

# High-Resolution Protein Design with Backbone Freedom

Pehr B. Harbury,\* Joseph J. Plecs, Bruce Tidor,  
Tom Alber, Peter S. Kim†

Recent advances in computational techniques have allowed the design of precise side-chain packing in proteins with predetermined, naturally occurring backbone structures. Because these methods do not model protein main-chain flexibility, they lack the breadth to explore novel backbone conformations. Here the *de novo* design of a family of  $\alpha$ -helical bundle proteins with a right-handed superhelical twist is described. In the design, the overall protein fold was specified by hydrophobic-polar residue patterning, whereas the bundle oligomerization state, detailed main-chain conformation, and interior side-chain rotamers were engineered by computational enumerations of packing in alternate backbone structures. Main-chain flexibility was incorporated through an algebraic parameterization of the backbone. The designed peptides form  $\alpha$ -helical dimers, trimers, and tetramers in accord with the design goals. The crystal structure of the tetramer matches the designed structure in atomic detail.

Proteins exhibit precise geometric packing of atoms in their interiors. Nevertheless, empirical protein design methods have achieved a measure of generality and simplicity by ignoring detailed interactions between amino acid residues. These design approaches rely instead on imitation of statistical sequence patterns in naturally occurring proteins, such as hydrophobic-polar residue patterns, and characteristic local interaction motifs (1). Perhaps as a direct consequence, many designed proteins exhibit fluctuating or "molten" interiors (2), and some assume unintended tertiary conformations (3).

Packing in proteins has been studied computationally by holding the protein main chain in the wild-type conformation (the "fixed-backbone" approximation) and asking which sets of amino acid side chains can efficiently fill the interior space (4). This technique has been used successfully to repack wild-type side chains into predeter-

mined backbone structures (5) and has more recently been extended to the design of amino acid sequences. Fixed-backbone design based on naturally occurring backbone templates has produced proteins that fold to the target structures with high thermal stabilities (6).

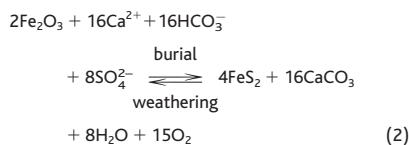
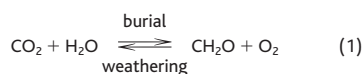
However, fixed-backbone methods are unable to model protein main-chain flexibility. Thus, main-chain adjustments known to occur in response to core mutations in proteins (7) are not allowed. Moreover, the fixed-backbone approach has a severe limitation when applied to backbone structures for which a naturally occurring example does not exist. Although naturally occurring backbone coordinates represent the ground-state conformation for at least one sequence (the naturally occurring sequence), this assumption is not necessarily valid for arbitrary backbone coordinates. When the backbone structure is designed *de novo*, a complete set of plausible backbone conformations must be sampled to identify structures that lie at a free-energy minimum in the sequence and conformational spaces.

These limitations may be overcome by treating the backbone as a parametric family of structures rather than as a static entity (8). A small but well-defined subset of main-chain conformations can then be exhaustively sampled in a finite time. For example, coupled searches of side-chain packing and main-chain conformation under a parametric coiled-coil backbone model have been used to reproduce detailed, crystallographically observed conformations for coiled-coil proteins (8).

A true test of the utility of parametric-backbone models in protein design would be to engineer a protein fold for which no structural example is known. We report here the computational design of a family of dimeric, trimeric, and tetrameric  $\alpha$ -helical bundles

(Geological Society Special Publication 101, 1996), pp. 401–441; T. J. Bralower *et al.*, *Geology* **25**, 963 (1997).

24. L. R. Kump and R. M. Garrels, *Am. J. Sci.* **286**, 337 (1986).
25. M. E. Raymo, W. F. Ruddiman, P. N. Froelich, *Geology* **16**, 649 (1988); R. A. Beck *et al.*, *Nature* **373**, 55 (1995); R. A. Beck, D. W. Burbank, W. J. Sercombe, T. L. Olson, A. M. Khan, *Geology* **23**, 387 (1995).
26. K. G. Miller, R. G. Fairbanks, G. S. Mountain, *Paleoceanography* **2**, 1 (1987); W. F. Ruddiman, *Nature* **394**, 723 (1998).
27. B. Peucker-Ehrenbrink, G. Ravizza, A. W. Hofmann, *Earth Planet. Sci. Lett.* **130**, 155 (1995).
28. R. A. Berner, *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **75**, 97 (1989).
29. W. H. Berger, V. S. Smetacek, G. Wefer, Eds., *Ocean Productivity and Paleoproductivity—An Overview* (Wiley, New York, 1989); M. L. Delaney and E. A. Boyle, *Paleoceanography* **3**, 137 (1988); G. H. Rau, in *Carbon Cycling in the Glacial Ocean Constrains on the Oceanic Role in Global Change: Quantitative Approaches in Paleoceanography*, R. Zahn, T. F. Pedersen, M. A. Kaminski, L. Labeyrie, Eds. (Springer-Verlag, Berlin, 1994), pp. 307–321; J. C. Zachos, L. D. Scott, K. C. Lohmann, *Paleoceanography* **9**, 353 (1994).
30. W. J. Shackleton, in *Marine Petroleum Source Rocks*, J. Brooks and A. J. Fleet, Eds. (Geological Society, London, 1987), pp. 423–434.
31. This is represented by the following two chemical equations



[H. D. Holland, *The Chemical Evolution of the Atmospheres and Oceans* (Princeton Univ. Press, Princeton, NJ, 1984); H. D. Holland, *Geochim. Cosmochim. Acta* **37**, 2605 (1973); R. M. Garrels and E. A. Perry Jr., in *The Sea*, 5, *Marine Chemistry*, E. D. Goldberg, Ed. (Wiley, New York, 1974); R. M. Garrels, A. Lerman, F. T. Mackenzie, *Am. Sci.* **63**, 306 (1976); M. Schidlowski, C. E. Junge, H. Pietrek, *J. Geophys. Res.* **82**, 2557 (1977); R. M. Garrels and A. Lerman, *Am. J. Sci.* **284**, 989 (1984); R. A. Berner, *ibid.* **287**, 177 (1987); F. T. Mackenzie, L. M. Ver, C. Sabine, M. Lane, A. Lerman, in *Interactions of C, N, P, and S Biogeochemical Cycles and Global Change* (NATO ASI Series, 1993), vol. 14, pp. 1–62].

