De Novo Design of Native Proteins: Characterization of Proteins Intended To Fold into Antiparallel, Rop-like, Four-Helix Bundles †

ARTICLE *in* BIOCHEMISTRY · APRIL 1997

Impact Factor: 3.02 · DOI: 10.1021/bi961704h · Source: PubMed

De Novo Design of Native Proteins: Characterization of Proteins Intended To Fold into Antiparallel, Rop-like, Four-Helix Bundles†

Stephen F. Betz,‡ Paul A. Liebman, and William F. DeGrado*

The Johnson Research Foundation, Department of Biochemistry and Biophysics, University of Pennsylvania, Philadelphia, Pennsylvania 19104-6059

Received July 12, 1996; Revised Manuscript Received October 22, 1996⊗

ABSTRACT: The *de novo* design and characterization of a series of 51-residue helix−turn−helix peptides intended to dimerize into antiparallel four-stranded coiled coils is described. The sequence is based on a coiled coil heptad repeat N_{cap}-$(A_aZ_bZ_cL_dZ_eZ_fZ_g)_3$−turn− $(X_aZ_bZ_cL_dZ_eZ_fZ_g)_3$-$C_{cap}$-$CONH_2$, where X is either Val or Ala. The overall topology was intended to be similar to that found in the *Escherichia coli* protein ROP. The design strategy included consideration of geometric complementarity of the packing of side chains within the hydrophobic core as well as the use of specific interfacial interactions, both of which were intended to favor the desired ROP-like topology. Additionally, the sequence was designed to destabilize potential alternative structures that might compete with the desired topology. The peptides (RLP-1, RLP-2, and RLP-3) assemble into stable α-helical dimers and exhibit the hallmarks of a native protein as judged by its spectroscopic properties, and the lack of binding to hydrophobic dyes. Also, the enthalpy and heat capacity changes upon denaturation were determined by measuring the temperature dependence of the CD spectra and confirmed by differential scanning calorimetry (DSC). The values determined by the two methods are in excellent agreement and are in the range of those of naturally occurring proteins of this size. These results suggest that it is now possible to design native-like helical proteins that should serve as templates for the further design of functional proteins.

Significant progress has been achieved in the design of proteins from first principles (Nautiyal et al., 1995; Olofsson et al., 1995; Raleigh et al., 1995; Betz & DeGrado, 1996). Recently, we described a series of fundamental principles for the hierarchic *de novo* design of α-helical bundles (Betz & DeGrado, 1996). Three principles guide the design process and include (i) the use of complementary packing within the hydrophobic core, (ii) the use of specific interfacial interactions to promote interhelical association, and (iii) the use of both core and interfacial interactions to destabilize potential alternate conformations.

Here, we describe the design and characterization of helix−turn−helix peptides designed to dimerize into a four-helix bundle that mimics the topology found in the *Escherichia coli* protein ROP (Banner et al., 1987). ROP is composed of two 63-residue peptides that form an antiparallel four-stranded coiled coil in which the two connecting loops are on opposite sides of the structure (Figure 1). The proteins described here are designed to be an abstraction of that structure and have the sequence N_{cap}-$(A_aZ_bZ_cL_dZ_eZ_fZ_g)_3$−turn−$(X_aZ_bZ_cL_dZ_eZ_fZ_g)_3$-$C_{cap}$-$CONH_2$, where the **a**−**g** subscripts refer to the canonical coiled coil heptad repeat position (Cohen & Parry, 1990) and X is either alanine or valine (Figure 2). Three different peptides, RLP-1, -2, and -3 (where the number refers to the number of valine residues in the second helix), were synthesized and shown to form dimers in solution, consistent with their design. The behavior of these proteins is compared to that of naturally occurring proteins and other designed α-helical bundles.

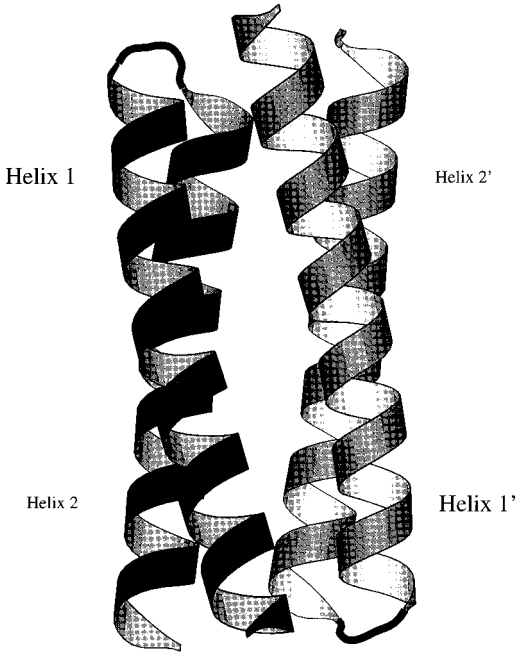

FIGURE 1: Molscript diagram (Kraulis, 1991) showing the structure and helix nomenclature of the *E. coli* protein ROP (Banner et al., 1987).

MATERIALS AND METHODS

Generation of α-Helical Bundles. The design of the RLP series of proteins followed the procedures described previ-

† S.F.B. and W.F.D. were supported in part by the MRSEC Program of the National Science Foundation under Award DMR96-32598 and by National Institutes of Health Grant GM54616-01. P.A.L. was supported by National Institutes of Health Grants EY00012 and EY01583.

* To whom correspondence should be addressed. Phone: 215-898-4590. Fax: 215-573-7229. E-mail: wdegrado@mail.med.upenn.edu.

‡ Current address: Abbott Laboratories, 247G, 100 Abbott Park Road, Abbott Park, IL 60064.

⊗ Abstract published in *Advance ACS Abstracts,* January 15, 1997.

S0006-2960(96)01704-7 CCC: $14.00

		Heptad Position						
Protein		a	b	c	d	e	f	g
RLP	S	A	Q	E	L	L	K	I
		A	R	R	L	R	K	E
		A	K	E	L	L	K	R
		A	E	H				
	G	G	P	E	L	L	K	E
		a	E	E	L	E	K	K
		b	D	K	L	Y	K	I
		c	E	H	G			
α_1B	G E	L	E	E	L	L	K	K
		L	K	E	L	L	K	G

	a	*b*	*c*
RLP-1	A	V	A
RLP-2	V	A	V
RLP-3	V	V	V

FIGURE 2: Sequence of the RLP series of peptides. The sequence is arranged by heptads, with variable **a** positions indicated by *a*, *b*, or *c*. The N terminus is a free amine, and the C terminus is amidated. The sequence of the early designed peptide α_1B (Ho & DeGrado, 1987) is shown for comparison.

ously (Betz & DeGrado, 1996). Briefly, single polyalanine helices were created using the programs PSSHOW and Insight II (Biosym) and then brought together and transformed to form coiled coils of various topologies. The hydrophobic core residues (by convention, positions **a** and **d**) were placed in low-energy conformations (McGregor et al., 1987) and their packing interactions critically examined. The interfacial and surface residues were then placed accordingly, and the entire model was subjected to energy minimization using the default force field in the program Discover (Biosym) (Betz & DeGrado, 1996). The resulting structures were examined for poor contacts, the appearance of which led to sequence refinement and further rounds of modeling and minimization.

