Annu. Rev. Biophys. Biomol. Struct. 1997. 26:597–627

PROTEIN FOLDS IN THE ALL-β AND ALL-α CLASSES

Cyrus Chothia, Tim Hubard, Steven Brenner, Hugh Barns, and Alexey Murzin

MRC Laboratory of Molecular Biology and Cambridge Centre for Protein Engineering, Hills Road, Cambridge CB2 2QH, United Kingdom

KEY WORDS: β-sandwiches, β-propellers, β-helices, β-barrels, β-prisms, α-helix polyhedra, α-helix bundles, α-helix layer structures

ABSTRACT

Analysis of the structures in the Protein Databank, released in June 1996, shows that the number of different protein folds, i.e. the number of different arrangements of major secondary structures and/or chain topologies, is 327. Of these folds, approximately 25% belong to the all-α class, 20% belong to the all-β class, 30% belong to the α/β class, and 25% belong to the $\alpha + \beta$ class.

We describe the types of folds now known for the all-β and all-α classes, emphasizing those that have been discovered recently. Detailed theories for the physical determinants of the structures of most of these folds now exist, and these are reviewed.

CONTENTS

1056-8700/97/0610-0597$08.00

INTRODUCTION

The fold of a protein is defined by the arrangement of its major elements of secondary structures and by the topology of the connections between them. Examination of the first known protein structures showed that these features have clear regularities (52, 81, 89). In most proteins the α-helices and β-sheets pack together in one of a small number of different ways. The connections between secondary structures obey a set of empirical topological rules in almost all cases. This means that it is common for sets of unrelated proteins to have similar, if not identical, folds. Subsequently, it was argued that these similarities arise from the intrinsic physical and chemical properties of proteins (15, 78), and a great deal of work was carried out to demonstrate that this is the case (reviewed in 11, 12, 26, 82).

Recently, among the many new protein structures that have been determined, some have been found to have folds that are different from any seen previously. Also, new theories have been developed that treat not only a particular type of packing or particular topological features, but whole assemblies of secondary structures.

We describe here this recent growth in our knowledge of protein structures. We give an overview of the folds now known for proteins built of β-sheets (the all-β class) and proteins built of α-helices (the all-α class). We also discuss what is known about the physical and chemical determinants of the structure of these folds. Although Richardson (82) described the fold of each protein structure known at the time of her publication in 1981, the large number of structures that are currently known and space limitations here prevent us from doing the same. We do, however, attempt to describe all the types of folds that are now known to occur in two or more unrelated proteins. Previous surveys of folds have been published by Orengo and colleagues (71), by Holm & Sander (33), and by Yee & Dill (102), and methods for classifying protein folds have recently been reviewed (34, 70).

THE KNOWN PROTEIN FOLDS

The Growth in Structural Information The first protein structure—that of sperm whale myoglobin—was solved in 1960 (41), but for decades thereafter, the number of new protein structure determinations increased relatively slowly. The Protein Databank (PDB) was founded in 1971 (5); Figure 1 traces the growth in the number of structures it holds. Ten years ago (1986), the PDB contained 235 protein domains, only 5% of the present number; during the past five years it has grown fourfold. The quantity of structural information about proteins has increased enormously within the past few years, but, if viewed in

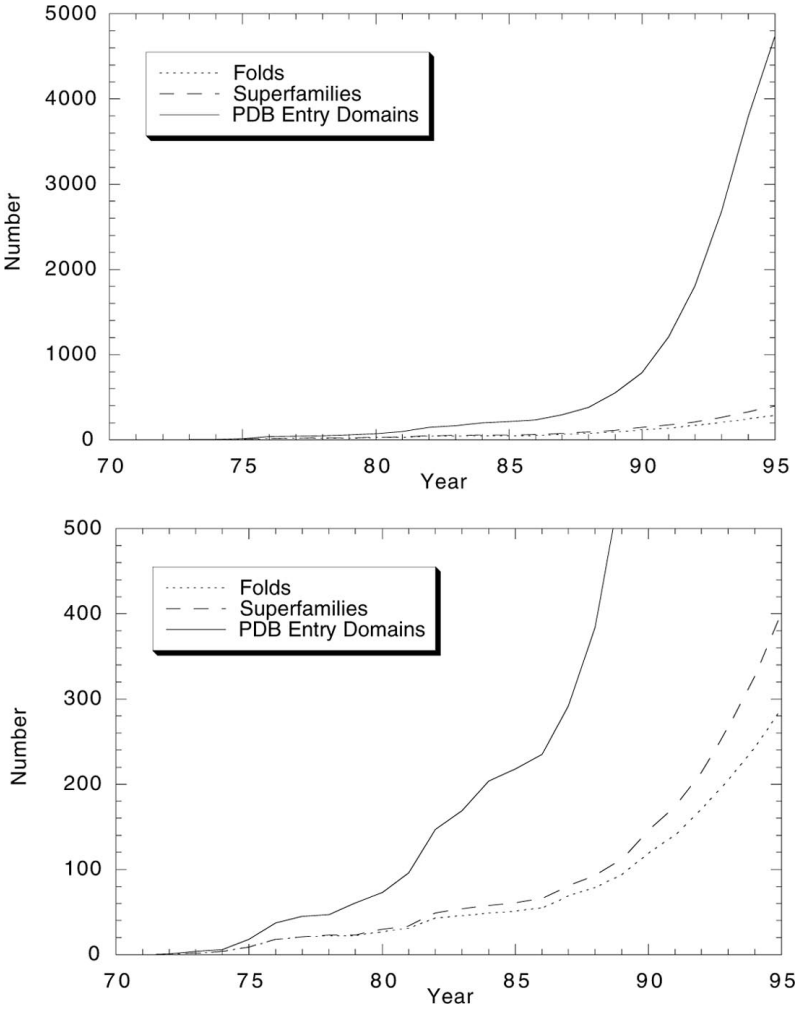


Figure 1 The growth in the data available on the atomic structures of proteins. Large proteins are often built of separate domains and it is useful to treat these individually. All small proteins and most proteins of medium size have a single domain; (*top*) The cumulative increase in the number of domains in the Protein Structure Data Bank (5) from its foundation to the present. Domains that have a common evolutionary origin can be clustered into superfamilies; those superfamilies with the same basic structure can be clustered into folds (see text). (*bottom*) The cumulative increase in the number of superfamilies and folds.

terms of protein folds, its "redundancy" has also increased: Entries involving variants of previously known structures have been increasing far more rapidly than those of entirely new structures (Figure 1).

This is demonstrated clearly by an analysis of the PDB entries for 1994, the most recent year for which complete data is available. The 773 protein structures that entered the PDB in that year contain 1112 domains. Of these, 70% were variants of a previously known protein structure and an additional 21% showed a clear similarity in sequence to a known protein structure. It was only the proteins that contained the remaining 103 domains that showed no clear sequence similarity to a known structure and, therefore, had the potential to be a new fold. Examination of these 103 structures showed, in fact, that only 38 had a fold different from any seen previously. The other 65 had a fold similar to one seen previously, and of these, 42 turned out to also be distant evolutionary relatives of a previously known structure.

Classification of Protein Structures and the Number of Folds

The Structural Classification of Proteins database (SCOP) provides a detailed and comprehensive description of the structural and evolutionary relationships of proteins whose three dimensional structures have been determined (61). It includes all proteins in the current version of the PDB and also many proteins whose structures have been published but whose coordinates are not available from the PDB. Information about the structural relationships of proteins is also available from three other databases: the Class Architecture Topology Homology database (CATH) (72); the Fold classification based on Structure-Structure alignment of Proteins (FSSP) database (35); and the Entrez database (69).

The unit of classification used by SCOP is the protein domain. Small proteins—and most proteins of medium size—have a single domain. The domains in large proteins are usually treated separately. Domains are classified on hierarchical levels that embody their evolutionary and structural relationships. These levels include:

1. Families: Proteins are clustered into families when evolutionary relationships are demonstrated by significant sequence similarities, or by very close similarities in both structure and function.

2. Superfamilies: Families are grouped into superfamilies when their proteins have low sequence identities, but their structural details and, in many cases, functional features suggest that a common evolutionary origin is highly probable.

3. Folds: Superfamilies and families are defined as having the same fold if their proteins contain the same major secondary structures in the same arrangement and with the same topological connections. Beyond this common

core, proteins with the same fold may have peripheral elements of secondary structure and turn regions that differ in size and conformation. For proteins placed together in the same fold category, it is likely, in most cases, that their structural similarities arise not from evolutionary relationships but from their physics and chemistry, which favor certain packing arrangements and chain topologies.

In June 1996 the PDB contained 4494 entries, of which 3978 were proteins. (The remainder were mostly fragments of proteins, nucleic acids, theoretical models, or carbohydrates.) These 3978 entries have, in all, 7673 domains. Elimination of repeat entries, the same proteins with different ligands or mutations, and multiple nuclear magnetic resonance (NMR) versions reduces this to 1192 different domains. Roughly half of these domains are represented in the PDB by a single entry and 90% are represented by fewer than 10 entries. However, a small number of proteins each have an extraordinary number of entries. For example, there were 212 sets of coordinates of T4 lysozyme in different crystal forms or with engineered mutations.

The protein domains cluster into 652 protein families. Of these, 285 (43%) contain features that indicate a distant evolutionary relationship to one or more of the other families, and they can be clustered into 96 true superfamilies. The remaining 367 families can be regarded as potential superfamilies, i.e. superfamilies that, at the moment, contain just one family. Thus the June 1996 PDB contained a total of 463 true or potential superfamilies.

Of these 463 superfamilies, 176 have a fold that is the same as that found in one or more of the other superfamilies, i.e. they have the same major secondary structures in the same arrangement and with the same topological connections. These 176 superfamilies cluster into 40 different folds, which, together with the 287 superfamilies that have unique folds, means that each of the protein domains in the June 1996 PDB had one of 327 different protein folds. [A calculation somewhat similar to this has been described previously (73) for a 1994 release of PDB.] Of these 327 different folds, approximately 25% belong to the all-α class, 20% belong to the all-β class, 30% belong to the α/β class, and 25% belong to the $\alpha + \beta$ class.

In the following sections we give an overview of the folds found in the all-β and the all-α classes, emphasizing new folds and recent theoretical advances in understanding the determinants of folds.

FOLDS IN THE ALL-β CLASS

β-Sheets can twist, coil, and bend. This allows them to pack together in a number of very different ways. This intrinsic flexibility of β-sheets is much greater than that of α-helices. It derives (*a*) from the hydrogen bonds imposing

essentially two-dimensional constraints rather than the three-dimensional constraints that occur in helices, and (*b*) from the residues in the β-conformations having fewer steric constraints on their main-chain torsion angles, ϕ and ψ, than do residues in an α-conformation.

