

Computational design of self-assembling cyclic protein homo-oligomers

Jorge A. Fallas^{1,2‡}, George Ueda^{1,2‡}, William Sheffler^{1,2‡}, Vanessa Nguyen², Dan E. McNamara^{3†}, Banumathi Sankaran⁴, Jose Henrique Pereira^{4,5}, Fabio Parmeggiani^{1,2}, T. J. Brunette^{1,2}, Duilio Cascio³, Todd R. Yeates³, Peter Zwart⁴ and David Baker^{1,2,6*}

Self-assembling cyclic protein homo-oligomers play important roles in biology, and the ability to generate custom homo-oligomeric structures could enable new approaches to probe biological function. Here we report a general approach to design cyclic homo-oligomers that employs a new residue-pair-transform method to assess the designability of a protein-protein interface. This method is sufficiently rapid to enable the systematic enumeration of cyclically docked arrangements of a monomer followed by sequence design of the newly formed interfaces. We use this method to design interfaces onto idealized repeat proteins that direct their assembly into complexes that possess cyclic symmetry. Of 96 designs that were characterized experimentally, 21 were found to form stable monodisperse homo-oligomers in solution, and 15 (four homodimers, six homotrimers, six homotetramers and one homopentamer) had solution small-angle X-ray scattering data consistent with the design models. X-ray crystal structures were obtained for five of the designs and each is very close to their corresponding computational model.

Cyclic homo-oligomers assembled from multiple identical protein subunits symmetrically arranged around a central axis play key roles in many biological processes, including catalysis, signalling and allostery^{1–3}. Despite their prevalence in natural systems, currently there is no systematic approach to design cyclic homo-oligomers starting from a monomeric protein structure. A number of prior design studies have relied on canonical structural motifs, such as α -helical coiled coils⁴, β -propeller motifs^{5,6}, unpaired β strands⁷ or metal-binding sites⁸. Recently, a C_2 dimer mediated by an α -helical interface was reported but the design protocol required extensive iteration between computation and experiment⁹. In contrast, there has been considerable progress in designing proteins that fold into predetermined target structures ranging from idealized versions of natural folds^{10–13} to topologies that appear not to have been explored during evolution^{14,15}. Particularly interesting from an engineering perspective are *de novo* designed α -helical repeat proteins with a wide range of shapes that can be readily shortened or lengthened simply by changing the number of sequence repeats¹⁵.

Here we present a general method to design cyclic homo-oligomers *in silico* and use it to design interfaces onto idealized proteins^{13,15,16} that direct their assembly into dimeric, trimeric, tetrameric and pentameric complexes. Structural characterization shows that many of the designs adopt the target oligomerization state and structure, demonstrating that we have a basic understanding of the determinants of oligomerization. The capability to design proteins with tunable shape, size and symmetry would allow for the rigid display of binding domains at arbitrary orientations and distances for a variety of biological applications.

Results

The self-assembly of naturally occurring complexes is driven by chemical and shape complementarity. Protein-protein interfaces

generally comprise a hydrophobic core that is buried on binding and surrounded by a rim of polar residues that prevent nonspecific aggregation^{17–21}. We developed a design strategy to generate such interfaces between protein monomers docked in a range of cyclic geometries. The strategy has two steps (Fig. 1): first, low-resolution docking to sample and rank symmetric arrangements of a given scaffold protein based on their designability (the likelihood of finding an amino acid sequence that can stabilize a given rigid body conformation) and, second, full atom RosettaDesign²² calculations to optimize the sequence at the protein-protein interfaces for high-affinity binding. To explore the generality of the method, symmetries that ranged from C_2 through to C_6 were designed. Of these designs, 96 were selected for experimental characterization, and four homodimers, six homotrimers, six homotetramers and one homopentamer were found to form stable monodisperse homo-oligomers in solution.

Computational design. Existing methods for protein-protein docking fall into three general categories: (1) voxelized rigid representations with fast Fourier transform (FFT)-based docking^{23,24}, (2) docking based on patches of high-resolution local shape complementarity²⁵ and (3) Monte Carlo sampling with soft centroid models^{26,27}. The first two categories are not ideal for the protein-design problem because the precise shape and chemical detail of the docked surfaces are unavailable, as the interface residues are not known prior to design. The approach we take is most similar to (3), in which docked backbones are generated and then scored using a low-resolution representation of the proteins (which requires only the backbone coordinates and secondary-structure assignments), but with two notable improvements. First, we employ a six-dimensional (6D) implicit side-chain scoring methodology, which predicts the result of subsequent full-atom

¹Department of Biochemistry, University of Washington, Seattle, Washington 98195, USA. ²Institute for Protein Design, University of Washington, Seattle, Washington 98195, USA. ³Department of Chemistry and Biochemistry, University of California Los Angeles, Los Angeles, California 90095, USA. ⁴Berkeley Center for Structural Biology, Molecular Biophysics and Integrated Bioimaging, Lawrence Berkeley Laboratory, Berkeley, California 94720, USA. ⁵Joint BioEnergy Institute, Emeryville, California 94608, USA. ⁶Howard Hughes Medical Institute, University of Washington, Seattle, Washington 98195, USA.

