Biochem. J. (1977) **163**, 45–57
Printed in Great Britain

Stereochemical Considerations for Constructing α-Helical Protein Bundles with Particular Application to Membrane Proteins

By A. KEITH DUNKER* and DAVID J. ZALESKE†
**Program in Biochemistry and Biophysics, Washington State University, Pullman, WA 99163, U.S.A., and †Massachusetts General Hospital, Fruit Street, Boston, MA 82114, U.S.A.*

(*Received* 12 *August* 1976)

The stereochemical constraints originally used to construct two- and three-stranded α-helical coiled-coils were generalized for aggregates of α-helices containing from 4 to 14 α-helices in tubular bundles. Certain features of bacteriorhodopsin show excellent correlations with these stereochemical constraints.

Formation of multistrand cables from protomer molecules seems to be a general feature of fibrous structures, as for example in muscle (Huxley & Brown, 1967; Pepe, 1967, 1971; Harrison *et al.*, 1971), in α-keratin (Astbury & Street, 1931; Bragg *et al.*, 1950; Pauling *et al.*, 1951; Crick, 1952) and in collagen (Rich & Crick, 1955, 1961). A recent interesting example of this phenomenon is provided by glutamine synthetase, in which hexagonal molecules are induced by Co^{2+} ions to form long strands; the long strands in turn form three-stranded and six-around-one-stranded cables (Frey *et al.*, 1975). Six-around-one cables and three-stranded coiled-coils are both structures that had been proposed as possible ways of aggregating α-helical protomers (Pauling & Corey, 1953; Crick, 1953). Of course the sizes of the proposed α-helical coiled-coils and the glutamine synthetase cables are very different; the repeating unit in glutamine synthetase has a mol.wt. of about 600000 whereas the α-helix repeating unit consists of just a few amino acids. This large disparity in size points to the concept that the same general principles are operating in both cases (Frey *et al.*, 1975).

Crick (1953) explained coiled-coil formation by noting that helical strands can have regular repeated favourable interstrand contacts if the strands coil about each other. He worked out in detail the consequences of this principle for the case of two- and three-stranded α-helical coiled-coils. Since the original work of Pauling & Corey (1953) and Crick (1953), it has become evident that there may be structures containing more than three α-helical chains in tubular aggregates. Thus it would be useful to work out the consequences of the packing principles devised by Crick to aggregates with more than three α-helices. In addition to certain of the α-helical fibrous proteins, a possibility of particular interest to us is that transmembrane protein channels are formed by aggregates of α-helices. This concept, originally proposed on general grounds by several workers (Wallach & Zahler, 1966; Lenard & Singer, 1966), was recently used as the basis for construction of molecular models containing up to 18 molecules of murein-lipoprotein in α-helical bundles (Inouye, 1974). In addition, the concept of transmembrane α-helical aggregates has received considerable support from work showing that protein from the purple membrane of *Halobacterium halobium* consists of seven aggregated transmembrane α-helical segments (Henderson, 1975; Unwin & Henderson, 1975; Henderson & Unwin, 1975). Especially because of these recent findings and suggestions with regard to membrane proteins, we have extended Crick's (1953) original analysis to tubes of α-helices containing up to 14 α-helices. Rather than presenting this work in the most general form possible, we have included constraints that seem to us to be reasonable for particular kinds of membrane proteins. The present stereochemical principles have been applied to several membrane-protein sequences to suggest possible structures for these aggregates and further, to propose a model for membrane transport through such aggregates (A. K. Dunker & T. Jones, unpublished work; Dunker *et al.*, 1976).

In addition to our particular interest in membrane proteins, the glutamine synthetase example suggests that our work may have potential application beyond the confines of membrane proteins and even, if suitably generalized, beyond the particular case of α-helical aggregates.

A preliminary account of these results has been published (Dunker *et al.*, 1976).

Theory

Tubes of α-helices

Crick (1953) proposed a method for packing two and three α-helical chains together so that the side chains of one α-helix fit into the spaces between the side chains of its neighbour. Crick called this knobs-

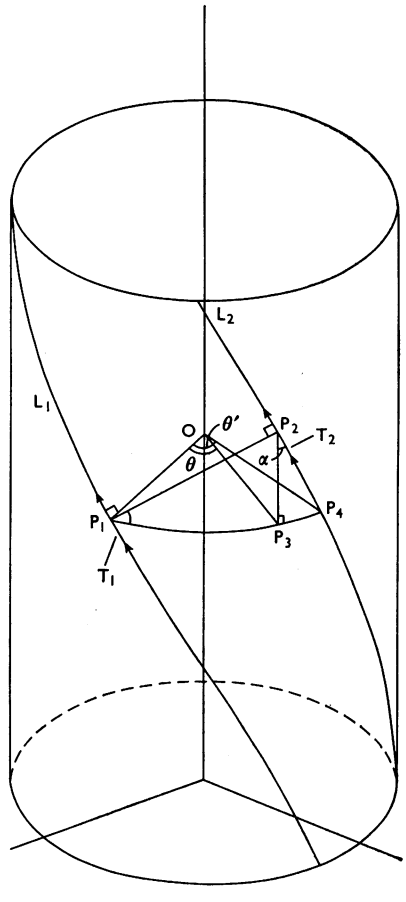

Fig. 1. *Packing of smooth cylinders*
On the surface of a cylinder of radius r_0 are drawn two helices L_1 and L_2 representing the axes of two smooth cylinders. Drop a perpendicular (on the cylindrical surface) from point P_1 on helix L_1 to point P_2 on helix L_2. Draw the radii from points P_1 and P_2 to the cylindrical axis and project the radius from P_2 on the plane P_1OP_3, which is perpendicular to the cylinder axis. This projected line makes an angle $\theta = P_1OP_3$ at the axis with the radius to point P_1. Draw tangents T_1 and T_2 to helices L_1 and L_2 at points P_1 and P_2. The angle ε is the angle between the (non-intersecting) lines T_1 and T_2, and is thus the angle at which two neighbouring α-helices cross.

into-holes packing. We have extended the principle of knobs-into-holes packing to bundles containing from 4 to 14 α-helices.

The structural principles fall logically into three categories. (*a*) The crossing angle between adjoining α-helices depends on the number of α-helices in the tube and the angle between the α-helix and tube axis. (*b*) The number of α-helices in a tube is determined by which sets of side chains in a lateral plane are involved in the knobs-into-holes packing. The packing arrangement in turn specifies which sets of side chains are located to the interior and exterior of the tube. (*c*) The requirement that the same set of knobs-into-holes interactions be used over-and-over throughout the tube provides the basis for a fitting index (*f*), which is determined by the relative longitudinal packing of neighbouring α-helices.

The first category of structural principles treats the proteins as smooth cylinders; the second and third categories take into account the knobs and holes resulting from the arrangement of the side chains in the structure.

(*a*) *Packing of smooth cylinders.* In Fig. 1 on the surface of a cylinder of radius r_0 are drawn two helices L_1 and L_2 with the same pitch, representing the axes of two helically distorted smooth cylinders. A line perpendicular (on the cylindrical surface of radius r_0) to helix L_1 can be drawn from point P_1 to the point P_2 on helix L_2. The crossing angle ε between the adjacent cylinder axes is then defined by the angle between the unit-length tangents T_1 drawn through point P_1 and T_2 drawn through point P_2. If θ is the angle between points P_1 and P_3 as shown in Fig. 1, then one can use the dot product between unit vectors T_1 and T_2 to show that:

$$\cos\varepsilon = \sin^2\alpha\cdot\cos\theta + \cos^2\alpha \qquad (1)$$

where α is the pitch angle of the helices L_1 and L_2 as shown in Fig. 1.

