34:167–339
7. Pauling, L., Corey, R. B., Branson, H. R. 1951. *Proc. Natl. Acad. Sci. USA* 37:205–11
8. Pauling, L., Corey, R. B. 1951. *Proc. Natl. Acad. Sci. USA* 37:729–40
9. Ramachandran, G. N., Sasisekharan, V. 1968. *Adv. Prot. Chem.* 23:283–437
10. Moews, P. C., Kretsinger, R. H. 1975. *J. Mol. Biol.* 91:201–28
11. Sawyer, L., Shotton, D. M., Campbell, J. W., Wendell, P. L., Muirhead, H., et al. 1978. *J. Mol. Biol.* 118:137–208
12. Baker, E. N. 1980. *J. Mol. Biol.* 141:441–84
13. Matsuura, Y., Takano, T., Dickerson, R. E. 1982. *J. Mol. Biol.* 156:389–409
14. Remington, S. J., Wiegand, G., Huber, R. 1982. *J. Mol. Biol.* 158:111–52
15. Holmes, M. A., Matthews, B. W. 1982. *J. Mol. Biol.* 160:623–39
16. Steigemann, W., Weber, E. 1979. *J. Mol. Biol.* 122:309–38
17. James, M. N. G., Sielecki, A. R. 1983. *J. Mol. Biol.* 163:299–361
18. Watson, H. C. 1969. *Prog. Sterochem.* 4:229–333
19. Artymiuk, P. J., Blake, C. C. F. 1981. *J. Mol. Biol.* 152:737–62
20. Deisenhofer, J., Steigemann, W. 1975. *Acta Crystallogr. Sect. B* 31:238–50
21. Takano, T., Dickerson, R. E. 1981. *J. Mol. Biol.* 153:79–94
22. Glover, I., Haneef, I., Pitts, J., Wood, S., Moss, D., et al. 1983. *Biopolymers* 22:293–304
23. Sakabe, N., Sakabe, K., Sasaki, K. 1981. *Structural Studies on Molecules of Biological Interest*, ed. G. Dodson, J. P. Glusker, D. Sayre, pp. 509–26. Oxford: Clarendon. 610 pp.
24. Bolin, J. T., Filman, D. J., Matthews, D. A., Hamlin, R. C., Kraut, J. 1982. *J. Biol. Chem.* 257:13650–72
25. Deleted in press
26. Fermi, G., Perutz, M. F. 1981. *Atlas of Molecular Structures in Biology. Vol. 2. Haemoglobin and Myoglobin*, ed. D. C. Phillips, F. R. Richards. Oxford: Clarendon. 104 pp.
27. Matthews, B. W., Weaver, L. H., Kester, W. R. 1974. *J. Biol. Chem.* 249:8030–44
28. Schellman, C. 1980. See Ref. 167, pp. 53–61
28a. Ravichandran, V., Subramanian, E. 1981. *Int. J. Peptide Prot. Res.* 18:121–26
28b. Marquart, M., Deisenhofer, J., Huber, R. 1980. *J. Mol. Biol.* 141:369–91
29. Venkatachalam, C. M. 1968. *Biopolymers* 6:1425–36
30. Lewis, P. N., Momany, F. A., Scheraga, H. A. 1973. *Biochim. Biophys. Acta* 303:211–39
31. Nemethy, G., Printz, M. P. 1972. *Macromolecules* 5:755–58
32. Crawford, J. L., Lipscomb, W. N., Schellman, C. G. 1973. *Proc. Natl. Acad. Sci. USA* 70:538–42