[†]Present address: Departments of Structural Biology and of Chemical Biology & Therapeutics, St. Jude Children's Research Hospital, Memphis, Tennessee 38105, USA. [‡]These authors contributed equally to this work. *e-mail: dabaker@u.washington.edu

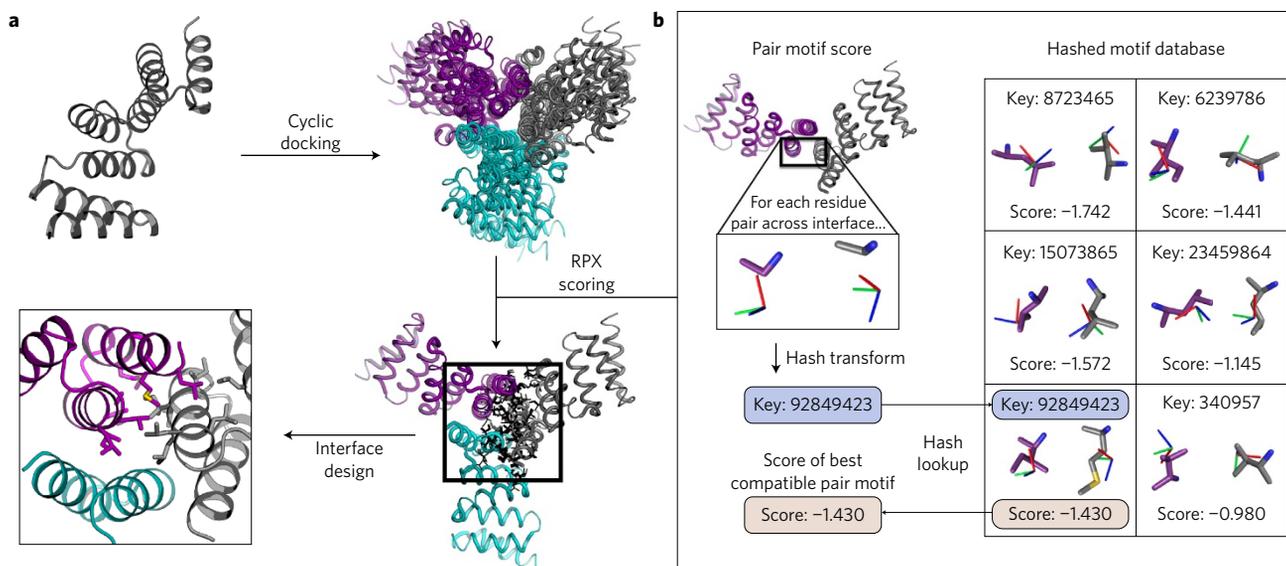


Figure 1 | Computational design protocol. **a**, Starting with a monomeric protein we exhaustively sample cyclic docked configurations, score them using the RPX method and generate sequences to drive the complex formation using a full-atom RosettaDesign²¹ calculation. **b**, Schematic representation of the RPX method scoring procedure.

design calculations better than a traditional coarse-grained model, and, second, we use an enumerative strategy to generate docked backbones, which samples the low-dimensional docking space more robustly than a Monte Carlo search.

In previous efforts, scoring at the docking stage has been accomplished using coarse-grained models in which the absent side chains are represented by one or two points in space, and the interaction potential between two amino acids is evaluated as a function of the distance or distances between these points and, in some cases, of an associated angle^{27–31}. These representations are incomplete because they do not capture the full 6D rigid-body relationship between pairs of side chains. To avoid loss of information, we developed a residue-pair transform (RPX) model that represents the interaction between two residues by the full 6D rigid-body transformation between their respective backbone N, Ca and C atoms. We employed a precompiled database of all the favourable residue-pair interactions found in structures from the Protein Data Bank (PDB) that involved alanine, isoleucine, leucine, valine and methionine, and binned these data based on the rigid-body transform between amino acids. The score of a given docked configuration is the sum, over each pair of residues across the interface, of the lowest Rosetta full-atom energy found in the associated spatial transformation bin of the database. This approach predicts the interface energy that results from the full-atom sequence-design calculation better than the Rosetta centroid energy function (Supplementary Fig. 1). As the residue-pair-transform database is compiled offline, arbitrary data selection (different subsets of amino acid identities) and processing (alternative smoothing and scoring schemes) can be employed with no impact on the runtime of the docking calculations. Details on the database utilized for this study are available in Methods and Supplementary Methods.

For the best leverage of the RPX scoring methodology described above, we employ deterministic sampling of the complete docking space. The configurational space for cyclic docking is 4D: the usual six degrees of freedom required to orient a rigid body minus translations along and rotations around the symmetry axis of the oligomer (to which the structure is invariant). These four degrees of freedom can be reduced effectively to three by the requirement that the subunits must be roughly in contact. We realize this dimensionality reduction with a fast slide-into-contact algorithm. To compute the translation along a slide vector which

brings two rigid clouds of atoms into contact, we create a pair of 2D arrays that contain the leading face of each cloud along the slide vector. Corresponding cells of each array are checked, and the pair of atoms with least separation along the slide vectors defines an upper bound on the slide distance. The final slide distance is calculated using a local octree-like data structure (Methods). This results in a significant savings in the total number of samples that must be evaluated compared with a simpler brute-force search.

For the ten best RPX-scoring docked arrangements of each monomer, low-energy and shape-complementary interfaces between the protomers were generated using Rosetta sequence design calculations employing a Monte-Carlo simulated annealing protocol (details on the RosettaScript³² that encodes the protocol are provided in Methods and Supplementary Methods). Designs were filtered on number of mutations, buried surface area, shape complementarity and computed interaction energy (Supplementary Fig. 2), and 96 were selected for experimental characterization. The 11 dimers, 34 trimers, 19 tetramers, 17 pentamers and 15 hexamers are named according to the following nomenclature: the first four letters refer to the scaffold protein (as described in the Supplementary Information), the symmetry is denoted as C_n and, finally, an integer is added to differentiate oligomers of identical symmetry and scaffold identity.

Protein expression and oligomerization state screening. Synthetic genes that encode each of the 96 designs were synthesized and cloned into a vector with a T7 promoter system and either an N- or C-terminal (His)₆ tag, and the corresponding proteins were expressed in *E. coli*. The proteins were purified by immobilized nickel-affinity chromatography (Ni²⁺ IMAC) and size-exclusion chromatography (SEC). Of these designs, 64 were soluble and amenable to purification (Supplementary Figs 3 and 4). The oligomerization states for 44 designs that eluted from SEC with a single predominant species were determined by SEC in tandem with multiangle light scattering (SEC-MALS). For 21 of the designs, the molecular weights determined by light scattering were consistent with the designed oligomerization state.

