Biochemistry

pubs.acs.org/biochemistry Article

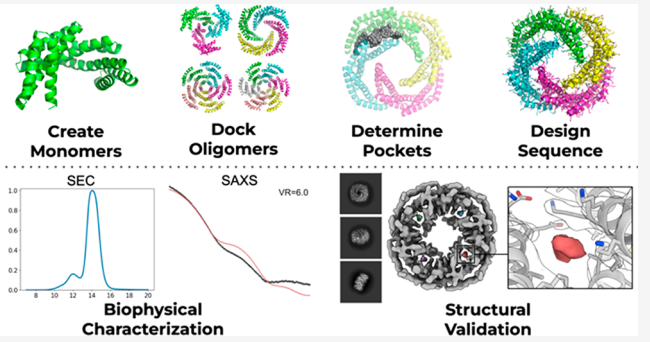
Design of Diverse Asymmetric Pockets in *De Novo* Homo-oligomeric Proteins

Stacey R Gerben, Andrew J Borst, Derrick R Hicks, Isabelle Moczygemba, David Feldman, Brian Coventry, Wei Yang, Asim K. Bera, Marcos Miranda, Alex Kang, Hannah Nguyen, and David Baker*

Cite This: *Biochemistry* 2023, 62, 358−368 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: A challenge for design of protein−small-molecule recognition is that incorporation of cavities with size, shape, and composition suitable for specific recognition can considerably destabilize protein monomers. This challenge can be overcome through binding pockets formed at homo-oligomeric interfaces between folded monomers. Interfaces surrounding the central homo-oligomer symmetry axes necessarily have the same symmetry and so may not be well suited to binding asymmetric molecules. To enable general recognition of arbitrary asymmetric substrates and small molecules, we developed an approach to designing asymmetric interfaces at off-axis sites on homo-oligomers, analogous to those found in native homo-oligomeric proteins such as glutamine synthetase. We symmetrically dock curved helical repeat proteins such that they form pockets at the asymmetric interface of the oligomer with sizes ranging from several angstroms, appropriate for binding a single ion, to up to more than 20 Å across. Of the 133 proteins tested, 84 had soluble expression in *E. coli*, 47 had correct oligomeric states in solution, 35 had small-angle X-ray scattering (SAXS) data largely consistent with design models, and 8 had negative-stain electron microscopy (nsEM) 2D class averages showing the structures coming together as designed. Both an X-ray crystal structure and a cryogenic electron microscopy (cryoEM) structure are close to the computational design models. The nature of these proteins as homo-oligomers allows them to be readily built into higher-order structures such as nanocages, and the asymmetric pockets of these structures open rich possibilities for small-molecule binder design free from the constraints associated with monomer destabilization.



De novo design of small-molecule binding proteins and enzymes is an ongoing challenge. Creating a new enzyme or small-molecule binder requires having a protein with a pocket of both a size and shape that can accommodate the desired function. While there are numerous successful cases of designing new small-molecule binding functions into natural proteins,[1−3] this approach is inherently limited by the finite set of natural proteins, which limits the range of pocket shapes available. Rather than screening through large numbers of natural proteins for one with specific desired properties, *de novo* design now allows the creation of new proteins with specific attributes. Previously, a range of pocket sizes and shapes have been designed in monomeric protein systems such as mixed alpha-beta proteins with NTF2-like folds[4,5] and beta barrels.[6,7] However, the position of the pocket in the center of the monomer restricts the number of positions that can be mutated without disrupting the overall fold geometry. Diverse pocket shapes and sizes have been designed surrounding the central symmetry axis of homodimeric systems,[8] but such pockets are C2 symmetric by construction and are hence not optimal for binding asymmetric substrates and small molecules.

We reasoned that it should be possible to design proteins containing a wide range of pocket sizes and shapes without destabilizing folding or requiring pocket symmetry. Specifically, we aimed to generate cyclic assemblies of protein monomers with pockets at the asymmetric (off-symmetry axis) interfaces formed between adjacent monomers in a manner analogous to naturally occurring enzymes such as proteobacterial carboxymuconolactone decarboxylase (PDB 2QEU) and human glutamine synthetase (PDB 2OJW). In addition, cyclic oligomers allow for the potential to be assembled into even higher-order structures, such as nanocages[9] or other self-

Special Issue: Protein Engineering

Received: August 28, 2022
Revised: November 28, 2022
Published: January 10, 2023

https://doi.org/10.1021/acs.biochem.2c00497
Biochemistry 2023, 62, 358−368

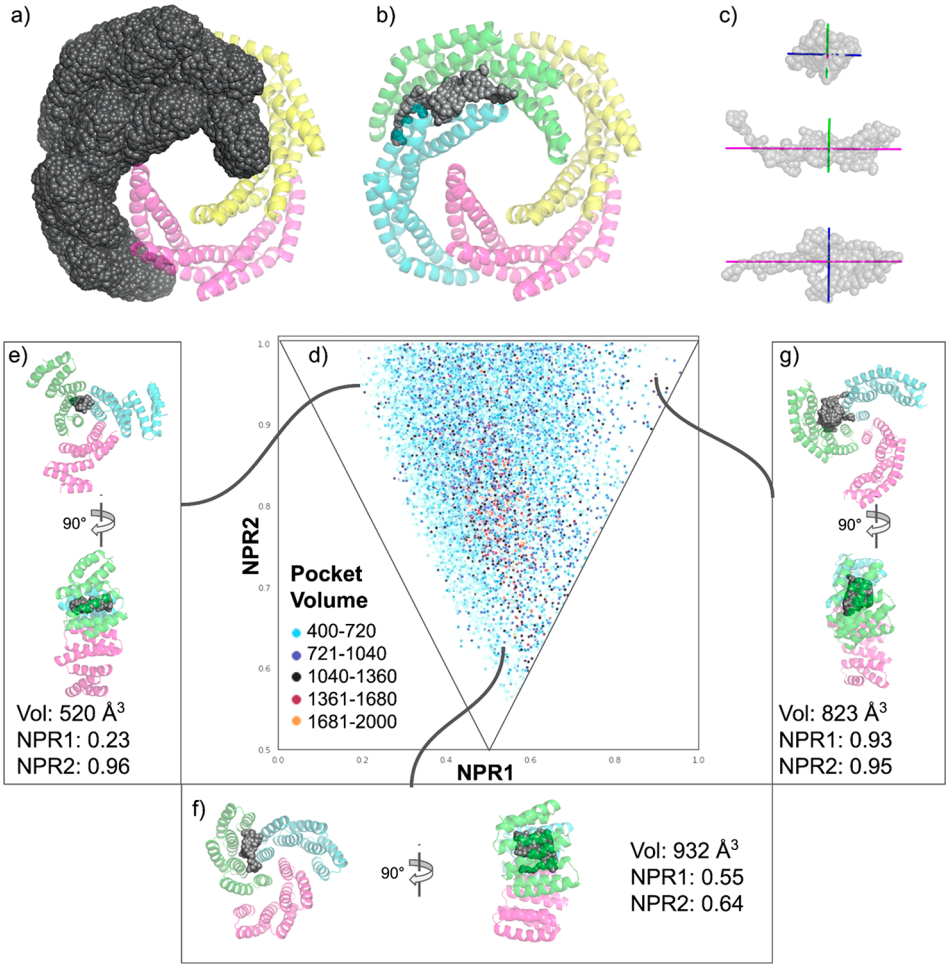


Figure 1. Design and characterization of pockets at asymmetric subunit interfaces. (a) SG135 sphere representation of N, C, O, and Cα atoms in chains A and B. (b) SG135 pocket filled by spheres. (c) Pocket *x*, *y*, and *z* axes shown as green, magenta, and blue lines. (d) The designed pockets span a wide range of shapes and volumes. (e–g) Examples of proteins with very different pocket shapes. (e) Rod-like pocket shaped dock. (f) Disk-like pocket shaped dock. (g) Sphere-like pocket shaped dock.

assembling arrays.[10] For modular reshaping of pocket sizes, we decided to employ designed helical repeat proteins as monomers because their curvature and twist can be readily modulated. Furthermore, these monomers can be easily extended or retracted simply by adding or removing terminal repeats, thus increasing or decreasing pocket size.[11] We anticipated that by combining these curved helical repeat proteins into cyclic homo-oligomers,[11,12] we could generate a library of designs encompassing a variety of cavities across the asymmetric oligomeric interface with differing sizes and shapes that can be used as a basis for enzyme or small molecule binding, and built into higher order symmetric assemblies.

MATERIALS AND METHODS

Monomer Filtering. Monomer sets were filtered on shape complementarity, lDDT,[13] and the fragment quality compared to a database of nine-residue fragments. Metrics are included in supplemental scripts. Bd4 and bd5 sets were separated based on the distance of Cα of each helix to its nearest helices. For each helix i, distance from both i and i + 1 was measured to each helix i + 3 and i + 4. The smallest of these four distance measurements is stored. Once all helices were measured this way, the largest of these distances was returned. If the distance was less than 7.5 Å, it is considered bd4. If it is greater than 7.5 Å, it is considered bd5.