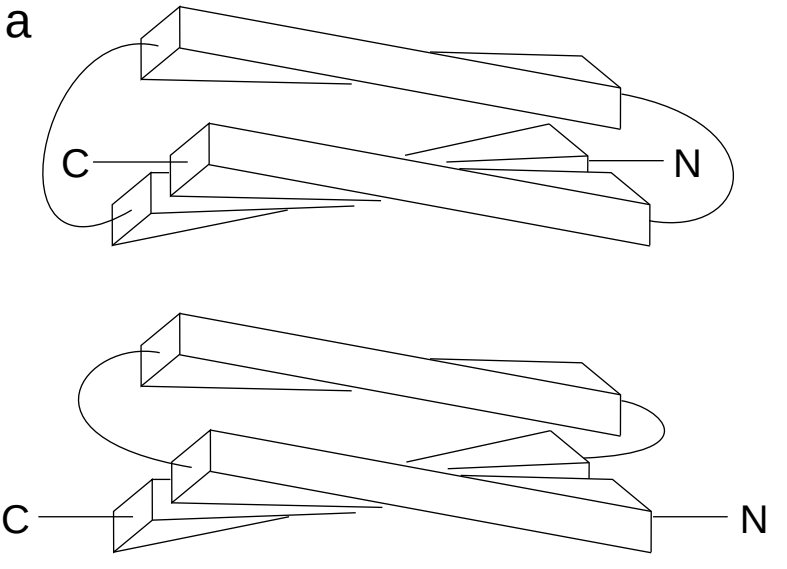
Chain Topology in All-β Proteins

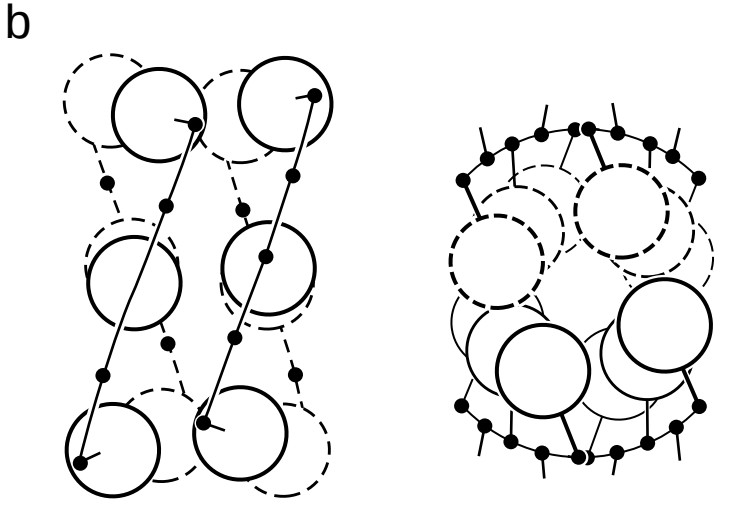
The connections between the strands in all-β proteins are largely governed by three empirical rules of topology: (*a*) strands adjacent in the sequence are often in contact in three dimensions; (*b*) in $\beta1$-$\beta2$-$\beta3$ units (where $\beta1$, $\beta2$ and $\beta3$ are three sequentially adjacent strands with $\beta1$ and $\beta3$ part of one β-sheet and $\beta2$ part of a second β-sheet), the connections between the strands are right-handed (Figure 2); and (*c*) connections neither cross each other nor do they make knots in the chain. The original work on these rules and their implications have been reviewed previously (11, 12, 26, 82). The combination of these rules make certain chain topologies common (10, 23, 24, 100). Thus, for two β-sheets, each with three anti-parallel strands, packed face to face, the rules allow for 24 different topologies; so far 6 have been observed (102).

β-Sandwiches

The simplest all-β fold is formed by two essentially independent β-sheets packing face to face to form a sandwich structure. The principles that govern this fold were described some time ago (13, 18) and are illustrated in Figure 2. The two β-sheets are usually twisted, and they pack with their interior side-chains aligned. As a result of these two features, the strand directions of the

Figure 2 (*top*) The handedness of $\beta1$-$\beta2$-$\beta3$ units. The three β-sheet strands are represented by thick ribbons and the connections between them by lines. Two $\beta1$-$\beta2$-$\beta3$ units are shown. $\beta1$ and $\beta3$ are to be seen as part of one β-sheet and $\beta2$ as part of a second β-sheet. The strands $\beta1$ and $\beta3$ are tilted relative to each other by the twist of the β-sheet. In the upper drawing the connections between the strands make the $\beta1$-$\beta2$-$\beta3$ unit left handed: if the eye of the viewer, starting from the closest end of the chain, follows it to the furthest end, it moves in a left-handed direction. In the lower drawing the connections between the strands make the $\beta1$-$\beta2$-$\beta3$ unit right handed: if the eye of the viewer, starting from the closest end of the chain, follows it to the furthest end, it moves in a right-handed direction. In right handed $\beta1$-$\beta1$-$\beta3$ units the connections between the strands bend through smaller angles than do connections in left handed units, and this makes them easier to form [see (26) and the text]. (*bottom*) Aligned packing of β-sheets in β-sandwich folds (13). Two orthogonal views of an ideal model for the packing in β-sandwich folds are shown. In the schematic drawings of two β-sheets, each with two strands, small filled circles represent the residue main-chain atoms; large open circles represent the side-chain atoms that point into the interior of the structure; side chains that point outward are not represented. The two sheets are packed so that the rows of the side chains that face each other in the interior are aligned. This maximizes the buried surface. Given this aligned arrangement of side chains, any twist in the β sheets rotates the main-chain–direction of the near-sheet relative to that in the far-β-sheet. The normal twist found in β-sheets gives a rotation angle of about -30° (See Figure 3 below). If there were no twist, the strands in the two β-sheets would be parallel to each other.


a
C
N
C
N


b

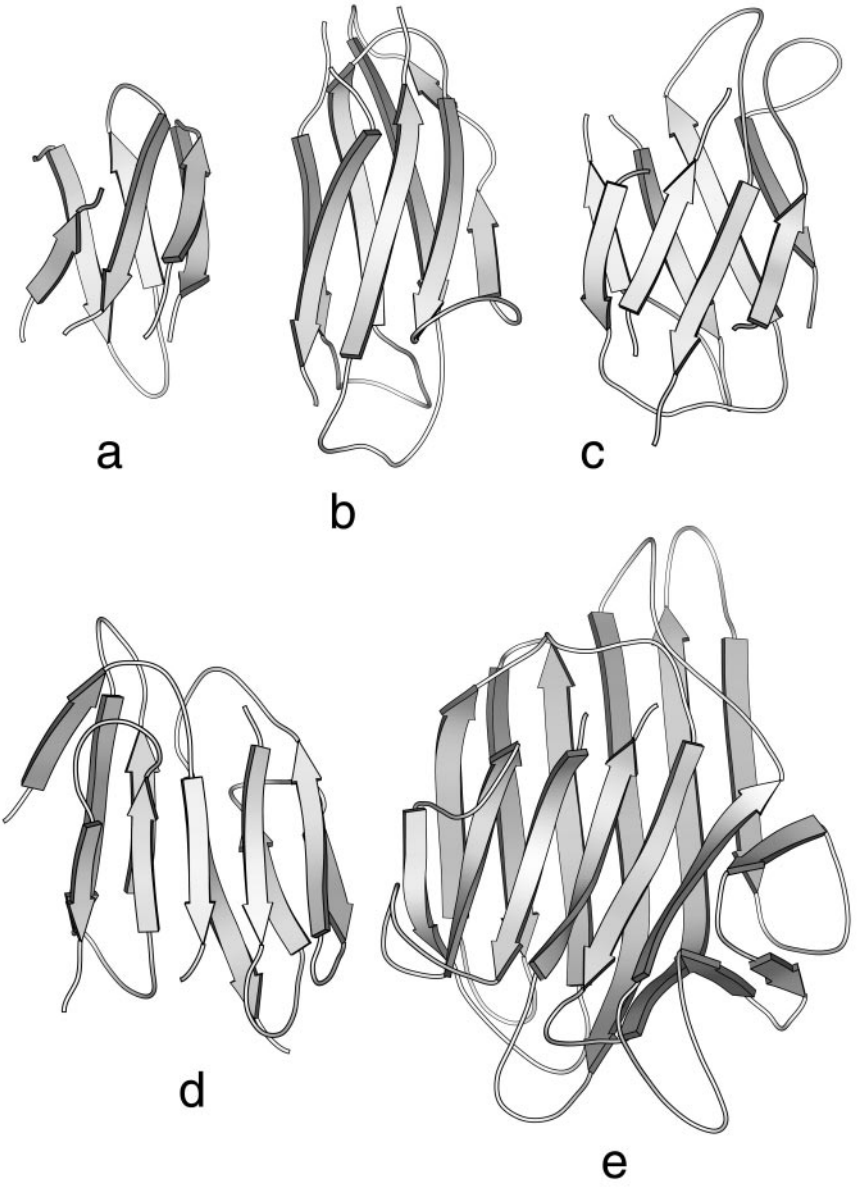


Figure 3 β-Sandwich folds: (*a*) amylase inhibitor (76); (*b*) the constant domain of T-cell receptor (4); (*c*) peptide:N-glycosidase F (45); (*d*) thaumatin (44); and (*e*) concanavalin A (30). The folds in these structures are well described by the aligned packing model in Figure 2 (13). In each case the folds have two twisted β-sheets packed face to face, with an angle of about -30° between the strand directions. The structures are built of between 6 and 13 strands.

two β-sheets are aligned at an angle of about -30° (Figure 2). Although this arrangement is found in most sandwich structures, it has been shown that misalignment of the interior side chains, small amounts of twist, or combinations of β-bulges can produce orientations with different angles that in a few cases can be small and positive (7, 57).

β-Propellers

In β-propeller folds, six, seven, or eight β-sheets, each with four anti-parallel strands, have a twist and a radial arrangement that has the appearance of a propeller (38, 95, 99) (Figure 4). An ideal theoretical model for these folds

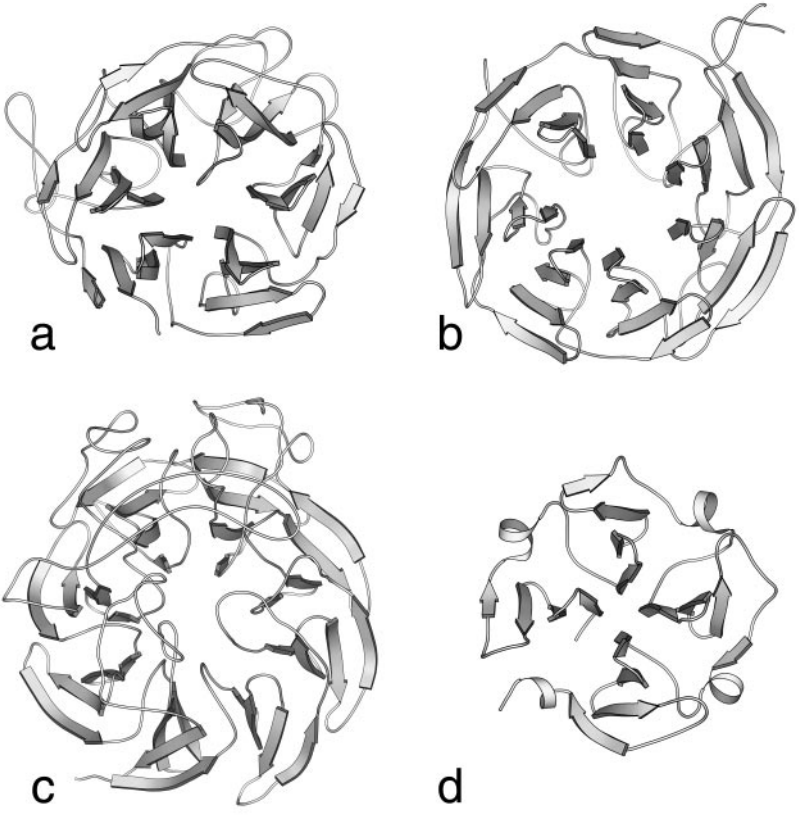


Figure 4 β-Propeller folds: (*a*) The six-fold propeller in sialidase (20), (*b*) the seven-fold propeller in galactose oxidase (38), (*c*) eight-fold propeller in methanol dehydrogenase (99). All have their β-sheets packed face to face, as β-sandwiches do. In contrast, the four-fold β-sheet propeller in the C-terminal domain of collagenase (*d*) (53) forms a closed structure by having residues in small helices that are packed between the β-sheets.

was made, assuming (*a*) the β-sheets have aligned packing like that found in β-sandwiches and (*b*) the fold is closed in that each β-sheet is packed against its two neighbors (59). The main structural parameters of the component β-sheets in β-propeller folds include:
(*a*) N, the number of β-sheets in the structure; (*b*) θ, the twist of the β-sheets measured as the dihedral angle between adjacent strands; (*c*) b, the distance between adjacent strands within a β-sheet; and (*d*) d, the mean perpendicular distance between adjacent β-sheets.

