PNAS RESEARCH ARTICLE | BIOCHEMISTRY BIOPHYSICS AND COMPUTATIONAL BIOLOGY OPEN ACCESS

De novo design of protein homodimers containing tunable symmetric protein pockets

Derrick R. Hicks[a,b], Madison A. Kennedy[a,c], Kirsten A. Thompson[a,b], Michelle DeWitt[a,b], Brian Coventry[a,b,d], Alex Kang[a,b], Asim K. Bera[a,b], T. J. Brunette[a,b], Banumathi Sankaran[e], Barry Stoddard[c], and David Baker[a,b,d,1]

Edited by Stephen Mayo, California Institute of Technology, Pasadena, CA; received July 20, 2021; accepted May 4, 2022

Function follows form in biology, and the binding of small molecules requires proteins with pockets that match the shape of the ligand. For design of binding to symmetric ligands, protein homo-oligomers with matching symmetry are advantageous as each protein subunit can make identical interactions with the ligand. Here, we describe a general approach to designing hyperstable C2 symmetric proteins with pockets of diverse size and shape. We first designed repeat proteins that sample a continuum of curvatures but have low helical rise, then docked these into C2 symmetric homodimers to generate an extensive range of C2 symmetric cavities. We used this approach to design thousands of C2 symmetric homodimers, and characterized 101 of them experimentally. Of these, the geometry of 31 were confirmed by small angle X-ray scattering and 2 were shown by crystallographic analyses to be in close agreement with the computational design models. These scaffolds provide a rich set of starting points for binding a wide range of C2 symmetric compounds.

protein design | homodimer | scaffold | repeat protein | symmetry

Significance

Proteins capable of binding arbitrary small molecules could enable the generation of new biosensors or medicines. While considerable progress has been made in recent years to design proteins from scratch capable of binding asymmetric molecules, little work has been done to facilitate the binding of symmetric molecules. Here, we present a method for generating libraries of C2 symmetric proteins with diverse central cavities that could be functionalized in the future to bind a range of C2 symmetric small molecules for applications such as ligand controllable cell engineering. We show that 31% of our designed proteins fold to the desired quaternary state, when experimentally characterized, and are hyperstable.

Cyclic two-fold (C2) symmetric molecules are common in biology and medicine, such as HIV protease inhibitors (1), iron sulfur clusters (2), and the chlorophyll special pair found in photosynthetic reaction centers (3). To bind such compounds, C2 symmetric protein homodimers are advantageous because each protein monomer can make identical interactions with an asymmetric unit of the small molecule. There are many C2 symmetric protein structures in nature, but it is not straightforward to re-engineer them to bind arbitrary C2 symmetric small molecules unless they happen to contain an interior cavity with the correct size and shape. 4-Helix bundles have been engineered to bind chlorophyll dimers (4), di-nuclear metals (5), and iron-sulfur clusters (6), and binding of a C3 symmetric molecule has been achieved using C3 symmetric helical bundles (7). However, the size and shape of binding pockets available in the interior of helical bundles is limited, as interactions between the helices are also important for stability, and hence they cannot be too far apart. A large set of protein scaffolds with C2 symmetric binding pockets spanning a wide range of sizes and shapes that can be functionalized without compromising stability could enable the creation of new enzymes, therapeutics, and light harvesting proteins, but neither such sets nor methods to generate them currently exist.

We set out to develop a general solution to the challenge of creating scaffold proteins for binding C2 symmetric ligands. Our approach builds on recent work in the field of de novo protein design centered around robust alpha-helical repeat proteins (8, 9) that have been adapted to create higher order cyclic oligomers (10) and nanocages (11), chemically induced protein switches (12), and proteins binding specific mineral surfaces. In particular, circular tandem repeat proteins, or "toroids," have been designed in which the curvature of the repeat proteins is chosen to enable closure into ring-like structures with a large circular central cavity (9). We aimed to design repeat proteins that could similarly house a central cavity, but with a wide range of elliptical, rather than perfectly circular, shapes to enable binding of a wider range of C2 symmetric ligands. We chose an overall design architecture consisting of repeat proteins which curve around a central axis that are docked into C2 symmetric homodimers surrounding an elliptical central cavity (Fig. 1*A*). By employing repeat proteins with minimal rise along the superhelical axis from one repeat unit to the next, we favor C2 arrangements with the ends of the two monomers in contact. Advantages of this conception are that the cavities can be vastly diverse in size, shape, and chemical composition (lined with different sidechain functional groups). Additionally, as the cavity lining residues are on the exterior of the monomers, the protein hydrophobic core is separated from the binding pocket; as such, functionalization to create binding interactions for specific compounds is unlikely to destabilize either the monomers or the dimer interface.