33. Huber, R., Steigemann, W. 1974. *FEBS Lett.* 48:235–37
34. Chou, P. Y., Fasman, G. D. 1977. *J. Mol. Biol.* 115:135–75

35. Smith, J. A., Pease, L. G. 1980. *CRC Crit. Rev. Biochem.* 8:315–99
36. Kuntz, I. D. 1972. *J. Am. Chem. Soc.* 94:4009–12
37. Levitt, M., Chothia, C. 1976. *Nature* 261:552–58
38. Rose, G. D., Young, W. B., Gierasch, L. M. 1983. *Nature* 304:655–57
39. Chothia, C. 1973. *J. Mol. Biol.* 75:295–302
40. Scott, R. A., Scheraga, H. A. 1966. *J. Chem. Phys.* 45:2091–2101
41. Chou, K. C., Pottle, M., Nemethy, G., Ueda, Y., Scheraga, H. A. 1982. *J. Mol. Biol.* 162:89–112
42. Chou, K. C., Scheraga, H. A. 1982. *Proc. Natl. Acad. Sci. USA* 79:7047–51
43. Chou, K. C., Nemethy, G., Scheraga, H. A. 1983. *J. Mol. Biol.* 168:389–407
44. Janin, J., Chothia, C. 1980. *J. Mol. Biol.* 143:95–128
45. Chothia, C., Janin, J. 1981. *Proc. Natl. Acad. Sci. USA* 78:4146–50
46. Chothia, C., Janin, J. 1982. *Biochemistry* 21:3955–65
47. Nishikawa, K., Scheraga, H. A. 1976. *Macromolecules* 9:365–407
48. Salemme, F. R., Weatherford, D. W. 1981. *J. Mol. Biol.* 146:101–17
49. Salemme, F. R., Weatherford, D. W. 1981. *J. Mol. Biol.* 146:119–41
50. Salemme, F. R. 1981. *J. Mol. Biol.* 146:143–56
51. Chothia, C. 1983. *J. Mol Biol.* 163:107–17
52. Richardson, J. S., Getzoff, E. D., Richardson, D. C. 1978. *Proc. Natl. Acad. Sci. USA* 75:2574–78
53. Chandrasekaran, R., Ramachandran, G. N. 1970. *Int. J. Prot. Res.* 2:223–33
54. Ponnuswamy, P. K., Sasisekharan, V. 1971. *Biopolymers* 10:565–82
55. Sasisekharan, V., Ponnuswamy, P. K. 1971. *Biopolymers* 10:583–92
56. Nemethy, G., Scheraga, H. A. 1977. *Quart. Rev. Biophys.* 10:239–352
57. Finkelstein, A. V. 1976. *Mol. Biol.* 10:879–86
58. Finkelstein, A. V., Ptitsyn, O. B. 1977. *Biopolymers* 16:469–95
59. Janin, J., Wodak, S., Levitt, M., Maigret, B. 1978. *J. Mol. Biol.* 125:357–86
60. Gelin, B. R., Karplus, M. 1979. *Biochemistry* 18:1256–68
61. Bhat, T. N., Sasisekharan, V., Vijayan, M. 1979. *Int. J. Peptide Prot. Res.* 13:170–84
62. Thornton, J. M. 1981. *J. Mol. Biol.* 151:261–87
63. Kauzmann, W. 1959. *Adv. Prot. Chem.* 14:1–64
64. Liquori, A. M. 1966. *Principles of Biomolecular Organization*, ed. G. E. W. Wolstenholme, M. O'Connor, pp. 40–68. London: Churchill. 491 pp.