Structural characterization. To assess further the configuration of the designed proteins in solution, small-angle X-ray scattering (SAXS) measurements were performed on designs that had

predominantly monodisperse SEC traces. A total of 26 designs (the 21 with consistent SEC–MALS data and five additional designs that had monodisperse SEC profiles) were characterized by SAXS and the measured scattering profile was compared with that expected from the computational model. Designs with a deviation of less than or equal to 3.1 a.u. using the χ measure³³ and a deviation of less than 11% between the computed and experimental radius of gyration were considered to be in the designed supramolecular arrangement (these thresholds were chosen based on the deviations between computed and measured values for designs with crystal structures consistent with the corresponding models (see below)).

Of the 26 designs, 15 fulfil these criteria—five dimers, six trimers, three tetramers and one pentamer. The docked configurations and designed interfaces of 13 of these are unique (three of the trimers have similar geometries with pairwise root mean squared deviation (r.m.s.d.) values between 1.9 and 2.5 Å; the lowest pairwise r.m.s.d. among the remaining designs is 5.3 Å with no similarity in designed interface). Computational models, *in silico* symmetric docking energy landscapes, SEC–MALS chromatograms and SAXS experimental and computed profiles for these 15 designs are summarized in Fig. 2 and Supplementary Fig. 5 (data on the full set of designs is provided in Supplementary Tables 1–4).

Crystal structures that contain the designed interface were obtained for five of the designed proteins (two dimers, two trimers and one tetramer), and are compared with the design models in Fig. 3. For each of the five cases the side-chain rotamers of the hydrophobic residues are similar to those in the design model. The two dimers, ank3C2_1 and ank1C2_1, are both built from idealized ankyrin repeat proteins and are shown in Fig. 3a,b. The ank3C2_1 design has a large hydrophobic patch (1,100 Å²) that is buried on binding; all the interface hydrophobic side chains are in the same rotameric state in the design model and the crystal structure with the exception of methionine 90 (Fig. 3a, right panel). The backbone r.m.s.d. between the design model and the crystal structure is 1.0 Å. The agreement between the model and the structure of ank1C2_1 (Fig. 3b) is even closer—both polar and hydrophobic side-chain rotamers are correct and the backbone r.m.s.d. to the model is only 0.9 Å.

The two trimeric designs with solved structures are 1na0C3_3 (Fig. 3c), built from a consensus designed TPR protein¹⁶, and HR00C3_2 (Fig. 3d), built from a *de novo* designed repeat protein. 1na0C3_3 has a hydrophobic core that lies on the threefold axis formed by residues in all subunits. The r.m.s.d. between the crystal structure and design model is 1.0 Å. HR00C3_2 contains a pore on the symmetry axis and is stabilized by three separate heterologous interfaces. This trimer was designed using the computational model of a designed repeat protein whose structure had not previously been confirmed by X-ray crystallography. Thus, the crystal structure, which has a backbone r.m.s.d. to the model of 0.9 Å, validates the design of both the monomer and oligomer simultaneously. This ability to design higher-order structures accurately based on design models of the monomers will considerably streamline the future computational design of nanomaterials using monomers with custom-designed properties.

For the two dimers and the two trimers, the χ values between the measured SAXS profiles and the profiles computed from either the corresponding design models or from crystal structures are less than 3.1. In contrast, the experimental SAXS data for the designed tetramer, ank1C4_2 (Fig. 3e), deviates considerably from that computed using the crystal structure (Supplementary Fig. 6). The ank1C4_2 crystal structure adopts a C₂ symmetric tetrameric structure in which two pairs of chains accurately match the design model (r.m.s.d. of 1.1 Å), but exhibit a clear overall distortion relative to the C₄ symmetric design model (r.m.s.d. of 4.5 Å). There are two distinct interfaces present in the structure, one of which corresponds

to the designed interface. The experimental SAXS profile is closer to the design model of the tetramer than to the crystal structure, and hence it seems likely that the symmetry breaking in the crystal results from lattice contacts.

A sixth structure was solved for the ank4C4 design, which shows a single symmetric peak by SEC and forms a tetrameric complex in solution, as determined by MALS. The SAXS profile of this design does not match that computed from the design model ($\chi = 3.8$), and the crystal structure exhibits D₂ symmetry rather than the target C₄ symmetry. The SAXS profile computed from the D₂ oligomer matches the measured scattering curve better than the target C₄ model ($\chi = 1.2$, Supplementary Fig. 8).

Subunit extensions. To explore the modularity of the designs and the robustness of the designed interfaces, we extended two of the designed oligomers by appending two additional repeats to the original constructs. Extended versions of ank1C2_1 and HR04C4_1 were expressed and characterized as described above. SEC–MALS traces of the long constructs show the expected shifts to larger apparent sizes compared with the original constructs (Fig. 4, third column), and the calculated molecular weights are close to those expected. Experimental SAXS profiles of the extended designs are in good agreement with those of the extended computational models (χ values are given in Supplementary Table 3), which suggests that the supramolecular arrangement of the subunits is maintained on extending the scaffold protein. This ability to maintain oligomer geometry while extending the length of the monomers will be very useful for systematically varying the distance between binding moieties and for nanomaterial design.

Resilience to guanidine denaturation. The repeat protein scaffolds used to construct the designed oligomers are very stable proteins, and thus guanidine denaturation can be used to probe the stability of the designed interfaces independent of the effects on the monomers. Four designed oligomers (one selected from each symmetry C₂–C₅) were purified in an initial round of IMAC and SEC, and subsequently run through SEC–MALS in Tris-buffered saline supplemented with 1 M or 2 M GuHCl. In both conditions, all four designs remained in their designed oligomeric state (as determined by MALS) without any indications of a smaller-assembly formation (Supplementary Fig. 7).