Analysis showed that in the ideal model, these parameters are related by the equation:

$$1/\cos 2\theta - 1/\cos\theta = 2\sin(\pi/N)b/d.$$

Examination of the propeller structures known at that time—those made from six and seven β-sheets—showed that their geometrical parameters do indeed closely fit this equation (59).

The intrinsic properties of β-sheets fix *b* at close to 4.5 Å, and they limit the range of values normally found for both the twist angle, θ, and the distance between close packed β-sheets, *d*. This means that the equation can be used to calculate the number of β-sheets, *N*, that are able to form closed propeller assemblies. This calculation suggests that closed assemblies are most easily formed by seven β-sheets but are also possible for six and eight β-sheets (59). At the time, only cases with six and seven β-sheets were known. Subsequently, a propeller structure with eight β-sheets was discovered (99). At present, one known superfamily has a six–β-sheets propeller; three superfamilies have seven β-sheets; and two superfamilies have eight β-sheets.

Right- and Left-handed β-Helices

Recently a number of structures, called β-helices, have been determined that have regions in which the polypeptide chain forms up to 16 helical turns, with each turn containing two or three β-sheet strands (3, 25, 79, 103) (Figure 5). These turns have 18 to 25 residues, if the occasional insertion is ignored. Strands in adjacent helical turns hydrogen bonds to form either two β-sheets, which pack in a manner similar to the aligned packing in sandwich structures, or three β-sheets, which pack in a triangular array with large residues filling the center and small residues filling the corners (Figure 5).

The β-helix structures are different from β-sandwich folds in three respects: Their β-sheets have little or no twist; the strand directions in the packed β-sheets are close to parallel (Figure 5), rather than tilted at about -30° (Figures 2 and 3); and, although the helix formed by the polypeptide is usually right-handed, in at least one structure (79) it is left-handed (Figure 5*d*). The left-handed helix produces $\beta1$-$\beta2$-$\beta3$ units in which all the connections are left-handed (see Figure 5*d*), whereas in sandwich folds the connections are always right-handed.

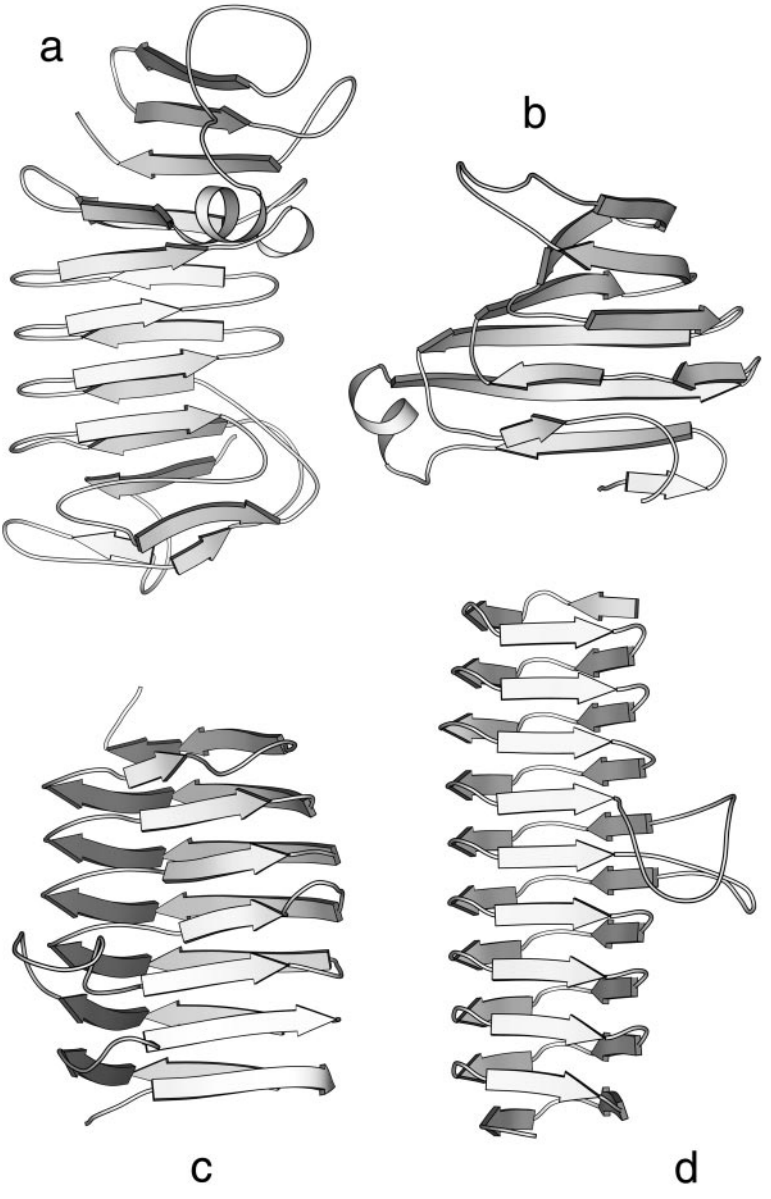


Figure 5 β-Helix folds. (*a*) The β-roll in *Serratia* metalloprotease (3); (*b*) the double β-helix in a canavalin domain (50); (*c*) the *right*-handed β-helix in pectate lyase (103); and (*d*) the *left*-handed β-helix in UDP N-acetylglucosamine acyltransferase (79). All of these folds are made of short β-strands, connected with predominantly short turns. The β-strands are organized in two or three sheets, packed in sandwich- or prism-like fashion. The β-sheets in single helices have almost no twist, which allows them to be left-handed as well as right-handed (see text).

These three differences between β-sandwiches and β-helix folds are related to one another.

If a β-sheet has a simple twist, the edge strands need to stretch further than the inner ones do, if good hydrogen bonds are to be maintained. This effect is not very significant for small β-sheets, but in those with many strands, it means that the β-sheets must be coiled (as well as twisted) or have only a small amount of twist. In β-helices the β-sheets have up to 16 strands and very little if any twist (25). (Coiling of the different β-sheets within one structure would make them splay apart.)

As we noted above, the angle between the strand directions of two packed β-sheets is proportional to their twist, to a first approximation (Figure 2). Thus, in the absence of significant twist, β-sheets tend to pack with parallel strand directions (17).

The right-handed nature of connections usually found in β-β-β units has been shown to be the net result of (*a*) the intrinsic flexibility of polypeptide chains, and (*b*) the structural features of β-sheets in folded proteins (26). The free energy (ΔG) required to bend a chain through an angle ($\Delta\theta$) is given by the equation

$$\Delta G = (\Delta\theta)^2 \, RTa/2L$$

where R is the gas constant, T the temperature, a is the persistence length (about 17 Å for proteins), and L is the chain length (equal to about 3.5 Å per residue). For β-β-β units in β-sheets with normal right-handed twist, connections in right-handed units require smaller bending angles than those in left-handed units do (Figure 2). This means that folding pathways that involve typical β-β-β units with right-handed connections will be favored over those that involve units with left-handed connections. [The energy difference is not large for one unit, but if the connections from different units are not to cross, then nearly all the connections must have the same handedness (26).] Thus, in the β-helix folds where the β-sheets have little or no twist, no preference from the sheet conformation for left- or right-handed β-β-β units would arise.

β-Barrels

The major feature of many proteins is a single large β-sheet that twists and coils to form a closed structure in which the first strand comes round to hydrogen bond to the last strand: the β-sheet barrel. The closed and simple geometrical nature of barrel structures has attracted a number of different theoretical studies (14, 40, 48, 49, 67, 82–85). In our opinion, however, the most coherent and comprehensive theory is the "*n*, *S* model" proposed initially in 1979 (58). Alhough it was used subsequently (51, 64), this model was only developed fully and shown to be generally valid some fifteen years later (65, 66).

The *n, S* model defines barrels in terms of two integral parameters: *n*, the number of strands in the β-sheet, and *S*, the "shear number". In β-barrel structures, neighboring strands are staggered and this allows the twisted and coiled β-sheet to close upon itself (Figure 6). The extent of the stagger of the strands is measured by the shear number, *S*, and can be determined in the manner shown in Figure 6.

The theoretical work on the *n, S* model (58, 65) demonstrated that in ideal barrels the structural features are determined by the values of *n* and *S*. If the following notation is used for the structural features of barrels (Figure 6):

α is the mean slope of the strands to the axis of the barrel,

R is the mean radius of the barrel,

a is the Cα to Cα distance along the strands, and

b is the interstrand distance,

then the *n, S* model showed that in ideal barrels:

1. the mean slope of the strands to the barrel axis, α, and the mean radius of the barrel, *R*, are related to *n* and *S* by:

$$\tan\alpha = Sa/nb,$$

$$R = [(Sa)^2 + (nb)^2]^{1/2}/[2n\sin(\pi/n)];$$

2. the mean twist and coiling angles of the β-sheet are related to *n, S*, and α;

3. packing constraints will favor structures with *n*2#*S*2#2*n*;

4. barrel structures in which residues in β-sheets fill the interior and have low conformational energies are likely to have structures corresponding to one of ten different *n, S* values.

The validity of these relationships was demonstrated by an analysis of all the protein structures then known to contain β-barrels (66). In Figure 7 we show that for β-barrels with the same number of strands (six) the values of *R* and α increase with *S* in a manner predicted by the model. In Figure 8, we show barrel structures that range from the smallest to the largest barrels currently known. In all cases, the geometrical parameters of the barrel structures are very close to those predicted by the *n, S* model (66).

