doi:10.1006/jmbi.2001.4900 available online at http://www.idealibrary.com on IDEAL® *J. Mol. Biol.* (2001) **311**, 1081–1090

JMB

D_n-Symmetrical Tertiary Templates for the Design of Tubular Proteins

B. North, C. M. Summa, G. Ghirlanda and W. F. DeGrado*

The Johnson Research Foundation, Department of Biochemistry & Biophysics School of Medicine, University of Pennsylvania, Philadelphia PA 19104-6059, USA

*Corresponding author

Antiparallel helical bundles are found in a wide range of proteins. Often, four-helical bundles form tube-like structures, with binding sites for substrates or cofactors near their centers. For example, a transmembrane four-helical bundle in cytochrome bc_1 binds a pair of porphyrins in an elongated central cavity running down the center of the structure. Antiparallel helical barrels with larger diameters are found in the crystal structures of TolC and DSD, which form antiparallel 12-helical and six-helical bundles, respectively. The backbone geometries of the helical bundles of cytochrome bc_1, TolC, and DSD are well described using a simple D_n-symmetrical model with only eight adjustable parameters. This parameterization provides an excellent starting point for construction of minimal models of these proteins as well as the *de novo* design of proteins with novel functions.

Keywords: de novo protein design; coiled coil; α-helical barrel; TolC, cytochrome bc_1; domain-swapped dimer

Introduction

De novo protein design provides an attractive approach to probe the features required for the folding and function of proteins. Recently, much work in this area has focussed on the design of parallel coiled coils,[1,2] which comprise a simple oligomeric folding motif that is frequently found in fibrous proteins as well as the oligomerization domains of globular structures. Parallel coiled coils have approximate or exact C_n symmetry, in which n parallel helices are arranged about a central superhelical axis. A number of C_n-symmetrical coiled coils have been designed and characterized crystallographically, including dimeric, trimeric, and tetrameric bundles of helices.[3–8] Antiparallel helical bundles represent a second class of folding motifs, which are functionally more diverse than parallel coiled coils. For example, antiparallel three-helix[9] and four-helical bundles[10–12] are found in hormones, cytokines, transcription factors, RNA-binding proteins and enzymes. Higher-order antiparallel helical bundles are present in proteins such as TolC, whose structure includes an elongated 12-helical bundle.[13,14]

An important step in *de novo* protein design is the development of a mathematical model that defines the backbone conformation of the structure.[6,15–17] While the parameterization of coiled coils has been shown to be very useful for the design of parallel coiled coils,[6,8,15,18] less has been accomplished in the area of antiparallel helical bundles. We recently developed a parameterization of antiparallel four-helical bundles, which provided an excellent starting point for the design of diiron proteins.[19,20] Here, we extend this parameterization to include a variety of structures that conform to D_n symmetry. D_n-symmetrical bundles have an n-fold rotational symmetry axis directed through the center of the bundle, as well as n 2-fold axes of rotational symmetry lying orthoganol to the bundle axis (Figure 1). Because of the orthoganol 2-fold axes, neighboring helices in D_n-symmetrical structures are oriented antiparallel to one another. Thus, for example, the application of a C_4 symmetry operator to a single helix would result in a parallel four-helix bundle, while the application of a D_2 symmetry operator results in an antiparallel four-helix bundle. Here, we consider three natural proteins whose structures show approximate D_2, D_3, and D_6 symmetry. In each case, the proteins have a cavity at the center of the structure that is capable of accommodating cofactors or substrates. Thus, these D_n-symmetrical bundles have the form of a tube.

