

that the time required for acceleration and transport to Earth does not greatly exceed the  $^{60}\text{Fe}$  half-life of 2.62 My. Our distance from the source of this nuclide cannot greatly exceed the distance that cosmic rays can diffuse over this time scale, which is  $\lesssim 1$  kpc.

*Note added in proof:* Additional detections of  $^{60}\text{Fe}$  in deep-sea crusts in all major oceans of the world have recently been reported (21), strengthening the conclusions reached in (9). Also, additional detections of  $^{60}\text{Fe}$  in lunar samples (22) strengthen the conclusion reached in (12, 13).

#### REFERENCES AND NOTES

1. G. Rugel et al., *Phys. Rev. Lett.* **103**, 072502 (2009).
2. M. E. Wiedenbeck et al., *Astrophys. J.* **523**, L61–L64 (1999).
3. S. E. Woosley, A. Heger, *Phys. Rep.* **442**, 269–283 (2007).
4. A. Chieffi, M. Limongi, *Astrophys. J.* **764**, 21 (2013).
5. E. C. Stone et al., *Space Sci. Rev.* **86**, 285–356 (1998).
6. W. Wang et al., *Astron. Astrophys.* **469**, 1005–1012 (2007).
7. W. R. Binns, *Science* **334**, 1071–1072 (2011).
8. K. Knie et al., *Phys. Rev. Lett.* **83**, 18–21 (1999).
9. K. Knie et al., *Phys. Rev. Lett.* **93**, 171103 (2004).
10. C. Fitoussi et al., *Phys. Rev. Lett.* **101**, 121101 (2008).
11. B. J. Fry, B. D. Fields, J. R. Ellis, *Astrophys. J.* **800**, 71 (2015).
12. L. Fimiani et al., *43rd Lunar and Planetary Science Conference*, 1279 (2012).
13. L. Fimiani et al., *45th Lunar and Planetary Science Conference*, 1778 (2014).
14. Materials and methods are available as supplementary materials on *Science Online*.
15. K. A. Lave et al., *Astrophys. J.* **770**, 117 (2013).
16. J. C. Higdon, R. E. Lingenfelter, *Astrophys. J.* **590**, 822–832 (2003).
17. J. C. Higdon, R. E. Lingenfelter, *Astrophys. J.* **628**, 738–749 (2005).
18. N. E. Yanasak et al., *Astrophys. J.* **563**, 768–792 (2001).
19. P. T. de Zeeuw, R. Hoogerwerf, J. H. J. de Bruijne, A. G. A. Brown, A. Blaauw, *Astron. J.* **117**, 354–399 (1999).
20. R. Voss, R. Diehl, J. S. Vink, D. H. Hartmann, *Astron. Astrophys.* **520**, A51 (2010).
21. A. Wallner et al., *Nature* **532**, 69–72 (2016).
22. L. Fimiani et al., *Phys. Rev. Lett.* **116**, 151104 (2016).

#### ACKNOWLEDGMENTS

This work was supported by NASA grants NNX08AI11G and NNX13AH66G for work performed at the California Institute of Technology, the Jet Propulsion Laboratory, and Washington University in St. Louis. Work done at Goddard Space Flight Center was funded by NASA through the ACE Project. We thank S. Woosley and J. Brown at the University of California Santa Cruz for providing modeling calculation data and for discussions about uncertainties in their calculations. We thank A. Chieffi at the Institute for Space Astrophysics and Planetary, Rome, Italy, for discussions on the uncertainties in the C&L calculations. Accelerator testing of the CRIS detectors was made possible by N. Anantaraman, R. Ronningen, and the staff at the National Superconducting Cyclotron Laboratory at Michigan State University, while H. Specht, D. Schardt, and the staff of the GSI heavy-ion accelerator in Darmstadt, Germany made possible the heavy-ion calibrations of the completed CRIS instrument. The data used are archived at NASA's Space Physics Data Facility (<http://spdf.gsfc.nasa.gov>) as data set ac\_h2\_criss and can be retrieved from this site by direct download or through the SPDF's CDAWeb data service.

#### SUPPLEMENTARY MATERIALS

[www.sciencemag.org/content/352/6286/677/suppl/DC1](http://www.sciencemag.org/content/352/6286/677/suppl/DC1)  
Materials and Methods  
Supplementary Text  
Figs. S1 to S6  
References (23–33)  
3 November 2015; accepted 18 March 2016  
Published online 21 April 2016  
[10.1126/science.aad6004](http://10.1126/science.aad6004)

## PROTEIN DESIGN

# De novo design of protein homo-oligomers with modular hydrogen-bond network-mediated specificity

Scott E. Boyken,<sup>1,2,3</sup> Zibo Chen,<sup>1,2,4</sup> Benjamin Groves,<sup>5</sup> Robert A. Langan,<sup>1,2,4</sup> Gustav Oberdorfer,<sup>1,2,7</sup> Alex Ford,<sup>1,2,4</sup> Jason M. Gilmore,<sup>1,2</sup> Chunfu Xu,<sup>1,2</sup> Frank DiMaio,<sup>1,2</sup> Jose Henrique Pereira,<sup>8,9</sup> Banumathi Sankaran,<sup>10</sup> Georg Seelig,<sup>5,6</sup> Peter H. Zwart,<sup>10,11</sup> David Baker<sup>1,2,3\*</sup>

In nature, structural specificity in DNA and proteins is encoded differently: In DNA, specificity arises from modular hydrogen bonds in the core of the double helix, whereas in proteins, specificity arises largely from buried hydrophobic packing complemented by irregular peripheral polar interactions. Here, we describe a general approach for designing a wide range of protein homo-oligomers with specificity determined by modular arrays of central hydrogen-bond networks. We use the approach to design dimers, trimers, and tetramers consisting of two concentric rings of helices, including previously not seen triangular, square, and supercoiled topologies. X-ray crystallography confirms that the structures overall, and the hydrogen-bond networks in particular, are nearly identical to the design models, and the networks confer interaction specificity *in vivo*. The ability to design extensive hydrogen-bond networks with atomic accuracy enables the programming of protein interaction specificity for a broad range of synthetic biology applications; more generally, our results demonstrate that, even with the tremendous diversity observed in nature, there are fundamentally new modes of interaction to be discovered in proteins.

**H**ydrogen bonds play key roles in the structure, function, and interaction specificity of biomolecules. There are two main challenges facing de novo design of hydrogen-bonding interactions: First, the partially covalent nature of the hydrogen bond restricts polar hydrogen-containing donors and electronegative acceptors to narrow ranges of orientation and distance, and second, nearly all polar atoms must participate in hydrogen bonds—either with other macromolecular polar atoms, or with solvent—if not, there is a considerable energetic penalty associated with stripping away water upon folding or binding (7). The DNA double helix elegantly resolves both challenges; paired bases come together such that all buried polar atoms make hydrogen bonds that are self-contained between the two bases and have

near-ideal geometry. In proteins, meeting these challenges is more complicated because backbone geometry is highly variable, and pairs of polar amino acids cannot generally interact as to fully satisfy their mutual hydrogen-bonding capabilities; hence, side-chain hydrogen bonding usually involves networks of multiple amino acids with variable geometry and composition, and there are generally very different networks at different sites within a single protein or interface preorganizing polar residues for binding and catalysis (2–6).

The modular and predictable nature of DNA interaction specificity is central to molecular biology manipulations and DNA nanotechnology (7, 8), but without parallels in nature, it has not been evident how to achieve analogous programmable specificity with proteins. There are more polar amino acids than DNA bases, each of which can adopt numerous side-chain conformations in the context of different backbones, which allows for countless network possibilities. We hypothesized that by systematically searching through these network possibilities, it could be possible to design protein interfaces specified by regular arrays of DNA-like central hydrogen-bond networks with modular specificity analogous to Watson-Crick base pairing.

We began by developing a general computational method, HBNet, to rapidly enumerate all side-chain hydrogen-bond networks possible in an input backbone structure (Fig. 1A). Traditional protein design algorithms are not well suited for this purpose; the total system energy is generally expressed as the sum of interactions between pairs of residues for computational

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

<sup>4</sup>Graduate Program in Biological Physics, Structure, and Design, University of Washington, Seattle, WA 98195, USA. <sup>5</sup>Department of Electrical Engineering, University of Washington, Seattle, WA 98195, USA. <sup>6</sup>Department of Computer Science and Engineering, University of Washington, Seattle, WA 98195, USA.