32. L. V. Berkner and L. C. Marshall, *J. Atmos. Sci.* **22**, 225 (1965); A. Watson, J. E. Lovelock, L. Margulis, *Biosystems* **10**, 293 (1978); T. P. Jones and W. G. Chaloner, *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **77**, 39 (1991); J. M. Robinson, *ibid.*, p. 51.
33. J. Veizer, W. T. Holser, C. K. Wilgus, *Geochim. Cosmochim. Acta* **44**, 579 (1980); T. B. Lindh, thesis, University of Miami (1983).
34. P. Van Cappellen and E. D. Ingall, *Science* **271**, 493 (1996); A. S. Colman, H. D. Holland, F. T. Mackenzie, *ibid.* **275**, 406 (1997).
35. S. J. Carpenter and K. C. Lohmann, *Geochim. Cosmochim. Acta* **61**, 4831 (1997).
36. M. Schidlowski and C. E. Junge, *ibid.* **45**, 589 (1981).
37. J. Hoefs, in *Hand Book of Geochemistry*, K. H. Wedepohl, Ed. (Springer-Verlag, Berlin, 1967), vol. I-6.
38. S. T. Petsch, *Geochim. Cosmochim. Acta*, in press.
39. R. A. Berner, *ibid.* **48**, 605 (1984).
40. ——— and R. Raiswell, *ibid.* **47**, 855 (1983); R. Raiswell and R. A. Berner, *ibid.* **50**, 1967 (1986).
41. S. C. Cande and D. V. Kent, *J. Geophys. Res.* **100**, 6093 (1995).
42. Supported by NSF grant EAR-9628479. Cores were provided by the Ocean Drilling Program. We thank T. Jackson for help with sample analysis.

27 July 1998; accepted 25 September 1998

P. B. Harbury is at the Whitehead Institute for Biomedical Research, Howard Hughes Medical Institute and Department of Biology, Massachusetts Institute of Technology, Nine Cambridge Center, Cambridge, MA 02142, USA, and Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA 02115, USA. J. J. Plecs is in the Department of Molecular and Cell Biology and Department of Physics, University of California at Berkeley, Berkeley, CA 94720, USA. B. Tidor is in the Department of Chemistry, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. T. Alber is in the Department of Molecular and Cell Biology, University of California at Berkeley, Berkeley, CA 94720, USA. P. S. Kim is at the Whitehead Institute for Biomedical Research, Howard Hughes Medical Institute and Department of Biology, Massachusetts Institute of Technology, Nine Cambridge Center, Cambridge, MA 02142, USA.

\*Present address: Department of Biochemistry, Stanford University, Stanford, CA 94305, USA.

†To whom correspondence should be addressed. E-mail: dvorak@wi.mit.edu

with a right-handed superhelical structure. Although a nuclear magnetic resonance (NMR) structure of a right-handed dimer of helices in detergent micelles has been reported (9), no structures of trimeric or tetrameric right-handed coiled coils exist. The x-ray crystal structure of the tetrameric bundle designed here matches the predicted structure in atomic detail, adopting an unprecedented, but deliberately engineered, right-handed superhelical fold.

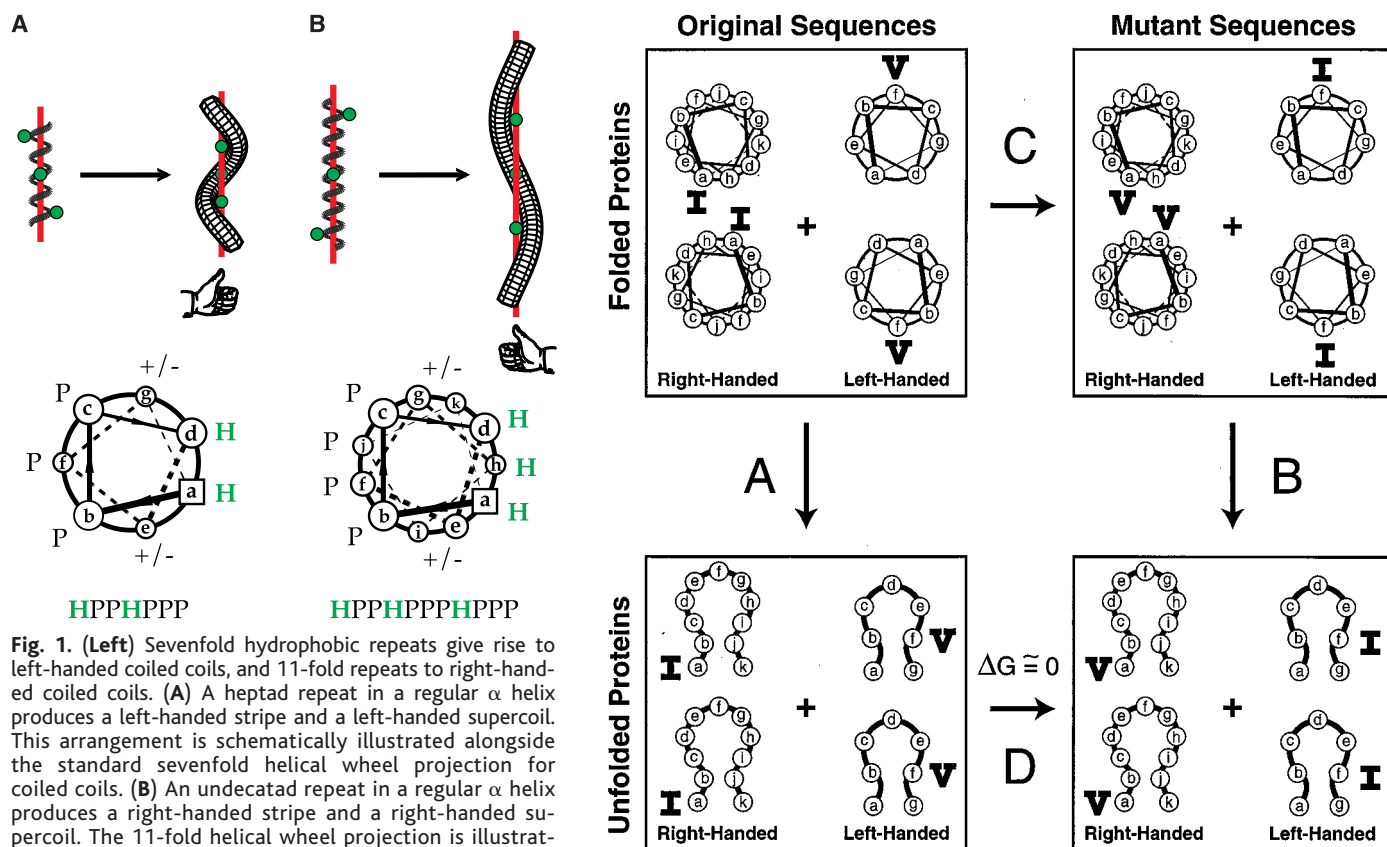
**Design principles.** The first step of our design is based on an analysis of the hydrophobic-polar residue pattern in left-handed coiled coils. The superhelical twist of left-handed coiled coils arises from a small difference between the integral frequency of the heptad repeat and the characteristic frequency of  $\alpha$  helices (10). Each amino acid in a straight helix rotates about  $100^\circ$  radially

around the helix axis ( $360^\circ$  for 3.6 residues). Seven amino acids rotate  $700^\circ$ , lagging two full turns ( $720^\circ$ ) by  $20^\circ$ . A sevenfold repeat, therefore, forms a left-handed stripe in a straight  $\alpha$  helix (Fig. 1). In a left-handed supercoiled conformation that evolves  $20^\circ$  every seven residues, this stripe can always face toward the axis of superhelical rotation (Fig. 1).

