

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

If you wish to distribute this article to others, you can order high-quality copies for your colleagues, clients, or customers by [clicking here](#).

Permission to republish or repurpose articles or portions of articles can be obtained by following the guidelines [here](#).

The following resources related to this article are available online at www.sciencemag.org (this information is current as of October 24, 2014):

Updated information and services, including high-resolution figures, can be found in the online version of this article at:

<http://www.sciencemag.org/content/346/6208/481.full.html>

Supporting Online Material can be found at:

<http://www.sciencemag.org/content/suppl/2014/10/22/346.6208.481.DC1.html>

A list of selected additional articles on the Science Web sites **related to this article** can be found at:

<http://www.sciencemag.org/content/346/6208/481.full.html#related>

This article **cites 31 articles**, 6 of which can be accessed free:

<http://www.sciencemag.org/content/346/6208/481.full.html#ref-list-1>

This article appears in the following **subject collections**:

Biochemistry

<http://www.sciencemag.org/cgi/collection/biochem>

CMG complex disassembly is not driven through Cdt1 but by direct removal of ubiquitylated helicase from chromatin. Moreover, data obtained from budding yeast presented in the accompanying paper by Maric *et al.* (28) indicate that this mechanism of replisome disassembly at the termination of replication is conserved throughout evolution. Polyubiquitylation of Mcm7 has also been observed in human embryonic kidney 293 cells, although a function for it has not yet been reported (29).

REFERENCES AND NOTES

1. J. Baxter, J. F. Diffley, *Mol. Cell* **30**, 790–802 (2008).
2. O. Cuvier, S. Stanojic, J. M. Lemaître, M. Mechali, *Genes Dev.* **22**, 860–865 (2008).
3. D. Fachinetti *et al.*, *Mol. Cell* **39**, 595–605 (2010).
4. C. J. Rudolph, A. L. Upton, A. Stockum, C. A. Nieduszynski, R. G. Lloyd, *Nature* **500**, 608–611 (2013).
5. K. Labib, J. A. Tercero, J. F. Diffley, *Science* **288**, 1643–1647 (2000).
6. D. Remus *et al.*, *Cell* **139**, 719–730 (2009).
7. C. Evin *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 20240–20245 (2009).
8. A. Gambus, G. A. Khoudoli, R. C. Jones, J. J. Blow, *J. Biol. Chem.* **286**, 11855–11864 (2011).
9. M. A. Kuipers *et al.*, *J. Cell Biol.* **192**, 29–41 (2011).
10. I. Ilves, T. Petojevic, J. J. Pesavento, M. R. Botchan, *Mol. Cell* **37**, 247–258 (2010).
11. A. Gambus *et al.*, *Nat. Cell Biol.* **8**, 358–366 (2006).
12. A. Nishiyama, L. Frappier, M. Méchali, *Genes Dev.* **25**, 165–175 (2011).
13. M. Jagannathan *et al.*, *Mol. Cell Biol.* **34**, 132–145 (2014).
14. D. T. Mahaffey, C. Gorbea, M. Rechsteiner, *Exp. Cell Res.* **288**, 225–234 (2003).
15. D. Komander, M. Rape, *Annu. Rev. Biochem.* **81**, 203–229 (2012).
16. J. Adams, M. Kauffman, *Cancer Invest.* **22**, 304–311 (2004).
17. J. Jin, E. E. Arias, J. Chen, J. W. Harper, J. C. Walter, *Mol. Cell* **23**, 709–721 (2006).
18. A. Li, J. J. Blow, *EMBO J.* **24**, 395–404 (2005).
19. A. M. Woodward *et al.*, *J. Cell Biol.* **173**, 673–683 (2006).
20. D. Remus, J. F. Diffley, *Curr. Opin. Cell Biol.* **21**, 771–777 (2009).
21. T. Germe, O. Hyrien, *EMBO Rep.* **6**, 729–735 (2005).
22. T. A. Soucy *et al.*, *Nature* **458**, 732–736 (2009).
23. N. P. Dantuma, T. Hoppe, *Trends Cell Biol.* **22**, 483–491 (2012).
24. S. Heubes, O. Stemmann, *J. Cell Sci.* **120**, 1325–1329 (2007).
25. A. Franz *et al.*, *Mol. Cell* **44**, 85–96 (2011).
26. M. Raman, C. G. Havens, J. C. Walter, J. W. Harper, *Mol. Cell* **44**, 72–84 (2011).
27. J. Mouysset *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 12879–12884 (2008).
28. M. Maric, T. Maculinski, G. De Piccoli, K. Labib, *Science* **346**, 1253596 (2014); DOI: 10.1126/science.1253596.
29. C. Kühne, L. Banks, *J. Biol. Chem.* **273**, 34302–34309 (1998).

