

Published in final edited form as:

*Science*. 2014 October 24; 346(6208): 481–485. doi:10.1126/science.1257481.

## High thermodynamic stability of parametrically designed helical bundles

Po-Ssu Huang<sup>#1,2</sup>, Gustav Oberdorfer<sup>#1,2,3</sup>, Chunfu Xu<sup>#1,2</sup>, Xue Y. Pei<sup>4</sup>, Brent L. Nannenga<sup>5</sup>, Joseph M. Rogers<sup>6,†</sup>, Frank DiMaio<sup>1,2</sup>, Tamir Gonen<sup>5</sup>, Ben Luisi<sup>4</sup>, and David Baker<sup>#1,2,7,‡</sup>

<sup>1</sup>Department of Biochemistry, University of Washington, Seattle, WA 98195, USA.

<sup>2</sup>Institute for Protein Design, University of Washington, Seattle, WA 98195, USA.

<sup>3</sup>Institute of Molecular Biosciences, University of Graz, Humboldtstrasse 50/3, 8010-Graz, Austria.

<sup>4</sup>Department of Biochemistry, University of Cambridge, Cambridge CB2 1GA, UK.

<sup>5</sup>Janelia Research Campus, Howard Hughes Medical Institute, Ashburn, VA 20147, USA.

<sup>6</sup>Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, UK.

<sup>7</sup>Howard Hughes Medical Institute, University of Washington, Seattle, WA 98195, USA.

# These authors contributed equally to this work.

### Abstract

We describe a procedure for designing proteins with backbones produced by varying the parameters in the Crick coiled coil–generating equations. Combinatorial design calculations identify low-energy sequences for alternative helix supercoil arrangements, and the helices in the lowest-energy arrangements are connected by loop building. We design an antiparallel monomeric untwisted three-helix bundle with 80-residue helices, an antiparallel monomeric right-handed four-helix bundle, and a pentameric parallel left-handed five-helix bundle. The designed proteins are extremely stable (extrapolated  $\Delta G_{\text{fold}} > 60$  kilocalories per mole), and their crystal structures are close to those of the design models with nearly identical core packing between the helices. The approach enables the custom design of hyperstable proteins with fine-tuned geometries for a wide range of applications.

‡Corresponding author. dabaker@u.washington.edu.

†Present address: Department of Chemistry, University of Tokyo, Graduate School of Science, 7-3-1, Hongo, Bunkyo, Tokyo 113-0033, Japan.

#### SUPPLEMENTARY MATERIALS

[www.sciencemag.org/content/346/6208/481/suppl/DC1](http://www.sciencemag.org/content/346/6208/481/suppl/DC1)

Computational Modeling

Materials and Methods

Figs. S1 to S12

Tables S1 and S2

Input Files and Command Lines for Computations

References (20–32)

Coiled coils consisting of two or more helices supercoiled around a central axis play important roles in biology, and their simplicity and regularity have inspired peptide-design efforts (1–4). Most studies have used sequence-based approaches, focusing on choosing optimal amino acids at core positions of the coiled-coil heptad repeat (5–7). The few structure-based efforts have used parametric equations first derived by Francis Crick (2) to design peptides that form right-handed coiled coils (8) or bind carbon nanotubes (9). Here we combine parametric backbone generation with the Rosetta protein-design methodology (10) to generate more complex and stable protein structures.

The Crick coiled-coil equation parameters for a bundle of  $n$  helices are  $\omega_0$ , the supercoil twist;  $\omega_1$ , the  $\alpha$ -helical twist;  $R_0$ , the supercoil radius;  $\phi_1, \phi_2, \dots, \phi_n$ , the phases of the individual helices; and  $z_2, \dots, z_n$ , their offsets along the superhelical axis relative to the first helix (2, 11, 12). As shown in the supplementary materials (12), successive C $\alpha$  atoms rotate about the  $\alpha$ -helical axis by  $\sim(\omega_0 + \omega_1)$ , and the protein backbone is strained when this sum deviates from the value of  $100^\circ$  found in ideal helices (which have  $\omega_0 = 0^\circ$  and  $\omega_1 = 100^\circ$ ) (fig. S1). Hence, supercoil ( $\omega_0$ ) and helical ( $\omega_1$ ) twist are coupled (fig. S1).

