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

Scott E. Boyken^{1,2,3}, Zibo Chen^{1,2}, Benjamin Groves⁴, Robert A. Langan^{1,2}, Gustav Oberdorfer^{1,2,6}, Alex Ford^{1,2}, Jason Gilmore^{1,2}, Chunfu Xu^{1,2}, Frank DiMaio^{1,2}, Jose Henrique Pereira^{7,8}, Banumathi Sankaran⁹, Georg Seelig^{4,5}, Peter H. Zwart^{9,10}, and David Baker^{1,2,3,*}

¹Department of Biochemistry, University of Washington, Seattle, WA 98195, USA

²Institute for Protein Design, University of Washington, Seattle, WA 98195, USA

³Howard Hughes Medical Institute, University of Washington, Seattle, WA 98195, USA

⁴Department of Electrical Engineering, University of Washington, Seattle, WA 98195, USA

⁵Department of Computer Science & Engineering, University of Washington, Seattle, WA 98195, USA

⁶Institute of Molecular Biosciences, University of Graz, Humboldtstrasse 50/3, 8010-Graz, Austria

⁷Joint BioEnergy Institute, Emeryville, CA 94608, USA

⁸Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA

⁹Berkeley Center for Structural Biology, Molecular Biophysics and Integrated Bioimaging, Lawrence Berkeley Laboratory, 1 Cyclotron Road, Berkeley, CA 94720, USA

¹⁰The Center for Advanced Mathematics for Energy Research Applications, Lawrence Berkeley National Laboratories, 1 Cyclotron Road, Berkeley, CA 94720, USA

Abstract

In nature, structural specificity in DNA and proteins is encoded quite differently: in DNA, specificity arises from modular hydrogen bonds in the core of the double helix, whereas in

*Correspondence to: David Baker, dabaker@u.washington.edu.

Author contributions: S.E.B., Z.C., and D.B. designed the research. 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.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.G. performed mass spectrometry. J.H.P. crystallized the designed proteins. B.S. and P.Z. collected and analyzed crystallographic data. B.S., P.Z., G.O., and F.D. solved structures with help from S.E.B. and Z.C. S.E.B. and D.B. wrote the manuscript. All authors discussed results and commented on the manuscript.

Supplementary Materials:

- I. Computational Methods
- II. Experimental Materials and Methods
- III. Figures S1–S20
- IV. Tables S1–S5
- V. Scripts and Command line input
- VI. References (58–78)

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 is a milestone for protein design and enables the programming of protein interaction specificity for a broad range of synthetic biology applications.

Hydrogen 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, hydrogen bonding atoms are geometrically restricted 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 {Fleming:2005cq}. 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 sidechain 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 pre-organizing polar residues for binding and catalysis {Chakrabarti:2005jw, Judd:2014io, SanchezAzqueta:2014gt} {Xu:1997wr, Sheinerman:2002hi}.

The modular and predictable nature of DNA interaction specificity is central to molecular biology manipulations and DNA nanotechnology {Chen:2015dr, Zhang:2014cj}, 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 sidechain conformations in the context of different backbones, allowing 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 sidechain hydrogen bond networks possible in an input backbone structure (Fig. 1A). Traditional protein design algorithms are not well suited for this purpose because the total system energy is generally expressed as the sum of interactions between pairs of residues for computational efficiency {Kuhlman:2000tc, Dahiyat:1996ia, Street:1999wb}, and hence 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 sidechains. These energies are stored in a graph data structure where 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 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{Gruber:2003by, Burkhard:2001wa}, the subject of many protein design studies{Huang:2014jt, Ogiwara:1997fr, Reinke:2010bp, Boyle:2011id}, and can be generated parametrically{Grigoryan:2011dr, Harbury:1998ub}, resulting in repeating geometric cross-sections. Coiled coil packing and oligomerization state is largely determined by position-specific identities of nonpolar residues that pack between the helices{Woolfson:1995gl, Harbury:1993ub, Acharya:2006ey}; salt bridge and hydrogen bonding interactions between residues on the periphery can provide additional specificity{Gradisar:2011ef, Kaplan:2014bk, Baker:2015bv}. In natural and designed coiled coils, buried polar interactions can also alter specificity; however, most of these cases involve at most one or two sidechain-sidechain hydrogen bonds with remaining polar atoms satisfied by water or ions{Eckert:1998fa, Akey:2001tr, Lumb:1995uh, Tatko:2006kq, Gonzalez:1996vl} – 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 using a generalization of the Crick coiled-coil parameterization{Crick:1953dq}. Wide ranges of backbones were generated by systematically sampling the radii and 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 sidechains (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). RosettaDesign{Kuhlman:2000tc, LeaverFay:2011fc} 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 using the Rosetta all atom force field{OMeara:2015fi}, filtering to remove designs with large cavities

