doi:10.1016/S0022-2836(02)00233-4 available online at http://www.idealibrary.com on IDEAL® J. Mol. Biol. (2002) 319, 243–253

JMB

A Hierarchic Approach to the Design of Hexameric Helical Barrels

Giovanna Ghirlanda[1], James D. Lear[1], Nancy L. Ogihara[2] David Eisenberg[2] and William F. DeGrado[1]*

[1]*Department of Biochemistry and Biophysics, The Johnson Research Foundation University of Pennsylvania School of Medicine, Stellar Chance Building, Room 1010 421 Curie Boulevard Philadelphia, PA 19104-6059 USA*

[2]*UCLA-DOE Laboratory of Structural Biology and Molecular Biology Institute P.O. Box 951570, University of California, Los Angeles, CA 90095-1570, USA*

**Corresponding author*

The design of large macromolecular assemblies is an endeavor with implications for protein engineering as well as nanotechnology. A hierarchic approach was used to design an antiparallel hexameric, tubular assembly of helices. In previous studies, a domain-swapped, dimeric three-helix bundle was designed from first principles. In the crystal lattice, three dimers associate around a 3-fold rotational axis to form a hexameric assembly. Although this hexameric assembly was not observed in solution, it was possible to stabilize its formation by changing three polar residues per monomer to hydrophobic (two Phe and one Trp) residues. Molecular models based on the crystallographic coordinates of DSD (PDB accession code 1G6U) show that these side-chains pack in the central cavity (the "supercore") of the hexameric bundle. Analytical ultracentrifugation, fluorescence spectroscopy, CD spectroscopy, and guanidine–HCl denaturation were used to determine the assembly of the hexamer. To probe the requirements for stabilizing the hexamer, we systematically varied the polarity and steric bulk of one of the Phe residues in the supercore of the hexamer. Depending on the nature of this side-chain, it is possible to modulate the stability of the hexamer in a predictable manner. This family of hexameric proteins may provide a useful framework for the construction of proteins that change their oligomeric states in response to binding of small molecules.

Keywords: protein design; coiled-coils; domain-swapped dimer; helical barrels; supramolecular assembly

Introduction

Symmetry is apparent at all levels of protein structure from the screw symmetry in β-strands and α-helices to the icosahedral symmetry observed in viruses.[1] Also, the elements of secondary structure within the tertiary structures of proteins often occur in quasi-symmetrical arrangements, as observed in antiparallel four-helix bundles,[2–6] porins,[7–11] β-propellers,[12] and TIM barrels.[13–18] The symmetrical arrangement of secondary structures in many proteins is a result of gene duplication.[16,18–20] Parallel coiled-coils represent a highly symmetrical class of oligomeric proteins in which one or more helices from each monomer inter-twine to form a superhelix.[21] Antiparallel coiled-coils and helical bundles show dihedral (D_n) symmetry, and the backbone geometries of a variety of four, six, and 12-helix bundles can be described to within approximately 1 Å RMSD using D_2, D_3, and D_6 symmetry operators, respectively.[20,22,23]

Symmetry has been used quite extensively to simplify the process of *de novo* protein design. For example, coiled-coils[24] show a seven-residue geometric repeat, which greatly facilitates the design of peptides that assumes this repeating structure.[25–33] Thus, the determinants for the formation of two, three, and four-stranded parallel coiled-coils have been systematically elucidated through protein design and structural studies.[28,31,34–36] Further, the requirements for folding into antiparallel four-helix bundles have been probed through the design of proteins with

Present address: N. L. Ogihara, Accelrys Inc., 9685 Scranton Road, San Diego, CA 92121, USA.

Abbreviations used: Fmoc, 9-fluorenylmethoxycarbonyl; PAL, 5[4-(aminomethyl)-3,5-bis(methoxy)-phenoxy]valeric acid; HOBt, *N*-hydroxybenzotriazole.

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

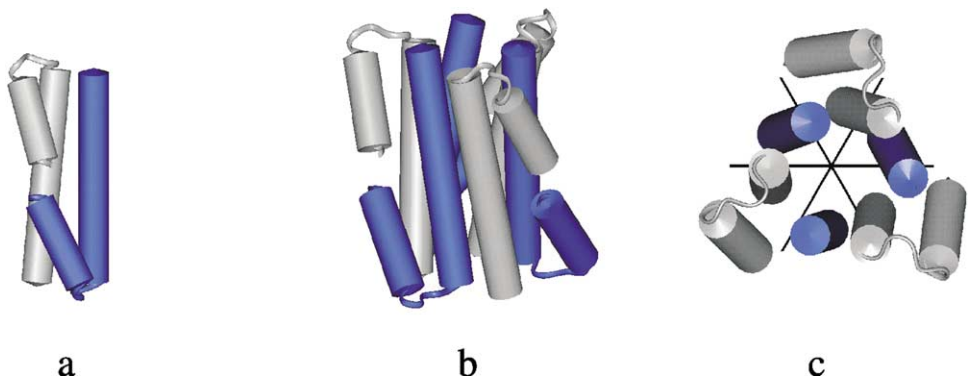


Figure 1. Schematic representation of the crystal structure of DSD: helices are represented as cylinders. (a) Domain-swapped dimeric three-helix bundle, in which two long helices, one from each monomer (in blue and gray, respectively) pack in an antiparallel manner, and two short helices dock against them. The N termini of the short helices define a cleft in the structure. (b) Side and (c) top view of the assembly formed by DSD. Two long helices from each of three DSD unit form a D_3-symmetric six helix bundle (axes displayed in (c)); the short helices pack on the exterior.

approximate D_2 (222) symmetry.[6,20,22] In one approach to automated protein design of four-helix bundles, the backbone is initially generated using this symmetry operator; as required for function, the sequence of the individual helices is next varied while maintaining the approximate symmetry of the overall fold.[20,22]

However, most natural proteins form assemblies considerably larger than those designed so far: the overall size of designed multimeric proteins is typically in the 10,000–20,000 Da range, with the largest protein being a five-helix bundle derived from HIV-1 gp 41.[37] Also, large multimeric proteins are often comprised of several independently folded domains that interact with exact relative orientations, while all designed proteins have either featured coiled-coils of single α-helices or monomeric single-domain structures. The goal of the present manuscript is to extend the use of symmetry in protein design to allow the design of complex assemblies of autonomously folded subunits. This is accomplished by stabilizing a D_3-symmetric arrangement observed in the unit cell of the crystal structure of a small, dimeric three-helix