Synthesis and Purification. Peptides were synthesized by standard automated methods on a Milligen 9050 peptide synthesizer using FMOC-protected amino acids and purified by reverse phase HPLC (Choma et al., 1994). The peptides were determined to be homogeneous by analytical HPLC and electrospray ionization mass spectrometry.

Analytical Ultracentrifugation. Sedimentation equilibrium analysis was performed with a Beckman XLA analytical ultracentrifuge (Harding et al., 1992). Loading peptide concentrations ranged from 25 to 750 μM in 20 mM MOPS (with or without 100 mM NaCl) at pH 6.9, 20 mM MES at pH 5.5, or 50 mM NaOAc at pH 4.0. (There is no discernible difference in behavior with added NaCl.) The samples were centrifuged at 35 000, 42 000, and 48 000 rpm. Equilibrium was determined when successive radial absorbance scans were indistinguishable.

The behavior of a single species at equilibrium can be described by the following equation:

$$M_b = M_w(1 - \bar{v}\rho) \quad (1)$$

where M_b is the measured buoyant molecular mass, M_w is the molecular mass in daltons, $\bar{v}$ the partial specific volume of the solute, and ρ is the density of the sample solution. The partial specific volume was calculated using the weighted average of the amino acid content using the method of Cohn and Edsall (1943).

The equilibrium aggregation state was determined at relatively low initial loading concentrations by fitting the data to various monomer−*n*mer equilibrium schemes. At higher initial loading concentrations, the association was complete so the data were fit to a single species using the program Igor Pro (WaveMetrics, Inc.).

Fluorescence Measurements. Fluorescence spectra of peptide complexes with 8-anilino-1-naphthalenesulfonic acid (ANS)[1] were measured using a Spex Fluorolog fluorimeter at 298 K. ANS−peptide complexes were monitored by titrating peptide into a constant amount (2 μM) of ANS in a 1 cm cuvette. Buffer conditions were 20 mM HEPES at pH 6.9. The excitation wavelength was 370 nm with emission spectra recorded from 420 to 500 nm.

Circular Dichroism. The circular dichroism of peptide solutions was monitored with an AVIV 62DS spectropolarimeter. The concentrations of peptide stock solutions were determined by absorbance at 275 nm in 6 M guanidinium chloride, using an extinction coefficient of 1420 cm^{-1} M^{-1} (there is one tyrosine per monomer) (Gill & von Hippel, 1989). Mean residue ellipticities were calculated using the equation

$$[\theta] = \theta_{obsd}/10lcn \quad (2)$$

where θ_{obsd} is the ellipticity measured in millidegrees, l is the length of the cell in centimeters, c is the concentration in moles per liter, and n is the number of residues in the protein. The mean residue ellipticity as a function of concentration was measured in two experiments using a dual-syringe titrator (AVIV Associates, Lakewood, NJ) by titrating buffer into a concentrated solution of peptide.

Thermodynamic Analysis. Circular dichroism data for thermal denaturations were collected at 222 nm, averaging 99 s per data point. Buffer conditions were 20 mM MES or 20 mM MOPS (pH range of 5.5−7.3). Data points were taken at 2 °C intervals while equilibrating for at least 3.5 min between points (longer incubation times produced no differences in the data). The experiments were performed at a protein concentration (75 μM) reasonably near the equilibrium constant for folding in this pH range, allowing both cold and heat denaturation to be observed. The presence of cold denaturation provides for a more accurate estimate of ΔC_p using the Gibbs−Helmholtz equation (Chen & Schellman, 1989), although this method assumes ΔC_p to be constant with respect to temperature. The data were fit to the Gibbs−Helmholtz equation using the program Igor Pro (WaveMetrics, Inc.) with the algorithm described in Raleigh et al. (1995). To confirm the validity of using the Gibbs−Helmholtz equation, ΔC_p was determined by calculating the slope of the plot of the van't Hoff enthalpy (ΔH_m, obtained at various pH values) versus T_m over a significantly smaller temperature range, and the two methods were found to agree within 10%.

Calorimetry. Excess heat capacity versus temperature for the RLP-3 dimer was determined using a model MC-2 differential scanning calorimeter (Microcal, Northampton,

[1] Abbreviations: ANS, 8-anilino-1-naphthalenesulfonic acid; DSC, differential scanning calorimetry; ΔC_p, change in heat capacity upon thermal denaturation; ΔG_d, free energy of denaturation; ΔH^{cal}, calorimetric enthalpy of denaturation; ΔH_m, van't Hoff enthalpy of denaturation; T_m, midpoint of thermal denaturation; $t_{1/2}$, midpoint of temperature of half-completion of the DSC thermal denaturation.

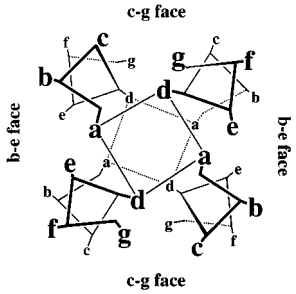

FIGURE 3: Helical wheel representation of the interactions within an antiparallel four-helix bundle. The orthogonal **c**−**g** and **b**−**e** interfaces are indicated.

MA). Protein concentrations varied between 50 and 260 μM (0.35−1.5 mg mL^{-1}). Prior to each measurement, the sample and buffer solutions were degassed under vacuum. Buffer scans were repeated until a steady and reproducible baseline was established. Experiments were performed with a constant heating rate of 90 °C h^{-1}, with the excess heat capacity C_p recorded from 25 to 115 °C. Each sample was subjected to three or four cycles of heating and cooling to judge reversibility. Data were fit using Origin-based software (Microcal).

NMR Spectroscopy. ^{1}H NMR spectra of each dimeric complex were recorded with a Bruker AMX-600 spectrometer at 298 K using presaturation of the water resonance. Sample conditions were 10 mM potassium phosphate and 10% D_2O at pH 6.9. Sample concentrations ranged from 600 μM to 1.2 mM. One-dimensional spectra were recorded with 8K points. Two-dimensional TOCSY experiments (Wuthrich, 1986) were collected with 2K points in the F_2 dimension and 512 increments in F_1. TOCSY mixing times were 61 and 102 ms. The raw data were transformed and phased using Felix 1.1 software (Hare). Hydrogen−deuterium exchange was measured by lyophilizing a 750 μM aqueous, buffered solution of protein and resuspending in D_2O. One-dimensional ^{1}H NMR spectra were recorded as a function of time after the addition of D_2O.

RESULTS

Topological Considerations. In the design of the RLP family of antiparallel dimers, an attempt was made to include successful aspects of previously designed α-helical bundles (Raleigh et al., 1995; Betz & DeGrado, 1996). Namely, the protein core should be able to pack in a geometrically complementary fashion, and the interfacial positions should be composed of an appropriate balance of hydrophobic and hydrophilic residues. Additionally, the sequence should direct the protein to assemble into the desired tertiary structure.