<sup>7</sup>Institute of Molecular Biosciences, University of Graz, Humboldtstrasse 50/3, 8010-Graz, Austria. <sup>8</sup>Joint BioEnergy Institute, Emeryville, CA 94608, USA. <sup>9</sup>Molecular Biophysics and Integrated Bioimaging, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA. <sup>10</sup>Berkeley Center for Structural Biology, Molecular Biophysics and Integrated Bioimaging, Lawrence Berkeley Laboratory, 1 Cyclotron Road, Berkeley, CA 94720, USA.

<sup>11</sup>The Center for Advanced Mathematics for Energy Research Applications, Lawrence Berkeley National Laboratories, 1 Cyclotron Road, Berkeley, CA 94720, USA.

\*Corresponding author. Email: [dabaker@u.washington.edu](mailto:dabaker@u.washington.edu)

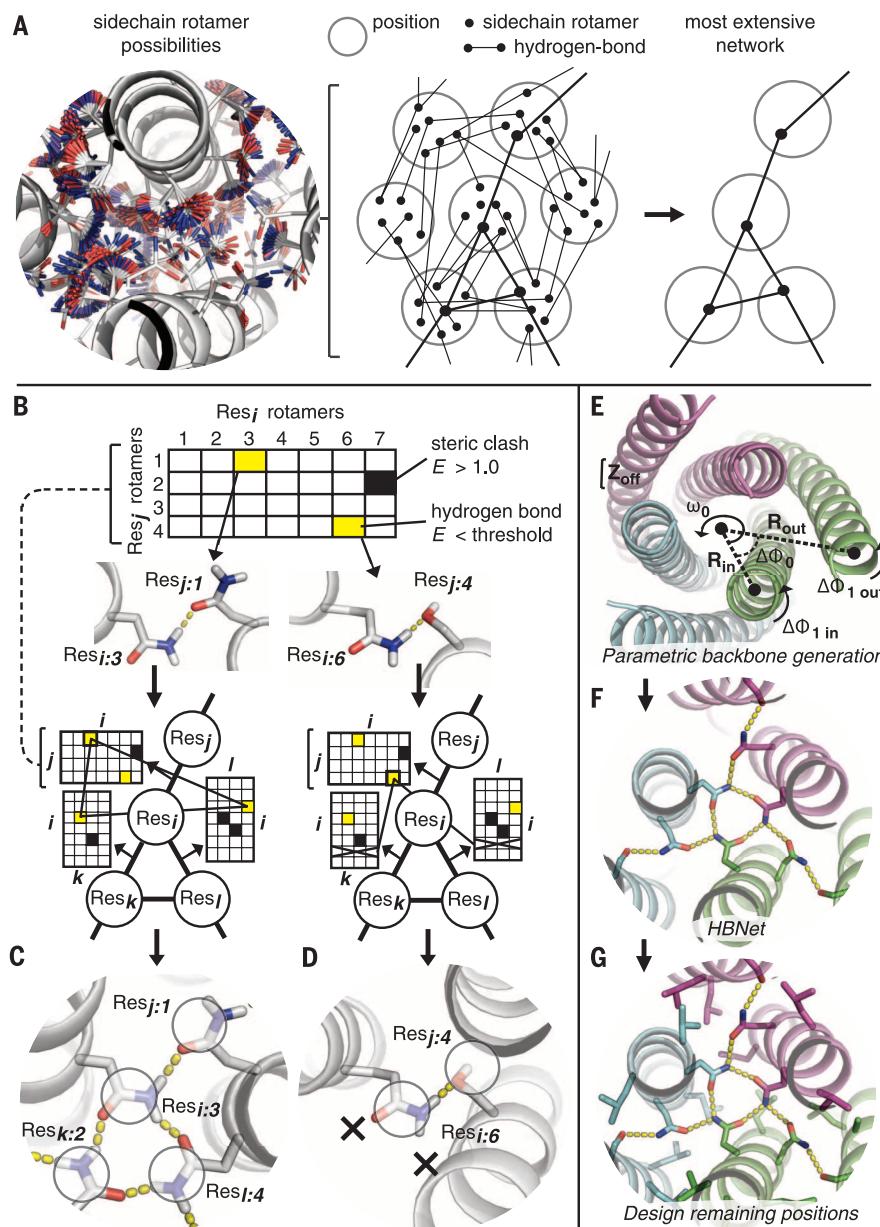
efficiency (9–11) and cannot clearly distinguish a connected hydrogen-bond network from a set of disconnected hydrogen bonds. HBNet starts by precomputing the hydrogen-bonding and steric repulsion interactions between all conformations (rotameric states) of all pairs of polar side chains. These energies are stored in a graph data structure in which the nodes are residue positions, positions close in three-dimensional space are connected by edges, and for each edge, there is a matrix representing the interaction energies between the different rotameric states at the two positions. HBNet then traverses this graph to identify all networks of three or more residues connected by low-energy hydrogen bonds with little steric repulsion (Fig. 1B). The most extensive and lowest-energy networks (Fig. 1C) are kept fixed in subsequent design calculations at the remaining

residue positions. Networks with buried donors and acceptors not making hydrogen bonds (unsatisfied) are rejected (Fig. 1D). Details of the method, as well as scripts for carrying out the design calculations, are provided in the supplementary materials.

Inspired by the DNA double helix, we aimed to host the hydrogen-bond networks in protein oligomers with an inherent repeat structure to enable networks to be reutilized within the same scaffold. We therefore turned our attention to coiled coils, which are abundant in nature (12, 13), the subject of many protein design studies (14–17), and can be generated parametrically (18, 19), which results in repeating geometric cross sections. Coiled-coil packing and oligomerization state are largely determined by position-specific identities of non-polar residues that pack between the helices (20–22); salt bridge and hydrogen-bonding interactions be-

tween residues on the periphery can provide additional specificity (23–25). In natural and designed coiled coils, buried polar interactions can also alter specificity; however, most of these cases involve, at most, one or two side-chain–side-chain hydrogen bonds with remaining polar atoms satisfied by water or ions (26–30). The relatively small cross-sectional interface area of canonical coiled coils limits the diversity and location of possible networks. To overcome these limitations, we decided to focus on oligomeric structures with two concentric rings of helices (Fig. 1E and fig. S1).

We built “two-ring” topologies from helical hairpin monomer subunits consisting of an inner and outer helix connected by a short loop by using a generalization of the Crick coiled-coil parameterization (31). Wide ranges of backbones were generated by systematically sampling the radii and



**Fig. 1. Overview of the HBNet method and design strategy.** (A) (Left) All side-chain conformations (rotamers) of polar amino acid types considered for design at each residue position (oxygen atoms colored red, nitrogen atoms blue); (middle) many combinations of hydrogen-bonding rotamers are possible, and the challenge is to traverse this space and extract (right) networks of connected hydrogen bonds. (B to D) HBNet. (B) HBNet precomputes the hydrogen bond and steric repulsive interaction energies between side-chain rotamers at all pairs of positions and stores them in a graph structure; nodes are residue positions, residue pairs close enough to interact are connected by edges, and for each edge there is an interaction energy matrix; yellow indicates rotamer pairs with energies below a specified threshold (hydrogen bonds with good geometry and little steric repulsion). Traversing the graph elucidates all possible connectivities of hydrogen-bonding rotamers (networks) that do not clash with each other. In the simple example shown, two pairs of side-chain rotamers at Res<sub>i</sub> and Res<sub>j</sub> make good-geometry hydrogen bonds, but graph traversal shows that only one of these (left) can be extended into a connected network: (C) Res<sub>i</sub> rotamer 3 (i:3) can hydrogen bond to both Res<sub>k</sub> rotamer 2 (k:2) and Res<sub>j</sub> rotamer 4 (j:4), yielding a “good” network of fully connected Asn residues with all heavy-atom donors and acceptors satisfied, whereas (D) would be rejected because the hydrogen-bonding rotamers i:6 (Gln) and j:4 (Ser) cannot form additional hydrogen bonds to nearby positions k and l, which leaves unsatisfied buried polar atoms. (E to G) Design strategy. (E) Parametric generation of two-ring coiled-coil backbones. For example, a C3 symmetric trimer (monomer subunits in different colors) is defined by the following parameters: supercoil radius of inner ( $R_{in}$ ) and outer ( $R_{out}$ ) helices, helical phase of the inner ( $\Delta\phi_{in}$ ) and outer ( $\Delta\phi_{out}$ ) helices, supercoil phase of the outer helix ( $\Delta\phi_0$ ), z-offset between the inner and outer helices ( $Z_{off}$ ), and the supercoil twist ( $\omega_0$ ). (F) HBNet is applied to parametric backbones to identify the best hydrogen-bond networks. (G) Networks are maintained; the remaining residue positions are designed in the context of the assembled symmetric oligomer.

helical phases of the inner and outer helices, the z-offset between inner and outer helices, and the overall supercoil twist (Fig. 1E). HBNet was then used to search these backbones for networks that span the intermolecular interface, have all heavy-atom donors and acceptors satisfied, and involve at least three side chains (Fig. 1F); because of these stringent requirements, only a small fraction of backbones can support such networks—but by systematically varying the degrees of freedoms of the two-ring structures, tens of thousands of backbones can readily be generated, and the efficiency of HBNet makes searching for networks in large numbers of backbones computationally tractable. Rosetta Design (II, 32) was then used to optimize rotamers at the remaining residue positions in the context of the cyclic symmetry of the oligomer (Fig. 1G). Designs were ranked based on the total oligomer energy by using the Rosetta all-atom force field (33) and were filtered to remove designs with large cavities or poor packing around the networks. The top-ranked designs were evaluated using Rosetta “fold-and-dock” calculations (34). Designs with energy landscapes shaped like funnels leading into the target-designed structure were identified, and a total of 114 dimeric, trimeric, and tetrameric designs spanning a broad range of superhelical parameters and hydrogen-bond networks were selected for experimental characterization [table S1; for design naming convention see (35)].