Repeating backbone geometries are good targets for design because there are fewer distinct side-chain packing problems to be solved. There are three repeating geometries that require deviation of less than  $3^\circ$  from an ideal unstrained helix. First, if  $\omega_1$  is increased to  $102.85^\circ$  from the ideal value of  $100.0^\circ$ , after seven residues the helix has completed two full turns ( $720^\circ$ ). Second, if  $\omega_1$  is reduced to  $98.2^\circ$ , after 11 residues the helix has completed three full turns ( $1080^\circ$ ) (8). Third, if  $\omega_1$  is kept at exactly  $100^\circ$ , after 18 residues the helix has completed five full turns ( $1800^\circ$ ). We refer to these three cases as two-layer, three-layer, and five-layer designs, respectively, corresponding to the number of distinct helix-helix-interacting layers that must be designed. Because of the coupling between  $\omega_0$  and  $\omega_1$ , two-layer designs are left-handed ( $\omega_0$  negative), three-layer designs are right-handed ( $\omega_0$  positive), and five-layer designs are untwisted ( $\omega_0$  close to zero) (3).

We explored the design of helix bundles with two-layer, three-layer, or five-layer geometries and different numbers of helices surrounding the supercoil axis. Once the number of helices in the bundle and the layer type were chosen, the Crick equation parameters were sampled on a grid, backbone conformations were generated, and Rosetta sequence design calculations were carried out. Finer grid searches were undertaken in the vicinity of the parameter sets yielding the lowest-energy designs. For the monomeric designs, the helices of the lowest-energy backbone solutions were connected using Rosetta loop modeling (13). Rosetta structure prediction calculations were used to investigate the extent to which the final designed sequences encode the desired structure (14); if the lowest-energy structures were similar to the design models, the designs were synthesized and experimentally characterized.

We designed antiparallel three-helix bundles with 80-residue helices and an 18-residue repeat unit ( $\omega_1 = 100^\circ$ ). Because a monomeric three-helix bundle contains both parallel and antiparallel helix interactions, we treated each of the three helices independently in the design calculations. Hence, there are seven degrees of freedom: the supercoil twist and radius, the phases of each of the three helices, and the displacements along the supercoil axis

of the second and third helices relative to the first. Successive grid searches yielded well-packed low-energy models. Following connection of the helices by loop modeling, the lowest-energy structures found in Rosetta@Home structure prediction calculations for the designed sequences had core packing arrangements very similar to those of the design models in the center of the bundle, with small deviations near the turns (fig. S2). Three designs—3H5L\_1, 3H5L\_2, and 3H5L\_4—of four tested were expressed and soluble at high levels in *Escherichia coli* and readily purified. All three proteins had helical circular dichroism (CD) spectra consistent with the design and were stable to thermal denaturation up to 95°C (fig. S3A), and negative-stain electron microscopy showed rodlike shapes with lengths (~12 nm) expected for 80-residue helices (Fig. 1C and fig. S3B).

More detailed thermodynamic characterization showed that 3H5L\_2 was exceptionally stable with a denaturation midpoint of 7.5 M guanidinium chloride (GdmCl) at 25°C and 7 M at 80°C (Fig. 1A). Fitting of a two-state model (15) yielded a  $\Delta G_{D-N}$  in the absence of denaturant of  $61 \pm 5$  kcal mol<sup>-1</sup> at 25°C (fig. S4). Because of the long extrapolation, sharp unfolding transition, and the limited unfolded protein CD baseline, the error in  $\Delta G_{D-N}$  may be significantly larger, but the fit *m*-value ( $m_{D-N} = 8.1 \pm 0.7$  kcal mol<sup>-1</sup> M<sup>-1</sup>; 25 °C) is that expected for the size of the protein (16). Even at 7.75 M GdmCl, 3H5L\_2 unfolded very slowly ( $k_{\text{unfold}} = 7.9 \pm 0.3 \times 10^{-5}$  s<sup>-1</sup> at 25°C) (Fig. 1B).

The 2.8 Å crystal structure of 3H5L\_2 (table S2) has the same topology as that of the design model (Fig. 1D) but less superhelical twist; the release of helical strain evidently outweighs the slightly improved packing in the design. Despite this untwisting, the core 18-residue repeat unit is nearly identical in the crystal structure and design [all-atom root mean square deviation (RMSD) 1.1 Å]. Figure 1E shows superpositions of the design and crystal structure for each of the five distinct core packing layers; in each layer, there is tight and complementary side-chain packing, with close agreement between the crystal structure and design model and between different repeats. In several of the layers, close complementary packing of methionine residues identified in the Rosetta combinatorial side-chain packing calculations differs from previously described helix packing motifs. The complexity of the design and hence the necessity for structure-based computer calculations rather than sequence-based rules is highlighted by comparison to classical parallel two-layer (heptad repeat) bundle designs: Whereas the latter have seven unique positions (heptad repeat positions a, b, c, d, e, f, g), every repeat of 3H5L\_2 is made up by three unique helix segments each with 18 unique positions, a total of 54 unique positions that must be designed. Further increasing the complexity, each layer involves packing between residues from two parallel helices and one antiparallel helix.

