PROTEINS: Structure, Function, and Genetics 7:1–15 (1990)

Review Article

α-Helical Coiled Coils and Bundles: How to Design an α-Helical Protein

Carolyn Cohen[1] and David A.D. Parry[2]
[1]Rosenstiel Basic Medical Sciences Research Center, Brandeis University, Waltham, Massachusetts 02254;
[2]Department of Physics and Biophysics, Massey University, Palmerston North, New Zealand

Key words: α-fibrous proteins, 4-α-helix bundle, membrane-spanning proteins

What I cannot create
I do not understand.
—R. Feynman[1]

INTRODUCTION

Fibrous proteins display in rather extravagant ways the structural designs that are common to a wide variety of other proteins. Astbury first recognized the essential simplicity of the fibrous proteins and their paradigmatic role in protein structure[2] (see also ref. 3). Following Pauling's proposal of the α-helix as a structural motif in proteins,[4] the first structures to be solved (in part on the basis of model-building studies) were two fibrous proteins—tropomyosin and α-keratin. In both cases, the molecules were shown to be α-helical coiled coils. During the intervening years, crystal structures of many globular proteins have confirmed the widespread occurrence of α-helices and have also revealed a number of previously unrecognized aspects of this secondary fold. Such features include the clear preferences exhibited by some amino acids for specific positions within the α-helix or at its termini.[5,6] As more structures have been determined, the close relationship between the α- and β-fibrous and the globular proteins has also become apparent. (Note, for example, that not only the β-sheet structure, but also its twist, were first recognized in fibrous proteins[7] (see also ref. 3). We now know also that globular proteins are often cunningly fitted mosaics of a few common folds found in the fibrous proteins. The recognition of these recurring stable structural motifs greatly simplifies the analysis of conformations as yet undetermined. The α-helical coiled coil is one such motif whose properties and occurrence are described in this review.

The α-helical coiled coil exemplifies both general principles of protein design, and special features of α-helical packing. A key point is that—unless it has a special sequence—an isolated α-helix is marginally stable in aqueous solution. The coiled coil is a structure stabilized by the interactions between α-helices: two (or more) right-handed α-helices wind around one another in a left-handed supercoil. This coiling not only optimizes interchain packing of apolar residues, but also allows specific interchain ionic interactions to be made which define the relative chain alignment and direction. Thus the essential aspect of the coiled-coil structure is not the bending of the axis of the α-helix in the supercoiled conformation, but the systematic side chain interactions that are thereby permitted. In globular proteins, as we show below, short regions of helix-helix packing related to the coiled coil do occur, but generally the interactions of α-helical segments are complex and often do not display a simple systematic scheme, such as knobs-into-holes packing.[8] In both fibrous and globular proteins, however, extensive hydrophobic bonding appears to be the driving force for tertiary fold formation in aqueous solution. In both these cases also, the bending of α-helices relates to specific interactions in the tertiary folding.

In a previous review[9] we showed that the α-helical coiled coil may be recognized in the primary sequences of a wide variety of proteins. As suggested by Crick in 1953,[8] the presence of a periodic disposition of apolar residues at positions *a* and *d* within a heptad repeat (of the form $(a\ b\ c\ d\ e\ f\ g)_n$[10]), signals the potential for the interlocking of groups of α-helices. This type of periodicity is common in fibrous proteins, is characteristic of most membrane proteins, and is found as well in a wide range of globular proteins. Here the structures of these three protein classes are related more closely and special features of design are noted. That a feature of tertiary structure may now be recognized directly from the primary sequence of a protein is an important step towards solving protein structures simply from sequence data alone. These observations are also

Received July 26, 1989; accepted August 23, 1989.

Address reprint requests to Dr. Carolyn Cohen, Rosenstiel Basic Medical Sciences Research Center, Brandeis University, P.O. Box 9110, Waltham, MA 02254-9110.

useful in the design of α-helical proteins and some guidelines are proposed.

THE COILED-COIL MOTIF IN α-FIBROUS PROTEINS

One of the key features of the coiled-coil design is the long-range regular disposition of apolar residues in the heptad substructure which allows meshing between these residues in different α-helical chains, thus generating a highly stabilized molecule with an appreciable axial ratio (Figs. 1a, 2). The number of residues per turn (N) in a right-handed α-helix has been shown to vary between 3.50 and 3.65 in globular proteins. Correspondingly, the inclination of the "apolar stripe" formed by the residues in positions *a* and *d* on the surface of a single α-helix with a heptad substructure could vary substantially, and lie anywhere within the range from 0 to about 15°, so that the crossing angle between two helices would be from 0° to 30°. In order for the meshing to be maintained over a long distance, it is also necessary for the axes of the chains to be slightly deformed into a left-handed supercoil. The opposite sense of coiling of the α-helix and the supercoil reduces the effective number of residues per turn of the α-helix to 3.5 for optimal knob/hole packing, thus allowing the apolar stripe to be "internalized," forming the contact line or "seam" between helices. The pitch length of the supercoil (P) depends on the crossing angle between helices, hence on both the geometry of the α-helices and the radius r_o of the coiled coil, and is calculated from the following expression:

$$P = -(n^2h^2 - 4\pi^2 r_o^2)^{1/2}$$

where $n = 3.5N/(3.5-N)$, and h is the axial rise per residue in the undistorted α-helix.[11] Note that just as the right-handed backbone of each α-helix is braced by three left-handed helices of hydrogen bond struts to provide rigidity,[12] so the left-handed sense of backbone supercoiling further stabilizes the coiled-coil structure. By contrast, amphiphilic α-helices interact over relatively limited regions determined by close packing of apolar residues.[13] Here the interacting α-helices cross at diverse angles and, in contrast to the coiled-coil structure, are not regularly inclined to one another.

The pitch length is crucial in specifying intermolecular interactions. In the case of tropomyosin, the average pitch length determined crystallographically is 14 nm, allowing each half turn of the tropomyosin coiled coil to make equivalent interactions with each actin monomer on the thin filament.[14] Other α-fibrous proteins are also believed to have similar pitch lengths,[15,16] and if this is indeed so, then the constituent α-helices will all have about 3.63–3.64 residues per turn and will have crossing angles of about 28°. The inclination of α-helices is thus maintained rather precisely in the α-fibrous proteins, compared with that in globular proteins, where the crossover angles between α-helices vary considerably (see below).

An important aspect of the coiling is the physical basis for the curvature in the helix axis. As Pauling pointed out in 1953, any periodic distribution of amino acid types along an α-helix generates a corresponding perturbation in the main chain hydrogen-bond lengths.[17] This in turn naturally leads to a supercoiling of the axis of the α-helix, and may be achieved with minimal energy cost.[8,18] In the case of the heptad periodicity, the hydrogen bonds are systematically shortened in the vicinity of the apolar residues, so that these residues lie on the concave side of the curved helix and the sense of the curvature is left-handed. Although it is not easily demonstrated, the supercoil pitch length is also defined by this argument and is the same as that based on the meshing of the apolar residues between the two (or more) α-helices constituting the rope-like structure. We thus see that the early and apparently different premises for the physical basis of supercoiling proposed by Crick and by Pauling are, in fact, complementary ways of analyzing the same problem.

It is important to note that although the meshing of apolar residues is essentially non-specific, these residue types are the most highly conserved feature of the structure. Of the apolar residues, leucine is by far the most frequent occupant of both "internal" positions *a* and *d* (see Table I). Other apolar residues do occur preferentially, however, in one or other of the two positions: for example, alanine is favored in position *d*, whereas isoleucine and valine occur most commonly in *a*.[19,20]

A second key feature of the coiled-coil structure is the specificity imposed by interchain ionic interactions and in this regard the charged residues in positions *e* and *g* are especially important. Studies of potential interchain ionic interactions calculated as a function of relative chain direction and stagger have shown that for all two-stranded molecules a parallel in-register arrangement of chains is highly preferred[21] (Fig. 2a). The distribution of acidic and basic residues in positions *e* and *g* in α-fibrous ropes is characteristic for different proteins, but the overall chain disposition is a general feature. (Note that in antiparallel coiled coils such as those in bundles described in a later section, interacting helices must be differently charged to allow stabilizing ionic interactions to be made.)

An early problem posed by the α-fibrous proteins was the number of helical chains in the molecule. There appeared to be no strong a priori argument to favor one possibility over another. Subsequent studies have shown, however, that most intracellular α-helical ropes are two-stranded (for example, tropomyosin, myosin, paramyosin, intermediate filament proteins[9]). Low resolution X-ray data and theoretical studies show that these coiled-coil molecules are

TABLE I. Heptad Repeats in α-Fibrous Proteins: The Average Percentage Occurrence of Residues in Each Position*

	Average % occurrence						
Averages	a	b	c	d	e	f	g
Ala	10.2	12.3	8.1	22.2	4.4	11.1	9.3
Cys	0.9	0.0	0.4	0.3	0.2	0.5	0.1
Asp	0.1	13.4	13.0	1.0	4.0	9.8	8.0
Glu	0.8	21.2	19.3	5.5	31.5	14.7	20.1
Phe	2.0	0.4	1.4	2.1	0.4	0.4	0.0
Gly	0.6	1.7	3.5	1.0	1.3	4.2	1.2
His	1.2	2.7	1.6	1.1	0.8	2.6	0.7
Ile	13.2	0.9	2.2	6.3	2.4	2.2	2.2
Lys	7.7	15.3	11.5	0.6	9.0	10.5	14.9
Leu	32.2	1.6	3.9	34.7	6.4	3.9	5.6
Met	4.9	0.9	0.9	2.3	0.9	1.0	0.4
Asn	3.6	4.3	4.6	1.1	5.7	4.8	2.7
Pro	0.0	0.2	0.1	0.0	0.0	0.0	0.0
Gln	0.8	8.7	7.5	4.1	14.0	6.1	13.2
Arg	5.5	6.2	8.2	0.9	6.6	13.2	10.7
Ser	2.1	3.6	8.1	2.2	5.2	8.1	4.5
Thr	1.2	3.8	3.7	2.2	5.1	3.6	3.5
Val	8.9	1.9	1.9	6.0	1.9	3.1	2.7
Trp	0.1	0.0	0.0	0.7	0.0	0.1	0.0
Tyr	4.1	0.9	0.3	5.7	0.3	0.2	0.2
	100	100	100	100	100	100	100