65. Klapper, M. H. 1971. *Biochem. Biophys. Acta* 229:557–66
66. Richards, F. M. 1974. *J. Mol. Biol.* 82:1–14
67. Finney, J. L. 1975. *J. Mol Biol.* 96:721–32
68. Finney, J. L. 1978. *J. Mol. Biol.* 119:415–41
69. Gellatly, B. J., Finney, J. L. 1982. *J. Mol. Biol.* 161:305–22
70. Chothia, C. 1975. *Nature* 254:304–8
71. Kuntz, I. D., Crippen, G. M. 1979. *Int. J. Peptide Prot. Res.* 13:223–28
72. Crippen, G. M., Kuntz, I. D. 1978. *Int. J. Peptide Prot. Res.* 12:47–56
73. Lee, B., Richards, F. M. 1971. *J. Mol. Biol.* 55:379–400
74. Richards, F. M. 1979. *Carlsberg Res. Commun.* 44:47–63
75. Lumry, R., Rosenberg, A. 1975. *Collect. Int. CNRS L'eau Syst. Biol.* 246:55–63
76. McCammon, J. A., Lee, C. Y., Northrup, S. H. 1983. *J. Am. Chem. Soc.* 105:2232–37
77. Liquori, A. M., Sadun, C. 1981. *Int. J. Biol. Macromol.* 3:56–59
78. Chothia, C., Levitt, M., Richardson, D. 1977. *Proc. Natl. Acad. Sci. USA* 74:4130–34
79. Sternberg, M. J. E. 1983. *Computing in Biological Science*, ed. M. Gerson, A. Barrett, pp. 143–77. Amsterdam: Elsevier/North-Holland
80. Crick, F. H. C. 1953. *Acta Crystallogr.* 6:689–97
81. Chothia, C., Levitt, M., Richardson, D. 1981. *J. Mol. Biol.* 145:215–50
82. Efimov, A. V. 1977. *Dokl. Acad. Nauk SSSR* 235:699–702
83. Dunker, A. K., Zaleske, D. J. 1977. *Biochem. J.* 163:45–57
84. Richmond, T., Richards, F. M. 1978. *J. Mol. Biol.* 119:537–55
85. Efimov, A. V. 1979. *J. Mol. Biol.* 134:23–40
86. Chothia, C., Lesk, A. M., Dodson, G. G., Hodgkin, D. C. 1983. *Nature* 302:500–5
87. Cohen, F. E., Sternberg, M. J. E., Taylor, W. R. 1982. *J. Mol. Biol.* 156:821–62
88. Cohen, F. E., Sternberg, M. J. E., Taylor, W. R. 1981. *J. Mol. Biol.* 148:253–72
89. Sternberg, M. J. E., Thornton, J. M. 1977. *J. Mol. Biol.* 115:1–17
90. von Heijne, G., Blomberg, C. 1978. *Biopolymers* 17:2033–37
91. Toniolo, C. 1978. *Macromolecules* 11:437–38

92. Lifson, S., Sander, C. 1979. *Nature* 282:109–11
93. Lifson, S., Sander, C. 1980. *J. Mol. Biol.* 139:627–39
94. Wetlaufer, D. E. 1974. *Proc. Natl. Acad. Sci. USA* 70:697–701
95. Schulz, G. E., Schirmer, R. H. 1974. *Nature* 250:142–44
96. Rao, S. T., Rossmann, M. G. 1973. *J. Mol. Biol.* 76:241–56
97. Richardson, J. S., Richardson, D. C. Thomas, K. A., Silverton, E. W., Davies, D. R. 1976. *J. Mol. Biol.* 102:221–35
98. Richardson, J. S. 1976. *Proc. Natl. Acad. Sci. USA* 73:2619–23
99. Sternberg, M. J. E., Thornton, J. M. 1976. *J. Mol. Biol.* 105:367–82
100. Sternberg, M. J. E., Thornton, J. M. 1977. *J. Mol. Biol.* 110:269–83
101. Nagano, N. 1977. *J. Mol. Biol.* 109: 235–50
102. Richardson, J. S. 1977. *Nature* 268: 495–500
103. Finkelstein, A. V., Ptitsyn, O. B., Bendzko, P. 1979. *Biofizika* 24:21–26
104. Ptitsyn, O. B., Finkelstein, A. V. 1979. *Biofizika* 24:27–31
105. Connolly, M. L., Kuntz, I. D., Crippen, G. M. 1980. *Biopolymers* 19:1167–82
105a. Klapper, M. H., Klapper, I. Z. 1980. *Biochim. Biophys. Acta* 626:97–105
105b. Creighton, T. E. 1974. *J. Mol. Biol.* 87:603–24
106. Ptitsyn, O. B., Finkelstein, A. V. 1980. *Quart. Rev. Biophys.* 13:339–86