From Fig. 1 construction shows that:

$$\theta' = \theta + (P_3P_4/r_0) \qquad (2)$$

where P_3P_4 is the line segment connecting points P_3 and P_4.

But from Fig. 1:

$$P_3P_4 = P_2P_4\sin\alpha \qquad (3)$$

but

$$P_2P_4 = P_1P_4\sin\alpha \qquad (4)$$

therefore:

$$\frac{P_3P_4}{r_0} = \frac{P_1P_4\sin^2\alpha}{r_0} \qquad (5)$$

Since

$$\frac{P_1P_4}{r_0} = \theta' \qquad (6)$$

it is evident that:

$$\theta' = \theta + \theta'\sin^2\alpha \qquad (7)$$

or

$$\theta = \theta'(1-\sin^2\alpha) = \theta'\cos^2\alpha \qquad (8)$$

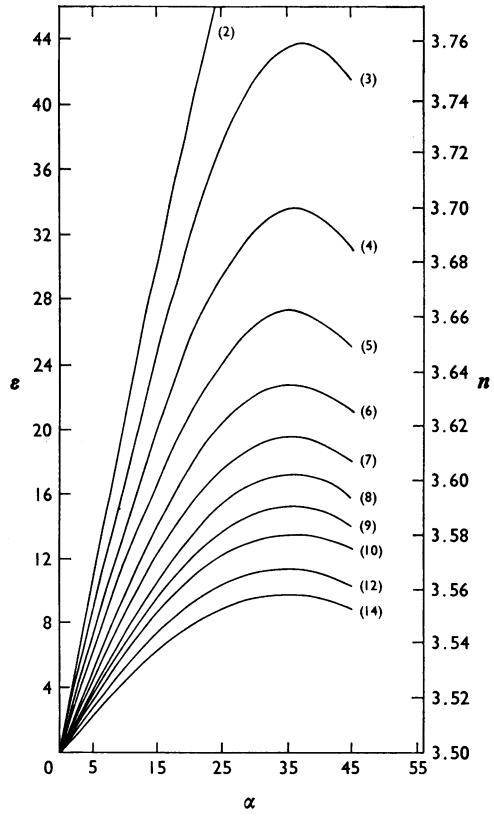

Fig. 2. *Crossing angle versus pitch angle*
Values of ε, the crossing angle of neighbouring α-helices, is plotted versus α, the pitch angle as defined in Fig. 1, for various values of S (in parentheses), the number of cylinders in the bundle.

For S equally spaced strands in the cylinder:

$$\theta' = (2\pi/S) \tag{9}$$

and therefore

$$\cos\varepsilon = \sin^2\alpha \cdot \cos(2\pi/S \cdot \cos^2\alpha) + \cos^2\alpha \tag{10}$$

By using the relationship $\cos\chi = 1-2\sin^2(\chi/2)$ and then substituting $2\pi/S\cdot\cos^2\alpha$ for χ, the above expression can be rewritten as:

$$\cos\varepsilon = 1-2\sin^2\alpha\cdot\sin^2(\pi/S\cdot\cos^2\alpha) \tag{11}$$

Thus, for a given number of strands (S) in a tube, one can graph ε versus α (Fig. 2).

A row of knobs and holes of an undistorted α-helix forms a gentle left-handed helix with an angle of about 10° with respect to the helix axis (see Crick, 1953; see also below). Crick (1953) showed that the consequence of knobs-into-holes packing would be that aggregated α-helices should form a gentle right-handed coiled-coil such that adjacent α-helices should cross at about twice the 10° pitch angle or at about 20°. From Fig. 2, it is readily evident that relatively undistorted α-helices could have crossing angles of approx. 20° only for bundles with S values of seven or below. However, if the α-helix is distorted somewhat, smaller crossing angles can be obtained. Thus, for α-helices distorted such that the row of knobs and holes lies at an angle of about 7.5° so that ε is approx. 15°, structures up to nine α-helices would be allowed, and so on. It should be evident that the same principles could be applied to protomers other than α-helices; all that is required is a repeating set of contacts forming a gentle helix about the axis of the protomer strand. Then bundles of such protomers could form with the appropriate local crossing angle.

(*b*) *Side-by-side packing considerations.* First consider the simplest case in which the α-helix axis is nearly parallel to the bundle axis, that is α approx. 0°. An α-helix contains seven equally spaced sets of rows of knobs and holes (Crick, 1953; see also below). Each row can be labelled by the residue number of the first amino acid residue in that row. Since the rows are equally spaced, the angle between adjacent rows is 360°/7 or 51.4°. The rows involved in the knobs-into-holes packing define the contact line, C_l, which can be specified by the numbers C_1, C_2, C_3, C_4, C_5, C_6, and C_7 corresponding to the seven rows of knobs and holes. Thus the angle between C_1 and C_5 is 51.4°; the angle between C_1 and C_2 is $2\times(360°/7) = 102.9°$; the angle between C_1 and C_6 is $3\times(360°/7) = 154.3°$. The angle between C_1 and C_3 is $4\times(360°/7) = 205.7°$, which is equivalent to an internal angle of $360°-205.7° = 154.3°$. The number of α-helices in the bundle can be used to define which rows of knobs will interleave simply by considering which polygon has an internal angle closest to 51.4°, 102.9° or 154.3°. These results are given in Table 1; a 14-sided polygon has an internal angle exactly the same as that between C_1 and C_6, and thus 14 was chosen as the upper limit for the present analysis. Obviously, this is an arbitrary choice since side-chain flexibility would permit an even greater number of α-helices to form knobs-into-holes packed tubular bundles. Some representative packing arrangements for α approx. 0° are shown in Fig. 3.

For example, since 90° is close to $2\times(360°/7)$, a fourfold aggregate with α approx. 0° should have knobs-into-holes packing between C_1 and C_2, which we choose to call a $C_1\times C_2$ interface. Fig. 3 also shows an alternative set of interactions in which the contact line is halfway between the lines of knobs-and-holes; these arrangements correspond to a second type of knobs-into-holes packing originally proposed by Crick (1953). Since for the fourfold aggregate the alternate packing arrangement leads to contact lines halfway between residues 1 and 5 and halfway

Table 1. *Comparison of polygon internal angles with angles between rows of knobs and holes*
For details of terminology see the text.