Oligomer Filtering. Oligomers were filtered on helix–helix packing assessed at the protein backbone level[11] and the number of residues in a docked protein model involved in the interactions between two chains ("ncontact"). Docks were filtered by

- "rpx_score" > 50
- 80 < "ncontact" < 120
- "rpx_score"^{2}/"ncontact_score" > 40

Filter ranges were set by binning these values and inspecting 10 random designs in each bin. Examples of graphs docks are shown in Figure S1.

Pocket Determination. Sequence-independent pocket determination was done using a Python script included in the supplemental scripts. This script takes in a pdb and returns (1) the *x*, *y*, and *z* axes of the pockets, (2) the pocket volume, (3) the normalized principal moments of inertia ratios (NPRs), and (4) the design name. Pockets were determined by converting chains A and B of each protein to a space-filling 3D model using voxels [Figure 1a] and then calculating a convex hull around the protein and excluding all space outside of that hull, leaving only empty space within the protein to be considered. The protein itself is converted to a backbone + Cα

https://doi.org/10.1021/acs.biochem.2c00497
Biochemistry 2023, 62, 358–368

only model. Each discrete empty space inside the protein is then filled, and the volume of the largest space is determined to be the pocket of interest [Figure 1b]. The moments of inertia were then calculated as though the pocket had mass of the voxels used to fill the space.

Sequence-dependent pocket determination was performed using the CASTp web server.[14]

NPR Calculation. The NPRs were calculated using the principal moments of inertia, disregarding the sum of the masses, as they cancel out in the ratios [Appendix 6, lines 352−392]. The principal moments of inertia were then sorted from smallest to largest [$I_1 < I_2 < I_3$], and the ratios of I_1/I_3 and I_2/I_3 were output in order to be used as the axes of a triangular graph that is commonly used in pharmacology and drug design in order to describe the overall shape of small molecules.[15]

Sequence Design. Two RosettaScripts[16,17] XML scripts were used to iteratively design the interface and the surface after docking. The first round of design only designed interface residues. An example script is located in Supplemental Scripts. The core positions of the monomer were also restricted from design. All positions were restricted by layer design to only design specific subsets of residues based on secondary structure and burial specifications [Supplemental scripts specified "DesignRestrictions" block]. Two rounds of FastDesign and FastRelax were run, with "beta_cart" weights of the beta score function. Designs were then filtered based on several relevant score terms, including the total score, probability of amino acid phi psi angles, the presence of unsatisfied buried charged atoms, the comparison of the quality of packing, and the quality of residue interactions as a monomer vs as a complex.

The second script consisted of a single round of FastDesign and FastRelax, as this was the third time that the oligomer or its components was being designed, and the overall structural change was expected to be minimal. These proteins were designed using "beta_genpot" weights with an overall net charge score term to ensure a negative overall charge in the protein and increased penalties for unsatisfied buried polar atoms (such as N and O). Protein sequence was designed using Rosetta FastDesign with a new spatial-aggregation-propensity (SAP) score term that was newly implemented into Rosetta. This score term was developed and integrated based on the Developability Index described in Lauer et al.[18]

RosettaScripts XML Scripts are included in Supplemental Scripts.

Expression. Genes encoding each designed protein were synthesized by IDT into PET29b+ plasmids. Each plasmid had a T7 promoter system, and proteins included either an N- or C-terminal His-tag. Proteins were transformed into BL21-(DE3) *E. coli* from New England Biolabs (NEB) and then expressed as 50 mL cultures in 250 mL flasks for initial validation in and then 0.5 L cultures in 2 L flasks if more protein was needed for crystallography. Proteins were expressed in Studiers M2 autoinduction media with 50 μg/mL kanamycin. Precultures were grown at 37 °C for ∼14 h, and cultures were inoculated with 1 mL of preculture per 100 mL of media. Cultures were grown at 37 °C for ∼14 h. Cells were pelleted at 4000*g* for 10 min, after which the supernatant was discarded. Pellets were resuspended in 25 mL of lysis buffer (25 mM Tris HCl pH 8, 150 mM NaCl, 30 mM imidazole, 1 mM PMSF, 1 mM DNase). Cell suspensions were lysed by sonication, and the lysate was clarified at 14 000*g* for 30 min. The His-tagged proteins were bound to Ni-NTA resin (Qiagen) during gravity flow and washed with a wash buffer (25 mM Tris HCl pH 8, 150 mM NaCl, 30 mM imidazole). Protein was eluted with an elution buffer (25 mM Tris HCl pH 8, 150 mM NaCl, 400 mM imidazole). The flow-through was collected and concentrated prior to further purification by SEC/FPLC on a Superdex 200 increase 10/300 GL column in TBS (25 mM Tris pH 8.0, 150 mM NaCl), with 0.5 mL fractionation between 8 and 22 mL.

Mass Spectrometry Barcoding. Eight residue barcode sequences were added at the N-terminus of each design, genes were pooled and transformed into *E. coli* and a single culture was grown and lysed, and the proteins were collectively purified by IMAC. The pooled proteins were then subjected to SEC fractionation, and the peptide barcodes released by proteolytic cleavage and analyzed by mass spectrometry. The frequency of each barcode in each fraction was determined, and SEC profiles were reconstructed for each individual design. Genes were ordered encoding 28 proteins with SEC elution peaks within 1 mL of that expected given the oligomerization state.

Size Exclusion Chromatography and Multiangle Light Scattering (SEC-MALS). After an initial SEC run, fractions containing peaks were pooled and concentrated to 2 mg/mL. A 100 μL aliquot of each sample was then run through a high-performance liquid chromatography system from Agilent using a Superdex 200 10/300 GL column. These fractionation runs were coupled to a Wyatte multiangle light scattering detector in order to determine the absolute molecular weights for each designed protein following the method described in refs 8 and 11.

Small-Angle X-ray Scattering (SAXS). This work was conducted at the Advanced Light Source (ALS), a national user facility operated by Lawrence Berkeley National Laboratory on behalf of the Department of Energy, Office of Basic Energy Sciences, through the Integrated Diffraction Analysis Technologies (IDAT) program, supported by DOE Office of Biological and Environmental Research. Additional support comes from the National Institute of Health project ALS-ENABLE (P30 GM124169) and a High-End Instrumentation Grant S10OD018483. Purified protein was concentrated to 1 and 5 mg/mL and concentrator flow through was collected for buffer subtraction. Proteins were sent in a 96-well plate in 4-well blocks of [buffer, 1 mg/mL protein, 5 mg/mL protein, buffer]. Data files were averaged using SAXS Frameslice Version 1.4.13. Data were compared to design models using the FOXS web server with a fixed *c*1 and *c*2 of 1.0.[19,20] VR was calculated with a maximum *q* value of 0.25, fixed *c*1 and *c*2 to 1.0.

Negative-Stain EM Sample Preparation. Samples were diluted to a range of concentrations between 0.02 and 0.1 mg/mL and 3 μL was negatively stained using Gilder Grids overlaid with a thin layer of carbon and 2% uranyl formate as previously described.[21]

Negative-Stain EM Data Collection and Processing (Three to Sixmer). Data were collected on a Talos L120C 120 kV electron microscope equipped with a CETA camera. An average total of ∼150−300 images was collected per sample by using a random defocus range of 1.3−2.3 μm, with a total exposure of between 30 and 50 $e^-/Å^2$, with a pixel size of 1.54 Å/pixel. All data were automatically acquired using EPU (ThermoFisher Scientific). All data processing was performed using CryoSPARC.[28] The parameters of the contrast transfer function (CTF) were estimated using CTFFIND4,[22] and particles were picked initially in a reference-free manner using

https://doi.org/10.1021/acs.biochem.2c00497
Biochemistry 2023, 62, 358−368

Blob Picker, followed by template picking using well-defined 2D classes of intact nanoparticles. Particles were extracted after correcting for the effect of the CTF for each micrograph. Extracted particles were sorted by reference-free 2D classification.

To create the 3D volume for SG122, data processing was performed using cisTEM.[23,24] The parameters of the contrast transfer function (CTF) were estimated using CTFFIND4,[22] and particles were picked in a reference-free manner in cisTEM. Particles were extracted after correcting for the effect of the CTF for each micrograph with cisTEM. Resulting particles were sorted by reference-free 2D classification over 25 iterations. 3D ab initio was performed in cisTEM using the presence of C1 symmetry, with the subsequent homogeneous refinement steps performed using C3 symmetry. 3D maps were visualized with rigid-body-docked models using ChimeraX.[25]

CryoEM Sample Preparation. A 2 μL aliquot of 1.0 mg/mL of *de novo* designed protein SG135 in 25 mM Tris 150 mM NaCl pH 8.0 was applied to glow-discharged 1.2/1.3 C-flat holey carbon grids. Vitrification was performed on a Mark IV Vitrobot with a wait time of 5 s, with blot times spanning 6.5 to 7.5 s, and a blot force of either 0 or −1 before being immediately plunged frozen into liquid ethane. The sample grids were clipped following standard protocols before loaded into the microscope for imaging.