*The heptad-containing segments of tropomyosin, myosin, paramyosin, and intermediate filament proteins have been analyzed separately; and the percentage occurrence of each residue type in each of the seven positions in the quasirepeat has been determined. The data above represent the average of these results.

dimers with a parallel dyad relating the two chains. Generally these molecules are homodimers. When more than one isoform of a chain is present, however, heterodimers can form, as shown for tropomyosin and cardiac myosin; in the case of keratin intermediate filament proteins, only heterodimers are found. Examples of three-stranded coiled-coil proteins are fewer in number, and all are extracellular[9] (Fig. 2b). They may be heterotrimers with imperfect heptad repeats (for example, fibrinogen[22,23] and laminin) or homotrimers with heptads of special character (for example, the gp17 tail fiber of bacteriophage T7 has alanine in *a* and phenylalanine in *d* over a major segment of its rod structure[24]). Until more three-stranded structures have been analyzed, generalizations regarding sequence requirements are not yet established. There is only one example of a four-stranded α-helical rope and that occurs in the silks of bees, wasps, and ants.[25]

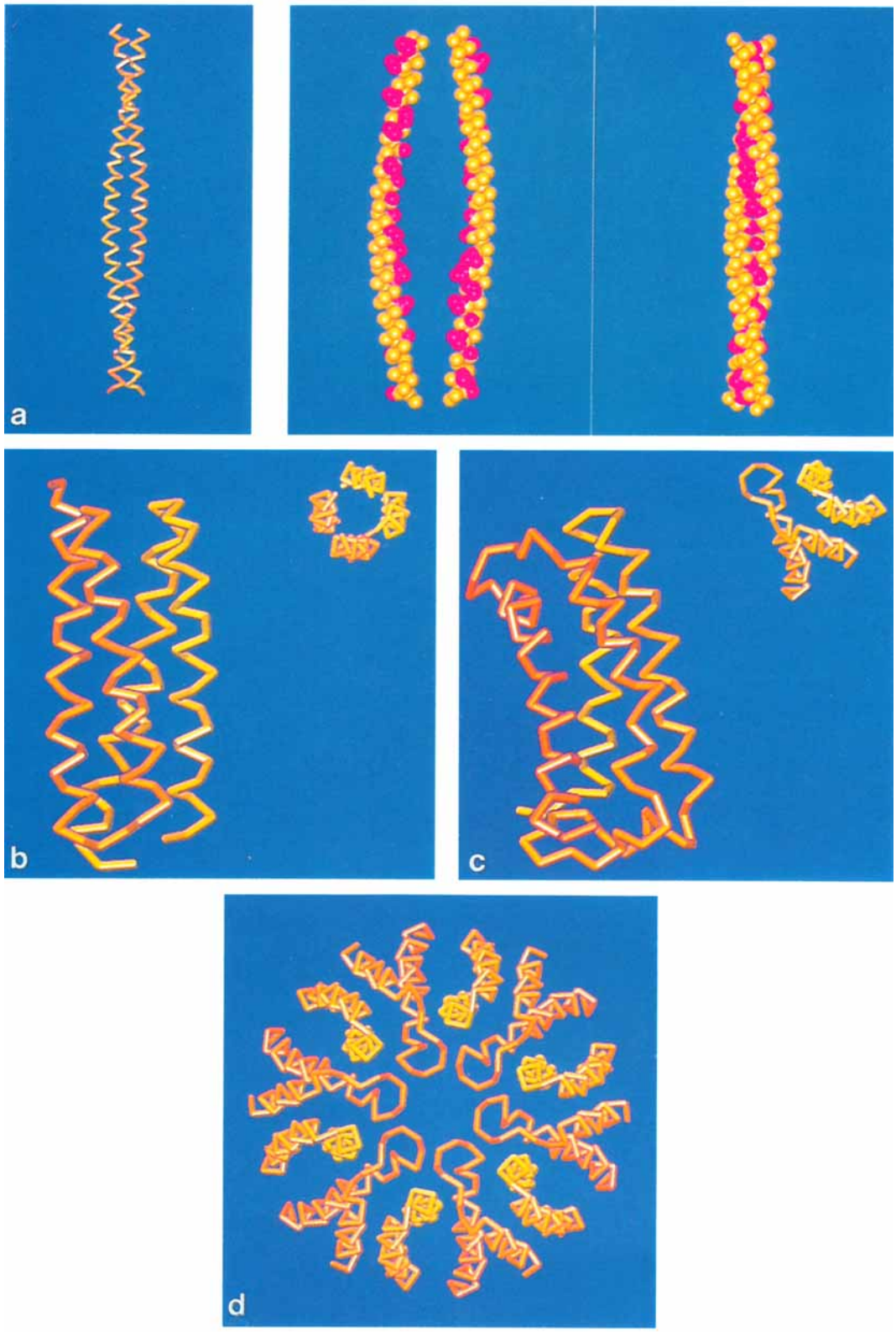
Non-Coiled-Coil Regions in α-Fibrous Proteins

There are now many instances of highly conserved breaks in the pattern of heptad repeats for these proteins. These breaks are known as "skip" residues or "stutters" and are found in all α-fibrous proteins except tropomyosin (Fig. 3). A variety of different but conserved stutters have been characterized: in paramyosin and myosin rod an insertion of one residue (or conversely a deletion of six residues) occurs on five occasions in an otherwise perfect heptad repeat[26–28]; in intermediate filament and lamin proteins, also M-proteins and the gp17 leg protein of bacteriophage T7 insertion(s) of four residues (or a deletion of three residues) occur[21,24,29–31]; in intermediate filament and lamin proteins there is an insertion of five residues (or a deletion of two residues)[21] and in myosin rod an insertion of six residues (or a deletion of one residue) is found.[26–28] The amino acid sequence of the C-terminus of dynein has recently been shown to consist of 68 almost identical heptad repeats (ile/leu-his-val-ile-gln-tyr-ser) interrupted by an unusually large number of skip residues.[32] (Note that the Chou-Fasman technique predicts a β structure for this region.) The effect of these insertions (or deletions) may be to produce a "kink" or local distortion in the axis of the coiled coil. In contrast to tropomyosin, many of these proteins function in vivo in highly ordered filaments consisting of a large number of molecules. The conserved skip residues may thus provide the means by which packing distortions between molecules can be overcome—and might even determine such features of organization as subfilament assemblies. Other weak points in the coiled coil occur where the hydrophobic core shows a local pattern of charged or polar residues, or where there are branched apolar side chains in the *d* position.[14] Such regions impart flexibility to the molecule, as in the case of tropomyosin, or can produce "hinge" regions as in the myosin rod. Thus each of the variety of modulations in the coiled-coil design may well have a simple functional significance.

It is noteworthy that the rod domains of the α-helical coiled-coil proteins often terminate in non-α-helical globular segments. These may be as short as 8 residues, as in the case of tropomyosin, or as large as the "head" domains of myosin or of the neurofilament heavy chain. The coiled coil thus provides a (relatively) stiff rod-shaped component, which positions other domains in the molecule for specific interactions or enzymatic activity (Fig. 3).

Design of a Parallel Two-Stranded Coiled-Coil Motif

From the observations that have been made on the parallel-stranded α-fibrous proteins, it is possible to postulate a number of idealized sequences which would favor such a structure. Hodges and coworkers have undertaken a series of important studies on the stability and properties of tropomyosin-like synthetic polypeptides. They have recently designed and synthesized a repeating heptapeptide (lys-leu-glu-ala-leu-glu-gly) which forms a two-stranded α-helical coiled coil, stabilized chiefly by leucines in the *a* and *d* positions[33] (see also references in ref. 33). Five such heptads appear to be required for conformational stability.


a
b
c
d

By considering additional factors, it may be possible to design a shorter (or longer) stable coiled coil. As one example, consider the sequence (adapted in part from DeGrado et al.[34]) gly-gly-asn-pro-(glu-leu-glu-glu-leu-lys-lys)$_n$-gly-pro-thr where n≥2. Such a sequence consists of an N terminal "initiator," a C-terminal "terminator" and a modular structure in between. The terminal residues become less important as n gets bigger. The key features are as follows:

1. the sequence consists of α-helix favoring residues;
2. the *a* and *d* positions are all occupied by large apolar residues—leucines;
3. positions *b*, *c*, and *f* (the outermost positions in the coiled-coil structure) are all occupied by charged residues;
4. positions *e* and *g* are filled by lysines and glutamic acids respectively, and strongly favor the formation of interchain ionic interactions that would stabilize a parallel-chain in-register arrangement;
5. the N- and C-terminal ends are acidic and basic respectively, to interact favorably with the superhelical macrodipole; see, for example, Hol.[35] (Note that the heptapeptide of Hodges et al.[33] has the opposite charge distribution);
6. intrahelix interactions of the type i → i + 3 and i → i + 4 are possible between ionic residues (when n = 2, these are 5 and 3, respectively) and could add additional stability to the structure (see, for example, ref. 36,37);
7. the sequences (asparagine-proline) and (glycine) are favored initiators and terminators respectively in an α-helix. (Two additional residues may also be added at each end in order to satisfy the hydrogen bonding potential at the helix termini.)[5,6]

Although this sequence could be only about half the length of 5 heptads considered by Hodges and coworkers to be necessary for a stable coiled coil in neutral aqueous solution, the presence of additional inter- and intrachain salt links might be sufficient to stabilize the structure. Based on our current knowledge of the coiled-coil structure, the design sequence described here has many positive attributes. Synthesis of the relevant peptides and their structural characterization will test some of these inferferences.