107. Efimov, A. V. 1982. *Mol. Biol.* 16:799–806
108. Baldwin, R. L. 1980. See Ref. 167, pp. 369–85
109. Weber, P. C., Salemme, F. R. 1980. *Nature* 287:82–84
110. Bello, J. 1977. *J. Theor. Biol.* 68:139–42
111. Bello, J. 1978. *Int. J. Peptide Prot. Res.* 12:38–41
112. Schellman, J. A. 1955. *C.R. Trav. Lab. Carlsberg Ser. Chim.* 29:223–29
113. Klotz, I. M., Franzen, J. S. 1962. *J. Am. Chem. Soc.* 84:3461–66
114. Susi, H., Timasheff, S. N., Ard, J. S. 1964. *J. Biol Chem.* 239:3051–54
115. Klotz, I. M., Farnham, S. B. 1968. *Biochemistry* 7:3879–82
116. Kresheck, G. C., Klotz, I. M. 1969. *Biochemistry* 8:8–12
117. Creighton, T. E. 1983. *Biopolymers* 22:49–58
118. Shrake, A., Rupley, J. A. 1973. *J. Mol. Biol.* 79:351–71
119. Greer, J., Bush, B. L. 1978. *Proc. Natl. Acad. Sci. USA* 75:303–7
120. Wodak, S. J., Janin, J. 1980. *Proc. Natl. Acad. Sci. USA* 77:1736–40
121. Chothia, C. 1974. *Nature* 248:338–39
122. Gelles, J., Klapper, M. H. 1978. *Biochim. Biophys. Acta* 533:465–77
123. Miller, K. W., Hildebrand, J. H. 1968. *J. Am. Chem. Soc.* 90:3001–4
124. Sinanoglu, O. 1968. *Molecular Associations in Biology*, ed. B. Pullman, pp. 427–45. New York: Academic
125. Hermann, R. B. 1972. *J. Phys. Chem.* 76:2754–59
126. Harris, M. J., Higuchi, T., Rutting, J. H. 1973. *J. Phys. Chem.* 77:2694–2703
127. Reynolds, J. A., Gilbert, D. B., Tanford, C. 1974. *Proc. Natl. Acad. Sci. USA* 71:2925–27
128. Hermann, R. B. 1977. *Proc. Natl. Acad. Sci. USA* 74:4144–45
129. Nandi, P. 1976. *Int. J. Peptide Prot. Res.* 8:253–64
130. Creighton, T. E. 1979. *J. Mol. Biol.* 129:235–64
131. Tanford, C. 1979. *Proc. Natl. Acad. Sci. USA* 76:4175–76
132. Karplus, M. 1980. *Biophys. J.* 32:45–46
133. Chothia, C. 1976. *J. Mol. Biol.* 105: 1–14
134. Manavalan, P., Ponnuswamy, P. K. 1977. *Biochem. J.* 167:171–82
135. Janin, J. 1979. *Nature* 277:491–92
136. Wolfenden, R., Anderson, L., Cullis, P. M., Southgate, C. C. B. 1981. *Biochemistry* 20:849–55
137. Rackovsky, S., Scheraga, H. A. 1977. *Proc. Natl. Acad. Sci. USA* 74:5248–51
138. Wertz, D. H., Scheraga, H. A. 1978. *Macromolecules* 11:9–15
139. Meirovitch, H., Rackovsky, S., Scheraga, H. A. 1980. *Macromolecules* 13:1398–1405
140. Meirovitch, H., Scheraga, H. A. 1980. *Macromolecules* 13:1406–14
141. Manavalan, P., Ponnuswamy, P. K. 1977. *Arch. Biochem. Biophys.* 184:476–87
142. Manavalan, P., Ponnuswamy, P. K. 1978. *Nature* 275:673–74
143. Ponnuswamy, P. K., Prabhakaran, M., Manavalan, P. 1980. *Biochim. Biophys. Acta* 623:301–16
143a. Kyte, J., Doolittle, R. F. 1982. *J. Mol. Biol.* 157:105–32
144. Black, J. A., Harkins, R. N., Stenzel, P. 1976. *Int. J. Peptide Prot. Res.* 8:125–30
144a. Klapper, M. H. 1977. *Biochem. Biophys. Res. Commun.* 78:1018–24
145. Warme, P. K., Morgan, R. S. 1978. *J. Mol. Biol.* 118:273–87
146. Warme, P. K., Morgan, R. S. 1978. *J. Mol. Biol.* 118:289–304
147. Richards, F. M., Richmond, T. 1978. *Molecular Interactions and Activity in Proteins*, pp. 23–45. Amsterdam: Excerpta Medica. 279 pp.