Number of sides of polygon	Polygon internal angle	Nearest packing interface angle	Packing interface
3	60.00	51.4	$C_1 \times C_5$ or $C_{1/5} \times C_{5/2}$
4	90.0	102.9	$C_1 \times C_2$ or $C_{1/5} \times C_{2/6}$
5	108.0	102.9	$C_1 \times C_2$ or $C_{1/5} \times C_{2/6}$
6	120.0	102.9	$C_1 \times C_2$ or $C_{1/5} \times C_{2/6}$
7	128.6	102.9	$C_1 \times C_2$ or $C_{1/5} \times C_{2/6}$
		or 154.3	or $C_1 \times C_6$ or $C_{1/5} \times C_{3/6}$
8	135.0	154.3	$C_1 \times C_6$ or $C_{1/5} \times C_{3/6}$
9	140.0	154.3	$C_1 \times C_6$ or $C_{1/5} \times C_{3/6}$
10	144.0	154.3	$C_1 \times C_6$ or $C_{1/5} \times C_{3/6}$
11	147.3	154.3	$C_1 \times C_6$ or $C_{1/5} \times C_{3/6}$
12	150.0	154.3	$C_1 \times C_6$ or $C_{1/5} \times C_{3/6}$
13	152.3	154.3	$C_1 \times C_6$ or $C_{1/5} \times C_{3/6}$
14	154.3	154.3	$C_1 \times C_6$ or $C_{1/5} \times C_{3/6}$

between residues 2 and 6 in the drawing as shown, we would denote such a packing interface as $C_{1/5} \times C_{2/6}$. For each packing interface, seven homologous packing arrangements can be generated by sevenfold rotations of each α-helix about its own axis. Thus, for example, the $C_5 \times C_6$, $C_2 \times C_3$, $C_6 \times C_7$, $C_3 \times C_4$, $C_7 \times C_1$, and $C_4 \times C_5$ are alternate possibilities for fourfold α-helical bundles of the $C_1 \times C_2$ type.

In addition to the $C_1 \times C_2$ and $C_{1/5} \times C_{2/5}$ contacts there is in theory a continuity of interactions, with contact lines anywhere between the $C_1 \times C_2$ and $C_{1/5} \times C_{2/5}$ extremes. Whether the $C_1 \times C_2$ and $C_{1/5} \times C_{2/5}$ contacts would have special stability compared with the intermediate contacts is uncertain, but for reasons of convenience in the sections below we will consider only these two limiting examples.

For the $C_1 \times C_2$ type of interaction there are three adjacent rows of interleaving knobs (a 'triple interleaving' type of packing). With more than three α-helical strands, these rows are on the outside, along the contact line, and on the inside of the bundle. Notice that for the twofold and threefold bundle there are only outside and contact rows. For the $C_{1/5} \times C_{2/6}$ type of interaction, there are just two adjacent rows of interleaving knobs (a 'double interleaving' type of packing); for structures with three or more α-helices, these are outside and inside rows; with two strands, both are outside rows.

As the bundles are drawn, the rows of knobs symmetrically to the exterior of the tube are solid coloured. We propose that in membranes these residues should be exclusively hydrophobic so that there would be no hydrogen donor or acceptor projecting directly into the lipid. The side chains indicated by an X could be either hydrophilic or hydrophobic. The key to the structure of membrane proteins may be that the lipid-adjacent interactions, which are usually confined to outside rows, but could also include the inside rows if the interior channel is filled with lipid, contain several hydrogen bonds extending from the side chains of one α-helix to the side chains of its neighbour (Dunker *et al.*, 1976; A. K. Dunker & T. Jones, unpublished work). Such hydrogen bonds would stabilize the α-helical bundle into the proper configuration. Of course there could also be important hydrophobic interactions that serve to stabilize the bundle. Thus the structure of the bundle would be determined by the amino acid sequence of the protein.

For structures with the α-helix axes inclined at angles relative to the axes of the bundles, the concepts described above also apply but with an additional complication. That is, rather than employing 51.4°, 102.9° or 154.3° as described above, the relevant angles are the projections of 51.4°, 102.9° or 154.3° on to a plane perpendicular to the axis of the bundle. The projected angles, which are equal to $51.4\cos\alpha$, $102.9\cos\alpha$, $154.3\cos\alpha$, indicate the most likely number of strands in the bundle by the same arguments as given above. Fig. 4 shows these three projected angles as functions of α. The horizontal lines on this graph are the internal angles of various polygons of up to 14 sides. If, as implied above for α approx. 0°, knobs-into-holes packing permits a range of angles between adjacent α-helices for each perpendicular packing interface, then the points of intersection between the projected angles and the lines defining the polygon internal angles would not have any special significance. That is, for example, an $S = 6$ bundle with α approx. 8° would be stereochemically feasible even though the projected angles and the polygon internal angle do not lie at a point of intersection. Of course, it is probable that, other factors being equal, structures closer to points of intersection would be preferred to structures farther from such points. Also, in a more general context, the more restrictive the geometry of interaction between two protomer strands, the closer to such points of intersection would lie the regions delimiting the allowed structures.

(*c*) *Longitudinal packing considerations.* A number of α-helices with axes on the surface of a cylinder will form a regular tube if every α-helix has the same packing environment; the same sets of knobs and holes must be used over-and-over. Therefore there must exist a symmetry axis coincident with the axis of the cylinder. If the S strands are all identical, the symmetry axis can be used to generate the whole tube by repeated operations of an S-fold screw on one α-helix. In other words, if the strands are all identical, the packing relationships between strands $(t+1)$ and $(t+2)$ must be the same as that between the strands t and $(t+1)$; also, the relationship

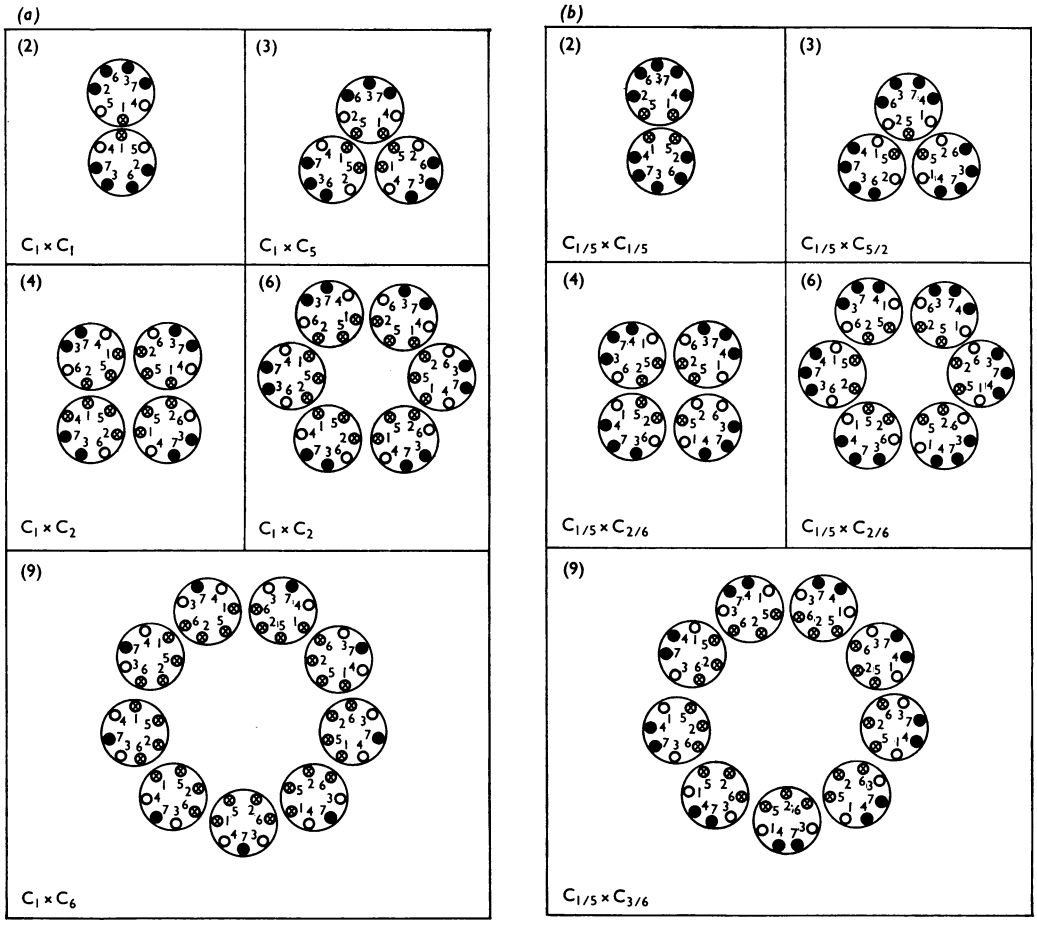

Fig. 3. *Packing interfaces between neighbouring α-helices*

Various clusters of α-helices (values in parentheses), viewed from the *N*-terminal end are shown. The first seven residue numbers are indicated, for details of shading see the text. In (*a*) the contact lines are directly through rows of knobs and holes; in (*b*) the contact lines lie half-way between adjacent rows of knobs and holes, corresponding to the two types of packing originally proposed by Crick (1953).