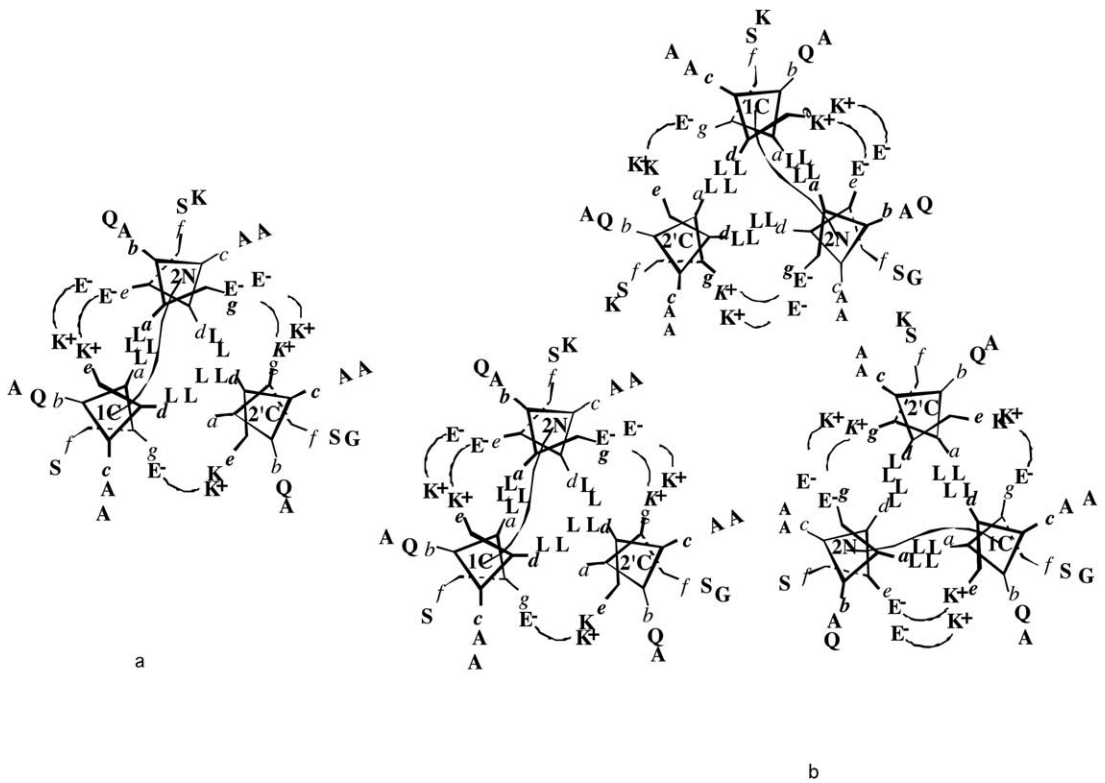


Figure 2. Helical wheel diagram of (a) the dimeric and (b) the hexameric assembly of DSD, showing the overall architecture. Because of the approximate 2-fold symmetry axes, only half of the structures are shown as viewed from the top of the assembly: specifically, residues 2–13 of helix 1, residues 21–33 of helix 2, and residues 35–47 of helix 2′.

bundle protein. The interactions observed in the crystal lattice were explicitly introduced in the design; salt-bridges stabilizing the inter-subunit contacts were replaced by specific packing of aromatic residues. The resulting assembly of six central helices forms a tubular protein of novel topology. This work has implications not only for the design of large proteins, but also for the design of molecular assemblies for nanotechnological applications.

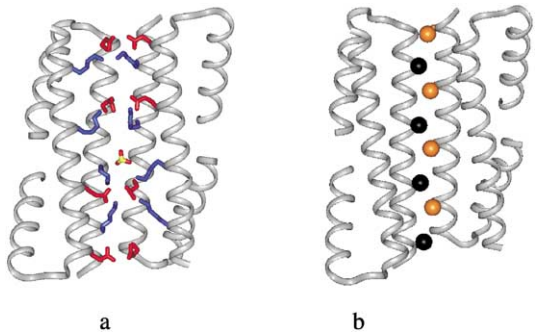


Figure 3. Interactions that stabilize the hexameric assembly, as observed in the crystal structure of DSD; for clarity, only two DSD units are displayed to expose the central cavity. (a) Eight alternating layers of Glu (red) and Lys (blue) residues stabilize the assembly through inter-twining electrostatic interactions clearly seen in the structure: Glu28 (Hel2) interacts with both Lys35′ and Lys42′ (Hel2′) with a distance of 2.8 Å between each Glu carboxylate oxygen atom and Lys N^{ε}. Also, corresponding interactions are observed between Lys35 (Hel2) and Glu28′ (Hel2′) (distance 3.8 Å). Interactions between residues on adjacent three-helix bundles are also observed: Lys35 (Hel2, dimer A) interacts with Glu28′ (Hel2′, dimer B) (observed distance 3.7 Å) and Glu21′ (Hel2′, dimer A) interacts with Lys42 (Hel2, dimer B) (observed distance 3.8 Å). Two Lys layers pair up in the middle because of the pseudo-symmetry of the structure: a sulfate ion is localized between these layers, along the central 3-fold symmetry axis. (b) Alacoil motif stabilizing the antiparallel interaction between pairs of helices: alanine residues from two helices (in orange and black, respectively) pack against each other.

Results

Design

The present design of a hexameric D_3-symmetrical bundle originated from an analysis of the packing of a domain-swapped three-helix bundle protein, DSD (Figure 1(a)). Because the crystal structure of DSD has not been described in detail, we will begin by discussing its structure, and the packing of this protein within a unit cell. Next, we will describe the strategy used to stabilize the hexameric arrangement observed in the solid state, such that it would also be stable in solution.

DSD[29] was originally designed as a model for domain swapping, a mechanism for dimerization/oligomerization observed in several natural proteins.[38] In domain swapping, a structural element of a protein is exchanged so that it interacts inter-molecularly rather than intra-molecularly, giving rise to an oligomer. The interactions between the swapped domain and the remaining structural elements of the protein are nearly identical in the oligomer and in the monomer. In DSD, a short helical element can be considered as the domain exchanged between two small monomeric three-helix bundles, resulting in the formation of a dimeric three-helix bundle of double length (Figure 1(a)). Thus, the monomeric unit of DSD is a helical hairpin, in which one of the helices, Hel2, is 28 residues long and the other, Hel1, is only 14 residues long. Hel1 and Hel2 are connected by a short loop. In the dimer, two long helices (Hel2 and Hel2′) pair up in an antiparallel manner, and Hel1 and Hel1′ dock against these to form an extended three-helix bundle, in which the third helix is interrupted. An approximate 2-fold axis of rotational symmetry is directed between the two short helices (Figure 1(a)).

The hydrophobic core of the dimer (Figure 2(a)), which provides the primary driving force for folding, is formed by leucine residues placed at the "a" and "d" positions of a heptad arrangement. The antiparallel arrangement of the design is stabilized by electrostatic interactions between Glu and Lys residues strategically placed on the sides of the helices. In particular, the residues at "g" positions of the two long helices (helix 2 and 2′, Figure 2(a)) form an extensive series of salt-bridges: the first two g residues in the long helices are Glu and the next two are Lys. Thus, they are able to interact favorably only when the long helices are docked in an antiparallel arrangement. Another feature included in the design was the incorporation of Ala at each c position. This residue was chosen for its favorable helix-forming tendency, and also to promote crystallization by forming lattice contacts similar to those observed in a parallel trimeric coiled-coil, coil-Val.[28] In coil-Val, the small Ala side-chains form a slightly apolar surface that is involved in inter-subunit contacts in the crystal. This surface is apparently hydrophilic enough to avoid unwanted aggregation in solution, but sufficiently hydrophobic to mediate weak crystal contact.