or poor packing around the networks. The top-ranked designs were evaluated using Rosetta “fold-and-dock” calculations [Das:2009bs]. 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).

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 evaluated by size-exclusion chromatography multi-angle light scattering (SEC-MALS). Sixty-six of the 101 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 α -helical spectra, and CD monitored unfolding experiments showed that more than 90% of these were stable at 95°C (Fig. 2 and figS2–8).

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). 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). The sequence of this inner helix is notable because it has four Asn residues at canonical *a* or *d* heptad packing positions (fig. S9E) where Asn is destabilizing [Acharya:2002vs, Fletcher:2012ei], and also because its other *a* and *d* positions are Leu and Ile respectively, which has been found to favor homotetramers [Fletcher:2012ei]. 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 non-canonical helical packing (fig. S9), permitting greater sequence diversity across larger interfaces.

Structural characterization

To assess the accuracy of the designs, we determined ten crystal structures spanning a range of oligomerization states, superhelical parameters, and hydrogen bond networks (Fig. 3A–F and fig. S10–S12). Designs for which crystals were not obtained were characterized by small angle x-ray scattering (SAXS) (Table S4, and Fig. 4, fig. S13–S14). 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. 3A,C–F) were also characterized by SAXS (fig. S14A), and the experimentally determined spectra were found to closely match those computed from the design models, suggesting that very similar structures are populated in solution.

The three left-handed trimer structures are remarkably similar to the design models with sub-angstrom RMSD across all backbone C α atoms and across all heavy atoms of the hydrogen bond networks (Fig. 3A–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 nine or twelve-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 resulted in substantially decreased thermostability (36), these two-ring trimers are stable up to 95°C and ~4.5M guanidinium chloride (fig. S3 and fig. S9) with numerous buried polar residues; 2L6HC3_13 has twelve completely buried asparagines, and 2L6HC3_6 has 24 buried polar residues confined to a small region of the interface, including six asparagines and six glutamines.

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 to 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 α atoms. The other two, 2L4HC2_11 (fig. S11A–B) and 2L4HC2_24 (fig. S11C–E), have slight structural deviations from the design models caused by water displacing designed network sidechains; in the former, the interface shifts ~2Å due to a buried water molecule bridging two network residues (fig. S11B), and in the latter, the backbone is nearly identical to the design model but sidechains 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 is in close agreement with the design model (fig. S14), but sidechain density was uncertain due to low (3.8Å) 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 quite 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). While 3L6HC2_2 was designed to

form a parallel dimer, its crystal structure revealed an antiparallel dimer interface, highlighting two design lessons (fig. S12): first, the importance of intermolecular hydrogen bonds at the binding interface (the 3L6HC2_2 design model has only two across the interface compared to 9 in 2L6HC3_6 (Fig. 3B)), and second, the importance of favorable hydrophobic contacts complementing the networks (the 3L6HC2_2 design model has mainly alanines at the interface).