E-mail address of the corresponding author: wdegrado@mail.med.upenn.edu

0022-2836/01/051081–10 $35.00/0

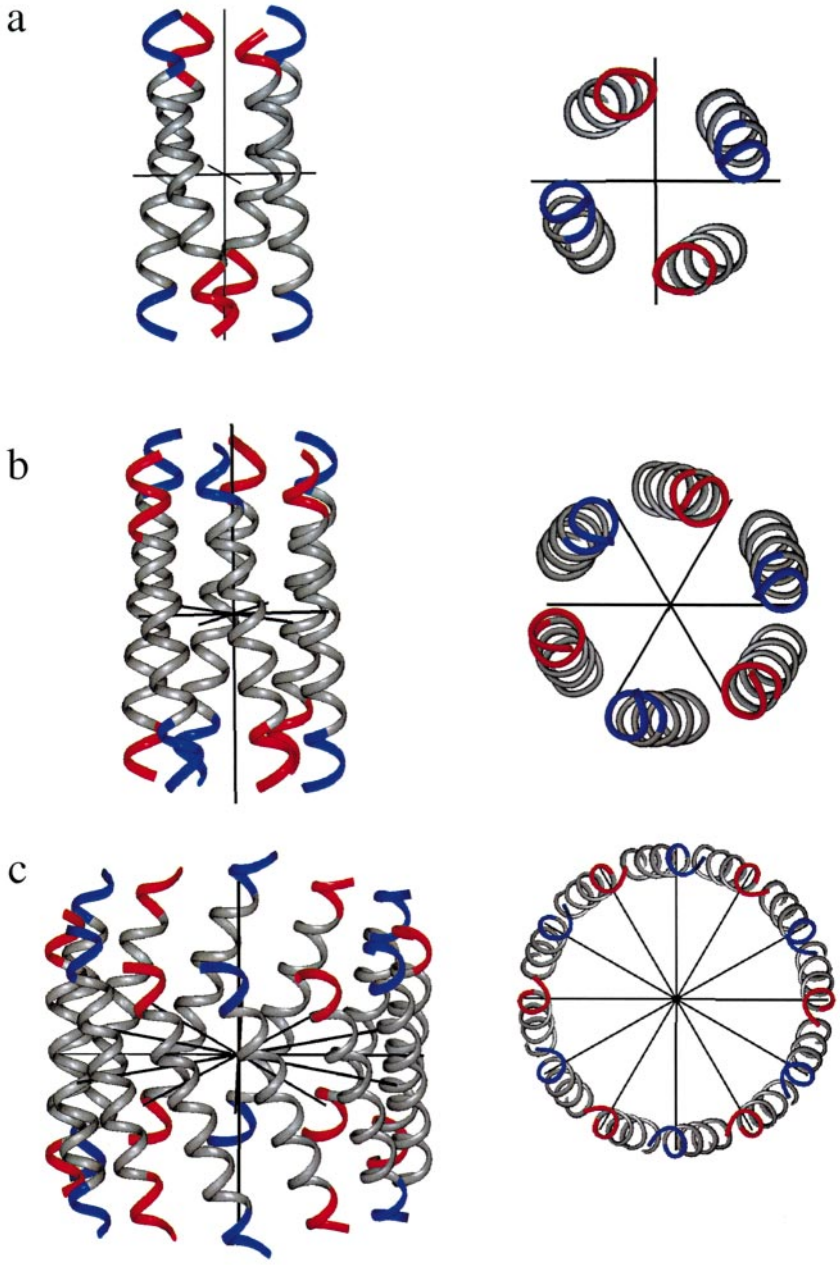

Figure 1. Side and top view of (a) four-helix bundle, (b) six-helix bundle, (c) 12-helix bundle with D_2, D_3 and D_6 symmetry respectively, showing the n-fold rotational symmetry axis aligned with the bundle and 2-fold axes of rotation orthogonal to the bundle axis. The helices are represented as ribbons; the N termini and the C termini are in blue and red, respectively, showing the antiparallel arrangement of the helices due to the 2-fold rotational symmetry.

Results

D_2 symmetry

Four-helical bundles often show approximate D_2 symmetry. Recently, we showed that models built with this symmetry reproduced the backbone geometries of six naturally occurring diiron proteins to a precision of approximately 1 Å r.m.s.d. Here, we extend this model to the central four-helical bundle in cytochrome bc_1[21,22] (Figure 2(a)), as an example of a heme-containing membrane protein. The core of bc_1 contains two bis-His ligated hemes, which are important for shuttling electrons across phospholipid membranes. These hemes are sandwiched between two diagonally opposed helices, which each donate two His ligands to the heme-binding sites. Previously, it was noted that this protein has quasi 2-fold symmetry with a helical pair as the asymmetric unit.[21,22] At the level of the backbone, the central four-helical bundle is well described by an even higher symmetry operator, D_2, with only a single helix as the asymmetric unit. Using a straight-helix model the r.m.s.d. is 1.16 Å with a total of 66 residues considered in the superposition. A coiled helix model results in an even better fit (Table 1) of 0.77 Å r.m.s.d with 85 residues included in the superposition (Figure 3(a)).

D_3 symmetry

As an example of D_3 symmetry, we considered the crystal structure of DSD,[24] a designed, hexame-

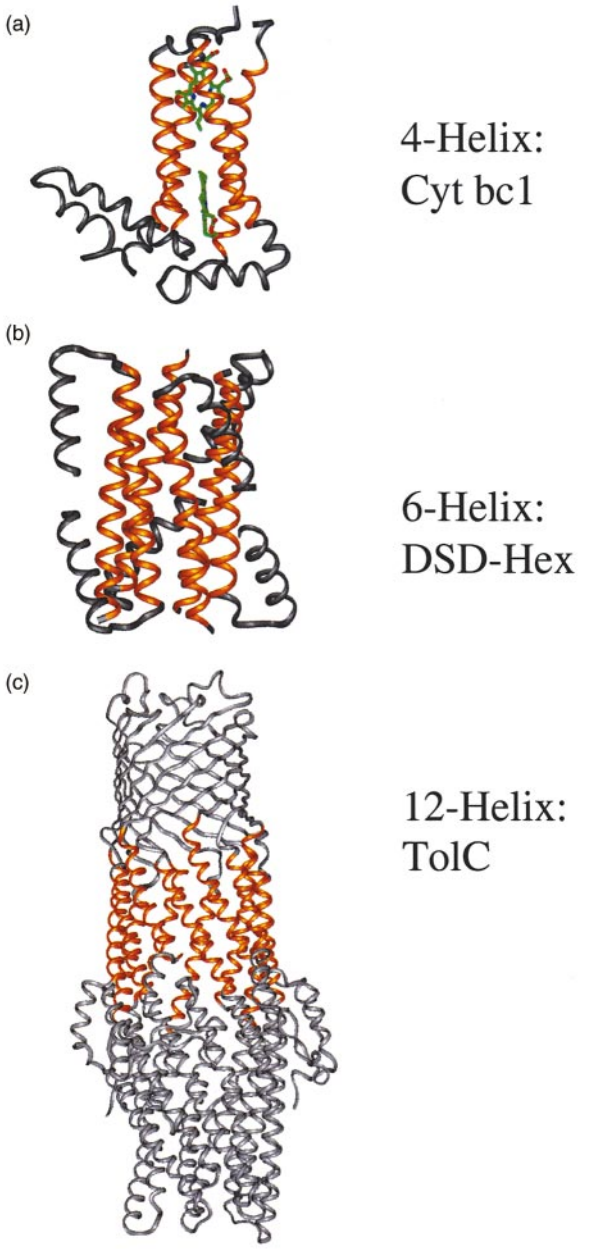

Figure 2. Experimental crystal structures of D_n helix bundles: the sections utilized for this work are highlighted in tan. (a) D_2, heme binding four-helix bundle in bc_1; (b) D_3, six-helix bundle at the core of the hexameric assembly of DSD; (c) D_6, a 12-helix bundle forms the tubular structure of TolC.