CryoEM Data Collection. SG135 data collection was performed automatically using Leginon[26] to control a ThermoFisher Titan Krios 300 kV TEM equipped with a standalone K3 Summit direct electron detector.[27] SG135 was collected using counting mode, with random defocus ranges spanning between −0.7 and −2.0 μm using image shift and multiple shots per hole. A total of 3714 movies were collected with a pixel size of 0.84 Å, with a total dose of 63 $e^-/Å^2$.

CryoEM Data Processing. All data processing was carried out in CryoSPARC.[28] Alignment of movie frames was performed using Patch Motion with an estimated B-factor of 500 $Å^2$, with a maximum alignment resolution set to 3. Defocus and astigmatism values were estimated using Patch CTF with default parameters. A total of 2 944 810 particles were initially picked in a reference-free manner using Blob Picker and extracted with a box size of 220 pixels. An initial round of reference-free 2D classification was performed in CryoSPARC with a maximum alignment resolution of 6 Å. A strong preferred orientation was observed from this subset of particle data, with only ∼100 000 particles corresponding to views down the fourfold symmetry axis with clearly defined secondary-structural elements. Upon closer inspection of micrographs, views down the fourfold symmetry axis appeared to be more common in regions of ice that were exceptionally thin and thus may have forced SG135 into this orientation as compared to thicker ice. Thus, an independent round of manual picking particles corresponding to this rare fourfold symmetry view was performed on a small subset of these "thinner ice" micrographs, for a total of 462 manually picked particles. These particles were subjected to an independent round of 2D classification, with their resulting 2D class averages low-pass filtered to 20 Å and used as inputs for a separate round of template picking, resulting in a new subset of 1 393 418 particles picked with templates focusing on the fourfold symmetry axis view angle. A separate round of 2D classification was performed using this subset of particles. Local motion correction was next performed using a select 291 956 particles presenting twofold axis "side views" and "tilted views" from the original Blob Picker set, which were combined with 171 083 fourfold axis "top views" from the template-picked set yielding defined 2D class averages with clear secondary-structural elements present. 3D ab initio was next performed with a maximum alignment resolution set to 8 Å using *C*4 symmetry, followed by 3D refinement resulting in a 4.26 Å map. Subsequent nonuniform refinements improved this global resolution estimation to ∼4.00 Å. A heterogeneous refinement consisting of three classes was initiated using C1 symmetry. The best two resulting classes which had clear fourfold symmetry and proper secondary-structural features were subjected to a final round of nonuniform refinement using C4 symmetry. This process generated a final map with a global resolution estimate of 3.85 Å. Local resolution estimates were in CryoSPARC using an FSC threshold of 0.143. 3D maps for the half maps, final unsharpened maps, and the final sharpened maps for SG135 were deposited in the EMDB under accession number EMD-27903.

CryoEM Model Building and Validation. The *de novo* predicted design model for SG135 (reported here) was used as initial reference for building the final cryoEM structure. The model was manually edited and trimmed using Coot.[29,30] We then further refined the structure in Rosetta using density-guided protocols.[31] EM density-guided molecular dynamics simulations were next performed using Interactive Structure Optimization by Local Direct Exploration (ISOLDE),[32] with manual local inspection and guided correction of rotamers and clashes. ISOLDE runs were performed at a simulated 25 K, with a round of Rosetta typically following this step. Multiple rounds of relaxation, minimization, and ISOLDE were performed on SG135 and were manually inspected for errors after each step. Throughout this process, we applied strict noncrystallographic symmetry constraints in Rosetta.[16] Phenix real-space refinement was subsequently performed as a final step before the final model quality was analyzed using Molprobity.[33,34] Figures were generated using either UCSF Chimera[35] or UCSF ChimeraX.[25] The final structure for SG135 was deposited under PDB accession number8E55.

Crystallography. Proteins were concentrated to a final concentration between 20 and 63 mg/mL, depending on the solubility of proteins at high concentrations. All crystallization trials were carried out at 20 °C in 96-well format using the sitting-drop method. Diffraction quality crystals appeared in 25 mM Tris, 150 mM NaCl Index well B11 with 2.1 M DL-malic acid pH 7.0 additive. Crystals were subsequently harvested in a cryo-loop and flash frozen directly in liquid nitrogen for synchrotron data collection.

Data collection from crystals of *de novo* designed protein SG122 was performed with synchrotron radiation at the Advanced Photon Source (APS) on 24ID-C. X-ray intensities and data reduction were evaluated and integrated using XDS[36] and merged/scaled using Pointless/Aimless in the CCP4 program suite.[37] Low-resolution structure determination and refinement starting phases were obtained by molecular replacement using Phaser[38] using the designed model for the structures. Following molecular replacement, the models were improved using phenix.autobuild. Structures were refined in Phenix.[39] Model building was performed using COOT.[29] The final model was evaluated using MolProbity.[34] The final structure for SG122 was deposited under the PDB accession number8E1E.

https://doi.org/10.1021/acs.biochem.2c00497
Biochemistry 2023, 62, 358−368

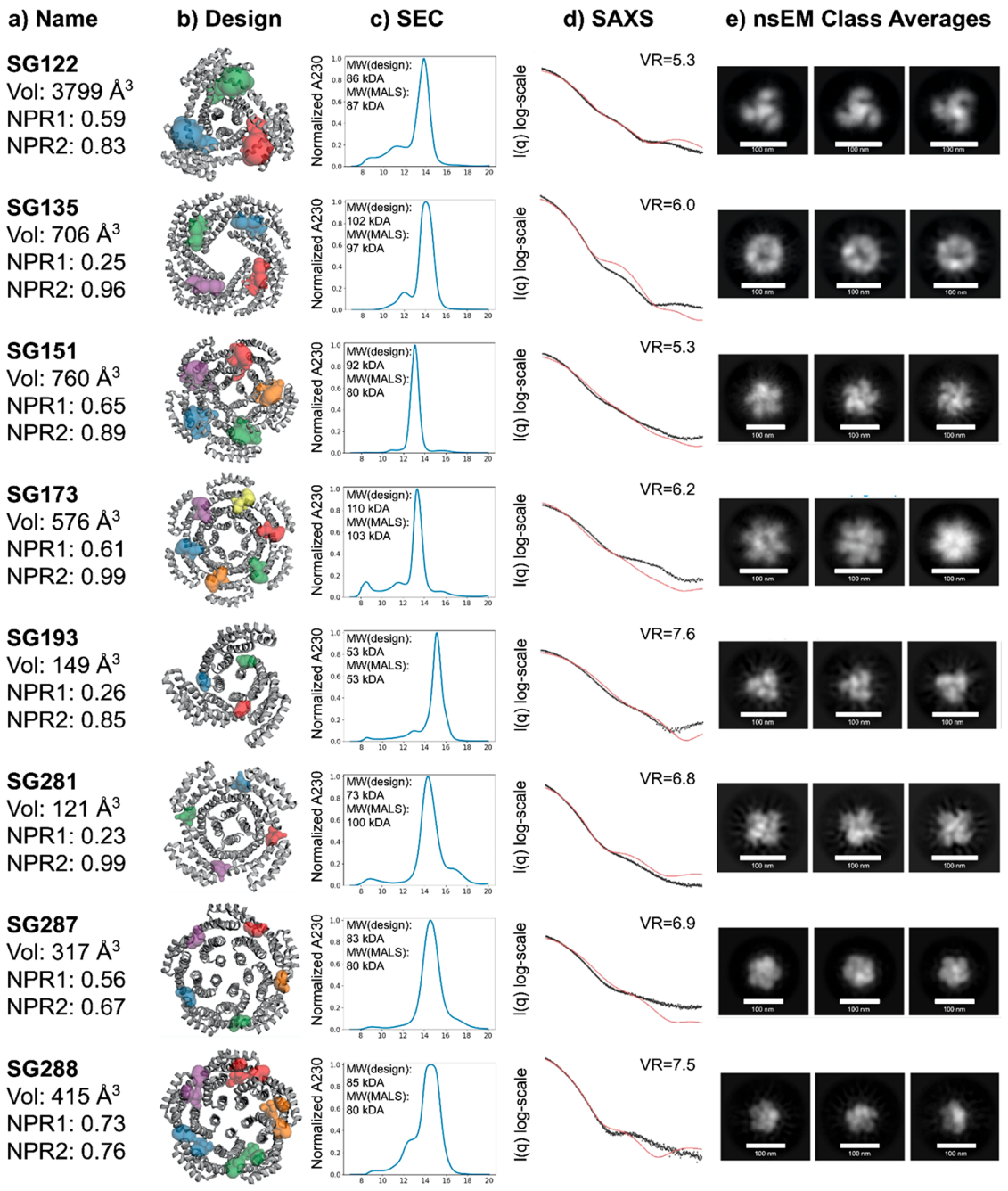


Figure 2. Experimental characterization of designed homo-oligomers. (a) Design name, pocket volume, and dimensions calculated in a sequence-independent manner using only Cα backbone positions [Supplemental Scripts: Pocket determination script]. (b) Image of the designed protein models (gray) with pockets highlighted in color as determined by the CASTp web server;[14] these parameters were computed using all designed amino acids and not just the backbone positions. (c) SEC traces from initial purification of design on a Superdex S200 Increase 10/300 GL column. (d) Experimental SAXS curve (right, black dots) and calculated SAXS curve from the design (right, red line) with volatility ratio (VR) listed in the top right corner. (e) Negative-stain EM averages of the designed protein.