SAXS data suggest that our untwisted dimer, trimer and tetramer designs assemble into the target triangular and square conformations (Fig. 4A–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, fig. S14). The SAXS data on the three untwisted tetramers were all in close agreement with the corresponding design models (Fig. 4A, fig. S14, and Table S4). 5L8HC4_6 has a distinctive network with a Trp making a buried hydrogen bond at one end of the network, which then propagates outwards towards solvent, connecting to an 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 α 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, with 0.41 Å and 0.48 Å RMSD over all network heavy-atoms. Like the supercoiled trimers, each of these networks contains sidechains from every helix, and helices were constructed to be uniformly symmetrical and equidistant (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/residue for the inner three helices and $\omega_0 = -0.137$ degrees/residues 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 sidechains contributing to the network. Networks with the highest connectivity and structural accuracy were those that spanned the entire cross-sectional interface, with each helix contributing at least one sidechain (Fig. 3AB,EF). Design 2L6HC3_13 also has two additional smaller networks consisting of a single symmetric Asn making two hydrogen bonds but with 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, yielding 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, eluting 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 2 ring structures than traditional single ring coiled coils, which have many fewer total hydrophobic residues and less inter-helical interface area).

We used an *in vivo* yeast-two-hybrid assay {Fields:1989dm} 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 such that binding of the designed protein interaction is necessary for cell growth. Designs in which the hydrogen bond networks partition hydrophobic interface area into relatively small regions are considerably 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 sidechain. 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 4 repeating cross-sections of the supercoil (Fig. 5C), we placed either the A, B, or X (Fig 4D) followed by the same design strategy and selection process as before. Six of these combinatorial designs were synthesized and 5/6 were found to be folded, thermostable, and assembled to the designed trimeric oligomerization state *in vitro* (Fig. 5F–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). 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{Kuhlman:2003kp, Koga:2012ej, King:2012gc, Huang:2007ge}. Unlike the multi-body problem of designing highly connected and satisfied hydrogen bond networks, hydrophobic packing is readily captured by established pairwise-decomposable potentials; consequently, most protein interface designs have been predominantly hydrophobic, and attempts to design buried hydrogen bonds across interfaces have routinely failed{Stranges:2012em}. Polar interfaces have been designed in specialized cases{Procko:2013eg, Joachimiak:2006dj, Kortemme:2004ia} but have been difficult to generalize, with many interface design efforts requiring directed evolution to optimize polar contacts and achieve desired specificity{Karanicolas:2011br, Fleishman:2011fx}. 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. Adding an outer ring of helices to enable hydrogen bond networks extends upon elegant studies from Keating, Woolfson, DeGrado and others demonstrating the designability of coiled coils with a wide range of hetero and homo-oligomeric specificities{Betz:1997gm, Grigoryan:2008ho, Grigoryan:2009iu, Reinke:2013jx, Fletcher:2012ei, Thomas:2013ks, Thomson:2014ks, Joh:2014jr}. The design models and crystal structures show that a wide range of hydrogen bond network composition and geometry are possible in repeating two-ring topologies, and that multiple networks can be engineered into the same backbone at varying positions without sacrificing thermostability, enabling 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{Rothmund:2006hb, Winfree:1998fs, Wei:2012ip}. 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.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Lauren Carter for assistance with SEC-MALS and protein production and Rosetta@Home volunteers for computing throughput for ab initio structure prediction. We thank Franziska Seeger for guidance with SAXS data analysis, and Matthew Bick and Peilong Lu for assistance with crystallographic refinement. This work was supported by the Howard Hughes Medical Institute, the Department of Energy, and the ERANET BioOrigami consortium, National Science Foundation (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 (ALS, Lawrence Berkeley National Laboratory, Berkeley, CA Department of Energy, contract no. DE-AC02-05CH11231); SAXS data was collected through the SIBYLS mail-in SAXS program under the aforementioned contract no., and we thank Kathryn Burnett and Greg Hura. The Berkeley Center for Structural Biology is supported in part by the National Institutes of Health, National Institute of General Medical Sciences, 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.G. are supported by Washington Research Foundation Innovation Postdoctoral Fellowships. David Baker, Scott Boyken, and Zibo Chen are inventors on a provisional application filed by the University of Washington. Coordinates and structure files have been deposited to the Protein Data Bank with accession codes: 5J0J (2L6HC3_6), 5J0I (2L6HC3_12), 5J0H (2L6HC3_13), 5J2S (5L6HC3_1), 5J73 (2L4HC2_9), 5J2L (2L4HC2_11), 5J0L (3L6HC2_2), 5J0K (2L4HC2_23), 5J10 (2L4HC2_24), and ____ (2L8HC4_12).