For a second test of the approach, we designed a three-layer connected four-helix bundle with helices 2 and 4 antiparallel to helices 1 and 3. Because of the relaxation of the supercoil twist ( $\omega_0$ ) to a value close to 0°—the ideal value for a five-layer bundle—observed in the crystal structure of 3H5L\_2, we fixed  $\omega_0$  at the ideal value given the layer type in subsequent designs. Thus, for the three-layer bundle, the helix twist  $\omega_1$  was set to 98.2° and the supercoil twist  $\omega_0$  to 1.8°. To reduce the size of the search space, we restricted sampling to C2 symmetric conformations in which helices 3 and 4 are related to helices 1 and 2 by a twofold rotation around the *z* axis—the helical phases and offsets for helices 3 and 4 are

then identical to those for helices 1 and 2. Iterative grid searches were carried out over the remaining four parameters (the supercoil radius, the phases of helices 1 and 2, and the  $z$  offset of helix 2). Symmetry between the first two and second two helices was maintained at both the sequence and structure level.

Genes were synthesized for three low-energy designs (4H3L\_1 to 4H3L\_3) that converged on the designed target structure in Rosetta structure prediction calculations (fig. S2). One of the proteins, 4H3L\_3, was solubly expressed as a monomer (fig. S12) at high levels in *E. coli*, had the expected  $\alpha$ -helical CD spectrum (Fig. 2A), and was stable to thermal denaturation with almost identical CD spectra at 25° and 95°C (Fig. 2A). No melting transition was observed by differential scanning calorimetry (DSC) at temperatures up to 130°C (fig. S5). The stability to chemical denaturation was even higher than for 3H5L\_2: Little or no unfolding was observed in 7.3 M GdmCl up to 130°C (Fig. 2B and fig. S5). In 5 M guanidinium thiocyanate (GdmSCN)—a stronger denaturant than GdmCl—the melting temperature is 97°C (Fig. 2C and fig. S5).

The 1.6 Å structure of 4H3L\_3 (table S2) is similar to that of the design model (Fig. 2D) with the predicted right-handed supercoil twist and the 11-residue three-layer repeat geometry. The core packing within individual repeats is virtually identical in the crystal structure and design model with an all-atom RMSD of 0.7 Å over the core repeating units. Superpositions of the side chains in the crystal structure and design model for each of the three unique layers are shown in Fig. 2E. The close and complementary side-chain packing arrangements at each layer are distinct, and the third layer again uses methionine residues.

An advantage of the repeat structure of the parametrically designed bundles is that their length can be readily controlled by varying the number of repeats. 3H5L\_2\_mini with one 18-residue repeat and 4H3L\_3\_mini with two 11-residue repeats had CD spectra identical to those of the full-length proteins and were stable for their size (fig. S7).

Both the 3H5L\_2 and 4H3L\_3 structures deviate from perfect supercoil geometry (Figs. 1D and 2D), and it is likely that the lowest-energy structures for monomeric antiparallel bundles more generally will not be confined to the space spanned by the Crick parameterization near the turns. Rosetta de novo structure prediction calculations are not confined to this space, and for both 3H5L\_2 and 4H3L\_3 the crystal structures are closer to the lowest-energy predicted structures than to the design models (fig. S2 legend). Hence, a final round of sequence optimization based on lowest-energy predicted structures could increase the accuracy of the design process.

As a third test, we designed parallel five-helix bundles with two-layer geometry ( $\omega_0 = 102.85^\circ$ ). In contrast to the three- and four-helix bundles, which are connected single-chain structures, the five-helix bundles consist of five copies of a single helical peptide arranged with fivefold cyclic symmetry ( $C_5$ ). With the  $C_5$  symmetry, the only degrees of freedom are  $R_0$ ,  $\omega_0$ , and  $\phi_1$  and hence the parameter space could be scanned in great detail. The energy landscapes following Rosetta sequence design have clear optima at  $R_0 = 8.7$  Å and  $\phi_1 = 43^\circ$  (fig. S8, A to C). In a five-helix bundle, each helix has two interaction surfaces at  $108^\circ$  from

each other; with this solution for  $\phi 1$ , both interfaces have close to optimal packing geometry (fig. S9).

The lowest-energy designs were tested *in silico* in docking calculations. The lowest-energy C5 arrangement sampled was nearly identical to that of the design model, and all the C4 and C6 arrangements had higher energies (fig. S8D). The designed interface also had lower energy than any other interface identified between two monomers in asymmetric docking calculations (fig. S8E). One of two experimentally tested designs, 5H2L\_2, was readily soluble in aqueous buffer and was found by CD to have a helical structure (Fig. 3A). 5H2L\_2 is stable at 95°C (Fig. 3B) and up to 4 M GdmCl (fig. S10) and sediments as a pentamer in analytical ultracentrifugation experiments (Fig. 3C).

