

# Structural Test of the Parameterized-backbone Method for Protein Design

Joseph J. Plecs<sup>1</sup>, Pehr B. Harbury<sup>2</sup>, Peter S. Kim<sup>2</sup> and Tom Alber<sup>3\*</sup>

<sup>1</sup>Department of Physics  
University of California  
Berkeley, CA 94720, USA

<sup>2</sup>Howard Hughes Medical  
Institute, Whitehead Institute  
Department of Biology  
Massachusetts Institute of  
Technology, 9 Cambridge Center  
Cambridge MA 02142, USA

<sup>3</sup>Department of Molecular and  
Cell Biology, University of  
California, Berkeley, CA  
94720-3206, USA

Designing new protein folds requires a method for simultaneously optimizing the conformation of the backbone and the side-chains. One approach to this problem is the use of a parameterized backbone, which allows the systematic exploration of families of structures. We report the crystal structure of RH3, a right-handed, three-helix coiled coil that was designed using a parameterized backbone and detailed modeling of core packing. This crystal structure was determined using another rationally designed feature, a metal-binding site that permitted experimental phasing of the X-ray data. RH3 adopted the intended fold, which has not been observed previously in biological proteins. Unanticipated structural asymmetry in the trimer was a principal source of variation within the RH3 structure. The sequence of RH3 differs from that of a previously characterized right-handed tetramer, RH4, at only one position in each 11 amino acid sequence repeat. This close similarity indicates that the design method is sensitive to the core packing interactions that specify the protein structure. Comparison of the structures of RH3 and RH4 indicates that both steric overlap and cavity formation provide strong driving forces for oligomer specificity.

© 2004 Elsevier Ltd. All rights reserved.

\*Corresponding author

**Keywords:** parameterized-backbone; RH3; protein design; structure

## Introduction

Protein design has developed rapidly in recent years, leading to important achievements such as *de novo* catalytic activity, an endonuclease with modified specificity, and protein-based sensors for small molecules.<sup>1–3</sup> While there have been several examples of redesigning backbone conformations, especially loop regions, most protein design methods work by placing side-chains in the context of a known backbone structure.<sup>4–17</sup> These fixed-backbone strategies employ detailed knowledge of a backbone fold to shrink the conformational space to be considered, and reduce the problem to choosing the most favorable sequence and rotamer set. A fixed backbone provides a powerful

simplification. Side-chains can be varied one at a time or in small groups, breaking down the problem of sequence design or structure prediction into a set of smaller problems that can be handled largely independently.

Even though a great deal of progress has been made using fixed-backbone techniques, treatment of backbone flexibility is essential for the design of new proteins, especially for entirely new folds. Fixed-backbone techniques can be used only in cases where an experimental structure is available, and the backbone is not optimized in response to changes in the packing of core side-chains. As a result, advantageous rearrangements that would accommodate new core sequences are not considered. In several cases, metal-binding proteins have been designed successfully using a backbone-parameterization method.<sup>18–20</sup> A striking example of the importance of backbone flexibility is the recent design of a novel 93 amino acid residue protein fold, in which it was found that optimizing the backbone during the design process was essential, even though the backbone changes were small.<sup>21</sup>

A flexible backbone introduces substantial difficulties for protein design. Side-chain sequences and

---

Present addresses: J. J. Plecs, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Mail Stop 6R2100, Berkeley, CA 94720, USA; P. B. Harbury, Department of Biochemistry, Stanford University, Stanford, CA 94305, USA; P. S. Kim, Merck Research Laboratories, 770 Sunnyside Pike, West Point, PA 19486, USA.

Abbreviations used: Glu,  $\gamma$ -carboxyglutamic acid; a-Ile, allo-isoleucine.

E-mail address of the corresponding author:  
tom@ucxray.berkeley.edu

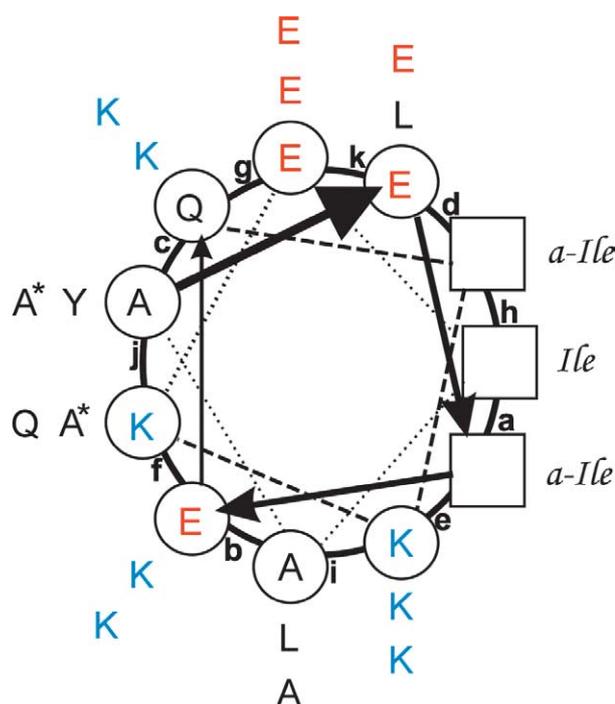
rotamer combinations alone can represent an enormous number of combinations to be considered by a design algorithm. When backbone variations are taken into account, a protein-design algorithm must consider an even larger conformational space. Backbone flexibility requires a method for relieving strain over sizeable regions of the chain, because main-chain atoms are coupled tightly to each other through covalent bonds. Also, a flexible backbone complicates the coupling of side-chains, because side-chain interactions will depend on the backbone conformation.

The problem of systematically managing backbone movements was initially addressed by introducing a mathematical parameterization to restrict the backbone conformation to a family of plausible folds.<sup>22,23</sup> Originally for the coiled-coil family, the backbone structure was defined by a small number of parameters. Backbone movements were represented in a global rather than a local fashion, avoiding the problem of unreasonable local distortions.

Coiled coils were described mathematically by Crick in 1953,<sup>24</sup> providing a natural starting-point for parameterization. We adapted Crick's approach, which uses only seven parameters to describe a regular coiled-coil backbone, to design the RH (right-handed) family of two, three, and four-helical bundles with an unprecedented right-handed superhelical twist.<sup>23</sup> The peptides are 33 amino acid residues long, containing three 11 residue repeats. The dimer, trimer and tetramer sequences are identical except for the residues in the hydrophobic core (Figure 1). The core residues included at least one allo-isoleucine residue (a-Ile), in which the C<sup>β</sup> chirality is opposite that of Ile. Strikingly, the predicted trimer and tetramer sequences are more than 90% identical, differing only in the placement of a single methyl group in each 11 residue repeat. Biological sequences with this level of similarity almost invariably adopt the same fold.