References

- Fleming PJ, Rose GD. Do all backbone polar groups in proteins form hydrogen bonds? *Protein Sci.* 2005; 14:1911–1917. [PubMed: 15937286]
- Chakrabarti R, Klibanov AM, Friesner RA. Sequence optimization and designability of enzyme active sites. *Proc Natl Acad Sci USA.* 2005; 102:12035–12040. [PubMed: 16103370]
- Judd ET, Stein N, Pacheco AA, Elliott SJ. Hydrogen bonding networks tune proton-coupled redox steps during the enzymatic six-electron conversion of nitrite to ammonia. *Biochemistry.* 2014; 53:5638–5646. [PubMed: 25137350]
- Sánchez-Azqueta A, et al. A hydrogen bond network in the active site of Anabaena ferredoxin-NADP(+) reductase modulates its catalytic efficiency. *Biochim Biophys Acta.* 2014; 1837:251–263. [PubMed: 24200908]
- Xu D, Tsai CJ, Nussinov R. Hydrogen bonds and salt bridges across protein-protein interfaces. *Protein Eng.* 1997; 10:999–1012. [PubMed: 9464564]
- Sheinerman FB, Honig B. On the role of electrostatic interactions in the design of protein-protein interfaces. *J Mol Biol.* 2002; 318:161–177. [PubMed: 12054776]
- Chen YJ, Groves B, Muscat RA, Seelig G. DNA nanotechnology from the test tube to the cell. *Nat Nanotechnol.* 2015; 10:748–760. [PubMed: 26329111]
- Zhang F, Nangreave J, Liu Y, Yan H. Structural DNA nanotechnology: state of the art and future perspective. *J Am Chem Soc.* 2014; 136:11198–11211. [PubMed: 25029570]
- Kuhlman B, Baker D. Native protein sequences are close to optimal for their structures. *Proc Natl Acad Sci USA.* 2000; 97:10383–10388. [PubMed: 10984534]
- Dahiyat BI, Mayo SL. Protein design automation. *Protein Sci.* 1996; 5:895–903. [PubMed: 8732761]
- Street AG, Mayo SL. Computational protein design. *Structure.* 1999; 7:R105–9. [PubMed: 10378265]
- Gruber M, Lupas AN. Historical review: another 50th anniversary—new periodicities in coiled coils. *Trends Biochem Sci.* 2003; 28:679–685. [PubMed: 14659700]
- Burkhard P, Stetefeld J, Strelkov SV. Coiled coils: a highly versatile protein folding motif. *Trends Cell Biol.* 2001; 11:82–88. [PubMed: 11166216]
- Huang PS, et al. High thermodynamic stability of parametrically designed helical bundles. *Science.* 2014; 346:481–485. [PubMed: 25342806]
- Ogihara NL, Weiss MS, DeGrado WF, Eisenberg D. The crystal structure of the designed trimeric coiled coil coil-VaLd: implications for engineering crystals and supramolecular assemblies. *Protein Sci.* 1997; 6:80–88. [PubMed: 9007979]
- Reinke AW, Grant RA, Keating AE. A synthetic coiled-coil interactome provides heterospecific modules for molecular engineering. *J Am Chem Soc.* 2010; 132:6025–6031. [PubMed: 20387835]