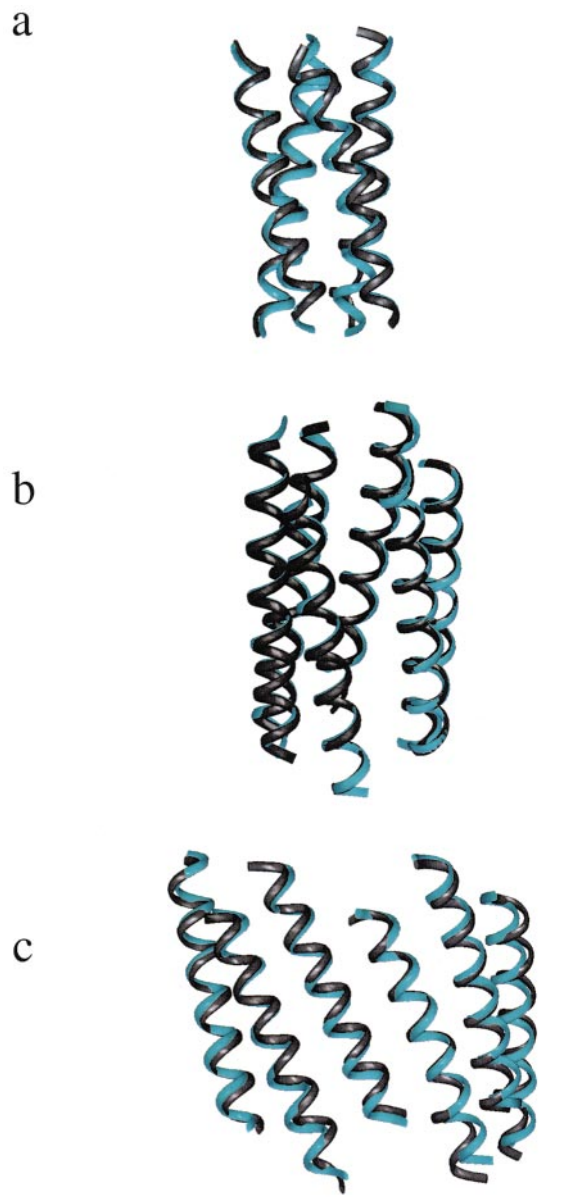

Figure 3. Superposition of coiled helix backbone models (blue) generated as described in Methods to experimental crystal structures (gray); (a) D_2-symmetrical four-helix bundle and bc_1; (b) D_3-symmetrical six-helix bundle and DSD; (c) D_6-symmetrical 12-helix bundle and TolC (for clarity, only half of the model is shown).

ric protein (Figure 2(b)). This protein was designed to form a domain-swapped dimer consisting of two identical three-helix bundles. In dilute aqueous solution, the protein indeed forms the desired domain-swapped dimer. However, at protein concentrations above approximately 2 mM these dimers show weak higher-order association, and a hexamer was observed from the crystal structure of the protein.

The unit cell of DSD contains three copies of a dimer, which forms the asymmetric unit in the crystal structure. The structure of the dimer is very close to that specified in its design. Each monomer contains a long helix, which packs against a second long helix from a neighboring subunit. N-terminal to the long helix is a second, shorter helix, which packs against the interface formed by the two long helices (Figure 2(b)). The two monomers in the dimer are related by approximate non-crystallographic symmetry (the r.m.s.d. of the backbone

Table 1. Parameters describing D_n-symmetrical coiled coils

Parameter	bc_1	DSD	TolC
r_1 (Å)	2.21	2.28	2.24
ω_1 (deg./residue)	−102.32	−102.37	−103.03
h, rise/residue (Å)	1.516	1.503	1.524
ϕ_1 (deg.)[a]	8.5	3.6	5.7
r_0 (Å)	7.4	9.4	18.6
P (Å)	171	288	308
ϕ_0 (deg.)[b]	13	10	−0.2
z_{trans} (Å)	0.47	1.81	3.65
Pitch angle α (deg.)[c]	15.3	11.6	20.7
Residues/turn (minor helix)	3.52	3.52	3.49
Residues/turn (α-helix)[d]	3.63	3.58	3.55
Number of residues included in fit	85[e]	156[f]	300[g]
r.m.s.d. (Å)	0.77	0.27	1.10

[a] ϕ_1 describes the rotation of the minor helix about its own axis. It is given relative to an "ideal" value in which a vector originating at the center of the helix and passing through the C^{α} atom of the a position is directed precisely towards the superhelical axis.
[b] The phase of the superhelix, ϕ_0, is expressed relative to the value required to provide equal spacing between each neighboring helix.
[c] The pitch angle α is calculated from the inverse sine of the circumference of the major cylinder *versus* its pitch.
[d] The number of residues per turn of the α-helix is obtained by solving equation (1) for ω_1.
[e] Residues C34-C52, C84-C105, C113-C134 and C183-C204 of the protein (pdb code, 1BCC) were included in the fit.
[f] Residues 21-46 and 71-96 from each of the subunits (as defined in pdb code, 1G6U) were included in the fit.
[g] Residues 15-39, 82-106, 221-245, and 300-324 of each of the subunits (pdb code, 1EK9) were included in the fit.

atoms is 0.16 Å). The long helices of the individual dimers further associate into a hexameric barrel-like structure, with the short helices packed against the outside of the barrel (Figure 2(b)). The central antiparallel six-helical bundle is very well described by D_3 symmetry, with the coiled coil model (Figure 3(b)) again providing a better fit (0.27 Å, Table 1) than the straight helix model (0.79 Å).

D_6 symmetry

TolC is one component of a bacterial transport system that is responsible for efflux of selected molecules from *Escherichia coli*.[14] TolC forms a long 12-stranded coiled coil, which forms a water-soluble extension of a porin-like transmembrane domain. The membrane-proximal half of this bundle forms a very regular barrel-like structure, while the packing deviates from this idealized geometry near the bottom of the protein as viewed in Figure 2(c). The TolC 12-helical bundle consists of a crystallographic trimer of four non-identical helices. It has been noticed that the structure has approximate C_6 symmetry with a two-helix structure in the asymmetric unit.[13] The N-terminal portion of the bundle (300 residues) can also be described by a D_6 symmetrical model (Figure 3(c)) with a single helix in the asymmetric unit. Again, a bundle with coiled helices fits better (1.10 Å r.m.s.d.) than a model with straight helices (1.24 Å r.m.s.d.).

Location of cofactors and bound ions

The approximate D_n symmetry of these proteins extends to the positions of bound cofactors and ions. In DSD, a sulfate ion is located along the crystallographic 3-fold axis 2.8 Å from the origin of the structure (Figure 4(b)). In bc_1 the heme iron atoms lie 0.5 Å and 2.8 Å from the central superhelical axis, respectively. However, the placement of the iron atoms relative to the origin of the structure is more consistent with C_n than D_n symmetry; one iron atom is located 8.1 Å above and one is located 12.2 Å below the origin (Figure 4(a)). The relatively large deviation from D_n symmetry arises from the fact that the His side-chains are located on only two, rather than all four of the helices.

Organization and packing in D_n-symmetrical bundles

Figure 5 illustrates idealized geometries for D_2, D_3 and D_6-symmetrical bundles. Each position in the heptad repeat is labeled in the usual nomenclature for coiled coils,[1] with the residue that projects most directly towards the central axis of the bundle being defined as a. In each case, the proteins show two geometrically distinct helix-helix pairings. In the D_2-symmetrical bundle (Figure 5(a)), the packing of residues at positions a and e from neighboring helices define one such interface (defined here as the a/e interface); residues at d and g pack at the other interface, denoted the d/g interface.