In the preparation for designing a dimeric four-helix bundle, it is important to consider possible alternate folding topologies. Figure 3 illustrates a cross section of an antiparallel four-helix bundle. The fully buried side chains emanate from positions **a** and **d**, and they segregate into layers containing two **a** and two **d** residues. The residues at **c** and **g** as well as those at **b** and **e** occupy more interfacial positions and are important for screening the interior hydrophobic residues (at **a** and **d**) from solvent. There are four possible topologies available to such a four-helix bundle. The loops can lie on the same (syn, Figure 4*b*,*c*) or opposite (anti, Figure 4*d*,*e*) sides of the bundle in either a left-turning or right-turning orientation. It is instructive to consider the features that stabilize the naturally occurring dimeric four-helix bundle ROP (which forms an anti-left-turning topology) in these four topologies. Regan and co-workers have shown that the interior **a** and **d** residues can be replaced with a variety of aliphatic side chains and still maintain a stable, native-like fold (Munson et al., 1994, 1996). The alternation of large and small hydrophobes within the sequence is important for specifying the antiparallel nature of the bundle. If helices with this alternating pattern were to pack in a parallel orientation, the small Ala residues would segregate into individual layers, leaving large holes in the protein interior. Thus, this pattern of large and small residues helps specify an antiparallel structure, although it cannot explain the adoption of a single topology; the core packs equivalently in each of the four antiparallel topologies, indicating that the packing of the **a** and **d** residues cannot account for the observed preference. Further, the loops in ROP can be replaced with a variety of different sequences (Vlassi et al., 1994; Predki et al., 1996), indicating that the sequences of the interhelical turns are also not the primary topological determinant. Therefore, the packing of the interfacial residues must determine the topology of the bundle.

The interfacial residues consist of positions **b** and **e**, which associate along two of the helix−helix interfaces, and **c** and **g**, which associate along the other interfaces (Figure 3). These residues must contribute to specificity and stability in several different ways. First, while they must shield the hydrophobes at the **a** and **d** positions from solvent, they must not be too hydrophobic or poor solubility and fluctuating structures will result (Betz et al., 1993). Second, they should undergo favorable interhelical hydrogen-bonding and electrostatic interactions in the desired topology (Betz & DeGrado, 1996). Finally, they should destabilize potential alternative folds. The manner in which this can be accomplished can be appreciated by considering Figure 4. In an ROP-like topology (anti, left-turning; Figure 4*d*), residues at **c** and **g** interact along the intermolecular interhelical interface between helix 1 and its symmetry-related partner 1′. Similar **c**−**g** interactions occur between helices 2 and 2′. The residues at positions **b** and **e** interact at the intramolecular interface between helices 1 and 2. By contrast, in the other anti orientation (right-turning, Figure 4*e*), the **b** and **e** residues form the intermolecular interface (the 1−1′ and 2−2′ interfaces), and residues from **c** and **g** now form the intramolecular interface between helices 1 and 2 (and 1′ and 2′).

One can manipulate such differences between each of the four geometries in designing a sequence to stabilize a unique topology. Along these lines, it is significant to note that the two syn topologies would be more difficult to engineer than either anti orientation. In the two syn orientations (Figure 4*b*,*c*), both the **c**−**g** and **b**−**e** helix−helix interfaces are identical, and the only difference between the two topologies is whether the interfaces are formed by inter- or intramolecular associations. Thus, for a syn homodimer, only the potential conformational preferences of the interhelical loops could be used to favor one of the two topologies. It is possible that the failure to explicitly consider and stabilize a single topology contributes to the dynamic behavior of previously designed α-helical bundles (Handel et al., 1993; Betz et al., 1993).

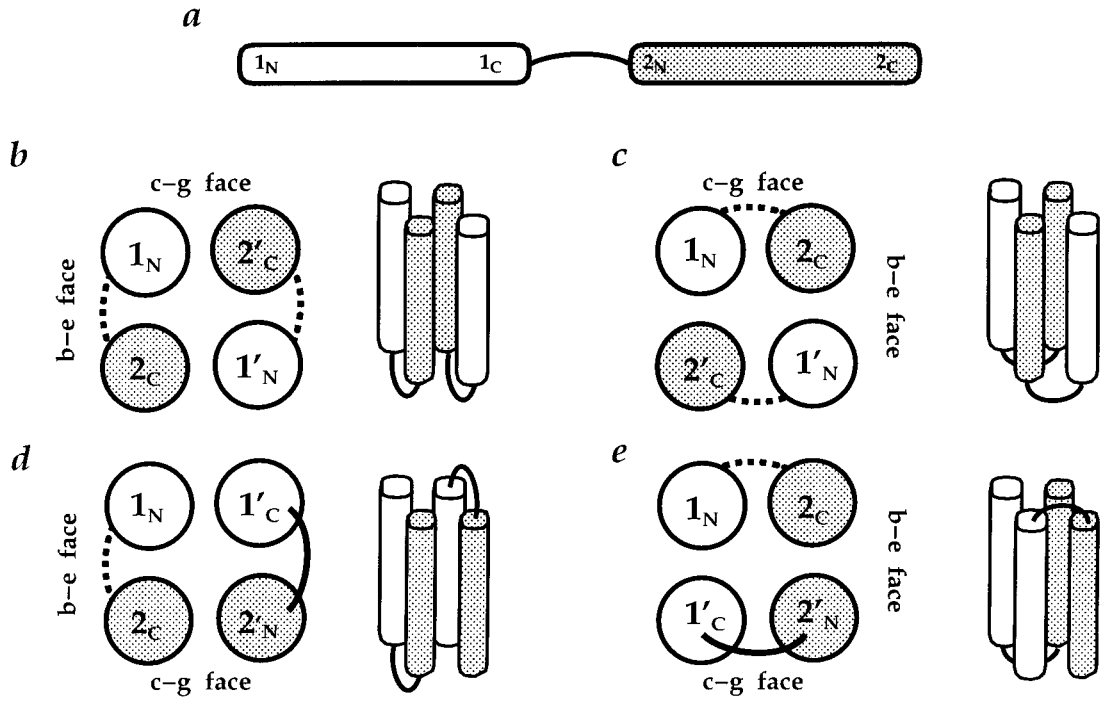

FIGURE 4: Schematic of the loop arrangements and interfacial packing within possible antiparallel (up−down−up−down) four-helix bundle topologies: (*a*) cartoon of a helix−turn−helix peptide indicating the N and C termini of each helix (*b*) syn left-turning, (*c*) syn right-turning, (*d*) anti left-turning, and (*e*) anti right-turning. Notice that the interfacial interactions between the syn topologies are identical, while those of the anti topologies are different from each other as well as from the syn topologies.

In addition to these antiparallel four-helix bundles, several other four-helix bundle orientations have also been classified (Harris et al., 1994), and the designed sequence should also destabilize these folds. Chief among these are the up−up−down−down topologies that are observed in many of the short chain α-helical cytokines (Wlodawer et al., 1993; Rozwarski et al., 1994; Mott & Campbell, 1995). In all up−up−down−down topologies, two of the four interfaces are composed of an antiparallel pair of helices, and the other interfaces are composed of parallel pairs. In these arrangements of helices, half of the **b**−**e** and **c**−**g** interactions found in the up−down−up−down are retained, while the other interfaces are dominated by the **e**−**g** and **b**−**c** interactions seen in parallel four-stranded coiled coils (Harbury et al., 1993; Betz & DeGrado, 1996). It is essential to design the sequence in such a manner that these potential parallel pairings of the helices are destabilized.