The 1.7 Å crystal structure of 5H2L\_2 with a surface substitution to promote crystal growth (5H2L\_2.1 in table S1) is nearly identical to that of the design model (0.4 Å all-atom RMSD; Fig. 3D). The two unique core packing layers are shown in Fig. 3E. For the layer indicated in red in Fig. 3D, two distinct packing solutions were found; one involving a hydrogen bond network and the other aliphatic packing. Both are closely recapitulated in the crystal structure. The combination of Leu at the heptad a position and Gln or Ile at the d position is very well packed in the pentamer, and the docking calculations suggest that these residues are not as compatible with other oligomerization states. In the crystal lattice, the helices pack end to end forming long crossing helical tubes, suggesting a route to nanowire design (Fig. 3F).

The stability to chemical denaturation of 3H5L\_2 and 4H3L\_3 stands out from those of the proteins collected in the ProTherm database (17) (Fig. 4 and fig. S11). This is notable given that the sequences and structures of the designs came directly from Rosetta calculations with no human modification or experimental optimization. That hyperstability is relatively easy to achieve by design (two out of nine designs tested), but very rarely observed [an example is described in (18)] for naturally occurring proteins, highlights the extent to which function trumped stability during natural evolution. Efforts to design new protein functions will likely move from repurposing native scaffolds to *de novo* design of hyperstable backbones with geometries optimal for the desired function.

Low-energy structures must have unstrained backbone conformations and complementary side-chain packing. The left-handed superhelical twist of the heptad repeat is traditionally attributed to “knobs into holes” side-chain packing; our approach highlights the less appreciated contribution of backbone strain: The left-handed supercoil compensates for the strain introduced by overtwisting the  $\alpha$  helix to achieve two full turns with seven residues. The combination of parametric generation of unstrained backbones and Rosetta combinatorial side-chain optimization should be extendible to the design of other classes of structures (19). The ability to readily generate hyperstable proteins with finely tuned geometries without relying on known sequence motifs should contribute to the next generation of designed protein-based nanostructures, therapeutics, and catalysts.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