RESULTS

Monomer Generation and Oligomeric Docking. A large set of designed helical repeat (DHR) proteins was generated using the Rosetta based helical extension method described previously,[40] which first assembles helical fragments into helical bundles by aligning overlapping helical and loop fragments before subsequently recombining the generated helical bundles into curved DHRs. Monomers were then filtered and separated into two groups based on measured distances between proximal helices, which ultimately determines the overall bend of the monomer. These two groups of DHRs are separated into a less curved set of DHRs named "bend distance 4" (bd4) and a more curved set "bend-distance 5" (bd5). In the bd4 group, curvature arises by shifting the overlap between helices such that the N-terminus of the helix interacts with central residues on helix i + 2 of the repeat protein [Figure S1a]. The bd5 group has larger distances between helices with the N- and C-terminal segments connected by single "hinge" helices at several points along their length. The monomers curve around the axis of the hinge

https://doi.org/10.1021/acs.biochem.2c00497
Biochemistry 2023, 62, 358−368

helices, which keeps the ends of the helices aligned across the length of the monomer [Figure S1b].

Monomers were docked into symmetric three, four, five, and sixfold cyclic assemblies using the Rosetta rpxDock method.[11,12] A set of 4618 helical repeat monomers were docked in 4 symmetries, and the top 100 docks of each symmetry were output, resulting in 1.85 million docks. Oligomers were then filtered based on the number of contacts in the interface between subunits and by the predicted quality of the interface interactions, as described in the Materials and Methods. Examples of high- and low-scoring docks as well as graphs illustrating filtering cutoffs can be found in Figure S1c–h.

Pocket Determination and Variety. Pockets were identified by filling all empty space between backbone + Cα only models of chains A and B of each oligomer [Figure 1a]. The largest filled space was determined to be the pocket of interest [Figure 1b], and the length, width, height [Figure 1c], and volume of the pocket were computed. The shape space spanned by the pockets [Figure 1d] was evaluated by calculating the normalized primary moments of inertia ratio (NPR), which were described by Sauer et al. as a method for visually plotting shape space diversity.[15] Rod-like [Figure 1e], disk-like [Figure 1f], and spherical [Figure 1g] pockets are found near (0,1), (0.5, 0.5), and (1,1) respectively in Figure 1.

With this broad coverage, appropriate starting scaffolds can be chosen for designing binders to a wide range of small molecules with diverse sizes and shapes. The pocket parameters in Figure 1d are computed using the backbone coordinates only and are not sequence-dependent. Figure 1 shows sequence-independent pocket determination metrics alongside designs with sequence-independent pockets highlighted [Figure 1e–g]. Figure 2 shows sequence-independent calculations [Figure 2a] alongside sequence-dependent pockets determined using CASTp[14] [Figure 2b]. During sequence redesign for binding a specific small-molecule target, shape complementarity will increase further as side chains are placed that make specific interactions with the ligand.

Protein Design and Experimental Validation. The monomer–monomer interfaces and oligomer surfaces were iteratively designed using RosettaScripts[16,17] xml scripts following the protocol used for C2 symmetric homo-oligomers.[8] Initially, our design process modified only the interface, as the initial monomers were already assigned a sequence. However, such designs were found to have a high rate of insoluble expression. Out of 25 designed proteins expressed in *E. coli*, only 9 had detectable soluble expression, and only 6 proteins had yields of >220 μg (from 50 mL cultures) needed for size exclusion chromatography coupled to multiangle light scattering (SEC-MALS) and mass spectrometry (MS).

To reduce surface hydrophobicity, which likely caused the aggregation and insoluble expression, we redesigned the oligomer surface, filtering out protein designs based on the number of hydrophobic residues with more than 35% of their surface exposed, a counter of the length longest hydrophobic stretch on the protein, and the percent hydrophobicity of the interface, among other metrics [Supplemental Scripts "Surface design script"]. A total of 104 designs were expressed and purified as described in the Materials and Methods, and only a single instance of insoluble expression was observed. A total of 58 cultures had a yield of >220 μg from a 50 mL expression for a total soluble expression rate of 58/105, or 54%. This is more than double the previous soluble expression rate.

Proteins with an expression yield of >220 μg were then analyzed using SEC-MALS. Designed proteins with a measured oligomeric mass with ≤13% discrepancy from design mass were considered to have a validated oligomeric state. Of the 58 proteins, 27 were found to be within this range (26% of all the designs and 47% of those that had soluble expression). An additional 10 proteins had a discrepancy between 13 and 24%, perhaps indicating a mix of oligomeric states. When correlating to later small-angle X-ray scattering (SAXS) data, it was found that six of these (SG138, SG162, SG165, SG191, SG192, and SG222) have SAXS curves with volatility ratios (VRs) indicating a close match to the design (VR ≤ 7.6) [Table S1]. This suggests that they adopt the correct overall structure. SEC curves of all soluble designs are divided based on SEC-MALS validation and listed in Figure S2 for matching oligomeric states and Figure S3 for deviating ones. Validated proteins show single major monodisperse peaks at the appropriate elution time for their oligomeric state [Figure 2c].

A total of 30 of the designs were sent to SYBLYS for SAXS analysis at the Advanced Light Source (ALS).[41−44] Of the 30, 20 had SAXS profiles close to those computed from the design models. VR was used, as it has been shown to be more accurate than X^2 in assessing SAXS data fit to designed proteins.[45,46] SAXS curves for a select set of validated oligomers can be compared to complementary validation data in Figure 2d. All SAXS data can be found filtered based on a VR < 7.6 cutoff in Figures S4 and S5. Overall, 19% of all designs and 67% of all purified proteins tested formed the designed structure based on the SAXS VR metric.

To enable larger scale testing of designs, we employed a peptide barcoding method to characterize a library of 4965 oligomers designed using the approach described in Feldman et al[47] and used in Kim et al.[48] This method uses short peptide sequences which are cleavable and distinguishable from each other via mass spectrometry in order to label protein libraries. A total of 28 proteins which were validated through this method were chosen to be tested individually. When expressed, all had some soluble expression, 26 of which had a soluble yield of >220 μg from a 50 mL expression. Of the 26 designed oligomers, 20 had a less than 13% deviation between the designed MW and the MW determined by SEC-MALS, and 15 had SAXS VR of < 7.6. The slightly lower SAXS success rate than in the nonbarcoded set reflects the lack of SEC-MALS based filtering; such filtering would have eliminated three designs with a VR < 7.6, including sg181, for which nsEM 2D class averages are close to the design model [Figure 2]. Full validation data for this set is listed in Table S1.

Designs with monodisperse SEC peaks and MALS data indicating a correct or near-correct oligomeric state were screened by negative-stain electron microscopy (nsEM). From these designs, we obtained 2D class averages for eight proteins that are clearly consistent with their designed oligomeric state [Figure 2e]. Designs where nsEM data are less clear or differ from the intended oligomeric state are shown in Figure S6. In most cases, the proteins have a preferred orientation along the symmetry axis, likely due to the overall cyclical disk shape. This made it difficult to get an accurate volume reconstruction of the designs, though it did enable clear determination of the oligomeric state by inspection in many cases. Together the SEC-MALS, SAXS, and nsEM data suggest that the scaffolds are assembling as designed.