α-HELICAL BUNDLES: COILED COILS IN GLOBULAR PROTEINS

We consider first those globular monomeric proteins which have a significant α-helix content. Examination of their sequences commonly reveals the presence of noncontiguous segments containing heptad repeats that are coincident with the regions which adopt (or are predicted to adopt) an α-helical conformation. In many cases such α-helical segments have been shown to mesh with one another in a coiled-coil-like manner. While it is not our thesis that all helix-helix interactions are of this type, we

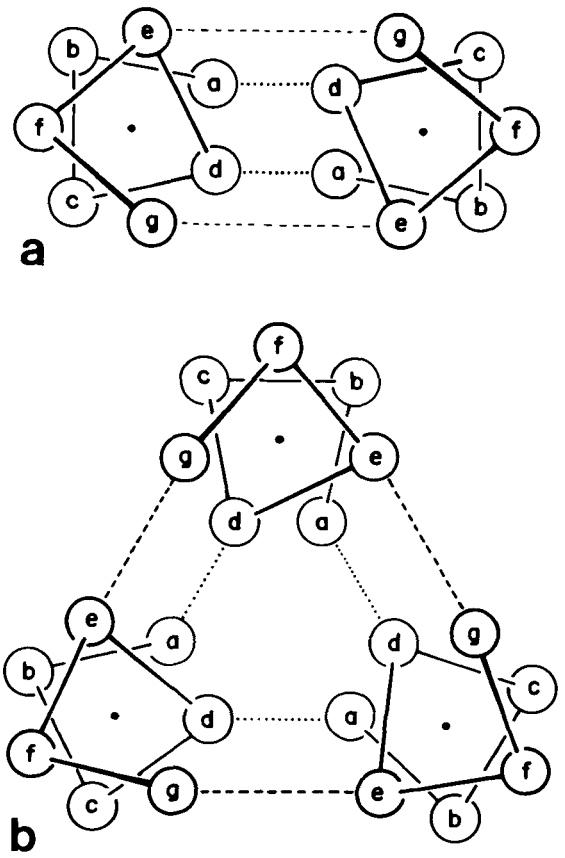


Fig. 2. Interlocking of α-helices to form a coiled coil. The sequences have a heptad pattern of residues of the form (*a b c d e f g*)$_n$, where *a* and *d* are usually apolar, and positions *e* and *g* are frequently charged. **a:** End-on view of two α-helices showing apolar interactions between residues *a* and *d*, and electrostatic interactions between residues *e* and *g*. The helices are parallel and in register. **b:** End-on view of three α-helices with a similar scheme of interlocking.

Fig. 1. α-helical coiled coils in diverse proteins. **a: Left**—Backbone model of two-stranded α-helical coiled coil from a portion of amino-terminal sequence of tropomyosin. **Right**—Space-filling models of the same structure showing the two chains separated (**middle**) and (right) with apolar sidechains (red) interlocking to form a coiled coil.[14] **b:** Backbone model of ColE1 Rop protein dimer.[51] **Side view:** Two monomers related by exact dyad perpendicular to the long axis. Note that pairs of termini are at opposite ends of the bundle, and that this unusually regular and tightly packed structure has a slight left-handed twist. **Top view:** Symmetry of structure. (The darker chains face viewer.) **c:** Backbone model of ferricytochrome c'.[106] This 4-α-helical bundle diverges from the top to accommodate the heme chromophore. **Side view:** Simple up-down right-handed connectivity of helices (as in Fig. 4, but with termini facing down). **End view:** The bend in the so-called "A" helix produced by a proline residue. (The darker chains face viewer.) **d:** Backbone model of gap junction membrane channel made from six ferricytochrome c' bundles which surround a central aqueous channel (adapted from ref. 95). Coordinates for these structures were obtained from the Protein Data Bank, Brookhaven, New York.

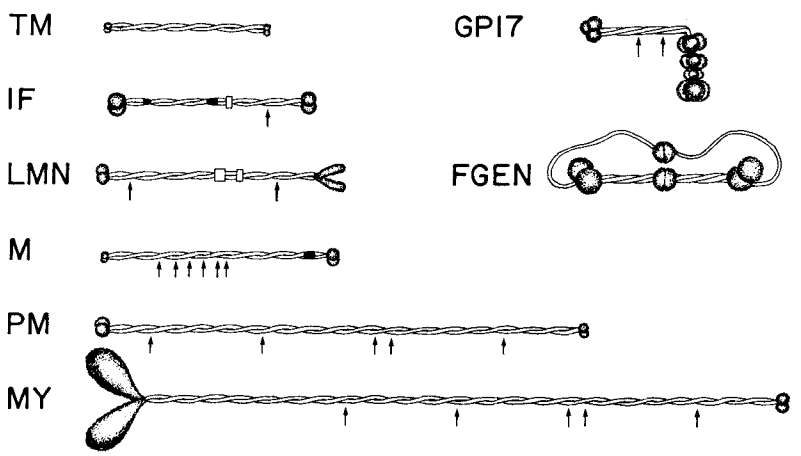


Fig. 3. Domains in some α-fibrous proteins. A schematic diagram of the molecular structure of a small selection of two-stranded and three-stranded α-fibrous proteins, each of which is characterized by an extensive coiled-coil rod domain. In each molecule the termini form a globular domain or a non α-helical region. (Some myosins do not have a non-helical C-terminal "tailpiece".) The two-stranded structures are tropomyosin (TM),[14,15] intermediate filament protein (IF),[21] lamin (LMN),[29] M-protein (M),[31] paramyosin (PM),[26,27] and myosin (MY)[28]; the three-stranded ones are the gp17 protein from bacteriophage T7[24] (GP17) and fibrinogen (FGEN).[22] In all cases the constituent strands are parallel and in axial register. The two-stranded structures generally exist as homodimers except for keratin IF; tropomyosin and cardiac myosin can also exist as a heterodimer. Of the three-stranded structures, fibrino-gen is a heterotrimer (α, β, and γ chains) and gp17 is a homotrimer. In the schematic representations of these proteins the N-terminal domain is drawn at the left. There is one exception: in fibrinogen the N-terminal disulfide knot in the dimeric structure formed is located at the center and the C-terminal ends of the α-, β-, and γ-chains fold independently as indicated. The lengths of the coiled-coil rod domains and of the non-helical terminal domains (stipled) are drawn approximately to scale. The boxed regions indicate a predicted α-helical conformation lacking the heptad repeat characteristic of the coiled-coil structures. Breaks in the heptad repeats ("skips" or "stutters") along the rod are indicated by upward arrows, except in the case of fibrinogen where they are very numerous. The solid black regions in the IF and M proteins represent non-α-helical segments within the coiled-coil rod domain.

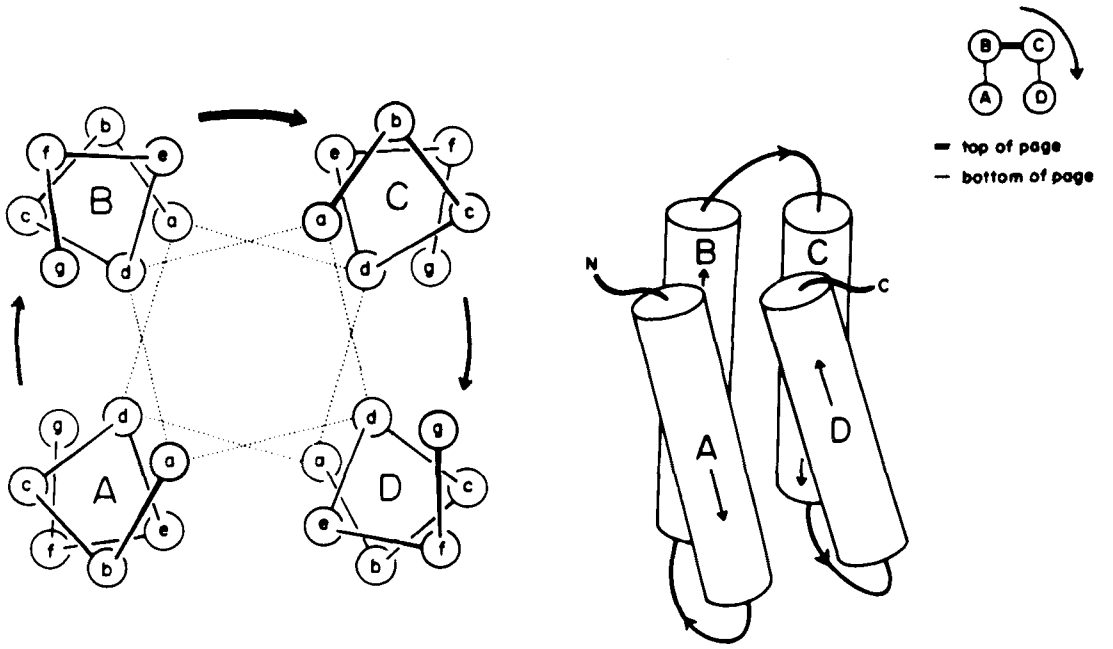
will show that such packing occurs with high frequency.

An important factor that influences the packing of α-helices is the dipole moment associated with each of them. The large dipole moment of the α-helix arises from the approximate orientation of the peptide dipoles along the helix axis. Because the effective dielectric constant is much lower in the protein core than it is in an aqueous environment, dipole-dipole interactions can be of considerable magnitude. Studies on the energetics of packing α-helices in limited bundles show that dipole interactions strongly favor antiparallel packing (see, for example, ref. 35,38–40). Chothia[41] has noted that linkages between elements of regular secondary structure are generally short and often have right-handed connectivity, and this again would favor antiparallel orientations of adjacent helices.

Whatever the packing and connectivity mode adopted, the resulting bundle of α-helices has a compact structure. Depending upon the molecular weight of the protein, it is possible in principle for any number of α-helical chain segments to form a bundle, and in this context it is interesting to note that adjacent elements of secondary structure within a sequence frequently lie in contact in the tertiary structure as well.[42] The basic designs of many globular proteins thus appear to be determined by the same principles as the coiled-coil proteins described above, but nonetheless some special features are likely to exist due to the limited extent of the heptad repeats, their unique sequences and the (generally) antiparallel orientations of neighboring α-helices.