DSD crystallized in the space group $P2_13$, which has a 3-fold axis within the unit cell. The asymmetric unit in the structure is a single domain-swapped dimer, whose non-crystallographic 2-fold rotational axis is directed perpendicular to the crystallographic 3-fold axis. Thus, DSD forms a quasi-D_3-symmetric hexamer in the unit cell (Figure 1(b) and (c)). A schematic diagram of this hexamer is shown in Figure 2(b). The two long helices from each of three dimers line a central tubular cavity (Figure 1(c)). The third, short helices pack on the exterior of the central tube. The Ala side-chains at the c positions of the long helices help mediate the association of the dimers into the

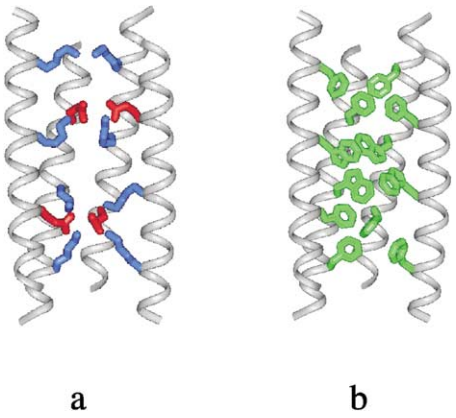


Figure 4. The alternating layers of Glu and Lys filling the central core of the hexameric assembly of DSD ((a) crystal structure) are replaced by layers of Trp and Phe in Hex-Phe ((b) computer model). For clarity, only the long helices of two DSD units are displayed.

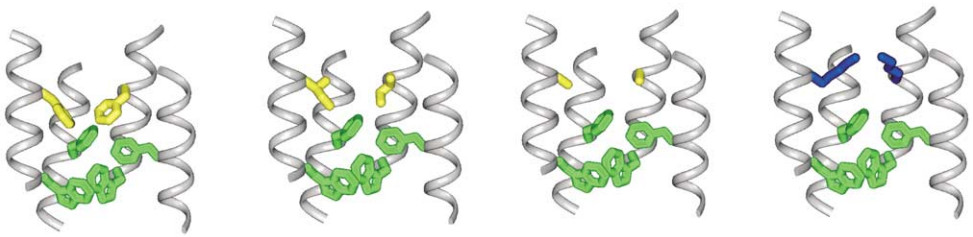
hexameric assembly (Figure 3(a)). The methyl groups along the long helices inter-digitate, forming a well-packed helix–helix interface. This packing motif, called the Alacoil, has been described previously;[23,39] many antiparallel coiled-coils and helical bundles exploit this sequence motif to allow close packing and stabilization of inter-helical contacts.[23,39]

The charged residues at the g positions of the long helices fill the central core of the hexameric structure (Figures 2(b) and 3(b)). We refer to this cylindrical core as a "supercore" to differentiate it from the hydrophobic Leu-rich core within the individual domain-swapped dimers. The charged residues in the supercore lie in alternate layers of positively (Lys) and negatively (Glu) charged residues, forming a well-defined hydrogen-bonded network. At the center of the hexamer, two layers of Lys are forced to lie adjacent to one other, because of the quasi-symmetry of the molecule; a crystallographic sulfate ion situated between these layers appears to minimize electrostatic repulsions between the Lys side-chains.

Although a stereochemically reasonable hexamer is observed in the crystal structure, DSD showed very little tendency to aggregate beyond a dimer in aqueous solution, even under conditions approaching those found in the crystallization buffer. We attribute the instability of the hexamer in solution to insufficient hydrophobic interactions in the supercore. Thus, while the core of the individual domain-swapped dimers is comprised of apolar Leu side-chains, the supercore of the hexamer is filled with charged Glu and Lys residues. As discussed previously, the interaction of buried salt-bridges is expected to be relatively unfavorable when compared to hydrophobic interactions between similarly sized residues.[40]

We generated a computer model based on the crystallographic coordinates of DSD, as a template for the design of a hexameric assembly that would be stable in solution. In the model, we introduced hydrophobic interactions in the supercore of the protein to replace the alternating layers of charged residues. The Alacoil was retained, and only a limited number of hydrophobic side-chains were introduced into the supercore. By minimizing the hydrophobicity of the supercore, we hoped to retain the dimer as the primary folding unit in the assembly of the hexamer. It was feared that excessive hydrophobicity in the supercore might lead to poor discrimination between the interfaces intended to stabilize the dimer *versus* the hexamer, thus leading to uncontrolled aggregation.

Based on the volumes of the side-chains within the supercore, aromatic residues were deemed to be the best suited to fill the central cavity of the hexamer. Beginning at the center of the supercore, Lys35 was mutated to Trp: this side-chain was found to be the only residue large enough to fill



	1	2	9	16	22	29	36	43
DSD	S	LAALKSE	LQALKKE	GFSPEE	LAALESE	LQALEKK	LAALKSK	LQALKG
Hex-Phe	S	LAALKSE	LQALKKE	GFSPEE	LAALESF	LQALEKW	LAALKSF	LQALKG
Hex-Ile	S	LAALKSE	LQALKKE	GFSPEE	LAALESF	LQALEKW	LAALKSI	LQALKG
Hex-Ala	S	LAALKSE	LQALKKE	GFSPEE	LAALESF	LQALEKW	LAALKSA	LQALKG
Hex-Lys	S	LAALKSE	LQALKKE	GFSPEE	LAALESF	LQALEKW	LAALKSK	LQALKG

Figure 5. Models of mutants in the Hex series: because of the symmetry, only half the core is displayed. In green, the conserved side-chains (one Trp and one Phe per monomer); in yellow, position 42 changed, respectively, to Phe in (a), Ile in (b), Ala in (c), and Lys, in blue, in (d). (e) shows the sequences of DSD, and of the Hex series.

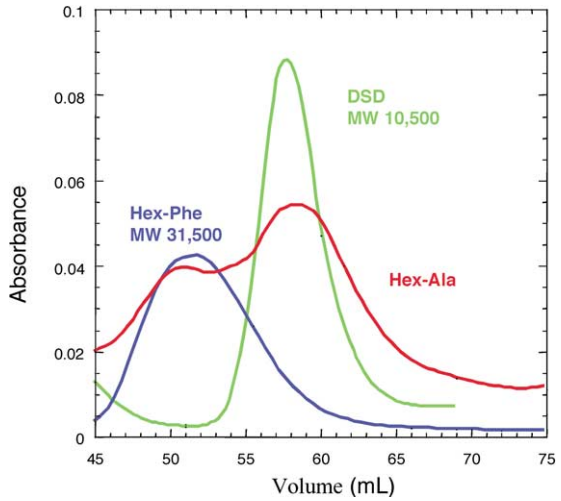


Figure 6. Gel filtration elution profile for Hex-Phe (blue) and Hex-Ala (red), contrasted with DSD (green), which forms a dimer with MW 10,500 Da. Hex-Phe elutes in one single peak at 50 ml, consistent with MW 31,500 Da, while the elution profile of Hex-Ala shows two peaks at 31,500 Da and 10,500 Da, respectively.

the cavity at this position, and it also serves as a fluorescent probe of structure and folding. Molecular modeling followed by rounds of minimization showed that the indole rings from each of the six side-chains tightly pack in the supercore. The layers immediately above and below position 35 were modified by replacing residue Glu28 with Phe, and a third supercore residue, Lys42 was changed to Phe. Because of the D_3 symmetry, these three changes result in the formation of six layers of aromatic residues: Figure 4 shows a comparison of the hydrophilic core of DSD and the redesigned aromatic core of Hex-Phe. We also designed three other mutants with decreasing buried hydrophobic surface by placing either Ile, Ala or Lys at position 42, in order to modulate the stability of the redesigned hexameric assemblies (Figure 5). These mutants will be referred to as Hex-Ile, Hex-Ala and Hex-Lys, respectively. Computer models were generated in which favorable rotamers were chosen for the Trp, Phe and Ile residues,[41] while Lys was left as observed in the structure; the geometry of the models was then optimized by energy minimization.