between strand S and strand 1, as one closes the cylindrical surface, must be the same as that between strands t and $(t+1)$. If the tube contains a non-identical strands, then the requirement for repeated use of the same bonding interfaces will be satisfied only if there exists an S/a-fold screw axis; therefore S/a must equal an integral value, including 1. In this limiting case, the bundle would be made up of S non-identical strands.

Repeated application of a screw operator usually results in very long structures. Thus, for the short segments of α-helical bundles appropriate for membrane bilayers, the polypeptide segments should probably be related by an S-fold axis of rotation rather than by an S-fold screw. Obviously, the foregoing comments are merely suggestive and not rigorous and so imposing a requirement for rotational rather than screw symmetry must be regarded as somewhat arbitrary. Nevertheless, for convenience, the additional requirement for rotational symmetry will be imposed here and in the following sections of this paper.

Generation of self-consistent structures

Development of the quantitative rules for the restrictions imposed by the requirement for symmetrical packing and then combining the results with

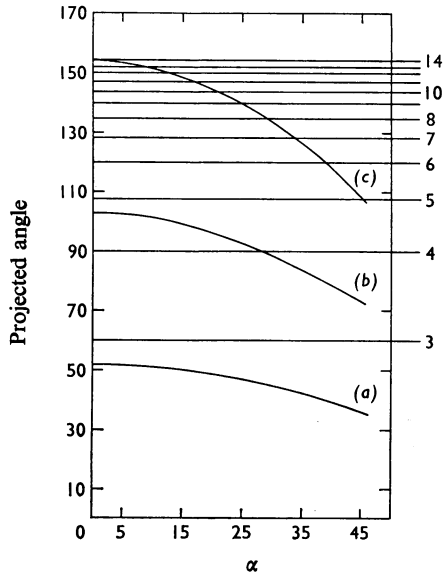

Fig. 4. *Projected angle versus pitch angle*
The projections of the three optimum internal angles are 51.4 cos α (curve *a*), 102.9 cos α (curve *b*), and 154.3 cos α (curve *c*). The horizontal lines numbered from 3 to 14 indicate the various internal angles of the corresponding polygons, having from 3 to 14 sides.

the two categories of structural principles outlined above to generate self-consistent structures is a fairly complex task. To simplify the explanation of this process, the overall scheme is presented in Fig. 5.

Step (1). Requiring regular packing between the α-helices leads to equations for α (the pitch angle of the α-helices in the bundles) in terms of n (the number of units per turn of the α-helix), P (the pitch of the α-helix), S (the number of α-helices in the bundle), f (a longitudinal packing index), a (the number of distinguishable sequences in the bundle) and r_c (the average packing radii of the α-helices). Thus, assuming values for n, P, S, f, a, and r_c, a value for α can be calculated.

Step (2). The smooth-cylinder-packing equation is then used to estimate ε, the crossing angle, from the values of α and S.

Step (3). Crick's (1953) hypothesis for knobs-into-holes packing is generalized so that the crossing angle ε can be used to specify a new n value for the α-helix. If the new value, $n_{(1+m)}$, is the same as the assumed value, n_m, then a self-consistent structure has been generated.

Step (4). When $n_{(m+1)}$ does not equal n_m, the structure is not self-consistent. In this case the equations developed and used previously by many workers (Ramakrishnan & Ramachandran, 1965; Schellman & Schellman, 1964; Arnott & Wonacott, 1966; Dickerson & Geis, 1969) provide a new value for the pitch, $P_{(1+m)}$. The $n_{(1+m)}$, $P_{(1+m)}$ values are then used to initiate reiterations of steps (1), (2), (3) and (4), until $n_{(1+m)}$ approx. equals $n_{(m)}$. A few examples of tubes generated by this algorithm are given in Table 2.

Determination of α from the packing considerations

For purposes of illustration, we will derive the equations for α for the tube with $C_1 \times C_2$, $S = 6$; the equations for the other tubes were determined in a similar fashion and are presented in Tables 3 and 4.

The radial projection of a tube with S strands is shown in Fig. 6 for $S = 6$.

By construction:

$$\tan \alpha = h/l \qquad (12)$$

For $\alpha = 0°$, $l = 2\pi r_0$, where r_0 as before is the radius of the bundle. It is simple to show that $r_0 = r_c/\sin(180/S)$, where r_c is the packing radius of the individual α-helices. Therefore, for $\alpha = 0°$, $l = 2\pi r_c/\sin(180/S)$. As α varies from zero, the curvilinear length of l remains very nearly constant.

For $C_1 \times C_2$, positioning the knobs symmetrically into the holes requires that strand S_{n+1} must be raised by an increment (x) relative to strand S_n. (For certain other contact with faces, the S_{n+1} strand must be lowered relative to strand S_n.) From Fig. 6 it is evident that h must obey the equation $h = (K2P - Sx) + 2Pf'$, where K is a fixed integer such that $K2P$ is the smallest possible value greater than Sx, and f' can take on any integral value. The value for x can be determined by positioning two α-helixes next to each other in the proper relative orientation; for $C_1 \times C_2$ packing, it then becomes evident that $x = P(1-1/n)$. For $n = 3.6$ and $S = 6$, $Sx = 4.33P$; since $2 \times 2P < 4.33P$ and $3 \times 3P > 4.33P$, it is determined that $K = 3$ (for $S = 4$ or 5, $K = 2$; for $S = 6$ or 7, $K = 3$). Thus by substituting for K and x and rearranging terms:

$$\tan \alpha = \frac{P[2(f'+3) - S(1-1/n)]}{2\pi r_c/\sin(180/S)} \qquad (13)$$

This derivation assumes that the α-helices are identical and have identical knobs and holes along their lengths. In general the second condition is not true for real proteins. We argued above that for a tube containing S α-helices with a distinct sequences, there should be an (S/a)-fold symmetry axis, thus, with $a = 1$, there should be a 6-fold symmetry axis; with $a = 2$, there should be a 3-fold symmetry axis, with $a = 3$, there should be a 2-fold symmetry axis. The $f' = -3$ structure in Fig. 6(*c*) has 2-, 3- and 6-fold symmetry axes and thus would be suitable for $a = 1$, 2, 3 or 6. Fig. 6(*b*) shows that the $f' = 0$ structure has a