Biochemical characterization of the dimer, trimer, and tetramer sequences by circular dichroism spectroscopy and analytical ultracentrifugation confirmed that they formed helical structures with the intended oligomeric states.<sup>23</sup> Further, the crystal structure of the tetramer confirmed that it forms a parallel bundle of helices with a right-handed, superhelical twist, and that the packing conformation and superhelical parameters match the predicted structure.

To test the parameterized backbone method in greater detail, we introduced a new metal-binding site into the designed RH trimer, RH3, and determined the structure at 2.0 Å resolution. This rationally designed, metal-binding site allowed us to calculate high-quality experimental phases, and the design is likely to be applicable to any solvent-exposed  $\alpha$  helix. The observed and predicted structures of RH3 are very similar, except for unanticipated asymmetry in the trimer core. The agreement of the design with the RH3 crystal structure implies that the parameterized-backbone method is sensitive to even subtle sequence



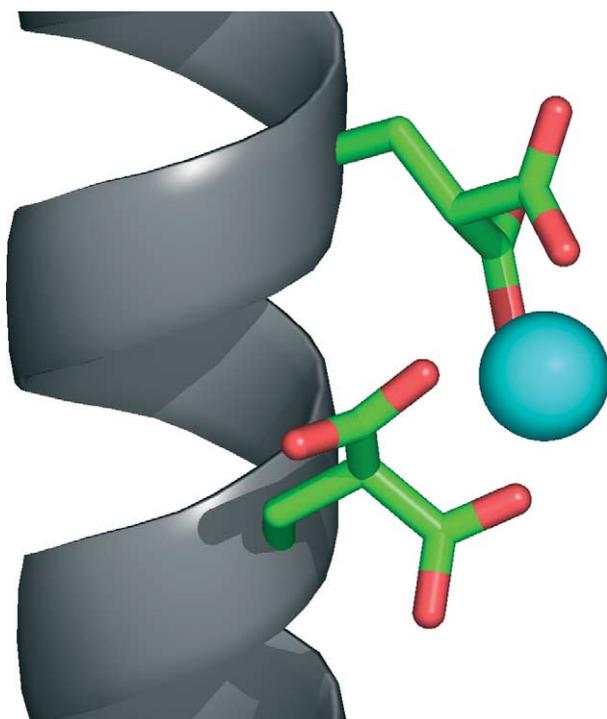
**Figure 1.** The RH3 sequence.<sup>22</sup> The 11-residue sequence repeat of the RH family is depicted in a helical wheel projection. Positions in the 11-residue repeat are labeled a through k. The amino acid residues on the hydrophobic face of the helix are denoted by squares and all others by circles. Allo-isoleucine residues, which have side-chain chirality opposite to that of Ile, are indicated by a-Ile. Blue amino acid residues are positively charged, and red are negatively charged. The two alanine residues of the RH3 sequence marked with asterisks were replaced with Glu residues in one repeat to produce a metal-binding derivative, RH3-Gla.

variations that control protein structure. A comparison between the RH3 and RH4 structures suggests that the overall fold of these coiled coils is very sensitive to both steric restrictions and cavity formation in the core.

## Results and Discussion

### Designed metal-binding site

A metal-binding site (Figure 2) was introduced onto the solvent-accessible surface of RH3 in order to allow experimental phasing of the crystallographic data. This site was designed on the basis of the coordination of calcium ions by  $\gamma$ -carboxyglutamic acid (Gla) residues in thrombin.<sup>25,26</sup> It has been shown that the calcium atoms in these sites can be replaced by heavier atoms, suggesting that a similar Gla motif could be used as a heavy-atom site to allow phasing of the RH3 structure. A model based on the predicted structure of RH3 indicated that Gla residues at positions 19 and 23 of an RH3 monomer are positioned to coordinate a metal ion (Figure 2). Alanine residues 19 and 23 of RH3 were



**Figure 2.** Designed metal-binding site. The Glu residues at positions 19 and 23 of the RH3 backbone form a binding site for divalent metal ions such as nickel and cobalt. A nickel ion is shown as a blue sphere.

replaced by Glu residues in a new peptide for crystallographic data collection. Crystals of this peptide containing bound nickel ions were used to determine the structure at 2.0 Å by multi-wavelength anomalous diffraction (MAD) analysis (Table 1; Figure 2).

**Table 1.** X-ray data collection and refinement statistics for the RH3 structure

Space group	$P3_2$		
	$a=b=25.42 \text{ \AA}$	$c=141.34 \text{ \AA}$	
	$\alpha=\beta=90^\circ$	$\gamma=120^\circ$	
Data sets	$F_p$	$F_{pp}$	$F_{high}$
$\lambda$ (Å)	1.4862	1.4851	1.3931
$R_{merge}^a$	0.084	0.081	0.084
$R_{anom}^b$	0.036	0.076	0.049
$I/\sigma$	20.8	21.4	16.4
Multiplicity	8.1	8.0	7.2
Completeness (%)	86.6	86.6	84.5
$B_{Wilson}$	29.2	34.1	30.5
Figure of merit	0.685 (0.882)		
before (after) dm			
Resolution (Å)	23.5–2.0		
$R$ ( $R_{free}$ ) <sup>c</sup>	23.26% (26.94%)		
RMS deviation from ideality			
Bond lengths (Å)	0.007		
Bond angles (deg.)	1.071		
Unique reflections	5854		
In working data set			
In $R_{free}$ data set	298		

<sup>a</sup>  $R_{merge} = \sum |I_{obs} - \langle I \rangle| / \sum \langle I \rangle$ .

<sup>b</sup>  $R_{anom} = \sum |I^+ - I^-| / \sum (\langle I^+ \rangle + \langle I^- \rangle)$ .