In the D_6-symmetrical structure (Figure 5(c)), one helix-helix interface is primarily defined by the packing of residues at b and e, while residues at d and g define the second interface. The structure of the α-helical coiled coil appears to be well suited to the formation of 12-sided structures; consider two lines running through the center of a helix and bisecting either b and e or d and g. These two lines will cross at an angle of 154° ($3/7 \times 360°$), which is very close to the angle of 150°, required for the generation of an equilateral dodecagon. This arrangement has also been noted by Walshaw &

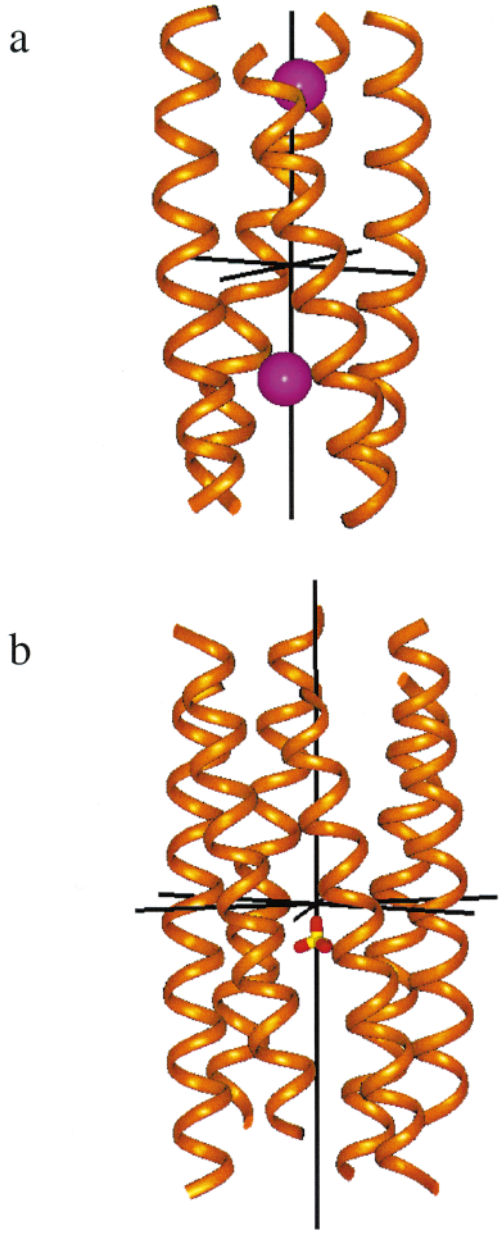

Figure 4. Position of bound heme iron ions in bc_1 (a) and sulfate ion in DSD (b) with respect to the symmetry axes.

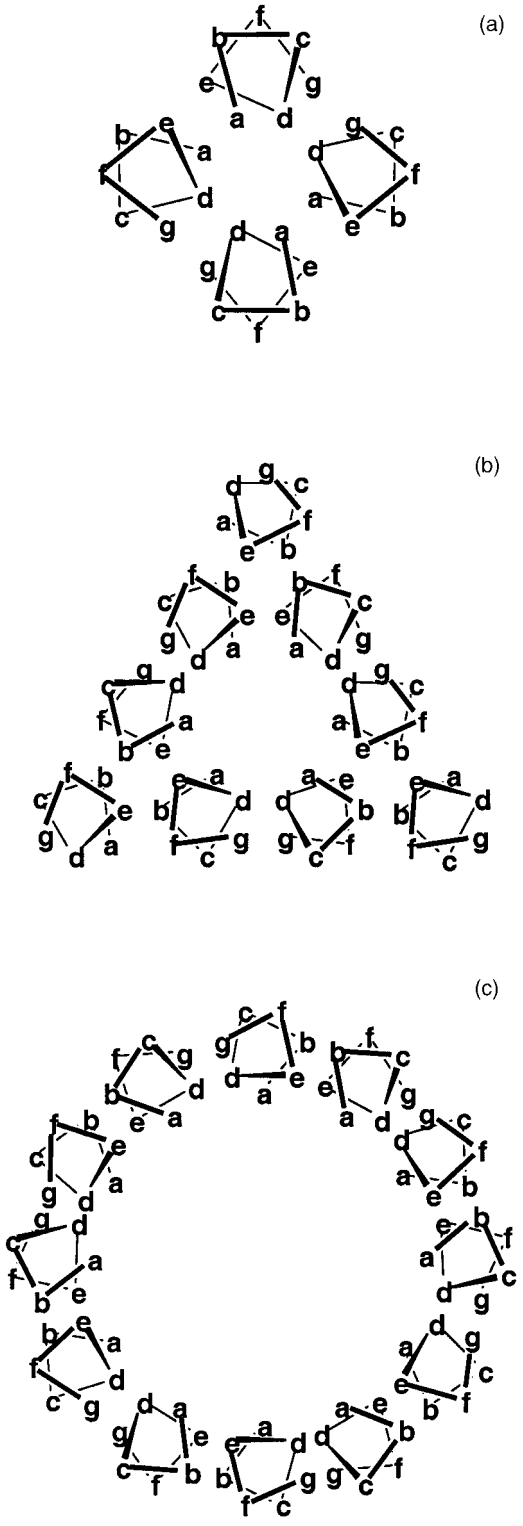

Figure 5. Helical wheel diagrams of antiparallel helix bundles, describing side-chain positions in the heptad repeat for (a) D_2, (b) D_3 and (c) D_6.

Woolfson[24] as well as by Calladine *et al.*[25] How then, is the six-sided D_3 bundle formed?

The D_3 bundle (Figure 5(b)) in DSD is a trimer of domain-swapped, three-helical bundles. The three-helix bundles are stabilized by a packing interaction between three helices, whereas the packing between different three-helix bundles involves a two-helix interface. Thus, as one progresses around the D_3-symmetrical inner ring of helices, each helix alternately forms a two-helix interaction along its d-g surface and a three-helix interaction along its b-e surface. As discussed,[5] different geometries are observed for two *versus* three-helix bundles, with ''acute'' packing being observed in the three-helix bundle. This difference in helix packing angles, together with small differences in helix-packing

radii, provide a hexameric rather than a 12-sided structure.