Synthetic genes encoding the selected designs were obtained and the proteins expressed in *Escherichia coli*. The ~90% (101/114) of designs that were expressed and soluble (table S2) were purified by affinity chromatography, and their oligomerization state was evaluated by size-exclusion chromatography multiangle light scattering (SEC-MALS). Of the 101 soluble designs, 66 were found to have the designed oligomerization state (table S2). The 101 soluble designs span eight different topologies (fig. S1); of these, the supercoiled tetramers have the largest buried interface area, yielded the fewest designs with all buried donors

and acceptors satisfied, and had the lowest success rate (only 3 of the 13 soluble designs properly assembled). Excluding supercoiled tetramers, 72% (63/88) assembled to the designed oligomeric state, and of these, 89% (56/63) eluted as a single peak from the SEC column. The designed proteins were further characterized by circular dichroism (CD) spectroscopy; all designs tested exhibited characteristic  $\alpha$ -helical spectra, and CD-monitored unfolding experiments showed that more than 90% of these were stable at 95°C (Fig. 2 and figs. S2 to S8).

To probe the energetic contribution of the outer ring of helices, we compared the stability of the two-ring designs to corresponding designs with only the inner ring; core interface positions of the inner helices, including hydrogen-bond network residues, were retained, and solvent-exposed surface positions were redesigned in the same manner as the surface of the two-ring designs. Design 2L4HC2\_9 (Fig. 2C), a supercoiled homodimer, is folded and thermostable (Fig. 2D); Its inner helix peptide, 2L4HC2\_9\_inner (Fig. 2E), also forms a homodimeric coiled coil (fig. S9), but with markedly decreased thermostability (Fig. 2F). Design 2L6HC3\_13 (Fig. 2G), a supercoiled homotrimer, is also folded and thermostable (Fig. 2H); however, the corresponding inner ring peptide (Fig. 2I) in isolation is unfolded (Fig. 2J) and monomeric (fig. S9D). This inner helix is internally frustrated: It has four Asn residues at canonical  $\alpha$  or  $\delta$  heptad-packing positions (fig. S9E), where Asn has been found to be destabilizing (36, 37), and Leu and Ile at other  $\alpha$  and  $\delta$  positions, respectively, which favors homotetramers (37). In the presence of the outer helix and designed hydrogen-bond networks, the two-ring design assembles to the intended trimeric structure, as elucidated by x-ray crystallography (Fig. 3A). Together, these results suggest that the outer ring of helices not only increases thermostability but also can drive coiled-coil assembly, even in the context of an inner helix with low helical propensity and noncanonical helical packing (fig. S9).

which permits greater sequence diversity across larger interfaces.

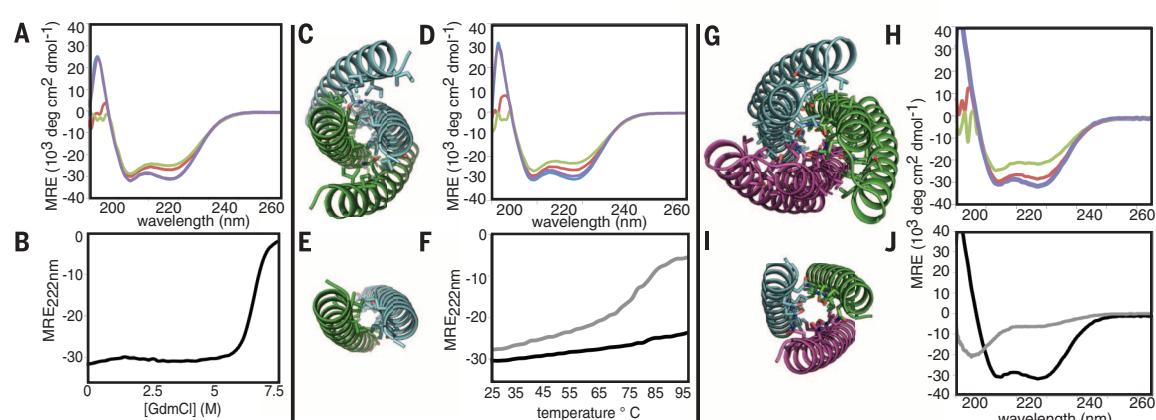
## Structural characterization

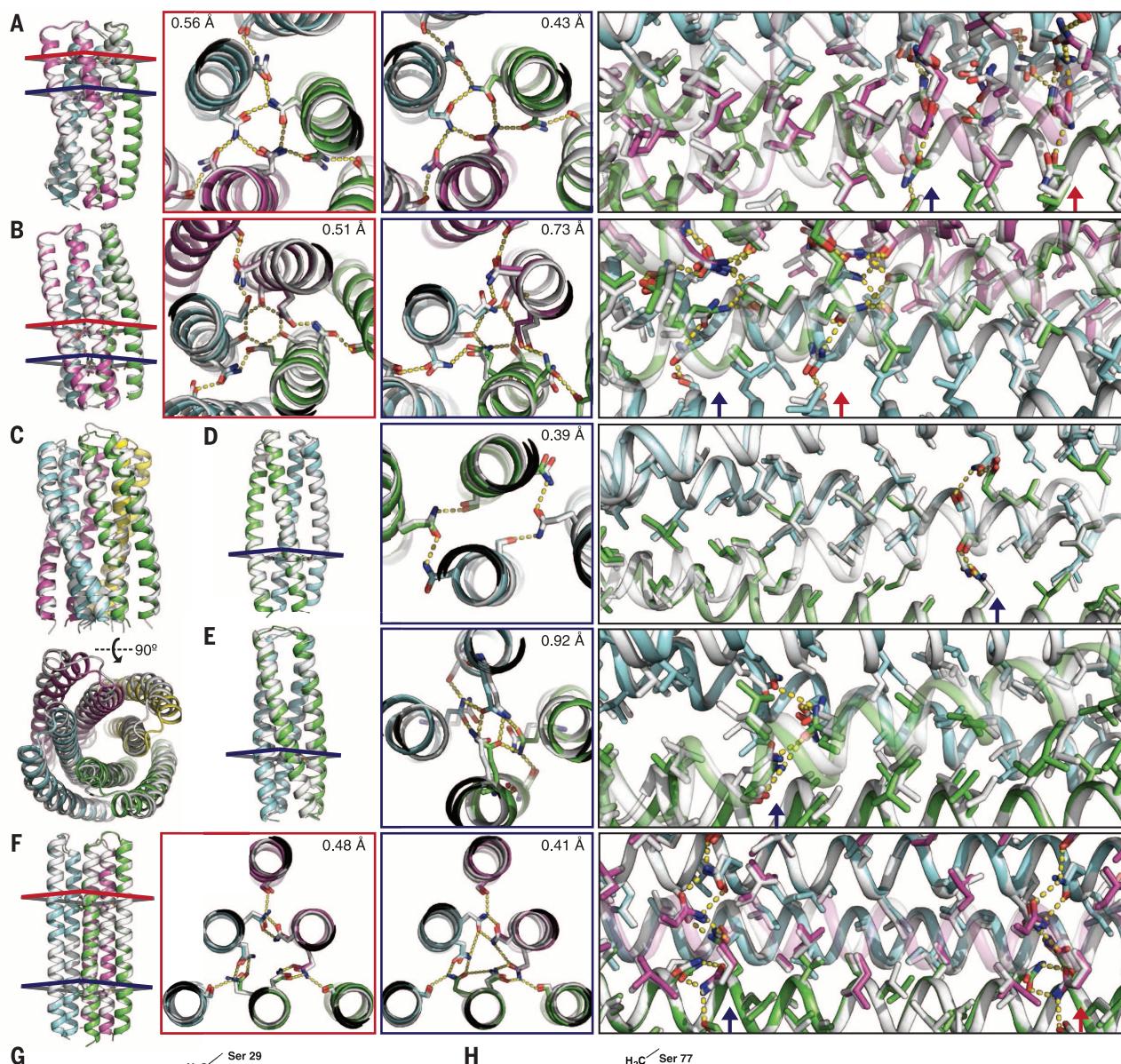
To assess the accuracy of the designs, we determined 10 crystal structures spanning a range of oligomerization states, superhelical parameters, and hydrogen-bond networks (Fig. 3, A to F, and figs. S10 to S12). Designs for which crystals were not obtained were characterized by small-angle x-ray scattering (SAXS) (Fig. 4, figs. S13 and S14, and table S4). We solved structures for three left-handed trimers, four left-handed dimers, a left-handed tetramer, and an untwisted triangle-shaped trimer. Additional topologies characterized by SAXS include square-shaped untwisted tetramers (Fig. 4A) and dimers (Fig. 4B), as well as six-helix dimers (two inner, one outer helix) with either parallel right-handed (Fig. 4C) or antiparallel left-handed (Fig. 4D) supercoil geometry. Five of the x-ray crystallography-verified designs (Fig. 3, A and C to F) were also characterized by SAXS (fig. S14A), and the experimentally determined spectra were found to closely match those computed from the design models, which suggests that very similar structures are populated in solution.