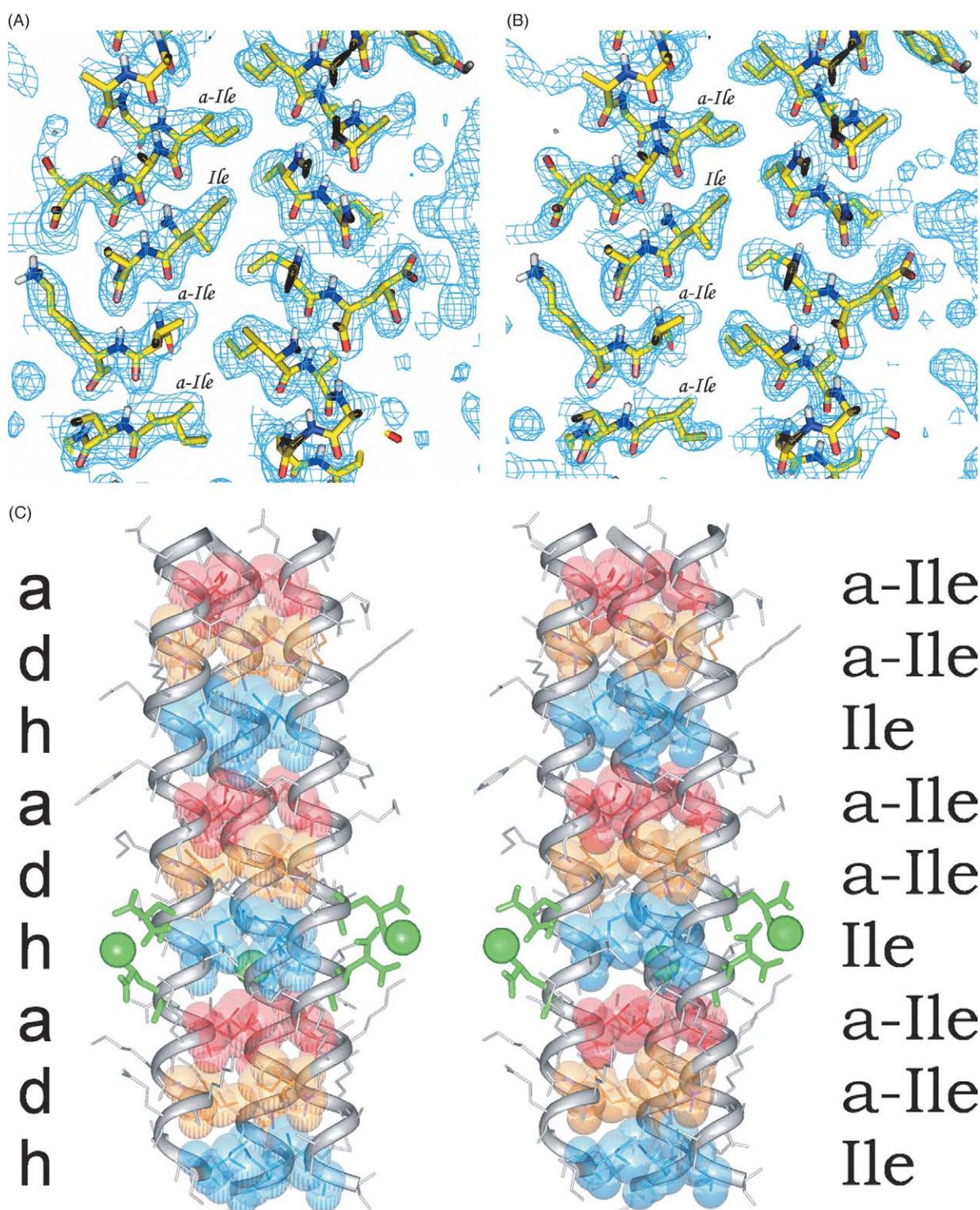
<sup>c</sup>  $R$  factor =  $\sum_{hkl} ||F_{obs}| - |F_{calc}|| / \sum_{hkl} |F_{obs}|$ .

## Structure overview

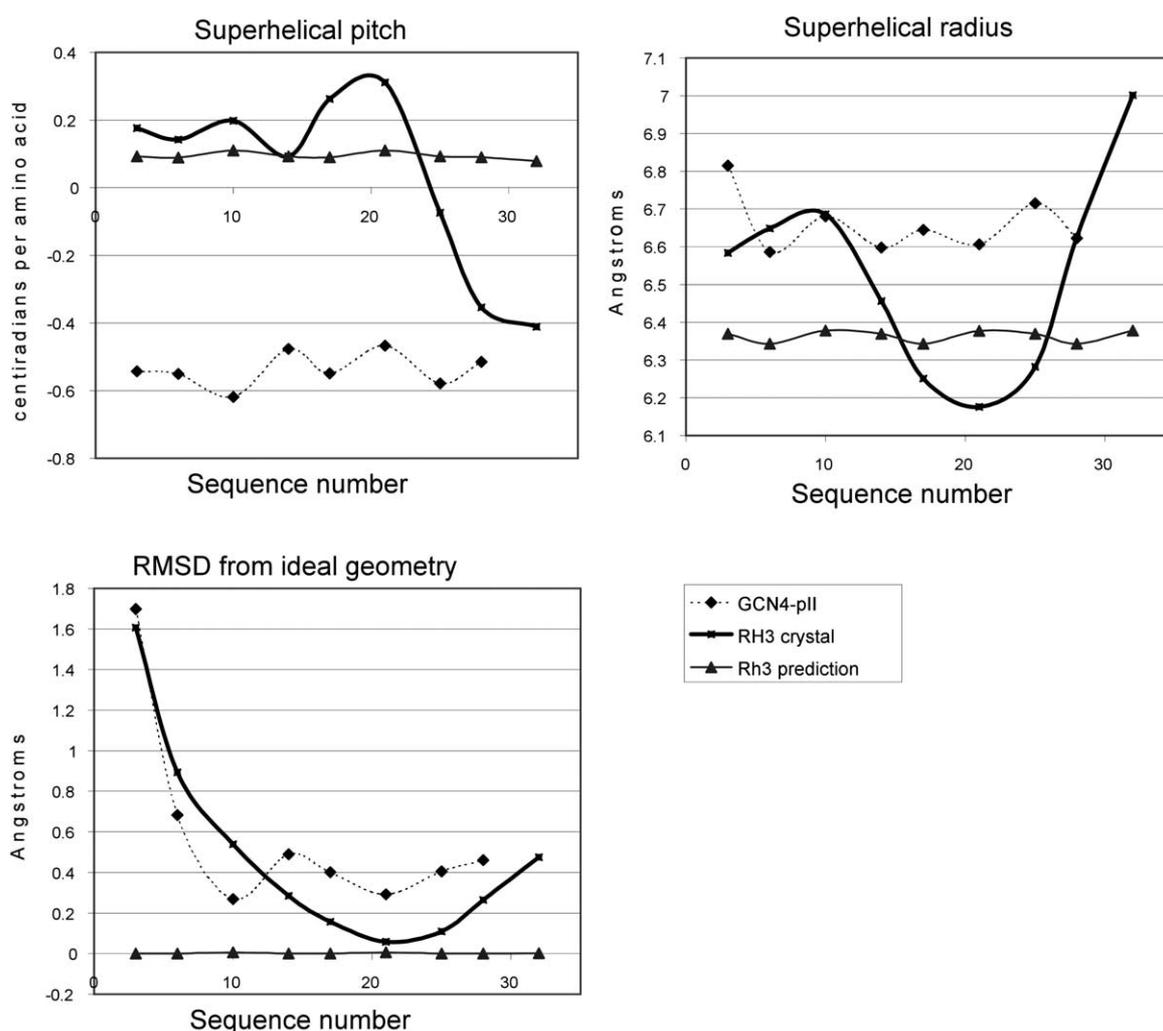
The experimental map was exceptionally clear and resolved most of the structural features of RH3 (Figure 3(A)). The final  $2F_o - F_c$  electron density map (Figure 3(B)) was very similar to the experimental map. RH3 forms a parallel, three-stranded coiled coil with a right-handed superhelical twist (Figure 3(C)). The helical bundle is approximately 50 Å long and 22 Å in diameter, somewhat longer and thinner than a left-handed helical bundle with the same number of residues. The overall superhelical twist is not as pronounced as that of a left-handed coiled coil trimer; the right-handed coiled coils based on an 11 residue repeat are less tightly twisted than left-handed coiled coils based on a seven residue repeat.