Discussion

Our results show that homo-oligomeric protein complexes with cyclic symmetry can be generated from repeat protein building blocks by computationally designing geometrically complementary low-energy interfaces. A key advance is the new fast method to assess designability that provides a reasonable estimate of the energy obtained after a full-atom combinatorial sequence design calculation with roughly six orders of magnitude less computational cost. This allows an exhaustive evaluation of the possible cyclically docked configurations of a monomer which would not be possible with a combinatorial, all-atom sequence design calculation. The broad applicability of the computational pipeline developed here is highlighted by the number of successful designs (15) and symmetries (C₂–C₅) (Table 1). Supplementary Fig. 9 provides an overview of all the experimentally validated dimers, trimers, tetramers and pentamer—the broad range of structures and the variety of interface geometries and architectures far exceeds that reported in any previous study (the elegant beta-propeller designs described in Voet *et al.*⁶ are shown for comparison). The combination of an RPX search for the designable interfaces followed by Rosetta all-atom design calculations can clearly generate a wide range of new interfaces that involve three to five alpha helices; the ability of the approach to design new beta-sheet- and loop-containing interfaces is an area for future investigation.

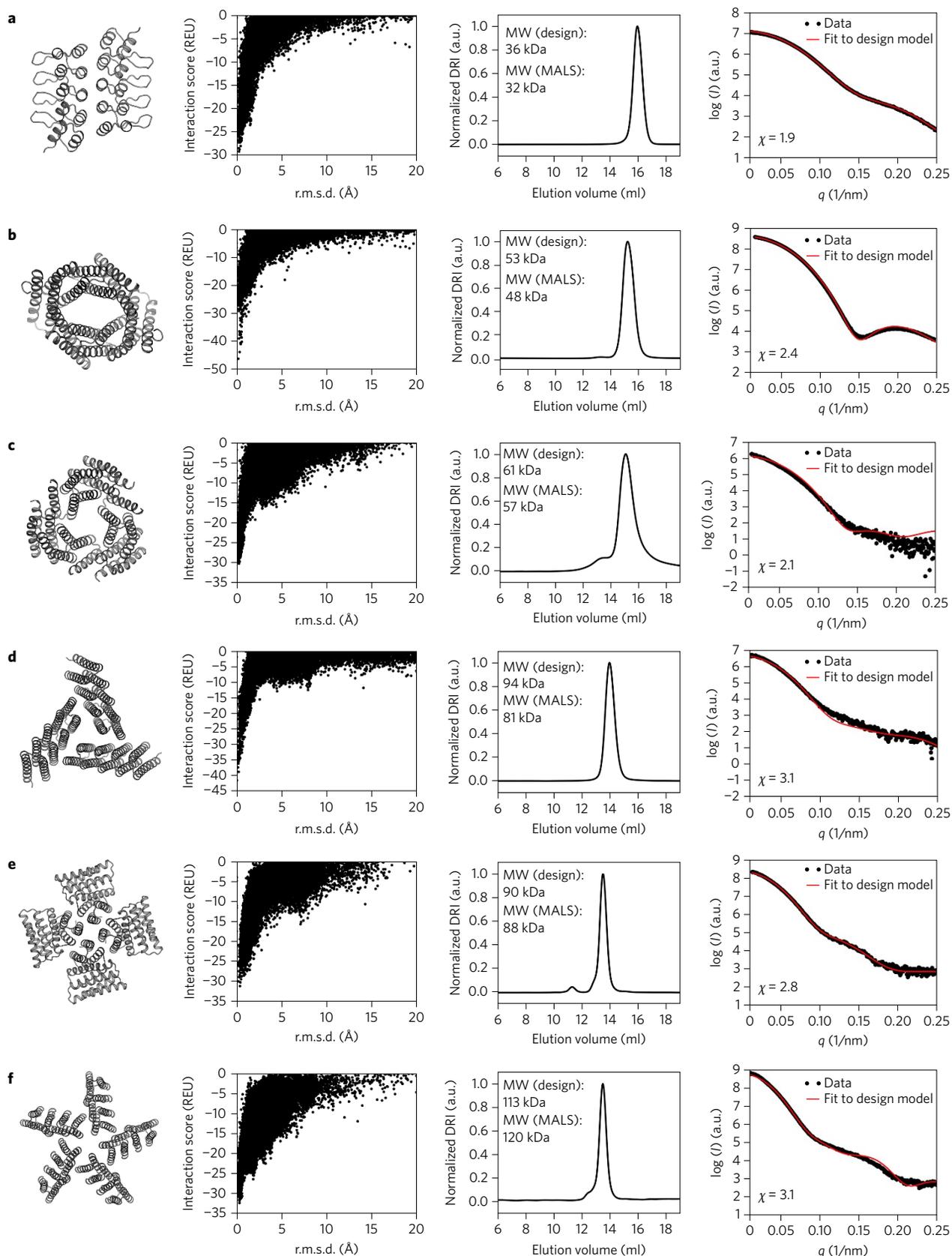


Figure 2 | Assessment of the solution conformation of selected cyclic oligomers. From left to right: computational model, symmetric docking energy landscape, size-exclusion chromatogram used for the molecular weight determination, and SAXS profiles experimentally measured (black dots) and computed from the model (red line). 'MW (design)' refers to the molecular weight of the oligomer design and 'MW (MALS)' refers to the experimentally determined molecular weight. **a**, ank3C2_1. **b**, HR79C2. **c**, HR08C3. **d**, HROOC3_2. **e**, HRO4C4_1. **f**, HR10C5_2. Analogous data for the nine other successful designs are provided in Supplementary Fig. 5. REU, Rosetta energy units; DRI, differential refractive index; I , scattering intensity; q , scattering vector; a.u., arbitrary units.

