De Novo Scaffold Design: From Membrane β-Barrels to α-helical Repeat Oligomers

Stacey Gerben

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2021

Reading Committee:

David Baker, Chair

Frank DiMaio

Sharona Gordon

Program Authorized to Offer Degree:

Biochemistry

University of Washington

Abstract

De Novo Scaffold Design: From Membrane β-Barrels to α-helical Repeat Oligomers

Stacey Gerben

Chair of the Supervisory Committee:

Professor David Baker

Biochemistry

De novo design of new protein folds, shapes, and features opens up new avenues for functional design. My thesis work covers the *de novo* design of membrane beta barrels, important considerations for making new metal binding proteins, and the design of large sets of homo-oligomers with a wide variety of pocket shapes and sizes. This work aims to shed structural and functional insights necessary to create well behaved proteins with specific desired features without being limited to the natural proteins that have been characterized. The quick and accurate generation of new protein scaffolds with a variety of folds or pocket sizes will contribute to enabling further functional design of *de novo* membrane pores, protein catalysts, and small-molecule binders.

Table of Contents

List of Figures

Introduction

The field of *de novo* protein design has come a long way in the past 20 years. Scientists have come from only being able to make the basic helical bundle proteins [1,2] to creating protein logic gates [3–5],enzymes [6–8], and more [9–11]. All of these discoveries rely on a robust understanding of the features needed to make well structured and well folded proteins. These important features share many similarities, such as the importance of the hydrophobic effect, but have many specificities that vary wildly between different folds, such as the importance of the a and d positions in alpha helical packing[12], glycine positioning and conserved beta hairpin turns in soluble beta barrels [13],or the beta strand length and positioning of helical lids in ntf2-like folds [14,15].

During my thesis I have worked on three different projects, all of which involved the design of large scale structural features of proteins, also known as protein scaffold design. In the first year of graduate school I was involved in a project elucidating the features necessary to make *de novo* membrane beta barrels. Transmembrane beta barrels perform a variety of functions in nature, often as porins or as a part of membrane biogenesis [16,17].Transmembrane beta barrels are also highly in demand as engineering subjects. Their ability to fold into a membrane without a chaperone [18,19] makes this fold incredibly useful when developing nanopores for a variety of functions, including single-molecule DNA sequencing [20], small-molecule sensing [21,22], or water-filtering [23]. We set out to develop design strategies for *de novo* Transmembrane beta barrels, and were successfully and able to publish our paper in 2021 [24].

The second portion of my thesis involves my work designing new protein scaffolds around putative metal binding sites in order to create *de novo* metalloenzymes. Metalloenzymes catalyse a wide variety of reactions in nature by taking advantage of the versatility and reactivity of transition metals. Despite the diversity of reactions catalysed by natural proteins, there is a demand for both reengineering of native enzymes and design of *de novo* enzymes to perform efficiently under the conditions required for industrial or medical applications [25–28]. In many cases, all that is needed is routine re-engineering of the native enzymes to perform efficiently under the demanded application conditions. In other cases, the reaction or reaction condition desired differs so much from natural conditions that mere redesign of natural proteins is not practical. *De novo* enzymes, which are generated entirely from first principles rather than modified from natural proteins, are ideal for these situations. While there are previous examples of designed metalloenzymes in *de novo* scaffolds, these designed proteins are often limited to incorporation of the metal binding motif into three and four helix bundles [29,30]. The limitations of this shape greatly limits the functionalities possible for these enzymes. The intent of this project was to combine structural and functional design into a single design process, creating a *de novo* metalloenzyme by designing the entire scaffold around a desired active site. This would allow for a generation of topologies custom designed for the reaction of interest. This project was not ultimately successful, however, the project illustrates a number of important lessons in protein design, especially the importance of having strong, stable interfaces in oligomeric proteins before proceeding with further functionalization efforts. These lessons lead me to focus

specifically on the design of oligomers with a pocket useful for future functionalization, rather than trying to simultaneously design both structure and function.