Application of this principle to an 11-fold (undecatad) amino acid repeat suggests that a right-handed supercoil should form. Eleven amino acids rotate about  $1100^\circ$ , which leads three full turns ( $1080^\circ$ ) by  $20^\circ$ . Thus, an undecatad repeat produces a right-handed stripe in a straight  $\alpha$  helix, which should give rise to a right-handed supercoil (Fig. 1). Examination of an 11-residue helical-wheel projection indicates that amino acids in the first, fourth, and eighth positions (positions a, d,

and h) fall on the same surface of the helix. These considerations suggest that a 3-4-4 hydrophobic repeat might specify a right-handed coiled coil.

The second step of our design consisted of determining which amino acids can pack the core of a right-handed bundle with a 3-4-4 hydrophobic repeat, and by their shapes direct dimer, trimer, or tetramer formation. Detailed dimer, trimer, and tetramer right-handed coiled coils were modeled for all possible core sequences made up of the small aliphatic amino acids alanine, valine, norvaline, leucine, isoleucine, and alloisoleucine. Alloisoleucine, the stereoisomer of isoleucine with inverted chirality at the C $\beta$  carbon, was included in the design calculation because preliminary models suggested the need for a residue that would orient side-chain volume into a trans  $\chi_1$  dihedral angle in its most



**Fig. 1. (Left)** Sevenfold hydrophobic repeats give rise to left-handed coiled coils, and 11-fold repeats to right-handed coiled coils. **(A)** A heptad repeat in a regular  $\alpha$  helix produces a left-handed stripe and a left-handed supercoil. This arrangement is schematically illustrated alongside the standard sevenfold helical wheel projection for coiled coils. **(B)** An undecatad repeat in a regular  $\alpha$  helix produces a right-handed stripe and a right-handed supercoil. The 11-fold helical wheel projection is illustrated. H, hydrophobic residue; P, polar residue; +/-, charged residue. **Fig. 2. (Right)** Calculating the effects of sequence changes on unfolding free energies, illustrated for an isoleucine-to-valine substitution at position a of a right-handed coiled-coil dimer. The top panels show axial projections of one right-handed and one left-handed coiled coil. The top left panel differs from the top right panel by interchange of the amino acid at position a of the right-handed coiled coil with the amino acid at position f of the left-handed coiled coil. The lower panels show the unfolded polypeptides. Each leg of the thermodynamic cycle is labeled with a letter. The legs labeled A and B correspond to unfolding of the wild-type and mutant coiled-coil sequences respectively. The legs labeled C and D correspond to residue permutation in the folded and unfolded states, respectively. Because the cycle is closed, the difference in the two unfolding free energies is equal to the difference in the two permutation free energies:  $A - B = C - D$ . The A and B legs

are a sum of two terms, the unfolding free energy for the right-handed coiled coil,  $\Delta G_{RH}^{unf,old}$ , and the unfolding free energy for the left-handed coiled coil,  $\Delta G_{LH}^{unf,old}$ . Expanding A and B and rearranging terms gives  $[\Delta G_{RH}^{unf,old}, (I \text{ at } a) - \Delta G_{RH}^{unf,old}, (V \text{ at } a)] = C - D + [\Delta G_{LH}^{unf,old}, (I \text{ at } f) - \Delta G_{LH}^{unf,old}, (V \text{ at } f)]$ . For the computational studies reported here (12, 16), D was assumed to be 0 kcal/mol, C was computed from the bonded, van der Waals and hydrogen bonding terms of the CHARMM19 potential (26) and a solvent accessible surface hydration potential (25), and  $[\Delta G_{LH}^{unf,old}, (V \text{ at } f) - \Delta G_{LH}^{unf,old}, (I \text{ at } f)]$  was taken from experimentally measured free energies of unfolding (15). Differences in calculated stability are dominated by the CHARMM19 potential, which accounts for 80% of the variance in the calculated stabilities of right-handed coiled-coil sequences. The surface hydration potential and left-handed coiled-coil unfolding free energies each account for  $\sim 10\%$  of the variance in the calculated stabilities.

commonly occurring rotamer. Norvaline (an *n*-propyl side chain) was used as a general straight-chain analog for methionine and lysine (11). These six amino acids could be placed at the core positions a, d, and h in 216 ( $6^3$ ) possible sequences. However, because most of the amino acids have multiple side-chain conformations, about 25 possible side-chain rotamers exist at each position. For simplicity, the design calculations were limited to the 11 lowest energy rotamers at each core position in each oligomeric state, as determined by single-level packing calculations (12). Thus, 3993 ( $3 \times 11^3$ ) structures (ignoring variations in backbone conformation) were generated for the design analysis.

**Computation.** To model the right-handed structures, we modified a technique previously used to predict crystal structures of natural left-handed coiled coils (8, 12). For each of the 3993 core rotamer conformations, main-chain coordinates were determined by exploring a parametric family of superhelix backbones described originally by Francis Crick (13). The parametric backbone algebra for left-handed coiled coils was altered to reflect an 11-fold amino acid repeat. Periodic boundary conditions were applied, and two-, three-, or fourfold rotational symmetry was imposed around the superhelix axis. For the computational studies, alanine residues were placed at the exterior positions of the unde-

catad repeat (positions b, c, f, g, and j), and  $\alpha$ -amino butyric acid residues (an ethyl side chain) were placed at positions on the boundary of the hydrophobic core (positions e, i, and k). As the backbone coordinate search for each core rotamer conformation required  $\sim 3$  min on a MIPS R3000 processor, the entire calculation took about 8 days.