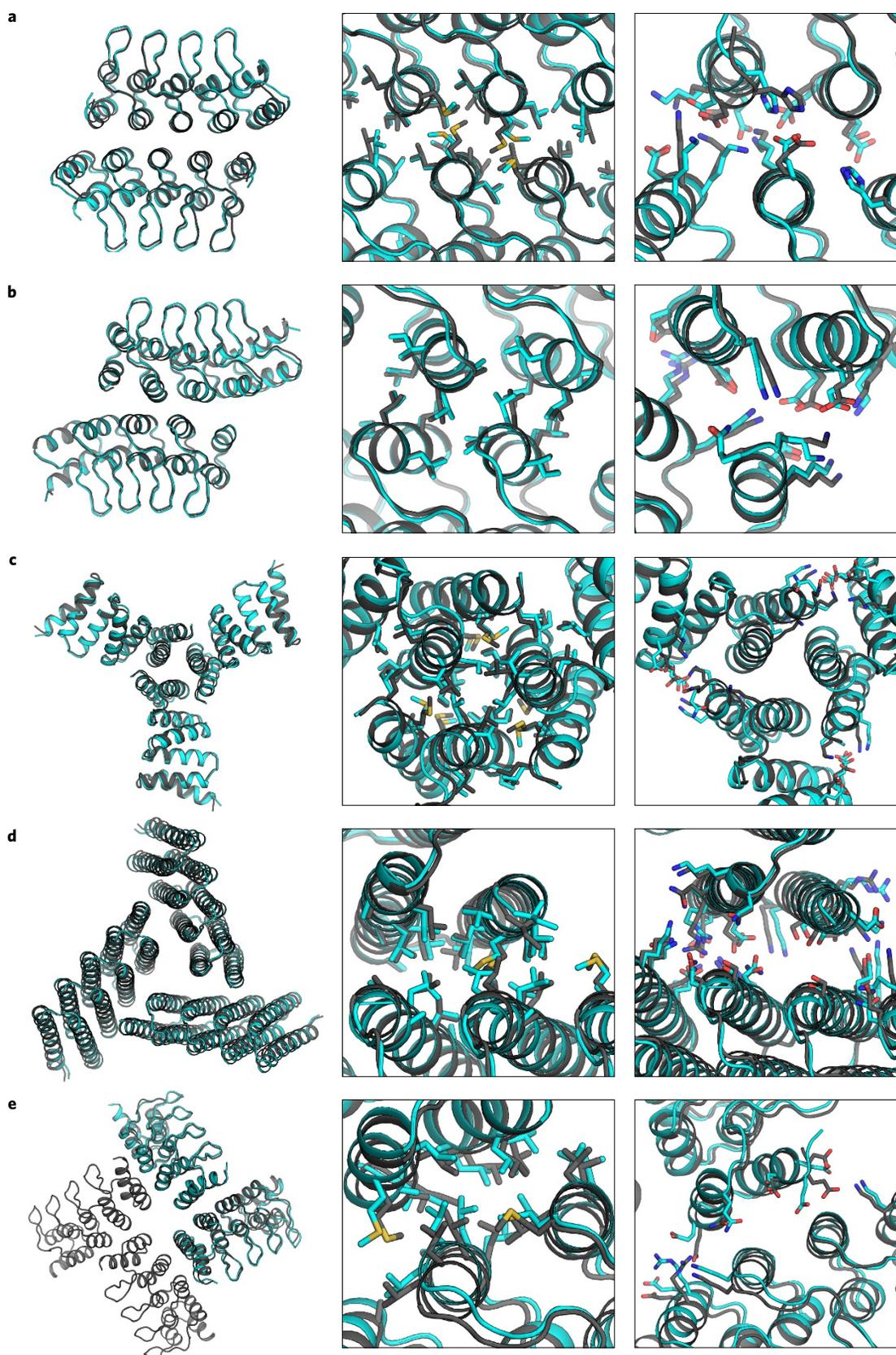


Figure 3 | Comparison between the experimentally determined crystal structures and corresponding design models. Crystal structures are shown in cyan and models in grey. Left column, full model and crystal structure superposition; central column, superposition showing the hydrophobic side chains at the designed interface; right column, superposition showing the hydrophilic side chains at the designed interface. **a**, ank3C2_1 (r.m.s.d. to model 1 Å). **b**, ank1C2_1 (r.m.s.d. to model 0.9 Å). **c**, 1na0C3_3 (r.m.s.d. to model 1 Å). **d**, HR00C3_2 (r.m.s.d. to model 0.9 Å). **e**, ank1C4_2 pair of chains (r.m.s.d. to model 1.1 Å).