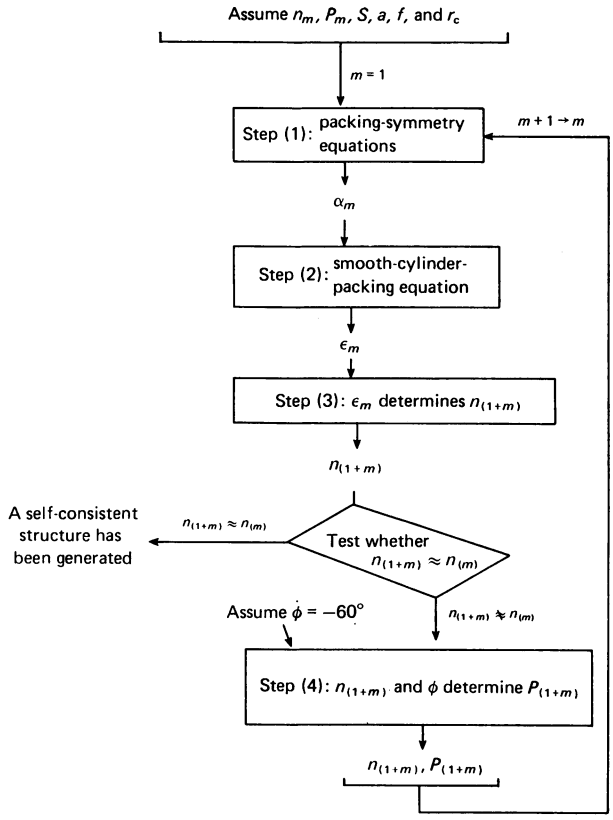

Fig. 5. *Schematic representation of the method used to generate self-consistent structures* For detailed explanation see the text.

3-fold axis, suitable for $a = 2$, and Fig. 6(*a*) shows that the $f' = +1$ has a 2-fold axis, suitable for $a = 3$. From these particular examples, it can be determined that the allowed f' values are given by the expression:

$$f' = -3 + f(S/a) \qquad (14)$$

where f is any integer.

The expressions defining the allowed f' values for the various structures are given in Table 4. The importance of f is that f defines which knobs pack into which holes along the length of the α-helices. For example, for $f = 0$ with the $C_1 \times C_2$ packing interface, the side chain of residue 8 on one helix is in the space defined by residues 2, 5, 6 and 9 of its neighbour; for $f = 1$, the relative positions of the α-helices are shifted by one set of knobs and holes with the result that the side chain of residues 8 is in the space defined by 9, 12, 13 and 16.

Determination of ε from α

From the equations given in Tables 3 and 4, the angle α is determined once n_m, P_m, S, a, f, and r_c are specified. The next step, determining ε from α, is accomplished analytically by using eqn. (11), or graphically from Fig. 2.

Determination of n from ε

Fig. 7 is a radial projection of an α-helix (viewed from the outside). Crick (1953) showed that the condition for knobs-into-holes packing is that the crossing angle ε is equal to twice the value of the angle given as β in Fig. 7.

By construction:

$$\tan \beta = y/g \qquad (15)$$

but since the rise/residue is simply P/n, $g = 7P/n$. The horizontal displacement/residue is simply $2\pi r_c/n$;

Table 2. *Representative tubes*

The values of α are to the nearest 0.1°; the values of inside diameter and outside diameter to the nearest 0.01 nm (0.1 Å). The values for α were calculated by using the equations given in Table 3. Inside diameter and outside diameter were calculated from the relations:

$$\text{Inside diam.} = 2r_0 - 2r_c$$
$$\text{Outside diam.} = 2r_0 + 2r_c,$$

where r_0 is the radius of the tube as defined by the axes of the α-helices, and r_c is the packing radius of the α-helix, taken to be 0.52 nm (5.2 Å) for these calculations.

By construction from Fig. 6:

$$\cos\alpha = \frac{2\pi r_c/\sin(180/S)}{2\pi r_0}$$

therefore

$$2r_0 = \frac{2r_c/\sin(180/S)}{\cos\alpha}$$

The values for n are to the nearest 0.05; the values for P, to the nearest 0.002 nm (0.02 Å), and the values for ψ to the nearest 0.5°. Assuming $\phi = -60°$, P and ψ were estimated from graphs of P versus n and ψ versus n constructed from the equations below (Fraser & MacRae, 1973).

$$P = n \cdot d$$
$$n = \frac{360}{\theta}$$
$$\cos(\theta/2) = -0.817\sin[(\phi+\psi)/2] + 0.045\sin[(\phi-\psi)/2]$$
$$d\sin(\theta/2) = +2.967\cos[(\theta+\psi)/2] - 0.664\cos[(\phi-\psi)/2]$$

	Parameters of the bundle					Parameters of the α-helix		
Packing interface	S	f'	α	Inside diam.	Outside diam.	n	P	ψ
$C_1 \times C_1$	2	−1	0	0	20.8	3.50	5.52	−41.0
	2	0	9.5	0.1	20.9	3.60	5.40	−45.5
$C_1 \times C_5$	3	0	3.5	1.6	22.4	3.55	5.46	−43.0
	3	1	18.6	2.3	23.0	3.70	5.26	−48.5
$C_1 \times C_2$	4	−1	−5.8	4.4	25.2	3.45	5.58	−39.5
	4	0	7.5	4.4	25.2	3.55	5.44	−43.5
$C_1 \times C_5$	5	−0	2.4	7.3	28.1	3.50	5.52	−42.0
	5	+1	13.1	7.7	28.5	3.60	5.40	−45.5
$C_1 \times C_5$	6	−1	−1.4	10.4	31.2	3.50	5.52	−41.0
	6	0	7.6	10.6	31.4	3.55	5.46	−43.0
$C_1 \times C_5$	7a	−2	−12.5	14.2	35.0	3.45	5.60	−38.5
$C_1 \times C_6$	7b	0	12.0	14.1	34.9	3.55	5.44	−43.5
$C_1 \times C_6$	8	0	12.2	17.4	38.2	3.55	5.44	−43.5
	8	2	24.9	19.5	40.3	3.60	5.40	−45.0
$C_1 \times C_6$	10	0	12.4	24.0	44.8	3.55	5.46	−43.0
	10	2	22.8	26.1	46.9	3.70	5.42	−44.0
$C_1 \times C_6$	12	0	12.5	30.7	51.5	3.55	5.46	−43.0
	12	2	21.3	32.7	53.5	3.55	5.44	−43.5
$C_1 \times C_6$	14	0	12.5	37.5	58.3	3.55	5.48	−42.6
	14	2	20.3	39.4	60.2	3.55	5.46	−43.0

thus the sum of the horizontal displacements is $7(2\pi r_c/n)$. But y is given by the difference between this value and twice the circumference, or $y = 2\times 2\pi r_c - 7(2\pi r_c/n)$. Thus by substituting for y and g and reorganizing:

$$\varepsilon = 2\tan^{-1}\frac{4\pi r_c(1-7/2n)}{7P/n} \qquad (16)$$

Pairs of n and P values were calculated for the allowed ranges for the α-helix by using formulae developed previously (Ramakrishnan & Ramachandran, 1965; Schellman & Schellman, 1964; Edsall *et al.*, 1966*a*,*b*; Fraser & MacRae, 1973; the formulae are given in Table 2). From these pairs of n and P, eqn. (16) was used to plot ε versus n; it was found that, to a very good first approximation, ε

Table 3. *Equations specifying α*

Equations for α were determined by construction as outlined in Fig. 5; $l = 2\pi r_c/\sin(180/S)$. Range of α is from Fig. 4. It is apparent that a given contact interface will be appropriate for a given S-fold aggregate only for certain values of α. Although there is no clear-cut division, to a first approximation low values for α refer to approx. 0–20°, medium values refer to approx. 15–35°, and high values refer to approx. 30–45°.