A difficult aspect of protein design is the need to compare unfolding free energies for different candidate amino acid sequences. The effects of sequence changes on unfolding free energies can be calculated as the difference in the free energy of mutation in the folded and unfolded states (14). To avoid explicitly modeling the unfolded state, which consists of a large ensemble of conformations, we modified this basic strategy by calculating an energy of permutation: the energy difference between two different covalent arrangements of the same amino acids (Fig. 2). Whereas energies of mutation are seldom zero, an energy of permutation will be zero if the amino acid side chains do not interact with each other, which we assume to be the case in the unfolded state. Importantly, conformation-independent energetic biases for one amino acid over another in the underlying potential energy function are canceled in a permutation difference. The permutation used here interchanges an amino acid at position f of a left-handed coiled coil with an

amino acid in the hydrophobic core of a right-handed coiled coil. Using experimentally measured unfolding free energies of f-substituted left-handed coiled coils (15), and assuming that the free energy of permutation in the unfolded state is zero, we derived calculated unfolding energy differences for different undecad sequences in the right-handed model conformation (16).

**Experimentally observed structures.** Two criteria were used to choose optimal core sequences for each oligomeric state. First, the stability of the sequence in the target conformation was required to be high. Second, the specificity of the sequence for forming the target conformation, instead of the two alternative oligomeric states, was maximized. We took the following approach: (i) Within each oligomeric state, the mean stability averaged over the entire family of 216 possible sequences was tabulated; stabilities of individual sequences were expressed as standard deviations from this mean. (ii) Because equilibria between different oligomerization states depend on monomer concentration and on residues outside of the hydrophobic core, we calculated specificities using an

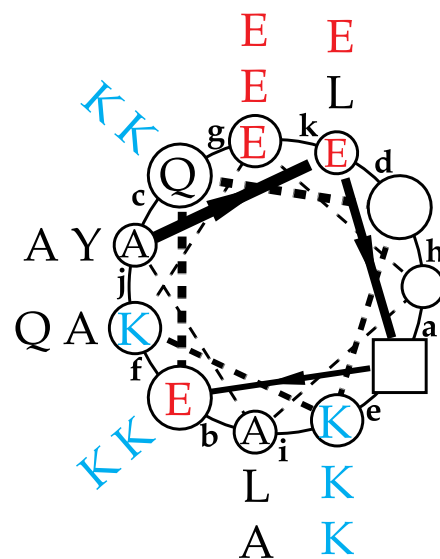
**Table 1.** Top sequence solutions for right-handed dimer, trimer, and tetramer coiled coils according to the packing calculation. The table consists of three lists, the first sorted according to dimer stability, the second according to trimer stability, and the third according to tetramer stability.

Sequence†			Stability‡	Specificity§	
a	d	h		Dimer-trimer	Dimer-tetramer
Dimer					
nV	L	nV	2.0	0.9	1.3
nV	nV	L	1.9	0.8	0.7
nV	L	L	1.9	0.7	0.9
I	L	nV*	1.8	1.8	1.0
Trimer				Trimer-dimer	Trimer-tetramer
nV	al	L	2.6	0.3	-0.5
I	al	L	2.4	0.5	0.2
I	al	nV	2.4	0.6	0.6
nV	al	nV	2.3	0.6	-0.4
L	al	L	2.0	-0.1	-0.8
al	al	I*	2.0	1.9	2.0
Tetramer				Tetramer-dimer	Tetramer-trimer
nV	al	I	3.1	2.2	1.7
nV	al	L	2.6	0.6	0.5
L	al	I*	2.5	1.9	2.0
L	al	L	2.3	0.3	0.8

\*The core sequences chosen for RH2, RH3, and RH4 are indicated by asterisks. These sequences were chosen on the basis of two criteria: high predicted stabilities in the target conformation, and high predicted preference for the target conformation over the two competing alternatives. Thus, large positive values in both the stability and the specificity columns were required. †Residues present at positions a, d, and h for each sequence in the table (28).

‡Within each oligomerization state, the stabilities of individual sequences are expressed in units of standard deviation from the mean stability of all 216 possible sequences (see text). More positive values indicate larger predicted stability.

§Specificity of each sequence for the target conformation over the two competing conformations (for example, the list sorted on dimer stability shows the dimer-over-trimer and dimer-over-tetramer specificities). More positive values indicate larger predicted specificities for the target conformation (see text).



**Fig. 3.** Helical wheel projection of residues Ala<sup>1</sup> to Ala<sup>33</sup> of the RH coiled-coil template sequence (28). View is from the NH<sub>2</sub>-terminus, and residues in the first three helical turns are boxed or circled. Undecad positions are labeled a through k. The full template sequence consists of three undecadats. The peptides RH2, RH3, and RH4 differ only by the amino acids present at positions a, d, and h (18). The sequences (with positions a, d, and h in italics; al, alioisoleucine; and nV, norvaline) are the following.

RH2-AE IEQ LKKE nVAYL IKK LKAERLAE IKK LKQERA  
 RH3-AEaIEQaIKKE IAYLaIKKaIKAE ILAEaIKKaIKQE IA  
 RH4-AE LEQaIKKE IAYL LKKaIKAE ILAE LKKaIKQE IA  
 The template sequence was chosen to contain positively charged lysine residues at position e on one side of the helix and negatively charged glutamate residues at positions g and k on the opposite side to favor a parallel association of helices in a helical bundle.

arbitrary standard state that factors out these considerations. For each equilibrium, the free energy of interconversion was assigned a value of zero when averaged over all 216 core sequences. Specificities were thus evaluated by taking the stability difference (in kilocalories per mole) for each sequence between the target conformation and the alternative oligomeric conformations, and expressing the differences in units of standard deviation.

Dimer, trimer, and tetramer core sequences thus identified (Table 1) were inserted into a common 33-residue template to generate the peptides RH2, RH3, and RH4 (Fig. 3). The template contains positively charged lysine residues on one side of the hydrophobic core and negatively charged glutamate residues on the other side to favor parallel helix arrangements over antiparallel arrangements (17). Two norvaline residues at positions h of the dimer sequence pack against glutamate residues at positions k in the neighboring helix and were therefore substituted with the isosteric amino acid lysine. Charged and polar residues were placed at the exterior posi-