Sequence Design. The first consideration in the design of the RLP peptides was the length of the helices. In previous work on single helices that were designed to associate into tetramers (Betz & DeGrado, 1996), a length of three heptads (plus helix caps) proved to be successful. The hydrophobic core of the most successful of these designs included geometric complementarity packing of the interior side chains using small and large hydrophobes at the **a** and **d** positions, respectively, and here, the same general motif is exploited. The large hydrophobe is consistently Leu, while the small hydrophobe is either Ala or Val. The inclusion of Val residues at some or all of the **a** positions of the second heptad was used to increase the amount of hydrophobic surface area buried while simultaneously disfavoring the syn topologies. For example, in RLP-3, the desired (anti) topology contains heptad layers that contain two Leu's, one Ala, and one Val (L_2AV) in the core and should have fairly uniform packing throughout the protein. The same sequence in the syn orientation would form three L_2A_2 and three L_2V_2 layers. Modeling indicates that the L_2V_2 layers are well-packed, but they would hold the helices too far apart for the intervening L_2A_2 layers to pack efficiently. The RLP-1 and RLP-2 sequences result in similar segregation, albeit to a lesser extent.

Interfacial positions were chosen to favor the desired left-turning anti topology using a combination of hydrophobic, hydrogen-bonding, and electrostatic interactions. As a starting point, the remaining residues were initially taken to be the same as in the previously designed peptide α_1B (Ho & DeGrado, 1987) (Figure 2), but numerous changes were made to encourage a unique topology. Next to the **a** and **d** positions, residues **e** and **g** are the most buried in the antiparallel four-helix bundle fold, and previous studies have shown that the hydrophobic residues at these positions can give rise to highly stable structures (Regan & DeGrado, 1988; Chin et al., 1992; Lutgring et al., 1994). However, too many hydrophobes at these positions can cause the bundles to have excessively dynamic behavior (Handel et al., 1993; Betz et al., 1996), so hydrophobic residues were placed sparingly at these positions in the RLP series. Most of the **e** and **g** positions consisted of amphiphilic residues (Glu, Arg, Tyr, and Lys) that could form electrostatic or hydrogen-bonding interactions with residues at **b** and **c**, respectively, while simultaneously shielding the apolar residues at **a** and **d** with the apolar portions of their side chains. With these considerations in mind, the residues at **b**, **c**, **e**, and **g** were introduced to favor a left-turning topology by (i) favorable interactions at the **c**−**g** interfaces between helices 1 and 1′, and 2 and 2′, (ii) favorable interactions at the **b**−**e** interface between helices 1 and 2 (and 1′ and 2′), and (iii) destabilizing interactions between alternative pairings that could possibly form [e.g. the **b**−**e** interface between helices 1 and 1′, and 2 and 2′ should be destabilizing as should the **c**−**g**

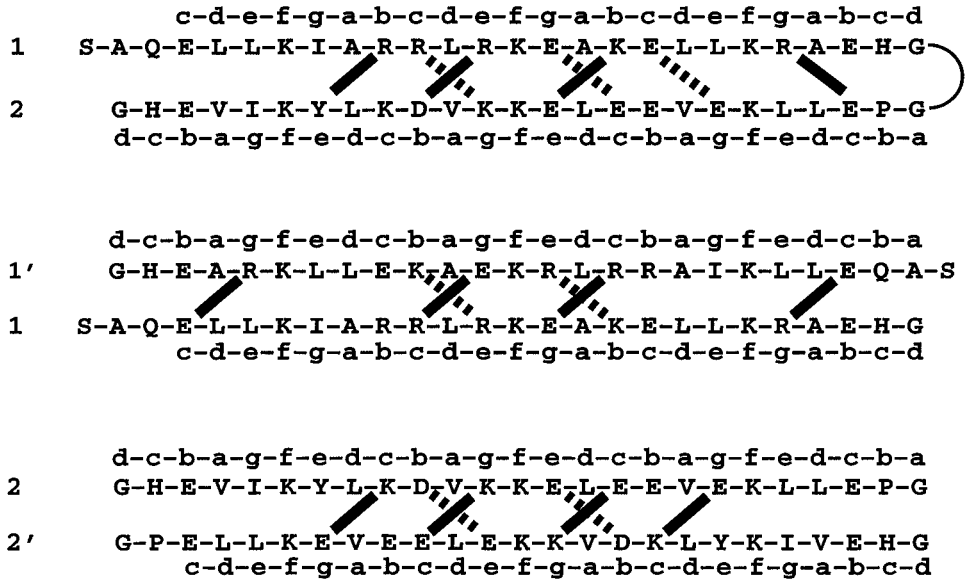

FIGURE 5: Two-dimensional representation of the interfacial interactions of the two anti orientations of RLP-3. The sequence of each helix is shown with the heptad repeat above or below it. The interactions in the anti left-turning topology (i.e. the desired topology) are shown as /. The interactions in the anti right-turning topology are shown as \. Potentially stabilizing interactions are shown as solid lines, and destabilizing interactions are shown as dashed lines. To stabilize the desired topology, favorable interactions should be included at the **c**−**g** interfaces between helices 1 and 1′ and 2 and 2′ and the **b**−**e** interfaces between helices 1 and 2 and 1′ and 2′ (the 1′−2′ interface is not shown because it is identical to the 1−2 interface). To disfavor the anti right-turning topology, the **b**−**e** interfaces between helices 1 and 1′, and 2 and 2′ should be destabilizing as should the **c**−**g** interfaces between helices 1 and 2 and 1′ and 2′. Note the desired topology contains only stabilizing interactions, while the alternate topology consists of mostly destabilizing ones. The favorable interaction between Arg and Tyr (at the 1−2 interface) is a result of complementary hydrophobic interactions and a potential hydrogen bond between Tyr-Oη and Arg-Nϵ.

interface between helices 1 and 2 (and 1′ and 2′)]. Similarly, all possible parallel pairings should be destabilized. Figure 5 provides a two-dimensional schematic of the interfacial interactions that were designed to simultaneously favor the left-turning anti bundle while destabilizing the right-turning topology.

The helix N_{cap} and C_{cap} sequences were chosen by evaluating possible sequences derived from data base searches (Richardson & Richardson, 1988; Harper & Rose, 1993; Seale et al., 1994) and model peptide studies (Zhou et al., 1994a,b; Doig & Baldwin, 1995) within the context of the RLP heptad sequence. The turn sequence was deduced by examining the results of a three-dimensional structure search of a protein structure data base for loops that connected two full turns of α-helix on each side that were of three to six residues in length. The nature and importance of turns with regard to the structure and stability of four-helix bundles have been of considerable interest (Chou et al., 1992; Brunet et al., 1993; Vlassi et al., 1994; Predki & Regan, 1995; Predki et al., 1996), and although the designed turn is not derived from a naturally occurring protein, it is of typical composition and length; modeling suggests that the turn should be able to adopt a low-energy structure of the $\gamma\alpha_L\beta_P\beta_P$ class (Efimov, 1991) that is consistent with the desired structure.