Packing interface	S	Equations for α	Range of α
$C_1 \times C_1$ and $C_{1/5} \times C_{1/5}$	2	$\alpha = \tan^{-1} \frac{P(f'+1)}{l}$	All
$C_1 \times C_5$	3	$\alpha = \tan^{-1} \frac{P(2f'+\frac{1}{2}S/n)}{l}$	All
$C_{1/5} \times C_{5/2}$	3	$\alpha = \tan^{-1} \frac{P[2(f'+2)-S(1-1/2n)]}{l}$	All
$C_1 \times C_2$	4, 5	$\alpha = \tan^{-1} \frac{P[2(f'+2)-S(1-1/n)]}{l}$	High
$C_1 \times C_2$	6, 7	$\alpha = \tan^{-1} \frac{P[2(f'+3)-S(1-1/n)]}{l}$	Low, medium
$C_{1/5} \times C_{2/6}$	4–7	$\alpha = \tan^{-1} \frac{P(2f'+S/n)}{l}$	All
$C_1 \times C_6$	7–14	$\alpha = \tan^{-1} \frac{P(2f'+\frac{3}{2}S/n)}{l}$	High, medium
$C_{1/5} \times C_{3/6}$	7–10	$\alpha = \tan^{-1} \frac{P[2(f'+3)-S(1-\frac{3}{2}S/n)]}{l}$	Low, medium
$C_{1/5} \times C_{3/6}$	11–13	$\alpha = \tan^{-1} \frac{P[2(f'+4)-S(1-\frac{3}{2}S/n)]}{l}$	Low
$C_{1/5} \times C_{3/6}$	14	$\alpha = \tan^{-1} \frac{P[2(f'+5)-S(1-\frac{3}{2}S/n)]}{l}$	Low, medium

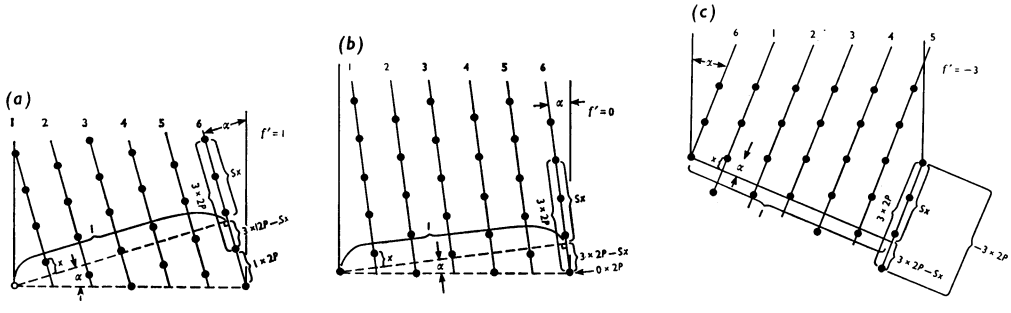

Fig. 6. *Radial projections of $S = 6$ tubes of α-helices*

If the cylinder on which lie the axes of the α-helices is cut parallel to the cylinder axis and the cylinder is opened and laid flat, a radial projection results; we are using the convention that the view is from the outside. The six numbered lines correspond to the axes of the six α-helices. The various symbols are explained in the text. The dots represent the spacings of every seventh residue along the axis. (*a*) $f' = 1$ packing; (*b*) $f' = 0$; (*c*) $f' = -3$.

depends on r_c and n only (Fig. 8). This is particularly significant because it allows a test of whether a given structure, with a definite ε, has a value for n which is within the range allowed for α-helices.

From several proteins of known structure, α-helical stretches of more than 3 turns have n values that range from 3.4 to about 3.8 (R. Henderson, personal communication). Also, this range for n is in fair agreement with that predicted by theory (Ramakrishnan & Ramachandran, 1965; Fraser & MacRae, 1973).

Given a choice for S, f, a and r_c, it turns out that α

Table 4. *Selection rules for f'*
For details see the text.

Contact	S	Selection rule
$C_1 \times C_1$	2	$f' = -1 + f(S/a)$
$C_{1/2} \times C_{1/2}$	2	$f' = -1 + f(S/a)$
$C_1 \times C_5$	3	$f' = f(S/a)$
$C_{1/5} \times C_{5/2}$	3	$f' = -2 + f(S/a)$
$C_1 \times C_2$	3	$f' = -1 + f(S/a)$
$C_{1/5} \times C_{2/6}$	3–7	$f' = f(S/a)$
$C_1 \times C_2$	4, 5	$f' = -2 + f(S/a)$
$C_1 \times C_2$	6, 7	$f' = -3 + f(S/a)$
$C_{1/5} \times C_{3/6}$	4–6	$f' = -2 + f(S/a)$
$C_1 \times C_6$	4–14	$f' = f(S/a)$
$C_{1/5} \times C_{3/6}$	7–10	$f' = -3 + f(S/a)$
$C_{1/5} \times C_{3/6}$	11–13	$f' = -4 + f(S/a)$
$C_{1/5} \times C_{3/6}$	14	$f' = -5 + f(S/a)$

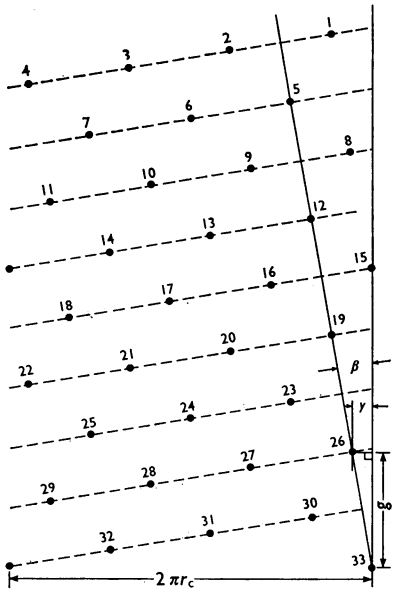

Fig. 7. *Radial projection of an α-helix*
By using the rules set down in Fig. 6, a radial projection of an α-helix is drawn. The broken lines connect successive residues, which are indicated by circles. Note that a 'hole' is the space defined by residues m, $m+3$, $m+4$ and $m+7$. The symbols are explained in the text.

changes only 1% or so for changes in n and P; thus one can use any allowed values for n and P along with the choices for the other parameters to determine an approximate α, which specifies ε, which in turn specifies an approximate value for n. Since further refinement changes this n only slightly the approximate value can be used to test the feasibility of the

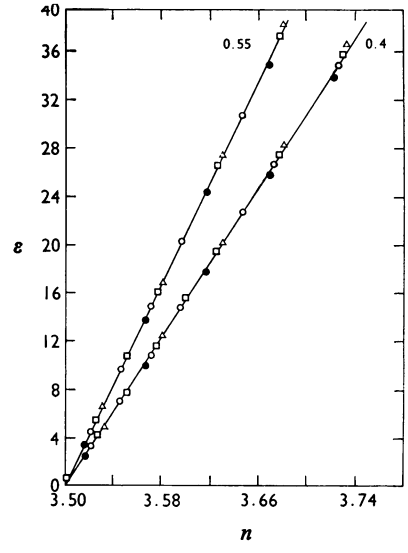

Fig. 8. *Crossing angle as a function of number of units per turn*
The calculations leading to this graph are explained in the text. The various points were calculated for constant ϕ as ψ varied; the symbols are $\phi = -70°$ (△); $\phi = -65°$, (□); $\phi = -60°$ (○); $\phi = -55°$ (●). The numbers 0.4 and 0.55 refer in nm to different choices for r_c. For a given choice of r_c, ε versus n is actually described by a family of lines that have slopes depending on ϕ and that all pass through the point $\varepsilon = 0$, $n = 3.50$. However, since the change in slope for these lines is very small for ϕ between $-70°$ and $-55°$, this family of lines can be replaced by a single line without significant error. For example, when $\varepsilon = 35°$, n changes by about 0.01 unit/turn as ϕ varies from $-70°$ to $-55°$, and for smaller values of ε, the uncertainty is even less.

proposed bundle. If $3.4 < n < 3.8$, the bundle specified by S, f, a, and r_c would be allowed; if $3.4 > n$ or if $3.8 < n$, the bundle would be forbidden.