Author contributions: D.R.H., M.A.K., K.A.T., B.C., B. Stoddard, and D.B. designed research; D.R.H., M.A.K., K.A.T., M.D., B.C., A.K., A.K.B., T.J.B., B. Sankaran, and B. Stoddard performed research; D.R.H. contributed new reagents/analytic tools; D.R.H., M.A.K., K.A.T., M.D., B.C., A.K., A.K.B., T.J.B., B. Sankaran, and B. Stoddard analyzed data; and D.R.H. and D.B. wrote the paper.

Competing interest statement: Provisional patent applications have been filed for proteins in this work.

This article is a PNAS Direct Submission.

Copyright © 2022 the Author(s). Published by PNAS. This open access article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

[1]To whom correspondence may be addressed. Email: dabaker@uw.edu.

This article contains supporting information online at http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2113400119/-/DCSupplemental.

Published July 21, 2022.

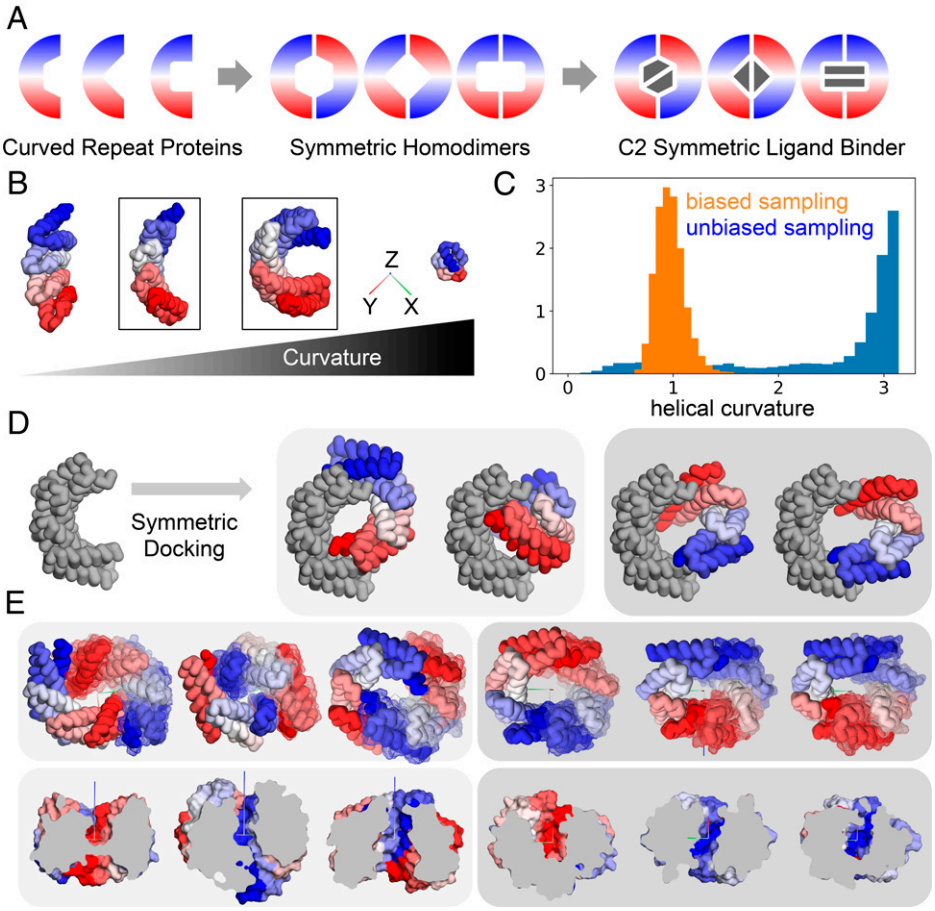

Fig. 1. Design strategy. (*A*) Schematic of design pipeline from curved repeat protein (*Left*) to symmetric homodimers (*Center*) to C2 symmetric ligand binders (*Right*). Color gradient represents the protein chain direction from N terminus (blue) to C terminus (red). Hypothetical C2 symmetric ligands are shown in gray. (*B*) Examples of repeat proteins sampling different curvatures. The helical symmetry axis of the proteins is aligned to the *z* axis, and an *xyz* axis is depicted to show that we are looking directly down the *z* axis. (*C*) A histogram of helical curvature for 1 million backbones made with our biased sampling method (orange) or without (blue). (*D*) A single monomer (*Left*) can be docked into various rigid body orientations to create homodimers (*Right*) with diverse central cavities. This docking approach can create head to tail (light gray box) or head to head + tail to tail (dark gray box) homodimers. (*E*) Examples of homodimers featuring a range of cavity sizes and shapes. Each dimer is based on a different monomeric curved repeat protein. The top row shows proteins looking down the central cavity, depicted as backbone ribbon representation for both chains and with surface mesh on chain B. The *Bottom* row shows a side view slicing through the protein to illustrate the shape of the central cavity, depicted as surface representation. Head to tail homodimers are shown in a light gray box and head to head + tail to tail homodimers are shown in a dark gray box. The C2 symmetry axis of the protein homodimer is indicated with a blue axis.