- Boyle AL, Woolfson DN. De novo designed peptides for biological applications. *Chem Soc Rev.* 2011; 40:4295–4306. [PubMed: 21373694]
- Grigoryan G, DeGrado WF. Probing designability via a generalized model of helical bundle geometry. *J Mol Biol.* 2011; 405:1079–1100. [PubMed: 20932976]
- Harbury PB, Plecs JJ, Tidor B, Alber T, Kim PS. High-resolution protein design with backbone freedom. *Science.* 1998; 282:1462–1467. [PubMed: 9822371]
- Woolfson DN, Alber T. Predicting oligomerization states of coiled coils. *Protein Sci.* 1995; 4:1596–1607. [PubMed: 8520486]
- Harbury PB, Zhang T, Kim PS, Alber T. A switch between two-, three-, and four-stranded coiled coils in GCN4 leucine zipper mutants. *Science.* 1993; 262:1401–1407. [PubMed: 8248779]
- Acharya A, Rishi V, Vinson C. Stability of 100 homo and heterotypic coiled-coil a-a' pairs for ten amino acids (A, L, I, V, N, K, S, T, E, and R). *Biochemistry.* 2006; 45:11324–11332. [PubMed: 16981692]
- Gradišar H, Jerala R. De novo design of orthogonal peptide pairs forming parallel coiled-coil heterodimers. *J Pept Sci.* 2011; 17:100–106. [PubMed: 21234981]
- Kaplan JB, Reinke AW, Keating AE. Increasing the affinity of selective bZIP-binding peptides through surface residue redesign. *Protein Science.* 2014; 23:940–953. [PubMed: 24729132]
- Baker EG, et al. Local and macroscopic electrostatic interactions in single α -helices. *Nat Chem Biol.* 2015; 11:221–228. [PubMed: 25664692]
- Eckert DM, Malashkevich VN, Kim PS. Crystal structure of GCN4-pIQI, a trimeric coiled coil with buried polar residues. *J Mol Biol.* 1998; 284:859–865. [PubMed: 9837709]
- Akey DL, Malashkevich VN, Kim PS. Buried polar residues in coiled-coil interfaces. *Biochemistry.* 2001; 40:6352–6360. [PubMed: 11371197]
- Lumb KJ, Kim PS. A buried polar interaction imparts structural uniqueness in a designed heterodimeric coiled coil. *Biochemistry.* 1995; 34:8642–8648. [PubMed: 7612604]
- Tatko CD, Nanda V, Lear JD, DeGrado WF. Polar networks control oligomeric assembly in membranes. *J Am Chem Soc.* 2006; 128:4170–4171. [PubMed: 16568959]
- Gonzalez L, Woolfson DN, Alber T. Buried polar residues and structural specificity in the GCN4 leucine zipper. *Nat Struct Biol.* 1996; 3:1011–1018. [PubMed: 8946854]
- Crick FHC. The packing of [alpha]-helices: simple coiled-coils. *Acta Cryst (1953).* Q6, 689–697. 1953; 6:1–9.
- Leaver-Fay A, et al. ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules. *Meth Enzymol.* 2011; 487:545–574. [PubMed: 21187238]
- O'Meara MJ, et al. A Combined Covalent-Electrostatic Model of Hydrogen Bonding Improves Structure Prediction with Rosetta. *J Chem Theory Comput.* 2015; 11:609–622. [PubMed: 25866491]
- Das R, et al. Simultaneous prediction of protein folding and docking at high resolution. *Proc Natl Acad Sci USA.* 2009; 106:18978–18983. [PubMed: 19864631]
- 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.
- Acharya A, Ruvinov SB, Gal J, Moll JR, Vinson C. A heterodimerizing leucine zipper coiled coil system for examining the specificity of a position interactions: amino acids I, V, L, N, A, and K. *Biochemistry.* 2002; 41:14122–14131. [PubMed: 12450375]
- Fletcher JM, et al. A basis set of de novo coiled-coil peptide oligomers for rational protein design and synthetic biology. *ACS Synth Biol.* 2012; 1:240–250. [PubMed: 23651206]
- Fields S, Song O. A novel genetic system to detect protein-protein interactions. *Nature.* 1989; 340:245–246. [PubMed: 2547163]