There is no simple rule for assessing the significance of short heptad runs in a sequence; since α-helices in globular proteins can be as long as 3–4 nm, however, a run of about two to three heptads that show at least 80% apolar occupancy of the *a* and *d* positions is noteworthy. For such short regions, the presence of helix-favoring residues in other positions in the heptad also becomes important. In some cases, it is likely that heptad regularity will induce α-helix formation.

An alternative approach to knob/hole packing suggested by Chothia[41] invokes the concept of ridges and grooves formed by the pattern of side chains involved in inter-helix contacts. In this scheme, the ridges formed by the side chains of one α-helix are envisaged as fitting into the grooves formed by the side chains of the second helix. There are three main types of ridges (or grooves) in an α-helix. These are formed by residues one apart (±1n), three apart (±3n) or four apart (±4n). Chothia postulates that the ±4n ridges may have special importance, since they are formed by side chains that splay apart less than for the other two arrangements. Some evidence to support Chothia's approach comes from the dis-



a b

Fig. 4. A 4-α-helical bundle. **a:** End-on view showing regular interlocking of apolar residues (modified from ref. 34 and 51). Further stabilization arises from dipole-dipole interactions of antiparallel chains, and *e-e* and *g-g* interactions (apolar and/or ionic) and possible amphiphilicity (see text). This packing is idealized as in ROP (Fig. 1b) where the interlocking of the apolar residues to stabilize the structure is near to optimal as a consequence of the alternately small and large apolar residues that are found in the *a* and *d* positions respectively of the heptad substructure. In general, however, the interlocking of apolar residues in 4-α-helical motifs is very much poorer. **b:** Right-handed connectivity of bundle in a.

tribution of crossing angles that α-helices make with one another in a variety of globular proteins. A clear peak is found at an angle of −50°, corresponding to ±4n ridges fitting into ±4n grooves. A minor peak at an angle of +20° can be accounted for in terms of ±4n ridges fitting into ±3n grooves. This same angle, however, corresponds to that predicted for a knob/hole packing of "standard" α-helices (i.e., having 3.6 residues per turn) with a heptad repeat. The ridges/grooves idea clearly has value when the ridges (and grooves) are well defined, but in many cases they are composed of side chains (or spaces between side chains) which render them little more than broken ripples (or corrugated undulations). In such cases the mode of packing is not specified with any degree of certainty, and this is reflected in the relatively uniform distribution of crossover angles (excluding those at −50 and +20° as previously discussed). It should also be noted that knob/hole packing is specified by a sequence repeat—the heptad substructure. In contrast, the ridge/groove packing is not simply specified on the basis of sequence substructure and there may be considerable subjectivity involved in deciding which ridge will fit into which groove.

The 4 α-Helical Motif

One α-helical bundle that has been found experimentally in a number of proteins is the 4-α-helical motif[43] (Fig. 4). This structure characterizes such proteins as myohaemerythrin, tobacco mosaic virus, cytochrome c′, apoferritin, and has also been predicted for both interferon[44,45] and a major class of the interleukins.[46] Two forms of the 4-α-helical motif exist. In the first, the interior of the bundle acts as a receptor for a prosthetic group, and in this case the bundle usually diverges from one end (Fig. 1c). In the second form (Figs. 1b, 4), no prosthetic group is involved, and the 4-α-helices diverge symmetrically from the central point at which they are closest packed. Each end of the bundle is an equivalent square in cross-section; each face of the bundle has a dyad axis of symmetry relating antiparallel helix pairs.[40,43]

Although the average length of an α-helix in a globular protein is only about 11 residues,[47] in the case of the 4-α-helical proteins (and related structures) the average length appears to be considerably longer, often in the range 15 to 24 residues (TMV protein—15 residues; cytochrome b_{562} and cy-

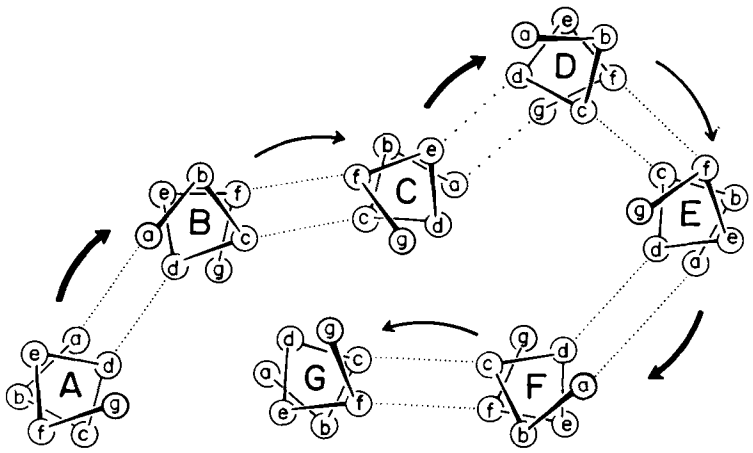


Fig. 5. Schematic diagram of bacteriorhodopsin (as viewed from the inside of the cell). Helices A and G are both pointing up and are at the N and C termini, respectively. The connectivity and helix arrangement shown are that predicted by Engelman et al.[107] and recently established by high resolution electron microscopy (R. Henderson, personal communication). The interlocking scheme is hypothetical, but illustrates that six out of seven contacts have the left-handed coiled-coil interaction; the C-D interlocking is different and is right-handed. There are also interactions (not shown) between helices A and G, as well as between the two layers of helices.

tochrome c′—17 residues; myohaemerythrin—18 residues; apoferritin—24 residues). It seems likely that stabilization is achieved in these cases by additional apolar interactions. The α-helical bundles tend to be somewhat elongated and display rotational symmetry at low resolution. In this respect, they are in contrast to those α-helical proteins which lack heptad repeats (such as myoglobin or hemoglobin). As Weber and Salemme have pointed out,[43] short linking segments (say less than or equal to 11 residues) will lead to a regular antiparallel arrangement of the four α-helices which generally have a connectivity of the up-down nearest-neighbor type (Figs. 4, 1c). Both left- and right-handed connections have been shown to occur with equal frequency.[47a] Longer linkers (say greater than 15 residues) allow, however, other packing modes and connectivities (Fig. 1b) (see, for example, porcine growth hormone[48] and ferritin[49] [see also refs. 47a and 50]). 4-α-Helical bundles also aggregate through hydrophobic interactions to form dimers, as in cytochrome c′ and apoferritin, and higher order structures, as in the tobacco mosaic virus disk and helix, and in membrane proteins (Fig. 1d, and see below).

Intra- and Intermolecular Stabilization by Coiled-Coil-Like Interactions

Similar considerations apply to multisubunit molecules forming α-helical bundles. ColE1 Rop (repressor of primer) protein is a very regular 4-α-helical bundle made up of two monomers related by an exact dyad perpendicular to its long axis[51] (Fig. 1b). Another well-known example is influenza virus hemagglutinin, where the molecular trimer is stabilized by an α-helical bundle formed from a single helix of each monomer.[52] (The leucine residues which pack between helices appear to have only limited contact, rather than interlocking, so that the bundle may be designed to have some dynamic properties (Wiley, personal communication). Larger bundles of α-helices are found, for example, in trypanosome VSGs (variable surface glycoproteins).[53–55] Recent crystallographic results of the dimeric molecule at 0.3 nm resolution show that two VSG monomers interact forming a structure about 7 nm long, which has regions of both 6- and 4-α-helix bundles. Heptad repeats characterize those portions of the sequence involved in interhelix packing contacts.[56] We note also that spectrin, a large fibrous protein, is a heterodimer made up of two elongated antiparallel subunits, which interact weakly except at the ends. Each monomer consists of a linear array of about 20 tandem repeats of a triple-stranded α-helical bundle.[57] Parts of α-actinin and dystrophin appear to be similarly arranged.[58] The average molecular weight of globular proteins, say 30 kDa, generally imposes a limit to the number of α-helical strands in an individual protein to about six or seven. Larger numbers of helices appear to pack in layers (see Fig. 5 and membrane section below), since open tubular structures (cf. ref. 59) are probably energetically unfavorable.

There are also other examples of interactions between globular protein monomers which involve coiled-coil-like interactions. In the case of SV40 and polyoma viruses, the icosahedral shell is stabilized by a short α-helical coiled coil at the local three- and two-fold positions. The same segment of polypeptide chain forms either the two- or the three-stranded structure depending on its location in the shell (Harrison, personal communication). The recently discovered "leucine zipper" characterizing a number of

DNA binding proteins[60] appears in certain cases to consist of a two-stranded coiled coil motif.[61] As the structures of more protein assemblies are determined, it is likely that this mode of stabilization will be recognized more widely (see, for example, ref. 62).

Curvature of α-Helices

Although α-helices in globular proteins have generally been considered to be straight, recent results indicate that the helices are, in fact, gently bent (as in the case of the α-fibrous proteins). A survey of globular proteins by Barlow and Thornton[47] has shown that about 70% of the α-helices have an average curvature of about 1° of arc per residue (i.e., they have a radius of curvature of about 6 nm). (Note that here the axes of the α-helices lie in a plane.) The α-fibrous proteins have a curvature of about 0.5° of arc per residue (or a radius of curvature of about 15 nm). (Here the planes defined by the helical segments rotate in a helical manner from one segment to the next.) This means that helices in the coiled-coil proteins are actually more linear than those in globular proteins.[47] The curvature of the α-helices in the bundles springs from the same physical basis as that discussed above for the α-fibrous proteins, i.e., in the differences which occur in peptide hydrogen bonding on opposite sides of the α-helix and in the packing of apolar side chains involved in stabilizing the hydrophobic core of the protein. Chothia has pointed out[41] that the internally located peptide groups are more closely aligned with the axis of the α-helix than those on the surface. In the latter case the carbonyl oxygen atoms are capable of forming additional hydrogen bonds with the solvent and this may explain in part the orientations of the C = O group and the resulting lengths of the N − H . . . O hydrogen bond. Furthermore, the larger curvature seen in α-helices in globular proteins compared with that in fibrous proteins is likely to arise from their larger radius and the stability gained by making the additional apolar interactions along the length of the strand. Barlow and Thornton[47] have also observed that about 17% of the α-helices in the globular proteins have a kink in them, such that the two pieces make an angle of about 26° with respect to one another. The cause of the kink can be related either to the presence of a (conserved) proline residue or to the requirement that the packing of the helix side chains be optimized. Such kinks are clearly structurally and functionally important and have been discussed previously in relation to the α-fibrous proteins.