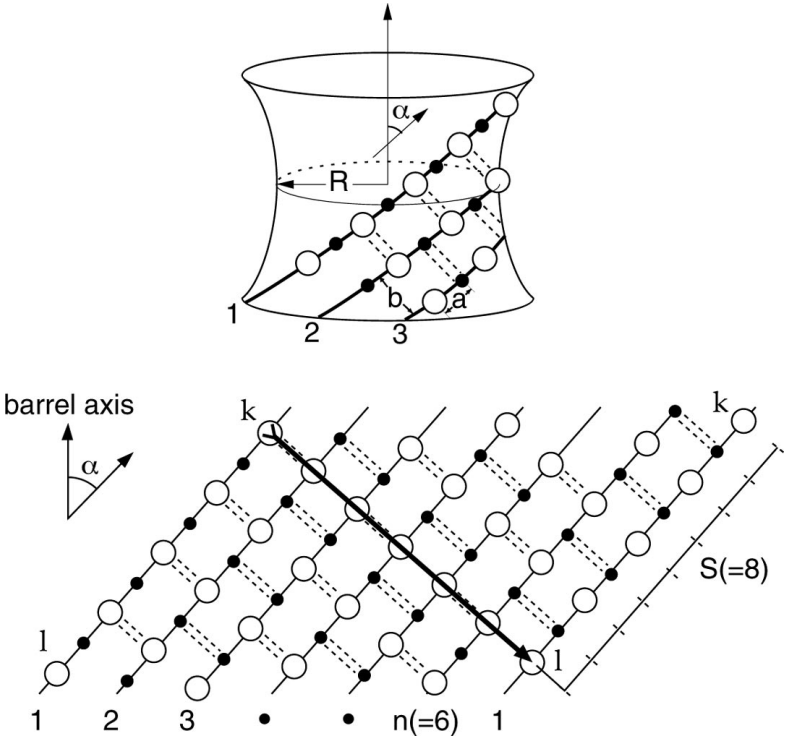


Figure 6 Geometric parameters in β-barrels (58, 65, 66) (*top*) Schematic drawing of a regular β-barrel. Strands are shown as thick lines; residues that point into the interior are shown as small closed circles; residues that point outward are shown as large open circles; and hydrogen bonds are shown as broken lines. Only three strands of the β-sheet is shown. In this drawing we indicate the barrel radius, *R*; the tilt of the strands to the barrel axis, α; and the Cα-to-Cα distance along the strands, *a*, and the interstrand distance, *b*. (*bottom*) A plan of the β-sheet in a barrel produced by cutting along strand 1 and unrolling the β-sheet on a flat surface. The first strand is shown twice to illustrate how its hydrogen bonding to the second and last strands closes the barrel structure. On the plan we indicate the number of strands, *n* (in this example $n = 6$).

The extent of the stagger of the β-sheet is given by the shear number, *S*. *S* can be determined by starting from residue *k* in strand 1, then moving around the barrel, in a direction perpendicular to the direction of the strand, until strand 1 is reached again. Because of the stagger, the point of return will not be residue *k* but a point displaced from it, l. The shear number is $|1 - k|$ (in this example $S = 8$). Note that along a strand, consecutive residues alternate in the direction of their hydrogen bonding; this means that S must be an even integer.

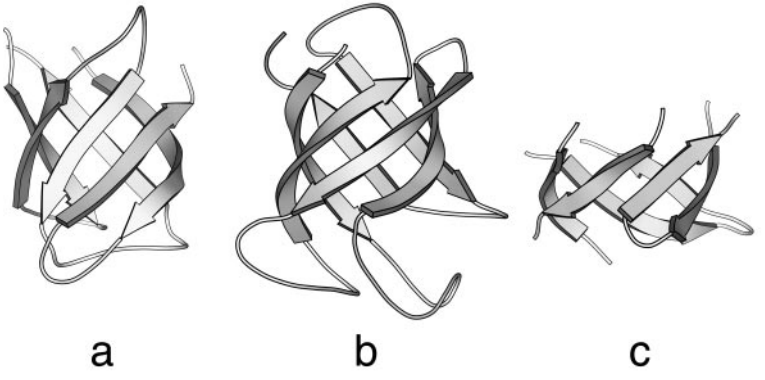


Figure 7 β-Barrels: the relation between R, the mean radius of the barrel; α, the mean slope of the strands to the axis of the barrel; and S, the shear number of the β-sheet (see Figure 6). Here we show three barrel structures with the same number of strands ($n = 6$) but different shear numbers (S): (*a*) a serine protease (29) where $S = 8$; (*b*) elongation factor Tu (42) where $S = 10$; and (*c*) interleukin 1-β (27) where $S = 12$. In ideal models, R and α are related to n and S by equations described in the text. The expected values for these three structures are: 6.2 Å and 45°; 7.0 Å and 51°; and 7.9 Å and 56°, respectively. The observed values are: 6.2 and 43°; 7.1 Å and 49°; and 8.0 Å and 56°, respectively.

β-Prisms

β-Prism folds have three sheets that are packed around an approximately threefold axis. Two orientations have been observed so far (32, 54, 88) (Figure 9). They differ in the direction of the β-sheet strand, relative to the threefold axis: In I the strands are parallel to the axis (Figure 9*a*), whereas in II they are orthogonal to it (Figure 9*b*). In both orientations, the β-sheets are twisted and coiled, and the concave surfaces that are produced by these effects pack together. The packing in the interior is facilitated by small side chains in the regions where the β-sheets are close together and by large side chains in the regions where they are farther apart.

In these structures, as in β-sheet barrels, the twist and coiling of the β-sheets must be interdependent, and these must also be related to the tilt angle between the strand directions and the prism axis. Our unpublished analysis suggests that the prism packing of β-sheets is possible at any tilt angle between 0° and 90°. Indeed, β-sheet packings at intermediate tilt angles of $45 \pm 15°$, and with a true threefold axis, are seen between subunits in trimeric structures (60, 90). The use of tilts of 0° in β-prism I and of 90° in β-prism II are likely to be due to the short connections between the strands of the different sheets, which will favor these orientations rather than intermediate ones.

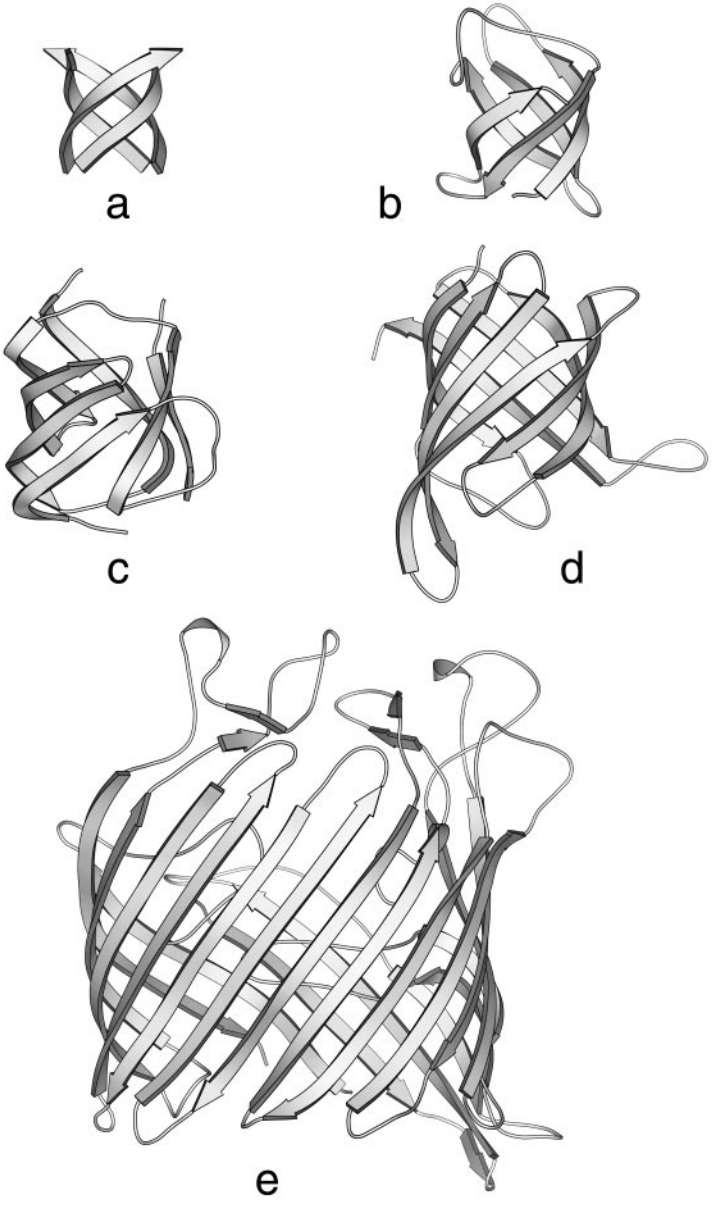


Figure 8 β-Barrel folds with different values of *n, S*. (*a*) The 4,8 barrel formed by the C-terminal tails of the four subunits of the ovomucoid inhibitor (75); (*b*) the 5,10 barrel in major cold shock protein (86); (*c*) the 7,10 barrel in pyruvate kinase (47); (*d*) the 8,12 barrel of the retinol binding protein (19); and (*e*) the 18,22 barrel in maltoporin (87).

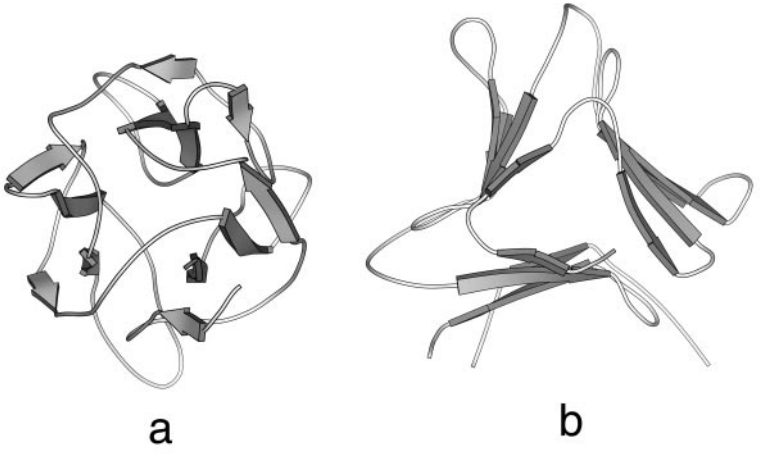


Figure 9 β-Prisms folds. (*a*) Orientation I, exemplified by vitelline-membrane outer-layer protein I (88). All the β-strands are approximately parallel to the pseudo-threefold axis. (*b*) Orientation II, exemplified by snowdrop lectin (32). The strands are nearly orthogonal to the axis. This internal structural symmetry does not always extend to sequences, although some β-prism proteins contain clear sequence repeats.

Other All-β Folds

A few of the known all-β protein structures contain essentially a single β-sheet that usually folds about its more hydrophobic side in a barrel-like fashion, as a result of the stagger, twist and coiling of its strands (Figure 10). Unlike barrel folds, however, this sheet is not closed and is covered partly by loops, which may contain helical segments.

Barrel-sandwich hybrids have two β-sheets, each in the shape of a half-barrel (22, 55). However, they do not associate to form a complete barrel but, instead, use part of the surface of each β-sheet to pack in a sandwich-like fashion (17).

The β-clip fold (9, 39) contains a structural core of three two-stranded β-sheets, which can be described as a long β-hairpin folded upon itself in two places.

FOLDS IN THE ALL-α CLASS

α-Helices have little intrinsic flexibility. Bends of a few degrees are not uncommon; kinked or broken α-helices, with angles of 30°–90° between the two parts, occur occasionally. This means that the folds formed by α-helices can be seen as solutions to the problem of packing cylinders that have a radius of about 10 Å, and surfaces that usually prefer to associate so that the axes of α-helices

making good contacts are inclined at angles of approximately -50° or $+20^\circ$ (Figure 11*c*).

The connections between the α-helices are like those between strands in β-sheets, in that pieces of secondary structure adjacent in the sequence are often in contact in three dimensions. Also, the connections between α-helices are usually short. This means that certain types of chain topologies occur commonly in all-α proteins (24, 77, 97).

Polyhedra Models for Assemblies of Three to Six Helices

The geometry of any assembly of helices packed around a center can be described in terms of a polyhedron: In Figure 11 we show how the geometry of a particular arrangement of three helices corresponds to a regular octahedron. The quasi-spherical polyhedra (QSP) model uses this type of geometrical description to derive a semi-quantitative theory for structures formed by assemblies of three to six helices (62, 63).