- Kuhlman B. Design of a Novel Globular Protein Fold with Atomic-Level Accuracy. *Science*. 2003; 302:1364–1368. [PubMed: 14631033]
- Koga N, et al. Principles for designing ideal protein structures. *Nature*. 2012; 491:222–227. [PubMed: 23135467]
- King NP, et al. Computational design of self-assembling protein nanomaterials with atomic level accuracy. *Science*. 2012; 336:1171–1174. [PubMed: 22654060]
- Huang PS, Love JJ, Mayo SL. A de novo designed protein protein interface. *Protein Sci*. 2007; 16:2770–2774. [PubMed: 18029425]
- Stranges PB, Kuhlman B. A comparison of successful and failed protein interface designs highlights the challenges of designing buried hydrogen bonds. *Protein Science*. 2012; doi: 10.1002/pro.2187
- Procko E, et al. Computational design of a protein-based enzyme inhibitor. *J Mol Biol*. 2013; 425:3563–3575. [PubMed: 23827138]
- Joachimiak LA, Kortemme T, Stoddard BL, Baker D. Computational design of a new hydrogen bond network and at least a 300-fold specificity switch at a protein-protein interface. *J Mol Biol*. 2006; 361:195–208. [PubMed: 16831445]
- Kortemme T, et al. Computational redesign of protein-protein interaction specificity. *Nat Struct Mol Biol*. 2004; 11:371–379. [PubMed: 15034550]
- Karanicolas J, et al. A de novo protein binding pair by computational design and directed evolution. *Mol Cell*. 2011; 42:250–260. [PubMed: 21458342]
- Fleishman SJ, et al. Computational design of proteins targeting the conserved stem region of influenza hemagglutinin. *Science*. 2011; 332:816–821. [PubMed: 21566186]
- Betz SF, Liebman PA, DeGrado WF. De novo design of native proteins: characterization of proteins intended to fold into antiparallel, rop-like, four-helix bundles. *Biochemistry*. 1997; 36:2450–2458. [PubMed: 9054549]
- Grigoryan G, Keating AE. Structural specificity in coiled-coil interactions. *Curr Opin Struct Biol*. 2008; 18:477–483. [PubMed: 18555680]
- Grigoryan G, Reinke AW, Keating AE. Design of protein-interaction specificity gives selective bZIP-binding peptides. *Nature*. 2009; 458:859–864. [PubMed: 19370028]
- Reinke AW, Baek J, Ashenberg O, Keating AE. Networks of bZIP protein-protein interactions diversified over a billion years of evolution. *Science*. 2013; 340:730–734. [PubMed: 23661758]
- Thomas F, Boyle AL, Burton AJ, Woolfson DN. A set of de novo designed parallel heterodimeric coiled coils with quantified dissociation constants in the micromolar to sub-nanomolar regime. *J Am Chem Soc*. 2013; 135:5161–5166. [PubMed: 23477407]
- Thomson AR, et al. Computational design of water-soluble α -helical barrels. *Science*. 2014; 346:485–488. [PubMed: 25342807]
- Joh NH, et al. De novo design of a transmembrane Zn^{2+} -transporting four-helix bundle. *Science*. 2014; 346:1520–1524. [PubMed: 25525248]
- Rothmund PWK. Folding DNA to create nanoscale shapes and patterns. *Nature*. 2006; 440:297–302. [PubMed: 16541064]
- Winfree E, Liu F, Wenzler LA, Seeman NC. Design and self-assembly of two-dimensional DNA crystals. *Nature*. 1998; 394:539–544. [PubMed: 9707114]
- Wei B, Dai M, Yin P. Complex shapes self-assembled from single-stranded DNA tiles. *Nature*. 2012; 485:623–626. [PubMed: 22660323]
- Schneidman-Duhovny D, Hammel M, Sali A. FoXS: a web server for rapid computation and fitting of SAXS profiles. *Nucleic Acids Res*. 2010; 38:W540–4. [PubMed: 20507903]
- Schneidman-Duhovny D, Hammel M, Tainer JA, Sali A. Accurate SAXS profile computation and its assessment by contrast variation experiments. *Biophys J*. 2013; 105:962–974. [PubMed: 23972848]
- Jacak R, Leaver-Fay A, Kuhlman B. Computational protein design with explicit consideration of surface hydrophobic patches. *Proteins*. 2012; 80:825–838. [PubMed: 2223219]
- Dimaio F, Leaver-Fay A, Bradley P, Baker D, André I. Modeling symmetric macromolecular structures in Rosetta3. *PLoS ONE*. 2011; 6:e20450. [PubMed: 21731614]