ACKNOWLEDGMENTS

We would like to thank K. Labib for sharing his data and critical discussions, J. Blow for support and help with egg extract, R. Hay for advice, O. Stemmann for vectors to express p97, and E. Petermann for help with fibers. A.G. would like to thank D. Tennant for support. This work was supported by U.K. Medical Research Council Career Development Award MR/K007106/1 to A.G. and Cancer Research UK Birmingham Centre award DF/270312 to A.G. S.P.M. was funded by ERASMUS scholarship and Wellcome Trust strategic award ISSFP47; N.C. was funded by Cancer Research UK Birmingham Centre award CAICS/05/12.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/346/6208/477/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S8
References (30–40)

19 March 2014; accepted 26 August 2014
10.1126/science.1253585

PROTEIN DESIGN

High thermodynamic stability of parametrically designed helical bundles

Po-Ssu Huang,^{1,2*} Gustav Oberdorfer,^{1,2,3*} Chunfu Xu,^{1,2*} Xue Y. Pei,⁴ Brent L. Nannenga,⁵ Joseph M. Rogers,^{6,†} Frank DiMaio,^{1,2} Tamir Gonen,⁵ Ben Luisi,⁴ David Baker^{1,2,7*†}

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.

Coiled coils consisting of two or more α 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 ω_0 , the supercoil twist; ω_1 , the α -helical twist; R_0 , the supercoil radius; $\varphi_1, \varphi_2, \dots, \varphi_n$, 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 α atoms rotate about the α -helical axis by $\sim(\omega_0 + \omega_1)$, and the protein backbone is strained when this sum deviates from the value of 100° found in ideal helices (which have $\omega_0 = 0^\circ$ and $\omega_1 = 100^\circ$) (fig. S1). Hence, supercoil (ω_0) and helical (ω_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° from an ideal unstrained helix. First, if ω_1 is increased to 102.85° from the ideal value of 100.0° , after seven residues the helix has completed two full turns (720°). Second, if ω_1 is reduced to 98.2° , after 11 residues the helix has completed three full turns (1080°) (8). Third, if ω_1 is kept at exactly 100° , after 18 residues the helix has completed five full turns (1800°). 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 ω_0 and ω_1 , two-layer designs are left-handed (ω_0 negative), three-layer designs are right-handed (ω_0 positive), and five-layer designs are untwisted (ω_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.

¹Department of Biochemistry, University of Washington, Seattle, WA 98195, USA. ²Institute for Protein Design, University of Washington, Seattle, WA 98195, USA. ³Institute of Molecular Biosciences, University of Graz, Humboldtstrasse 50/3, 8010-Graz, Austria. ⁴Department of Biochemistry, University of Cambridge, Cambridge CB2 1GA, UK. ⁵Janelia Research Campus, Howard Hughes Medical Institute, Ashburn, VA 20147, USA. ⁶Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, UK. ⁷Howard Hughes Medical Institute, University of Washington, Seattle, WA 98195, USA. *These authors contributed equally to this work. †Present address: Department of Chemistry, University of Tokyo, Graduate School of Science, 7-3-1, Hongo, Bunkyo, Tokyo 113-0033, Japan. ‡Corresponding author. E-mail: dabaker@u.washington.edu

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 ΔG_{D-N} in the absence of denaturant of 61 ± 5 kcal mol⁻¹ at 25°C (fig. S4). Because of the long extrapolation, sharp unfolding transition, and the limited unfolded protein CD baseline, the error in ΔG_{D-N} may be significantly larger, but the fit m -value ($m_{D-N} = 8.1 \pm 0.7$ kcal mol⁻¹ M⁻¹; 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⁻¹ 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 (ω_0) to a value close to 0°—the ideal value for a