The third portion of my thesis covers the design of new protein oligomers with asymmetric pockets. The goal of this project is to expand the space of proteins with pockets available for future functionalization. Currently, functional design of protein small molecule binders and catalysts is limited by the finite set of natural proteins with pockets of a size and shape amenable to design for small molecule binding or enzymatic activity. *De novo* protein design allows us to create new sets of proteins with specific desired features without being limited to natural proteins that have been previously characterised. This method allows for the generation of a large number of stable inert homo-oligomeric proteins with a variety of pocket shapes and sizes at the asymmetric interface. The quick generation of new protein scaffolds, as well as the potential range of pocket sizes will contribute to enabling further functional design of *de novo* small-molecule binders and protein catalysts.

In order to create new sets of oligomers with diverse pockets, we symmetrically dock curved helical repeat proteins such that they form pockets at the asymmetric interface of the oligomer. We then design both the interface for proper oligomerization, and then re-design the surface to optimize expression. The pocket sizes created in this way vary between the size amenable to bind a single ion, up to more than 20 angstroms across, and encompass a wide variety of shapes and sizes. Of the 138 proteins designed, 58 had soluble expression, 32 had correct oligomeric states in solution, 16 independent

designed proteins and 2 modifications to a design had small angle x-ray scattering (SAXS) data largely consistent with design models, and 4 have negative stain electron microscopy (nsEM) 2d class averages showing the structures coming together as designed. When comparing alphafold predictions to experimental designed proteins, 13 of the 16 SAXS validated designed proteins had a root mean squared deviation (rmsd) between the alphafold monomer prediction and the design model of less than 2 angstroms. We were able to get 2D class averages of a single design, sg135_C4 which clearly show it's similarity to the design state. Finally we were able to get a 4.2 Å crystal structure of sg122_C3 which has an interface rmsd to design of 0.84 Å, and a full oligomer rmsd of 3.4 Å

Chapter 1. Design of *De Novo* Membrane Beta Barrels

1.1 Introduction and Clarification of Work

My contributions to the published paper (Vorobieva et. al. 2021) included 1) the initial round of designs, 2) expression of designed proteins and purification from inclusion bodies, 3) re-expression of designed proteins when hybridized with OmpA loops and scrambled loops, and 4) a supplemental figure on the alignment of natural transmembrane and soluble beta barrels. The initial design process was developed using relevant information from established design strategies for both soluble beta barrels [10,13] and other non-beta barrel membrane proteins, such as α-helical membrane proteins [31,32], and some conserved motifs of natural transmembrane beta barrels (TMBs).

In order to design membrane beta barrels, we decided to use the structure of the transmembrane domain of the model outer membrane protein A (tOmpA) of *E. coli* as a basis for design. This domain is a useful model because its folding is well characterised [33,34]. The design goal was twofold. First, to create well folded transmembrane proteins with a similar backbone structure to the naturally occurring protein, which proves that we were able to accurately control the fold. Second, to have a low sequence similarity to the natural protein, Which proves that we were creating genuinely novel designed proteins rather than merely modifying a naturally occurring protein. Blueprint files were generated for Rosetta BlueprintBuilder based on the structural connections of tOmpA, and structural considerations already learned from the design soluble beta barrels [13], including proper placement of glycines, which allows beta strands to properly curve around the barrel, as well as the ramachandran angles of conserved beta turns on one

side of the beta barrel. I designed, expressed, and purified the initial round of transmembrane beta barrels (TMB0.1-TMB0.9), as well as the re-expression of TMB0.1-TMB0.9 that were each hybridized with both 1) tOmpA loops and 2) long unstructured loops, which were both done in order to try to troubleshoot lack of expression in the initial designed proteins. I also wrote a script to calculate the frequency distribution of each amino acid in specific positions along conserved periplasmic loops of natural transmembrane beta barrels. This script was modified to separate loops based on type of beta turn[35] and used to create supplemental figure S2 on the alignment of natural transmembrane and soluble beta barrels.