Sedimentation Equilibrium. The association of the RLP peptides was examined using sedimentation equilibrium centrifugation with initial loading concentrations ranging from 25 to 750 μM. All the members of the RLP series behave very similarly and form dimers in a concentration-dependent manner. They sediment as single homogeneous species with molecular masses consistent with dimer formation at loading concentrations of 100 μM or greater (Figure 6). At lower loading concentrations, more monomer is

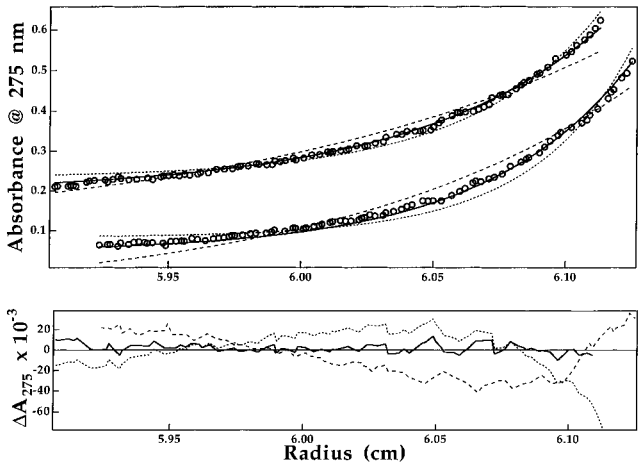

FIGURE 6: Sedimentation equilibrium. (Top) Representative analytical ultracentrifugation runs for RLP-3 (initial loading concentration of 400 μM in 25 mM MOPS and 100 mM NaCl at pH 6.9 and 298 K) showing absorbance at 275 nm versus radius for raw data (circles) and different theoretical fits: (solid line), fit to dimer, (dashed line) fit to monomer, and (dotted line) fit to trimer. (Bottom) Residuals of the lower of the two data scans in the upper panel. The residuals for dimer are random and centered around zero, while those of the other association states exhibit a radial-dependent difference.

present, allowing estimates of the dissociation constants for dimerization: K_{diss} of 1.2 μM (±0.8 μM) for RLP-1, corresponding to a free energy of assembly (ΔG) of −8.1 kcal mol^{-1} (± 0.2 kcal mol^{-1}), a K_{diss} of 2.0 μM for RLP-2 ($\Delta G = -7.8$ kcal mol^{-1}), and a K_{diss} of 0.3 μM for RLP-3 ($\Delta G = -9.0$ kcal mol^{-1}) [the value of K_{diss} for wild-type ROP is estimated to be between 0.5 and 1.0 μM (Stief et al., 1995; Munson et al., 1996)]. These values are in reasonable agreement with the free energies derived from

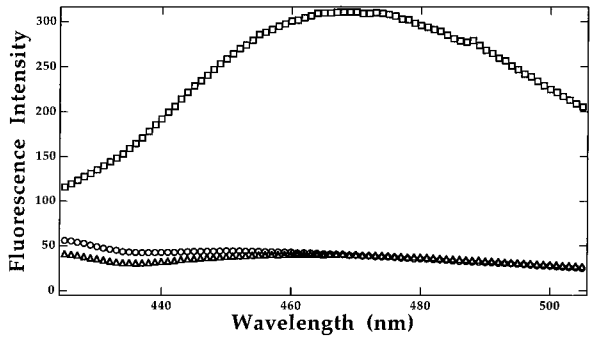

FIGURE 7: Emission fluorescence spectrum of 2 μM 8-anilino-1-naphthalenesulfonic acid (ANS) in the presence and absence of peptide complexes. Buffer conditions were 20 mM HEPES at pH 6.9: ANS alone (circles), ANS with 800 μM RLP-3 (triangles), and ANS with 750 μM α_2B (squares) (Ho & DeGrado, 1987) which exhibits molten globule-like behavior. The excitation wavelength is 370 nm.

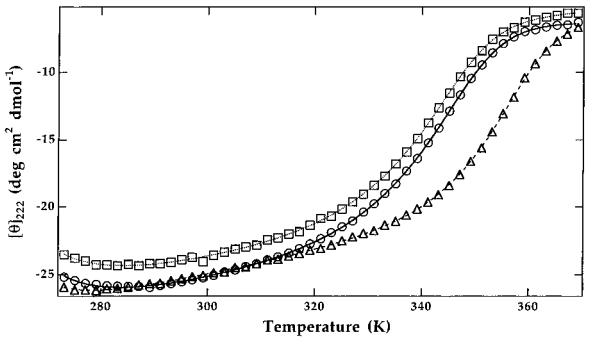

FIGURE 8: Mean residue ellipticity at 222 nm versus absolute temperature and calculated fits to the Gibbs−Helmholtz equation for the RLP series of peptides. The peptides were at 75 μM in 20 mM MOPS at pH 6.97: RLP-1 (circles, solid line), RLP-2 (squares, dotted line), and RLP-3 (triangles, dashed line).

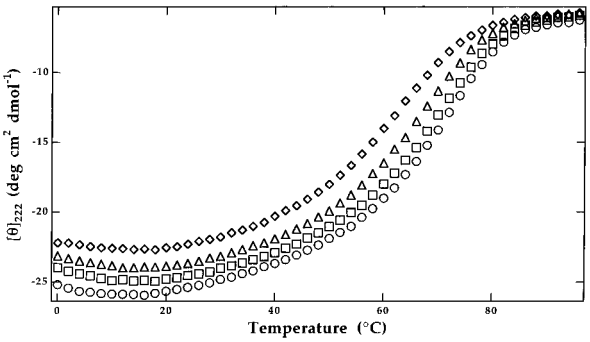

FIGURE 9: Mean residue ellipticity at 222 nm versus temperature for RLP-1 as a function of pH. The sample concentration was 75 μM: pH 6.97 (circles), pH 6.39 (squares), pH 5.97 (triangles), and pH 5.61 (diamonds).

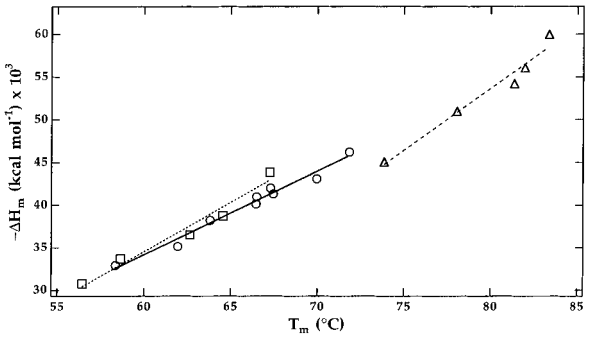

FIGURE 10: $-\Delta H_m$ versus T_m for the RLP series of peptides. The slope of the line is equal to ΔC_p: RLP-1 (circles, solid line), ΔC_p = 1020 cal mol^{-1} K^{-1} (10.0 cal mol^{-1} K^{-1} res^{-1}); RLP-2 (squares, dotted line), ΔC_p = 1150 cal mol^{-1} K^{-1} (11.4 cal mol^{-1} K^{-1} res^{-1}); and RLP-3 (triangles, dashed line), ΔC_p = 1290 cal mol^{-1} K^{-1} (12.6 cal mol^{-1} K^{-1} res^{-1}).

thermal denaturation (see below). Even at the lowest experimentally accessible concentration, however, approximately 85% of the peptide was still associated. This relatively small amount of monomer peptide present causes considerable uncertainty in the value of K_{diss} obtained by this technique. Thus, the values obtained from thermal unfolding should be considered more accurate. At pH 4.0, as assessed by both sedimentation and circular dichroism, the RLP peptides are predominantly monomeric, presumably because of the protonation of several acidic residues which affects the stability of the interfacial packing scheme.