To implement this strategy to design C2 symmetric protein homodimers containing central cavities, we first set out to create a diverse library of monomeric units to dock into various symmetric homodimer orientations. Previous efforts to design diverse repeat proteins (8) did not produce many structures with shapes suitable to create the C2 symmetric homodimers with the elliptical central cavities that we envisioned. Because of the lack of existing protein monomers, we began by generating a set of helical repeat protein monomers with structures specifically tailored for building C2 symmetric binding pockets. We selected a range of superhelical curvature, rise, and radius parameters, such that a four-unit repeat protein would approximate a half-circle and the resulting dimer would form an ellipse. Superhelical curvature and rise correspond to the rotation around and translation along the central superhelical axis per repeat unit respectively, and radius is the distance of the protein from this axis; these quantities are calculated from the center of mass of one repeat unit to the center of mass of the next repeat unit (see *SI Appendix*, Fig. S1 for schematic) using the RepeatParameterFilter within the Rosetta macromolecular modeling suite (13). Model building to approximate a half circle suggested the rotation between each repeat should be between 0.7 rad and 1.1 rad, the rise less than 1.5 Å per repeat, and the radius between 10 Å and 22 Å. We hypothesized that, when docked into dimers, proteins with these parameters would create pockets that could accommodate ligands of diverse sizes and shapes. Fig. 1*D* and *E* and *SI Appendix*, Fig. S2 illustrate how radius, curvature, and rise control the shape of repeat proteins and highlight the type of monomeric proteins we aimed to make.

Repeat protein design as previously described (8) relies on Rosetta Monte Carlo fragment assembly approaches to explore repeat protein space (14). Unbiased sampling rarely yields repeat proteins with our desired helical parameters and only in cases where the lengths of the two helices in a repeat differ by 6–7 residues (*SI Appendix*, Fig. S3). We thus developed methods for biasing fragment assembly toward desired regions of repeat protein parameter space; at each fragment insertion (made identically in each repeat unit), the deviation from a target set of superhelical parameters is computed and the sum of these deviations is added to the coarse-grained score function previously used. With this biased assembly protocol, we were able to focus sampling on repeat protein structures with the desired superhelical parameters at all combinations of helix length (see Fig. 1*C* and *SI Appendix*, Figs. S3 and S4); almost

all trajectories with the biased fragment assembly protocol generate proteins with superhelical parameters in the desired range (see *SI Appendix*, Fig. S3). After filtering out poorly packed structures (*SI Appendix*, Fig. S5), we obtained far more curved repeat proteins at all helix length combinations using the biased fragment assembly protocol (see *SI Appendix*, Fig. S3; there were a higher fraction of poorly packed structures with the helical parameter bias terms on, and a lower weight on these terms could increase efficiency). We were able to generate 100,000 curved repeat protein backbones to use for subsequent design, a large increase over the handful of curved repeat proteins previously published (8).

These backbones were subjected to combinatorial sequence optimization using a RosettaScripts FastDesign protocol with repeat protein symmetry (applied through the RepeatProteinRelaxMover), which makes identical moves to each repeat unit during sequence design and minimization. The designs were then extended or shortened by up to half a repeat unit based on the energy per residue of the terminal helix to eliminate terminal helices that may be disordered due to limited contacts to the rest of the structure. The top 12,000 designs based on a combination of energy, packing, and sequence-structure agreement, were submitted for Ab Initio folding simulations. Designs for which the sequence strongly encoded the structure in de novo structure prediction calculations (see *Methods*; the lowest energy structures are close to the designed structure), 2,500 in total, were used in subsequent docking and design calculations.

We next set out to create C2 symmetric homodimers with central cavities using these 2,500 curved repeat proteins. We extended a previous symmetric docking approach (10) by adding a requirement that the docks create one of two classes of closed circular structure with either two N to C-terminal interfaces (head to tail dimer) or both N to N and C to C-terminal interfaces (head to head + tail to tail dimer). This docking protocol generated 1 million docked structures. We subsequently removed docks that had small interfaces (less than 10 contacting residues) that were likely to form weak interfaces, as well as docks that had excessively large interfaces (greater than 24 contacts) that could lead to poor behavior before dimerization due to having many exposed hydrophobic residues. This yielded a set of about 100,000 docks for both classes of homodimers that were subjected to interface sequence optimization using a RosettaScripts FastDesign protocol (15) with C2 symmetry (16). Fig. 1*D* shows how a single monomer can be docked into many distinct orientations creating diverse central cavities, and Fig. 1*E* shows examples of the diversity of proteins and pockets that can be achieved by docking diverse monomers into various C2 symmetric orientations. The top 1,200 (head to tail) and 2,000 (head to tail + tail to tail) homodimer designs were selected based on a combination of interface energy, interface shape complementarity, and buried unsatisfied hydrogen bonds (17) for both classes of closed circular homodimers (see *SI Appendix*, Fig. S6 for design flowchart).