Characterization by X-ray Crystallography. Crystal trays were set up for all 37 proteins with oligomeric states

https://doi.org/10.1021/acs.biochem.2c00497
Biochemistry 2023, 62, 358−368

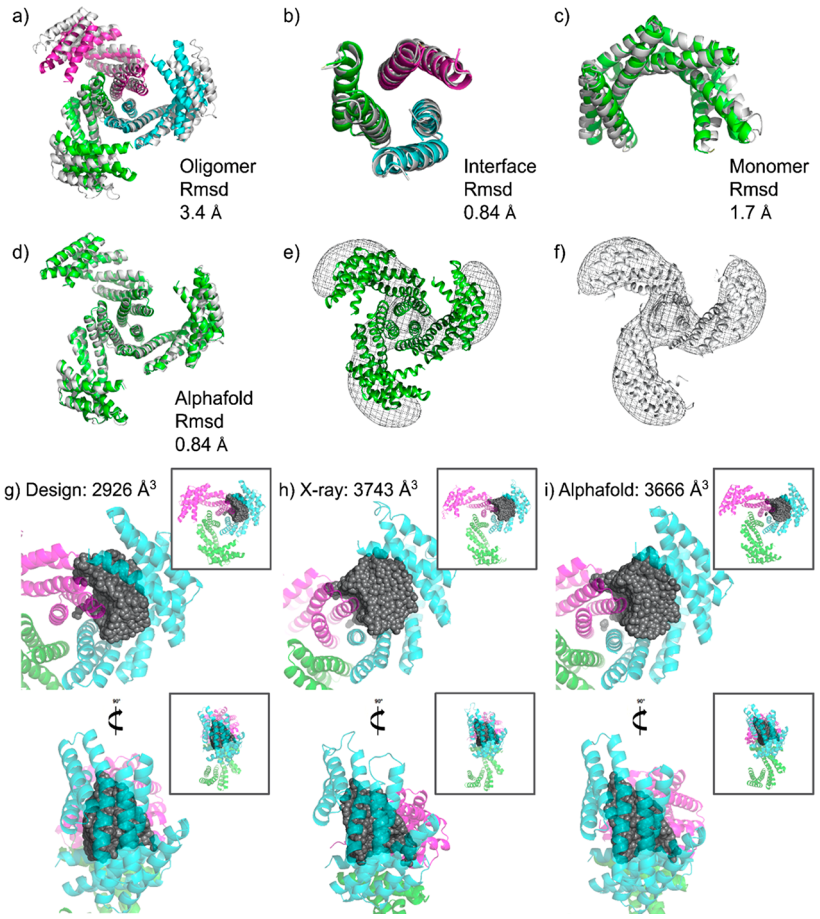


Figure 3. Crystal structure of SG122. (a) Alignment of designed model (green/cyan/magenta by chain) to crystal structure (white). (b) Alignment of designed interface (green/cyan/magenta by chain) to crystal structure (white). (c) Alignment of chain A of design (green) to chain A of crystal structure (white). (d) Alignment of Alphafold-2 predicted oligomer (green) to crystal structure (white). (e) Designed model (green) fit to negative-stain EM density map (gray mesh). (f) Crystal structure (white) fit to negative-stain EM density map (gray mesh). (g–i) Sequence-independent pockets are shown in gray, with chains A and B in magenta and cyan. Pockets are viewed from the axis of symmetry or rotated 90°. Full structure shown in the top right section. The widening of the arms causes the pocket to be larger in the structure than as designed, though the Alphafold2 model also predicts a larger pocket size and is far more accurate to the structure.

determined by SEC-MALS close to the computational design models. The crystal structure of SG122 at 4.2 Å revealed an oligomer very similar to the design model [Figure 3a]. The interface and each monomeric subunit had a root mean squared deviation (RMSD) of less than 1 Å and less than 2 Å, respectively [Figure 3b,c]. This combination shows the threefold structure to be highly accurate along the axis of symmetry, with deviations from design increasing along the arms. This widening of the arms increases the sequence-independent pocket volume from 2926 to 3743 Å^3 while retaining the overall largely spherical shape of the pocket [Figure 3g–h]. The overall structure was predicted with even higher accuracy using Alphafold2[49] [Figure 3d], and the model generated using Alphafold2 also had a more accurate sequence-independent pocket volume, at 3666 Å^3 [Figure 3i]. These data match closely to ∼20 Å volume generated from negative-stain data [Figure 3e] consistent with the crystal structure [Figure 3f]. Full crystallographic data collection and refinement statistics are listed in Table S2.

Characterization by Cryo-Electron Microscopy. While the assembly states for many of our designed homo-oligomers were confidently validated by nsEM, a subset of our designs proved more difficult to characterize by this structural method alone. Most designs were observed to readily adopt an easily identifiable propeller-like configuration [Figure 2e], but designs such as SG135 were observed to adopt a more circular and compact oligomeric profile in-line with design models. 2D classification by nsEM confirmed the designed ring-like organization of SG135, though the unique geometry of this assembly prevented confident assignment to the C4 symmetry group. We thus further characterized SG135 by cryo-electron microscopy (cryoEM). SG135 behaved well in ice and allowed for confident identification of the designed C4 symmetry following 2D classification [Figure S7]. A cryoEM 3D reconstruction was solved for SG135 to a global resolution estimate of 3.85 Å, with the core of the protein resolving as high as 3.2 Å. Full cryoEM data collection information is listed in Table S3. At this resolution range, we were able to unambiguously place the backbone for the entire protein as well as many side-chain amino acid residues near the core. Model building confirmed that both the monomeric subunits and overall oligomeric geometry of SG135 were nearly identical to the corresponding design model, with RMSDs of 2.14 and 3.95 Å, respectively [Figure 4b,c].

Using sequence-independent volume calculations, we found that the narrow rod-like shape of the designed pocket remained conserved in the cryoEM model, with the volume of pocket for the original design calculated to be 706 and 2008

https://doi.org/10.1021/acs.biochem.2c00497
Biochemistry 2023, 62, 358−368

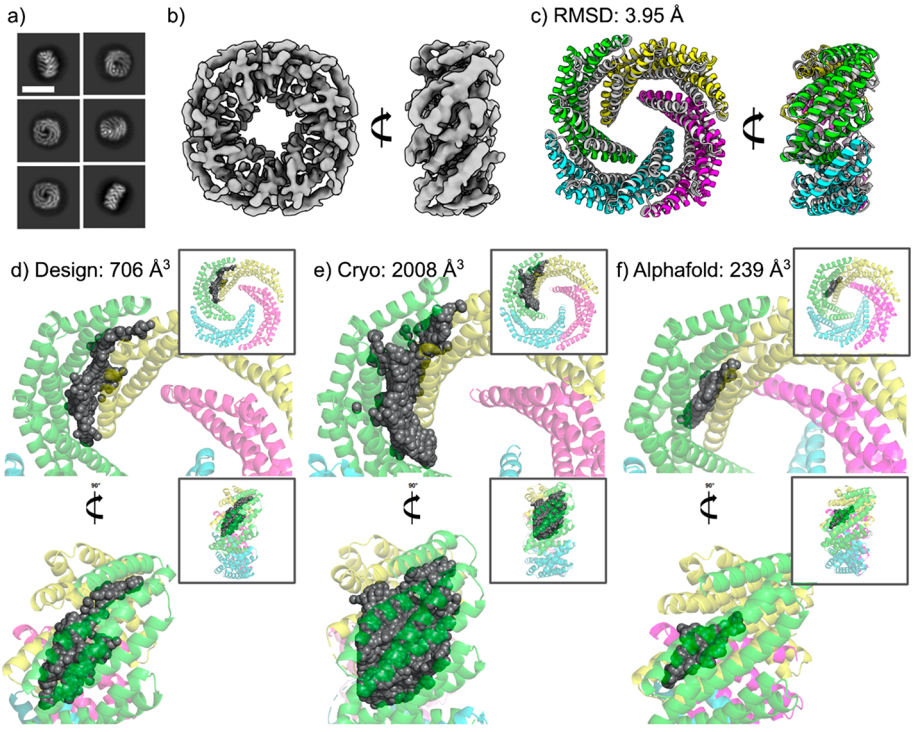


Figure 4. CryoEM structure of SG135. (a) CryoEM 2D class averages showing clear secondary-structural information for SG135 particles. (b) 3.85 Å CryoEM map visualized along orthogonal view axes. (c) SG135 cryoEM structure (gray) aligned to designed structure (colored by chain) viewed along orthogonal axes. (d–f) Sequence-independent pockets shown in gray with protein colored by chain. Protein pockets are viewed from the axis of symmetry (top) or rotated 90° (bottom). Full structure shown in the top right section. The cryoEM structure shows an increase in pocket size from the design model, while the Alphafold-Multimer[50] model predicts a significant reduction in size.