Notice also that for $3.4 < n < 3.5$ the α-helix is deformed to form a right-handed coiled-coil; the lines of Fig. 6 can be extrapolated to determine ε versus n for this range; likewise eqn. (11) and Fig. 1 can also be extended to these regions.

Determination of P from n

Given a value for n, the equations given in Table 2 can be used to estimate the pitch; since there are four equations relating to six variables, P is determined only if two variables are specified. It is convenient to specify ϕ arbitrarily in addition to the value for n [ψ and ϕ are used as defined in Fraser & MacRae (1973)]. By arbitrarily specifying ϕ, however, n decreases while P increases (see, e.g., Dickerson &

Geiss, 1969, p. 28); this results in rather extreme values for d, the rise per residue, whenever the values for n approach upper and lower limits. This could be corrected by optimizing the α-helix stereochemistry rather than just arbitrarily specifying ϕ; however, for reasons of simplicity and also for the reasons given below, we have for the present chosen the much easier option of simply specifying ϕ.

The equations in Table 4 were derived for straight α-helices, but the α-helices in the coiled-coils are gently curved. Errors arising from this discrepancy are minimized if the number of units/turn and pitch of the α-helix are made to refer to a curved frame of reference based on the α-helix axis; the values of ϕ and ψ are then found to vary systematically as the residue of interest is chosen from the inside or outside of the bundle.

Thus for the two reasons just presented the structures generated by the algorithm presented in Fig. 5 are not fully optimized. If required in any particular application, the α-helical bundles could be further optimized by methods described previously (Diamond, 1966).

Discussion

We regard the structures derived in the present paper only as guides; they demonstrate that given the structural parameters of an α-helix, it is possible to derive a set of tubular structures with knobs-into-holes packing. Even if our approach is essentially correct, for at least two reasons, real structures in membranes can be anticipated to be slightly different from our structures.

First, the parameters for our structures were in effect derived for infinitely long α-helices, for which even a slight mismatch between bundle and α-helix parameters would lead to impossibly close contacts between atoms if one travelled far enough along the helix. However, membrane bilayers are only on the order of 5–5.5 nm (50–55 Å) thick (Levine & Wilkins, 1971; Casper & Kirshner, 1971) and thus one needs good packing over a short segment of only about 10–12 turns of the α-helix. Thus, even if the actual crossing angle ε does not mesh precisely with that prescribed by the stereochemistry of the α-helix in the bundle, the knobs-into-holes interactions could still be acceptable over such a short length. This point is illustrated by the work of Inouye (1974). By using CPK molecular models of murein-lipoprotein in an α-helical conformation, Inouye (1974) studied the interactions between two α-helices positioned in space so as to be members of a bundle containing six α-helices. From the interactions described in Fig. 1 of his paper, we deduce his packing arrangement to be $C_1 \times C_6$, $f = 0$; the pitch angle he used was +25°.

First, the $C_1 \times C_6$ interface would be appropriate for an $S = 6$ bundle only for large values of α, but the choice of $f = 0$ for this packing interface leads to a small value for α. Therefore, the packing chosen by Inouye (1974) is not suitable for an $S = 6$ bundle. Secondly, for various bundles containing from 8–14 α-helices, bundles that could employ the $C_1 \times C_6$, $f = 0$ packing scheme, the equations of Tables 3 and 4 lead to values of α near 12°, not 25°. Further, from Fig. 2, the appropriate α-helices would contain from about 3.55–3.53 units/turn for bundles with 8–14 α-helices.

But despite the substantial deviation from the self-consistent structures, in the packing actually observed in the CPK models (Inouye, 1974), the knobs were quite close to the appropriate holes. This observation has special importance for structures near the theoretical limits, since it suggests that the strain on the α-helices could be relieved somewhat by allowing a slight mismatch between the parameters of the bundle and the parameters of the α-helices.

A second reason for anticipating slight differences from our structures is that along the length of real α-helices there are variations in the amino acid side chains which in turn could lead to concomitant longitudinal variations in the crossing angle and separation of the neighbouring α-helices. For example, in the current best structure for the filamentous bacterial virus, the crossing angle for adjacent α-helices varied from 9° to 12° and the packing distance $2r_c$ varied from 0.95 to 1.05 nm (9.5 to 10.5 Å) (Marvin & Wachtel, 1975). Also, for bacterial rhodopsin, crossing angles from 0° to 20° and $2r_c$ values from 0.9 to 1.2 nm (9 to 12 Å) were observed although the larger $2r_c$ values may indicate helices that are not in contact (Henderson & Unwin, 1975).

It is interesting to compare our predicted structures with that observed for bacterial rhodopsin. Each bacterial rhodopsin contains seven membrane spanning α-helical segments; three of the segments are approximately perpendicular to the membrane, the remaining four form a fan-shaped aggregate. Together the seven α-helices form an irregular bundle surrounding an irregular pore (Unwin & Henderson, 1975; Henderson & Unwin, 1975). Such aggregates are very different from the kinds of structures predicted here. This is not surprising since the assumption underlying our work is that the α-helices would be in identical environments, but there would be no requirement for the seven distinct α-helical segments to be in identical environments. On the other hand, three bacterial rhodopsin molecules aggregate via interactions between the three almost perpendicular segments; the result is a nine-membered bundle that is similar to, but not identical with, the tubular bundles proposed here. The nine-membered bundle contains three distinct internal angles, each centered on one of the three non-identical chains. The three internal angles are

Table 5. *Possible packing for $S = 9$, $a = 3$ bundles containing the specified non-homologous contact interfaces*
For details see the text.