Comparison of the structural parameters describing D_n-symmetrical coiled coils

Table 1 summarizes the relevant parameters defining the three types of coiled coils investigated in this work. In each structure, the minor helix has a nearly perfect 3.5-residue repeat, as expected for coiled coils. This observed 3.5-residue repeat is a result of applying a left-handed superhelical twist to an α-helix. Dunker & Zaleske[26] discussed the relationships between the pitch (P), superhelical radius (r_0), and the pitch angle, (α, the dihedral angle between the minor and major helical axes) of a coiled coil. The pitch angle is defined by equation (1):

$$\alpha = \tan^{-1}[(\omega_1 - \omega_\alpha)(r_0)/(h)] \qquad (1)$$

in which ω_1 is the minor helical angular frequency in its own reference frame (which rotates about the superhelical axis) and ω_α is the α-helical angular frequency in the global (fixed) reference frame. Both are expressed in radians per residue ($2\pi/3.5$ for an ideal coiled coil, and $2\pi/3.62$ for an idealized α-helix). The parameter r_0 is the radius of the superhelix, and h is the rise per residue of the alpha-helix. For coiled coils with relatively small radii, idealized values for the geometric parameters of the α-helix can be used to predict the value of α. For example, the value predicted for bc_1 ($r_0 = 7.4$ Å) is 16.3° using 3.5 residues/turn for the coiled coil minor helix, 3.62 residues/turn for the α-helix, and $h = 1.51$ Å/residue. This value is in very good agreement with the observed value of 15.3°, and this pitch angle allows highly favorable interactions between the side-chains of neighboring helices. However, as the radius of the helix is increased, the predicted pitch angle increases beyond the range of values that are consistent with reasonable helical geometries and inter-helical packing interactions. For example, the value predicted for the TolC bundle ($r_0 = 18.6$ Å) is 36.2°, considerably greater than the observed value of 20.7°. Instead, it has been predicted[26] that as the radius of a coiled coil is increased, the α-helices will slightly unwind, thereby providing a more reasonable value of α. Indeed, as the radii of the bundles in Table 1 increase from 7.4 Å to 18.6 Å, the repeat of the α-helix decreases from 3.63 residues/turn to 3.55 residues/turn. A similar value of 3.54 was calculated for the TolC bundle by Calladine *et al.*[25] using a different method. This deviation from an ideal α-helical geometry appears to represent a relatively low-energy deformation, based on the fact that helices with 3.5 to 3.6 residues/turn occur quite frequently in globular proteins.[27]

A second consequence of the unwinding of the α-helix is that the pitch of the superhelix changes according to equation (2):

$$P = [(2\pi h/(\omega_1 - \omega_\alpha))^2 - (2\pi r_0)^2]^{1/2} \qquad (2)$$

The first term in equation (2) increases with increasing radius (because of the concomitant decrease in the difference between ω_1 and ω_α), while the second term becomes more negative with increasing r_0. In this series of bundles, the first term dominates; as r_0 changes from 7.4 in bc_1 to 18.6 Å in tolC, P increases from 171 Å to 308 Å.

Helix pairwise packing interactions

It is instructive to consider the similarities and differences in the geometries of the helix-helix packing interactions in D_n-symmetrical bundles. To allow a comparison of the interhelical packing arrangements in proteins with different symmetry, it is useful to introduce a heptad nomenclature that depends only on the relationship of the two interacting helices, and is independent of the aggregation number of the bundle. In this convention, the residues projecting towards the neighboring helices are designated a′ and d′, respectively (Figure 6). Thus, if the b-e interface of the D_6 bundle

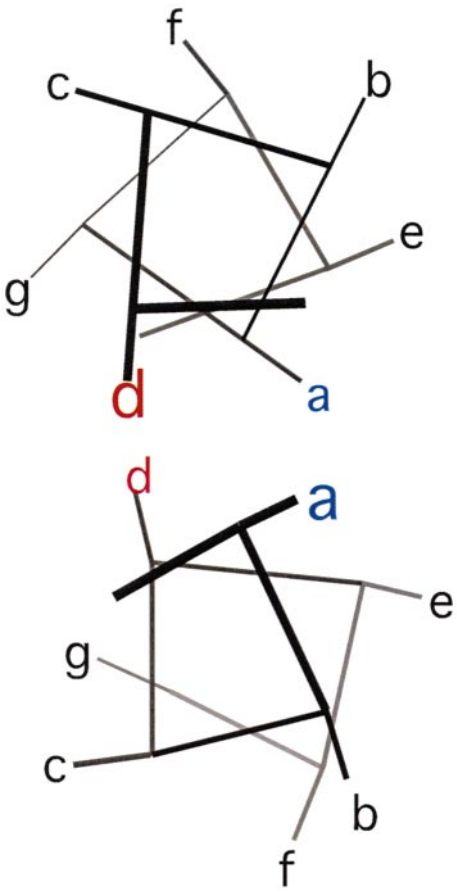

Figure 6. Helical wheel diagram illustrating the side-chain interactions between a pair of neighboring helices in antiparallel helix bundles. The residues projecting towards the helix-helix interface are designated a′ and d′, respectively, independent of their position in the parent bundle.

(Figure 5(c)) is considered as a pairwise interaction, the b and e positions become a′ positions and d′ positions, respectively. Similarly, the d and g positions of a d-g interface become a′ and d′, respectively, in the pairwise nomenclature.

Figure 6 displays a typical antiparallel helical pair with the positions labeled a′ through g′. The helical pair can be divided into two halves, with positions a′, b′, and e′ on one side of the structure, and positions d′, g′ and c′ on the opposite side. Following the convention of Efimov,[28] if residues a′, b′ and e′ face the interior of the protein, the helix pair is defined as left-handed; the pair is considered right-handed if d′, g′ and c′ face the interior of the bundle. D_n-symmetrical bundles have an equal number of left-handed and right-handed helical pairs, forming a closed, circular structure when viewed from the top of the bundle (Figure 5).

Table 2 provides the interhelical distances and packing angles, associated with the helical pairs in the D_n-symmetrical bundles. Because a rotational axis relates the two individual helices in each pair, only two angles and two translations are required to approximately characterize their geometries: the interhelical distance d, the interhelical packing angle δ, the rotation of the helix about its own axis, ε, and the vertical displacement of the helices, d_z. In Table 2, the value of d_z is also expressed as the distance between the a residues expressed in fractions of a heptad, as described.[29] The angle ε is defined by measuring the angle between a line connecting the center of the two helices and a vector from the center of an individual helix that bisects its a′ and d′ C^α atoms.