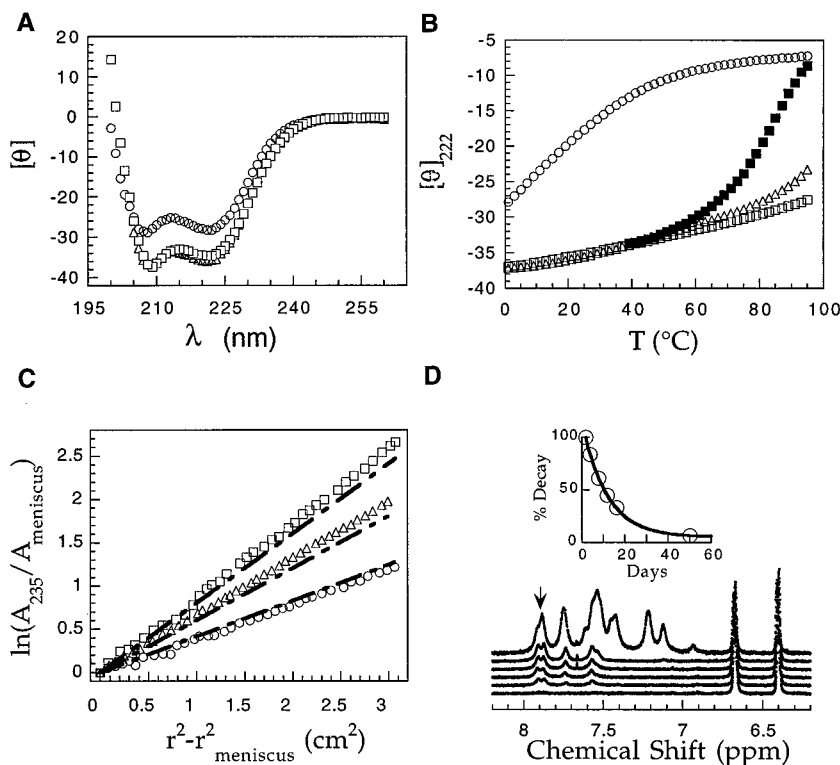
tions of the template to ensure high solubility, and a single tyrosine residue was included to facilitate concentration determination. The RH peptides were prepared by solid-phase synthesis and purified by reversed-phase high-performance liquid chromatography (18).

On the basis of circular dichroism (CD) measurements at 10  $\mu\text{M}$  peptide concentration in neutral pH phosphate-buffered saline (PBS) at 4°C, RH2 appears to be  $\sim 80\%$  helical, whereas RH3 and RH4 appear to be  $>95\%$  helical (19) (Fig. 4A). Under these conditions, RH2 displays a broad thermal unfolding transition, RH3 exhibits a cooperative melt with an apparent melting temperature ( $T_m$ ) of 95°C, and RH4 has a thermal stability that exceeds 100°C (Fig. 4B). In the presence of the denaturant guanidinium hydrochloride (GdmCl) at 3 M concentration, RH4 melts cooperatively with an apparent  $T_m$  of 90°C. Thus, whereas RH2 appears to be incompletely folded under physiologic conditions, RH3 and RH4 form well-structured and extremely stable helical structures.

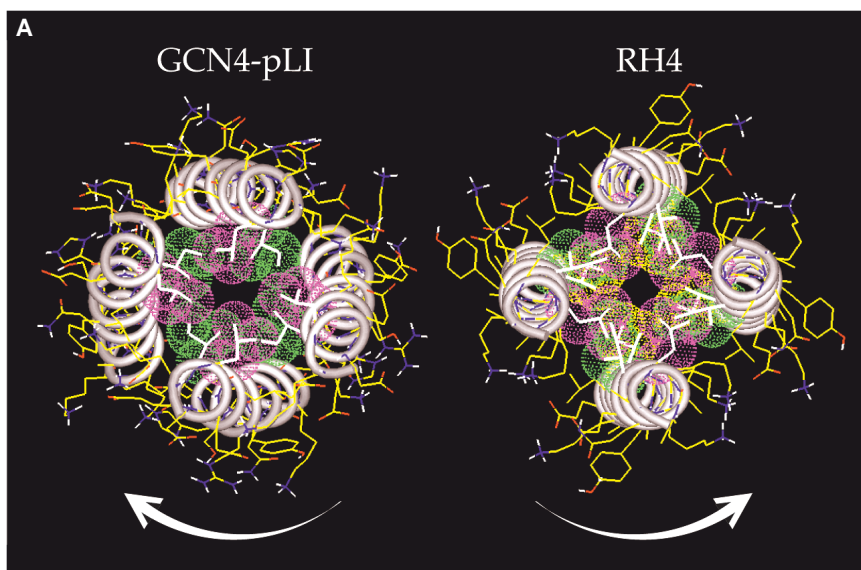
Sedimentation equilibrium measurements (20) were used to determine the oligomerization state of each peptide under native conditions (Fig. 4C). As intended in the design, the RH2 peptide sediments approximately as a dimer, the RH3 peptide sediments approximately as a trimer, and the RH4 peptide sediments approximately as a tetramer. The RH3 and RH4 peptides exhibit no systematic dependence of molecular weight on concentration between 20  $\mu\text{M}$  and 200  $\mu\text{M}$ . The molecular weight of the RH2 peptide systematically decreases at concentrations below 100  $\mu\text{M}$  (presumably because of dissociation in the low micromolar concentration range) but exhibits no systematic deviation from dimer molecular weight between 200  $\mu\text{M}$  and 2 mM. Thus, each of the RH peptides assumes the oligomerization state for which it was designed.

To assess whether the designed oligomers associate in a “molten” fashion or with fixed tertiary structures, we studied the dynamic properties of RH3 and RH4 by hydrogen-deuterium amide-proton exchange (21) (Fig. 4D). Relative to poly D,L-alanine, the most slowly exchanging protons in RH3 were protected by a factor of  $>10^5$ , whereas the most slowly exchanging protons in RH4 were protected by a factor of  $>10^7$ . These protection factors are comparable to protection factors observed for the native states of naturally occurring proteins. The trimer and tetramer structures thus appear to assemble with native-like rigidity.

To evaluate the high-resolution features of the design, we determined the x-ray crystal structure of the RH4 peptide at 1.9 Å resolution (22). The RH4 structure was refined to a conventional  $R$ -factor of 20.4% with a free  $R$ -factor of 24.8% and root mean square deviations from ideal bond lengths and bond angles of 0.009 Å and 1.97°, respectively. The designed peptide forms the intended parallel, right-handed superhelix structure (Fig. 5A). Moreover, the side-chain packing observed in the RH4 crystal structure closely matches that predicted by the design calculation (Fig. 5B). The engineered and observed rotamers for the core side chains are identical. The crystal structure differs from the designed conformation primarily at the COOH-terminus where crystal contacts cause the superhelix to be locally underwound. Such end effects were deliberately ignored by the design method in the interest of simplicity (8). Core side-chain and main-chain atom positions in the central undecad of the crystal structure differ from atom positions in the calculated model by a root mean square deviation of 0.20 Å. The superhelical parameters (8, 13) (radius  $R_0$ , frequency  $\omega_0$ , and phase angle  $\phi$ ) for the designed structure and the  $\text{NH}_2$ -terminal two undecads of the crystal structure are as follows:  $R_0 = 7.29$  Å,  $\omega_0$



**Fig. 4.** The RH2 (open circles), RH3 (open triangles), and RH4 (open squares) peptides form two-, three-, and four-stranded helical bundles. (A) CD spectra at 4°C in PBS (pH 7.0) and 10  $\mu\text{M}$  peptide concentration (19). The mean residue ellipticity,  $[\theta]$ , is reported in units of  $10^3$  degrees  $\text{cm}^2 \text{dmol}^{-1}$ . (B) Thermal melts monitored by CD at 222 nm (19). The filled squares show data for RH4 collected in the presence of 3 M GdmCl, a denaturant. (C) Analytical ultracentrifugation data (32 krpm) collected at 4°C in PBS (pH 7.0) at  $\sim 100$   $\mu\text{M}$  peptide concentration (22). The natural logarithm of the absorbance at 235 nm is plotted against the square of the radial position. Dashed lines with increasing slopes indicate, respectively, the predicted data for dimer, trimer, and tetramer bundles. (D) Aromatic and amide-proton NMR spectra of the RH4 peptide at different times after transfer into  $\text{D}_2\text{O}$  (21). The inset shows the volume of one amide resonance (the peak labeled with an arrow) plotted against exchange time. The data closely fit a single exponential decay with a half-life of  $\sim 10$  days.