With the ability to generate these proteins computationally, we set out to characterize a diverse set of examples (see *SI Appendix*, Fig. S7) with pockets varying in volume and shape (see *SI Appendix*, Fig. S8), approximated through the three principal axes, which were calculated on poly-alanine backbones to represent the maximum possible size of the pockets. In total, we characterized 101 designs including 77 head to tail dimers and 24 head to head + tail to tail dimers. Forty-four of the designs expressed enough soluble, well-behaved, protein for further characterization. Five of these designs were determined to be soluble aggregate by subsequent analysis. Of the 39 remaining proteins, 38 were characterized by circular dichroism (CD) (18), and of these, 36 were found to be helical and hyperstable (maintaining 80% helicity on average at 95 °C). All 36 of these proteins had nearly identical CD spectrums upon cooling back to 25 °C (data for seven designs are shown in Fig. 2). One design characterized by CD appeared helical as expected, but was not hyperstable, while another had low helical signal.

We used small angle X-ray scattering (SAXS) (19, 20) to characterize the 37 designs that appeared helical by CD (including the one with low stability); of these, 31 had experimental SAXS profiles that closely matched profiles predicted for the corresponding design model (19, 21), suggesting that they have the correct ellipsoidal shape in solution (the first six rows of Fig. 2 show representative examples). Taken together, the CD and SAXS data suggest that 31% of designs (31 of 101) are well-expressing soluble dimers with shapes consistent with the design models. See *SI Appendix*, Table S1 for characterization of all 101 designs.

We determined crystal structures for three designs (see Fig. 3). One of these, design D_3_337, is hyperstable and helical by CD, but the SAXS data suggests it dimerizes to a different shape than designed (Fig. 2, bottom row; Table S1). For this design, the monomer rather than the dimer crystalized (Fig. 3*C*; Protein Data Bank PDB: 7RMY); the Ca rmsd to the design model of the monomer structure is 2.75 Å, demonstrating control over the shape of repeat proteins in the absence of the designed protein-protein interfaces. In the crystal structure, lattice contacts are formed from the hydrophobic residues intended to form the homodimer interface.

Two of the designs, D_3_212 (Fig. 3*A*) and D_3_633 (Fig. 3*B*), fold and assemble to the desired dimeric ellipsoidal architecture with central cavities along the axis of symmetry, albeit with some deviations between the experimentally determined structures and design models (Ca rmsd of 2.59 Å and 2.04 Å, respectively). D_3_212 (PDB: 7RMX) is formed from a 224 amino acid-long monomer with helical rise, radius, and curvature of 0.23 Å, 19.6 Å, and 0.73 rad, respectively, that drifted to 1.6 Å, 19.5 Å, and 0.77 rad as the backbone adjusted during subsequent sequence design steps. This near-ideal repeat protein has helix lengths of 19 and 27 amino acids, which was found to be a near optimal combination to create curved repeat proteins. The D_3_212 homodimer has a central cavity with maximum (calculated without sidechains) height and width of 29 Å and 37 Å and a volume of 3,132 $Å^3$. In comparison, D_3_633 (PDB: 7RKC) is formed from a 228 amino acid-long monomer with helical rise, radius, and curvature of 0.68 Å, 20.1 Å, and 0.75 rad, respectively, that drifted to 2.2 Å, 22.1 Å, and 0.68 rad during sequence design. The D_3_633 design has repeating helix lengths of 24 and 29 amino acid that surround a central cavity with maximum height and width of 28 Å and 37 Å and a volume of 4,062 $Å^3$. In order to test the strength of the protein-protein interface of this structure, We performed a 500× dilution series from 80 μM to 160 nM (see *SI Appendix*, Fig. S12), and found that it remained 100% dimeric by SEC.