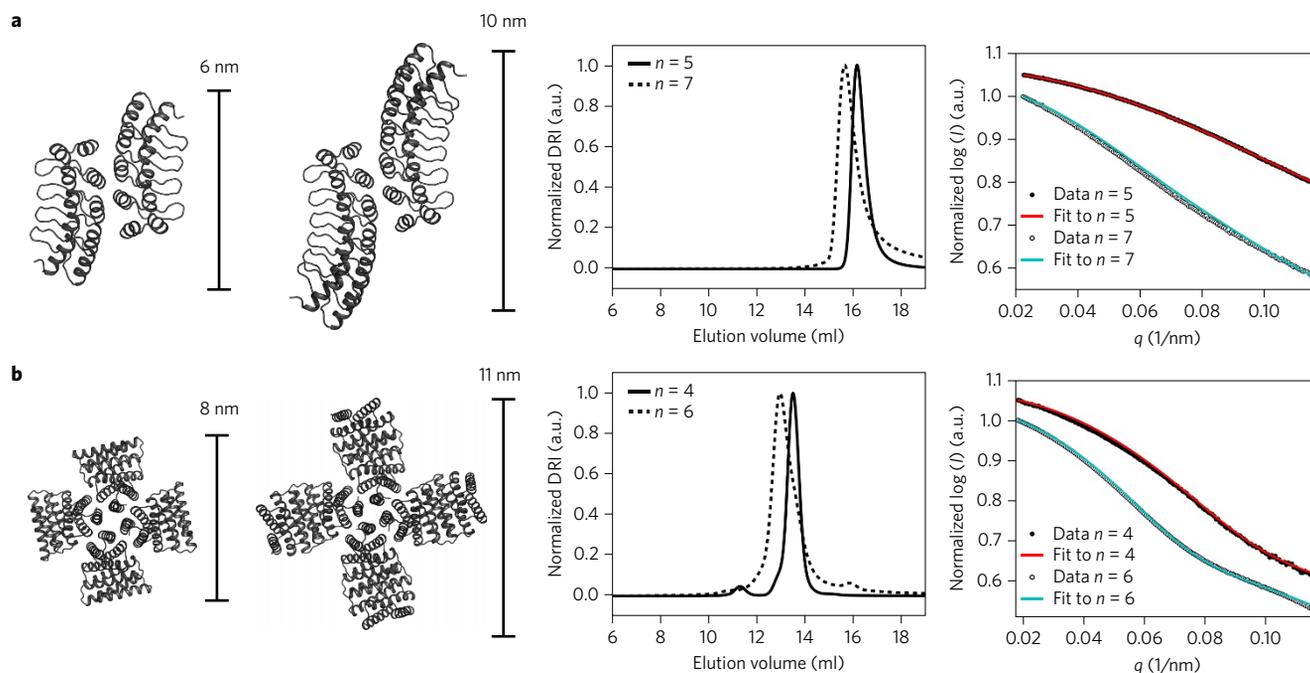


Figure 4 | Robustness of designs to subunit extension by repeat addition. From left to right: computational model of the original design, computational model of the extended design, SEC-MALS chromatogram used for molecular weight determination (n represents the number of repeat modules in each monomer; original design, solid line; extended design, dotted line), SAXS profiles (original design, experimental data in black circles, computed profile in red; extended design, experimental data open circles, computed profile in cyan). **a**, ank1C2_1. **b**, HR04C4_1.

Progress in protein design will require the study not only of the successes but also of the failures. The results reported in this paper provide a valuable resource to understand failure modes, as the input scaffolds are all very stable designed proteins (in previous design studies, the often unknown stability of the starting native scaffolds and the robustness to amino acid substitutions were potentially confounding factors). We are able to distinguish distinct failure modes for the designs reported: 32 were not expressed solubly in *E. coli*, 24 adopted multiple oligomerization states, four were monomeric, 15 were monodisperse, but had an oligomerization state different from that designed, and six occupied the designed oligomerization state but had unanticipated configurations based on SAXS data. Analysis of the properties of the design models revealed that designs with (1) a high total charge (greater than -50), (2) small ($<750 \text{ \AA}^2$) interfaces, (3) low shape complementarity (<0.625) or (4) for which asymmetric pairwise-docking calculations found alternative arrangements of much lower energy than the two-body interaction energies in the design model were generally unsuccessful. Furthermore, despite the success with HR00C3_2, designs based on monomers with crystal structures had higher success rates (19%) than those based on monomers validated only by SAXS (4%). The fraction of designs experimentally confirmed to be in the designed state increased from 15/96 in the overall population to 14/45 when restricted to models that satisfy the above criteria (low

electrostatic repulsion, larger shape-complementary interfaces, absence of competing dimeric states of much lower energy and crystallographically validated monomer structures). Evidently, we currently understand some, but not all, the factors that determine the accuracy of the design calculations. As this is clearly an important area for future investigation, we provide all the experimental data for both unsuccessful and successful designs, the design models and sequences, and a variety of metrics computed from the models in the Supplementary Information.

The design success rate also clearly decreases with increasing oligomerization state—indeed, there were no successes with hexamers. Higher oligomerization states present several challenges: an increase in translational entropy loss (the formation of three dimers from six subunits results in three independently translating bodies, whereas the formation of a single hexamer results only in one), an increase in electrostatic repulsion and a decrease in the difference in interface geometry between competing alternative oligomerization states (smaller reorientations are required to convert a pentamer into a hexamer than a dimer into a trimer). There are clear ways forward to address the second and third challenges: the total charge of the designs can be adjusted to be close to zero at pH 7.0 by suitable redesign of the surface (although some experimentation may be required to maintain solubility), and employing hydrogen-bond networks³⁴ could provide the conformational specificity required to distinguish between higher-order oligomerization states.

Our robust design pipeline can be combined with the modularity of computationally designed repeat proteins to control the 3D arrangement of the protomers at multiple length scales. The designed homo-oligomeric interfaces control the nanoscale 3D arrangement and extensions of the repeat proteins allow for the placement of functional motifs with subnanometre resolution in each of the interacting proteins. Designed proteins can remain folded under strongly denaturing conditions¹⁴, and the design process provides unparalleled control over their geometry^{15,35} and amino acid composition, which allows for reactive chemical moieties, such as thiols or aromatic rings, to be reserved to engineer function in

Table 1 | Summary of experimental results.

Symmetry	Designs	Soluble expression	Target molecular weight	Structural validation
C ₂	11	11/11	7/11	5/11
C ₃	34	20/34	6/34	6/34
C ₄	19	13/19	6/19	3/19
C ₅	17	9/17	1/17	1/17
C ₆	15	11/15	1/15	0/15
Total	96 (100%)	64 (67%)	21 (22%)	15 (16%)

downstream applications. An immediate use for these designed oligomers is to probe how the geometry and valency of tethered signalling molecules affects the clustering of receptors and the cellular response. The relationship between ligand valency, spatial orientation and signalling outcome is not well understood, and designed homo-oligomerization with systematically tunable lengths should be very well suited to investigating this and other basic biological questions.

Methods

Scaffold set. A set of 17 monomeric designed repeat proteins with high-resolution crystal structures as well as six computational models that were validated by SAXS were used as a scaffold set for our design protocol. PDB IDs of the scaffolds used are available in Supplementary Methods.

Motif database and scoring. We construct Cartesian frames given two N–Ca–C backbone segments across the symmetric interface. The relative position and orientation of the two N–Ca–C segments form a 6D space that can be divided into bins, with any possible position/orientation assigned to a bin index. The best-scoring superimposable residue pair available in a large database of candidates can then be found with a single memory lookup keyed on the bin index. The database of residue-pair motifs was constructed from residue pairs observed in a set of high-quality structures from the PDB, filtered for energetic favourability, separated by at least ten residues in sequence and a residue composition of only alanine, isoleucine, leucine, valine and methionine. To compute an aggregate score for each conformation, we consider all pairs of N–Ca–C backbone segments across the newly formed symmetric interface within 9 Å of one another. For each such pair, the score of the best superimposable residue-pair motif is looked up, and the results are summed.