Oligomerization

The oligomerization states of the four mutants were assessed by gel filtration and by equilibrium analytical ultracentrifugation. Figure 6 shows the elution profiles for Hex-Phe and Hex-Ala. The proteins display clearly different behavior: Hex-Phe elutes as a single species at 50 ml, while Hex-Ala shows an additional peak at lower MW (60 ml) of equal intensity. An identical elution volume of 60 ml was obtained for the native DSD peptide, which is a dimer in solution in the 0.1–2 mM range, as assessed by analytical ultracentrifugation (MW 10,500). Thus, Hex-Ala forms two species in solution, one of which has a MW consistent with a dimeric three-helix bundle, and the other is a higher-order oligomer. Hex-Phe preferentially forms high-order oligomers of the same apparent MW as those formed by Hex-Ala.

The aggregation state in solution of the whole series was more rigorously determined by equilibrium analytical ultracentrifugation. At loading concentrations between 0.1 mM and 0.5 mM, the sedimentation curves for Hex-Ile and Hex-Phe are best described by a single species model with an apparent molecular weight of 32,700 Da and 33,500 Da, respectively, consistent within experimental error with the formation of a hexamer (31,212 Da and 31,308 Da, respectively). On the other hand, the data obtained for Hex-Lys and Hex-Ala were well described by a model in which a species with an apparent molecular mass of 10,500 Da is in equilibrium with its trimer (Figure 7). The dissociation constant for trimer formation are 1.2×10^{-7} M^2 for Hex-Lys and 6.3×10^{-10} M^2 for Hex-Ala, respectively. Thus, the midpoint of the oligomerization curve occurs approximately at 58 μM for Hex-Ala, and around 0.8 mM for Hex-Lys, respectively. The higher stability of Hex-Ile and Hex-Phe prevented an accurate evaluation of the equilibrium constant, but nevertheless indicates that the K_d for these peptides is below 10^{-13} M^2. Thus the peptides are mainly hexameric at peptide concentrations in the low micromolar range.

Fluorescence

A tryptophan residue at the center of the supercore provides a convenient fluorescence probe for investigating the structure and folding of the Hex series. In all the mutants, the midsection of the supercore is very similar; two central layers of Trp35 are sandwiched between a layer of Phe at both sides (Figures 4 and 5). The emission spectra of the Hex series at 2 μM, shown in Figure 8, are remarkably different, reflecting their different oligomerization states; the emission maximum ranges from 342 nm (Lys and Ala) to 334 nm (Ile) to 329 nm (Phe), indicating that Trp35 experiences an increasingly more apolar and/or rigid environment in the series. At this concentration Hex-Ala and Hex-Lys are more than 95% dimeric, while for Hex-Ile and Hex-Phe the hexameric form is predominant. Taking Hex-Phe as the most stable hexamer in the series, we provisionally assigned the value of 329 nm as λ_{max} for the hexamer, and approximately 342 nm as λ_{max} for the dimer. Both values are significantly lower than the values expected for a fully exposed tryptophan residue (352 nm), indicating that the indole groups are partially shielded from solvent, even in the dimer. Further, the significantly greater blue shift in λ_{max} for Hex-Phe relative to Hex-Lys is consistent with the greater burial of the Trp in the core of the hexameric bundle, relative to the dimer.

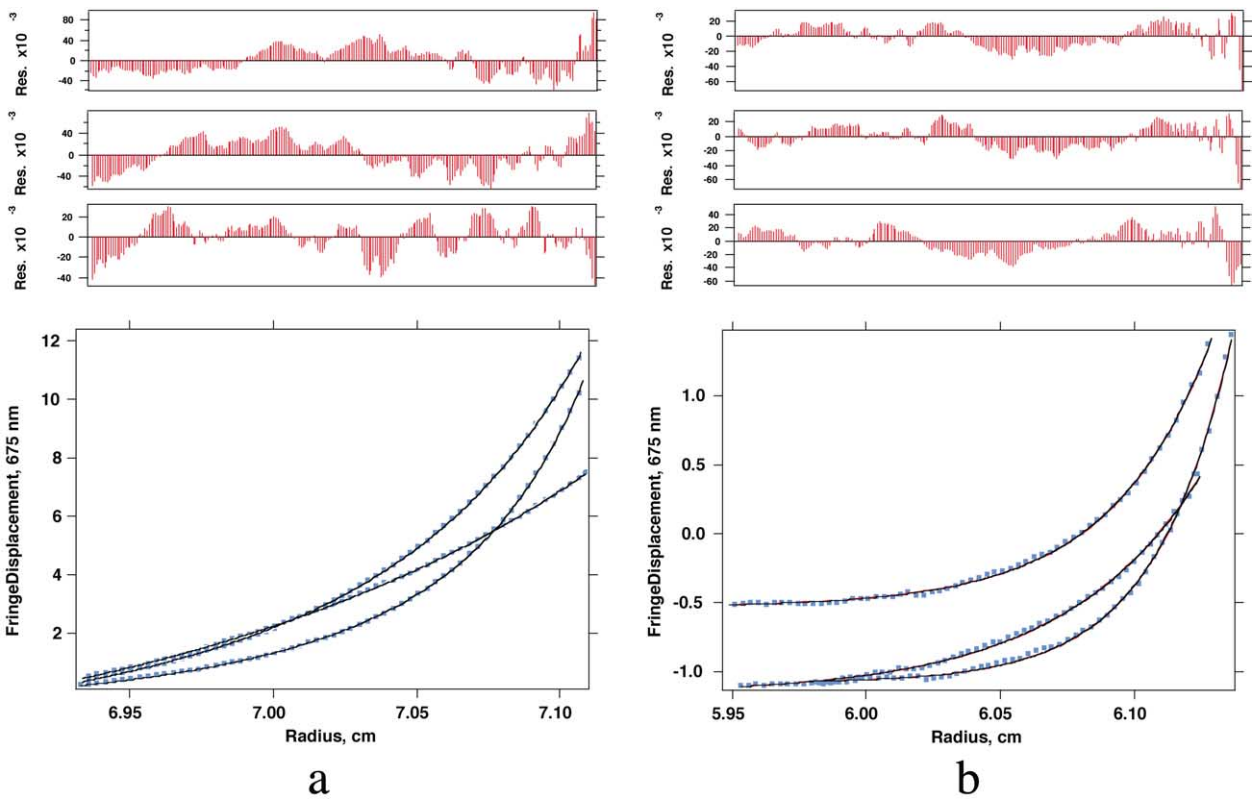


Figure 7. Analytical ultracentrifugation sedimentation profile for Hex-Lys and Hex-Phe. The data for (a) Hex-Ala could be analyzed with a dimer–hexamer equilibrium model, obtaining a K_d of 1.2×10^{-7} M^2; a single species model was used for (b) Hex-Phe, yielding an apparent molecular mass of 33,500 Da. A global fit to three data sets, collected at 35,000, 40,000 and 45,000 K, respectively, is shown. Conditions: peptide loading concentration 0.1 mM, 0.05 M sodium phosphate buffer (pH 7.2), 0.1 M NaCl.