Stabilization of α-Helical Bundles

Stabilization of an α-helical bundle can be achieved by ionic interactions between residues in neighboring strands. In α-fibrous proteins charged residues in positions *e* and *g* of the heptad lead naturally to parallel packing of identical chains. In bundles, as in all globular proteins, the chain segments are unique and both parallel and antiparallel combinations may occur in principle, depending on the precise charge relations, but as pointed out above, dipole-dipole interactions strongly favor antiparallel arrangements. We might also note that covalent stabilization through disulfide bonds will not in general be a feature of either the bundles or the coiled coils, since the severe steric requirements of this linkage effectively prevent its formation between elements of regular secondary structure.[63-65] Further stabilization of an α-helix does appear achievable by the formation of salt bridges between oppositely charged residues separated by either three or four residues. The central, isolated α-helix in the troponin-C chain appears to be stabilized by such interactions[36] which may also be present in a number of other globular proteins.[66] Marqusee and Baldwin[37] have designed and synthesized model compounds to assess the relative importance of i,i+3 and i,i+4 interactions in an α-helix. Their conclusions are that the highest α-helical contents are achieved for the i,i+4 combinations (with glu and lys residues) rather than the i,i+3 peptides. Recently, Perutz and Fermi[67] have shown that only the former group are further stabilized by favorable Van der Waals interactions.

Design of a 4-α-Helical Motif

The 4-α-helical bundle is a particularly interesting structure for de novo synthesis. The design of such a motif has been undertaken by DeGrado and his colleagues in an important series of studies.[34] Their structure, formed by four 16-residue peptides (gly-glu-leu-glu-glu-leu-leu-lys-lys-leu-lys-glu-leu-leu-lys-gly) linked by three (pro-arg-arg) sequences, has been produced by gene synthesis and is currently being analyzed. The "protein" is highly α-helical and is very stable in aqueous solution, showing, in fact, a more cooperative unfolding transition than most native proteins of this size. The structure has a heptad repeat with leucines in the *a* and *d* positions, and is also amphiphilic with leucines in the *e* positions. The many charged residues are capable of forming a number of i,i + 3 and/or i,i + 4 intrahelical salt links and the acidic and basic residues at the termini interact favorably with the helix backbone dipole. We note that the leucine residues in the *e* positions also lead to the possibility of a ridges-and-grooves packing (±4n and ±3n) which is similar to the knobs-into-holes interaction. The presence of leucine in position *e*, however, effectively eliminates the possibility of interhelical ionic stabilizing interactions. This structure has not yet been crystallized so that the detailed packing in the multimer has yet to be established. A highly α-helical aggregate of a closely related 12-residue peptide, has been crystallized, however, and shown to form a structure consistent with the 222 symmetry of a 4-

α-helical bundle.[68] Further X-ray analysis of these crystals is in progress.

The de novo design of a complete 4-α-helical repeat, including linker segments ("felix"), has also been undertaken by Hecht et al.[69] They chose a nonrepetitive 79-residue sequence, which they have produced by gene synthesis, and have shown it to be high in α-helix content. This polypeptide has a regular heptad repeat, and is amphiphilic with additional apolar residues in both the *e* and *g* positions. Felix has not yet been crystallized.

In comparing felix to the DeGrado polypeptide, there is a striking difference in the ratio of apolar-to-charged residues. Felix resembles most natural 4-α-helical bundles in that about 50% of the heptad residues are apolar and about 25% are charged. In contrast, the DeGrado polypeptide has only 38% apolar residues but 50% charged ones. The degree of charge is thus higher than that found in α-fibrous proteins (40% charged and 35% apolar). In general, proteins with such a high content of charged residues form elongated structures, and not compact bundles.[9] The regular pattern of leucine residues producing amphiphilicity in the DeGrado polypeptide thus allows additional stabilizing apolar interactions of the core residues.

Although it seems likely that amphiphilicity is an essential aspect of a highly charged α-helix in a 4-α-helical bundle,[68] it is surprising that the elevated apolar content found in a variety of natural 4-α-helical bundles (compared with that in the α-fibrous proteins) appears to be distributed more or less equally in positions *b, c, e,* and *g* with *f* being relatively unaffected. This finding implies that in naturally occurring 4-α-helical bundles, the proportion and degree of regularity of the large apolar residues in the *a* and *d* positions are generally sufficient to drive the folding process. For α-helices that are highly charged, however, such as those in ColE1Rop protein, amphiphilicity appears to be an essential and additional requirement for bundle stability. A key difference then in the design of α-fibrous and 4-α-helical proteins is that complementary charged residues (which stabilize and align parallel α-helices in axial register) are located largely in positions *e* and *g* in the former, whereas in the latter the dipole moment is of especial importance in aligning adjacent α-helices antiparallel to one another. Further stabilization (non-specific) arises from amphiphilicity of the α-helices when these are highly charged, or from the general elevation of the apolar content (leading to a less regular kind of amphiphilicity) in those less highly charged α-helices.

MEMBRANE-SPANNING PROTEINS: STRUCTURES AND PREDICTIONS

There is now considerable evidence that α-helical bundles are the predominant structural motif in membrane-spanning proteins. With the exception of the photosynthetic reaction centers, however, neither the conformation of this tertiary fold at high resolution, nor the forces which lead to its stabilization have yet been established. A number of recent reviews have analyzed methods for predicting the conformation of such transmembrane segments.[70–72] On the basis of the low resolution (7 Å) structure of bacteriorhodopsin[73] (see also ref. 74), and more recently, the detailed crystallographic analysis of two photosynthetic reaction centers,[75–77] together with the availability of many sequences derived from cloning, some of these predictive schemes can now be assessed,[72] and general design principles established. This information may be related to the structures of channel and pump proteins involving transport through aqueous pores, which involve additional design problems.[34,78] In this section we attempt to extend our previous arguments to this class of structures.

The membrane-spanning segments of proteins consist largely of groups of α-helices that traverse the lipid bilayer. α-helices appear to be selected as the principal motif for these proteins because of the ease of partitioning this fully hydrogen-bonded structure into the lipid medium from the aqueous cytoplasm.[79] Membrane proteins are characterized by α-helical segments of lengths of about 23 residues, made up of a high proportion of apolar residues. The lengths of the linker sequences which join consecutive α-helices vary considerably in these proteins, but consecutive membrane spanning segments are often oppositely directed, and frequently (although not necessarily) adjacent to one another within the structure. Transmembrane sequences are generally identified by such methods as the hydropathy profile[71,80] or the calculation of hydrophobic moments.[70] Turn identification procedures are also useful when linker segments are short.[81] Although there are differences in the results obtained by these predictive schemes, by and large there is good agreement in identifying highly hydrophobic transmembrane regions.[72] By contrast, application of the Chou-Fasman and Robson techniques is often unsuccessful in predicting α-helices in membrane proteins,[82] since these algorithms are based on data from proteins in aqueous solution.

Amphiphilic helices, or other elements of structure that span the membrane, are more difficult to predict.[72] The hydrophobic moment method,[70] turn-prediction,[81] Fourier analysis of hydrophobicities,[70,83] or real space measures of amphipathy[84] have been used to identify such regions. More recently, Fourier analysis of conserved residues in the sequences of photosynthetic reaction centers has been shown to reveal the environment of helices in the membrane.[85,86] These methods generally do not have a high reliability when taken individually; when taken together, greater confidence can be attached to the results.

We have suggested that α-helical membrane proteins have a structural design related to the coiled coil,[9] so that transmembrane segments, including those that are amphipathic, may readily be identified by scanning a sequence for segments of three consecutive heptads. Since transmembrane segments are very rich in apolar residues, they will necessarily display this repeat. In these segments the percentage of apolar residues in positions *a* and *d* is higher than that in the α-fibrous proteins, and is generally in excess of 80%. Hence, many positions other than *a* and *d* must also be occupied by apolar residues. Amphiphilic helices, not detected by methods such as the hydropathy profile, will also be revealed by this simple method. In membrane proteins the apolar residues appear to have two distinct roles: those in positions *a* and *d* stabilize the meshing of adjacent α-helices; those in the other positions interact with the lipid side chains and anchor the molecule within the membrane (Fig. 5). On this basis, one would expect the apolar character of residues in the *a* and *d* positions to be highly conserved, whereas those interacting with the lipid side chains are likely to vary more widely. These notions are well illustrated by results on the bacterial photosynthetic reaction centers.[85–87]

Stabilization of Membrane-Spanning Proteins: Are They Inside Out?