The three left-handed trimer structures are remarkably similar to the design models with subangstrom root mean square deviation (RMSD) across all backbone  $C\alpha$  atoms and across all heavy atoms of the hydrogen-bond networks (Fig. 3, A and B, and fig. S10). These structures are constructed with supercoil phases of 0, 120, and 240 degrees for the inner helices, and 60, 180, and 300 degrees for the outer helices (fig. S1); loops connect outer N-terminal helices to inner C-terminal helices (at -60 degrees from the outer helix). Extensive 9- or 12-residue networks form the intended hydrogen bonds in the crystal structures (Fig. 3, A and B, middle, and fig. S10). Unlike previously designed single-ring trimers where three buried asparagines (Asns) resulted in substantially decreased thermostability (38), these two-ring trimers are stable up to 95°C and ~4.5 M

**Fig. 2. The outer ring of helices increases thermostability and can overcome the poor helical propensity of the inner helices.** (A) CD spectrum (260 to 195 nm) of design 2L4HC2\_23 at 25°C (blue), 75°C (red), 95°C (green), and 25°C after cooling (purple).

(B) Design 2L4HC2\_9 unfolds at 6.5 M GdmCl. (C) Design 2L4HC2\_9, a supercoiled C2 homodimer colored by chain, view down the supercoil axis. (D) CD spectrum of 2L4HC2\_9 as in (A). (E) Inner ring design of 2L4HC2\_9. (F) CD temperature melt monitoring absorption at 222 nm; 2L4HC2\_9 (black) is considerably more stable than 2L4HC2\_9\_inner (gray). (G) Design 2L6HC3\_13, a supercoiled C3 homotrimer. (H) CD spectrum of 2L6HC3\_13 at different temperatures as in (A). (I) 2L6HC3\_13\_inner. (J) CD spectra of 2L6HC3\_13 (black) and 2L6HC3\_13\_inner (gray); in the absence of the outer helix, the inner helix is unfolded. All CD data are plotted in mean residue ellipticity (MRE)  $10^3 \text{ deg cm}^2 \text{ dmol}^{-1}$ .





**Fig. 3. X-ray crystal structures are in close agreement with the design models.** (A to F) Crystal structures (white) are superimposed onto the design models, monomer subunits colored green, cyan, magenta, yellow for six different topologies; (left) the full backbone is shown with colored cross-sections corresponding to the (middle) designed hydrogen-bond networks (yellow dashed lines); outline color corresponds to cross-section color on the left; RMSD over all network residue heavy atoms is reported inside each panel; (right)

hydrophobic core packing surrounding the networks, which are indicated by colored arrows. (A) 2L6HC3\_13 (1.64 Å resolution; RMSD = 0.56 Å over all C<sub>α</sub> atoms) and (B) 2L6HC3\_6 (2.26 Å resolution; RMSD = 0.77 Å over all C<sub>α</sub> atoms) are left-handed C3 homotrimers, each with two identical networks at different locations that span the entire interface, contacting all six helices. (C) 2L8HC4\_12, a left-handed C4 homotetramer with two different hydrogen-bond networks (fig. S4D); the low (3.8 Å) resolution does not allow assessment of the hydrogen-bond network side chains. (D) 2L4HC2\_9 (2.56 Å resolution; 0.39 Å RMSD over all C<sub>α</sub> atoms) and (E) 2L4HC2\_23 (1.54 Å resolution; RMSD = 1.16 Å over all C<sub>α</sub> atoms) are left-handed C2 homodimers, each with one network. (F) 5L6HC3\_1 (2.36 Å resolution; RMSD = 0.51 Å over all C<sub>α</sub> atoms) is a C3 homotrimer with straight, untwisted helices and two identical networks at different cross sections. (G and H) Schematics of hydrogen-bond networks from 2L6HC3\_13 (A) and 5L6HC3\_1 (F). The indicated hydrogen bonds are present in both design model and crystal structure.

guanidinium chloride (fig. S3 and fig. S9) with numerous buried polar residues; 2L6HC3\_13 has 12 completely buried Asns, and 2L6HC3\_6 has 24 buried polar residues confined to a small region of the interface, including six Asns and six glutamines (Glns).

The four left-handed dimer crystal structures all have the designed parallel two-ring topology. Two of the dimer structures have hydrogen-bond networks in close agreement with the designs: 2L4HC2\_9 (Fig. 3D) and 2L4HC2\_23 (Fig. 3E) have 0.39 Å and 0.92 Å RMSD across all network residue heavy atoms, respectively, and 0.39 Å and 1.16 Å RMSD over all C<sub>α</sub> atoms. The other two, 2L4HC2\_11 (fig. S11, A and B) and 2L4HC2\_24 (fig. S11, C to E), have slight structural deviations from the design models caused by water displacing designed network side chains; in the former, the interface shifts ~2 Å because of a buried water molecule bridging two network residues (fig. S11B), and in the latter, the backbone is nearly identical to the design model, but side chains of the designed network are displaced by ordered water molecules (fig. S11E). These two cases highlight the need for high connectivity and satisfaction (all polar atoms participating in hydrogen bonds) of the networks. The left-handed tetramer structure has the designed overall topology (Fig. 3C), and SAXS data are in close agreement with the design model (fig. S14), but side-chain density was uncertain because of low (3.8 Å) resolution. The amino acid sequence is unrelated to any known sequence, and the top hit in structure-based searches of the Protein Data Bank (PDB) has a different helical bundle arrangement (fig. S15D).

The five antiparallel six-helix dimers were soluble and assembled to the designed oligomeric state (table S2), with SAXS data in agreement with the design models (Fig. 4D and fig. S14). Design

2L6Hanti\_3 contains a hydrogen-bond network with a buried Tyr at the dimer interface (Fig. 4D). Of the three right-handed six-helix dimers characterized by SAXS, 3L6HC2\_4 (Fig. 4C) and 3L6HC2\_7 (fig. S14) exhibited scattering in agreement with the design models, whereas 3L6HC2\_2 did not (fig. S14). Although 3L6HC2\_2 was designed to form a parallel dimer, the crystal structure revealed an antiparallel dimer interface, which highlights two design lessons (fig. S12): (i) the importance of intermolecular hydrogen bonds at the binding interface (the 3L6HC2\_2 design model has only two across the interface compared with nine in 2L6HC3\_6) (Fig. 3B) and (ii) the importance of favorable hydrophobic contacts complementing the networks (the 3L6HC2\_2 design model has mainly Alas at the interface).