Stability

The four mutants are highly helical (>90%) as assessed from circular dichroism spectroscopy, showing minima at 208 nm and 222 nm, typical of α-helical structure. Similar to the native DSD peptide, the entire series of variants is extremely stable to thermal denaturation: solutions of the peptides at approximately 2 μM are more than 90% structured at 94 °C. The overall stability of the peptides was evaluated by chemical denaturation, measuring the variation of $[\theta_{222}]$ as a function of the concentration of added guanidinium hydrochloride (Gdn). At 2 μM, the stability of the series increases in the order DSD < Hex-Lys < Hex-Ala < Hex-Ile < Hex-Phe (Figure 9). In multimeric systems, folding and oligomerization are often thermodynamically linked,[42] so that the stability of the protein to denaturation is concentration dependent. A global analysis of the denaturation curves at two concentrations for each mutant was used to obtain a more accurate value of the free energy of folding, $\Delta G°$, for the association (Figure 9).[43] Hex-Lys and Hex-Ala are dimers at the experimental conditions, thus a simple monomer–dimer model was used to fit the data (Figure 9(a) and (b)): the values of $\Delta G°$ extrapolated to 0 M Gdn are −17.8(± 0.3) kcal/mol and −18.7(± 0.4) kcal/mol, and the corresponding m values are 1.8 kcal/mol M and 1.7 kcal/mol M, respectively (1 cal = 4.184 J).

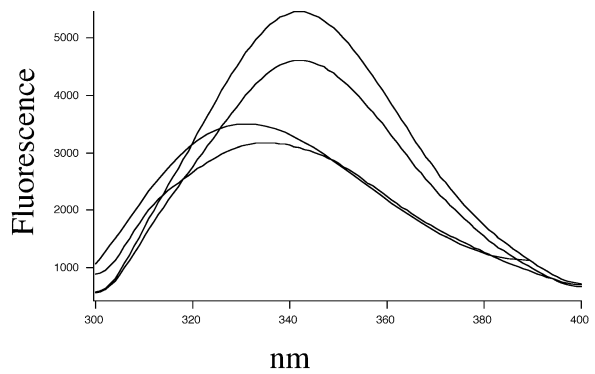


Figure 8. Fluorescence spectra of mutants: the emission maximum is 342 nm for Hex-Ala and Hex-Lys, 334 nm for Hex-Ile and 329 nm for Hex-Phe. Conditions: peptide concentration 2 μM, 0.05 M sodium phosphate buffer (pH 7.2), 0.1 M NaCl.

At the low micromolar range used for the Gdn-dependent denaturation, Hex-Ile and Hex-Phe exist in solution as a mixture of hexamer and dimer, with a K_d for trimerization of the dimer estimated to be lower than 10^{-13} M^2. Thus, the model to be used to fit the denaturation curves at two concentrations should include a consideration of both equilibria, monomer–dimer and

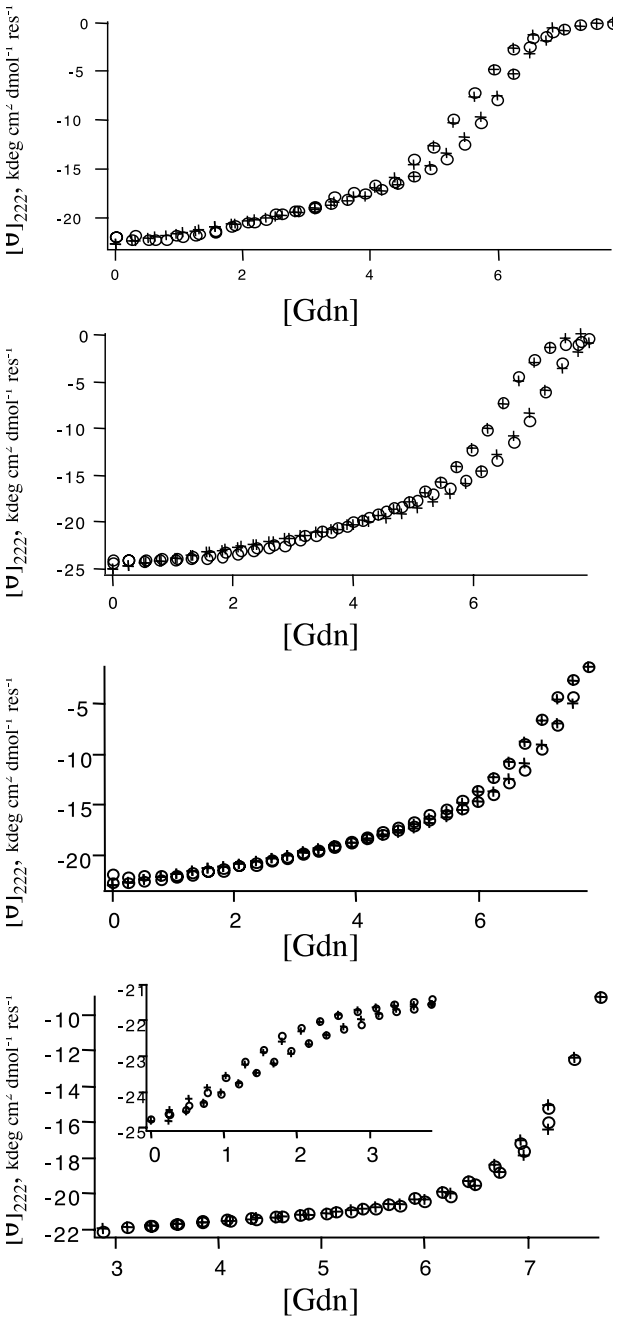


Figure 9. Guanidinium hydrochloride denaturation curves showing the experimental data (open circles) and the theoretical fits (crosses); the data were collected at two peptide concentrations, approximately 2 μM and 10 μM, for each mutant and analyzed globally to obtain a more accurate value of ΔG°. A monomer–dimer equilibrium model was used for (a) Hex-Lys, (b) Hex-Ala, and (c) Hex-Ile; (d) the same model could be used for Hex-Phe to fit data collected at [Gdn] above 3 M, while a dimer–hexamer equilibrium, consistent with a dimer–hexamer association was used to fit data at [Gdn] below 3 M ((d) inset).

dimer–hexamer. It is possible that the peptides would dissociate from hexamers to dimers, prior to the main transition. Indeed, the data for Hex-Ile are well described by a monomer–dimer equilibrium (Figure 9(c)), obtaining a ΔG° of $-19.4(\pm 0.3)$ kcal/mol, and an m value of 1.7 kcal/mol M, but not by a monomer–hexamer scheme. Thus, the transition from hexamer to dimer appears to occur significantly before the main transition, and does not involve a large enough change in $[\theta_{222}]$ to be experimentally observed.