148. Chou, P. Y., Fasman, G. D. 1978. *Adv. Enzymol.* 47:45–148
148a. Levitt, M. 1978. *Biochemistry* 17: 4277–85
148b. Kabsch, W., Sander, C. 1983. *FEBS Lett.* 151:179–82
148c. Ptitsyn, O. B., Finkelstein, A. V. 1983. *Biopolymers* 22:15–25
148d. Taylor, W. R., Thornton, J. M. 1983. *Nature* 301:540–42
149. Lesk, A. M., Chothia, C. 1980. *Biophys. J.* 32:35–47
149a. Eisenberg, D., Weiss, R. M., Terwilliger, T. C. 1982. *Nature* 299:371–74
150. Janin, J. 1976. *J. Mol. Biol.* 105:13–14
151. Teller, D. C. 1976. *Nature* 260:729–31
152. Gates, R. E. 1979. *J. Mol. Biol.* 127:345–51
153. Fisher, H. F. 1964. *Proc. Natl. Acad. Sci. USA* 51:1285–91
154. Sprang, S., Yang, D., Fletterick, R. J. 1979. *Nature* 280:333–35
155. Janin, J., Chothia, C. 1979. *FEBS: 12th Meet. Dresden 1978*, ed. E. Hofmann, W. Pfeil, H. Aurich, 52:227–37. Oxford: Pergamon. 522 pp.
156. Chothia, C., Janin, J. 1975. *Nature* 256:705–8
157. Janin, J., Chothia, C. 1976. *J. Mol. Biol.* 100:197–221
158. Chothia, C., Wodak, S., Janin, J. 1976. *Proc. Natl. Acad. Sci. USA* 73:3793–97
159. Jaenicke, R., Rudolph, R. 1980. See Ref. 167, pp. 525–46
160. Bränden, C. I. 1980. *Quart. Rev. Biophys.* 13:317–38
161. Janin, J. Wodak, S. J. 1983. *Prog. Biophys. Mol. Biol.* 42:21–78
162. Goldberg, M. E. 1969. *J. Mol. Biol.* 46:441–46
163. Wetlaufer, D. B. 1981. *Adv. Prot. Chem.* 34:61–92
164. McLachlan, A. D. 1979. *J. Mol. Biol.* 128:49–79
165. McLachlan, A. D. 1980. See Ref. 167, pp. 79–99
166. Rossmann, M. G., Liljas, A., Bränden, C. I., Banaszak, L. J. 1975. *Enzymes* 11:61–102
167. Jaenicke, R., ed. 1980. *Protein Folding.* Amsterdam: Elsevier/North-Holland. 587 pp.
168. Hildebrand, J. H. 1979. *Proc. Natl. Acad. Sci. USA* 76:1
169. Nozaki, Y., Tanford, C. 1971. *J. Biol. Chem.* 246:2211–17
170. Dayhoff, M. D. 1976. *Atlas of Protein Sequence and Structure*, Vol. 5, Suppl. 2. Washington DC: Natl. Biomed. Res. Found.
171. Kendrew, J. C. 1961. *Sci. Am.* 205(6): 96–110

Annual Reviews of Biochemistry
Volume 53, 1984

CONTENTS

[1] As noted in the preface (p. vi) to Volume 52 (1983), typesetting problems resulted in the omission of an author index from that volume. That author index appears as an extra index here in Volume 53.