1.2 *De Novo* Design of Transmembrane β barrels (Vorobieva et. al. 2021)

The following is a copy of Vorobieva et al, published in Science in 2021. References are numbered separately, and are listed starting on page 140.

Abstract

Transmembrane β-barrel proteins (TMBs) are of great interest for single-molecule analytical technologies because they can spontaneously fold and insert into membranes and form stable pores, but the range of pore properties that can be achieved by repurposing natural TMBs is limited. We leverage the power of *de novo* computational design coupled with a "hypothesis, design, and test" approach to determine TMB design principles, notably, the importance of negative design to slow β-sheet assembly. We design new eight-stranded TMBs, with no homology to known TMBs, that insert and fold reversibly into synthetic lipid membranes and have nuclear magnetic resonance and x-ray crystal structures very similar to the computational

models. These advances should enable the custom design of pores for a wide range of applications.

Advances in *de novo* protein design have yielded water-soluble proteins of increasing complexity (*1*–*5*), and several examples of α-helical membrane proteins (*6*, *7*). However, the *de novo* design of an integral transmembrane β barrel (TMB) has not yet been achieved. The unassisted folding of TMBs into lipid bilayers in vitro likely involves concerted membrane insertion and folding of β hairpins (*8*, *9*), and how this is encoded in the sequences of TMBs is not well understood because of experimental challenges in characterizing their rugged folding pathways (*10*, *11*). To prevent misfolding and aggregation in vivo, an array of chaperones assist TMB folding and assembly in the outer membranes of prokaryotes, mitochondria, and chloroplasts (*12*). The lipid-folding–water-aggregation trade-off places poorly understood constraints on the global sequence properties of TMBs, slowing down the development of *de novo* design methods. Instead, TMB engineering has proceeded by modification of naturally occurring TMBs, which has yielded nanopores for single-molecule DNA sequencing (*13*), small-molecule sensing (*14*, *15*), or water-filtering bioinspired membranes (*16*).

To shed light on the sequence determinants of folding and stability of TMBs, and to enable the custom design of TMBs for specific applications, we set out to design

TMBs *de novo*. We started by studying the constraints membrane embedding puts on both the backbone geometry and the sequence of transmembrane β barrels.

Geometric constraints on transmembrane β-barrel backbones

TMBs are formed from a single β sheet that twists and bends to close on itself, so that all membrane-embedded backbone polar groups are hydrogen-bonded and shielded from the lipid environment. Insertion of TMBs into the lipid membrane is oriented (*17*), with β strands usually connected with long loops on the translocating (trans) side of the β barrel (extracellular in bacteria) and short β-turns on the nontranslocating (cis) side (Fig. 1.1A). The β-barrel architecture is characterized by two discrete parameters: the number of strands (*n*) and the shear number (*S*)—the sum of residue offsets (register shifts) between the neighbor strands, starting at any strand and tracing around the β barrel (fig. S1A) (*18*). The ideal β-barrel radius *r* (eq. S1) and angle of the strands with the main barrel axis θ (eq. S2) are functions of *S* and *n* (table S1) (*19*). *S* and *n* also define the packing arrangement of side chains in the β barrel. There are *S* continuous strips of side chain Cβ atoms perpendicular to the β strands (fig. S1, B and C); half of these Cβ strips point toward the lumen and the other half toward the β-barrel exterior. We focused on the simplest and smallest β-barrel architecture of eight β strands. We first considered a shear number of eight ($n = S$). In this configuration, the total register shift is distributed equally among the four β hairpins (two-residue offset between each β hairpin), and the four Cβ strips pointing toward the

lumen of the barrel are arranged in fourfold-symmetric rungs with the Cα-Cβ vectors (which indicate the direction of the side chains) pointing at each other (Fig. 1.1B, left).