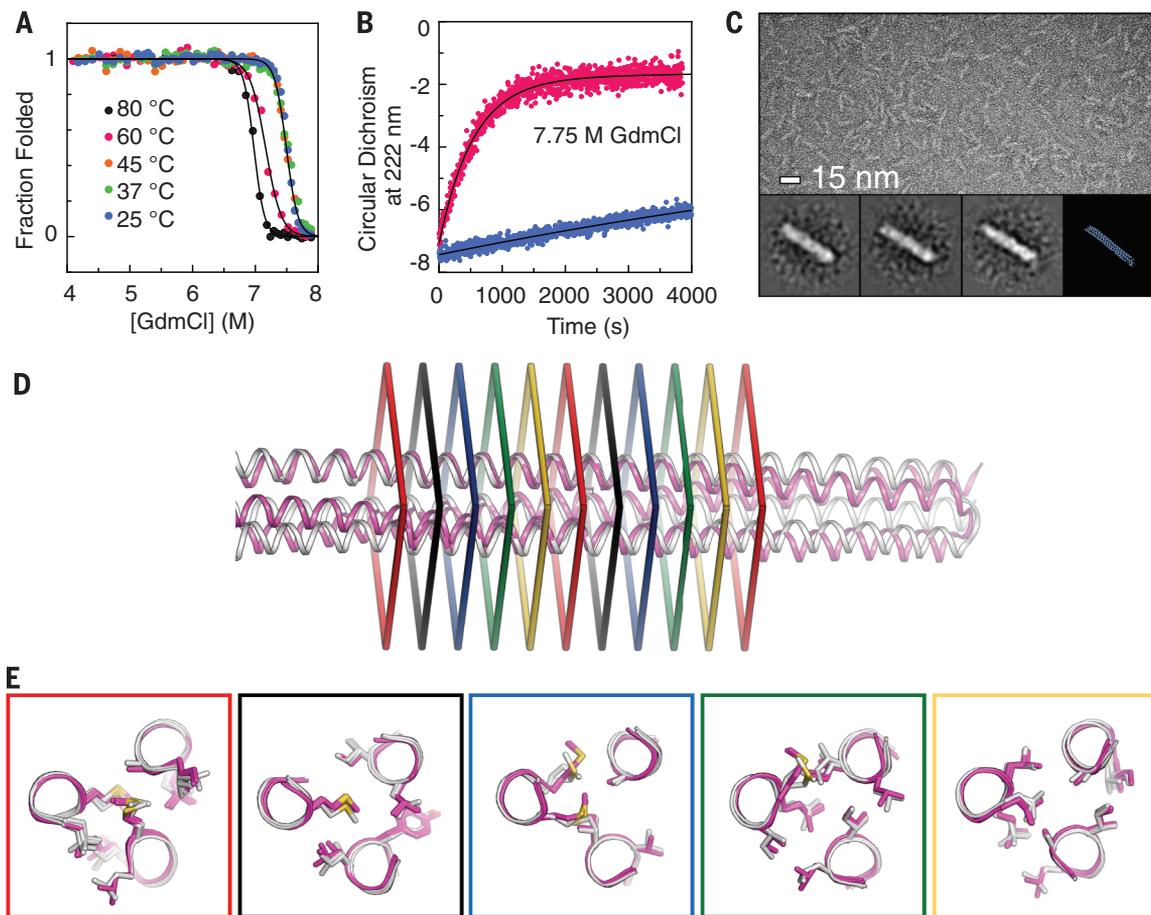


Fig. 1. Stability and structure of designed monomeric three-helix bundle 3H5L_2. (A) GdmCl denaturation monitored by CD. At 80°C, the midpoint of the folding transition is ~7 M GdmCl. (B) Kinetics of unfolding in 7.75 M GdmCl at 25°C (blue) and 60°C (red). (C) Negative-stain electron micrographs of 3H5L_2; 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 C α 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.