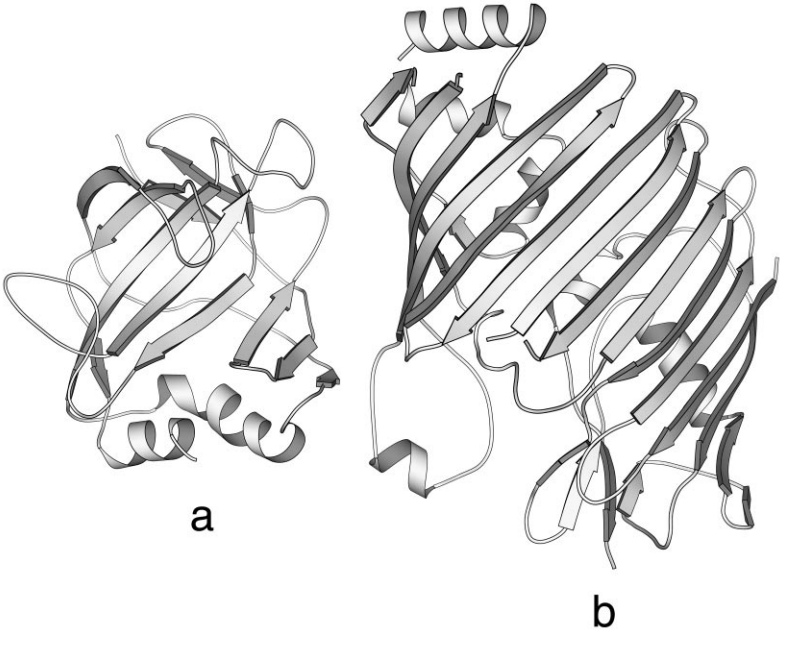


Figure 10 Single β-sheet folds. In the absence of extensive support of other secondary structures, a single β-sheet tends to coil and to fold upon itself and has the appearance of an incomplete barrel. (*a*) Carbonic anhydrase (46). (*b*) Bacteriochlorophyl A protein (94); in this structure the large β-sheet and the small α-helices behind it enclose seven chlorophyl A molecules.

The QSP model assumes that helices in proteins:

(*a*) have the general shape of cylinders;

(*b*) pack with one face directed into the interior of the protein and one face exposed to the solvent;

(*c*) pack around a core whose diameter is the length of two residues;

(*d*) close pack with similar numbers of inter-helical contacts; and

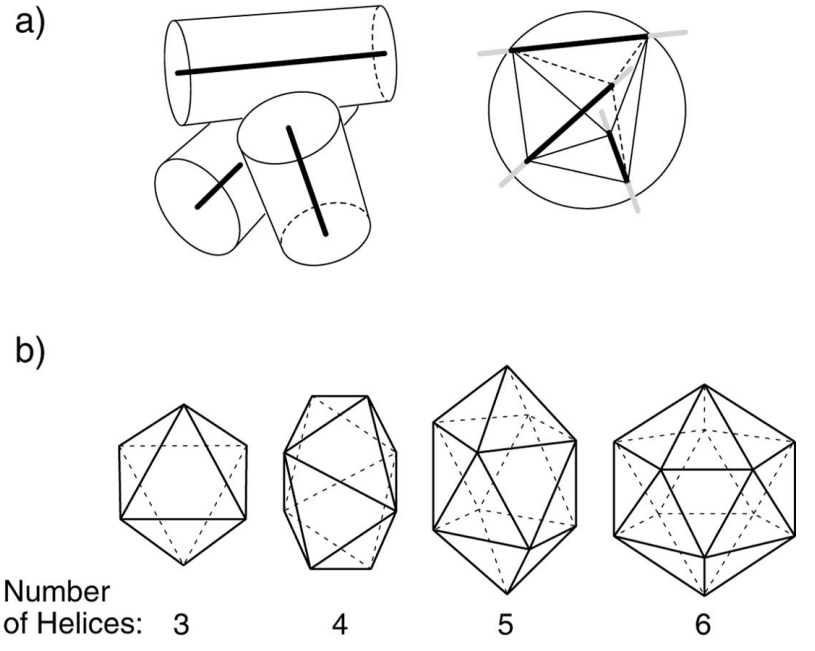
(*e*) form assemblies that are roughly spherical.

These assumptions might seem excessive but they do, in fact, embody features observed to a large extent in real protein structures. The main reason for this is the general distribution of hydrophilic and hydrophobic residues in sequences that usually form soluble globular proteins (26).

The first paper on the QSP model (62) demonstrated that the constraints of these five assumptions result in assemblies with a given number of helices fitting only one polyhedron: All assemblies of three helices are described by an octahedron, those with four helices are described by a dodecahedron, those with five helices are described by a hexadecahedron, and those with six helices are described by an icosahedron. This does not mean, of course, that all structures with the same number of helices have the same structure. There are different ways in which helices can be arranged on a polygon. Three helices can be placed on the octahedron in two different arrangements. Four helices can be placed on the dodecahedron, and five on the hexadecahedron, in ten different arrangements. Six helices can be placed on an icosahedron in eight different ways (63) (Figures 12 and 13). For a given polyhedral arrangement of helices, further variations arise from differences in the direction of the helices and in their connectivity.

An extensive comparison was made of the geometries of ideal models, built for assemblies of three, four, five, or six helices, and geometries observed in the all-α structures known at that time (63). Helices in real protein structures have properties that deviate, to at least some extent, from the assumptions of the QSP model. Nevertheless, in almost all cases, the observed geometries of the helix assemblies have close fits to one of the polyhedron models: The observed position and orientations of the helices differ from those in the corresponding ideal model by 2 Å and 20°, on average (63). The larger deviations occur in proteins where the spherical nature of the packing is distorted by helices clustering, not around a point but around a more elongated region (see below Section).

The shapes of helix surfaces that commonly occur result in pairs of close-packed helices usually having their axes inclined in one of two directions: at


a)
b)
Number
of Helices: 3 4 5 6
c)
Number of cases
30
25
20
15
10
5
0
90 70 50 30 10 -10 -30 -50 -70 -90
Relative orientation of helix axes(°)

angles of approximately -50° or $+20^\circ$ (16, 21) (Figure 11*c*). Polyhedron models in which pairs of neighboring helices are in orientations close to these values are those that most commonly correspond to the observed structures.

Assemblies of Long Helices

Sets of long helices pack together in bundles, i.e. stacks of α-helices whose axes are approximately parallel. In fact, angles between their axes are usually close to $+20^\circ$. This angle is the same as that in coiled coils (21), but little or no coiling of the helices is observed in globular proteins. The most common such structure is the four-helical bundle, but two-, three-, and five-helical bundles are also observed (Figure 14). The geometries of these bundles are close to those that would be predicted by the QSP model if the ideal polyhedra were elongated along one axis.

Large Assemblies: Packing Around a Central Helix and Layer Structures

Assemblies with more than six helices can not simultaneously satisfy all the conditions that form the basis of the QSP model. For more than six helices, formation of a single close-packed globular domain requires one or more helices to be completely buried inside; or it requires the formation of layer structures.

One central helix can accommodate four to six helices around it; larger assemblies have two or more central helices (Figure 15). Central helices must be almost entirely hydrophobic. In an ideal model of four helices packed around a central helix, all the helices make contacts with one another at an ideal angle of -50°. In an ideal model of six helices packed around a central helix, all the helices make contacts with one another at ideal angle of $+20^\circ$. In the observed structures, most of the interactions approximate to these values (our

←

Figure 11 The quasi-spherical–polyhedra models for assemblies helices (62, 63). (*a*) The use of polyhedra to describe the geometry of helix packings. (*a, left*) three α-helices packed together are shown as cylinders with a diameter of 10 Å and the axes as thick lines. (*a, right*) The construction of a polyhedron that describes the geometry of this packing: A sphere is drawn from the center of the packing with a radius of 11 Å. The parts of the helix axes enclosed within the sphere form one set of ribs of the polyhedron. The other set of ribs is formed by lines joining the nearest ends of these parts of the helix axes. Note that, for the three helices shown here, the polyhedron that describes their packing geometry is an octahedron. (*b*) The quasi-spherical polyhedra (QSP) model demonstrates that, for ideal packings of three to six helices, assemblies with a given number of helices fit only one polyhedron. All assemblies of three helices are described by octahedrons, those with four helices are described by a dodecahedron, those with five helices are described by a hexadecahedron and those with six helices are described by an icosahedron (see text). (*c*) The observed relative orientation of pairs of close-packed α-helices in proteins. All orientations are observed but, for both parallel and anti-parallel packings, most pairs have their axes inclined at angles of approximately $+20^\circ$ or -50°.

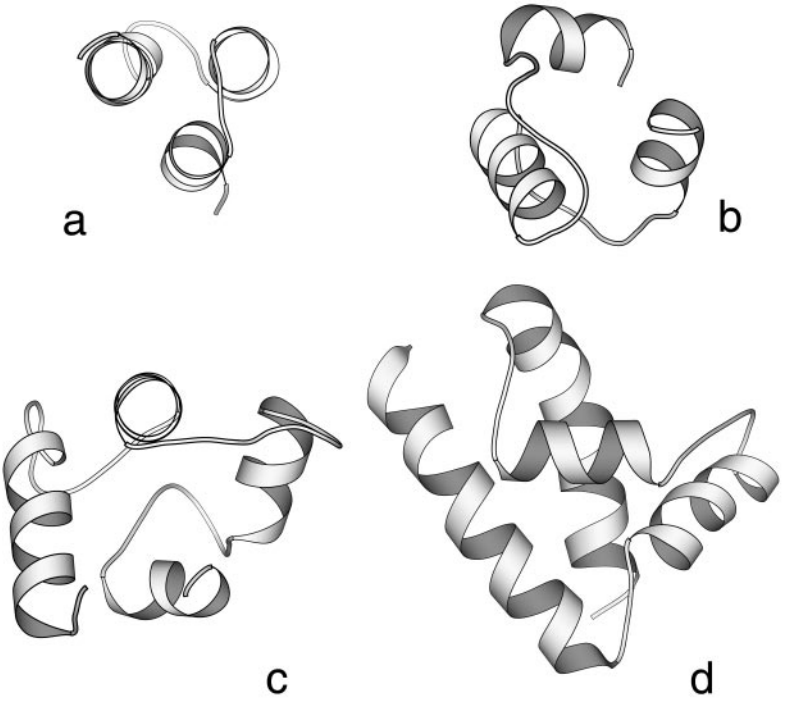


Figure 12 Assemblies of three and four helices. In accordance with the quasi-spherical polyhedra (QSP) model, three helices can be placed on the octahedron model in two different ways: one gives a bundle with the left-handed twist and the other gives a bundle with the right-hand twist. Both twists are favored by the non-zero values of optimal angles for helix-to-helix packing, but their magnitudes are different. (*a*) The left-handed twist is usually small [the interaxial angle in this orientation is $+20^\circ$ (16, 21)], as in pheromone protein ER-1 (98). (*b*) The right-handed twist is larger [the ideal interaxial angle in this orientation is -50° (16)], as in (*b*) the paired domain (101). Both bundles are viewed along their axes.