Discussion

The C2 symmetric homodimer proteins created in this study have central cavities with diverse shapes that can accommodate a range of C2 symmetric ligands. The designs have high thermal stability and solubility in a range of buffer conditions; many remained soluble for several months at 100 mg/ml in diverse crystal screening buffers. Because the protein core is distinct from the pocket, they should have high mutation

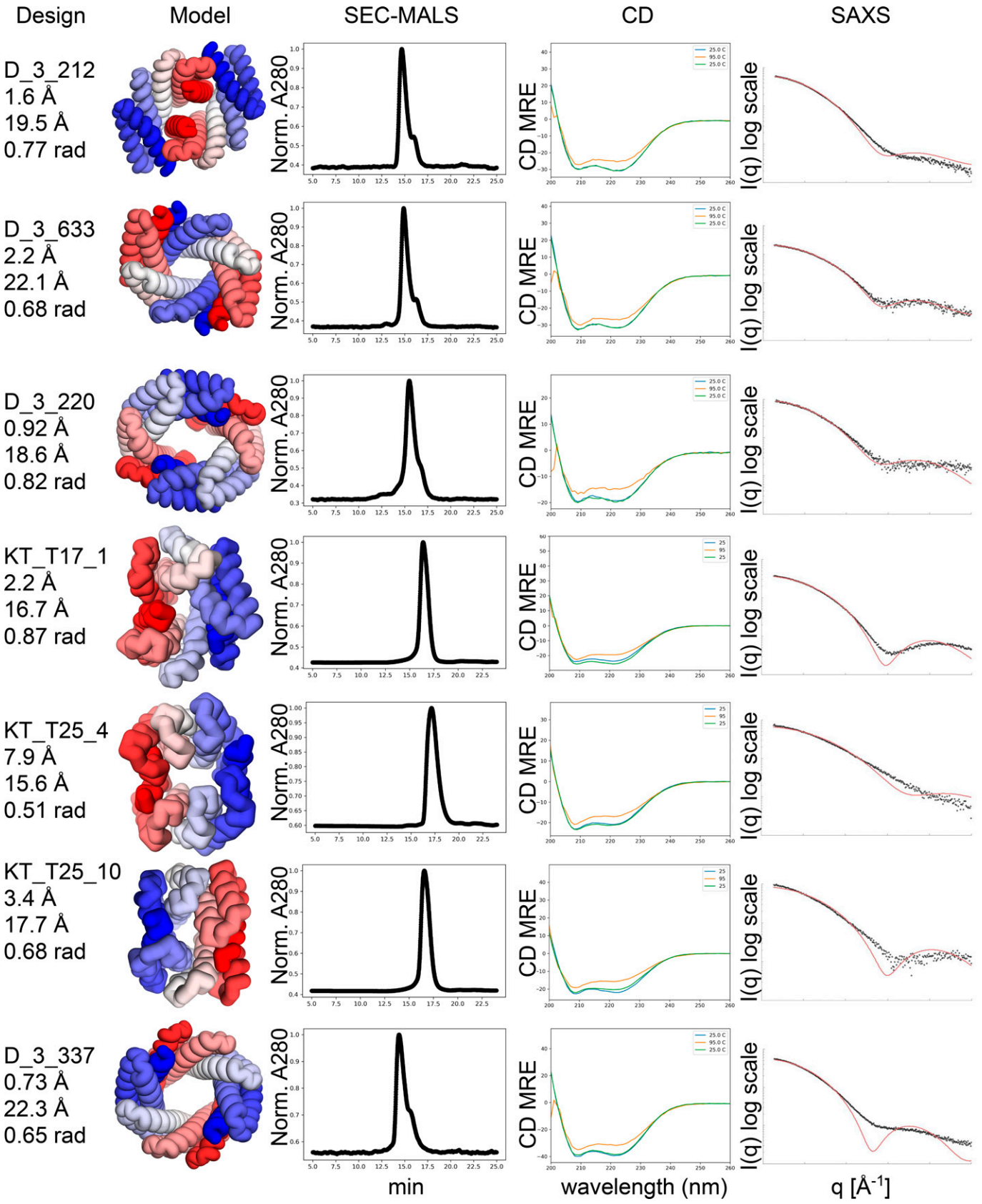

Fig. 2. Biophysical characterization. *Left*: design names along with the helical parameters (rise, radius, and curvature) of the associated protein monomer. Second column; design models depicted as ribbon backbones colored from blue (N terminus) to red (C terminus). Third column; normalized ultraviolet absorbance (A280) obtained during SEC-MALS, followed by circular dichroism scans from 200 to 260 nm at 25 °C, 95 °C, and 25 °C post heating. Fourth column; predicted SAXS profiles overlaid on experimental SAXS data points for scattering vector (q, from 0 to 0.25) vs. intensity (I).