five-layer bundle—observed in the crystal structure of 3H5L_2, we fixed ω_0 at the ideal value given the layer type in subsequent designs. Thus, for the three-layer bundle, the helix twist ω_1 was set to 98.2° and the supercoil twist ω_0 to 1.8° . To reduce the size of the search space, we restricted sampling to C_2 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 α -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 , ω_0 , and φ_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 $\varphi_1 = 43^\circ$ (fig. S8, A to C). In a five-helix bundle, each helix has two interaction surfaces at 108° from each other; with this solution for φ_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 C_5 arrangement sampled was nearly identical to

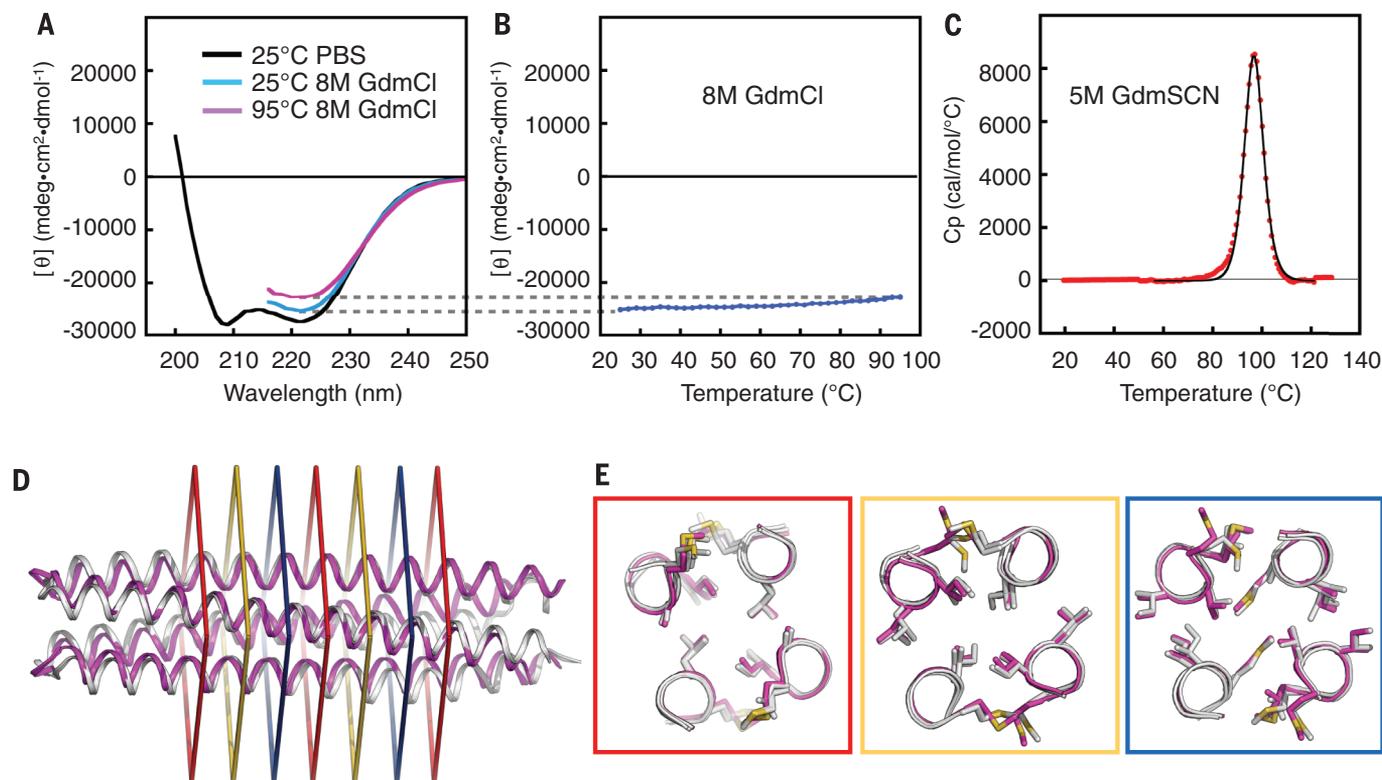


Fig. 2. Stability and structure of designed monomeric four-helix bundle 4H3L_3. (A) CD spectra of 4H3L_3 in the presence and absence of GdmCl. (B) Temperature dependence of CD signal at 222 nm in 8 M GdmCl. No unfolding transition is observed at temperatures up to 95° C. (C) DSC of 4H3L_3 in 5 M GdmSCN. 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.