For assemblies of four helices, the model polyhedron is a dodecahedron. This gives not only helices packed with axes inclined at angles of about -50°, as in calbindin D_{9k} (*c*) (91), but it also arrays with combinations of different helix orientations, as in the POU domain (*d*) (43). For another discussion of four helix assemblies, see Reference 31.

unpublished calculations). For an ideal model of five helices packed around a central helix, the helix packing must involve mixed orientations of about $+20^\circ$ and -50°, as has been observed, or it must use nonideal angles.

The most complex multihelical assemblies are composed of great numbers of helices, of which some may be very long and others may be relatively short. Multihelical assemblies form layer structures of various sizes. They differ in the number of buried helices and can, in principle, have a variety of different folds. However, two general principles are observed in the known structures:

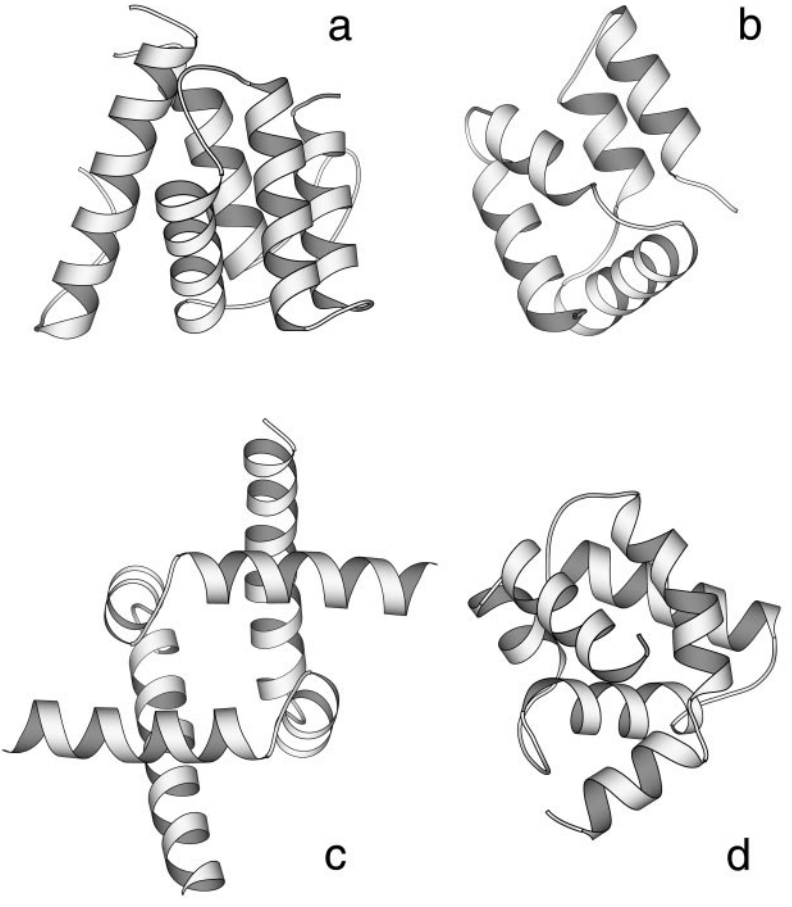


Figure 13 Assemblies of five and six helices. The QSP models for these assemblies allow formations in which helices pack against each other at both positive and negative interaxial angles. The helices in these assemblies do not have a uniform twist, unlike the three-helical assemblies and the four-helix bundles. The structures shown here are from: (*a*) thermolysin (36); (*b*) annexin (37); (*c*) tryptophan repressor (74), and (*d*) glutamyl-tRNA synthetase (68).

(*a*) helices are organized in three or more layers with the longest helices usually in the center, surrounded by shorter helices; and (*b*) the arrangement of the helices often contains internal pseudo-symmetry (Figure 16). Indeed, most multihelical structures contain two or more structural repeats, each with one long and several shorter helices. It is most likely that the observed pseudosymmetries reflect ancient gene duplications that appear to have played a major role in the creation of most multihelical domains.

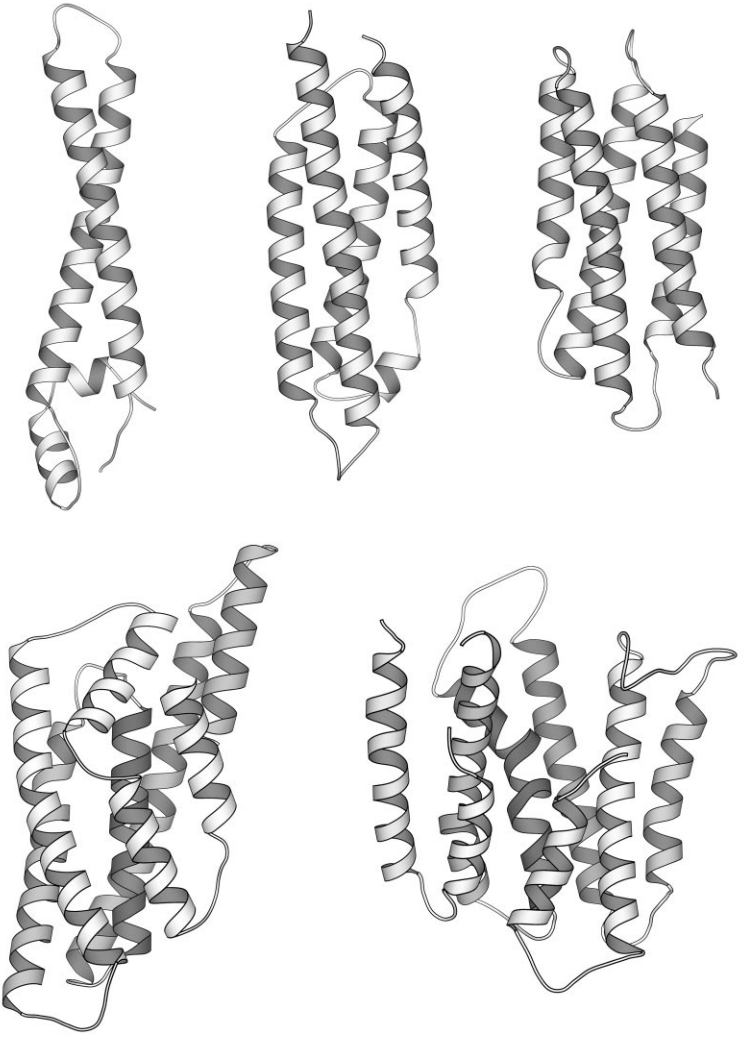


Figure 14 Assemblies of long helices. In these assemblies, helices pack together at the interaxial angle of approximately $+20^{\circ}$ (16, 21). This packing can stabilize a domain with a long hairpin of two helices, as in seryl-tRNA synthetase (*top left*) (28). It is also the dominant feature of different bundle and layer structures: (*top center*) the four-helical bundle of aspartate receptor (6); (*top right*) the five-helical bundle of apolipophorin-III (8); (*d*) the seven-helical bundle with a central helix in bottom left-endotoxin (54); and (*bottom right*) the eight-helical two-layer structure of farnesyl diphosphate synthase (92).

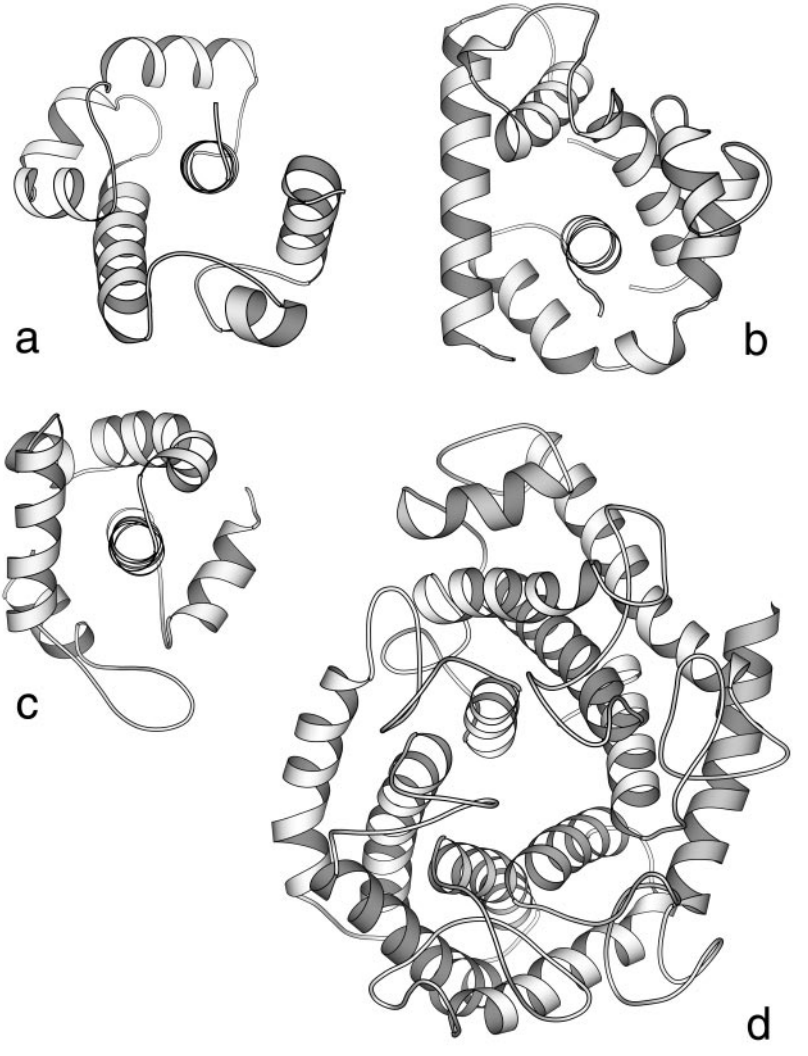


Figure 15 Multihelical assemblies with central helices. The QSP model predicts that it is not possible to pack more than six helices in a globule with a single core. Such assemblies will have one or more helices buried inside the globule, or they will form layer structures. In fact, a central helix can occur in 5-helical globule, as in the HIV matrix protein (*a*) (56). This is a common feature of larger α-helical assemblies, like those of endonuclease III (*b*) (93) and goose lysozyme (*c*) (96). (*d*) The 13-helical assembly in glucoamylase contains as many as six central helices (2).