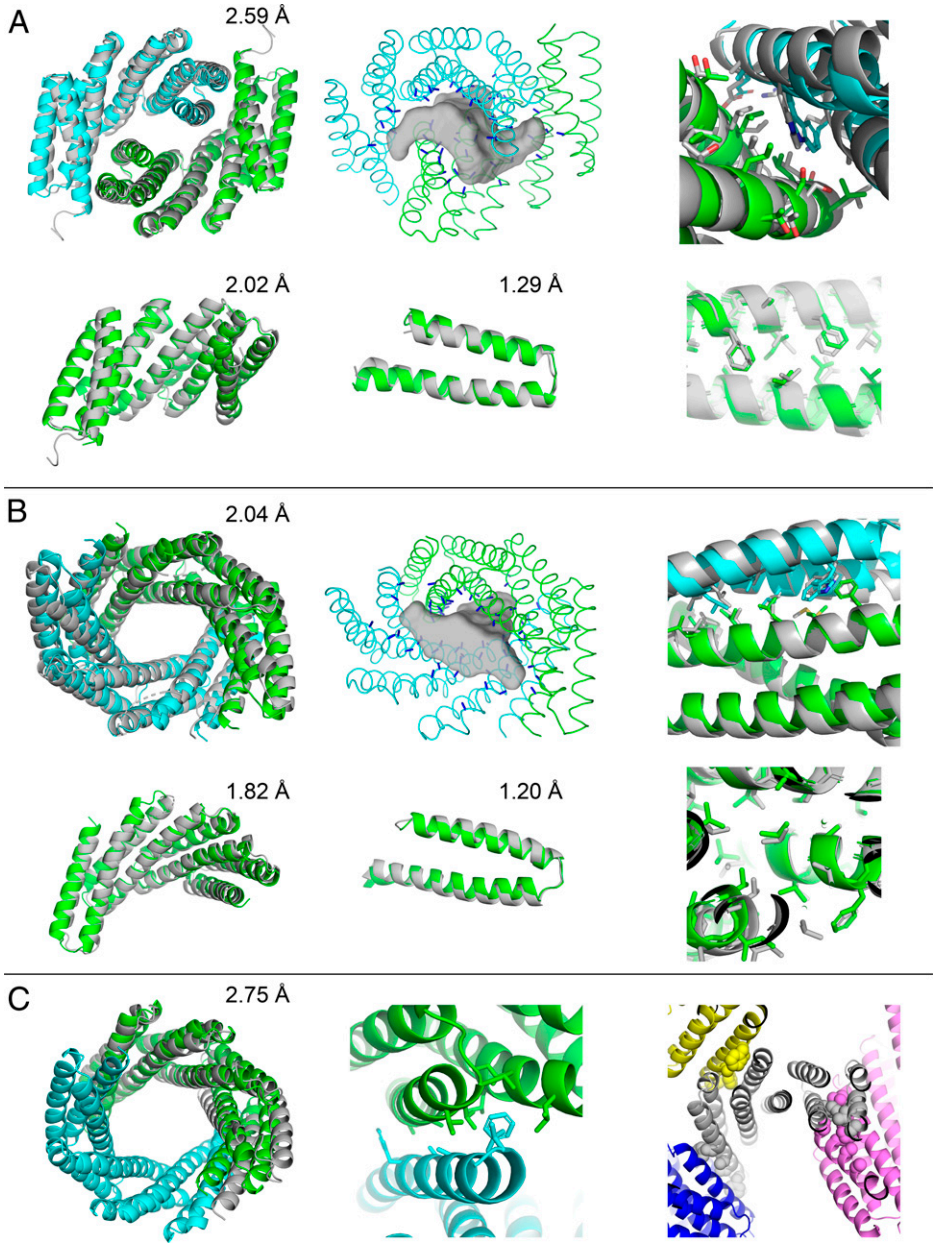

Fig. 3. Crystallographic analysis. (*A*) Overlay of D_3_212 (PDB: 7RMX) design model (green and cyan) and crystal structure (gray). *Top* from *Left* to *Right*: superposition of the homodimer, potential binding cavity (gray surface) with functionalizable sidechain positions highlighted in blue, superposition of protein–protein interface with sidechain residues shown as sticks (oxygen atoms are red, nitrogen atoms are blue); *Bottom* from *Left* to *Right*: superposition of the monomer, a repeat unit, and a section of the hydrophobic core with sidechain residues shown as sticks. Associated rmsds are indicated. (*B*) Overlay of design D_3_633 (PDB: 7RKC) design model (green and cyan) and crystal structure (gray); panels are as in (*A*). (*C*) D_3_337 design model and crystal structure. *Left*: overlay of design D_3_337 (PDB: 7RMY) design model (green and cyan) and monomer crystal structure (gray). *Middle*: designed homodimer interface, with hydrophobic residues shown as sticks and the two chains colored green and cyan. *Right*: crystal structure showing the central asymmetric unit in gray and its crystal lattice neighbors colored blue, pink, and yellow. The hydrophobic residues which were intended to form the homodimer interface are shown in spheres forming key crystal contacts.

tolerance during functionalization. The methods described here enable focused sampling of repeat protein conformational space beyond perfectly closing toroid structures; we used these methods to create a library of curved repeat proteins, but they could also be used to create a library of perfectly flat repeat proteins or to match the helical parameters of DNA or other helical biomolecules.