The three helices of the trimer (denoted A, B, and C) are not crystallographically equivalent. If the trimer is rotated  $120^\circ$  about the superhelical axis, so that each monomer is superimposed on another, the root-mean-square deviation (rmsd) of the  $C^\alpha$  atoms in the central 11 residue segment is 1.40 Å. Individual comparisons show that while monomers B and C are relatively similar, with a pairwise  $C^\alpha$  rmsd of 0.69 Å, monomer A is different, with a  $C^\alpha$  rmsd of 1.22 Å and 1.16 Å to strands B and C, respectively. The fact that the three identical subunits display three slightly different conformations suggests that crystal contacts trap the structural variations.

Despite the fact that the sequence comprises three repeats, the structure displays appreciable local variations in the superhelical parameters over the length of the RH3 trimer (Figure 4). The superhelical pitch, for example, varies significantly around its mean value of 0.098 centiradians per amino acid. (Superhelical pitch is defined as  $2\pi$  radians divided by the number of residues per superhelical turn. Positive values indicate a right-handed superhelix.) The N-terminal two-thirds of the structure has nearly double the mean pitch. The C-terminal third of the structure, which contains a 3,4 hydrophobic repeat, locally displays left-handed supercoiling. The radius of the supercoil is larger at the ends, reaching maxima of more than 7 Å at the C terminus and more than 6.6 Å at the N terminus. The minimum radius of the bundle, observed about two-thirds of the way from the N terminus, is less than 6.2 Å. The deviation of the local structure from an idealized superhelix is greatest at the ends. This deviation follows a strikingly smooth, approximately parabolic curve with a minimum near residue 20. The experimental electron density and the atomic  $B$ -factors display a similar pattern, suggesting unraveling of the ends of the trimer. In all three helices, the main-chain  $B$ -factors are lowest from residue 17 to 25, and the experimental electron density for both main-chain and side-chain atoms has the highest values in the same area. This region includes the Glu residues that ligate the nickel ions used for phasing. The structural variations may reflect local differences in stability, end effects that



**Figure 3.** RH3 forms a right-handed, trimeric coiled coil. (A) The experimental 2.0 Å electron density map after solvent flattening, contoured at  $1\sigma$ , is superimposed on the final refined model. The refined structure is shown in a stick representation, with carbon atoms in gold, nitrogen atoms in blue, and oxygen atoms in red. (B) The final  $2F_o - F_c$  map after refinement of the crystal structure. The final map is strikingly similar to the experimental map, confirming the accuracy of the experimental map. (C) Stereo view of the overall structure, seen from the side. The van der Waals surfaces are shown for side-chains at **a** (red), **d** (gold), and **h** (blue) positions. The N termini of the helices are at the top. The Glu residues and the coordinated nickel ions are depicted in green.



**Figure 4.** Variation of superhelical parameters. Local superhelical parameters<sup>13</sup> were calculated by a least-squares fit of the parameterized backbone to five-residue segments centered on the hydrophobic core amino acid residues of RH3. Residue 32 was fit using a four-residue window, residues 30 to 33. Local unraveling of the RH3 crystal structure at the C terminus precluded a meaningful fit for the most C-terminal region. Positive values of the superhelical pitch ( $\omega_0$ ) indicate a right-handed supercoil; negative values are left-handed. The average values of the superhelical radius (6.36 Å predicted, 6.50 Å observed) and the superhelical pitch (0.100 centiradians per amino acid residue predicted, 0.098 observed) are in good agreement with the prediction.