Interestingly, a different result is observed for Hex-Phe, which was expected to form the most stable hexamer. The data for this variant conforms to a monomer–dimer equilibrium only if data are considered at concentrations of Gdn greater than 3 M (Figure 9(d)). The resulting fit gives a ΔG° of $-25.0(\pm 1)$ kcal/mol, and an m value of 2.2 kcal/mol M. The treatment is complicated by the stability of the proteins: at 2 μM, in fact, both proteins are still partially structured even at 7.8 M Gdn. The unfolded baselines are well defined only for Hex-Lys; the other mutants are still partially folded at 7.8 M Gdn. Thus, we utilized the experimental values obtained for Hex-Lys in fitting the denaturation curves of Hex-Ala, Hex-Ile and Hex-Phe as well (see Materials and Methods).

The pre-transition in the Gdn denaturation curve observed for Hex-Phe (Figure 9(d), inset) was treated separately. The midpoint for this transition depends on the total peptide concentration, confirming that the observed transition involves a change in the aggregation state of the peptide. Thus, the hexamer of Hex-Phe would undergo a discreet transition to dimer at low Gdn concentrations; in turn, the dimer would unfold at much higher [Gdn]. It is possible to confirm the aggregation state of the peptide by globally analyzing [Gdn] denaturation curves at multiple peptide concentration.[44] A global analysis of the pre- and main-transitions indeed supported the proposed hexamer–dimer–monomer scheme. The ΔG° obtained for the dimer–hexamer equilibrium is -17.8 kcal/mol, with an m value of 0.47 kcal/mol M on a per dimer basis, consistent with a relatively small change in the solvent accessibility of the hydrophobic core between the hexamer and the dimer (as opposed to the larger value expected for full denaturation of the entire assembly). The corresponding K_d of trimerization derived from the Gdn denaturation, 6×10^{-14} M^2, is in reasonable agreement with the limit of $<10^{-13}$ M^2 obtained from the ultracentrifugation experiments.

To further explore this hypothesis, we contrasted the fluorescence spectra of Hex-Phe and Hex-Ala at different Gdn concentrations in the 0–7 M range, reasoning that Gdn should destabilize the hexamer at lower concentration than the dimer. The spectrum of Hex-Ala shows very little change until the Gdn concentration is above 6 M, after which an abrupt red shift occurs (Figure 10(b)); on the other hand, Hex-Phe undergoes a gradual red shift at Gdn concentrations as low as 2 M (Figure 10(a)). The spectra of the unfolded peptides at 8 M Gdn are very similar. Thus, while Hex-Ala unfolds

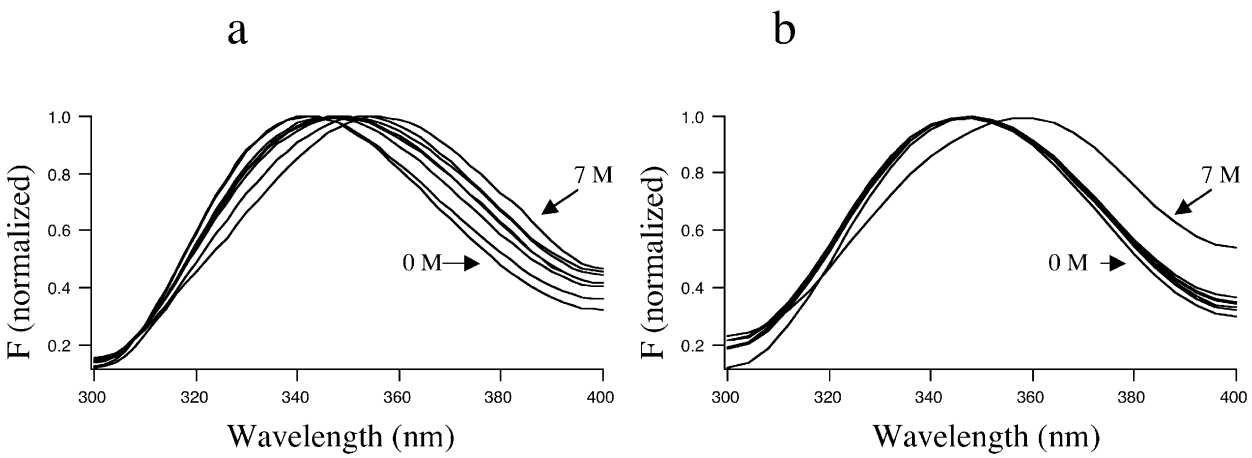


Figure 10. Fluorescence emission spectra of a 2 μM solution of (a) Hex-Phe and (b) Hex-Ala at increasing Gdn concentrations in the 0–7 M range.

with a simple two-state equilibrium, Hex-Phe shows more transitions.

It is interesting to compare the stability of the dimeric forms of variants. The ΔG obtained for Hex-Lys, the least stable of the series, is −17.5 kcal/mol; the corresponding free energy of unfolding for the native DSD, also a dimer, was −13.6 kcal/mol. There are two mutations between the sequence of DSD and Hex-Lys: a Lys and a Glu in DSD are substituted with a Trp and a Phe, respectively. In the dimer, these residues are only partially buried at interfacial positions near the surface of the protein. At higher concentrations, this aromatic cluster mediates the further association into a hexameric assembly.

Conclusions

Here we describe a hierarchic approach to the design of large molecular assemblies. A domain-swapped dimer is the structural unit of the design. The packing of this dimeric three-helix bundle in the crystal lattice provided valuable clues to the potential formation of higher-order oligomers, which might be obtained from the dimeric unit by applying a C_3 symmetry operator. This results in the formation of a hexameric assembly, which was not observed in solution. To stabilize the hexameric state in solution, it was necessary to build a second hydrophobic supercore. However, it was deemed important to not make the supercore too hydrophobic, otherwise the protein might have low solubility or lack conformational specificity. Previous surveys of protein oligomerization sites have shown that they tend to be significantly less hydrophobic than the interiors of individual subunits or of monomolecularly folded proteins (for a review, see Jones & Thornton[45]). Indeed, the introduction of only three hydrophobic side-chains per monomer results in the formation of a stable hexamer. This issue was further explored by designing a series of single-point mutants in which the hydrophobicity of the core was varied and monitoring how it affected the stability of the hexameric structure.