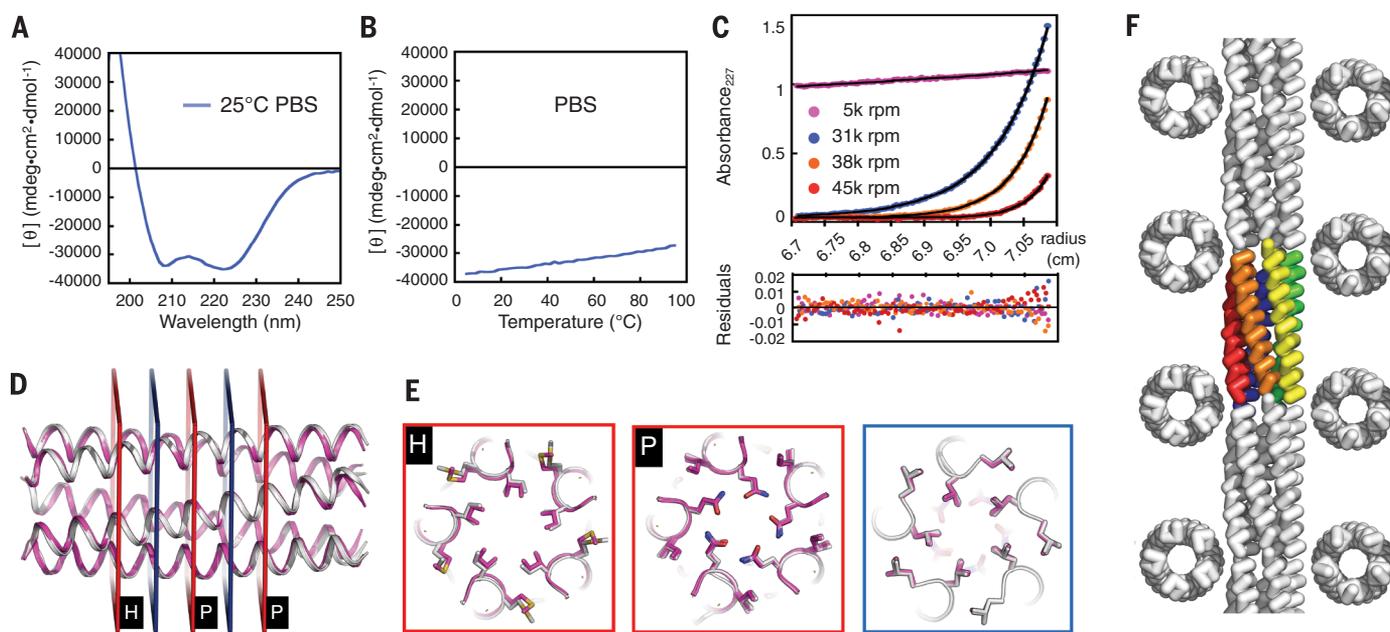


Fig. 3. Stability and structure of designed pentameric five-helix bundle 5H2L_2. (A) CD spectrum and (B) CD-monitored temperature melt of 5H2L_2 (0.2 mg/ml in PBS, pH 7.4). (C) Representative analytical ultracentrifugation sedimentation-equilibrium curves at four different rotor speeds for 5H2L_2 0.5 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.

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 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. 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 -value $\times C_m = 25$ kcal/mol (gray) separates almost all native proteins from the two designs. 4H3L_3 does not denature in GdmCl.