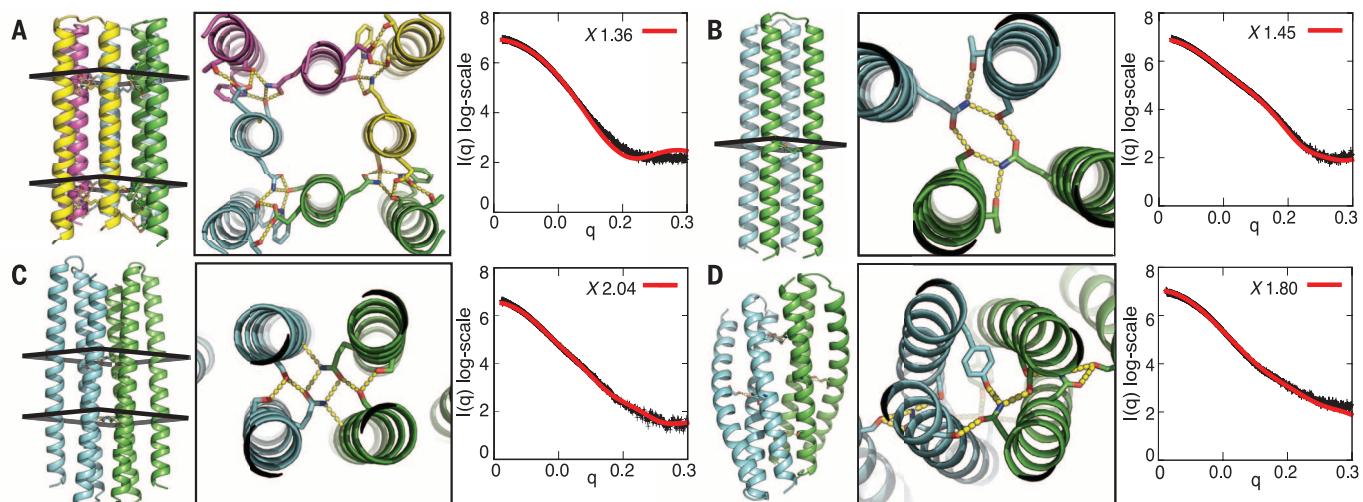
SAXS data suggest that our untwisted dimer, trimer, and tetramer designs assemble into the target triangular and square conformations (Fig. 4, A and B, and fig. S14). Guinier analysis (table S4) and fit of the low-q region of the scattering vector indicates that the seven untwisted dimers tested are in the correct oligomeric state, four of which have very close agreement between the experimental spectra and design models (Fig. 4B and fig. S14). The SAXS data on the three untwisted trimers were all in close agreement with the corresponding design models (Fig. 4A, fig. S14, and table S4). Design 5L8HC4\_6 has a distinctive network with a Trp making a buried hydrogen bond at one end of the network, which then propagates outwards toward solvent and connects to a Glu on the surface (Fig. 4A). To the best of our knowledge, oligomers with such uniformly straight helices do not exist in nature, nor have these topologies been designed previously.

The 2.36 Å crystal structure of the untwisted trimer reveals straight helices with 0.51 Å RMSD

to the design model over all C<sub>α</sub> atoms (Fig. 3F). The two hydrogen-bond networks (Fig. 3F, middle), as well as the hydrophobic packing residues surrounding the networks (Fig. 3F right), are nearly identical between the crystal structure and design model. Like the supercoiled trimers, each of these networks contains side chains from every helix, and helices were constructed to be uniformly spaced (fig. S1). The helices are nearly perfectly straight in the crystal structure, with supercoil twist values very close to the idealized design value of zero:  $\omega_0 = -0.036$  degrees per residue for the inner three helices and  $\omega_0 = -0.137$  degrees per residue for the outer three helices. Blast searches with the amino acid sequence returned no matches with E-values better than 10, and the top hit in a search for similar structures in the PDB has three supercoiled helices flanked by long extended regions (fig. S15E).

### Comparison of successful versus unsuccessful network designs

Several trends emerged distinguishing successful designs. First, in successful designs, nearly all buried polar groups made hydrogen bonds. We selected designs with all heavy-atom donors and acceptors satisfied, but the networks had varying numbers of polar hydrogens unsatisfied. Networks with the largest fraction of satisfied polar groups generally had relatively high connectivity, both with respect to the total number of hydrogen bonds and number of side chains contributing to the network. The networks with the highest connectivity and structural accuracy spanned the entire cross-sectional interface, with each helix contributing at least one side chain (Fig. 3, A, B, E, and F). Design 2L6HC3\_13 also has two additional smaller networks consisting of a single symmetric Asn making two hydrogen bonds, but with



**Fig. 4. Structural characterization by SAXS.** (Left) backbones and (middle) hydrogen-bond networks for the design models are displayed as in Fig. 3; (right) design models (red) were fit to experimental scattering data (black) using FoXS (59, 60); quality of fit ( $X$ ) is indicated inside each panel. **(A)** Design 5L8HC4\_6 ( $X = 1.36$ ), an untwisted C4 homotetramer with two identical hydrogen-bond networks. **(B)** Design 5L4HC2\_12 ( $X = 1.45$ ), an untwisted C2 homodimer with a single hydrogen-bond network. **(C)** Design 3L6HC2\_4 ( $X = 2.04$ ), a parallel right-handed C2 homodimer with two repeated networks, two inner helices, and one outer helix. **(D)** Design 2L6Hanti\_3 ( $X = 1.80$ ), a left-handed antiparallel homodimer with two inner helices and one outer helix; because of the antiparallel geometry, the same network occurs in two locations.

one polar hydrogen unsatisfied; in the crystal structure, these residues move away from the design model, displaced by water molecules (fig. S16).

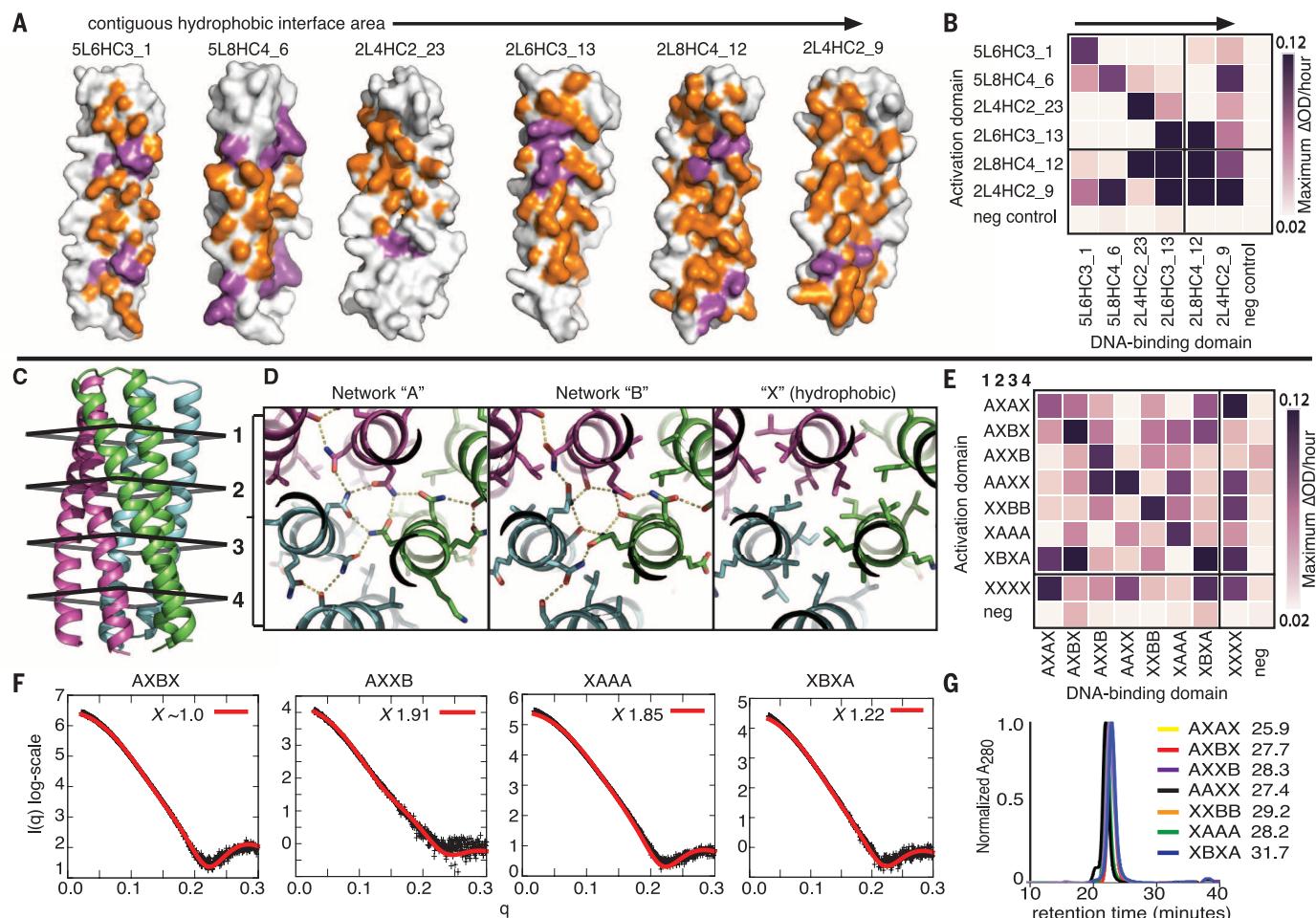
### The designed hydrogen-bond networks confer specificity