Å^3 in the solved structure [Figure 4d,e]. Alphafold2, in contrast, predicted a collapsing of the pocket, with a volume of only 239 Å^3 [Figure 4f]. This discrepancy, which shows a much larger pocket in the cryoEM structure, could be explained by an observed region of difficult-to-identify floating rod-shaped density in the designed pocket of SG135. We hypothesize that this may correspond to the presence of an unknown ligand serendipitously bound to the computationally designed pocket [Figure S8a]. If this is the case, this serendipitous ligand could be influencing the structure toward a larger pocket size. Density-guided molecular dynamics simulations, however, predicted that this could also be explained by an uncommon rotameric assignment for Arg190 where the amine groups of Arg190 occupy a portion of this space [Figure S8b]. We highlight that these final MD-derived coordinates do not explain a large portion of the surrounding rod-like density and also place the hydrocarbon chain of Arg190 outside of the map [Figure S8c]. As a result, we believe both explanations of the density are plausible and thus interpret them as such. Taken together, these results highlight the robustness of our design protocol for generating homo-oligomers with asymmetric pockets.

Comparison to Alphafold Predictions. Given the success of Alphafold2 for accurately predicting the general structures for both SG122 and SG135, we decided to use Alphafold-Multimer[50] to generate oligomer models for all pocketed protein designs for which we were able to express. We found that designs where the predicted structures were close to the design models were more likely to assemble to the target designed state following expression. Indeed, below a 7.5 A RMSD between the design model and the Alphafold-Multimer prediction, there was a considerable enrichment of experimentally validated designs [Tables S1 and S4 and Figure S9]. Based on the agreement between the Alphafold predicted and Rosetta designed structure, and both X-ray crystallography and cryoEM validating these Alphafold predictions, future design efforts using our design approach could incorporate Alphafold-Multimer scoring metrics as an additional filter prior to experimental testing.

CONCLUSION

We have developed a general method to generate a large set of proteins with a variety of asymmetric pockets which cover a wide area of the possible shape space [Figure 1d]. As the residues forming the pockets are distinct from the residues in the core and interface of the protein, these scaffolds are primed for sequence modification for the purpose of introducing functionality with minimal disruption of structurally important core features. Because the pockets are located at the asymmetric interface between adjacent chains, rather than along the symmetry axis at the center, they are not constrained by symmetry. Thus, in principle, almost any shape can be achieved with this design pipeline to help guide binding of a wide variety of target ligands and substrates. While there is variation between the designed pocket size and the experimentally determined pocket size that may necessitate characterization before downstream applications, 53% of designs tested by SEC-MALS were determined to have the correct oligomerization state, and 65% of tested designs have a VR < 7.6 by SAXS. Crystal and cryoEM structures show that designed pocket shapes are recapitulated in the empirically determined structures, with a putative ligand potentially bound within the pocket of our cryoEM structure of SG135. The computational methods we describe here enable us to turn a

https://doi.org/10.1021/acs.biochem.2c00497
Biochemistry 2023, 62, 358−368

small number of stable base monomers into a diverse set of inert oligomeric scaffolds with a wide range of asymmetric pocket sizes and shapes that can quickly and easily be filtered to fit functional design specifications. Thus, these validated scaffolds can be used for the functional design of small-molecule binding, cyclic peptide binding, serine hydrolase design, and more.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.biochem.2c00497.

Filtering metrics for creating homo-oligomers, all SEC and SAXS traces for all proteins tested, additional nsEM class averages, additional cryoEM data and X-ray crystallography data, comparison of Alphafold predicted structures to the design structure (PDF)

Protein validation data (XLSX)

Protein sequences of all the tested *de novo* proteins as well as example scripts for design and filtering (PDF)

■ AUTHOR INFORMATION

Corresponding Author

David Baker − *Department of Biochemistry, University of Washington, Seattle, Washington 98195, United States; Institute for Protein Design and Howard Hughes Medical Institute, University of Washington, Seattle, Washington 98195, United States*; Email: dabaker@uw.edu

Authors

Stacey R Gerben − *Department of Biochemistry, University of Washington, Seattle, Washington 98195, United States; Institute for Protein Design, University of Washington, Seattle, Washington 98195, United States;* orcid.org/0000-0003-0313-6248

Andrew J Borst − *Department of Biochemistry, University of Washington, Seattle, Washington 98195, United States; Institute for Protein Design, University of Washington, Seattle, Washington 98195, United States;* orcid.org/0000-0003-4297-7824

Derrick R Hicks − *Department of Biochemistry, University of Washington, Seattle, Washington 98195, United States; Institute for Protein Design, University of Washington, Seattle, Washington 98195, United States*

Isabelle Moczygemba − *Institute for Protein Design, University of Washington, Seattle, Washington 98195, United States*

David Feldman − *Department of Biochemistry, University of Washington, Seattle, Washington 98195, United States; Institute for Protein Design, University of Washington, Seattle, Washington 98195, United States*

Brian Coventry − *Department of Biochemistry, University of Washington, Seattle, Washington 98195, United States; Institute for Protein Design, University of Washington, Seattle, Washington 98195, United States*

Wei Yang − *Department of Biochemistry, University of Washington, Seattle, Washington 98195, United States; Institute for Protein Design, University of Washington, Seattle, Washington 98195, United States*

Asim K. Bera − *Department of Biochemistry, University of Washington, Seattle, Washington 98195, United States; Institute for Protein Design, University of Washington, Seattle, Washington 98195, United States*

Marcos Miranda − *Department of Biochemistry, University of Washington, Seattle, Washington 98195, United States; Institute for Protein Design, University of Washington, Seattle, Washington 98195, United States;* orcid.org/0000-0002-2839-6997

Alex Kang − *Department of Biochemistry, University of Washington, Seattle, Washington 98195, United States; Institute for Protein Design, University of Washington, Seattle, Washington 98195, United States*

Hannah Nguyen − *Department of Biochemistry, University of Washington, Seattle, Washington 98195, United States; Institute for Protein Design, University of Washington, Seattle, Washington 98195, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.biochem.2c00497

Author Contributions

The project was conceptualized by S.G., D.R.H., and D.B. Protein backbone generation and initial sequence design were done by W.Y., B.C., and D.R.H. RPX docking and filtering were done by I.M., D.R.H., and S.G. Sequence redesign was done by S.G. and D.R.H. Proteins were expressed and purified by S.G. Large-scale screening by mass spectrometry was done by D.F. SEC-MALS was performed by M.M. Negative-stain EM was done by S.G. with training from A.J.B. CryoEM was done by A.J.B. Crystal tray setup and screening were done by A.K. and H.N. Crystal structure determination was done by A.B. The original draft was prepared by S.G. and A.J.B and reviewed and edited by all authors.

Funding

Defense Threat Reduction Agency (DTRA) Grant No HDTRA1−19−1−0003; National Institute on Aging Grant No 5U19AG065156−02; Amgen Donation Agreement TED Audacious Project.

Notes

The authors declare the following competing financial interest(s): A record of innovation (ROI) has been filed for the proteins described in this manuscript.

■ ACKNOWLEDGMENTS

We would like to thank the staff at the Advanced Light Source 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) and NIGMS supported ALS-ENABLE (GM124169-01) and National Institute of Health project MINOS (R01GM105404). Crystallography work was conducted at the Advanced Photon Source and Advanced Light Source. Advanced Photon Source (APS) Northeastern Collaborative Access Team beamlines are funded by the National Institute of General Medical Sciences from the National Institutes of Health (P30 GM124165). This research used resources of the Advanced Photon Source, a U.S. Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under Contract No. DE-AC02-06CH11357. Advanced Light Source (ALS) is a national user facility operated by Lawrence Berkeley National Laboratory on behalf

of the Department of Energy, Office of Basic Energy Sciences, through the Integrated Diffraction Analysis Technologies (IDAT) program, supported by DOE Office of Biological and Environmental Research. Additional support comes from the National Institute of Health project ALS-ENABLE (P30 GM124169) and a High-End Instrumentation Grant S10OD018483. We would like to thank the IPD Core Research Staff, Including Michelle DeWitt, Analisa Murray, and Piper Heine for production of proteins in quantities necessary for crystallography and Xinting Li for mass spectrometry that was used to verify the proteins characterized.