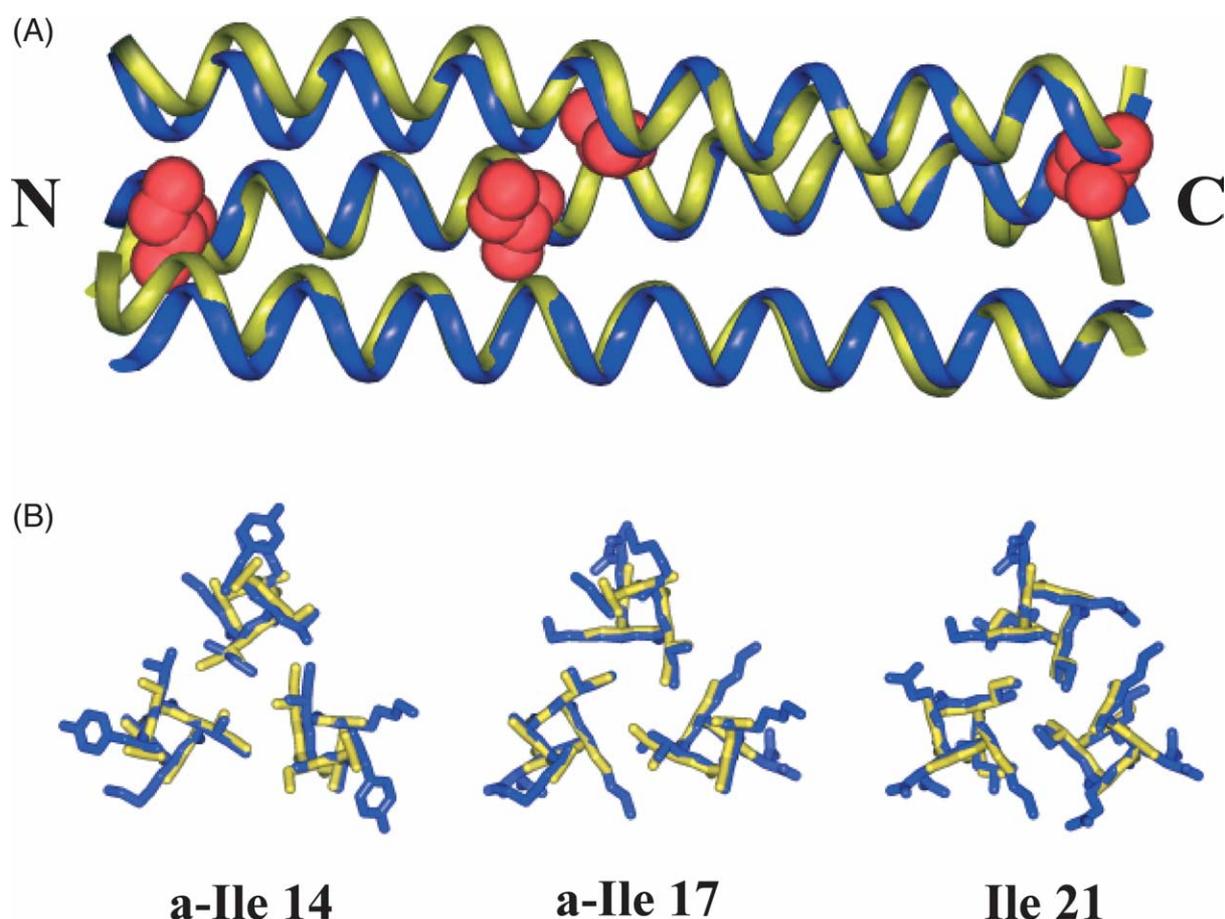
permit superhelical excursions or differences in crystal contacts of the three 11 residue repeats.

### The hydrophobic core

Like the backbone, the hydrophobic core of RH3 lacks 3-fold rotational symmetry. In several core layers, equivalent residues in different monomers adopt different rotamers. Examination of these rotamers reveals why some alternate rotamers are tolerated. These changes affect the overall distribution of side-chain mass without disrupting the knobs-into-holes packing typical of coiled coils. The allo-isoleucine (a-Ile) residues at the **a** positions (residues 3, 14, and 25), for example, occur in two different rotamers, (*t*, *t*) and (*-*, *-*). The superposition of these two rotamers defines a hole into which the opposite side-chain packs, and either rotamer forms a suitable pocket without clashing with neighboring atoms.

The residues that form the hydrophobic core of RH3 have substantial surface area exposed to solvent. The 27 core amino acid side-chains expose a total of 371.9 Å<sup>2</sup> of surface area, an average of 13.8 Å<sup>2</sup> each. By comparison, the 27 core isoleucine residues in GCN4-pII, a left-handed coiled coil trimer,<sup>27</sup> expose 118.7 Å<sup>2</sup> in total (4.4 Å<sup>2</sup> per core residue). The RH3 structure is straighter than that of GCN4-pII, and this superhelical unwinding opens the groove between helices. In addition, few inter-helical salt-bridges are formed in RH3, leaving the sides of the hydrophobic core exposed.

The RH3 core contains several cavities large enough to accommodate water molecules, although the electron density map does not indicate that ordered water is present. These cavities occur between every **a** layer and the following **d** layer. The cavities in RH3 are about 50% larger than those between **a** and **d** layers in GCN4-pII. The average



**Figure 5.** Close agreement between the predicted and observed RH3 structures. (A) The predicted backbone model (gold) and the observed backbone structure (blue) were superimposed by a least-squares fit of the  $\alpha$ -carbon atoms. Helix A is in the middle (behind the other two), helix C is at the top, and helix B is at the bottom. Red van der Waals surfaces show the four core side-chains that adopted rotamers that were ambiguous or differed from the predicted model. (B) Superposition of the core side-chains in the **a**, **d** and **h** layers of the central 11-residue repeat are depicted looking along the superhelix axis. The difference between the predicted and observed rotamers at a-IleA14 is accompanied by a shift in the backbone. Helices A (top center) and C (bottom right) are locally farther apart, and a-IleA14 fills this extra space. The packing of a-IleA3 (not shown) follows the same pattern.

cavity volumes (inside the contact surface<sup>28</sup>) calculated using a 1.4 Å probe sphere were 29.6 Å<sup>3</sup> in GCN4-pII and 45.9 Å<sup>3</sup> in RH3.

### Molecular surface

Each Ni<sup>2+</sup> is coordinated between the Gla residues at positions 19 and 23 of a single RH3 monomer (Figure 2). The coordination geometry is octahedral. Each Gla residue ligates the Ni<sup>2+</sup> with two oxygen atoms, and the remaining two coordination positions are occupied by water molecules or chloride ions.

Metal coordination using Gla residues does not appear to strain the  $\alpha$ -helical backbone. The main-chain dihedral angles in the region of the Gla residues are not significantly different from those in other parts of the structure, and the stereochemistry in this region exhibits no more strain than elsewhere. On the contrary, the lowest *B*-factors in the structure occur in the region encompassing the Gla residues. This apparent rigidity may be due to the

placement of the Gla groups near the center of the trimer, but it may indicate that the Gla side-chains reduce the mobility of the backbone by linking two amino acid residues on successive turns of the helix through metal coordination.