To test the role of the designed hydrogen-bond networks in conferring specificity for the target oligomeric state, we carried out control design calculations using the same protein backbones without HBNet, which yielded uniformly hydrophobic interfaces. In silico, despite having lower total energy in the designed oligomeric state, these designs exhibit more pronounced alternative energy-minima in fold-and-dock and asymmetric docking calculations (fig. S17), consistent with the

much less restrictive geometry of nonpolar packing interactions. Experimentally, these hydrophobic designs exhibited less soluble expression than their counterparts with hydrogen-bond networks (fig. S18A) and tended to precipitate during purification; of those that remained in solution long enough to collect SEC-MALS data, all but one formed higher-molecular-weight aggregates and eluted as multiple peaks from the SEC column (fig. S18). These results suggest that the designed hydrogen-bond networks confer specificity for the target oligomeric state and resolve the degeneracy of alternative states observed with purely hydrophobic packing (this degeneracy is considerably more pronounced for our two-ring structures than traditional single-ring coiled coils, which

have many fewer total hydrophobic residues and less interhelical interface area).

We used an *in vivo* yeast two-hybrid assay (38) to further probe the interaction specificity of the designed oligomers. Sequences encoding a range of dimers, trimers, and tetramers were crossed against each other in all-by-all binding assays (Fig. 5 and fig. S19): Synthetic genes for the designs were cloned in frame with both DNA binding domains and transcriptional activation domains in separate vectors, and the extent of binding between the different designs assessed by cell growth, which requires juxtaposition of the DNA binding domain with the activation domain. Even without explicit negative design, the designed homo-oligomeric interactions are stronger than the (unintended)



**Fig. 5. The hydrogen-bond networks confer specificity.** (A) Interaction surfaces of monomer subunits for six structurally verified designs, ordered by increasing contiguous hydrophobic interface area (orange), as calculated by hpatch (61); hydrogen-bond network residues are colored magenta. (B) Binding heat-map from yeast two-hybrid assay. Designs in (A) were fused to both DNA binding domain and the activation domain constructs and binding measured by determining the cell growth rate [maximum change in optical density ( $\Delta\text{OD}$ ) per hour]; darker cells indicate more rapid growth, hence stronger binding; values are the average of at least three biological replicates with standard deviations reported in fig. S19. The heat-map is ordered as in (A), and designs with more extensive networks and better-partitioned hydrophobic interface area exhibit higher interaction specificity. (C to G) Modular networks confer specificity in a programmable fashion. (C) The backbone corresponding to designs 2L6HC3\_13 (Fig. 3A) and 2L6HC3\_6 (Fig. 3B) can accommodate different networks at each of four repeating geometric cross sections. (D) Three possibilities for each cross section: Network "A," network "B," or hydrophobic, "X." (E) Combinatorial designs using this three-letter combination were tested for interaction specificity using the yeast two-hybrid assay as in (B). Axis labels denote the network pattern; for example, "AXBX" indicates network A at cross section 1, network B at cross section 3, and X (hydrophobic) at the two others. (F) SAXS profiles for combinatorial designs as in Fig. 4. (G) SEC chromatograms monitoring absorbance at 280 nm ( $A_{280}$ ) and estimated molecular masses (from MALS); designs range from ~27 to 30 kD. AXXX, XXBB, and XXXX correspond to designs 2L6HC3\_13, 2L6HC3\_6, and 2L6HC3\_hydrophobic\_1, respectively.

programmable fashion. (C) The backbone corresponding to designs 2L6HC3\_13 (Fig. 3A) and 2L6HC3\_6 (Fig. 3B) can accommodate different networks at each of four repeating geometric cross sections. (D) Three possibilities for each cross section: Network "A," network "B," or hydrophobic, "X." (E) Combinatorial designs using this three-letter combination were tested for interaction specificity using the yeast two-hybrid assay as in (B). Axis labels denote the network pattern; for example, "AXBX" indicates network A at cross section 1, network B at cross section 3, and X (hydrophobic) at the two others. (F) SAXS profiles for combinatorial designs as in Fig. 4. (G) SEC chromatograms monitoring absorbance at 280 nm ( $A_{280}$ ) and estimated molecular masses (from MALS); designs range from ~27 to 30 kD. AXXX, XXBB, and XXXX correspond to designs 2L6HC3\_13, 2L6HC3\_6, and 2L6HC3\_hydrophobic\_1, respectively.

competing hetero-oligomeric interactions (Fig. 5B). Designs in which the hydrogen-bond networks partition hydrophobic interface area into relatively small regions are more specific than designs with large contiguous hydrophobic patches at the helical interface (Fig. 5, A and B). The designs with the best-partitioned hydrophobic area had networks spanning the entire oligomeric interface, with each helix contributing at least one side chain. This unifying design principle can readily be enforced using HBNet.

To test if regular arrays of networks can confer specificity in a modular, programmable manner, we designed an additional set of trimers, each with identical backbones and hydrophobic packing motifs, the only difference being placement and composition of the hydrogen-bond networks. The designs are based on 2L6HC3\_13 (Fig. 3A) and 2L6HC3\_6 (Fig. 3B), which originated from the same superhelical parameters but have unique networks we will refer to as "A" and "B," respectively; cross sections with only nonpolar residues are labeled "X." We used this three-letter code to generate new designs in combinatorial fashion: At each of the four repeating cross sections of the supercoil (Fig. 5C), we placed the A, B, or X (Fig. 4D) followed by the same design strategy and selection process as before. The design names indicate network placement; for example, "AXAX" has network "A" at cross sections one and three and hydrophobic packing ("X") at two and four. Six of these combinatorial designs were synthesized, and five out of six were found to be folded, thermostable, and assembled to the designed trimeric oligomerization state in vitro (Fig. 5, F and G, and fig. S20). These five, along with the two parent designs (2L6HC3\_13 = AAXX and 2L6HC3\_6 = XXBB) and an all-hydrophobic control (XXXX), were crossed in all-by-all yeast two-hybrid binding experiments (Fig. 5E). Again, the designed self interactions were found to be the strongest. Overall, the combinatorial designs exhibit a level of specificity that is striking, given that all have identical backbones and high overall sequence similarity (fig. S20), whereas the hydrophobic control is relatively promiscuous; the central hydrogen-bond networks are clearly responsible for mediating specificity.

## Conclusions

Previous de novo protein design efforts have focused on jigsaw-puzzle-like hydrophobic core packing to design new structures and interactions (39–42). Unlike the multibody problem of designing highly connected and satisfied hydrogen-bond networks, hydrophobic packing is readily captured by established pairwise-decomposable potentials; because of this and the inherent challenge of designing buried polar interactions, most protein interface designs have been predominantly hydrophobic, and attempts to design buried hydrogen bonds across interfaces have routinely failed (43). Polar interfaces have been designed in specialized cases (44–46) but have been difficult to generalize, with many interface design efforts requiring directed evolution to optimize polar contacts and achieve desired specificity (47, 48).

HBNet now provides a general computational method to accurately design hydrogen-bond networks. This ability to precisely preorganize polar contacts without buried unsatisfied polar atoms should be broadly useful in protein design challenges such as enzyme design, small molecule binding, and polar protein interface targeting.

Our two-ring structures are a new class of protein oligomers that have the potential for programmable interaction specificity analogous to that of Watson-Crick base pairing. Whereas Watson-Crick base pairing is largely limited to the antiparallel double helix, our designed protein hydrogen-bond networks allow the specification of two-ring structures with a range of oligomerization states (dimers, trimers, and tetramers) and supercoil geometries. Elegant studies have demonstrated a wide range of interaction specificity with standard single-ring coiled coils (37, 49–55); with an outer ring of helices to enable extensive hydrogen-bond networks, it should be possible to generate a much larger range of orthogonal interactions than has been achieved previously. Our results demonstrate that a wide range of hydrogen-bond network compositions and geometries is possible in repeating two-ring topologies and multiple networks can be engineered into the same backbone at varying positions without sacrificing thermostability, and that network combination enables stable building blocks with uniform shape but orthogonal binding interfaces (Fig. 5). The DNA nanotechnology field has demonstrated that a spectacular array of shapes and interactions can be built from a relatively limited set of hydrogen-bonding interactions (56–58). It should now become possible to develop new protein-based materials with the advantages of both polymers: DNA-like programmability and tunable specificity coupled with the geometric variability, interaction diversity, and catalytic function intrinsic to proteins.