The assembly of the hexamers reflects the hierarchic nature of the design: two of the mutants (Hex-Ala and Hex-Lys) display a concentration-dependent dimer/hexamer equilibrium, while the two most stable mutants, Hex-Ile and Hex-Phe, are hexameric at all the experimentally accessible concentrations. However, for Hex-Phe a similar behavior can be brought out in the presence of denaturant. A pre-transition occurs at low concentration of guanidinium. The peptide concentration dependence of this pre-transition indicates that it is associated with a change in the aggregation state. Thus, as the concentration of denaturant is increased, the hexamers dissociate to dimers in an initial step, and the dimers unfold at significantly higher concentrations of denaturant. This is similar to the assembly of the hexameric enzyme 4-oxalocrotonate tautomerase,[46] which shows discrete intermediate states of association in a pH-dependent manner; specifically, a stable dimer was observed at pH 4.8 (M.C. Fitzgerald, personal communication). Similarly, many dimeric proteins unfold in a three-step process with a folded monomer as an intermediate.[47]

In conclusion, these studies establish methods for the design of high-order assemblies of helical bundles. It is interesting to note that it was not necessary to pack the entire supercore to achieve a hexameric structure. Thus, there may be sufficient space at the ends of the bundle to accommodate small apolar ligands. It will be particularly interesting to determine whether Hex-Ala or a redesigned version of this protein will undergo a dimer-to-hexamer transition in response to the addition of small molecule ligands. The proteins designed in this work also may have implications for the construction of nanotechnological devices. Along these lines, we note that Padilla *et al.*[48] have described a related different approach to the design of protein polyhedra. By combining different naturally occurring oligomeric proteins, these

workers were able to construct hybrid proteins that form tubular and virus-like assemblies.

Materials and Methods

Design

Computer models were generated starting from the coordinates from the crystal structure of DSD (PDB entry 1G6U) using InsightII and minimized using cvff as implemented in Discover (Accelrys).

Materials

Fmoc-protected amino acids (Fmoc: 9-fluorenylmethoxycarbonyl), 5[4-(aminomethyl)-3,5-bis(methoxy)-phenoxy]valeric acid (PAL) resin, *N*-hydroxybenzotriazole (HOBt), and 2-(1H-benzotriazole-1-yl)-1,1,3,3-tetramethyluroniumhexafluorophosphate (HBTU) were purchased from NovaBiochem. All solvents and chemicals used in peptide synthesis and purification were of the highest available grade and were used without further purification.

Synthesis

The peptides were synthesized with standard solid phase procedures on an ABI 433 synthesizer (PE Applied Biosystems) equipped with a UV detector to monitor Fmoc deprotection (Alltech) and purified by reverse-phase HPLC on a semipreparative C_{18} column (Vydac). All peptides were acetylated at the amino terminus. The peptides were determined to be at least 95% pure by analytical HPLC; MALDI mass spectrometry confirmed the expected molecular mass and purity.

Determination of aggregation state

Sedimentation equilibrium analysis was performed using a Beckman XLI analytical ultracentrifuge. Initial peptide concentrations were 0.1 mM in 0.01 M sodium phosphate (pH 7.2), 0.05 M NaCl. The samples were centrifuged at 35,000, 40,000 and 45,000 rpm; equilibrium was determined when successive interference radial scans at the same speed were indistinguishable. Partial specific volumes were determined by the residue-weight average method of Cohn & Edsall.[49,50] Solution densities were estimated using solute concentration-dependent density tables in the CRC Handbook of Chemistry and Physics. The aggregation state was determined by dimer–hexamer equilibria; the software allows quantification of the components.[51] The data were also treated as single species to provide an estimate of the aggregation states. Curve fitting to the data was done using Igor Pro (WaveMetrics, Inc.) with procedures adapted from Brooks *et al.*[52]

Gel filtration elution profiles were obtained on a Superdex 75 column on an FPLC system (Amersham Pharmacia Biosystems); typically, 0.5 mg of peptide was loaded and eluted with 0.01 M sodium phosphate (pH 7.2), 0.2 M NaCl at 0.5 ml/minute.

CD measurements

Gdn denaturations were carried out at 25 °C using a CD spectrometer equipped with a dual-syringe automated titrator (Aviv Associates) recording the ellipticity at 222 nm *versus* Gdn concentration at two peptide concentrations for each mutant (conditions: 0.01 M sodium phosphate buffer (pH 7.0), 0.1 M NaCl; peptide concentrations were approximately 1.5 μM and 4.5 μM for Hex-Lys, 2.3 μM and 10.5 μM for Hex-Ala, 1.7 μM and 4.1 μM for Hex-Ile, and 3.9 μM and 7.9 μM for Hex-Phe). The curves were analyzed globally using a dimerization-linked folding model in Igor Pro (WaveMetrics, Inc.).[43] The unfolded baselines of Hex-Phe and Hex-Ile were not well defined, as the peptides were still partially folded at 7.8 M Gdn. Therefore, we chose to use the unfolded baseline experimentally obtained for Hex-Lys as fixed parameters for all the mutants. The data between 0 M and 3 M guanidinium for Hex-Phe were additionally fit to a dimer–hexamer equilibrium.

Fluorescence

The fluorescence intensity of the Hex series was monitored with a Fluorolog spectrofluorometer (model 3) equipped with a Peltier thermostated cell holder at 25 °C with excitation at 280 nm (5 nm band-pass); emission scans were acquired in the 300–400 nm range (2 nm band-pass) with 0.5 nm steps, and an average time of two seconds. The position of the emission maximum was determined by calculating the first derivative of the intensities *versus* wavelengths. An Aviv Associates fluorometer equipped with a dual-syringe automated titrator was used for the Gdn-dependent experiment; emission scans were acquired in the 300–400 nm range (4 nm band-pass) with 2 nm steps, exciting the sample at 280 nm (4 nm band-pass) and averaging the signal for one second. A solution of *N*-acetyl-L-tryptophanamide (5 μM) in phosphate buffer at pH 7 was used as standard to calibrate the instrument (emission max at 352 nm).

Acknowledgments

We thank Jane Vanderkooi and Marcos Milla for use of the fluorimeters, Michael C. Fitzgerald for sharing unpublished data, and Vikas Nanda for helpful discussions. This work was supported by grants from the NIH (GM. 54616) and NSF (DMR79909 and MCB 94-20769).

References

1. Goodsell, D. S. & Olson, A. J. (2000). Structural symmetry and protein function. *Annu. Rev. Biophys. Biomol. Struct.* **29**, 105–153.
2. Kamtekar, S. & Hecht, M. H. (1995). Protein motifs. 7. The four-helix bundle: what determines a fold? *FASEB J.* **9**, 1013–1022.
3. Hill, R. B., Raleigh, D. P., Lombardi, A. & DeGrado, W. F. (2000). *De novo* design of helical bundles as models for understanding protein folding and function. *Acc. Chem. Res.* **33**, 745–754.
4. Betz, S. F., Liebman, P. A. & DeGrado, W. F. (1997). *De novo* design of native proteins: characterization of proteins intended to fold into antiparallel, rope-like, four-helix bundles. *Biochemistry*, **36**, 2450–2458.
5. Schafmeister, C. E., LaPorte, S. L., Miercke, L. J. & Stroud, R. M. (1997). A designed four helix bundle protein with native-like structure. *Nature Struct. Biol.* **4**, 1039–1046.