Hydrophobic Dye Binding. The binding of the hydrophobic dye ANS has been used often as a probe for molten globular behavior in proteins (Semisotnov et al., 1991; Ptitsyn, 1992). The molecule has a very low fluorescence intensity in water but fluoresces strongly in the 440−480 nm range when in an apolar environment. The increase in fluorescence of ANS while it is bound to a protein in the molten globule state is presumed to occur because the dye can access the hydrophobic core. Earlier generations of designed proteins have exhibited significant ANS binding (DeGrado et al., 1991; Betz et al., 1993). Each of the RLP series of peptides was titrated into a constant amount of ANS (2 μM) and displayed no discernible increase in fluorescence with peptide concentrations up to 1 mM (Figure 7). In this respect, each member of this family behaves as a native-like protein.

Thermal Denaturation. The-far UV CD spectra of the RLP series were measured as a function of pH and temperature at a concentration of 75 μM. The dimers exhibited thermal denaturation curves like those of native proteins. Figure 8 displays the thermal denaturation curves and theoretical fits of the three RLP peptides at pH 7. Each peptide monomer contains two His residues with pK_a's expected to be in the range 6−7. The pH was varied (Figure 9) to perturb the equilibrium constant for unfolding such that variations in the enthalpy of unfolding could be readily assessed to permit the determination of the change in heat capacity upon denaturation (ΔC_p) as described previously (Stief et al., 1993; Raleigh et al., 1995; Betz & DeGrado, 1996). The van't Hoff enthalpies (ΔH_m) were determined from a plot of R ln K versus 1/T evaluated near the midpoint of the transition. ΔH_m for each peptide varies linearly with respect to T_m and yields ΔC_p from the slope of the line (Figure 10).

Additionally, the entire thermal unfolding curve can be evaluated to determine both ΔH_m and ΔC_p using the Gibbs−Helmholtz equation:

$$\Delta G_d = \Delta H_m(1 - T/T_m) - \Delta C_p[T_m - T + T[\ln(T/T_m)]] \quad (3)$$

Typically, it is difficult to accurately determine ΔC_p from the Gibbs−Helmholtz equation because the quality of the fit is generally insensitive to changes in ΔC_p. However, the presence of cold denaturation observed for the RLP peptides (Figures 8 and 9) permits a more accurate determination of ΔC_p (Chen & Schellman, 1989). Calculations of ΔC_p from van't Hoff and Gibbs−Helmholtz methods have distinct

Table 1: Thermodynamic Parameters of the RLP Series of Proteins

protein	method	T_m (°C)	ΔH_m (kcal mol^{-1})[a]	ΔG_d, 300 K (kcal mol^{-1})[a]	ΔC_p (cal mol^{-1} K^{-1})[a]	ΔC_p per residue (cal mol^{-1} K^{-1})[a]
RLP-1	vH[b]	69.9	−45.6	−8.7	1020	10.0
	GH[c]	71.9	−46.2	−8.7	1030	10.1
RLP-2	vH	66.3	−45.8	−8.4	1150	11.3
	GH	67.2	−44.9	−8.5	1090	10.7
RLP-3	vH	81.6	−62.0	−9.7	1290	12.6
	GH	83.4	−62.9	−9.9	1210	11.9

[a] Based on the total amount of monomer. [b] van't Hoff analysis; the T_m and ΔH_m listed are from a from a single denaturation at pH 6.97. ΔC_p was calculated from linear regression of the van't Hoff enthalpies (Figure 10) versus the corresponding thermal denaturation midpoint temperatures (T_m) over a range of pHs. ΔG_d was calculated by inserting these values of T_m, ΔH_m, and ΔC_p into eq 3. [c] Gibbs−Helmholtz analysis of an entire single thermal denaturation at pH 6.97 was used to determine T_m, ΔH_m, and ΔC_p simultaneously. ΔG_d was calculated by inserting these values into eq 3.

Table 2: Comparison of Enthalpy Changes for RLP-3 Determined by DSC and CD[a]

method	protein concentration (μM)	ΔH_m (kcal mol^{-1})	ΔH^{cal} (kcal mol^{-1})	T_m (°C)	$\Delta H^{cal}/\Delta H_m$	ΔH_m(DSC)/ΔH_m(CD)
DSC	58	−47.6	−22.9	78.6	0.48	1.0
CD	58	−46.8		77.7	0.49	
DSC	260	−55.0	−25.0	83.5	0.45	1.1
CD	260	−50.3		82.4	0.50	

[a] Sample buffer of 20 mM NaP_i at pH 6.25.

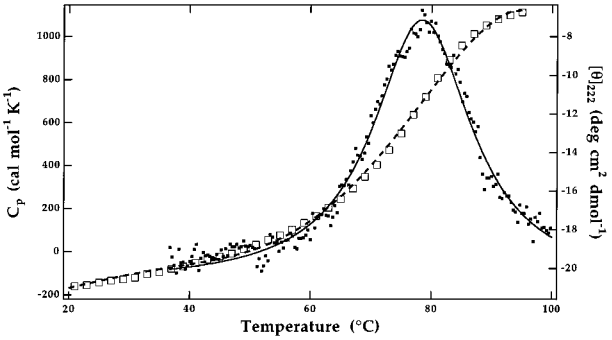

FIGURE 11: Comparison of thermal denaturation of RLP-3 monitored by DSC (dots, left axis) and mean residue ellipticity monitored at 222 nm (open squares, right axis). The conditions for both experiments were 58 μM peptide and 20 mM NaP_i at pH 6.0. Theoretical fits are shown as solid and dashed lines, respectively.

advantages and disadvantages. The Gibbs−Helmholtz method does not require perturbing the solution conditions but assumes that ΔC_p is constant over a large temperature range (>60 °C, in this case). By contrast, the use of van't Hoff enthalpies obtained at various solution conditions requires a less extreme temperature range but introduces possible deviation associated with different protonation states of the protein. Therefore, we were pleased to find that both methods yielded very similar values for ΔH_m, ΔC_p, and T_m (Table 1).

Finally, differential scanning calorimetry (DSC) was used to further confirm that the CD experiments were an accurate measure of the enthalpy change and two-state behavior of the RLP dimers. The data are summarized in Table 2, and Figure 11 displays the superposition of the CD and DSC denaturations of 58 μM RLP-3 in 20 mM NaP_i at pH 6.0. The T_m for the CD-monitored denaturation (77.7 °C) is in excellent agreement with the $t_{1/2}$ determined by DSC (78.6 °C). Further, the ratio of the values of ΔH_m for the two methods is 1.02. The cooperativity of the transition as determined by the ratio of ΔH_m to calorimetric enthalpies (ΔH^{cal}) is 2, indicating that the species being denatured is dimeric (Privalov & Potekhin, 1986). The DSC experiments were repeated on a more concentrated sample (260 μM), and the spectroscopic and DSC methods of determining ΔH_m were again in excellent agreement (Table 2). As expected for a monomer−dimer equilibrium, the value of T_m was increased at the higher concentration. The value of ΔH_m was also increased, although the calorimetrically determined value was somewhat uncertain due to the lack of resolution of the high-temperature baseline.