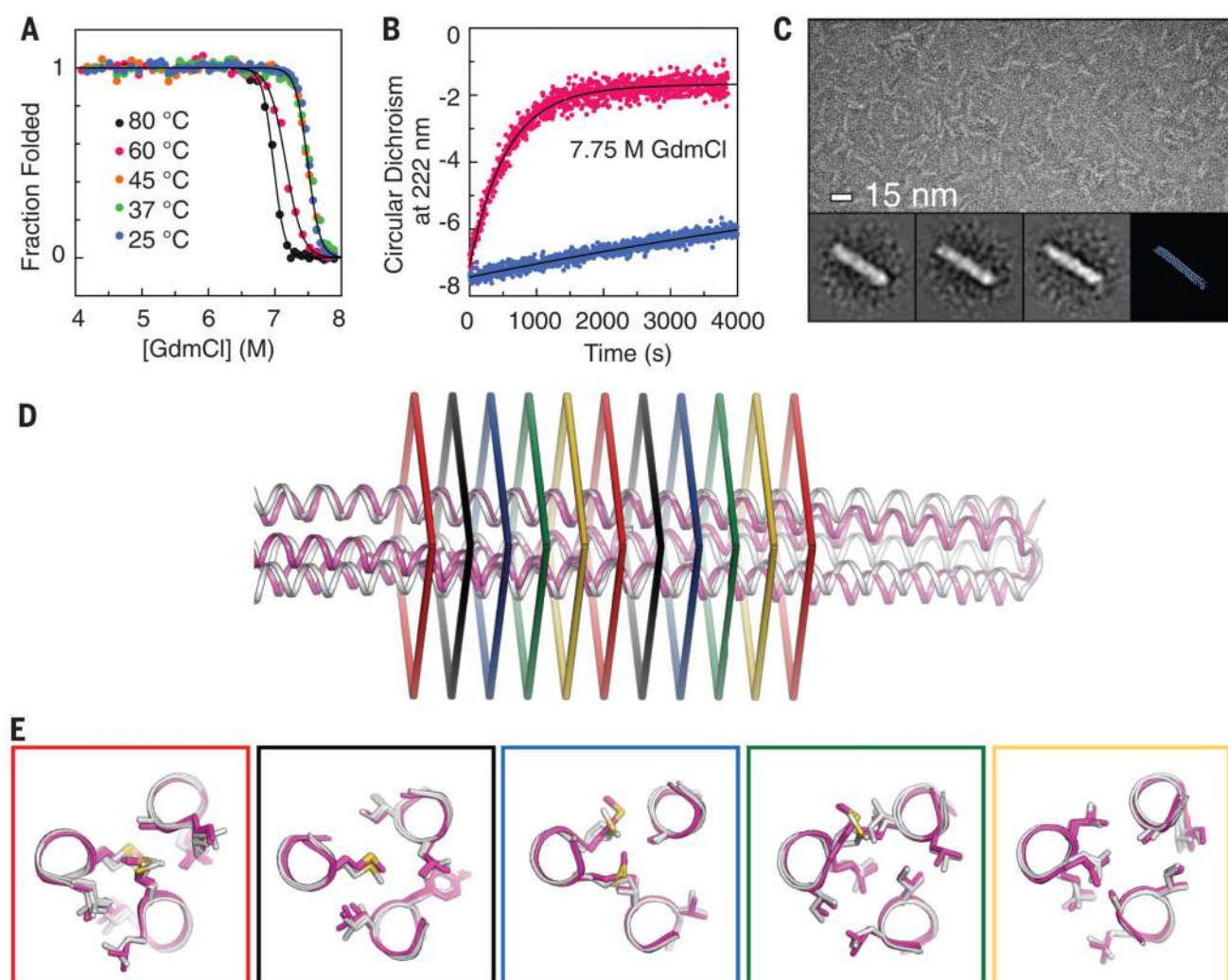
## ACKNOWLEDGMENTS

We are indebted to G. Grigoryan and W. DeGrado for their paper on coiled-coil geometry and to G. Grigoryan for the CCCP Web server that was used in initial exploratory calculations, and for helpful advice. We thank S. Gordon for protein production, Rosetta@Home volunteers for providing the computing throughput to rigorously test designs by ab initio structure prediction, and the staff at Diamond Light source and at the Advanced Light Source (U.S. Department of Energy contract no. DE-AC02-05CH11231) for access and assistance with x-ray data collection. This work was supported by Howard Hughes Medical Institute, Defense Threat Reduction Agency, and the Wellcome Trust. G.O. is a Marie Curie International Outgoing Fellowship fellow (332094 ASR-CompEnzDes FP7-People-2012-IOF). J.M.R. was supported by a studentship from the Biotechnology and Biological Sciences Research Council. Coordinates and structure factors have been deposited in the Protein Data Bank with the accession codes 4TQL (3H5L\_2), 4UOS (4H3L\_3), and 4UOT (5L2L\_2.1). P.S.H., G.O., C.X., and D.B. designed the research. D.B. wrote the parametric backbone generation code; P.S.H. wrote the loop modeling code; and P.S.H., G.O., C.X., and D.B. carried out design calculations and biophysical analyses. J.M.R. characterized 3H5L\_2 and 3H5L\_2\_mini. With help from G.O. and F.D., X.Y.P. and B.L. solved structures of 4H3L\_3 and 5H2L\_2, and B.L.N. and T.G. solved the 3H5L\_2 bundle structure and did electron microscopy analysis.