The positively and negatively charged faces of the helices align with each other as intended in the design, but many of the possible interhelical salt-bridges are not formed. This pattern suggests that the overall distribution of charged residues, rather than particular ion pairs, restricts the RH3 fold. Even without specific salt-bridge formation, the charge distribution of the RH peptides favors parallel arrangements and strongly disfavors anti-parallel arrangements. This conclusion agrees well with previous observations that unfavorable electrostatic interactions influence helix direction and specificity.<sup>29–32</sup>

### Agreement with prediction

The observed structure was a close match to the

structure predicted using the parameterized-backbone method (Figure 5). The rmsd between the predicted and observed positions for all main-chain and core side-chain atoms was 1.22 Å for the whole structure and 0.74 Å in the middle third of the structure where end effects are minimized. These values are comparable to those describing the rotational asymmetry within the RH3 structure. Most of the overall superhelical parameters also showed close agreement with the design. The observed/predicted values were 6.50/6.36 Å for the superhelical radius ( $R_0$ ), 0.098/0.100 centiradians per amino acid for the pitch ( $\omega_0$ ), and 0.227/0.103 radians in  $\phi$  arises from the variations at the N and C-terminal ends that were not treated by the periodic boundary conditions imposed in the design.<sup>22</sup>

Of 27 core side-chains, 23 adopted the predicted rotamers. Two of the incorrectly predicted rotamers (a-Ile A3 and a-Ile C32) make crystal contacts at the ends of the structure. The other two unexpected rotamers (a-Ile A14 and a-Ile C217) are adjacent, near the center of the trimer (Figure 5(A)). Three of the four incorrectly predicted core rotamers showed poor density ( $\rho < 1\sigma$  for several atoms) in the experimental map. This weak density suggests that these side-chains may exist in two or three distinct rotamers in the crystal, or they may be distributed broadly in the neighborhood of a canonical rotamer. The 3-fold rotational symmetry was imposed in the design, and this requirement did not account for the variations in rotamers observed within layers of the RH3 structure. This suggests that more detailed calculations that permit symmetry-breaking might lead to more accurate predictions, at the expense of greater computational complexity.

In coiled coils, the hydrophobic core is partly buried by the side-chains of flanking amino acid residues. In the RH3 design, the  $\chi_1$  angles of eight of these flanking amino acid residues (residues 2, 7, 13, 18, 22, 24, 29, and 33) were predicted. For the 24 predicted  $\chi_1$  angles (eight on each chain), three residues have no observable side-chain density in the crystal structure. Of the remaining 21  $\chi_1$  angles, 14 were predicted correctly. In particular, all nine  $\chi_1$  angles in the most regular region (residues 17 to 25) matched the prediction. Many of the deviations appear to be associated with the increased supercoil radius at the ends of the trimer. The partly buried side-chains were expected to be restricted to the  $\chi_1 = 180^\circ$  rotamer by steric contacts with core residues but, because of the increased interhelix spacing at the ends, the side-chains swing inward ( $\chi_1 = -60^\circ$ ) to pack against the core. In general, residues neighboring the core were found in rotamers that pack against the core as closely as possible, given the local backbone geometry.

### Comparison with RH4

The RH3 and RH4 sequences differ only in the placement of a single buried methyl group in each

11 residue repeat. Where RH4 has leucine residues at positions 3, 14, and 25, RH3 contains alloisoleucine residues. Because the total volumes of the hydrophobic cores are identical, the different oligomerization states of RH3 and RH4 arise from the difference in the shape of a single core side-chain in each repeat.

The backbone structures of individual RH3 and RH4 monomer helices are very similar. The  $C^\alpha$  rmsds of the different RH3 monomers are 0.56 Å, 0.88 Å, and 1.05 Å, while the matches between RH3 monomers and the single RH4 monomer in the 4-fold symmetric tetramer are 0.52 Å, 0.56 Å, and 0.50 Å.

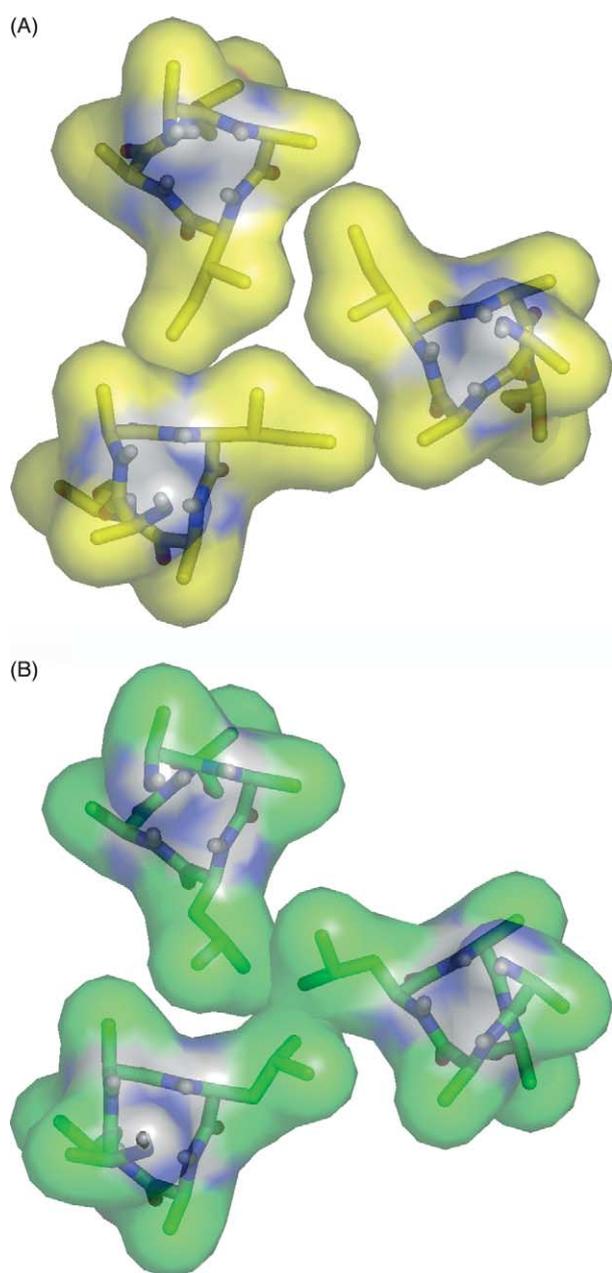
Because the **d** and **h** layers are identical in RH3 and RH4, these residues are compatible with both the trimer and tetramer states. The leucine residues at position **a** in RH4, however, would overlap in the context of the trimeric backbone (Figure 6(A)). Similarly, replacing the leucine residues at **a** positions in the RH4 structure with the alloisoleucine residues of the RH3 sequence would result in a large cavity in the core of the tetramer (not shown). These results suggest that somewhat different packing considerations drive the formation of RH3 trimers and RH4 tetramers. Steric overlaps block trimer formation by the RH4 sequence, and a core cavity disfavors tetramer formation by the RH3 sequence. Rotations of the individual helices (represented in the backbone parameterization as a change in  $\phi$ ) could remedy these packing defects in the **a** layers, but these rotations would disrupt the packing of the other two core layers. Consequently, favorable packing in the coiled-coil core is not simply a matter of choosing suitable side-chains in each position individually, but of assuring that the various side-chains packing into different core positions will all be compatible with a single continuous backbone structure. Avoidance of packing defects, whether steric overlaps of core residues or cavities in the hydrophobic core, provides a strong driving force for oligomer choice and structural uniqueness.