6. Schafmeister, C. E. & Stroud, R. M. (1998). Helical protein design. *Curr. Opin. Biotechnol.* **9**, 350–353.
7. 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.
8. Stowell, M. H. & Rees, D. C. (1995). Structure and stability of membrane proteins. *Advan. Protein Chem.* **46**, 279–311.
9. Koebnik, R., Locher, K. P. & Van Gelder, P. (2000). Structure and function of bacterial outer membrane proteins: barrels in a nutshell. *Mol. Microbiol.* **37**, 239–253.
10. Koronakis, V., Andersen, C. & Hughes, C. (2001). Channel-tunnels. *Curr. Opin. Struct. Biol.* **11**, 403–411.
11. Postle, K. & Vakharia, H. (2000). TolC, a macromolecular periplasmatic "channel". *Nature Struct. Biol.* **7**, 527–531.
12. Fulop, V. & Jones, D. T. (1999). Beta propellers: structural rigidity and functional diversity. *Curr. Opin. Struct. Biol.* **9**, 715–721.
13. Branden, C.-I. (1991). The TIM barrel—the most frequently occurring folding motif in proteins. *Curr. Opin. Struct. Biol.* **1**, 978–983.
14. Murzin, A. G., Lesk, A. M. & Chothia, C. (1994). Principles determining the structure of beta-sheet barrels in proteins. II. The observed structures. *J. Mol. Biol.* **236**, 1382–1400.
15. Murzin, A. G., Lesk, A. M. & Chothia, C. (1994). Principles determining the structure of beta-sheet barrels in proteins. I. A theoretical analysis. *J. Mol. Biol.* **236**, 1369–1381.
16. Silverman, J. A., Balakrishnan, R. & Harbury, P. B. (2001). Reverse engineering the (beta/alpha)8 barrel fold. *Proc. Natl Acad. Sci. USA*, **98**, 3092–3097.
17. Wierenga, R. K. (2001). The TIM-barrel fold: a versatile framework for efficient enzymes. *FEBS Letters*, **492**, 193–198.
18. Farber, G. K. & Petsko, G. A. (1990). The evolution of alpha/beta barrel enzymes. *Trends Biochem. Sci.*, **15**, 228–234.
19. McLachlan, A. D. (1987). Gene duplication and the origin of repetitive protein structures. *Cold Spring Harbor Symp. Quant. Biol.*, vol. 52
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. Crick, F. H. C. (1953). The packing of alpha-helices: simple coiled-coils. *Acta Crystallog.* **6**, 689–697.
22. 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.
23. North, B., Summa, C. M., Ghirlanda, G. & DeGrado, W. F. (2001). D_n-symmetrical tertiary templates for the design of tubular proteins. *J. Mol. Biol.* **311**, 1081–1090.
24. Lupas, A. (1996). Coiled coils: new structures and new functions. *Trends Biochem. Sci.*, **21**, 375–382.
25. 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.
26. Harbury, P. A. B. (1998). Springs and zippers: coiled coils in SNARE-mediated membrane fusion. *Structure*, **6**, 1487–1491.
27. Hodges, R. S. (1996). *De novo* design of alpha-helical proteins: basic research to medical applications. *Biochem. Cell. Biol.* **74**, 133–154.
28. 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.
29. 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 proteins. *Proc. Natl Acad. Sci. USA*, **98**, 1404–1409.
30. Lovejoy, B., Choe, S., Cascio, D., McRorie, D. K., DeGrado, W. F. & Eisenberg, D. (1993). Crystal structure of a synthetic triple-stranded alpha-helical bundle. *Science*, **259**, 1288–1293.
31. Gonzalez, L., Jr, Brown, R. A., Richardson, D. & Alber, T. (1996). Crystal structures of a single coiled-coil peptide in two oligomeric states reveal the basis for structural polymorphism. *Nature Struct. Biol.* **3**, 1002–1009.
32. Lumb, K. J. & Kim, P. S. (1995). Measurement of interhelical electrostatic interactions in the GCN4 leucine zipper. *Science*, **268**, 436–439.
33. Sharma, V. A., Logan, J., King, D. S., White, R. & Alber, T. (1998). Sequence-based design of a peptide probe for the APC tumor suppressor protein. *Curr. Biol.* **8**, 823–830.
34. 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.
35. Nautiyal, S. & Alber, T. (1999). Crystal structure of a designed, thermostable, heterotrimeric coiled coil. *Protein Sci.* **8**, 84–90.
36. Eckert, D. M., Malashkevich, V. N. & Kim, P. S. (1998). Crystal structure of GCN4-pIQI, a trimeric coiled coil with buried polar residues. *J. Mol. Biol.* **284**, 859–865.
37. Root, M. J., Kay, M. S. & Kim, P. S. (2001). Protein design of an HIV-l entry inhibitor. *Science*, **291**, 884–888.
38. Schlunegger, M. P., Bennett, M. J. & Eisenberg, D. (1997). Oligomer formation by 3D domain swapping: a model for protein assembly and misassembly. *Advan. Protein Chem.* **50**, 61–122.
39. 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.
40. Hendsch, Z. S. & Tidor, B. (1994). Do salt bridges stabilize proteins? A continuum electrostatic analysis. *Protein Sci.* **3**, 211–226.
41. Dunbrack, R. L., Jr & Karplus, M. (1994). Conformational analysis of the backbone-dependent rotamer preferences of protein sidechains. *Nature Struct. Biol.* **1**, 334–340.
42. Zitzewitz, J. A., Bilsel, O., Luo, J., Jones, B. E. & Matthews, C. R. (1995). Probing the folding mechanism of a leucine zipper peptide by stopped-flow circular dichroism spectroscopy. *Biochemistry*, **34**, 12812–12819.
43. Ghirlanda, G., Lear, J. D., Lombardi, A. & DeGrado, W. F. (1998). From synthetic coiled coils to functional proteins: automated design of a receptor for the calmodulin-binding domain of calcineurin. *J. Mol. Biol.* **281**, 379–391.
44. 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.

45. Jones, S. & Thornton, J. M. (1996). Principles of protein–protein interactions. *Proc. Natl Acad. Sci. USA*, **93**, 13–20.
46. Silinski, P., Allingham, M. J. & Fitzgerald, M. C. (2001). Guanidine-induced equilibrium unfolding of a homo-hexameric enzyme 4-oxalocrotonate tautomerase (4-OT). *Biochemistry*, **40**, 4493–4502.
47. Neet, K. E. & Timm, D. E. (1994). Conformational stability of dimeric proteins: quantitative studies by equilibrium denaturation. *Protein Sci.* **3**, 2167–2174.
48. Padilla, J. E., Colovos, C. & Yeates, T. O. (2001). Nanohedra: using symmetry to design self-assembling protein cages, layers, crystals and filaments. *Proc. Natl Acad. Sci. USA*, **98**, 2217–2221.
49. Harding, S. E., Rowe, A. J. & Horton, J. C. (1992). *Analytical Ultracentrifugation in Biochemistry and Polymer Science*, The Royal Society of Chemistry, Cambridge, UK.
50. Cohn, E. J. & Edsall, J. T. (1943). *Proteins, Amino acids and Peptides as Ions and Dipolar Ions*, Reinhold Publishing Corp, New York.
51. Scott, C. P., Kashlan, O. B., Lear, J. D. & Cooperman, B. S. (2001). A quantitative model for allosteric control of purine reduction by murine ribonucleotide reductase. *Biochemistry*, **40**, 1651–1661.
52. Brooks, I. S., Soneson, K. K. & Hensley, P. (1993). Developement and use of a Mac based data analysis package for equilibrium sedimentation data from the analytical ultracentrifuge. *Biophys. J.* **64**, 244.

Edited by P. Wright

(Received 7 November 2001; received in revised form 3 March 2002; accepted 7 March 2002)