## REFERENCES AND NOTES

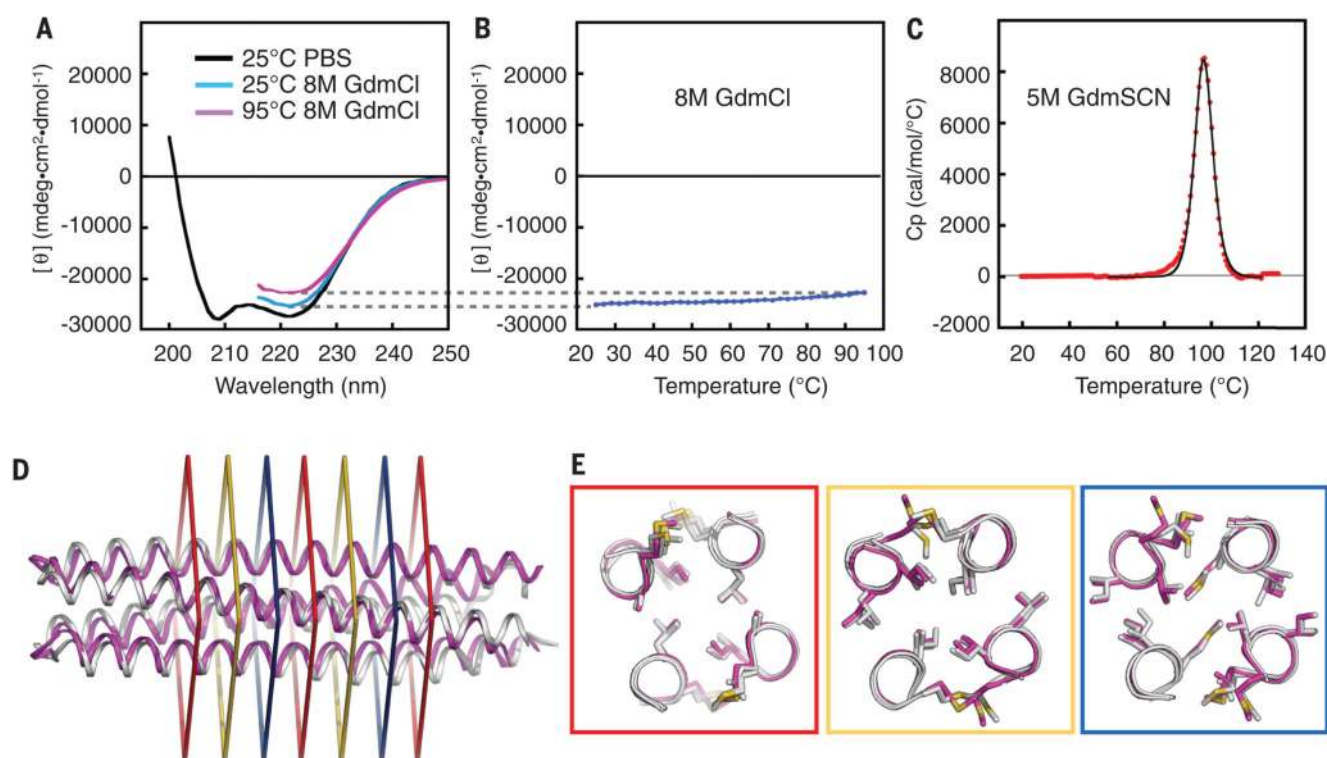
1. Boyle AL, Woolfson DN. *Chem. Soc. Rev.* 2011; 40:4295–4306. [PubMed: 21373694]
2. Crick FHC. *Acta Crystallogr.* 1953; 6:685–689.
3. Gruber M, Lupas AN. *Trends Biochem. Sci.* 2003; 28:679–685. [PubMed: 14659700]
4. Hill RB, Raleigh DP, Lombardi A, DeGrado WF. *Acc. Chem. Res.* 2000; 33:745–754. [PubMed: 11087311]
5. Harbury PB, Zhang T, Kim PS, Alber T. *Science.* 1993; 262:1401–1407. [PubMed: 8248779]
6. Liu J, et al. *Proc. Natl. Acad. Sci. U.S.A.* 2006; 103:15457–15462. [PubMed: 17030805]
7. Zaccai NR, et al. *Nat. Chem. Biol.* 2011; 7:935–941. [PubMed: 22037471]
8. Harbury PB, Plecs JJ, Tidor B, Alber T, Kim PS. *Science.* 1998; 282:1462–1467. [PubMed: 9822371]
9. Grigoryan G, et al. *Science.* 2011; 332:1071–1076. [PubMed: 21617073]
10. Das R, Baker D. *Annu. Rev. Biochem.* 2008; 77:363–382. [PubMed: 18410248]
11. Grigoryan G, DeGrado WF. *J. Mol. Biol.* 2011; 405:1079–1100. [PubMed: 20932976]
12. Materials and methods are available as supplementary materials on *Science* Online.
13. Huang PS, et al. *PLOS ONE.* 2011; 6:e24109. [PubMed: 21909381]
14. Bradley P, Misura KMS, Baker D. *Science.* 2005; 309:1868–1871. [PubMed: 16166519]
15. Pace CN. *Methods Enzymol.* 1986; 131:266–280. [PubMed: 3773761]
16. Geierhaas CD, Nickson AA, Lindorff-Larsen K, Clarke J, Vendruscolo M. *Protein Sci.* 2007; 16:125–134. [PubMed: 17123959]
17. Kumar MD, et al. *Nucleic Acids Res.* 2006; 34:D204–D206. [PubMed: 16381846]
18. Peters J, Baumeister W, Lupas A. *J. Mol. Biol.* 1996; 257:1031–1041. [PubMed: 8632466]
19. Chou KC, Caracci L, Maggiora GM. *J. Mol. Biol.* 1990; 213:315–326. [PubMed: 2342109]