**Fig. 5.** Crystal structure of RH4. **(A)** Axial views of the left-handed GCN4-pLI tetramer (27) next to the right-handed RH4 tetramer (22). The view is from the  $\text{NH}_2$ -terminus looking toward the  $\text{COOH}$ -terminus. Purple van der Waals surfaces identify residues at the a positions, green surfaces identify residues at the d positions, and yellow surfaces identify residues at the h positions. **(B)** Superposition of the central undecad of the calculated (red) and crystallographically observed (blue) structures of RH4. Three cross-sections of the superhelix, centered at positions a, d, and h, are shown. In each case the calculated side-chain packing conformation matches the conformation observed in the crystal structure (11). **(C)** Superposition of the calculated (red) and crystallographically observed (blue) backbone conformations of the right-handed tetramer.

$= 1.5$  centiradians per residue, and  $\phi = 11^\circ$  for the model, and  $R_0 = 7.49 \text{ \AA}$ ,  $\omega_0 = 1.6$  centiradians per residue, and  $\phi = 10^\circ$  for the crystal structure. Thus, the experimentally observed structure for the RH4 peptide matches the designed structure in atomic detail.

**Practical implications.** Our results demonstrate that empirical protein design principles combined with computational protein engineering methods can be used to predict and design novel backbone structures, with root mean square coordinate errors that approach  $0.2 \text{ \AA}$ . This level of precision is encouraging with regard to the feasibility of designing protein catalysts, which may require the accurate positioning of reactive groups. Methods to engineer buried electrostatic interactions and to calculate accurately their energetic effects (23) should provide a second high-resolution design tool that is effectively orthogonal to engineered van der

Waals packing. Efforts to parameterize the backbones of more complex protein folds have begun (24), and it will be interesting to see the extent to which the parametric-backbone approach to protein design presented here can be generalized.

#### References and Notes

- W. F. DeGrado, *Adv. Protein Chem.* **39**, 51 (1988); M. H. Cordes, A. R. Davidson, R. T. Sauer, *Curr. Opin. Struct. Biol.* **6**, 3 (1996); R. Aurora, T. P. Creamer, R. Srinivasan, G. D. Rose, *J. Biol. Chem.* **272**, 1413 (1997); C. E. Schafmeister, S. L. LaPorte, L. J. Miercke, R. M. Stroud, *Nature Struct. Biol.* **4**, 1039 (1997).
- W. F. DeGrado, D. P. Raleigh, T. Handel, *Curr. Opin. Struct. Biol.* **1**, 984 (1991); S. F. Betz, D. P. Raleigh, W. F. DeGrado, *ibid.* **3**, 601 (1993); J. W. Bryson *et al.*, *Science* **270**, 935 (1995).
- C. P. Hill, D. H. Anderson, L. Wesson, W. F. DeGrado, D. Eisenberg, *Science* **249**, 543 (1990); B. Lovejoy *et al.*, *ibid.* **259**, 1288 (1993).
- K. E. Drexler, *Proc. Natl. Acad. Sci. U.S.A.* **78**, 5275 (1981); C. Pabo, *Nature* **301**, 200 (1983); J. W. Ponder and F. M. Richards, *J. Mol. Biol.* **193**, 775 (1987).
- P. E. Correa, *Proteins* **7**, 366 (1990); C. Lee and S. Subbiah, *J. Mol. Biol.* **217**, 373 (1991); L. Holm and C. Sander, *ibid.* **218**, 183 (1991); P. Tuffery, C. Etchebest, S. Hazout, R. Lavery, *J. Biomol. Struct. Dyn.* **8**, 1267 (1991); J. Desmet, M. D. Maeyer, B. Hazes, I. Lasters, *Nature* **356**, 539 (1992); P. Koehl and M. Delarue, *J. Mol. Biol.* **239**, 249 (1994); C. Lee, *ibid.* **236**, 918 (1994).
- J. R. Desjarlais and T. M. Handel, *Protein Sci.* **4**, 2006 (1995); G. A. Lazar, J. R. Desjarlais, T. M. Handel, *ibid.* **6**, 1167 (1997); B. I. Dahiyat and S. L. Mayo, *Science* **278**, 82 (1997); B. I. Dahiyat, C. A. Sarisky, S. L. Mayo, *J. Mol. Biol.* **273**, 789 (1997); S. M. Malakauskas and S. L. Mayo, *Nature Struct. Biol.* **5**, 470 (1998); J. W. Bryson, J. R. Desjarlais, T. M. Handel, W. F. DeGrado, *Protein Sci.* **7**, 1404 (1998); H. W. Hellinga, *Nature Struct. Biol.* **5**, 525 (1998).
- T. Alber *et al.*, *Science* **239**, 631 (1988); E. P. Baldwin, O. Hajiseyidjavad, W. A. Baase, B. W. Matthews, *ibid.* **262**, 1715 (1993).
- P. B. Harbury, B. Tidor, P. S. Kim, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 8408 (1995).
- The transmembrane helices of the erythrocyte protein glycophorin A have been shown to form a right-handed dimer in detergent micelles [K. R. MacKenzie, J. H. Prestegard, D. M. Engelman, *Science* **276**, 131 (1997)]. There are hints that other right-handed superhelical assemblies exist in nature. Repeating 11-fold motifs proposed to form right-handed coiled coils have been found in several naturally occurring proteins, including D-7 Lea proteins from cotton seeds and the tetrabrachion membrane surface protein of *Staphylothermus marinus* [L. Dure, *Plant J.* **3**, 363 (1993); J. Peters, H. Baumeister, A. Lupas, *J. Mol. Biol.* **257**, 1031 (1996)]. Helices M4 and M6 of SERCA1a, the  $\text{Ca}^{2+}$  adenosine triphosphatase of fast-twitch skeletal muscle sarcoplasmic reticulum, have been proposed on the basis of disulfide cross-linking data to form a right-handed coiled coil that lines the  $\text{Ca}^{2+}$  channel [W. J. Rice, N. M. Green, D. H. MacLennan, *J. Biol. Chem.* **272**, 31412 (1997)]. Our right-handed superhelical structures may provide useful models for studying this family of proteins.
- F. H. C. Crick, *Acta Crystallogr.* **6**, 689 (1953).
- For the nonnatural amino acids norvaline and allo-isoleucine, six possible rotamers ( $-$  and  $t$  for  $\chi_1$ ;  $-$ ,  $+$ , and  $t$  for  $\chi_2$ ) were tested. Notation:  $-$ ,  $-60^\circ$ ;  $+$ ,  $+60^\circ$ ;  $t$ ,  $180^\circ$ .
- Design calculations were performed on coiled-coil models containing alanine at positions b, c, f, g, and j;  $\alpha$ -amino butyric acid at positions e, i, and k; and all possible combinations of the residues alanine, valine, leucine, isoleucine, alioisoleucine and norvaline at positions a, d, and h. Optimal conformations were identified as described in (8) with the following modifications. First, the  $\alpha$  helix frequency in Crick's parameterization ( $\omega_1$ ) was set to  $6\pi/11$  radians per amino acid to ensure a regular undecad repeat. Second, the superhelix frequency ( $\omega_0$ ) of the initial backbone coordinates was set to 2 centiradians per amino acid and then explored in both the positive and negative directions until the potential minimum with respect to superhelix frequency had been located. Third, single-site enumerations (for example, with alanine at positions a and d and a variable residue at position h) were performed initially to identify good symmetrical single-level packing rotamers. The values of the a-position orientation angles ( $\phi$ ) were required to lie between  $-3^\circ$  and  $32^\circ$  for these calculations (for an undecad repeat, one full asymmetric unit in  $\phi$  covers  $360/11 = 32.7^\circ$ ; larger changes in  $\phi$  produce symmetry-related structures that differ only in the assignment of the undecad repeat). The 11 single-level packings with the largest calculated stabilities at each position (a, d, and h) were then combined in all possible combinations to construct undecad packings. Each of the six amino acids under consideration was represented at least once at each core position.
- F. H. C. Crick, *Acta Crystallogr.* **6**, 685 (1953).
- B. L. Tembe and J. A. McCammon, *Comput. Chem.* **8**, 281 (1984); D. L. Beveridge and F. M. DiCapua, *Annu. Rev. Biophys. Biophys. Chem.* **18**, 431 (1989).
- K. T. O'Neil and W. F. DeGrado, *Science* **250**, 646 (1990).
- Unfolding free energy differences ( $\Delta G$ ) between se-