This symmetric arrangement combined with a small β-barrel radius does not allow tight jigsaw-puzzle–like packing—the side chains clash with each other rather than interdigitating. To enable better packing, we broke the symmetry in the core by increasing the register shift between two β hairpins from two to four residues, resulting in a shear number of 10. In this case, there are five intertwined Cβ strips that spiral around the barrel axis, and the Cα-Cβ vectors point between rather than at each other so that the side chains can pack in a more interdigitated fourfold screw-like pattern (Fig. 1.1B, right).

The uneven distribution of register shifts between β hairpins complicates interactions with the lipid membrane, which can be approximated as two planes that must be parallel to ensure constant membrane thickness. In natural TMBs, the cis (periplasmic) β-turns are close to the periplasmic lipid–water boundary (fig. S2, A to D). The β-turn residues closely match the sequence preferences observed in water-soluble β barrels (mostly polar residues), but the lipid-exposed residues flanking these β-turns are predominantly hydrophobic (fig. S2, H to K) (*20*) and define the cis boundary of the transmembrane region (“membrane anchor residues”; fig. S2, A to D). The geometric challenge is that differences in the register

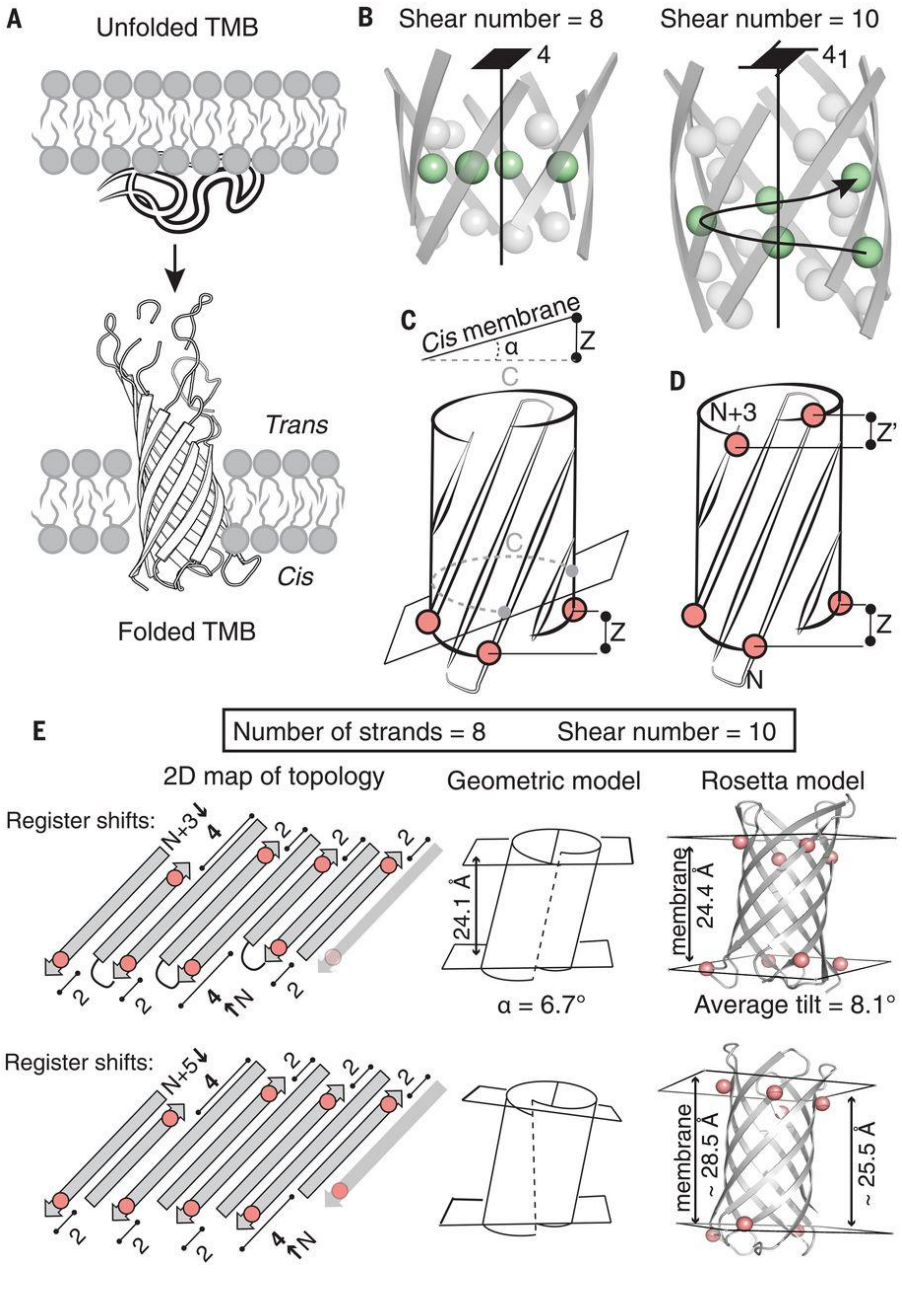

Fig. 1.1 Geometric principles for TMB backbone design.

(A) The trans side of the β barrel is the side that translocates through the lipid membrane during TMB folding and insertion. The cis β-turns remain on the initial protein–lipid interaction side of the membrane. (B) Comparison of side-chain packing arrangements in eight-strand β barrels with shear numbers of 8 (left, fourfold-symmetric packing) and 10 (right, fourfold screw resulting in a jigsaw-like packing). In (C) to (E), the membrane-anchoring residues are shown as pink spheres. (C and D) Geometric model of membrane-association constraints on the β-barrel architecture. (C) Asymmetric register shifts between the β hairpin can be accommodated by tilting the β barrel to the transmembrane axis by an angle α = arctan(Z/C). (D) The change of level Z between