Cyclic docking. To generate cyclic homo-oligomeric arrangements of n copies of a protein monomer, we centre it at the origin, finely sample the three rotational degrees of freedom, generate a symmetric copy by $(360/n)^\circ$ rotation around the z axis and slide the two bodies into contact along the x axis, allowing a small range of x offsets close to the contact value. For each of these, the axis of symmetry is determined from the relative orientation of the two subunits, and the full oligomer is generated and evaluated using the database of residue-pair motifs. A rapid slide-into-contact operation is required for this sampling strategy. Computing the slide distance along a given slide vector is accomplished using two 2D arrays perpendicular to the slide direction into which the atoms along the leading face of each body are placed. Corresponding cells are checked, and the pair with the least separation provides an estimate of the slide distance. The bodies are placed according to this estimate, but may still have clashes. All contacting pairs of atoms across the bodies are checked using an octree-like data structure, and the bodies are backed off to relieve the largest clash found. This process is repeated until no clashes are found. In practice, only one or two iterations through the fast clash check are required in most cases, which makes the slide-move rapid.

Interface design. An interface-design protocol was implemented in RosettaScripts and is described briefly here and extensively in the Supplementary Methods. In each design trajectory, the protomer was initially perturbed by a small translation perpendicular to the axis of symmetry, as well as by a random rotation around its centre of mass. An oligomer with the specified cyclic symmetry was then generated using the information stored in the symmetry definition file (described in the Supplementary Methods). Amino acids at the interface were optimized using the Monte Carlo simulated annealing protocol available in the Rosetta Macromolecular Modeling suite. An initial optimization step was executed with a modified score function with a soft repulsive term. Once a sequence was converged on, designable positions were allowed to minimize side-chain torsion angles using the same weight of reduced repulsive term. A subsequent round of design and minimization was conducted, but with the standard score function to obtain a sequence that corresponds to a local minimum of the energy function. Initially, the extended rotamer library available in Rosetta was utilized, but in later design rounds it was augmented with the rotamers available in the database of residue-pair motifs. Individual design trajectories were filtered by the following criteria: the difference between the Rosetta energy of the bound (oligomeric) and unbound (monomeric) states less than -20.0 Rosetta energy units, interface surface area greater than 700 \AA^2 , Rosetta shape complementarity greater than 0.65 and fewer than 45 mutations made from the respective native scaffold. Designs that passed these criteria were manually inspected and refined by single-point reversion for mutations that were deemed not to contribute to stabilizing the bound state of the interface. The design with the best overall scores for each docked configuration was then added to a set of finalized proteins to be validated experimentally.

Details on protein expression, purification, SEC, molecular weight determination and structural characterization of the proteins characterized in this study are available in the Supplementary Methods.

Code availability. The source code and pre-compiled executable along with the scoring tables and motif database are available on request. Crystal structures have

been deposited in the Research Collaboratory for Structural Bioinformatics PDB with the accession numbers 5HRY (ank3C2_1), 5HRZ (1na0C3_3), 5HSO (ank1C4_2), 5KBA (ank1C2_1), 5K7B (HR00C3_2) and 5KWD (ank4C4).

Received 29 January 2016; accepted 13 October 2016;
published online 5 December 2016

References

- Ali, M. H. & Imperiali, B. Protein oligomerization: how and why. *Bioorg. Med. Chem.* **13**, 5013–5020 (2005).
- Goodsell, D. S. & Olson, A. J. Structural symmetry and protein function. *Annu. Rev. Biophys. Biomol. Struct.* **29**, 105–153 (2000).
- Nishi, H., Hashimoto, K., Madej, T. & Panchenko, A. R. Evolutionary, physicochemical, and functional mechanisms of protein homooligomerization. *Prog. Mol. Biol. Transl.* **117**, 3–24 (2013).
- Fletcher, J. M. *et al.* A basis set of *de novo* coiled-coil peptide oligomers for rational protein design and synthetic biology. *ACS Synth. Biol.* **1**, 240–250 (2012).
- Smock, R. G., Yadid, I., Dym, O., Clarke, J. & Tawfik, D. S. *De novo* evolutionary emergence of a symmetrical protein is shaped by folding constraints. *Cell* **164**, 476–486 (2016).
- Voet, A. R. D. *et al.* Computational design of a self-assembling symmetrical β -propeller protein. *Proc. Natl Acad. Sci. USA* **111**, 15102–15107 (2014).
- Stranges, P. B., Machius, M., Miley, M. J., Tripathy, A. & Kuhlman, B. Computational design of a symmetric homodimer using β -strand assembly. *Proc. Natl Acad. Sci. USA* **108**, 20562–20567 (2011).
- Der, B. S. *et al.* Metal-mediated affinity and orientation specificity in a computationally designed protein homodimer. *J. Am. Chem. Soc.* **134**, 375–385 (2012).
- Mou, Y., Huang, P. S., Hsu, F. C., Huang, S. J. & Mayo, S. L. Computational design and experimental verification of a symmetric protein homodimer. *Proc. Natl Acad. Sci. USA* **112**, 10714–10719 (2015).
- Huang, P. S. *et al.* *De novo* design of an ideal TIM-barrel scaffold. *Protein Sci.* **24**, 186–186 (2015).
- Lin, Y. -R. *et al.* Control over overall shape and size in *de novo* designed proteins. *Proc. Natl Acad. Sci. USA* **112**, E5478–E5485 (2015).
- Koga, N. *et al.* Principles for designing ideal protein structures. *Nature* **491**, 222–227 (2012).
- Parmeggiani, F. *et al.* A general computational approach for repeat protein design. *J. Mol. Biol.* **427**, 563–575 (2015).
- Huang, P. -S. *et al.* High thermodynamic stability of parametrically designed helical bundles. *Science* **346**, 481–485 (2014).
- Brunette, T. J. *et al.* Exploring the repeat protein universe through computational protein design. *Nature* **528**, 580–584 (2015).
- Main, E. R. G., Xiong, Y., Cocco, M. J., D'Andrea, L. & Regan, L. Design of stable alpha-helical arrays from an idealized TPR motif. *Structure* **11**, 497–508 (2003).
- Pechmann, S., Levy, E. D., Tartaglia, G. G. & Vendruscolo, M. Physicochemical principles that regulate the competition between functional and dysfunctional association of proteins. *Proc. Natl Acad. Sci. USA* **106**, 10159–10164 (2009).
- Levy, E. D. & Teichmann, S. Structural, evolutionary, and assembly principles of protein oligomerization. *Prog. Mol. Biol. Transl.* **117**, 25–51 (2013).
- Bahadur, R. P., Chakrabarti, P., Rodier, F. & Janin, J. Dissecting subunit interfaces in homodimeric proteins. *Proteins* **53**, 708–719 (2003).
- Bahadur, R.P., Chakrabarti, P., Rodier, F. & Janin, J. A dissection of specific and non-specific protein–protein interfaces. *J. Mol. Biol.* **336**, 943–955 (2004).
- Janin, J., Bahadur, R. P. & Chakrabarti, P. Protein–protein interaction and quaternary structure. *Q. Rev. Biophys.* **41**, 133–180 (2008).
- Leaver-Fay, A. *et al.* in *Computer Methods, Part C* Vol. 487 (eds Johnson, M. L. & Brand, L.) 545–574 (Elsevier, 2011).
- Chen, R. & Weng, Z. P. Docking unbound proteins using shape complementarity, desolvation, and electrostatics. *Proteins* **47**, 281–294 (2002).
- Pierce, B., Tong, W. W. & Weng, Z. P. M-ZDOCK: a grid-based approach for C_n symmetric multimer docking. *Bioinformatics* **21**, 1472–1478 (2005).
- Schneidman-Duhovny, D., Inbar, Y., Nussinov, R. & Wolfson, H. J. Patchdock and symmDock: servers for rigid and symmetric docking. *Nucleic Acids Res.* **33**, W363–W367 (2005).
- Gray, J. J. & Baker, D. Protein–protein docking predictions with RosettaDock. *Bioinformatics* **20**, 306A–306A (2004).
- Gray, J. J. *et al.* Protein–protein docking with simultaneous optimization of rigid-body displacement and side-chain conformations. *J. Mol. Biol.* **331**, 281–299 (2003).
- Wei, H. *et al.* Lysozyme-stabilized gold fluorescent cluster: synthesis and application as Hg^{2+} sensor. *Analyst* **135**, 1406–1410 (2010).
- Lu, M., Douis, A. D. & Ma, J. OPUS-PSP: an orientation-dependent statistical all-atom potential derived from side-chain packing. *J. Mol. Biol.* **376**, 288–301 (2008).
- Zheng, W. H., Schafer, N. P., Davtyan, A., Papoian, G. A. & Wolynes, P. G. Predictive energy landscapes for protein–protein association. *Proc. Natl Acad. Sci. USA* **109**, 19244–19249 (2012).