A puzzle posed by the membrane proteins is why sequences incorporate the heptad repeat, since the gain in free energy by removing apolar residues from an aqueous environment appears not to apply here. One possibility is that at least parts of the chain are folded prior to insertion into the membrane and maintain much of this conformation when within the membrane. The "helical hairpin" hypothesis of Engelman and Steitz,[79] for example, posits an antiparallel pair of helices in the cytoplasm which then inserts into the membrane. In their view, no interactions between these helices occur in the cytoplasm; they also suggest that key polar interactions, together with van der Waals' forces, are required to stabilize the protein in the lipid milieu of the membrane. This notion of polar forces determining the tertiary structure of membrane proteins (hence their "inside-out" organization compared with soluble proteins) has also been suggested on the basis of neutron diffraction studies of bacteriorhodopsin.[88] Various hydrogen-bonding networks involving polar residues have also been advanced as critical stabilizing interactions between helices in membranes.[81,89,90]

We would suggest that the helical hairpin is stabilized chiefly by apolar knobs-into-holes packing, together with helix-dipole forces, both in the cytoplasm and after insertion into the membrane, and that the additional hairpins would also associate in this way. (The linkers may help solubilize the hairpins in the cytoplasm and may also act to stabilize the tertiary folding in the membrane.) Such a picture fits in with the conservation of buried residues in the reaction centers cited above,[77] as well as the similarity of both residue hydrophobicity and volumes of buried atoms in these membrane proteins and in water soluble proteins.[86,87] Recent findings have, in fact, shown that, although the buried residues in membrane-spanning proteins are more polar than those on the surface, there is not a simple reversal of relative polarities between surface and interior residues in water-soluble and membrane proteins.[91] The driving force for folding in membrane-spanning proteins can thus be ascribed to van der Waals or dispersion forces, which result from the close fit between apolar side chains achieved by knobs-into-holes packing. It is also possible that the characteristic differences in the types of residues in the *a* and *d* positions, and those elsewhere in the heptad, will result in a (small) net free energy gain upon meshing, even in an apolar environment. On this view, the interior packing of membrane proteins is determined by the same interactions as those in soluble proteins, although the surface interactions with lipid are different. To quote a recent review by Rees et al.,[86] "This behavior suggests that water-soluble proteins may be considered as modified membrane proteins with covalently attached polar groups that make the proteins soluble in aqueous solutions." One distinctive feature of the α-helices in the membrane-spanning proteins, however, in contrast to those in soluble proteins, is that they are likely to be relatively straight or possibly curved toward the lipid, on the basis of our analysis of hydrogen bond lengths described above.

The recent structure determination of the membrane-pore-forming fragment of colicin A is of special interest here.[92] This water-soluble protein has been shown to consist of a bundle of ten α-helices organized into a three-layer structure. Colicin A has been described as an "inside-out" membrane protein: two buried helices (forming a hairpin turn) are very apolar, but the protein as a whole has an apolar/charged residue composition similar to other water-soluble α-helical bundles (50/25%). It has been suggested[92] that colicin A is rather loosely folded and penetrates the membrane by a reorganization involving pairwise insertions of helical hairpins, and perhaps oligomerization. The structure in the membrane is not yet known. It is of interest that a very good prediction of the helices in the protein can be made based on the heptad repeats in the sequence. Moreover, all the helices except the highly apolar 8 and 9 are amphiphilic. Because of the very high charge on the protein, it is therefore also possible that pore formation may be produced by insertion (or partial insertion) of only the C-terminal pair of apolar helices, which then form an oligomer.[93] It seems unlikely that the protein can exist in two

folded states of equal helix content (one the inside-out version of the other), which occur alternately in water or in a lipid environment. Unlike the highly apolar helical hairpins in membrane proteins, those in colicin A are too charged to be stable in a lipid milieu.

Molecular Organization of α-Helical Membrane-Spanning Proteins

Although there are few membrane proteins whose structures are known in detail, both sequence prediction and electron microscope studies reveal that there appears to be some common design features for these and many channel and pump proteins. In the case of bacteriorhodopsin and the photosynthetic reaction centers, where there is no aqueous pore, two levels of organization appear to be present. For both kinds of proteins, the monomers consist of somewhat irregular bundles of seven and five α-helices, respectively, with an overall left-handed twist (Fig. 5). Within the bundle, the helices are not related by rotational symmetry, but are arranged in two layers, with the outer layer of helices (consisting of four in the case of bacteriorhodopsin, and three in the case of the photosynthetic reaction centers) tilted out of the plane of the layer with respect to one another by about 20°. Such a structure suggests the presence of a heptad repeat with 4 apolar residues (*a, b, d,* and *f,* or *a, c, d,* and *f*). This feature would provide the physical basis for the twist of the bundle. We have, in fact, identified this type of heptad in the sequences of both bacteriorhodopsin and the photosynthetic reaction centers.[9] The unusual curvature of the helices follows from the arguments cited above. At a higher level of organization, rotational symmetry relates the monomers, so that a closed shell (or somewhat distorted barrel-like structure) results. At this level, the three inner helices of three monomers interlock with one another, in the case of bacteriorhodopsin, or, in the case of the photosynthetic reaction center, a 4-α-helical bundle is formed from the two helices of each monomer.[85,86] In the former, a large (20 Å) channel of lipid results, and in the latter, a smaller pore containing the light harvesting pigments (or cofactors).

As Unwin has pointed out[94] this kind of design seems to characterize other large multisubunit channel and pump proteins, consisting of four or more homologous or identical subunits, such as gap junctions (suggested to be a hexamer of four α-helical bundles[90,95]; Fig. 1d), the acetylcholine receptor,[83,96,97] the sodium channel,[98] and the potassium channel.[99] (Additional examples and references can be found in ref. 34 and 78.) For these proteins, the outer helices are likely to be similar to those in the membrane proteins described above, while the inner helices which do not contact lipid would be amphipathic.[70] In the case of two of these membrane proteins, electron-image maps also show similar features of constriction at one end of the pore, and a more open diameter at the other.[97,100] The shapes of the channels could depend on kinks in the helices, produced, for example, by a proline residue, as in the pore formed by alamethicin molecules[101]; or by the presence of shorter helices, as suggested by Guy and Seetharamulu[98] for the sodium channel; or even by other elements of structure.[78] In the case of transport proteins, which may function as monomers, such as the lactose permease of *Escherichia coli,* a related design has been suggested, involving an outer ring of ten membrane spanning helices in contact with lipid, surrounding an inner ring of (shorter) hydrophilic α-helices forming the transport channel and sugar-binding site.[84,102] The topology of the erythrocyte band 3 protein has been similarly described.[78,103] Such a picture is analogous to the α-fibrous proteins where the terminal domains, usually non-helical, carry out many of the molecule's functions.

Designs for Membrane-Spanning Proteins

The close relationship between the structures of membrane proteins and those of α-helical bundles suggest some design principles for this class of proteins. Any amino acid sequence consisting only of apolar residues is likely to give rise to a stable α-helix in an apolar environment. The starting design would then be such a sequence at least 23 residues in length. In general, specificity is required for functional properties and may be achieved by introducing hydrophilic residues into the sequence. Tandem repeats of these sequences connected by linkers would then assemble into multi-stranded α-helical bundles. Relatively few such residues are required for the bundle to exhibit special packing characteristics. For example, a sequence with a regular heptad repeat, containing, however, additional apolar residues in positions *c* and *f* (or *b* and *f*) would be likely to interlock to form an antiparallel bundle. The hydrophilic residues in such sequences would be largely confined to positions *b, e,* and *g* (or *c, e,* and *g*), but some of these must also be apolar in order to interact with the lipid side chains. Closed rings formed of several such bundles (related by a rotation axis) would then be produced, thus shielding the polar side chains from the apolar milieu. The size and character of the central channel could be determined by the location of charged residues within the heptad substructure, or by the introduction of additional structures derived from the linker regions.[78]

Related sequences have, in fact, recently been synthesized by DeGrado and colleagues and their properties characterized.[104,34] Two heptad triplets were chosen: $(\text{leu-ser-leu-leu-leu-ser-leu})_3$ and $(\text{leu-ser-ser-leu-leu-ser-leu})_3$, based on the recently determined 4-α-helical bundle structure of ColE1 Rop protein,[51] described above. Although the α-helical structure of these peptides in phospholipid mem-

branes is established, the precise number of helices and their interactions are not yet known. By rewriting these sequences, it is easy to show that the first can be both inside-out and conforming as well to the double apolar heptad repeat described above; whereas the second fits neither concept perfectly. These two peptides do, however, appear to form channels with distinctive properties: $(leu\text{-}ser\text{-}leu\text{-}leu\text{-}ser\text{-}leu)_3$ conducts protons efficiently and $(leu\text{-}ser\text{-}ser\text{-}leu\text{-}leu\text{-}ser\text{-}leu)_3$ conducts protons and cations less than 8 Å in diameter. As designs are perfected for 4-α-helical[105] and other bundles for synthetic water-soluble proteins, modifications for stabilization in a lipid environment may be tested. It appears that the "minimalist" approach adopted by DeGrado and associates, may be particularly promising in aiding the rational design for such proteins.

In this short review we have tried to show how the α-helical coiled-coil structure exemplifies general principles of protein design and illuminates also aspects of the structure of certain globular and membrane proteins not previously appreciated. The presence of heptad substructure in a sequence reveals not only important features of secondary structure, but of tertiary structure as well.

ACKNOWLEDGMENTS

We thank Bill DeGrado for criticism of this paper, Xiaoling Xie and Tom Tibbitts for the graphics in Figure 1, Judith Black for assistance with photography, and Louise Seidel for help with the manuscript. This work was supported by grants (to C.C.) from the National Institutes of Health (AR17346), National Science Foundation (DMB85-02233), and the Muscular Dystrophy Association.