### Conclusions

Exposed  $\gamma$ -carboxyglutamic acid residues in an ( $i, i+4$ ) arrangement may provide a general motif for obtaining a well-ordered, high-occupancy, metal-binding site in helical peptides. The metal-binding site in RH3 did not distort the helical structure of the peptide backbone. This motif offers a new approach to MAD phasing for coiled coils and other synthetic proteins with solvent-exposed helices.

The crystal structure of RH3 revealed that the designed sequence folds into the intended structure, a right-handed, trimeric coiled-coil. The small deviations of the experimental structure from the prediction are dominated by end effects, crystal contacts and the absence of the 3-fold symmetry assumed in the design calculations.

Comparison of the RH3 and RH4 structures



**Figure 6.** Steric interactions specify oligomeric states in the RH family. (A) Axial view of a hypothetical trimer constructed by superimposing the RH4 monomer structure on the backbone observed in the RH3 crystal structure. Placing leucine at a positions in the RH3 backbone causes a steric clash between the leucine side-chains, shown as semi-transparent van der Waals surfaces. (B) At the equivalent position in RH3 (yellow), the allo-isoleucine residues pack the core with no clash or large cavity.

indicates that the oligomerization of RH coiled coils is influenced strongly by avoidance of both steric overlaps and cavities in the core. A similar sensitivity to core packing is displayed by alternate left-handed coiled-coil oligomers.<sup>33</sup>

The structures of RH3 and RH4 show that the parameterized-backbone method enables the design of a protein fold without the use of a

known structure as a template. The predicted and observed structures matched in detail. The discrimination between RH3 and RH4 sequences indicates that the parameterized-backbone method is highly sensitive to the placement of key packing elements. These results suggest the possibility of extending backbone parameterization to other structures for protein modeling and design.

## Materials and Methods

The design of the RH3 structure using the parameterized backbone has been described in detail.<sup>22</sup> The RH3 sequence is AEaIEQaIKKEIAYL aIKKaIKAEILAEaIK-KaIKQEIA. The RH3 and RH3-Gla 33 amino acid residue peptides comprising three 11 residue repeats with acetylated and amidated N and C termini, respectively, were synthesized and purified as described.<sup>23,34</sup> RH3-Gla was crystallized by the hanging-drop, vapor-diffusion method from 100 mM Tris (pH 8.5), 0.3 M sodium bromide, 15 mM NiCl<sub>2</sub>, 15% (v/v) methyl pentanediol, and 15% (w/v) polyethylene glycol 4000. The asymmetric unit of the crystals contained one trimer. Crystal growth required the presence of divalent metal ions such as NiCl<sub>2</sub> or CoCl<sub>2</sub>. Data were collected at 70 K at Beamline 1-5 at SSRL and processed using MOSFLM,<sup>35</sup> the CCP4 data software,<sup>36</sup> and an early version of the Elves automation scripts.<sup>37</sup> Three metal sites were refined using MLPHARE and, after density modification with DM,<sup>36</sup> the map was readily interpretable. The predicted model was aligned with the density and manually adjusted to conform to the map. The structure was refined using cycles of manual rebuilding with O,<sup>38</sup> and refinement with XPLOR<sup>39</sup> or, at later stages, with CNS.<sup>40</sup> Molecular superpositions were carried out using XPLOR.<sup>39</sup> Surface areas and volumes were calculated using a 1.4 Å radius probe using the Connolly molecular-surface software suite.<sup>28</sup> Helical and superhelical parameters were calculated as described.<sup>22,23</sup>

## Protein Data Bank accession code

The coordinates have been deposited with the Protein Data Bank (accession code 1TGG).

## References

1. Bolon, D. N. & Mayo, S. L. (2001). Enzyme-like proteins by computational design. *Proc. Natl Acad. Sci. USA*, **98**, 14274–14279.
2. Chevalier, B. S., Kortemme, T., Chadsey, M. S., Baker, D., Monnat, R. J. & Stoddard, B. L. (2002). Design, activity, and structure of a highly specific artificial endonuclease. *Mol. Cell*, **10**, 895–905.
3. Benson, D. E., Conrad, D. W., de Lorimier, R. M., Trammell, S. A. & Hellinga, H. W. (2001). Design of bioelectronic interfaces by exploiting hinge-bending motions in proteins. *Science*, **293**, 1641–1644.
4. Kuhlman, B., O'Neill, J. W., Kim, D. E., Zhang, K. Y. & Baker, D. (2002). Accurate computer-based design of a new backbone conformation in the second turn of protein L. *J. Mol. Biol.* **315**, 471–477.
5. Desjarlais, J. R. & Handel, T. M. (1999). Side-chain and backbone flexibility in protein core design. *J. Mol. Biol.* **290**, 305–318.
6. DeGrado, W. F., Summa, C. M., Pavone, V., Nastro, F. &