- Fleishman SJ, et al. RosettaScripts: A Scripting Language Interface to the Rosetta Macromolecular Modeling Suite. *PLoS ONE*. 2011; 6:e20161. [PubMed: 21731610]
- Liu P, Agrafiotis DK, Theobald DL. Fast determination of the optimal rotational matrix for macromolecular superpositions. *J Comput Chem*. 2010; 31:1561–1563. [PubMed: 20017124]
- Gish W, States DJ. Identification of protein coding regions by database similarity search. *Nat Genet*. 1993; 3:266–272. [PubMed: 8485583]
- Minami S, Sawada K, Chikenji G. MICAN: a protein structure alignment algorithm that can handle Multiple-chains, Inverse alignments, C(α) only models, Alternative alignments, and Non-sequential alignments. *BMC Bioinformatics*. 2013; 14:24. [PubMed: 23331634]
- Gibson DG, et al. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods*. 2009; 6:343–345. [PubMed: 19363495]
- Otwinowski Z, Minor W Jr, W CC. Processing of X-ray Diffraction Data Collected in Oscillation Mode, *Methods in Enzymology* Vol. 276 Part A. 1997; 276:307–326.
- McCoy AJ, et al. Phaser crystallographic software. *Journal of applied crystallography*. 2007; 40:658–674. [PubMed: 19461840]
- Adams PD, et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr*. 2010; 66:213–221. [PubMed: 20124702]
- Afonine PV, et al. Joint X-ray and neutron refinement with phenix.refine. *Acta Crystallogr D Biol Crystallogr*. 2010; 66:1153–1163. [PubMed: 21041930]
- Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr*. 2004; 60:2126–2132. [PubMed: 15572765]
- Davis IW, et al. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res*. 2007; 35:W375–83. [PubMed: 17452350]
- Dyer KN, et al. High-throughput SAXS for the characterization of biomolecules in solution: a practical approach. *Methods Mol Biol*. 2014; 1091:245–258. [PubMed: 24203338]
- Rambo RP, Tainer JA. Characterizing flexible and intrinsically unstructured biological macromolecules by SAS using the Porod-Debye law. *Biopolymers*. 2011; 95:559–571. [PubMed: 21509745]
- Huang PS, et al. RosettaRemodel: A Generalized Framework for Flexible Backbone Protein Design. *PLoS ONE*. 2011; 6:e24109. [PubMed: 21909381]
- Chien CT, Bartel PL, Sternglanz R, Fields S. The two-hybrid system: a method to identify and clone genes for proteins that interact with a protein of interest. *Proc Natl Acad Sci USA*. 1991; 88:9578–9582. [PubMed: 1946372]
- Bartel PL, Roecklein JA, SenGupta D, Fields S. A protein linkage map of *Escherichia coli* bacteriophage T7. *Nat Genet*. 1996; 12:72–77. [PubMed: 8528255]
- Eng JK, McCormack AL, Yates JR. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *Journal of the American Society for Mass Spectrometry*. 1994; 5:976–989. [PubMed: 24226387]
- Käll L, Canterbury JD, Weston J, Noble WS, MacCoss MJ. Semi-supervised learning for peptide identification from shotgun proteomics datasets. *Nat Methods*. 2007; 4:923–925. [PubMed: 17952086]
- Zhang B, Chambers MC, Tabb DL. Proteomic parsimony through bipartite graph analysis improves accuracy and transparency. *J Proteome Res*. 2007; 6:3549–3557. [PubMed: 17676885]
- The PyMOL Molecular Graphics System, version 1.7.4 (Schrödinger, LLC).

One Sentence Summary

We designed a new class of protein oligomers with regular arrays of hydrogen bond networks that enable programming of interaction specificity.