## REFERENCES AND NOTES

- P. J. Fleming, G. D. Rose, *Protein Sci.* **14**, 1911–1917 (2005).
- R. Chakrabarti, A. M. Klibanov, R. A. Friesner, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 12035–12040 (2005).
- E. T. Judd, N. Stein, A. A. Pacheco, S. J. Elliott, *Biochemistry* **53**, 5638–5646 (2014).
- A. Sánchez-Azqueta et al., *Biochim. Biophys. Acta* **1837**, 251–263 (2014).
- D. Xu, C. J. Tsai, R. Nussinov, *Protein Eng.* **10**, 999–1012 (1997).
- F. B. Sheinerman, B. Honig, *J. Mol. Biol.* **318**, 161–177 (2002).
- Y.-J. Chen, B. Groves, R. A. Muscat, G. Seelig, *Nat. Nanotechnol.* **10**, 748–760 (2015).
- F. Zhang, J. Nangreave, Y. Liu, H. Yan, *J. Am. Chem. Soc.* **136**, 11198–11211 (2014).
- A. G. Street, S. L. Mayo, *Structure* **7**, R105–R109 (1999).
- B. I. Dahiyat, S. L. Mayo, *Protein Sci.* **5**, 895–903 (1996).
- B. Kuhlman, D. Baker, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 10383–10388 (2000).
- M. Gruber, A. N. Lupas, *Trends Biochem. Sci.* **28**, 679–685 (2003).
- P. Burkhardt, J. Stetefeld, S. V. Strelkov, *Trends Cell Biol.* **11**, 82–88 (2001).
- P.-S. Huang et al., *Science* **346**, 481–485 (2014).
- N. L. Ogihara, M. S. Weiss, W. F. DeGrado, D. Eisenberg, *Protein Sci.* **6**, 80–88 (1997).
- A. W. Reinke, R. A. Grant, A. E. Keating, *J. Am. Chem. Soc.* **132**, 6025–6031 (2010).
- A. L. Boyle, D. N. Woolfson, *Chem. Soc. Rev.* **40**, 4295–4306 (2011).
- G. Grigoryan, W. F. DeGrado, *J. Mol. Biol.* **405**, 1079–1100 (2011).
- P. B. Harbury, J. J. Plecs, B. Tidor, T. Alber, P. S. Kim, *Science* **282**, 1462–1467 (1998).
- D. N. Woolfson, T. Alber, *Protein Sci.* **4**, 1596–1607 (1995).
- P. B. Harbury, T. Zhang, P. S. Kim, T. Alber, *Science* **262**, 1401–1407 (1993).
- A. Acharya, V. Rishi, C. Vinson, *Biochemistry* **45**, 11324–11332 (2006).
- H. Gradišar, R. Jerala, *J. Pept. Sci.* **17**, 100–106 (2011).
- J. B. Kaplan, A. W. Reinke, A. E. Keating, *Protein Sci.* **23**, 940–953 (2014).
- E. G. Baker et al., *Nat. Chem. Biol.* **11**, 221–228 (2015).
- D. M. Eckert, V. N. Malashkevich, P. S. Kim, *J. Mol. Biol.* **284**, 859–865 (1998).
- D. L. Akey, V. N. Malashkevich, P. S. Kim, *Biochemistry* **40**, 6352–6360 (2001).
- K. J. Lumb, P. S. Kim, *Biochemistry* **34**, 8642–8648 (1995).
- C. D. Tatko, V. Nanda, J. D. Lear, W. F. DeGrado, *J. Am. Chem. Soc.* **128**, 4170–4171 (2006).
- L. Gonzalez Jr., D. N. Woolfson, T. Alber, *Nat. Struct. Biol.* **3**, 1011–1018 (1996).
- F. H. C. Crick, *Acta Crystallogr.* **6**, 689–697 (1953).
- A. Leaver-Fay et al., *Methods Enzymol.* **487**, 545–574 (2011).
- M. J. O'Meara et al., *J. Chem. Theory Comput.* **11**, 609–622 (2015).
- R. Das et al., *Proc. Natl. Acad. Sci. U.S.A.* **106**, 18978–18983 (2009).
- Design nomenclature: The first two characters indicate supercoil geometry: "2L" refers to a two-layer heptad repeat that results in a left-handed supercoil; "3L" refers to a three-layer 11-residue repeat with a right-handed supercoil; and "5L" refers to untwisted designs with a five-layer 18-residue repeat and straight helices (no supercoiling), where "layer", in this context, is the number of unique repeating geometric slices along the supercoil axis. The middle two characters indicate the total number of helices, and last two characters indicate symmetry and oligomeric assembly. For example, "2L6HC3" denotes a left-handed, six-helix trimer with C3 symmetry.
- A. Acharya, S. B. Ruvinov, J. Gal, J. R. Moll, C. Vinson, *Biochemistry* **41**, 14122–14131 (2002).
- J. M. Fletcher et al., *ACS Synth. Biol.* **1**, 240–250 (2012).
- S. Fields, O. Song, *Nature* **340**, 245–246 (1989).
- B. Kuhlman et al., *Science* **302**, 1364–1368 (2003).
- N. Koga et al., *Nature* **491**, 222–227 (2012).
- N. P. King et al., *Science* **336**, 1171–1174 (2012).
- P.-S. Huang, J. J. Love, S. L. Mayo, *Protein Sci.* **16**, 2770–2774 (2007).
- P. B. Stranges, B. Kuhlman, *Protein Sci.* **22**, 74–82 (2013).
- E. Procko et al., *J. Mol. Biol.* **425**, 3563–3575 (2013).
- L. A. Joachimiak, T. Kortemme, B. L. Stoddard, D. Baker, *J. Mol. Biol.* **361**, 195–208 (2006).
- T. Kortemme et al., *Nat. Struct. Mol. Biol.* **11**, 371–379 (2004).
- J. Karanicolas et al., *Mol. Cell* **42**, 250–260 (2011).
- S. J. Fleishman et al., *Science* **332**, 816–821 (2011).
- S. F. Betz, P. A. Liebman, W. F. DeGrado, *Biochemistry* **36**, 2450–2458 (1997).
- G. Grigoryan, A. E. Keating, *Curr. Opin. Struct. Biol.* **18**, 477–483 (2008).
- G. Grigoryan, A. W. Reinke, A. E. Keating, *Nature* **458**, 859–864 (2009).
- A. W. Reinke, J. Baek, O. Ashenberg, A. E. Keating, *Science* **340**, 730–734 (2013).
- F. Thomas, A. L. Boyle, A. J. Burton, D. N. Woolfson, *J. Am. Chem. Soc.* **135**, 5161–5166 (2013).
- A. R. Thomson et al., *Science* **346**, 485–488 (2014).
- N. H. Joh et al., *Science* **346**, 1520–1524 (2014).
- P. W. K. Rothemund, *Nature* **440**, 297–302 (2006).
- E. Winfree, F. Liu, L. A. Wenzler, N. C. Seeman, *Nature* **394**, 539–544 (1998).
- B. Wei, M. Dai, P. Yin, *Nature* **485**, 623–626 (2012).
- D. Schneidman-Duhovny, M. Hammel, A. Sali, *Nucleic Acids Res.* **38** (Web Server), W540–W544 (2010).
- D. Schneidman-Duhovny, M. Hammel, J. A. Tainer, A. Sali, *Biophys. J.* **105**, 962–974 (2013).
- R. Jacak, A. Leaver-Fay, B. Kuhlman, *Proteins* **80**, 825–838 (2012).