31. DeBartolo, J., Dutta, S., Reich, L. & Keating, A. E. Predictive Bcl-2 family binding models rooted in experiment or structure. *J. Mol. Biol.* **422**, 124–144 (2012).
32. Fleishman, S. J. *et al.* RosettaScripts: a scripting language interface to the Rosetta macromolecular modeling suite. *PLoS One* **6**, e20161 (2011).
33. Schneidman-Duhovny, D., Hammel, M. & Sali, A. FoXS: a web server for rapid computation and fitting of SAXS profiles. *Nucleic Acids Res.* **38**, W540–W544 (2010).
34. Boyken, S. E. *et al.* *De novo* design of protein homo-oligomers with modular hydrogen-bond network-mediated specificity. *Science* **352**, 680–687 (2016).
35. Park, K. *et al.* Control of repeat-protein curvature by computational protein design. *Nat. Struct. Mol. Biol.* **22**, 167–174 (2015).

Acknowledgements

This work was supported by the Howard Hughes Medical Institute (HHMI), Air Force Office for Scientific Research (AFOSR FA950-12-10112), the National Science Foundation (NSF MCB-1445201 and CHE-1332907), the Bill and Melinda Gates Foundation (OPP1120319) and the Defense Threat Reduction Agency (HDTRA1-11-C-0026 AM06). We thank R. Koga and L. Carter for assistance with SEC-MALS. We thank M. Collazo and M. Sawaya supported by Department of Energy (DOE Grant DE-FC02-02ER63421). We thank M. Capel, K. Rajashankar, N. Sukumar, J. Schuermann, I. Kourinov and F. Murphy at Northeastern Collaborative Access Team supported by grants from the National Center for Research Resources (5P41RR015301-10) and the National Institute of General Medical Sciences (NIGMS P41GM103403-10) from the National Institutes of Health (NIH). Use of the APS is supported by the DOE under Contract DE-AC02-06CH11357. X-ray crystallography and SAXS data were collected at the Advanced Light Source (ALS),

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 no. and is funded by DOE BER IDAT, NIH MINOS (RO1GM105404) and the ALS, and we thank K. Burnett and G. Hura. The Berkeley Center for Structural Biology is supported in part by the NIH, NIGMS and the HHMI. The ALS is supported by the Director, Office of Science, Office of Basic Energy Sciences of the US DOE under contract no. DE-AC02-05CH11231.

Author contributions

J.A.F., G.U., W.S. and D.B. designed the research. W.S. developed the RPX method and wrote the program code. J.A.F., G.U. and V.N. carried out design calculations, and purified and biophysically characterized the designed proteins. F.P. and T.J.B. designed and characterized the monomeric repeat proteins used as scaffolds. D.E.M., D.C., T.R.Y., J.H.P., G.U. and J.A.F. crystallized the designed proteins. D.E.M., D.C., B.S. and P.Z. collected and analysed crystallographic data. J.A.F., D.E.M., D.C., B.S. and P.Z. solved the structures. All the authors discussed the results and commented on the manuscript.

Additional information

Supplementary information is available in the [online version of the paper](#). Reprints and permissions information is available online at www.nature.com/reprints. Correspondence and requests for materials should be addressed to D.B.

Competing financial interests

The authors declare no competing financial interests.