**Fig. 1. Stability and structure of designed monomeric three-helix bundle 3H5L\_2**

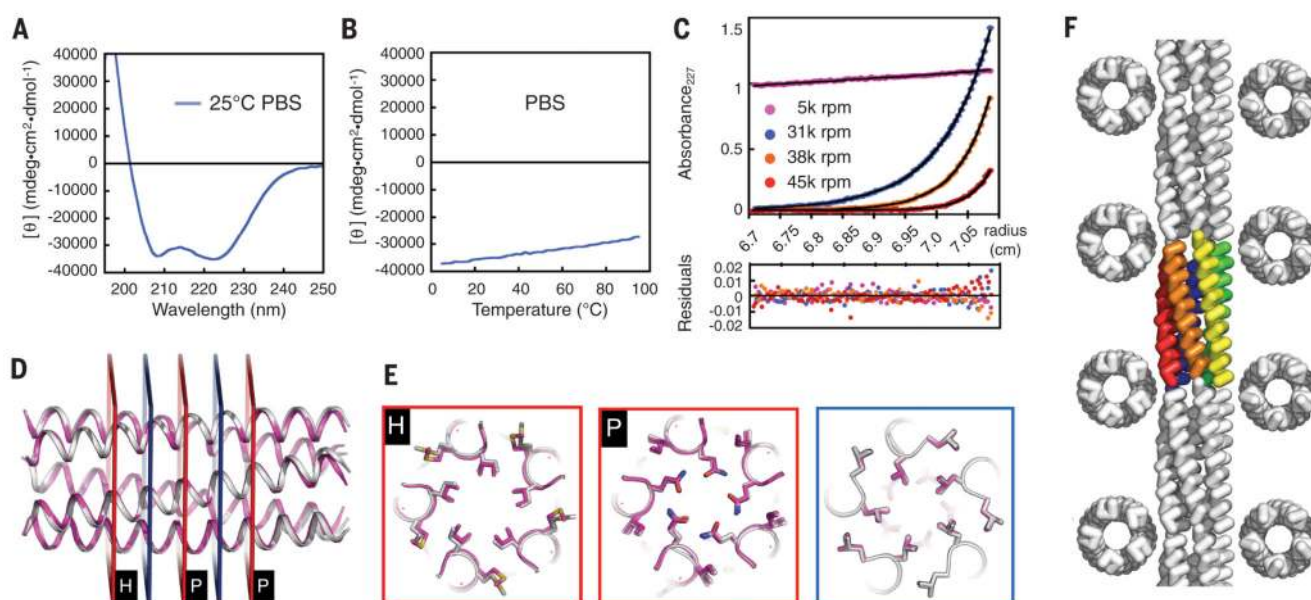
(A) GdmCl denaturation monitored by CD. At 80°C, the midpoint of the folding transition is ~7 M GdmCl. (B) Kinetics of unfolding in 7.75 M GdmCl at 25°C (blue) and 60°C (red). (C) Negative-stain electron micrographs of 3H5L\_2; particle averages are in the inset. The rods are ~12 nm in length, consistent with the 3H5L\_2 design model. (D) Superposition of 3H5L\_2 crystal structure and design model (RMSD = 3.1 Å over all Cα atoms). Colored rectangles represent the five distinct packing layers in the 18-residue repeat of the structure. (E) Side-chain packing arrangements in each of the five unique layers. Magenta, design model; gray, crystal structure. For each layer, the very similar solutions found by Rosetta in the two central 18-residue repeats are shown.