Type of packing	Contact interface	Theoretical internal angle	Observed internal angle	Total number per structure
Triple interleaving	$C_1 \times C_6$	154.3°	158°	3
	$C_1 \times C_6$	154.3	159	3
	$C_1 \times C_2$	102.9	103	3
Double interleaving	$C_{1/5} \times C_{3/6}$	154.3	158	3
	$C_{1/5} \times C_{3/6}$	154.3	159	3
	$C_{1/5} \times C_{2/6}$	102.9	103	3

Calculated values for α

Type of packing	f	f'	α
Triple interleaving	−1	−3	−18.5°
	0	0	+1.5
	+1	+3	+21.1
Double interleaving	0	−2	−8.7
	1	+1	+11.6
	2	+4	+29.5

approximately 103°±1°, 158°±1°, and 159°±1°; these values are averages of three measurements from Plate IV(*a*) of Unwin & Henderson (1975). These values agree with those predicted for knobs-into-holes packed α-helices. Unlike our simplified structures, however, these angles imply that the non-identical chains use non-homologous contact interfaces. We had purposely ignored such a possibility in order to decrease to a manageable size the number of structures to be considered. It would be very straightforward to modify our derivations to take into account the general possibility of such structures, but for the present it seems sufficient to present the necessary modifications to fit the particular case of bacterial rhodopsin. For this particular example, the defining equation for α is given by:

$$\tan\alpha = \frac{\sum_{i=1}^{9} X_i + k2P + 2Pf'}{l}$$

where $k = 0$ if $\sum_{i=1}^{9} X_i > 0$, but

$$|k2P| - |\sum_{i=1}^{9} X_i| > 0 \text{ if } \sum_{i=1}^{9} X_i < 0 \quad \text{(see Fig. 6)}$$

If the bundle were a perfect cylinder, then the value for l would be defined by $l = 2\pi c_c/\sin(180/9)$; it turns out that insignificant error is introduced by using this relationship even though the bundle is not a perfect cylinder. It is relatively easy to show that, from the values of the internal angles, the contact interfaces are either all double interleaving or all triple interleaving. By substituting the appropriate values of x_i for these combinations of contact interfaces, carrying out the summation and solving the k values gives the following results. For triple-interleaving interfaces, the relationship for α is given by:

$$\tan\alpha = \frac{-3P + 12P/n + 2Pf'}{l}$$

where $f' = 3f$.
For double-interleaving interfaces, the relationship for α is given by:

$$\tan\alpha = \frac{-2P + 12P/n + 2Pf'}{l}$$

where $f' = -2 + 3f$.

Table 5 presents several calculated values for α as defined for different f and f' values. Since the observed value for α ranges from about 2–3° (R. Henderson, personal communication) it is evident that only one of the packing arrangements (e.g. triple-interleaving interfaces with f, $f' = 0$, α approx. 1.5) seems to be appropriate for the central bundle in the bacterial rhodopsin structure.

We wish to point out that the derivation leading to the results of Table 5 does not consider the possibility of contacts between the extremes represented by double-interleaving and triple-interleaving contact interfaces. The agreement between the observed structure and our calculations suggests that such an omission was warranted in this particular example.

The agreement between the theoretical internal angles and the observed internal angles and between the theoretical and observed values for α are remarkable. This may be entirely coincidental or it may suggest that our packing hypothesis does indeed account for certain features of an observed membrane protein.

Superficially any α-helix would seem much like any other, and thus one would argue that although there may be bundles of α-helices in the membrane, such simple structures cannot account for the various functions of membrane proteins. However, we wish to point out that about 35 amino acids are required for an α-helix to cross the membrane, and thus with 19 amino acids (excluding proline), about 19^{35} different permutations of sequence would be possible. Since the sequence would determine the number of α-helices in the bundle and the character of the interior pore, it would seem that there is more than sufficient polymorphism inherent in such models to account for the various protein-mediated membrane functions, including the many different types of membrane transport.

This work is the direct outgrowth of unpublished work (e.g. Figs. 1, 2, and 6*a*) of Dr. D. A. Marvin, whose constructive comments greatly improved this manuscript. This work was completed in outline while A. K. D. was a postdoctoral fellow in the Department of Molecular Biophysics, Yale University, 1970–1972, in the laboratory of Dr. D. A. Marvin.

References

Arnott, S. & Wonacott, A. J. (1966) *J. Mol. Biol.* **21**, 371–383

Astbury, W. T. & Street, A. (1931) *Philos. Trans. R. Soc. London Ser. A* **230**, 75–101

Bragg, W. L., Kendrew, J. C. & Perutz, M. F. (1950) *Proc. R. Soc. London Ser. A* **203**, 321–357

Casper, D. L. D. & Kirshner, D. A. (1971) *Nature (London) New Biol.* **231**, 46–52

Crick, F. H. C. (1952) *Nature (London)* **176**, 915–916

Crick, F. H. C. (1953) *Acta Crystallogr.* **6**, 689–697

Diamond, R. (1966) *Acta Crystallogr.* **21**, 253–266

Dickerson, R. E. & Geis, I. (1969) *The Structure and Action of Protein*, Harper and Row, New York

Dunker, A. K., Marvin, D. A., Zaleske, D. J. & Jones, T. (1976) *Biophys. J.* **16**, 102a

Edsall, J. T., Flory, J. C., Kendrew, J. C., Liquori, A. M., Nemethy, G., Ramachandran, G. N. & Scheraga, H. A. (1966*a*) *J. Mol. Biol.* **15**, 399–407

Edsall, J. T., Flory, J. C., Kendrew, J. C., Liquori, A. M., Nemethy, G., Ramachandran, G. N. & Scheraga, H. A. (1966*b*) *J. Mol. Biol.* **20**, 589

Fraser, R. D. B. & MacRae, T. P. (1973) *Conformation in Fibrous Proteins*, Academic Press, New York

Frey, T. G., Eisenberg, D. & Eiserling, F. A. (1975) *Proc. Natl. Acad. Sci. U.S.A.* **72**, 3402–3406

Harrison, R. G., Lowey, S. & Cohen, C. (1971) *J. Mol. Biol.* **59**, 531–535

Henderson, R. (1975) *J. Mol. Biol.* **93**, 123–128

Henderson, R. & Unwin, P. N. T. (1975) *Nature (London)* **257**, 28–32

Huxley, H. E. & Brown, W. (1967) *J. Mol. Biol.* **30**, 383–434

Inouye, M. (1974) *Proc. Natl. Acad. Sci. U.S.A.* **71**, 2396–2400

Jarosch, R. (1969) *Z. Naturforsch.* **56**, 1828–1835

Lenard, J. & Singer, S. J. (1966) *Proc. Natl. Acad. Sci. U.S.A.* **56**, 1828–1835

Levine, Y. K. & Wilkins, M. H. F. (1971) *Nature (London) New Biol.* **230**, 69–72

Marvin, D. A. & Wachtel, E. (1975) *Nature (London)* **253**, 19–23

Neville, D. M. (1969) *Biochem. Biophys. Res. Commun.* **34**, 60–64

Pauling, L. & Corey, R. B. (1953) *Nature (London)* **171**, 59–61

Pauling, L., Corey, R. B. & Branson, H. R. (1951) *Proc. Natl. Acad. Sci. U.S.A.* **37**, 205–211

Pepe, F. A. (1967) *J. Mol. Biol.* **27**, 203–225

Pepe, F. A. (1971) *Prog. Biophys. Mol. Biol.* **22**, 77–96

Ramakrishnan, C. & Ramachandran, G. N. (1965) *Biophys. J.* **5**, 909–933

Rich, A. & Crick, F. H. C. (1955) *Nature (London)* **176**, 915–916

Rich, A. & Crick, F. H. C. (1961) *J. Mol. Biol.* **3**, 483–506

Schellman, J. A. & Schellman, C. (1964) in *The Proteins* (Neurath, H., ed.), vol. 2, pp. 1–35, Academic Press, New York

Unwin, P. N. T. & Henderson, R. (1975) *J. Mol. Biol.* **94**, 425–440

Wallach, D. F. H. & Zahler, P. H. (1966) *Proc. Natl. Acad. Sci. U.S.A.* **56**, 1552–1559