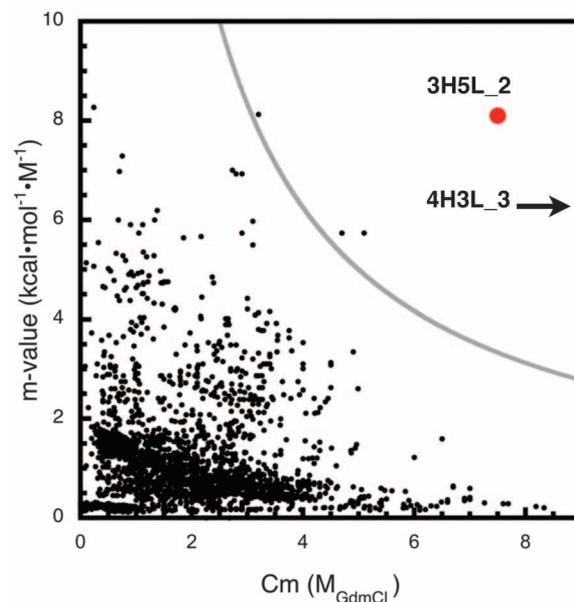


fig. S11). This is notable given that the sequences and structures of the designs came directly from Rosetta calculations with no human modification or experimental optimization. That hyperstability is relatively easy to achieve by design (two out of nine designs tested), but very rarely observed [an 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 α 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.

REFERENCES AND NOTES

1. A. L. Boyle, D. N. Woolfson, *Chem. Soc. Rev.* **40**, 4295–4306 (2011).
2. F. H. C. Crick, *Acta Crystallogr.* **6**, 685–689 (1953).
3. M. Gruber, A. N. Lupas, *Trends Biochem. Sci.* **28**, 679–685 (2003).
4. R. B. Hill, D. P. Raleigh, A. Lombardi, W. F. DeGrado, *Acc. Chem. Res.* **33**, 745–754 (2000).
5. P. B. Harbury, T. Zhang, P. S. Kim, T. Alber, *Science* **262**, 1401–1407 (1993).
6. J. Liu *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 15457–15462 (2006).
7. N. R. Zaccari *et al.*, *Nat. Chem. Biol.* **7**, 935–941 (2011).
8. P. B. Harbury, J. J. Plecs, B. Tidor, T. Alber, P. S. Kim, *Science* **282**, 1462–1467 (1998).
9. G. Grigoryan *et al.*, *Science* **332**, 1071–1076 (2011).
10. R. Das, D. Baker, *Annu. Rev. Biochem.* **77**, 363–382 (2008).
11. G. Grigoryan, W. F. DeGrado, *J. Mol. Biol.* **405**, 1079–1100 (2011).
12. Materials and methods are available as supplementary materials on Science Online.
13. P. S. Huang *et al.*, *PLOS ONE* **6**, e24109 (2011).
14. P. Bradley, K. M. S. Misura, D. Baker, *Science* **309**, 1868–1871 (2005).
15. C. N. Pace, *Methods Enzymol.* **131**, 266–280 (1986).
16. C. D. Geierhaas, A. A. Nickson, K. Lindorff-Larsen, J. Clarke, M. Vendruscolo, *Protein Sci.* **16**, 125–134 (2007).
17. M. D. Kumar *et al.*, *Nucleic Acids Res.* **34**, D204–D206 (2006).
18. J. Peters, W. Baumeister, A. Lupas, *J. Mol. Biol.* **257**, 1031–1041 (1996).
19. K. C. Chou, L. Carlacci, G. M. Maggiora, *J. Mol. Biol.* **213**, 315–326 (1990).

ACKNOWLEDGMENTS

We are indebted to G. Grigoryan and W. DeGrado for their paper on coiled-coil geometry and to G. Grigoryan for the CCPP 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-AC02-05CH11213) 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-2012-10F). 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.

SUPPLEMENTARY MATERIALS

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)

16 June 2014; accepted 17 September 2014
 10.1126/science.1257481

PROTEIN DESIGN

Computational design of water-soluble α -helical barrels

Andrew R. Thomson,^{1†} Christopher W. Wood,^{1,2*} Antony J. Burton,^{1*} Gail J. Bartlett,¹ Richard B. Sessions,² R. Leo Brady,² Derek N. Woolfson^{1,2†}

The design of protein sequences that fold into prescribed de novo structures is challenging. General solutions to this problem require geometric descriptions of protein folds and methods to fit sequences to these. The α -helical coiled coils present a promising class of protein for this and offer considerable scope for exploring hitherto unseen structures. For α -helical barrels, which have more than four helices and accessible central channels, many of the possible structures remain unobserved. Here, we combine geometrical considerations, knowledge-based scoring, and atomistic modeling to facilitate the design of new channel-containing α -helical barrels. X-ray crystal structures of the resulting designs match predicted in silico models. Furthermore, the observed channels are chemically defined and have diameters related to oligomer state, which present routes to design protein function.