The values of the interhelical distance and the packing angles are reasonably well clustered for four of the six interfaces described in Table 2, including both interfaces of TolC, the a-e interface of bc_1, and the b-e interface of DSD. These helical pairs each represent variants of an Alacoil α-packing motif, in which Ala residues (or other small side-chains) occupy either the a′ or d′ position of the helix.[29] A particularly striking example of a ''ferritin-like'' Alacoil is encountered in the d-g interface of DSD (Figure 7(a)). In this Alacoil, the Ala residues occur in the a′ positions on the inner side of the barrel. Their small size allows efficient packing of the helices, resulting in a separation of 8.0 Å. A second ROP-like Alacoil, with the Ala side-chains occurring at the d′ positions, is found in a d-g interface of TolC (Figure 7(b)). As described by Gernert and co-workers,[29] the offset of the helices is different for the ferritin-type *versus* the ROP-type packing (Figure 7(b)). This difference in z_{trans} is a result of the fact that the C^α-C^β bonds do not radiate directly outward from the helix, when viewed axially (see Figure 6). Also, the C^α-C^β bonds are angled towards the N terminus of the helix when viewed from the side of the helix (not shown). Calladine *et al.*[25] and Walshaw & Woolfson[24] have conducted a very extensive analysis of side-chain packing in TolC, and the role that sterics play in fine-tuning the geometry of the helix-helix interface.

Two helical interfaces deviate significantly from the others in terms of their packing parameters (Table 2). The bc_1 protein has one helix-helix interface with a relatively long (11.6 Å) interhelical distance and large values of δ and ε, which result from the insertion of the heme between the two helices. There is relatively little interaction between the helices along this interface, which is dominated by the interactions of the helices with the porphyrin ring. In contrast, the helices interact extensively along the d-g interface of bc_1, in a variant of the Alacoil motif in which the helices are slightly further apart than in a canonical structure.

The b-e interface of DSD is involved in a three-helix bundle interaction that gives rise to differences in its inter-helical packing parameters. The core of the three-helix bundle is packed with Leu side-chains, which gives rise to a relatively large inter-helical separation (10.8 Å). The acute inter-helical packing angle found in three-helix bundles[5]

Table 2. Interhelical packing distances and angles

Interface	a′ residue[a]	d (Å)[b]	d_z (Å)[c]	δ (deg.)[d]	ε (deg.)[e]
bc_1 (a-e interface)	e	9.2	2.0 (0.19)	18.9	3.5
bc_1 (d-g interface)	d	11.6	4.9 (0.48)	23.9	15.8
DSD (b-e interface)	b	10.8	0.69 (0.067)	13.4	17.6
DSD (d-g interface)	d	8.0	3.7 (0.36)	9.7	−14.3
TolC (b-e interface)	b	9.6	1.3 (0.13)	10.5	−4.4
TolC (d-g interface)	d	9.8	4.4 (0.45)	10.6	−8.4

The distances and angles are defined by first orienting the bundles in a Cartesian coordinate system with the superhelical axis along z, and the orthoganol 2-fold axes in the *x*-*y* plane.

[a] The residue in the heptad nomenclature used in Figure 6, corresponding to the a′ position in the pairwise helical nomenclature.

[b] d is the distance of closest approach between the axes of the individual helices.

[c] d_z is the distance along the α-helical axis from the *x*-*y* plane to the closest a′ residue. The values in parentheses are expressed in factional distance along a heptad unit, following the convention of Gernert *et al.*[29]

[d] δ is the interhelical crossing angle defined by the dot product between tangents to the helical axes of neighboring helices. It is defined to be zero when the helices are antiparallel.

[e] ε describes the rotation of the helix about its own axis. It is given relative to an ''ideal'' value expected if a vector originating at the center of the helix and bisecting the C^α atoms of the central a′ and d′ positions (in a projection onto the *x*-*y* plane) were directed precisely towards the axis of the neighboring helix. The angle is measured at the heptad position closest to the *x*-*y* plane.

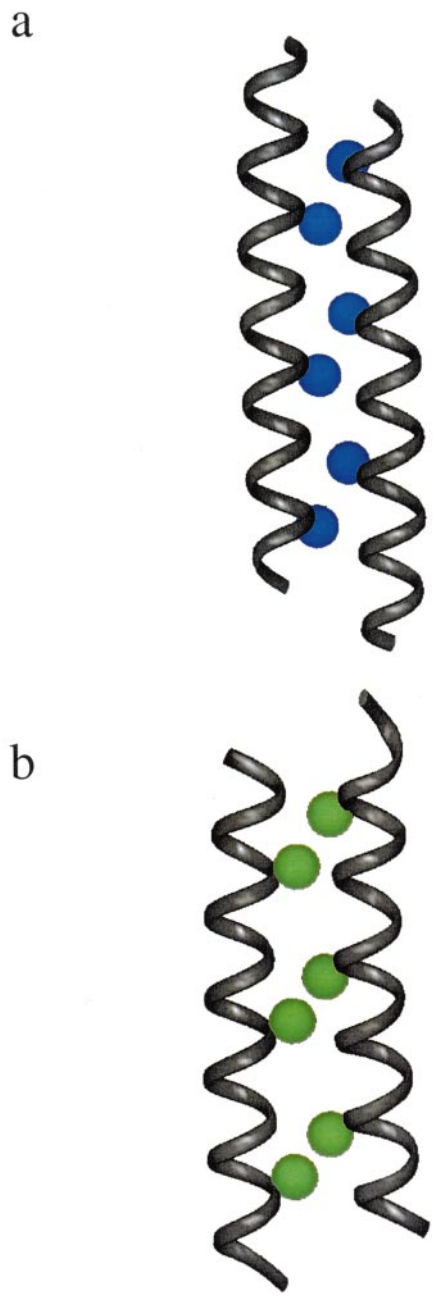

Figure 7. Side view of Alacoil motifs in DSD and TolC. (a) In DSD the Ala side-chains lie at the a′ positions, forming a ferritin-type Alacoil. (b) A ROP-type Alacoil motif with Ala side-chains at d′ positions is found in one of the two interfaces of TolC.

is reflected in a value of ε angle that is larger from the values observed for the other helix-helix packings (Table 2).

Conclusions

The parameterization of protein three-dimensional structures is an important tool for *de novo* protein design. A very early attempt at protein design used a D_2 symmetry operator to define the backbone structure of a four-helical bundle.[30–32] Although this protein was subsequently found to adopt a molten globule conformation,[17] refinements in side-chain packing[33–35] have made this general approach more suitable for *de novo* protein design of complex, functional proteins.[20,36] Here, we demonstrate that a very simple mathematical formalism is able to capture the essential features of the structures of several complex proteins. The D_n-symmetrical model used in this work is particularly well suited to the analysis and design of tubular proteins that contain binding sites within their central cores. Depending on the aggregation number, the tubes may contain tightly packed cofactors, as in bc_1 and diiron proteins,[19,20] or large cavities capable of accommodating a variety of substrates as in TolC. The parameters that have been obtained from these models can be used for the design of idealized versions of these folds, or they can be varied to allow the design of entirely novel structures.