REFERENCES

1. Feynman, R. Last blackboard. Phys. Today 42:88. Courtesy of the Archives, California Institute of Technology. Photograph by Robert Paz.
2. Astbury, W.T. Croonian Lecture On the structure of biological fibres and the problem of muscle. Proc. R. Soc. 134B:303–328, 1947.
3. Cohen, C. Design and function of fibrous proteins. In: "CIBA Foundation Symposium: Principles of Biomolecular Organization." Wolstenholme, G.E.W., O'Connor, M., eds. London: J. & A. Churchill Ltd., 1966:101–129.
4. Pauling, L., Corey, R.B., Branson, H.R. The structure of proteins: Two hydrogen-bonded helical configurations of the polypeptide chain. Proc Natl. Acad. Sci. USA 37:205–211, 1951.
5. Presta, L.G., Rose, G.D. Helix signals in proteins. Science 240:1632–1641, 1988.
6. Richardson, J.S., Richardson, D.C. Amino acid preferences for specific locations at the ends of α helices. Science 240:1648–1652, 1988.
7. Fraser, R.D.B., MacRae, T.P., Parry, D.A.D., Suzuki, E. The structure of feather keratin. Polymer 12:35–56, 1971.
8. Crick, F.H.C. The packing of α-helices: Simple coiled-coils. Acta Crystallogr. 6:689–697, 1953.
9. Cohen, C., Parry, D.A.D. α-helical coiled coils—A widespread motif in proteins. Trends Biochem. Sci. 11:245–248, 1986.
10. McLachlan, A.D., Stewart, M. Tropomyosin coiled-coil interactions: Evidence for an unstaggered structure. J. Mol. Biol. 98:293–304, 1975.
11. Fraser, R.D.B., MacRae, T.P. "Conformation in Fibrous Proteins and Related Synthetic Polypeptides." London: Academic Press, 1973.
12. Salemme, F.R. Engineering aspects of protein structure. Ann. NY Acad. Sci. 439:97–106, 1985.
13. Murzin, A.G., Finkelstein, A.V. General architecture of the α-helical globule. J. Mol. Biol. 204:749–769, 1988.
14. Phillips, G.N., Jr., Fillers, J.P., Cohen, C. Tropomyosin crystal structure and muscle regulation. J. Mol. Biol. 192: 111–131, 1986.
15. Parry, D.A.D. Analysis of the primary structure of α-tropomyosin. J. Mol. Biol. 98:519–535, 1975.
16. Quinlan, R.A., Stewart, M. Crystalline tubes of myosin subfragment-2 showing the coiled-coil and molecular interaction geometry. J. Cell Biol. 105:403–415, 1987.
17. Pauling, L., Corey, R.B. Compound helical configurations of polypeptide chains: Structure of proteins of the α-keratin type. Nature 171:59–61, 1953.
18. Parry, D.A.D., Suzuki, E. Intrachain potential energy of the α-helix and of the coiled-coil strand. Biopolymers 7: 189–197, 1969.
19. Parry, D.A.D. Coiled-coils in α-helix-containing proteins: Analysis of the residue types within the heptad repeat and the use of these data in the prediction of coiled-coils in other proteins. Biosci. Rep. 2:1017–1024, 1982.
20. Conway, J.F., Parry, D.A.D. Intermediate filament structure: 3. Analysis of sequence homologies. Int. J. Biol. Macromol. 10:79–98, 1988.
21. Parry, D.A.D., Fraser, R.D.B. Intermediate filament structure. 1. Analysis of IF protein sequence data. Int. J. Biol. Macromol. 7:203–313, 1985.
22. Doolittle, R.F., Goldbaum, D.M., Doolittle, L.R. Designation of sequences involved in the "coiled-coil" interdomainal connections in fibrinogen: Construction of an atomic scale model. J. Mol. Biol. 120:311–325, 1978.
23. Parry, D.A.D. Fibrinogen: A preliminary analysis of the amino acid sequences of the portions of the α, β and γ-chains postulated to form the interdomainal link between globular regions of the molecule. J. Mol. Biol. 120:545–551, 1978.
24. Steven, A.C., Trus, B.L., Maizel, J.V., Unser, M., Parry, D.A.D., Wall, J.S., Hainfeld, J.F., Studier, F.W. Molecular substructure of a viral receptor-recognition protein. The gp17 tail-fiber of bacteriophage T7. J. Mol. Biol. 200: 351–365, 1988.
25. Atkins, E.D.T. A four-strand coiled-coil model for some insect fibrous proteins. J. Mol. Biol. 24:139–141, 1967.
26. Cohen, C., Lanar, D.E., Parry, D.A.D. Amino acid sequence and structural repeats in schistosome paramyosin match those of myosin. Biosci. Rep. 7:11–16, 1987.
27. Kagawa, H., Gengyo, K., McLachlan, A.D., Brenner, S., Karn, J. The paramyosin gene (*unc-15*) of *Caenorhabditis elegans:* Molecular cloning, nucleotide sequence and models for thick filament assembly. J. Mol. Biol. 207:311–333, 1989.
28. Dibb, N.J., Maruyama, I.N., Krause, M., Karn, J. Sequence analysis of the complete *Caenorhabditis elegans* myosin heavy chain gene family. J. Mol. Biol. 205:603–613, 1989.
29. McKeon, F.D., Kirschner, M.W., Caput, D. Homologies in both primary and secondary structure between nuclear envelope and intermediate filament proteins. Nature 319:463–468, 1986.
30. Parry, D.A.D., Conway, J.F., Steinert, P.M. Structural studies on lamin: Similarities and differences between lamin and intermediate filament proteins. Biochem. J. 238:305–308, 1986.
31. Fischetti, V.A., Parry, D.A.D., Trus, B.L., Hollingshead, S.K., Scott, J.R., Manjula, B.N. Conformational characteristics of the complete sequence of group A streptococcal M6 protein. Proteins Struct. Funct. Genet. 3:60–69, 1988.
32. Garber, A.T., Retief, J.D., Dixon, G.H. Isolation of dynein heavy chain cDNAs from trout testis which predict an extensive carboxyl-terminal α-helical coiled-coil domain. EMBO J. 8:1727–1734, 1989.
33. Hodges, R.S., Semchuk, P.D., Taneja, A.K., Kay, C.M., Parker, J.M.R., Mant, C.T. Protein design using model synthetic peptides. Pept. Res. 1:19–30, 1988.