- quences were calculated with the thermodynamic cycle illustrated in Fig. 2. An advantage of using a thermodynamic pathway is that it avoids double counting of contributions to stability. The  $\Delta G$ 's of permutation in the folded state, corresponding to step C, were approximated in three steps. First, transfer free energies of the native coiled coils from water to vacuum were estimated from the accessible hydrophobic surface areas of their lowest energy conformations [B. Lee and F. M. Richards, *J. Mol. Biol.* **55**, 379 (1971)], assuming an average transfer free energy of 0.012 kcal/(mol·Å<sup>2</sup>) (25). Second, free energies of permutation in vacuo were approximated as CHARMM19 (26) potential differences between the lowest energy packings of the native coiled coils and the lowest energy packings of the permuted coiled coils. All bonded potentials, the Lennard-Jones potential (switched off between 6.0 and 6.5 Å), and the explicit hydrogen-bond potential in the polar EF2 configuration [L. Nilsson and M. Karplus, *J. Comput. Chem.* **7**, 591 (1986)] were used. Finally, transfer free energies of the permuted coiled-coil sequences from vacuum to water were estimated from the accessible hydrophobic surface areas of their lowest energy conformations. Our approximation of relative stability derives from the assumptions that (i) translational, rotational, and electronic contributions to the classical partition function roughly cancel in the difference, (ii) the lowest potential energy-packing conformations dominate the partition sums, and (iii) the entropic contributions from the vibrational properties of the lowest potential energy-packing conformations roughly cancel in the difference.
17. O. D. Monera, C. M. Kay, R. S. Hodges, *Biochemistry* **33**, 3862 (1994).
  18. Peptides were synthesized and purified as described [D. J. Lockhart and P. S. Kim, *Science* **257**, 947 (1992)]. The identities of the peptides were confirmed by mass spectrometry (Perceptive Biosystems Voyager Elite), and all molecular masses were found to be within 1 dalton of the expected mass. The peptides have acetylated NH<sub>2</sub>-termini and amidated COOH-termini.
  19. CD spectra were measured on an Aviv 60DS spectropolarimeter. Measurements of  $[\theta]_{222}$  were made at 4°C in 50 mM phosphate (pH 7.0), 150 mM NaCl, and 10 μM peptide concentration. A  $[\theta]_{222}$  value of -35,000 degrees cm<sup>2</sup> dmol<sup>-1</sup> was taken to correspond to 100% helix [P. Luo and R. L. Baldwin, *Biochemistry* **36**, 8413 (1997)]. Thermal melts were performed in the same buffer and also, for RH4, with the addition of 3 M GdmCl to facilitate unfolding. Values of  $T_m$  were estimated from  $[\theta]_{222}$  versus temperature data (in 2° steps) by evaluating the maximum of  $d[\theta]/dT^{-1}$  [C. R. Cantor and P. R. Shimmel, *Biophysical Chemistry* (Freeman, New York, 1980), vol. 3, p. 1132]. Peptide concentrations were determined spectrophotometrically in 6 M GdmCl [H. Edelhoch, *Biochemistry* **6**, 1948 (1967)] with an extinction coefficient of 1280 cm<sup>-1</sup>M<sup>-1</sup> at 280 nm.
  20. Analytical ultracentrifugation measurements were carried out as described in (27). All experiments were performed in 50 mM sodium phosphate (pH 7.0) and 150 mM NaCl at 4°C. Data for RH3 and RH4 were collected at peptide concentrations of 20, 70, and 200 μM, and showed no systematic dependence of molecular weight on concentration. At concentrations below 100 μM, the molecular weight of the RH2 peptide varied systematically between monomer and dimer values. The data reported here for RH2 were therefore collected at 200 μM, 700 μM, and 2 mM peptide concentrations and show no systematic dependence of molecular weight on concentration. The molecular weights derived from the complete data sets (followed by expected molecular weight) and [the rotor speeds in thousand revolutions per minute at which data were collected] are as follows: RH2—8623 (7792) [30, 35]; RH3—13,087 (11,637) [32]; RH4—15,833 (15,516) [27, 32].
  21. One-dimensional <sup>1</sup>H-NMR spectra were recorded on a Bruker AMX 500 spectrometer as described [E. K. O'Shea, K. J. Lumb, P. S. Kim, *Curr. Biol.* **3**, 658 (1993)]. Measurements were made in D<sub>2</sub>O containing 50 mM phosphate (pH 7.15), 150 mM NaCl (the pH is uncorrected for the isotope effect), and 1 mM peptide. Spectra were normalized on the volumes of a tyrosine aromatic proton resonance, and peaks corresponding to the most slowly exchanging amide protons were integrated and fit to a single exponential in time. Half-lives of 161 min and 16,700 min were measured for the RH3 and RH4 peptides, respectively. Following [S. W. Englander, N. W. Downer, H. Teitelbaum, *Annu. Rev. Biochem.* **41**, 903 (1972)], poly D,L-alanine was calculated to have a half-life of  $8.3 \times 10^{-4}$  min under the conditions of the experiment. Thus the most slowly exchanging amide protons in the RH3 and RH4 peptides are protected by factors of  $2 \times 10^5$  and  $2 \times 10^7$ , respectively, relative to poly D,L-alanine. Although these protection factors are quite large by comparison with protection factors for most naturally occurring proteins, it is likely that they are smaller than would be predicted by a global unfolding mechanism. It is possible that for coiled coils in the extremely slow exchange regime, a local exchange mechanism dominates [see S. W. Englander, T. R. Sosnick, J. J. Englander, L. Mayne, *Curr. Opin. Struct. Biol.* **6**, 18 (1996) and references therein for a discussion of exchange mechanisms].
  22. The RH4 peptide was crystallized by vapor diffusion from 100 mM tris (pH 8.1), 0.3 M NaBr, 20% isopropanol, and 20% polyethylene glycol 1450. The crystals had P4<sub>2</sub>,2 symmetry (unit cell dimensions:  $a = b = 37.56$  Å,  $c = 53.38$  Å) with one monomer in the asymmetric unit. Data to 1.9 Å resolution were collected at room temperature with an R-Axis image plate detector (3175 unique reflections,  $R_{\text{merge}} = 3.4\%$ ;  $R_{\text{merge}} = \sum |I - \langle I \rangle| / \sum I$ , where  $I$  is the intensity of a reflection). The predicted structure, which included core side chains at positions a, d, and h, α-amino butyric acid at positions e, i, and k, and alanine elsewhere, was used in a combined rotation-translation search (with 6 to 2.9 Å data) along the crystallographic fourfold axis with the program X-PLOR [A. T. Brünger, J. Kuriyan, M. Karplus, *Science* **235**, 458 (1987)]. The coordinates corresponding to the peak solution (initial  $R = 49.6\%$ ) were annealed and rebuilt into simulated-annealing omit maps with the program O [T. A. Jones, J. Y. Zou, S. W. Cowan, M. Kjeldgaard, *Acta Crystallogr.* **A47**, 110 (1991)]. When the model reached an R-factor of ~22.0% ( $R$  factor =  $\sum |F_{\text{obs}} - F_{\text{calc}}| / F_{\text{obs}}$ , where  $F_{\text{obs}}$  and  $F_{\text{calc}}$  are the observed and calculated structure factors, respectively), annealing no longer resulted in improvement. Alternating cycles of manual rebuilding and Powell minimization led to a final R-factor of 20.4% and a free R-factor (calculated for 10% of the data not used for refinement) of 24.8% [A. T. Brünger, *Nature* **355**, 472 (1992)]. The final model, which includes 33 residues of peptide sequence (Glu<sup>2</sup>, Glu<sup>4</sup>, Lys<sup>15</sup>, and Glu<sup>24</sup> do not have clear side-chain density and are modeled as alanine), 8 water molecules, and 3 isopropanol molecules, exhibits root mean square deviations from ideal bond lengths and bond angles of 0.009 Å and 1.97°, respectively. All main-chain torsional angles fall within the helical region of a Ramachandran plot. The coordinates have been deposited in the Brookhaven Protein Data Bank (accession code #1RH4).
  23. B. Honig and A. Nicholls, *Science* **268**, 1144 (1995).
  24. A. G. Murzin, A. M. Lesk, C. Chothia, *J. Mol. Biol.* **236**, 1369 (1994).
  25. P. L. Privalov and G. I. Makhatadze, *J. Mol. Biol.* **232**, 660 (1993).
  26. B. R. Brooks et al., *J. Comput. Chem.* **4**, 187 (1983).
  27. P. B. Harbury, T. Zhang, P. S. Kim, T. Alber, *Science* **262**, 1401 (1993).
  28. Abbreviations for the amino acid residues are A, Ala; E, Glu; I, Ile; K, Lys; L, Leu; Q, Gln; Y, Tyr; nV, nor-Val; al, allo-Ile.
  29. We gratefully acknowledge M. Burgess for peptide synthesis, C. Lockshin, L. Stern, and A. Rich for use of their rotating anode and image-plate detector, J. Harris for assistance with the preparation of figures, A. Keating and R. Varadarajan for comments on the manuscript, and Z. Y. Peng, D. J. Lockhart, and members of the Kim laboratory for insightful discussions and criticism throughout the course of this work. This research was supported by NIH grants (GM55758 to B.T., GM48598 to T.A., and GM44162 to P.S.K.).