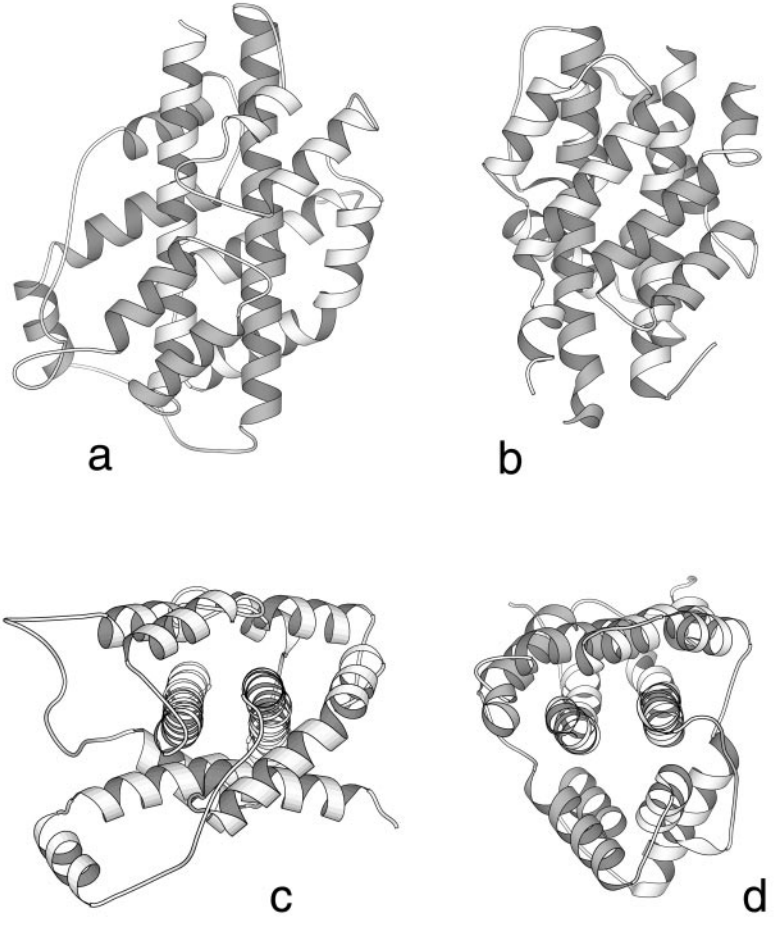


Figure 16 Layer structures and internal pseudosymmetries in multihelical assemblies. (*a* and *c*) Orthogonal views of the large domain of six-phosphogluconate dehydrogenase (1). (*b* and *d*) Orthogonal views of the large domain of citrate synthase (80). Each of these domains are composed of two structural repeats. Their symmetry-related helices are differently shaded. Views *a* and *b* are along the pseudo-twofold axes. Views *c* and *d* show the pairs of long central helices that form the central layer in each structure. The top and the bottom layers in both structures are made of shorter helices.

CONCLUSIONS

We have described the variety of the different folds that are now known for two classes of proteins. What is not apparent from the discussion is the extent to which different fold types occur in proteins. For the currently known protein structures this can be calculated from the fold entries in the SCOP database. β-Sandwiches form just over a third of the all-β folds, as β-barrels do, whereas β-propellers, β-helices, and β-prisms together form less than one third of the all-β folds. Those all-α folds that fit a polyhedron model form half the entries; bundles of long helices form a fifth and axial and layer packing each form one tenth of the entries. Thus, the types of folds that became familiar 20 years ago from the first known protein structures—β-sandwiches, β-barrels, bundles of long α-helices, and what are now known as polyhedral assembles of α-helices—are found in more than two thirds of the currently known folds.

As was mentioned at the beginning of this review, of the protein structures that were determined in 1994 and had no clear sequence similarity to a previously known structure, close to 40% did have a distant evolutionary relation with a previously known structure, and another 20% had the same fold as a previously known structure. The proportion of protein structures with a new fold was 40%. A preliminary calculation for the equivalent structures determined in 1995 shows only about 33% as having a new fold.

Because certain types of proteins are rarely subject to structure determination—membrane proteins and those with repetitive sequences, for example—it is probable that the proportion with new folds would be somewhat higher if new structures came from a more representative selection. Nevertheless, this does indicate that most proteins contain one of a relatively small number of folds: probably fewer than a thousand.

If this is the case, it is not because of physical or chemical limitations on the number of possible folds. It can be observed, for example, that many more folds than a thousand can be generated just by changing, within the limits of the topology rules, the connectivity of the secondary structures in the known folds. The reason for the small number of folds is most likely to be historical. The present evidence suggests that, at a point very early in evolution, a wide range of general functional and catalytic properties arose in a relatively small number of proteins and it became easier to produce new proteins with more specific properties by the duplication, divergence and, in some cases, recombination of old proteins than by *ab initio* invention.

Literature Cited

1. Adams MJ, Ellis GH, Gover S, Naylor, CE, Phillips, C. 1994. Crystallographic study of coenzyme, coenzyme analog and substrate-binding in 6-phosphogluconate dehydrogenase-implications for NADP specificity and the enzyme mechanism. *Structure* 2:651–68
2. Aleshin AE, Golubev A, Firsov LM, Honzatko RB. 1992. Crystal structure of glucoamylase from *Aspergillus Awamori* var. X100. *J. Biol. Chem.* 267:9291–98
3. Baumann U, Wu S, Flaherty KM, Mckay DB. 1993. Three dimensional structure of the alkaline protease from *Ps. aeruginosa*: a two domain protein with a calcium binding parallel β-roll motif. *EMBO J.* 12:3357–64
4. Bentley GA, Boulot G, Karjalainen K, Mariuzza RA. 1995. Crystal structure of the β-chain of a T cell antigen receptor. *Science* 267:1984–87
5. Bernstein FC, Koetzle TF, Williams GJ, Meyer EJ, Brice MD, et al. 1977. The Protein Data Bank: a computer-based archival file for macromolecular structures. *J. Mol. Biol.* 112:535–42
6. Bowie JU, Pakula AA, Simon MI. 1995. The 3-dimensional structure of the aspartate receptor from *Escherichia coli*. *Acta Cryst. D* 51:145–54
7. Brannigan JA, Dodson G, Duggleby HJ, Moody P, Smith JL, et al. 1995. A protein catalytic framework with an N-terminal nucleophile is capable of self-activation. *Nature* 378:416–19
8. Breiter DR, Kanost MR, Benning MM, Wesenberg G, Law JH, et al. 1991. Molecular-structure of an apolipoprotein determined at 2.5 Å resolution. *Biochemistry* 30:603–8
9. Cedergren-Zeppezauer ES, Larsson G, Nyman PO, Dauter Z, Wilson KS. 1992. Crystal structure of a dutpase. *Nature* 335:740–43
10. Chirgadze YN. 1987. Deduction and systematic classification of spatial motifs of the antiparallel β-structure in globular proteins. *Acta Cryst. A* 43:405–17
11. Chothia C. 1984. Principles that determine the structure of proteins. *Annu. Rev. Biochem.* 53:537–72
12. Chothia C, Finkelstein AV. 1990. The classification and origins of protein folding patterns. *Annu. Rev. Biochem.* 59:1007–39
13. Chothia C, Janin J. 1981. Relative orientation of close-packed β-pleated sheets in proteins. *Proc. Natl. Acad. Sci. USA* 78:4146–50
14. Chothia C, Janin J. 1982. Orthogonal packing of β-pleated sheets in proteins. *Biochemistry*. 21:3955–65
15. Chothia C, Levitt M, Richardson D. 1977. Structure of proteins: packing of α-helices and β-sheets. *Proc. Natl. Acad. Sci. USA* 74:4130–34
16. Chothia C, Levitt M, Richardson D. 1981. Helix to helix packing in proteins. *J. Mol. Biol.* 145:215–50
17. Chothia C, Murzin AG. 1993. New folds for all-β proteins. *Structure* 1:217–22
18. Cohen FE, Sternberg M, Taylor WR. 1981. Analysis of the tertiary structure of protein β-sheet sandwiches. *J. Mol. Biol.* 148:253–72
19. Cowan SW, Newcomer ME, Jones TA. 1990. Crystallographic refinement of human serum retinol binding-protein at 2 Å resolution. *Proteins* 8:44–61
20. Crennell SJ, Garman EF, Philippon C, Vasella A, Laver WG, et al. 1996. The structures of *Salmonella typhimurium* LT2 neuraminidase and its complexes with three inhibitors at high resolution. *J. Mol. Biol.* 259:264–80
21. Crick F. 1953. The packing of α-helices: simple coiled coils. *Acta Crystallogr.* 6:689–97
22. Dardel F, Davis AL, Laue ED, Perham RN. 1993. Three-dimensional structure of the lipoyl domain from *Bacillus-stearothermophilus* pyruvate dehydrogenase multienzyme complex. *J. Mol. Biol.* 229:1037–48
23. Efimov AV. 1993. Supersecondary involving triple-strand β-sheets. *FEBS Lett.* 334:253–56
24. Efimov AV. 1994. Favored structural motifs in globular proteins. *Structure* 2:999–1002
25. Emsley P, Charles IG, Fairweather NF, Isaacs NW. 1996. Structure of *Bordetella pertussis* virulence factor P.69 pertactin. *Nature* 381:90–92
26. Finkelstein AV, Ptitsyn OB. 1987. Why do globular proteins fit the limited set of folding patterns. *Prog. Biophys. Mol. Biol.* 50:171–90
27. Finzel BC, Clancy LL, Holland DR, Muchmore SW, Watenpaugh KD, Einspahr HM. 1989. Crystal-structure of recombinant human interleukin-1-β at 2.0 Å resolution. *J. Mol. Biol.* 209:779–91
28. Fujinaga M, Berthetcolominas C, Yaremchuk AD, Tukalo MA, Cusack S. 1993.

Refined crystal-structure of the seryl-transfer RNA-synthetase from *Thermus thermophilus* at 2.5 Å resolution. *J. Mol. Biol.* 234:222–33
29. Fujinaga M, Delbaere L, Brayer GD, James M. 1985. Refined structure of α-lytic protease at 1.7 Å resolution-analysis of hydrogen-bonding and solvent structure. *J.Mol. Biol.* 184:479–502
30. Hardman KD, Ainsworth CF. 1972. Structure of Concanavalin at 2.4 Å resolution. *Biochemistry* 11:4910–19
31. Harris NL, Presnell SR, Cohen FE. 1994. Four helix bundle diversity in globular proteins. *J. Mol. Biol.* 236:1356–68
32. Hester G, Kaku H, Goldstein IJ, Wright CS. 1995. Structure of mannose-specific snowdrop (galanthus-nivalis) lectin is representative of a new plant lectin family. *Nature Struct. Biol.* 2:472–79
33. Holm L, Sander C. 1993. Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.* 233:123–38
34. Holm L, Sander C. 1994. Searching protein structure databases has come of age. *Proteins* 19:65–73
35. Holm L, Sander C. 1996. The FSSP database of structurally aligned protein fold families. *Nucl. Acids Res* 24:206–10
36. Holmes MA, Matthews BW. 1982. Structure of thermolysin refined at 1.6 Å resolution. *J. Mol. Biol.* 160:623–39
37. Huber R, Roemisch J, Paques EP. 1990. The crystal and molecular structure of human annexin V, an anticoagulant protein that binds to calcium and membranes. *EMBO J.* 9:3867–74
38. Ito N, Phillips S, Stevens C, Ogel ZB, McPherson MJ, et al. 1991. Novel thioether bond revealed by a 1.7 Å crystal structure of galactose oxidase. *Nature* 350:87–90
39. Jabri E, Carr MB, Hausinger RP, Karplus PA. 1995. The crystal structure of urease from *Klebsiella aerogenes. Science* 268:998–1004
40. Ke H. 1992. Similarities and differences between human cyclophilin A and other β-barrel structures. *J. Mol. Biol.* 228:539–50
41. Kendrew JC, Dickerson RE, Strandberg BE, Hart RG, Davies DR, et al. 1960. Structure of myoglobin. *Nature* 185:422–30
42. Kjeldgaard M, Nissen P, Thirup S, Nyborg J. 1993. The crystal-structure of elongation-factor EF-TU from *Thermus Aquaticus* in the GTP conformation. *Structure* 1:35–50
43. Klemm JD, Rould MA, Aurora R, Herr W, Pabo CO. 1994. Crystal-structure of the oct-1 pou domain bound to an octamer site-DNA recognition with tethered DNA-binding modules. *Cell* 77:21–32
44. Ko TP, Day J, Greenwood A, McPherson A. 1994. Structures of 3 crystal forms of the sweet protein thaumatin. *Acta Cryst. D* 50:813–25
45. Kuhn P, Tarentin, AL, Plummer TH, Van Roey P. 1994. Crystal structure of peptide-N-4-(N-acetyl-beta-D-glucosaminyl)asparagine amidase-F at 2.2 Å resolution. *Biochemistry* 33:11699–706
46. Kumar V, Kannan KK. 1994. Enzyme-substrate interactions-structure of human carbonic-anhydrase-I complexed with bicarbonate. *J. Mol. Biol.* 241:226–32
47. Larsen TM, Laughlin LT, Holden HM, Rayment I, Reed GH. 1994. Structure of rabbit muscle pyruvate-kinase complexed with Mn^{2+}, K^+, and pyruvate. *Biochemistry* 33:6301–9
48. Lasters I, Wodak SJ, Alard P, Van Cutsem E. 1988. Structural principles of parallel β-barrels in proteins. *Proc. Natl. Acad. Sci. USA* 85:3338–42
49. Lasters I, Wodak SJ, Pio F. 1990. The design of idealized α/β-barrels: analysis of β-sheet closure requirements. *Proteins* 7:249–56
50. Lawrence MC, Suzuki E, Varghese JN, Davis PC, van Donkelaar A, et al. 1990. The three dimensional structure of the seed storage protein phaseolin at 3 Å resolution. *EMBO J* 9:9–15
51. Lesk AM, Bränden CI, Chothia C. 1989. Structural principles of a/β-barrel proteins: the packing of the interior of the sheet. *Proteins* 5:139–48
52. Levitt M, Chothia C. 1976. Structural patterns in globular proteins. *Nature* 261:552–58
53. Li J, Brick P, Ohare MC, Skarzynski T, Lloyd LF, et al. 1995. Structure of full-length porcine synovial collagenase reveals a C-terminal domain-containing a calcium-linked, 4-bladed beta-propeller. *Structure* 3:541–49
54. Li JD, Carroll J, Ellar DJ. 1991. Crystal-structure of insecticidal d-endotoxin from *Bacillus Thuringiensis* at 2.5 Å resolution. *Nature* 353:815–21
55. Liao DI, Kapadia G, Reddy P, Saier MH, Reizer J, Herzberg O. 1991. Structure of the IIA domain of the glucose permease of *Bacillus Subtilis* at 2.2 Å resolution. *Biochemistry* 30:9583–94
56. Massiah MA, Starich MR, Paschall C, Summers MF, Christensen AM, Sund-