■ ABBREVIATIONS

SAXS, small-angle X-ray scattering; VR, volatility ratio; nsEM, negative-stain electron microscopy; cryoEM, cryogenic electron microscopy; DHR, designed helical repeat; bd4, less curved DHRs "bend distance 4"; bd5, more curved DHRs "bend distance 5"; NPR, normalized primary moments of inertia ratio; SEC-MALS, size exclusion chromatography coupled to multiangle light scattering; MS, mass spectrometry; ALS, Advanced Light Source; RMSD, root mean squared deviation

■ REFERENCES

(1) Feldmeier, K.; Höcker, B. Computational Protein Design of Ligand Binding and Catalysis. *Curr. Opin. Chem. Biol.* **2013**, *17* (6), 929−933.
(2) Yang, W.; Jones, L. M.; Isley, L.; Ye, Y.; Lee, H.-W.; Wilkins, A.; Liu, Z.-R.; Hellinga, H. W.; Malchow, R.; Ghazi, M.; Yang, J. J. Rational Design of a Calcium-Binding Protein. *J. Am. Chem. Soc.* **2003**, *125* (20), 6165−6171.
(3) Regan, L. Protein Design: Novel Metal-Binding Sites. *Trends Biochem. Sci.* **1995**, *20* (7), 280−285.
(4) Marcos, E.; Basanta, B.; Chidyausiku, T. M.; Tang, Y.; Oberdorfer, G.; Liu, G.; Swapna, G. V. T.; Guan, R.; Silva, D.-A.; Dou, J.; Pereira, J. H.; Xiao, R.; Sankaran, B.; Zwart, P. H.; Montelione, G. T.; Baker, D. Principles for Designing Proteins with Cavities Formed by Curved β Sheets. *Science* **2017**, *355* (6321), 201−206.
(5) Basanta, B.; Bick, M. J.; Bera, A. K.; Norn, C.; Chow, C. M.; Carter, L. P.; Goreshnik, I.; Dimaio, F.; Baker, D. An Enumerative Algorithm for de Novo Design of Proteins with Diverse Pocket Structures. *Proc. Natl. Acad. Sci. U. S. A.* **2020**, *117* (36), 22135−22145.
(6) Dou, J.; Vorobieva, A. A.; Sheffler, W.; Doyle, L. A.; Park, H.; Bick, M. J.; Mao, B.; Foight, G. W.; Lee, M. Y.; Gagnon, L. A.; Carter, L.; Sankaran, B.; Ovchinnikov, S.; Marcos, E.; Huang, P.-S.; Vaughan, J. C.; Stoddard, B. L.; Baker, D. De Novo Design of a Fluorescence-Activating β-Barrel. *Nature* **2018**, *561* (7724), 485−491.
(7) Klima, J. C.; Doyle, L. A.; Lee, J. D.; Rappleye, M.; Gagnon, L. A.; Lee, M. Y.; Barros, E. P.; Vorobieva, A. A.; Dou, J.; Bremner, S.; Quon, J. S.; Chow, C. M.; Carter, L.; Mack, D. L.; Amaro, R. E.; Vaughan, J. C.; Berndt, A.; Stoddard, B. L.; Baker, D. Incorporation of Sensing Modalities into de Novo Designed Fluorescence-Activating Proteins. *Nat. Commun.* **2021**, *12* (1), 856.
(8) Hicks, D. R.; Kennedy, M. A.; Thompson, K. A.; DeWitt, M.; Coventry, B.; Kang, A.; Bera, A. K.; Brunette, T. J.; Sankaran, B.; Stoddard, B.; Baker, D. De Novo Design of Protein Homodimers Containing Tunable Symmetric Protein Pockets. *Proc. Natl. Acad. Sci. U. S. A.* **2022**, *119* (30), No. e2113400119.
(9) Divine, R.; Dang, H. V.; Ueda, G.; Fallas, J. A.; Vulovic, I.; Sheffler, W.; Saini, S.; Zhao, Y. T.; Raj, I. X.; Morawski, P. A.; Jennewein, M. F.; Homad, L. J.; Wan, Y.-H.; Tooley, M. R.; Seeger, F.; Etemadi, A.; Fahning, M. L.; Lazarovits, J.; Roederer, A.; Walls, A. C.; Stewart, L.; Mazloomi, M.; King, N. P.; Campbell, D. J.; McGuire, A. T.; Stamatatos, L.; Ruohola-Baker, H.; Mathieu, J.; Veesler, D.; Baker, D. Designed Proteins Assemble Antibodies into Modular Nanocages. *Science* **2021**, *372* (6537), eabd9994.
(10) Laniado, J.; Yeates, T. O. A Complete Rule Set for Designing Symmetry Combination Materials from Protein Molecules. *Proc. Natl. Acad. Sci. U. S. A.* **2020**, *117* (50), 31817−31823.
(11) Fallas, J. A.; Ueda, G.; Sheffler, W.; Nguyen, V.; McNamara, D. E.; Sankaran, B.; Pereira, J. H.; Parmeggiani, F.; Brunette, T. J.; Cascio, D.; Yeates, T. R.; Zwart, P.; Baker, D. Computational Design of Self-Assembling Cyclic Protein Homo-Oligomers. *Nat. Chem.* **2017**, *9* (4), 353−360.
(12) Sheffler, W.; Yang, E. C.; Dowling, Q.; Hsia, Y.; Fries, C. N.; Stanislaw, J.; Langowski, M.; Brandys, M.; Khmelinskaia, A.; King, N. P.; Baker, D. Fast and Versatile Sequence-Independent Protein Docking for Nanomaterials Design Using RPXDock. *bioRxiv* **2022**, 2022.10.25.513641.
(13) Mariani, V.; Biasini, M.; Barbato, A.; Schwede, T. lDDT: A Local Superposition-Free Score for Comparing Protein Structures and Models Using Distance Difference Tests. *Bioinformatics* **2013**, *29* (21), 2722−2728.
(14) Tian, W.; Chen, C.; Lei, X.; Zhao, J.; Liang, J. CASTp 3.0: Computed Atlas of Surface Topography of Proteins. *Nucleic Acids Res.* **2018**, *46* (W1), W363−W367.
(15) Sauer, W. H. B.; Schwarz, M. K. Molecular Shape Diversity of Combinatorial Libraries: A Prerequisite for Broad Bioactivity. *J. Chem. Inf. Comput. Sci.* **2003**, *43* (3), 987−1003.
(16) DiMaio, F.; Leaver-Fay, A.; Bradley, P.; Baker, D.; André, I. Modeling Symmetric Macromolecular Structures in Rosetta3. *PLoS One* **2011**, *6* (6), No. e20450.
(17) Fleishman, S. J.; Leaver-Fay, A.; Corn, J. E.; Strauch, E.-M.; Khare, S. D.; Koga, N.; Ashworth, J.; Murphy, P.; Richter, F.; Lemmon, G.; Meiler, J.; Baker, D. RosettaScripts: A Scripting Language Interface to the Rosetta Macromolecular Modeling Suite. *PLoS One* **2011**, *6* (6), No. e20161.
(18) Lauer, T. M.; Agrawal, N. J.; Chennamsetty, N.; Egodage, K.; Helk, B.; Trout, B. L. Developability Index: A Rapid in Silico Tool for the Screening of Antibody Aggregation Propensity. *J. Pharm. Sci.* **2012**, *101* (1), 102−115.
(19) Schneidman-Duhovny, D.; Hammel, M.; Tainer, J. A.; Sali, A. Accurate SAXS Profile Computation and Its Assessment by Contrast Variation Experiments. *Biophys. J.* **2013**, *105* (4), 962−974.
(20) Schneidman-Duhovny, D.; Hammel, M.; Tainer, J. A.; Sali, A. FoXS, FoXSDock and MultiFoXS: Single-State and Multi-State Structural Modeling of Proteins and Their Complexes Based on SAXS Profiles. *Nucleic Acids Res.* **2016**, *44* (W1), W424−W429.
(21) Veesler, D.; Cupelli, K.; Burger, M.; Gräber, P.; Stehle, T.; Johnson, J. E. Single-Particle EM Reveals Plasticity of Interactions between the Adenovirus Penton Base and Integrin $\alpha V\beta 3$. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111* (24), 8815−8819.
(22) Rohou, A.; Grigorieff, N. CTFFIND4: Fast and Accurate Defocus Estimation from Electron Micrographs. *J. Struct. Biol.* **2015**, *192* (2), 216−221.
(23) Grant, T.; Rohou, A.; Grigorieff, N. cisTEM, User-Friendly Software for Single-Particle Image Processing. *Elife* **2018**, *7*, e35383.
(24) Grigorieff, N.; Grant, T.; Rohou, A. IUCr. cisTEM: User-Friendly Software for Single-Particle Image Processing. *Acta Crystallographica Section A: Foundations and Advances* **2017**, *73*, C1368−C1368.
(25) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Meng, E. C.; Couch, G. S.; Croll, T. I.; Morris, J. H.; Ferrin, T. E. UCSF ChimeraX: Structure Visualization for Researchers, Educators, and Developers. *Protein Sci.* **2021**, *30* (1), 70−82.
(26) Suloway, C.; Pulokas, J.; Fellmann, D.; Cheng, A.; Guerra, F.; Quispe, J.; Stagg, S.; Potter, C. S.; Carragher, B. Automated Molecular Microscopy: The New Leginon System. *J. Struct. Biol.* **2005**, *151* (1), 41−60.
(27) Sun, M.; Azumaya, C. M.; Tse, E.; Bulkley, D. P.; Harrington, M. B.; Gilbert, G.; Frost, A.; Southworth, D.; Verba, K. A.; Cheng, Y.; Agard, D. A. Practical Considerations for Using K3 Cameras in CDS