34. DeGrado, W.F., Wasserman, Z.R., Lear, J.D. Protein design, a minimalist approach. Science 243:622–628, 1989.
35. Hol, W.G.J. The role of the α-helix dipole in protein function and structure. Prog. Biophys. Mol. Biol. 45:149–195, 1985.
36. Sundaralingam, M., Drendel, W., Greaser, M. Stabilization of the long central helix of troponin C by intrahelical salt bridges between charged amino acid side chains. Proc. Natl. Acad. Sci. USA 82:7944–7947, 1985.
37. Marqusee, S., Baldwin, R.L. Helix stabilization by Glu^-...Lys^+ salt bridges in short peptides of *de novo* design. Proc. Natl. Acad. Sci. USA 84:8898–8902, 1987.
38. Hol, W.G.J., Halie, L.M., Sander, C. Dipoles of the α-helix and β-sheet: Their role in protein folding. Nature 294: 532–536, 1981.
39. Shoemaker, K.R., Kim, P.S., York, E.J., Stewart, J.M., Baldwin, R.L. Tests of the helix dipole model for stabilization of α-helices. Nature 326:563–567, 1987.
40. Chou, K.C., Maggiora, G.M., Nemethy, G., Scheraga, H.A. Energetics of the structure of the four-α-helix bundle in proteins. Proc. Natl. Acad. Sci. USA 85:4295–4299, 1988.
41. Chothia, C. Principles that determine the structure of proteins. Annu. Rev. Biochem. 53:537–572, 1984.
42. Levitt, M., Chothia, C. Structural patterns in globular proteins. Nature 261:552–558, 1976.
43. Weber, P.C., Salemme, F.R. Structural and functional diversity in 4-α-helical proteins. Nature 287:82–84, 1980.
44. Sternberg, M.J.E., Cohen, F.E. Prediction of the secondary and tertiary structures of interferon from four homologous amino acid sequences. Int. J. Biol. Macromol. 4: 137–144, 1982.
45. Raj, N.B.K., Israeli, R., Kelley, K.A., Leach, S.J., Minasian, E., Sikaris, K., Parry, D.A.D., Pitha, P.M. Synthesis, antiviral activity and conformational characterization of mouse-human α-interferon hybrids. J. Biol. Chem. 263:8943–8952, 1988.
46. Parry, D.A.D., Minasian, E., Leach, S.J. Conformational homologies among cytokines: Interleukins and colony stimulating factors. J. Mol. Recog. 3:107–110, 1988.
47. Barlow, D.J., Thornton, J.M. Helix geometry in proteins. J. Mol. Biol. 201:601–619, 1988.
47a. Presnell, S.R., Cohen, F.E. Topological distribution of four-α-helix bundles. Proc. Natl. Acad. Sci. USA 86: 6592–6596, 1989.
48. Abdel-Meguid, S.S., Shieh, H.-S., Smith, W.W., Dayringer, H.E., Violand, B.N., Bentle, L.A. Three-dimensional structure of a genetically engineered variant of procine growth hormone. Proc. Natl. Acad. Sci. USA 84: 6434–6437, 1987.
49. Rice, D., Ford, G., White, J., Smith, J., Harrison, P. The spatial structure of horse spleen apoferritin. Adv. Inorg. Biochem. 5:39–50, 1983.
50. Richardson, J.S., Richardson, D.C. Principles and patterns of protein conformation. In: "Prediction of Protein Structure and the Principles of Protein Conformation." Fasman, G. (ed.). New York: Plenum Publishing Corp., 1989: 1–98.
51. Banner, D.W., Kokkinidis, M., Tsernoglou, D. Structure of the ColE1 Rop protein at 1.7 Å resolution. J. Mol. Biol. 196:657–675, 1987.
52. Wilson, I.A., Skehel, J.J., Wiley, D.C. Structure of the haemagglutinin membrane glycoprotein of influenza virus at 3 Å resolution. Nature 289:366–373, 1981.
53. Freymann, D.M., Metcalf, P., Turner, M., Wiley, D.C. 6 Å-resolution X-ray structure of a variable surface glycoprotein from *Trypanosoma brucei*. Nature 311:167–169, 1984.
54. Cohen, C., Reinhardt, B., Parry, D.A.D., Roelants, G.E., Hirsch, W., Kanwé, B. α-helical coiled-coil structures of *Trypanosoma brucei* variable surface glycoproteins. Nature 311:169–171, 1984.
55. Metcalf, P., Blum, M., Freymann, D., Turner, M., Wiley, D.C. Two variant surface glycoproteins of *Trypanosoma Brucei* of different sequence classes have similar 6 Å resolution X-ray structures. Nature 325:84–86, 1987.
56. Freymann, D.M. "The Partial Crystal Structure of a VSG From *Trypanosoma Brucei*." Ph.D. Thesis, Harvard University, 1988.
57. Speicher, D.W., Marchesi, V.T. Erythrocyte spectrin is comprised of many homologous triple helical segments. Nature 311:177–180, 1984.
58. Davison, M.D., Critchley, D.R. α-actinins and the DMD protein contain spectrin-like repeats. Cell 52:159–160, 1988.
59. Dunker, A.K., Zaleske, D.J. Stereochemical considerations for constructing α-helical protein bundles with particular application to membrane proteins. Biochem. J. 163:45–57, 1977.
60. Landschulz, W.H., Johnson, P.F., McKnight, S.L. The DNA binding domain of the rat liver nuclear protein C/EBP is bipartite. Science 243:1681–1688, 1989.
61. O'Shea, E.K., Rutkowski, R., Kim, P.S. Evidence that the leucine zipper is a coiled coil. Science 243:538–542, 1989.
62. Buckland, R., Wild, F. Leucine zipper motif extends. Nature 338:547, 1989.
63. Thornton, J.M. Disulphide bridges in globular proteins. J. Mol. Biol. 151:261–287, 1981.
64. Richardson, J.S. The anatomy and taxonomy of protein structure. Adv. Protein Chem. 34:168–339, 1981.
65. Fraser, R.D.B., MacRae, T.P., Sparrow, L.G., Parry, D.A.D. Disulphide bonding in α-keratin. Int. J. Biol. Macromol. 10:106–112, 1988.
66. Sundaralingam, M., Sekharudu, Y.C., Yathindra, N., Ravichandran, V. Ion pairs in alpha helices. Proteins Struct. Funct. Genet. 2:64–71, 1987.
67. Perutz, M.F., Fermi, G. Stereochemistry of salt-bridge formation in α-helices and β-strands. Proteins Struct. Funct. Genet. 4:294–295, 1988.
68. Eisenberg, D., Wilcox, W., Eshita, S.M., Pryciak, P.M., Ho, S.P., DeGrado, W.F. The design, synthesis and crystallization of an alpha-helical peptide. Proteins Struct. Funct. Genet. 1:16–22, 1986.
69. Hecht, M.M., Richardson, D.C., Richardson, J.S., Ogden, R. Design, expression, and preliminary characterization of felix: A model protein. J. Cell. Biochem. 13a:86, 1989.
70. Eisenberg, D. Three-dimensional structure of membrane and surface proteins. Annu. Rev. Biochem. 53:595–623, 1984.
71. Engelman, D.M., Steitz, T.A., Goldman, A. Identifying nonpolar transbilayer helices in amino acid sequences of membrane proteins. Annu. Rev. Biophys. Biophys. Chem. 15:321–353, 1986.
72. von Heijne, G. "Sequence Analysis in Molecular Biology. Treasure Trove or Trivial Pursuit." San Diego: Academic Press, Inc., 1987.
73. Henderson, R., Unwin, P.N.T. Three-dimensional model of purple membrane obtained by electron microscopy. Nature 257:28–32, 1975.
74. Baldwin, J.M., Henderson, R., Beckman, E., Zemlin, F. Images of purple membrane at 2.8 Å resolution obtained by cryo-electron microscopy. J. Mol. Biol. 202:585–591, 1988.
75. Diesenhofer, J., Epp, O., Miki, K., Huber, R., Michel, H. Structure of the protein subunits in the photosynthetic reaction centre of *Rhodopseudomonas viridis* at 3 Å resolution. Nature 318:618–623, 1985.
76. Allen, J.P., Feher, G., Yeates, T.O., Komiya, H., Rees, D.C. Structure of the reaction center from *Rhodobacter sphaeroides* R-26: The protein subunits. Proc. Natl. Acad. Sci. USA 84:6162–6166, 1987a.
77. Allen, J.P., Feher, G., Yeates, T.O., Komiya, H., Rees, D.C. Structure of the reaction center from *Rhodobacter sphaeroides* R-26: The cofactors. Proc. Natl. Acad. Sci. USA 84:5730–5734, 1987b.
78. Lodish, H.F. Multi-spanning membrane proteins: How accurate are the models? Trends Biochem. Sci. 13:332–334, 1988.
79. Engelman, D.M., Steitz, T.A. On the folding and insertion of globular membrane proteins. In: "The Protein Folding Problem." AAAS Selected Symposium 89. Wetlaufer, D.B. (ed.). Boulder, CO: Westview Press, Inc., 1984:87–113.
80. Kyte, J., Doolittle, R.F. A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. 157: 105–132, 1982.
81. Paul, C., Rosenbusch, J.P. Folding patterns of porin and bacteriorhodopsin. EMBO J 4:1593–1597, 1985.
82. Wallace, B.A., Cascio, M., Mielke, D.L. Evaluation of methods for the prediction of membrane protein second-

ary structures. Proc. Natl. Acad. Sci. USA 83:9423–9427, 1986.
83. Finer-Moore, J., Stroud, R.M. Amphipathic analysis and possible formation of the ion channel in an acetylcholine receptor. Proc. Natl. Acad. Sci. USA 81:155–159, 1984.
84. Vogel, H., Wright, J.K., Jähnig, F. The structure of the lactose permease derived from Raman spectroscopy and prediction methods. EMBO J. 4:3625–3631, 1985.
85. Komiya, H., Yeates, T.O., Rees, D.C., Allen, J.P., Feher, G. Structure of the reaction center from *Rhodobacter sphaeroides* R-26 and 2.4.1: Symmetry relations and sequence comparisons between different species. Proc. Natl. Acad. Sci. USA 85:9012–9016, 1988.
86. Rees, D.C., Komiya, H., Yeates, T.O., Allen, J.P., Feher, G. The bacterial photosynthetic reaction center as a model for membrane proteins. Annu. Rev. Biochem. 58: 607–633, 1989.
87. Yeates, T.O., Komiya, H., Rees, D.C., Allen, J.P., Feher, G. Structure of the reaction center from *Rhodobacter sphaeroides* R-26: Membrane-protein interactions. Proc. Natl. Acad. Sci. USA 84:6438–6442, 1987.
88. Engelman, D.M., Zaccai, G. Bacteriorhodopsin is an inside-out protein. Proc. Natl. Acad. Sci. USA 77: 5894–5898, 1980.
89. Burres, N., Dunker, A.K. Membrane transport through α-helical bundles. IV. Preliminary model building investigation of helix-helix interactions. J. Theor. Biol. 87: 723–736, 1980.
90. Milks, L.C., Kumar, N.M., Houghten, R., Unwin, N., Gilula, N.B. Topology of the 32-kd liver gap junction protein determined by site-directed antibody localizations. EMBO J. 7:2967–2975, 1988.
91. Rees, D.C., DeAntonio, L., Eisenberg, D. Hydrophobic organization of membrane proteins. Science 245:510–513, 1989.
92. Parker, M.W., Pattus, F., Tucker, A.D., Tsernoglou, D. Structure of the membrane-pore-forming fragment of colicin A. Nature 337:93–96, 1989.
93. Pattus, F., Heitz, F., Martinez, C., Provencher, S.W., Lazdunski, C. Secondary structure of the pore-forming colicin A and its C-terminal fragment. Experimental fact and structure prediction. Eur. J. Biol. 152:681–689, 1985.
94. Unwin, N. Is there a common design for cell membrane channels? Nature 323:12–13, 1986.
95. Tibbitts, T.T., Caspar, D.L.D., Phillips, W.C., Goodenough, D.A. Diffraction diagnosis of protein folding in gap junction connexons. Biophys. J. (in press).
96. Brisson, A., Unwin, P.N.T. Quaternary structure of the acetylcholine receptor. Nature 315:474–477, 1985.
97. Toyoshima, C., Unwin, N. Ion channel of acetylcholine receptor reconstructed from images of postsynaptic membranes. Nature 336:247–250, 1988.
98. Guy, H.R., Seetharamulu, P. Molecular model of the action potential sodium channel. Proc. Natl. Acad. Sci. USA 83:508–512, 1986.
99. Tempel, B.L., Papazian, D.M., Schwarz, T.L., Jan, Y.N., Jan, L.Y. Sequence of a probable potassium channel component encoded at *Shaker* locus of *Drosophila*. Science 237:770–775, 1987.
100. Unwin, P.N.T., Zampighi, G. Structure of the junction between communicating cells. Nature 283:545–549, 1980.
101. Fox, R.O., Jr., Richards, F.M. A voltage-gated ion channel model inferred from the crystal structure of alamethicin at 1.5-Å resolution. Nature 300:325–330, 1982.
102. Overath, P., Wright, J.K. Lactose permease: A carrier on the move. Trends Biochem. Sci. 8:404–408, 1983.
103. Kopito, R.R., Lodish, H.F. Primary structure and transmembrane orientation of the murine anion exchange protein. Nature 316:234–238, 1985.
104. Lear, J.D., Wasserman, Z.R., DeGrado, W.F. Synthetic amphiphilic peptide models for protein ion channels. Science 240:1177–1181, 1988.
105. Ohlendorf, D.H., Finzel, B.C., Weber, P.C., Salemme, F.R. Some design principles for structurally stable proteins. In: "Protein Engineering." Oxender, D.L., Fox. C.F. (eds.). New York: Alan R. Liss, Inc., 1987:165–173.
106. Finzel, B.C., Weber, P.C., Hardman, K.D., Salemme, F.R. Structure of ferricytochrome c′ from *Rhodospirillum molischianum* at 1.67 Å resolution. J. Mol. Biol. 186:627–643, 1985.
107. Engelman, D.M., Henderson, R., McLachlan, A.D., Wallace, B.A. Path of the polypeptide in bacteriorhodopsin. Proc. Natl. Acad. Sci. USA 77:2023–2027, 1980.