quist WI. 1994. 3-Dimensional structure of the human-immunodeficiency-virus type-1 matrix protein. *J. Mol. Biol.* 244:198–223
57. Matthews DA, Appelt K, Oatley SJ. 1989. Stacked β-bulges in thymidylate synthase account for a novel right-handed rotation between opposing β-sheets. *J. Mol. Biol.* 205:449-54
58. McLachlan AD. 1979. Gene duplications in the structural evolution of chymotrypsin. *J. Mol. Biol.* 128:49–79
59. Murzin AG. 1992. Structural principles for the propeller assembly of β-sheets—the preference for 7-fold symmetry. *Proteins* 14:191–201
60. Murzin AG. 1994. New-protein folds. *Curr. Opin. Struct. Biol.* 4:441–49
61. Murzin AG, Brenner SE, Hubbard T, Chothia C. 1995. SCOP-a structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.* 247:536–40
62. Murzin AG, Finkelstein AV. 1983. Polyhedrons describing packing of helices in a protein globule. *Biofizika* 28:905–911
63. Murzin AG, Finkelstein AV. 1988. General architecture of the α-helical globule. *J. Mol. Biol.* 204:749–69
64. Murzin AG, Lesk AM, Chothia C. 1992. β-trefoil fold patterns of structure and sequence in the kunitz inhibitors interleukins-1-β and 1-α and fibroblast growth-factors. *J. Mol. Biol.* 223:531–43
65. Murzin AG, Lesk AM, Chothia C. 1994. Principles determining the structure of β-sheet barrels in proteins. I. A theoretical-analysis. *J. Mol. Biol.* 236:1369–81
66. Murzin AG, Lesk AM, Chothia C. 1994. Principles determining the structure of β-sheet barrels in proteins. II. The observed structures. *J. Mol. Biol.* 236:1382–400
67. Novotny J, Bruccoleri RE, Newell J. 1984. Twisted hyperboloid (Strophoid) as a model of β-barrels in proteins. *J. Mol. Biol.* 177:567–73
68. Nureki O, Vassylyev DG, Katayanagi K, Shimizu T, Sekine S, et al. 1995. Architectures of class-defining and specific domains of glutamyl-transfer-RNA synthetase. *Science* 267:1958–65
69. Ohkawa H, Hogue C, Bryant S, Kans J, Epstein J, et al. 1996. MMDB: Entrez's Structure Database. *WWW*: http://www.-ncbi.nlm.nih.gov/Structure//struchelp.html
70. Orengo C. 1994. Classification of protein folds. *Curr. Opin. Struct. Biol.* 4:429–40
71. Orengo C, Flores TP, Taylor WR, Thorton JM. 1993. Identifying and classifying protein fold families. *Protein Eng.* 6:485–500
72. Orengo C, Michie AD, Thornton JM, Taylor WR. 1995. Class, architecture, topology and homology (CATH) classification of proteins. *WWW*: http://www.-biochem.ucl.ac.uk/bsm/cath
73. Orengo CA, Jones DT, Thorton JM. 1994. Protein superfamilies and domain superfolds. *Nature* 372:631–34
74. Otwinowski Z, Schevitz RW, Zhang RG, Lawson CL, Joachimiak A, et al. 1988. Crystal-structure of Trp repressor operator complex at atomic resolution. *Nature* 335:321–29
75. Papamokos E, Weber E, Bode W, Huber R, Empie MW, et al. 1982. Crystallographic refinement of japanese quail ovomucoid, a kazal-type inhibitor, and model-building studies of complexes with serine proteases. *J. Mol. Biol.* 158:515–37
76. Pflugrath JW, Wiegand G, Huber R, Vertesy L. 1986. Crystal structure determination, refinement and the molecular model of the α-amylase inhibitor Hoe-467A. *J. Mol. Biol.* 189:383–86
77. Presnell SR, Cohen FE. 1989. Topological distribution of four α-helix bundles. *Proc. Natl. Acad. Sci. USA* 86:6592–96
78. Ptitsyn OB, Finkelstein AV. 1981. Similarities in protein topologies: evolutionary divergence, functional convergence or principles of folding. *Q. Rev. Biophys.* 13:339–86
79. Raetz C, Roderick SL. 1996. A left-handed parallel β-helix in the structure of UDP-N-acetylglucosamine acetyltransferase. *Science* 270:997–1000
80. Remington S, Wiegand G, Huber R. 1982. Crystallographic refinement and atomic models of 2 different forms of citrate synthase at 2.7 Å and 1.7 Å resolution. *J. Mol. Biol.* 158:111–52
81. Richardson JS. 1976. Handedness of crossover connections in β-sheets. *Proc. Natl. Acad. Sci. USA* 73:2619–23
82. Richardson JS. 1981. The anatomy and taxonomy of protein structure. *Adv. Protein Chem.* 34:167–339
83. Salemme FR. 1981. Conformational and geometrical properties of β-sheets in proteins: III isotropically stressed configurations. *J. Mol. Biol.* 146:143–56
84. Salemme FR, Weatherford DW. 1981. Conformational and geometrical properties of β-sheets in proteins. I. Parallel β-sheets. *J. Mol. Biol.* 146:101–17
85. Salemme FR, Weatherford DW. 1981. Conformational and geometrical properties of β-sheets in proteins. II. Antipar-

allel and mixed β-sheets. *J. Mol. Biol.* 146:119–41
86. Schindelin H, Marahiel MA, Heinemann U. 1993. Universal nucleic acid-binding domain revealed by crystal-structure of the *Bacillus subtilis* major cold-shock protein. *Nature* 364:164–68
87. Schirmer T, Keller TA, Wang YF, Rosenbusch JP. 1995. Structural basis for sugar translocation through maltoporin channels at 3.1 Å resolution. *Science* 267:512–14
88. Shimizu T, Vassylyev DG, Kido S, Doi Y, Morikawa K. 1994. Crystal-structure of vitelline membrane outer layer protein-I (vmo-I)—a folding motif with homologous greek key structures related by an internal 3-fold symmetry. *EMBO J.* 13:1003–10
89. Sternberg MJE, Thornton JM. 1976. On the conformation of proteins: the handedness of the β-strand-α-helix-β-strand unit. *J. Mol. Biol.* 105:367–82
90. Suzuki M, Sugimoto H, Nakagawa A, Tanaka I, Nishihira J, Sakai M. 1996. Crystal structure of the macrophage migration inhibitory factor from rat liver. *Nature Struct. Biol.* 3:259–66
91. Svensson LA, Thulin E, Forsen S. 1992. Proline cis-trans isomers in calbindin D9k observed by x-ray crystallography. *J. Mol. Biol.* 223:601–6
92. Tarshis LC, Yan MJ, Poulter CD, Sacchettini JC. 1994. Crystal-structure of recombinant farnesyl diphosphate synthase at 2.6 Å resolution. *Biochemistry* 33:10871–77
93. Thayer MM, Ahern H, Xing DX, Cunningham RP, Tainer JA. 1995. Novel DNA-binding motifs in the DNA-repair enzyme endonuclease-III crystal-structure. *EMBO J.* 14:4108–20
94. Tronrud DE, Schmid MF, Matthews BW. 1986. Structure and x-ray amino-acid sequence of a bacteriochlorophyll-A protein from *Prosthecochloris aestuarii* refined at 1.9 Å resolution. *J. Mol. Biol.* 188:443–54
95. Varghese JN, Laver WG, Colman PM. 1983. Structure of the influenza virus glycoprotein antigen Neuraminidase at 2.9 Å resolution. *Nature* 303:35–40
96. Weaver LH, Grutter MG, Matthews BW. 1995. The refined structures of goose lysozyme and its complex with a bound trisaccharide show that the goose-type lysozymes lack a catalytic aspartate residue. *J. Mol. Biol.* 245:54–68
97. Weber PC, Salemme, FR. 1980. Structural and functional diversity in 4-α-helical proteins. *Nature* 287:82–84
98. Weiss MS, Anderson DH, Raffioni S, Bradshaw RA, Ortenzi C, et al. 1995. A cooperative model for receptor recognition and cell-adhesion—evidence from the molecular packing in the 1.6 Å crystal-structure of the pheromone ER-1 from the ciliated protozoan euplotes-raikovi. *Proc. Natl. Acad. Sci. USA* 92:10172–76
99. White S, Boyd G, Mathews FS, Xia ZX, Dai WW, et al. 1993. The active site structure of the calcium-containing quinoprotein methanol dehydrogenase. *Biochemistry* 32:12955–58
100. Woolfson DN, Evans PA, Hutchinison EG, Thornton J. 1993. Topological and stereochemical restrictions in β-sandwich protein structures. *Protein Eng.* 6:461–70
101. Xu WG, Rould MA, Jun S, Desplan C, Pabo CO. 1995. Crystal-structure of a paired domain-dna complex at 2.5 Å resolution reveals structural basis for pax developmental mutations. *Cell* 80:639–50
102. Yee DP, Dill KA. 1993. Families and the structural relatedness among globular proteins. *Protein Sci.* 2:884–99
103. Yoder MD, Lietzke SE, Jurnak F. 1993. Unusual structural features in the parallel β-helix in pectate lyases. *Structure* 1:241–25