The dimeric ellipsoidal architecture of the proteins created here leads to a wide range of pocket sizes and geometries. A similar approach could be applied to the generation of higher order symmetric complexes to create pockets suitable for binding higher order symmetric molecules. Furthermore, many of the head to tail C2 symmetric designs described here could be connected into single chain proteins with short structured loops or flexible linkers to enable design of asymmetric small molecule binding and catalytic sites. The number of design models presented here already rivals the size of common fold classes found in the Protein Data Bank, and there is nearly unlimited ability to create more. Docking C2 symmetric compounds into these scaffolds can be carried out efficiently by superimposing the symmetry axes of the small molecule and protein scaffold, sampling the two remaining rigid body degrees of freedom (the translation along and rotation around the symmetry axes), and, for each dock, designing the protein interface to maximize

interactions with the ligand. We are currently exploring this approach for designing binders to a variety of C2 compounds, which ultimately could be useful for ligand induced dimerization for biological control among other applications.

Methods

Genes were ordered from IDT with N-terminal his-tags and cloned into pET-29b+ between NdeI/XhoI restriction sites. Proteins were expressed in *E. coli* using autoinduction and purified by IMAC followed by SEC. CD measurements were conducted on an AVIV Model 420 DC or Jasco J-1500 CD spectrometer with proteins at 0.25 mg/mL SEC-MALS was conducted as described previously (10) Small-angle X-ray scattering (SAXS) was collected at the SIBYLS High Throughput SAXS Advanced Light Source in Berkeley, California (20). Crystal screening was performed using Mosquito Crystal by STP Labtech, and data were collected on ALS beamline 8.2.1. Please see *SI Appendix* for more detailed wet laboratory methods.

Proteins were designed using the Rosetta macromolecular modeling suite (13) Backbones were generated with RosettaRemodel using a coarse-grained energy function supplemented with scoring terms that bias the trajectories toward desired helical parameters. Monomers were subsequently designed with a FastDesign protocol with repeat symmetry enforced and top scoring monomers with docked into C2 symmetric homodimer geometries using the Rosetta app sicdock as previously published (10) with the added requirement that the proteins form closed circular architectures. Interfaces were then designed using a FastDesign protocol with C2 symmetry enforced, and top scoring designs were ordered for experiment characterization. Please see *SI Appendix* for more detailed computational methods.

Design scripts are available at: https://github.com/drhicks/donut_protein_manuscript.git.

Additionally, rosetta_scripts, pyrosetta, and sicdock can be downloaded from the main rosetta Github repository: https://github.com/RosettaCommons.