The RLP proteins are similar in stability to other dimeric four-helix bundles. Munson et al. (1996) determined the ΔG_d of wild-type ROP (wt-ROP) to be −7.7 kcal mol^{-1} by chemical denaturation, and the dimeric helix−loop−helix protein E47 has a free energy of association of −7.3 kcal mol^{-1} at 300 K (Fairman et al., 1993). RLP-3 has the highest T_m of the series (Figure 8) and is the most highly stabilized enthalpically and in terms of free energy.

One of the hallmarks of small, soluble proteins is the small range of ΔC_p exhibited upon thermal denaturation (Privalov & Gill, 1988; Makhatadze & Privalov, 1990; Pace et al., 1989). The values are generally in the range of 10−15 cal mol^{-1} K^{-1} on a per residue basis. Each of the RLP dimers has a value of ΔC_p in the range expected for natural proteins of this size, and these represent the largest such values determined for designed proteins to date. There is an incremental increase in ΔC_p with increasing hydrophobic content, consistent with the hypothesis that ΔC_p is proportional to the change in apolar surface area upon denaturation (Eisenberg et al., 1989; Spolar et al., 1992; Xie & Freire, 1994).

Nuclear Magnetic Resonance. The NMR characteristics of the RLP dimer complexes were also examined. The dimers are soluble (up to ∼1 mM), and yield high-quality spectra. In the one-dimensional spectrum of RLP-1 (data not shown), the amide protons are dispersed and the aliphatic methyl resonances show a few singularly resolved peaks. In the fingerprint region of the TOCSY spectrum (Figure 12), about 80% of the expected NH−C_α cross-peaks are resolved. Overall, the spectra are quite good considering the α-helical

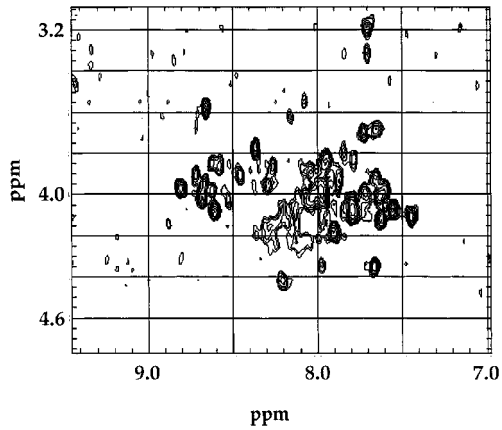

FIGURE 12: Two-dimensional TOCSY spectrum of RLP-1 at 298 K in 10 mM potassium phosphate and 10% D_2O at pH 6.9. The mixing time was 61 ms. Spectra were recorded on a Bruker AMX-600 MHz spectrometer with weak irradiation to saturate residual HOD. Chemical shifts are given in parts per million from DSS.

nature of the complex and the lack of aromatic residues (one tyrosine per 51-residue monomer).

One criterion for native-like behavior of designed proteins that we have suggested is amide proton−deuterium exchange rates on the order of those found for naturally occurring proteins (Betz et al., 1993). At pH 6 and 15 °C, the dissociation constants for the RLP dimers are in the low micromolar range. Even at NMR concentrations (750 μM to 1 mM), a significant fraction (2−5%) of the population is unstructured monomer which exchanges with a half-time of approximately 5 min at this pH. Rapid exchange between monomer and dimer overwhelms any protection that would occur in the structured dimer. In principle, the proton−deuterium exchange could be slowed by lowering the pH, but this also decreases the stability of the dimer and concomitantly further increases the concentration of the unstructured monomer.

DISCUSSION

The RLP series of proteins represents another step in the evolution of designed proteins toward native-like behavior (Raleigh et al., 1995; Nautiyal et al., 1995; Olofsson et al., 1995; Betz & DeGrado, 1996). These proteins behave in all biophysical criteria examined as native proteins: single association state, lack of ANS binding, native-like thermodynamic behavior, and spectroscopic properties.

The systematic study of the naturally occurring protein ROP has provided a solid foundation for the *de novo* design of proteins with such a topology. In previous studies, all or part of the **a** and **d** residues were idealized, allowing a better understanding of their roles in protein folding and stability (Munson et al., 1994, 1996). In this work, we have focused primarily on the interfacial interactions at **b**, **c**, **e**, and **g** and the role they play in specifically stabilizing the helical bundle. We then tested these insights through the design and characterization of idealized antiparallel four-helix bundles.

The design of these molecules followed a strategy that attempts to use a combination of geometric interior packing and interfacial interactions to achieve a singular topology while disfavoring competing folds (Betz & DeGrado, 1996). The idea of uniqueness being coupled to the distinction between a protein's native state and its next nearest energy state has been postulated recently in theoretical models (Sali et al., 1994a,b; Dill et al., 1995). Similarly, the presence of several nearly isoenergetic states has been considered as a major cause of molten globule-like behavior of some designed proteins (Raleigh et al., 1995; Betz et al., 1996).

The three proteins of the RLP series all behave as native-like proteins with RLP-3 being the most thermodynamically stable. In its design, RLP-3 possesses six layers of hydrophobic core that consist of two Leu's, one Ala, and one Val residue (L_2AV). The other proteins contain two (RLP-2) or four (RLP-1) L_2A_2 layers with the remainder being L_2AV. The regular packing throughout its entire hydrophobic core may contribute to the significant enthalpic stabilization of the RLP-3 complex.

The effects of the packing of the interior side chains on the stability of the naturally occurring ROP protein have been explored (Munson et al., 1994, 1996). In these studies, the residues at the **a** and **d** positions were systematically varied while maintaining the topology-determining **b**, **c**, **e**, and **g** residues as they are in ROP. The individual repacked core variants displayed a wide variety of characteristics from unfolded, to fully native-like, to aggregated tetramers. The most native-like molecules had cores composed of Ala and Leu. By comparison, the global stabilities (at 300 K) of the RLP series are very similar to those determined for wt-ROP and the most well-behaved repacked core variants (Munson et al., 1996).

The RLP series of *de novo* designed dimers demonstrates that it is now possible to engineer sequences that fold into stable native-like structures that possess the characteristics of naturally occurring proteins. By regulation of the hydrophobic driving force through the manipulation of the residues at **a** and **d** positions and the specificity through the manipulation of the interfacial **b**, **c**, **e**, and **g** positions, native-like protein structures can now be produced. Several series of experiments are now underway to determine how closely the dimers' structures match those of their design, including NMR, crystallography, and the generation of a recombinant, single-chain version that should maintain the same interhelical interactions. The strategies outlined here should serve as a blueprint for the generation of functional molecules as well as those with more complex tertiary structures.

ACKNOWLEDGMENT

We thank S. Walter Englander, Jim Bryson, and Blake Hill for helpful discussions, Gregg Dieckmann for advice on modeling, Leland Mayne and Scott Walsh for NMR assistance, Nancy Thornton for DSC assistance, and Francesc Rabanal and Tobin Sosnick for critical reading of the manuscript.

REFERENCES

Banner, D. W., Kokkinidis, M., & Tsernoglou, D. (1987) *J. Mol. Biol. 196*, 657−675.

Betz, S. F., & DeGrado, W. F. (1996) *Biochemistry 35*, 6955−6962.

Betz, S. F., Raleigh, D. P., & DeGrado, W. F. (1993) *Curr. Opin. Struct. Biol. 3*, 601−610.