14 August 1998; accepted 14 October 1998

# POWERSURGE

**NEW! Science Online's Content Alert Service:** Knowledge is power. If you'd like more of both, there's only one source that delivers instant updates on breaking science news and research findings: *Science's* Content Alert Service. This free enhancement to your *Science* Online subscription delivers e-mail summaries of the latest news and research articles published weekly in *Science* – **instantly**. To sign up for the Content Alert service, go to *Science* Online – but make sure your surge protector is working first.

**Science**  
www.sciencemag.org

For more information about Content Alerts go to www.sciencemag.org. Click on Subscription button, then click on Content Alert button.



## High-Resolution Protein Design with Backbone Freedom

Pehr B. Harbury, Joseph J. Plecs, Bruce Tidor, Tom Alber and Peter S. Kim (November 20, 1998)

*Science* **282** (5393), 1462-1467. [doi: 10.1126/science.282.5393.1462]

Editor's Summary

---

This copy is for your personal, non-commercial use only.

---

- Article Tools** Visit the online version of this article to access the personalization and article tools:  
<http://science.sciencemag.org/content/282/5393/1462>
- Permissions** Obtain information about reproducing this article:  
<http://www.sciencemag.org/about/permissions.dtl>

*Science* (print ISSN 0036-8075; online ISSN 1095-9203) is published weekly, except the last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. Copyright 2016 by the American Association for the Advancement of Science; all rights reserved. The title *Science* is a registered trademark of AAAS.