1. J. Erickson, D. Kempf, Structure-based design of symmetric inhibitors of HIV-1 protease. *Arch Virol Suppl.* **9**, 19–29 (1994).
2. S. Bandyopadhyay, K. Chandramouli, M. K. Johnson, Iron-sulfur cluster biosynthesis. *Biochem. Soc. Trans.* **36**, 1112–1119 (2008).
3. T. Oie, G. M. Maggiora, R. E. Christoffersen, Structural characterization of a special-pair chlorophyll dimer model of P700. *Int. J. Quantum Chem.* **22**, 157–171 (1982).
4. I. Cohen-Ofri *et al.*, Zinc-bacteriochlorophyllide dimers in de novo designed four-helix bundle proteins. A model system for natural light energy harvesting and dissipation. *J. Am. Chem. Soc.* **133**, 9526–9535 (2011).
5. M. Faiella *et al.*, An artificial di-iron oxo-protein with phenol oxidase activity. *Nat. Chem. Biol.* **5**, 882–884 (2009).
6. B. R. Gibney, S. E. Mulholland, F. Rabanal, P. L. Dutton, Ferredoxin and ferredoxin-heme maquettes. *Proc. Natl. Acad. Sci. U.S.A.* **93**, 15041–15046 (1996).
7. J. Park *et al.*, De novo design of a homo-trimeric amantadine-binding protein. *eLife* **8**, e47839 (2019).
8. T. J. Brunette *et al.*, Exploring the repeat protein universe through computational protein design. *Nature* **528**, 580–584 (2015).
9. L. Doyle *et al.*, Rational design of α-helical tandem repeat proteins with closed architectures. *Nature* **528**, 585–588 (2015).
10. J. A. Fallas *et al.*, Computational design of self-assembling cyclic protein homo-oligomers. *Nat. Chem.* **9**, 353–360 (2017).
11. G. Ueda *et al.*, Tailored design of protein nanoparticle scaffolds for multivalent presentation of viral glycoprotein antigens. *eLife* **9**, e57659 (2020).
12. G. W. Foight *et al.*, Multi-input chemical control of protein dimerization for programming graded cellular responses. *Nat. Biotechnol.* **37**, 1209–1216 (2019).
13. J. K. Leman *et al.*, Macromolecular modeling and design in Rosetta: Recent methods and frameworks. *Nat. Methods* **17**, 665–680 (2020).
14. L. An, G. R. Lee, De Novo protein design using the blueprint builder in rosetta. *Curr. Protoc. Protein Sci.* **102**, e116 (2020).
15. J. B. Maguire *et al.*, Perturbing the energy landscape for improved packing during computational protein design. *Proteins* **89**, 436–449 (2021).
16. F. DiMaio, A. Leaver-Fay, P. Bradley, D. Baker, I. André, Modeling symmetric macromolecular structures in Rosetta3. *PLoS One* **6**, e20450 (2011).
17. B. Coventry, D. Baker, Protein sequence optimization with a pairwise decomposable penalty for buried unsatisfied hydrogen bonds. *PLOS Comput. Biol.* **17**, e1008061 (2021).
18. N. J. Greenfield, Using circular dichroism spectra to estimate protein secondary structure. *Nat. Protoc.* **1**, 2876–2890 (2006).
19. G. L. Hura *et al.*, Comprehensive macromolecular conformations mapped by quantitative SAXS analyses. *Nat. Methods* **10**, 453–454 (2013).
20. K. N. Dyer *et al.*, "High-throughput SAXS for the characterization of biomolecules in solution: A practical approach" in *Structural Genomics*, Y. W. Chen, Ed. (Springer, 2014), pp. 245–258.
21. D. Shin, SAXS FrameSlice (2017). https://bl1231.als.lbl.gov/ran. Accessed 13 July 2022.
22. D. R. Hicks *et al.*, Data for "De novo design of protein homodimers containing tunable symmetric protein pockets." GitHub. https://github.com/drhicks/donut_protein_manuscript. Deposited 7 February 2022.
23. D. R. Hicks *et al.*, Data for "De novo design of protein homodimers containing tunable symmetric protein pockets." GitHub. https://github.com/RosettaCommons. Accessed 13 July 2022.

Data Availability. Design scripts data(rosetta_scripts, pyrosetta, and sicdock) have been deposited in GitHub (https://github.com/drhicks/donut_protein_manuscript; https://github.com/RosettaCommons) (22, 23). All other study data are included in the article and/or *SI Appendix*.

ACKNOWLEDGMENTS. We want to thank the Advanced Light Source (ALS) beamline 8.2.2/8.2.1 at Lawrence Berkeley National Laboratory for X-ray crystallography data collection. The Berkeley Center for Structural Biology is supported in part by the National Institutes of Health (NIH), National Institute of General Medical Sciences, and the Howard Hughes Medical Institute. The ALS is supported by the Director, Office of Science, Office of Basic Energy Sciences and US Department of Energy (DOE) (DE-AC02-05CH11231). We also thank the staff at the ALS SIBYLS beamline at Lawrence Berkeley National Laboratory, including K. Burnett, G. Hura, M. Hammel, J. Tanamachi, and J. Tainer for the services provided through the mail-in SAXS program, which is supported by the DOE Office of Biological and Environmental Research Integrated Diffraction Analysis program DOE BER IDAT grant (DE-AC02-05CH11231), National Institute of General Medical Sciences (NIGMS) supported ALS-ENABLE (GM124169-01), and NIH project MINOS (R01GM105404). This work was supported by the National Science Foundation (NSF) (CHE-1629214 to D.B. and GRFP DGE-1762114 to M.A.K.), NIGMS (R01GM115545), NIH (R01GM139752 to B. Stoddard), the National Institute of Aging (U19AG065156 to D.B. and D.R.H.), a generous gift from the Audacious Project (to D.B., A.K., and A.K.B.), and the Open Philanthropy Project at the Institute for Protein Design (to D.B., B.C., and T.J.B.).

Author affiliations: [a]Department of Biochemistry, University of Washington, Seattle, WA 98195; [b]Institute for Protein Design, University of Washington, Seattle, WA 98195; [c]Division of Basic Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA 98109; [d]Howard Hughes Medical Institute, University of Washington, Seattle, WA 98195; and [e]Molecular Biophysics and Integrated Bioimaging, Berkeley Center for Structural Biology, Lawrence Berkeley National Laboratory, Berkeley, CA 94720