In automated approaches to protein design, the parameters defining the backbone structure are allowed to vary, providing a solution to the flexible docking problem.[6,15,16,33–35,37] In this manner, it is possible to search both conformational space as well as sequence space to find a sequence that can stabilize a given fold. The use of D_n-symmetry operators should be particularly valuable in the initial phases of computational design. Because a specific function often requires a loss of symmetry, the D_n symmetry of the original bundle can be lifted as the design progresses, as in our previous work on diiron proteins.[20] The design of a variety of D_2-symmetrical heme-binding proteins, D_3-symmetrical hexamers, and D_6-symmetrical tubes is in progress, and will be reported elsewhere.

An interesting finding of this and other papers[24,25] is the observation that the helical faces used in the helix-helix packings help to define the aggregation number of the bundle. The a-e and d-g interfaces are oriented at nearly 90° angles, fitting nicely with the rectangular cross-section of an antiparallel four-helix bundle. Similarly, the b-e and d-g interfaces subtend a wider angle, nearly ideal for a 12-helix bundle. Finally, by alternating two-helix and three-helix interfaces, it is possible to engineer a hexameric bundle. The actual angles within the bundle are further fine-tuned by the nature of the steric interactions between the side-chains.[26] It would be interesting to determine whether further modifications in the bulk of the interacting side-chains could result in a change in the aggregation number, similar to the changes observed in variants of GCN4 P1[4].

Methods

Fitting idealized backbones to experimental structures

We examined two mathematical models to describe the backbone conformations of bc_1, DSD, and TolC. The first is a D_n-symmetrical bundle of straight α-helices,[19] while the second uses a coiled coil model.[38] The fitting parameters for the first method were the three Eulerian rotations and the three translations required to define the location of a monomeric helix within the coordinate system. A modification of the Crick parameterization was used for the coiled-coil model, which employs six parameters to define the path of an individual chain: the radius (r_1), angular frequency (ω_1), rise per residue (h) and phase (ϕ_1) of the minor helix, as well as the radius (r_0), and pitch (P) of the superhelix. While these parameters suffice to describe a C_n-symmetrical coiled coil, two additional parameters are necessary to describe the interhelical orientations of antiparallel D_n-symmetrical coiled coils. These parameters include the phase of the superhelix (ϕ_0) and the translation of the supercoiled helix along the superhelical axis (z_{trans}).

The parameters, r_0, r_1, P, ω_1, ϕ_1, ϕ_0, h, and z_{trans} were fit to the C^α coordinates of the proteins using the equations:

$$x = r_0 \cos \omega_0 t + r_1 \cos \omega_0 t \times \cos(\omega_1 t + \phi_1) - r_1(P/(2\pi r_0^2 + P^2)^{1/2}) \times \sin \omega_0 t \times \sin(\omega_1 t + \phi_1) \quad (3a)$$

$$y = r_0 \cos \omega_0 t + r_1 \sin \omega_0 t \times \cos(\omega_1 t + \phi_1) - r_1(P/(2\pi r_0^2 + P^2)^{1/2}) \times \sin \omega_0 t \times \sin(\omega_1 t + \phi_1) \quad (3b)$$

$$z = -Pt + r_1(2\pi r_0/(2\pi r_0^2 + P^2)^{1/2}) \times \sin(\omega_1 t + \phi_1) \quad (3c)$$

in which t is the residue number (for the examples in Table 1, the center residue was defined to be the zeroth position). The parameter ω_0 is the angular frequency of the major helix (in radians per residue). It is related to the fitting parameters according to equation (4):

$$\omega_0 = 2\pi h/(r_0^2 + P^2) \quad (4)$$

These equations generate a superhelix, which rotates around the z-axis. To generate the remaining helices, the superhelix is translated in the z direction by the value z_{trans}, rotated by ϕ_0, and the appropriate symmetry operations are performed to generate the bundle.

For each system, the helical regions of the peptides were fit to equations (3) (Table 1). Eight parameters are fit using a genetic algorithm (the remaining parameters are derived from these variables). The algorithm encodes the eight parameters as a contiguous binary number, which is then modified by either swapping with a portion of another string or by randomly changing a single digit in the sequence, which is referred to as ''mutation''. One ''immortal'' chromosome, corresponding to the best fit in that generation, is passed unmodified to the next generation. The algorithm has a mutation rate of 0.08, a single crossover rate of 0.65, and a double crossover rate of 0 with the C^α r.m.s.d. fit used as a scoring function. The algorithm is stopped once the fitting parameters and the r.m.s.d. are constant for at least 20 generations. Tests on fitting the bc_1 peptide indicate that after 20 generations of being constant, the fit does not change after 150 additional generations. Once the alpha carbon atoms are in place, the rest of the backbone is generated by assuming that the peptide is approximately helical. This preliminary backbone is then refined (with fixed alpha carbon positions) through energy minimization using the cvff force-field in the program Discover.

Acknowledgments

We thank David Eisenberg and Kim Sharp for helpful discussions, and grants from the NIH (GM56423, GM. 54616) and NSF (DMR79909).