Betz, S. F., Raleigh, D. P., DeGrado, W. F., Lovejoy, B., Anderson, D., Ogihara, N., & Eisenberg, D. (1996) *Folding Des. 1*, 57−64.

Brunet, A. P., Huang, E. S., Huffine, M. E., Loeb, J. E., Weltman, R. J., & Hecht, M. H. (1993) *Nature 364*, 355−358.

Chen, B. L., & Schellman, J. A. (1989) *Biochemistry 28*, 685−691.

Chin, T.-M., Berndt, K. D., & Yang, N.-c. C. (1992) *J. Am. Chem. Soc. 114*, 2279−2280.
Choma, C. T., Lear, J. D., Nelson, M. J., Dutton, P. L., Robertson, D. E., & DeGrado, W. F. (1994) *J. Am. Chem. Soc. 116*, 856−865.
Chou, K. C., Maggiora, G. M., & Scheraga, H. A. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 7315−7319.
Cohen, C., & Parry, D. A. D. (1990) *Proteins: Struct., Funct., Genet. 7*, 1−15.
Cohn, E. J., & Edsall, J. T. (1943) *Proteins, Amino Acids, and Peptides as Ions and Dipolar Ions*, pp 370−377, Reinhold Publishing Corp., New York.
DeGrado, W. F., Raleigh, D. P., & Handel, T. (1991) *Curr. Opin. Struct. Biol. 1*, 984−993.
Dill, K. A., Bromberg, S., Yue, K., Kiebig, K. M., Yee, D. P., Thomas, P. D., & Chan, H. S. (1995) *Protein Sci. 4*, 561−602.
Doig, A. J., & Baldwin, R. L. (1995) *Protein Sci. 4*, 1325−1336.
Efimov, A. V. (1991) *Protein Eng. 4*, 245−250.
Eisenberg, D., Wesson, M., & Yamashita, M. (1989) *Chem. Scr. 29A*, 217−221.
Fairman, R., Beran-Steed, R. K., Anthony-Cahill, S. J., Lear, J. D., III, DeGrado, W. F., Benfield, P. A., & Brenner, S. L. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 10429−10433.
Gill, S. C., & von Hippel, P. H. (1989) *Anal. Biochem. 182*, 319−326.
Handel, T. M., Williams, S. A., & DeGrado, W. F. (1993) *Science 261*, 879−885.
Harbury, P. B., Zhang, T., Kim, P. S., & Alber, T. (1993) *Science 262*, 1401−1407.
Harding, S. E., Rowe, A. J., & Horton, J. C. (1992) *Analytical Ultracentrifugation in Biochemistry and Polymer Science*, The Royal Society of Chemistry, Cambridge.
Harper, E. T., & Rose, G. D. (1993) *Biochemistry 32*, 7605−7609.
Harris, N. L., Presnell, S. R., & Cohen, F. E. (1994) *J. Mol. Biol. 236*, 1356−1368.
Ho, S. P., & DeGrado, W. F. (1987) *J. Am. Chem. Soc. 109*, 6751−6758.
Kraulis, P. J. (1991) *J. Appl. Crystallogr. 24*, 946−950.
Lutgring, R. A., & Chmielewski, J. A. (1994) *J. Am. Chem. Soc. 116*, 6451−6452.
Makhatadze, G. I., & Privalov, P. L. (1990) *J. Mol. Biol. 213*, 385−391.
McGregor, M. J., Islam, S. A., & Sternberg, M. J. E. (1987) *J. Mol. Biol. 198*, 295−310.
Mott, H. R., & Campbell, I. D. (1995) *Curr. Opin. Struct. Biol. 5*, 114−121.
Munson, M., O'Brien, R., Sturtevant, J. M., & Regan, L. (1994) *Protein Sci. 3*, 2015−2022.
Munson, M., Balasubramanian, S., Fleming, K. G., Nagi, A. D., O'Brien, R., Sturtevant, J. M., & Regan, L. (1996) *Protein Sci. 5*, 1584−1593.
Nautiyal, S., Woolfson, D. N., King, D. S., & Alber, T. (1995) *Biochemistry 34*, 11645−11651.
Olofsson, S., Johansson, G., & Baltzer, L. (1995) *J. Chem. Soc., Perkin Trans. 2*, 1−10.
Pace, C. N., Shirley, B. A., & Thomson, J. A. (1989) in *Protein Structure: a Practical Approach*, pp 287−310, IRL Press, Oxford.
Predki, P. F., & Regan, L. (1995) *Biochemistry 34*, 9834−9839.
Predki, P. F., Agrawal, V., Brunger, A. T., & Regan, L. (1996) *Nat. Struct. Biol. 3*, 54−58.
Privalov, P. L., & Potekhin, S. A. (1986) *Methods Enzymol. 131*, 4−51.
Privalov, P. L., & Gill, S. J. (1988) *Adv. Protein Chem. 39*, 191−234.
Ptitsyn, O. B. (1992) *Protein Folding*, pp 243−299, Freeman, New York.
Raleigh, D. P., Betz, S. F., & DeGrado, W. F. (1995) *J. Am. Chem. Soc. 117*, 7558−7559.
Regan, L., & DeGrado, W. F. (1988) *Science 241*, 976−978.
Richardson, J. S., & Richardson, D. C. (1988) *Science 240*, 1648−1652.
Rozwarski, D. A., Gronenborn, A. M., Clore, G. M., Bazan, J. F., Bohm, A., Wlodawer, A., Hatada, M., & Karplus, P. A. (1994) *Structure 2*, 159−173.
Sali, A., Shakhnovich, E., & Karplus, M. (1994a) *Nature 369*, 248−251.
Sali, A., Shakhnovich, E., & Karplus, M. (1994b) *J. Mol. Biol. 235*, 1614−1636.
Seale, J. W., Srinivasan, R., & Rose, G. D. (1994) *Protein Sci. 3*, 1741−1745.
Semisotnov, G. V., Rodionova, N. A., Razgulyaev, O. I., Uversky, V. N., Gripas, A. F., & Gilmanshin, R. I. (1991) *Biopolymers 31*, 119−128.
Spolar, R. S., Livingston, J. R., & Record, M. T., Jr. (1992) *Biochemistry 31*, 3947−3955.
Steif, C., Weber, P., Hinz, H.-J., Flossdorf, J., Cesareni, G., & Kokkinidis, M. (1993) *Biochemistry 32*, 3867−3876.
Steif, C., Hinz, H.-J., & Cesareni, G. (1995) *Proteins: Struct., Funct., Genet. 23*, 83−96.
Vlassi, M., Steif, C., Weber, P., Tsernoglou, D., Wilson, K. S., Hinz, H. J., & Kokkinidis, M. (1994) *Nat. Struct. Biol. 1*, 706−716.
Wlodawer, A., Pavlovsky, A., & Gustchina, A. (1993) *Protein Sci. 2*, 1373−1382.
Wuthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley, New York.
Xie, D., & Freire, E. (1994) *Proteins: Struct., Funct., Genet. 19*, 291−301.
Zhou, H. X., Lyu, P., Wemmer, D. E., & Kallenbach, N. R. (1994a) *Proteins: Struct., Funct., Genet. 18*, 1−7.
Zhou, H. X., Lyu, P. C., Wemmer, D. E., & Kallenbach, N. R. (1994b) *J. Am. Chem. Soc. 116*, 1139−1140.

BI961704H