- Lombardi, A. (1999). *De novo* design and structural characterization of proteins and metalloproteins. *Annu. Rev. Biochem.* **68**, 779–819.
7. Domingues, H., Cregut, D., Sebald, W., Oschkinat, H. & Serrano, L. (1999). Rational design of a GCN4-derived mimetic of interleukin-4. *Nature Struct. Biol.* **6**, 652–656.
  8. Benson, D. E., Wisz, M. S., Liu, W. & Hellinga, H. W. (1998). Construction of a novel redox protein by rational design: conversion of a disulfide bridge into a mononuclear iron-sulfur center. *Biochemistry*, **37**, 7070–7076.
  9. Malakauskas, S. M. & Mayo, S. L. (1998). Design, structure and stability of a hyperthermophilic protein variant. *Nature Struct. Biol.* **5**, 470–475.
  10. Kortemme, T., Ramirez-Alvarado, M. & Serrano, L. (1998). Design of a 20-amino acid, three-stranded beta-sheet protein. *Science*, **281**, 253–256.
  11. Tuffery, P., Etchebest, C. & Hazout, S. (1997). Prediction of protein side-chain conformations: a study on the influence of backbone accuracy on conformation stability in the rotamer space. *Protein Eng.* **10**, 361–372.
  12. Dahiyat, B. I. & Mayo, S. L. (1997). *De novo* protein design: fully automated sequence selection. *Science*, **278**, 82–87.
  13. Dahiyat, B. I., Sarisky, C. A. & Mayo, S. L. (1997). *De novo* protein design: towards fully automated sequence selection. *J. Mol. Biol.* **273**, 789–796.
  14. Lazar, G. A., Desjarlais, J. R. & Handel, T. M. (1997). *De novo* design of the hydrophobic core of ubiquitin. *Protein Sci.* **6**, 1167–1178.
  15. Desjarlais, J. R. & Handel, T. M. (1995). *De novo* design of the hydrophobic cores of proteins. *Protein Sci.* **4**, 2006–2018.
  16. Koehl, P. & Delarue, M. (1994). Application of a self-consistent mean field theory to predict protein side-chains conformation and estimate their conformational entropy. *J. Mol. Biol.* **239**, 249–275.
  17. Hellinga, H. W. & Richards, F. M. (1994). Optimal sequence selection in proteins of known structure by simulated evolution. *Proc. Natl Acad. Sci. USA*, **91**, 5803–5807.
  18. Lombardi, A., Summa, C. M., Geremia, S., Randaccio, L., Pavone, V. & DeGrado, W. F. (2000). Retrostructural analysis of metalloproteins: application to the design of a minimal model for diiron proteins. *Proc. Natl Acad. Sci. USA*, **97**, 6298–6305.
  19. Summa, C. M., Rosenblatt, M. M., Hong, J. K., Lear, J. D. & DeGrado, W. F. (2002). Computational *de novo* design and characterization of an A(2)B(2) diiron protein. *J. Mol. Biol.* **321**, 923–938.
  20. DeGrado, W. F., Di Costanzo, L., Geremia, S., Lombardi, A., Pavone, V. & Randaccio, L. (2003). Sliding helix and change of coordination geometry in a model di-MnII protein. *Angew. Chem. Int. Ed. Engl.* **42**, 417–420.
  21. Kuhlman, B., Dantas, G., Ireton, G. C., Varani, G., Stoddard, B. L. & Baker, D. (2003). Design of a novel globular protein fold with atomic-level accuracy. *Science*, **302**, 1364–1368.
  22. Harbury, P. B., Tidor, B. & Kim, P. S. (1995). Repacking protein cores with backbone freedom: structure prediction for coiled coils. *Proc. Natl Acad. Sci. USA*, **92**, 8408–8412.
  23. 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.
  24. Crick, F. (1953). The Fourier transform of a coiled-coil. *Acta Crystallog.* **6**, 685–689.
  25. Stenflo, J. & Suttie, J. W. (1977). Vitamin K-dependent formation of gamma-carboxyglutamic acid. *Annu. Rev. Biochem.* **46**, 157–172.
  26. Seshadri, T. P., Skrzypczak-Jankun, E., Yin, M. & Tulinsky, A. (1994). Differences in the metal ion structure between Sr- and Ca-prothrombin fragment 1. *Biochemistry*, **33**, 1087–1092.
  27. Harbury, P. B., Kim, P. S. & Alber, T. (1994). Crystal structure of an isoleucine-zipper trimer. *Nature*, **371**, 80–83.
  28. Connolly, M. L. (1983). Solvent-accessible surfaces of proteins and nucleic acids. *Science*, **221**, 709–713.
  29. O'Shea, E. K., Rutkowski, R. & Kim, P. S. (1992). Mechanism of specificity in the Fos-Jun oncoprotein heterodimer. *Cell*, **68**, 699–708.
  30. Zhou, N. E., Kay, C. M. & Hodges, R. S. (1994). The net energetic contribution of interhelical electrostatic attractions to coiled-coil stability. *Protein Eng.* **7**, 1365–1372.
  31. Vinson, C. R., Hai, T. & Boyd, S. M. (1993). Dimerization specificity of the leucine zipper-containing bZIP motif on DNA binding: prediction and rational design. *Genes Dev.* **7**, 1047–1058.
  32. Lumb, K. J. & Kim, P. S. (1995). Measurement of interhelical electrostatic interactions in the GCN4 leucine zipper. *Science*, **268**, 436–439.
  33. Harbury, P. B., Zhang, T., Kim, P. S. & Alber, T. (1993). A switch between two-, three-, and four-stranded coiled coils in GCN4 leucine zipper mutants. *Science*, **262**, 1401–1407.
  34. Lockhart, D. J. & Kim, P. S. (1992). Internal stark effect measurement of the electric field at the amino terminus of an alpha helix. *Science*, **257**, 947–951.
  35. Leslie, A. G. W. (1992). *Joint CCP4 ESF-EACMB Newsletter Protein Crystallography*, vol. 26. Warrington, UK: Daresbury Laboratory.
  36. Collaborative Computational Project Number 4. (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallog. sect. D*, **50**, 760–763.
  37. Holton, J. M. & Alber, T. (2004). Automated protein structure determination with Elves. *Proc. Natl Acad. Sci. USA*, **101**, 1539–1542.
  38. Jones, T. A., Zou, J. Y. & Cowan, S. W. (1991). Kjeldgaard improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallog. sect. A*, **47**, 110–119.
  39. Brunger, A. T., Kuriyan, J. & Karplus, M. (1987). Crystallographic R factor refinement by simulated annealing. *Science*, **235**, 458–460.
  40. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W. *et al.* (1998). Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallog. sect. D*, **54**, 905–921.

Edited by I. Wilson

(Received 27 January 2004; received in revised form 11 June 2004; accepted 15 June 2004)