**Fig. 2. Stability and structure of designed monomeric four-helix bundle 4H3L\_3**

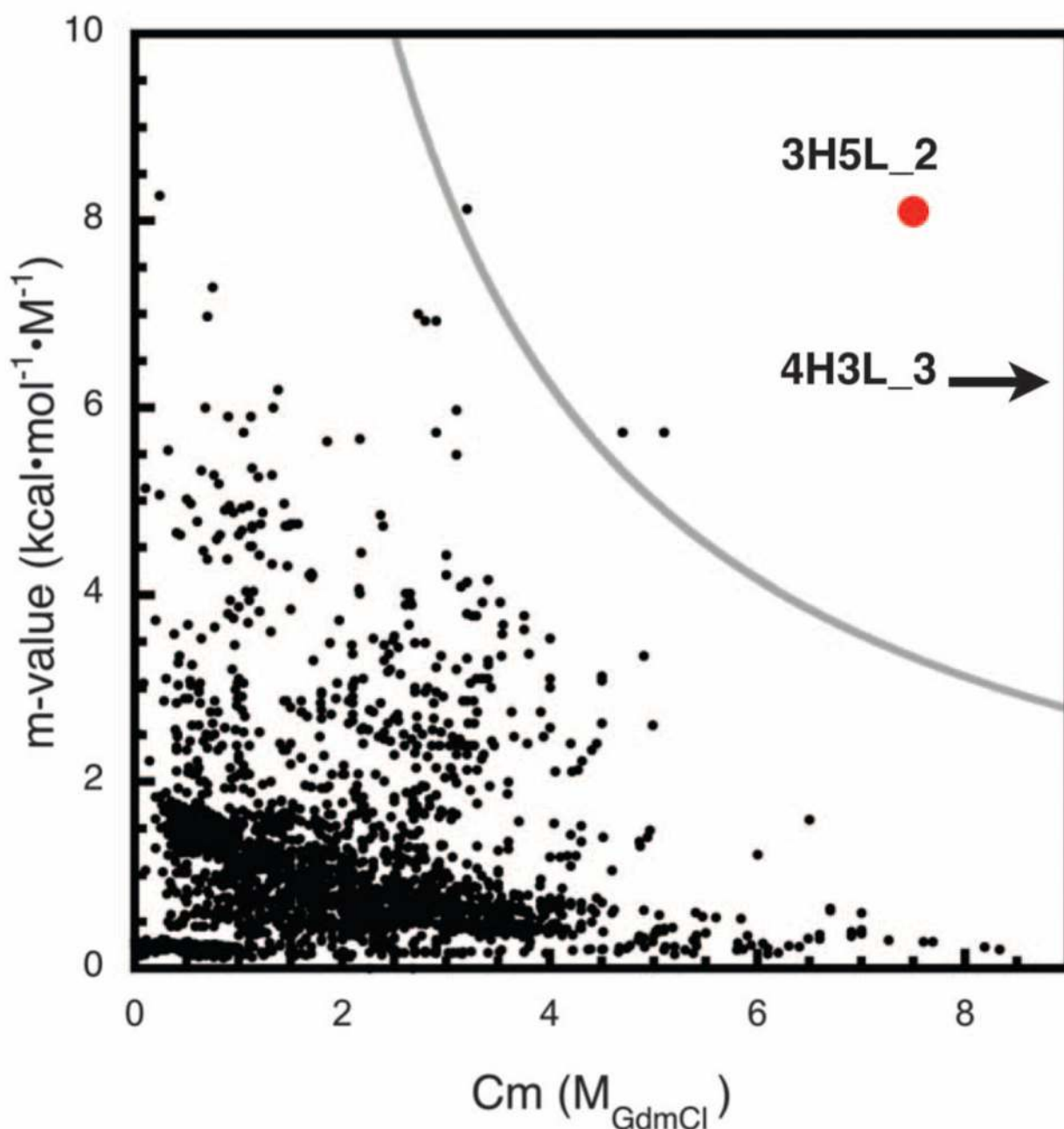
(A) CD spectra of 4H3L\_3 in the presence and absence of GdmCl. (B) Temperature dependence of CD signal at 222 nm in 8 M GdmCl. No unfolding transition is observed at temperatures up to 95°C. (C) DSC of 4H3L\_3 in 5 M GdmSCN. An endothermic transition is observed at 97°C ( $\Delta H = 95$  kcal/mol). No transition is observed at temperatures up to 130°C in GdmCl or phosphate-buffered saline (PBS) (fig. S5). (D) Superposition of 4H3L\_3 crystal structure and design model. At points where the crystal structure deviates from the design model and the helical axis changes direction, peptide backbone carbonyl groups are tipped outward toward the bulk solvent, where they contribute to entrained hydration networks (fig. S6). Colored rectangles indicate the three distinct layers in the 11-residue repeat of the protein. (E) Superposition of 4H3L\_3 crystal structure and design model for each of the three unique packing layers for both of the central repeats. Magenta, design model; gray, crystal structure.





**Fig. 3. Stability and structure of designed pentameric five-helix bundle 5H2L\_2**

(A) CD spectrum and (B) CD-monitored temperature melt of 5H2L\_2 (0.2 mg/ml in PBS, pH 7.4). (C) Representative analytical ultra-centrifugation sedimentation-equilibrium curves at four different rotor speeds for 5H2L\_2 0.5 mg/ml in PBS, pH 7.4. The data fit (black lines) to a single ideal species in solution corresponding to the pentameric complex of 5H2L\_2. (D) Superposition of backbone of crystal structure and design model. The all-atom RMSD between computational model and experimental structure is 0.4 Å. (E) Comparison of side-chain packing in crystal structure (gray) and design model (magenta) at the two unique layers in the 5H2L\_2 structure. Two solutions were found for the red layer—a simple aliphatic packing (H) and a polar hydrogen bonding network (P)—and are shown in the two red panels. Both computed solutions were accurately recapitulated in the crystal structure. (F) Packing of the pentamers into straight filaments in the crystal. The colored pentamers occupy one asymmetric unit of the crystal, and the gray pentamers are from adjacent units.



**Fig. 4. High thermodynamic stability of 3H5L\_2 and 4H3L\_3**

X axis, GdmCl denaturation midpoint ( $C_m$ ); y axis, dependence of folding free energy on GdmCl concentration ( $m$  value); black dots, data on previously described proteins from ProTherm database (17); red circle, 3H5L\_2; black arrow, lower bound for 4H3L\_3  $C_m$ . The free energy of folding in the absence of denaturant is the product of the  $m$ -value and the  $C_m$ ; the curve  $m\text{-value} \times C_m = 25$  kcal/mol (gray) separates almost all native proteins from the two designs. 4H3L\_3 does not denature in GdmCl.