Defining protein sequences that fold into specified three-dimensional structures is called the “inverse protein-folding problem” (1). Mostly, this has been applied to mimic existing folds (2). However, the design of structures not yet seen in nature can also be considered (3). Repeat proteins are of interest here, as extrapolation from known structures can provide geometric parameters and sequence-to-structure relations to guide design (4, 5), and proteins with cyclic symmetry offer possibilities for systematic variation of repeating elements to produce families of proteins (6). One example is the α -helical coiled coil (7, 8). Classical coiled coils comprise bundles of two to four α helices, account for >98% of known coiled-coil structures (9, 10), and have well-understood sequence-to-structure relations (7, 8). Unusually for proteins, the conformations of coiled-coil backbones are well described by a small number of parameters (11–14). Consequently, a relatively large number of successful coiled-coil structures have been designed (8), although, with a few exceptions (8, 13), these have largely mimicked natural precedents.

The α -helical barrels present an intriguing subset of coiled coils to move beyond known structures (15). These have more-complex helical packing (15, 16), which results in the assembly of five or more α helices into cylindrical bundles with central channels or pores (15). The few current examples include natural parallel 5- and 10-helix structures, and antiparallel 10- and 12-helix bundles (17–20); a de novo parallel hexamer, achieved partly serendipitously (21); and a mutant leucine-zipper peptide that forms an unusual staggered parallel 7-helix arrangement (22). For these, there is a near-linear relation be-

tween lumen size and oligomer state (Fig. 1A), which opens possibilities for designing channel proteins. However, because of the scarcity of α -helical barrels and because these are usually parts of larger membrane-spanning proteins, it is difficult to derive rules to design new examples. To overcome this, we describe a geometrical and computational framework for designing α -helical barrels from first principles and apply this to deliver discrete, water-soluble assemblies with five to seven parallel and identical helices.

First, we required a means to map structural and sequence relations between the new targets and the plentiful examples of classical coiled coils. The latter have heptad sequence repeats, $(haxhxxx)_n$, where h and x are hydrophobic and polar residues, respectively; often labeled $abcdefg$; this places h -type residues at a and d . A resulting hydrophobic seam mediates helix association and packing, with the interface often buttressed by polar interactions between e and g (Fig. 1, B and C). These are type-N interfaces, and the residues at the “ $gade$ ” positions determine oligomer state and partner selection (8, 23).

In an α -helical barrel, each helix interacts with two neighbors via independent hydrophobic seams (Fig. 1, D and E). There are three ways to achieve this within a heptad repeat: The two seams can share one residue (type I interfaces); be adjacent (type II) (Fig. 1D); or be separated by an intervening residue (type III) (15, 16). We hypothesized that type II and III interfaces can lead to α -helical barrels, with the oligomer state determined by the angular offset between the two interfaces. Because differences in this contact angle between helices become smaller with increasing oligomeric state (fig. S1A), we anticipated that controlling barrel size would be more tractable for smaller assemblies. Therefore, we concentrated on type II interfaces and “ $hhaxhha$ ” repeats (read $g-f$), which should define oligomers with five to seven helices.

Also with increasing oligomer state, the helix-helix interfaces in α -helical barrels become more

¹School of Chemistry, University of Bristol, Cantock’s Close, Bristol, BS8 1TS, UK. ²School of Biochemistry, University of Bristol, Medical Sciences Building, University Walk, Bristol, BS8 1TD, UK.

*These authors contributed equally to this work. †Corresponding author. E-mail: chzart@bristol.ac.uk (A.R.T.); d.n.woolfson@bristol.ac.uk (D.N.W.)