Fig. 1. Overview of the HBNet method and design strategy

(A) (left) All sidechain 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-D) HBNet. (B) HBNet precomputes the hydrogen bond and steric repulsive interaction energies between sidechain 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 sidechain rotamers at Res_i and Res_j make good-geometry hydrogen bonds, but graph traversal shows that only one of these (*left*) can be extended into a connected network: **(C)** Res_i rotamer 3 ($i:3$) can also hydrogen bond to Res_k rotamer 2 ($k:2$) and Res_j 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 , leaving unsatisfied buried polar atoms. **(E-G) Design strategy:** **(E)** Parametric backbone generation of two-ring coiled coils: a C_3 symmetric trimer is shown, colored by monomer subunit, labeled with parameters sampled: supercoil radius of inner (R_{in}) and outer (R_{out}) helices, helical phase of the inner ($\phi_{1\text{ in}}$) and outer ($\phi_{1\text{ out}}$) helices, supercoil phase of the outer helix (ϕ_0), z-offset between the inner and outer helices (Z_{off}), and the supercoil twist (ω_0). **(F)** HBNet is applied to parametric backbones to identify the best hydrogen bond networks. **(G)** Networks are maintained while remaining residue positions are designed in context of the assembled symmetric oligomer.

Fig. 2. The outer ring of helices increase thermostability and can overcome poor helical propensity of the inner helices

(A) CD spectrum (260–195nm) of design 2L4HC2_23 at 25°C (blue), 75°C (red), 95°C (green), and 25°C after cooling (purple). (B) 2L4HC2_23, denaturation by GdmCl monitoring 222nm; (C) 2L4HC2_9, a supercoiled C2 homodimer colored by chain, looking 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 significantly more stable than 2L4HC2_9_inner (gray). (G) 2L6HC3_13, a supercoiled C3 homotrimer. (H) CD spectrum of 2L6HC3_13 at different temperatures as in (A). (I) 2L6HC3_13_inner. (J) CD spectrum of 2L6HC3_13 (black) versus 2L6HC3_13_inner (gray) shows that the inner helix by itself is primarily unfolded. All CD data is plotted in Mean Residue Ellipticity (MRE) $10^3 \text{ deg cm}^2 \text{ dmol}^{-1}$.

Fig. 3. Structural characterization by x-ray crystallography

(A–F) Crystal structures (white) are superimposed onto the design models (colored by chain ID: 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); panel 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 = 0.51 Å over all Ca atoms) and (B) 2L6HC3_6

(2.26 Å resolution; RMSD = 0.77 Å over all Cα 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 sidechains. **(D)** 2L4HC2_9 (2.56 Å resolution; 0.39 Å RMSD over all Cα atoms) and **(E)** 2L4HC2_23 (1.54 Å resolution; RMSD = 1.16 Å over all Cα atoms) are left-handed C2 homodimers, each with one network. **(F)** 5L6HC3_1 (2.36 Å resolution; RMSD = 0.51 Å over all Cα atoms) is a C3 homotrimer with straight, untwisted helices and two identical networks at different cross-sections. **(G, 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.

Fig. 4. Structural characterization by small angle x-ray scattering (SAXS)

(*left*) backbones and (*middle*) h-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* {SchneidmanDuhovny:2010km, SchneidmanDuhovny:2013er}; Chi2 values of fit (X) indicated inside each panel. **(A)** 5L8HC4_6 ($X=1.36$), an untwisted C_4 homotetramer with two identical h-bond networks. **(B)** 5L4HC2_12 ($X=1.45$), an untwisted C_2 homodimer with a single h-bond network. **(C)** 3L6HC2_4 ($X=2.04$), a parallel right-handed C_2 homodimer with two repeated networks and two inner helices, one outer helix. **(D)** 2L6Hanti_3 ($X=1.80$), a left-handed antiparallel homodimer with two inner helices, one outer helix; because of the anti parallel geometry, the same network occurs in two locations.

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 h-patch{Jacak:2012ip}; 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 Activation domain constructs and binding measured by determining the cell growth rate (maximum $\Delta OD/\text{hour}$); darker cells indicate more rapid growth, hence stronger binding; values are the average of at least 3 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–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 "alphabet" 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 and estimated molecular weights (from

MALS); designs range from ~27–30 kDa. AAXX, XXBB, and XXXX correspond to designs 2L6HC3_13, 2L6HC3_6, and 2L6HC3_hydrophobic_1, respectively.