References

1. Cohen, C. & Parry, D. A. D. (1990). a-Helical coiled coils and bundles: how to design an α-helical protein. *Proteins: Struct. Funct. Genet.* **7**, 1-15.
2. Hodges, R. S. (1996). *De novo* design of alpha-helical proteins: basic research to medical applications. *Biochem. Cell Biol.* **74**, 133-154.
3. Gonzalez, L., Jr, Plecs, J. J. & Alber, T. (1996). An engineered allosteric switch in leucine-zipper oligomerization. *Nature Struct. Biol.* **3**, 510-515.
4. Harbury, P. B., Zhang, T., Kim, P. S. & Alber, T. (1993). A switch between two-, three-, and four-stranded coiled coils. *Science,* **262**, 1401-1407.
5. Harbury, P. B., Kim, P. S. & Alber, T. (1994). Crystal structure of an isoleucine-zipper trimer. *Nature,* **371**, 80-83.
6. Harbury, P. B., Plecs, J. J., Tidor, B., Alber, T. & Kim, P. S. (1998). High-resolution protein design with backbone freedom. *Science,* **282**, 1462-1467.
7. Nautiyal, S. & Alber, T. (1999). Crystal structure of a designed, thermostable, heterotrimeric coiled coil. *Protein Sci.* **8**, 84-90.
8. Ogihara, N. L., Weiss, M. S., DeGrado, W. F. & Eisenberg, D. (1997). The crystal structure of the designed trimeric coiled coil coil-V_aL_d: implications for engineering crystals and supramolecular assemblies. *Protein Sci.* **6**, 80-88.
9. Schneider, J. P., Lombardi, A. & DeGrado, W. F. (1998). Analysis and design of three-stranded coiled coils and three-helix bundles. *Fold. Des.* **3**, R29-R40.
10. Presnell, S. R. & Cohen, F. E. (1989). Topological distribution of four-a-helix bundles. *Proc. Natl Acad. Sci. USA,* **86**, 6592-6596.
11. Kamtekar, S. & Hecht, M. H. (1995). Protein motifs. 7. The four-helix bundle: what determines a fold? *FASEB J.* **9**, 1013-1022.
12. Weber, P. C. & Salemme, F. R. (1980). Structural and functional diversity in 4-α-helical proteins. *Nature,* **287**, 82-84.
13. Koronakis, V., Sharff, A., Koronakis, E., Luisi, B. & Hughes, C. (2000). Crystal structure of the bacterial membrane protein TolC central to multidrug efflux and protein export. *Nature,* **405**, 914-919.
14. Postle, K. & Vakharia, H. (2000). TolC, a macromolecular periplasmic ''chunnel''. *Nature Struct. Biol.* **7**, 527-530.
15. Harbury, P. B., Tidor, B. & Kim, P. S. (1995). Repacking protein cores with backbone freedom: structure prediction for coiled coils. *Proc. Natl Acad. Sci. USA,* **92**, 8408-8412.

16. Su, A. & Mayo, S. L. (1997). Coupling backbone flexibility and amino acid sequence selection in protein design. *Protein Sci.* **6**, 1701-1707.
17. DeGrado, W. F., Summa, C. M., Pavone, V., Nastri, F. & Lombardi, A. (1999). *De novo* design and structural characterization of proteins and metalloproteins. *Annu. Rev. Biochem.* **68**, 779-819.
18. Boice, J. A., Dieckmann, G. R., DeGrado, W. F. & Fairman, R. (1996). Thermodynamic analysis of a designed three-stranded coiled coil. *Biochemistry,* **35**, 14480-14485.
19. Summa, C. M., Lombardi, A., Lewis, M. & DeGrado, W. F. (1999). Tertiary templates for the design of diiron proteins. *Curr. Opin. Struct. Biol.* **9**, 500-508.
20. Lombardi, A., Summa, C. M., Geremia, S., Randaccio, L., Pavone, V. & DeGrado, W. F. (2000). Inaugural article: retrostructural analysis of metalloproteins: application to the design of a minimal model for diiron proteins. *Proc. Natl Acad. Sci. USA,* **97**, 6298-6305.
21. Xia, D., Yu, C. A., Kim, H., Xia, J. Z., Kachurin, A. M., Zhang, L., Yu, L. & Deisenhofer, J. (1997). Crystal structure of the cytochrome *bc1* complex from bovine heart mitochondria. *Science,* **277**, 60-66.
22. Iwata, S., Lee, J. W., Okada, K., Lee, J. K., Iwata, M., Rasmussen, B. *et al*. (1998). Complete structure of the 11-subunit bovine mitochondrial cytochrome *bc1* complex. *Science,* **281**, 64-71.
23. Ogihara, N. L., Ghirlanda, G., Bryson, J. W., Gingery, M., DeGrado, W. F. & Eisenberg, D. (2001). Design of three-dimensional domain-swapped dimers and fibrous oligomers. *Proc. Natl Acad. Sci. USA,* **98**, 1404-1409.
24. Walshaw, J. & Woolfson, D. N. (2001). Open-and-shut cases in coiled-coil assembly: alpha-sheets and alpha-cylinders. *Protein Sci.* **10**, 668-673.
25. Calladine, C. R., Sharff, A. & Luisi, B. (2001). How to untwist an alpha-helix: structural principles of an alpha- helical barrel. *J. Mol. Biol.* **305**, 603-618.
26. Dunker, K. & Zaleske, D. J. (1977). Stereochemical considerations for constructing alpha-helical proteins with particular application to membrane proteins. *Biochem. J.* **163**, 43-57.
27. Barlow, D. J. & Thornton, J. M. (1988). Helix geometry in proteins. *J. Mol. Biol.* **201**, 601-619.
28. Efimov, A. V. (1993). Patterns of loop regions in proteins. *Curr. Opin. Struct. Biol.* **3**, 379-384.
29. Gernert, K. M., Surles, M. C., Labean, T. H., Richardson, J. S. & Richardson, D. C. (1995). The Alacoil: a very tight, antiparallel coiled-coil of helices. *Protein Sci.* **4**, 2252-2260.
30. Eisenberg, D., Wilcox, W., Eshita, S. M., Pryciak, P. M., Ho, S. P. & DeGrado, W. F. (1986). The design, synthesis, and crystallization of an alpha-helical peptide. *Proteins: Struct. Funct. Genet.* **1**, 16-22.
31. Ho, S. P. & DeGrado, W. F. (1987). Design of a four-helix bundle protein: synthesis of peptides which self-associate into a helical protein. *J. Am. Chem. Soc.* **109**, 6751-6758.
32. Regan, L. & DeGrado, W. F. (1988). Characterization of a helical protein designed from first principles. *Science,* **241**, 976-978.
33. Lazar, G. A. & Handel, T. M. (1998). Hydrophobic core packing and protein design. *Curr. Opin. Chem. Biol.* **2**, 675-679.
34. Desjarlais, J. R. & Clarke, N. D. (1998). Computer search algorithms in protein modification and design. *Curr. Opin. Struct. Biol.* **8**, 471-475.
35. Street, A. G. & Mayo, S. L. (1999). Computational protein design. *Struct. Fold. Des.* **7**, R105-R109.
36. Betz, S. F. & DeGrado, W. F. (1996). Controlling topology and native-like behavior of *de novo*-designed peptides: design and characterization of antiparallel four-stranded coiled coils. *Biochemistry,* **35**, 6955-6962.
37. Desjarlais, J. R. & Handel, T. M. (1999). Side-chain and backbone flexibility in protein core design. *J. Mol. Biol.* **290**, 305-318.
38. Crick, F. H. C. (1953). The fourier transform of a coiled-coil. *Acta Crystallog.* **6**, 685-689.

Edited by J. Thornton

(Received 18 April 2001; received in revised form 18 June 2001; accepted 18 June 2001)