Mode for High-Resolution and High-Throughput Single Particle Cryo-EM. *J. Struct. Biol.* **2021**, *213* (3), 107745.
(28) Punjani, A.; Rubinstein, J. L.; Fleet, D. J.; Brubaker, M. A. cryoSPARC: Algorithms for Rapid Unsupervised Cryo-EM Structure Determination. *Nat. Methods* **2017**, *14* (3), 290−296.
(29) Emsley, P.; Cowtan, K. Coot: Model-Building Tools for Molecular Graphics. *Acta Crystallogr. D Biol. Crystallogr.* **2004**, *60* (12), 2126−2132.
(30) Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K. Features and Development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **2010**, *66* (4), 486−501.
(31) Wang, R. Y.-R.; Song, Y.; Barad, B. A.; Cheng, Y.; Fraser, J. S.; DiMaio, F. Automated Structure Refinement of Macromolecular Assemblies from Cryo-EM Maps Using Rosetta. *Elife* **2016**, *5*, e17219.
(32) Croll, T. I. ISOLDE: A Physically Realistic Environment for Model Building into Low-Resolution Electron-Density Maps. *Acta Crystallogr. D Struct Biol.* **2018**, *74* (6), 519−530.
(33) Chen, V. B.; Arendall, W. B., 3rd; Headd, J. J.; Keedy, D. A.; Immormino, R. M.; Kapral, G. J.; Murray, L. W.; Richardson, J. S.; Richardson, D. C. MolProbity: All-Atom Structure Validation for Macromolecular Crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **2010**, *66* (1), 12−21.
(34) Williams, C. J.; Headd, J. J.; Moriarty, N. W.; Prisant, M. G.; Videau, L. L.; Deis, L. N.; Verma, V.; Keedy, D. A.; Hintze, B. J.; Chen, V. B.; Jain, S.; Lewis, S. M.; Arendall, W. B., 3rd; Snoeyink, J.; Adams, P. D.; Lovell, S. C.; Richardson, J. S.; Richardson, D. C. MolProbity: More and Better Reference Data for Improved All-Atom Structure Validation. *Protein Sci.* **2018**, *27* (1), 293−315.
(35) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera-a Visualization System for Exploratory Research and Analysis. *J. Comput. Chem.* **2004**, *25* (13), 1605−1612.
(36) Kabsch, W. XDS. *Acta Crystallogr. D Biol. Crystallogr.* **2010**, *66* (2), 125−132.
(37) Winn, M. D.; Ballard, C. C.; Cowtan, K. D.; Dodson, E. J.; Emsley, P.; Evans, P. R.; Keegan, R. M.; Krissinel, E. B.; Leslie, A. G. W.; McCoy, A.; McNicholas, S. J.; Murshudov, G. N.; Pannu, N. S.; Potterton, E. A.; Powell, H. R.; Read, R. J.; Vagin, A.; Wilson, K. S. Overview of the CCP4 Suite and Current Developments. *Acta Crystallogr. D Biol. Crystallogr.* **2011**, *67* (4), 235−242.
(38) McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J. Phaser Crystallographic Software. *J. Appl. Crystallogr.* **2007**, *40* (4), 658−674.
(39) Adams, P. D.; Afonine, P. V.; Bunkóczi, G.; Chen, V. B.; Davis, I. W.; Echols, N.; Headd, J. J.; Hung, L.-W.; Kapral, G. J.; Grosse-Kunstleve, R. W.; McCoy, A. J.; Moriarty, N. W.; Oeffner, R.; Read, R. J.; Richardson, D. C.; Richardson, J. S.; Terwilliger, T. C.; Zwart, P. H. PHENIX: A Comprehensive Python-Based System for Macromolecular Structure Solution. *Acta Crystallogr. D Biol. Crystallogr.* **2010**, *66* (2), 213−221.
(40) Maguire, J. B.; Haddox, H. K.; Strickland, D.; Halabiya, S. F.; Coventry, B.; Griffin, J. R.; Pulavarti, S. V. S. R. K.; Cummins, M.; Thieker, D. F.; Klavins, E.; Szyperski, T.; DiMaio, F.; Baker, D.; Kuhlman, B. Perturbing the Energy Landscape for Improved Packing during Computational Protein Design. *Proteins* **2021**, *89* (4), 436−449.
(41) Dyer, K. N.; Hammel, M.; Rambo, R. P.; Tsutakawa, S. E.; Rodic, I.; Classen, S.; Tainer, J. A.; Hura, G. L. High-Throughput SAXS for the Characterization of Biomolecules in Solution: A Practical Approach. *Structural Genomics: General Applications* **2014**, *1091*, 245−258.
(42) Hura, G. L.; Menon, A. L.; Hammel, M.; Rambo, R. P.; Poole, F. L., 2nd; Tsutakawa, S. E.; Jenney, F. E., Jr; Classen, S.; Frankel, K. A.; Hopkins, R. C.; Yang, S.-J.; Scott, J. W.; Dillard, B. D.; Adams, M. W. W.; Tainer, J. A. Robust, High-Throughput Solution Structural Analyses by Small Angle X-Ray Scattering (SAXS). *Nat. Methods* **2009**, *6* (8), 606−612.
(43) Classen, S.; Hura, G. L.; Holton, J. M.; Rambo, R. P.; Rodic, I.; McGuire, P. J.; Dyer, K.; Hammel, M.; Meigs, G.; Frankel, K. A.; Tainer, J. A. Implementation and Performance of SIBYLS: A Dual Endstation Small-Angle X-Ray Scattering and Macromolecular Crystallography Beamline at the Advanced Light Source. *J. Appl. Crystallogr.* **2013**, *46* (1), 1−13.
(44) Putnam, C. D.; Hammel, M.; Hura, G. L.; Tainer, J. A. X-Ray Solution Scattering (SAXS) Combined with Crystallography and Computation: Defining Accurate Macromolecular Structures, Conformations and Assemblies in Solution. *Q. Rev. Biophys.* **2007**, *40* (3), 191−285.
(45) Brunette, T. J.; Parmeggiani, F.; Huang, P.-S.; Bhabha, G.; Ekiert, D. C.; Tsutakawa, S. E.; Hura, G. L.; Tainer, J. A.; Baker, D. Exploring the Repeat Protein Universe through Computational Protein Design. *Nature* **2015**, *528* (7583), 580−584.
(46) Courbet, A.; Hansen, J.; Hsia, Y.; Bethel, N.; Park, Y.-J.; Xu, C.; Moyer, A.; Boyken, S. E.; Ueda, G.; Nattermann, U.; Nagarajan, D.; Silva, D.-A.; Sheffler, W.; Quispe, J.; Nord, A.; King, N.; Bradley, P.; Veesler, D.; Kollman, J.; Baker, D. Computational Design of Mechanically Coupled Axle-Rotor Protein Assemblies. *Science* **2022**, *376* (6591), 383−390.
(47) Feldman, D.; Li, X.; Johnson, R.; Gerben, S.; Kim, D.; Sims, J.; Richardson, C.; Hicks, D.; Koepnick, B.; Goreshnik, I.; Allen, A.; Stewart, L.; MacCoss, M.; Baker, D. Screening Large Protein Libraries Using Mass Spectrometry Barcoding. *bioRxiv* (manuscript in preparation).
(48) Kim, D. E.; Jensen, D. R.; Feldman, D.; Tischer, D.; Saleem, A.; Chow, C. M.; Li, X.; Carter, L.; Milles, L.; Kang, A.; Bera, A. K.; Peterson, F. C.; Volkmand, B. F.; Ovchinnikov, S.; Baker, D. De Novo Design of Small Beta Barrel Proteins. *Proc. Natl. Acad. Sci. U. S. A.* (manuscript in review).
(49) Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; Bridgland, A.; Meyer, C.; Kohl, S. A. A.; Ballard, A. J.; Cowie, A.; Romera-Paredes, B.; Nikolov, S.; Jain, R.; Adler, J.; Back, T.; Petersen, S.; Reiman, D.; Clancy, E.; Zielinski, M.; Steinegger, M.; Pacholska, M.; Berghammer, T.; Bodenstein, S.; Silver, D.; Vinyals, O.; Senior, A. W.; Kavukcuoglu, K.; Kohli, P.; Hassabis, D. Highly Accurate Protein Structure Prediction with AlphaFold. *Nature* **2021**, *596* (7873), 583−589.
(50) Evans, R.; O'Neill, M.; Pritzel, A.; Antropova, N.; Senior, A.; Green, T.; Žídek, A.; Bates, R.; Blackwell, S.; Yim, J.; Ronneberger, O.; Bodenstein, S.; Zielinski, M.; Bridgland, A.; Potapenko, A.; Cowie, A.; Tunyasuvunakool, K.; Jain, R.; Clancy, E.; Kohli, P.; Jumper, J.; Hassabis, D. Protein Complex Prediction with AlphaFold-Multimer. *bioRxiv* **2022**, 2021.10.04.463034.

https://doi.org/10.1021/acs.biochem.2c00497
Biochemistry 2023, 62, 358−368