## ACKNOWLEDGMENTS

We thank L. Carter for assistance with SEC-MALS and protein production and Rosetta@Home volunteers for contributing computing resources to enable rigorous testing of designs by ab initio structure prediction, F. Seeger for guidance with SAXS data analysis, and M. Bick and P. Lu for assistance with crystallographic refinement. This work was supported by the Howard Hughes Medical Institute, the U.S. Department of Energy, the European Research Area (ERA)-NET BioOrigami consortium,

and NSF (MCB-1445201). Design calculations were facilitated through the use of advanced computational, storage, and networking infrastructure provided by the Hyak supercomputer system at the University of Washington. X-ray crystallography and SAXS data were collected at the Advanced Light Source (Lawrence Berkeley National Laboratory, Berkeley, California Department of Energy, contract no. DE-AC02-05CH11231); SAXS data were collected through the SIBYLS mail-in SAXS program under the aforementioned contract number, and we thank K. Burnett and G. Hura. The Berkeley Center for Structural Biology is supported in part by the National Institute of General Medical Sciences (NIH), and the Howard Hughes Medical Institute. G.O. is a Marie Curie International Outgoing Fellowship fellow (332094 ASR-CompEnzDes FP7-People-2012-IOF). B.G and J.M.G. are supported

by Washington Research Foundation Postdoctoral Fellowships. Coordinates and structure files have been deposited to the Protein Data Bank with accession codes: 5J0J (2L6HC3\_6), 5J0I (2L6HC3\_12), 5J0H (2L6HC3\_13), 5IZS (5L6HC3\_1), 5J73 (2L4HC2\_9), 5J2L (2L4HC2\_11), 5J0L (3L6HC2\_2), 5J0K (2L4HC2\_23), 5J1O (2L4HC2\_24). S.E.B., Z.C., and D.B. designed the research and S.E.B. and D.B. wrote the manuscript. S.E.B. developed the HBNet method and wrote the program code. D.B. wrote the parametric backbone generation code with help from C.X. and G.O. A.F. wrote the loop closure program code. S.E.B., Z.C., R.A.L., and D.B. carried out design calculations. S.E.B. and Z.C. purified and biophysically characterized the designed proteins. B.G. performed yeast two-hybrid assays. J.M.G. performed mass spectrometry. J.H.P. crystallized the designed proteins. B.S.

and P.H.Z. collected and analyzed crystallographic data. B.S., P.H.Z., G.O. and F.D. solved structures with help from S.E.B. and Z.C. All authors discussed results and commented on the manuscript.

## SUPPLEMENTARY MATERIALS

[www.science.org/content/352/6286/680/suppl/DC1](http://www.science.org/content/352/6286/680/suppl/DC1)

Materials and Methods

Supplementary Text

Figs. S1 to S20

Tables S1 to S5

References (62–82)

4 December 2015; accepted 23 March 2016

10.1126/science.aad8865

## REPORTS

### PROTEIN DESIGN

# Design of structurally distinct proteins using strategies inspired by evolution

T. M. Jacobs,<sup>1</sup> B. Williams,<sup>2</sup> T. Williams,<sup>2</sup> X. Xu,<sup>3,4\*</sup> A. Eletsky,<sup>3,4</sup> J. F. Federizon,<sup>3</sup> T. Szyperski,<sup>3</sup> B. Kuhlman<sup>2,5†</sup>

Natural recombination combines pieces of preexisting proteins to create new tertiary structures and functions. We describe a computational protocol, called SEWING, which is inspired by this process and builds new proteins from connected or disconnected pieces of existing structures. Helical proteins designed with SEWING contain structural features absent from other de novo designed proteins and, in some cases, remain folded at more than 100°C. High-resolution structures of the designed proteins CA01 and DA05R1 were solved by x-ray crystallography (2.2 angstrom resolution) and nuclear magnetic resonance, respectively, and there was excellent agreement with the design models. This method provides a new strategy to rapidly create large numbers of diverse and designable protein scaffolds.

**M**ost efforts in de novo protein design have been focused on creating idealized proteins composed of canonical structural elements. Examples include the design of coiled coils, repeat proteins, TIM barrels, and Rossmann folds (1–6). These studies elucidate the minimal determinants of protein structure, but they do not aggressively explore new regions of structure space. Additionally, idealized structures may not always be the most effective starting points for engineering novel protein functions. Functional sites in proteins are often created from nonideal structural elements, such as kinks, pockets, and bulges.

The lack of nonideal structural elements from de novo designed proteins highlights a key difference between natural protein evolution and current design methods. Specifically, protein design methods universally begin with a target structure in mind. Therefore, the space of designable structures that can accommodate these nonideal protein elements is limited by the imagination of the designer. In contrast, natural evolution is based not on design but on cellular fitness provided by the evolved protein function. This lack of a predetermined target fold is a powerful feature of protein evolution that holds significant potential for the design of novel structures and functions. In an effort to tap this potential, we sought to develop a method of computational protein design inspired by mechanisms of natural protein evolution.

Gene duplication and homologous recombination mix and match elements of protein structure to give rise to new structures and functions (7–9). This phenomenon is most evident at the level of independently folding protein domains (10–12), but recent studies have shown that

these same principles function at a smaller scale during the evolution of distinct, globular protein folds (13). Insertions, deletions, and replacement of secondary and supersecondary structural elements sample alternative tertiary structures (14–16). Our design strategy, called SEWING (structure extension with native-substructure graphs), is motivated by this process and builds new protein structures from pieces of naturally occurring protein domains. The process is not dictated by the need to adopt a specific target fold but rather is aimed at creating large sets of alternative structures that satisfy predefined design requirements. One of the strengths of this approach is that it ensures that all of the structural elements of the protein are inherently designable, at the same time allowing for the incorporation of structural oddities unlikely to be found in idealized proteins. Here, we apply SEWING to the design of helical proteins. We show that designed structures are diverse and contain structural features absent from alternative design strategies.

SEWING begins with the extraction of small structural motifs, or substructures, from existing protein structures. These serve as the basic building blocks for all generated models. We aimed to identify substructures that were large enough to carry information regarding structural preference yet small enough to allow combinations that can generate novel globular structures. Ultimately, we chose to extract two distinct types of substructures. The first is composed of continuous stretches of protein structure that encompass two secondary structural elements separated by a loop (Fig. 1). These substructures capture the relative orientation between adjacent secondary structure elements and maintain local packing interactions. In addition, there is evidence that substructures of this size adopt a relatively limited number of conformations that have already been sampled exhaustively in known protein structures (14). The second type is composed of groups of three to five secondary structural elements, where each element makes van der Waals contacts with every other element, but the elements are not necessarily continuous in primary sequence (Fig. 1, supplementary methods). Nonadjacent, or discontinuous, substructures maintain longer-range tertiary interactions that provide valuable stability and are often conserved during protein evolution (17).

\*Program in Bioinformatics and Computational Biology, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA. <sup>†</sup>Department of Biochemistry and Biophysics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA. <sup>‡</sup>Department of Chemistry, State University of New York at Buffalo, Buffalo, NY 14260, USA. <sup>§</sup>Northeast Structural Genomics Consortium. <sup>§</sup>Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA.

<sup>†</sup>Present address: Department of Biochemistry and Molecular Biology, University of Georgia, Athens, GA 30602, USA.  
†Corresponding author. Email: bkuhlman@email.unc.edu

## De novo design of protein homo-oligomers with modular hydrogen-bond network-mediated specificity

Scott E. Boyken, Zibo Chen, Benjamin Groves, Robert A. Langan, Gustav Oberdorfer, Alex Ford, Jason M. Gilmore, Chunfu Xu, Frank DiMaio, Jose Henrique Pereira, Banumathi Sankaran, Georg Seelig, Peter H. Zwart and David Baker

Science 352 (6286), 680-687.  
DOI: 10.1126/science.aad8865

### Building with designed proteins

General design principles for protein interaction specificity are challenging to extract. DNA nanotechnology, on the other hand, has harnessed the limited set of hydrogen-bonding interactions from Watson-Crick base-pairing to design and build a wide range of shapes. Protein-based materials have the potential for even greater geometric and chemical diversity, including additional functionality. Boyken *et al.* designed a class of protein oligomers that have interaction specificity determined by modular arrays of extensive hydrogen bond networks (see the Perspective by Netzer and Fleishman). They use the approach, which could one day become programmable, to build novel topologies with two concentric rings of helices.

Science, this issue p. 680; see also p. 657

### ARTICLE TOOLS

<http://science.sciencemag.org/content/352/6286/680>

### SUPPLEMENTARY MATERIALS

<http://science.sciencemag.org/content/suppl/2016/05/04/352.6286.680.DC1>

### RELATED CONTENT

<http://science.sciencemag.org/content/sci/352/6286/657.full>  
<http://science.sciencemag.org/content/sci/352/6286/687.full>  
<http://science.sciencemag.org/content/sci/352/6288/aag1318.full>

### REFERENCES

This article cites 80 articles, 13 of which you can access for free  
<http://science.sciencemag.org/content/352/6286/680#BIBL>

### PERMISSIONS

<http://www.sciencemag.org/help/reprints-and-permissions>

Use of this article is subject to the [Terms of Service](#)