two anchor residues on the cis side of the β barrel must be matched by the change of level Z' between the two stacking anchor residues on the trans side. Because of β strands staggering to the main β-barrel axis, an anchor residue on the cis side of a strand N stacks with an anchor residue on the trans side of strand N+3. (E) The geometric model (center) and Rosetta model (right) predict similar tilt angles (α) of the β barrel to the membrane axis and constant hydrophobic thickness, for β-strand arrangements with matching Z and Z' (double register shifts located on strand N in cis and on strand N+3 in trans) (top). Both models show inconsistent hydrophobic thickness for β-barrel architectures with double register shifts located on strand N in cis and on strand N+5 in trans (bottom)

.

shifts between β hairpins result in a screw-like arrangement of the four anchor residues with a translation *Z* along the main β-barrel axis (eqs. S3 to S5); hence, if the β-barrel axis is along the membrane normal, the anchor residues cannot all be in the same plane. The vertical offset of the anchor residues can be made more compatible with the planarity requirement by tilting the β barrel in the membrane by an angle $\alpha = \arctan(Z/C)$, where the denominator is the length of the arc between anchor residues 1 to 4 projected onto the plane perpendicular to the main axis (eq. S6) (Fig. 1.1C and fig. S3B). In the case of a β barrel with symmetry ($n = 8$, $S = 8$), the vertical offset between each anchor residue, approximated from the geometric model, is close to zero (supplementary text), and no tilt is required. When *S* is increased to 10 by increasing the register shift between one pair of hairpins to four residues, the barrel must be tilted by ~6.7° to the transmembrane axis (Fig. 1.1E, top) to bring the anchor residues into the same plane. To test the validity of this simple geometric model, we explicitly assembled TMB backbones using Rosetta (*21*) and predicted their placement in the membrane [materials and methods (*22*)], which yielded an average tilt angle (8.1°; Fig. 1.1E, top) close to that of the geometric model. Placing the four-residue register shift after each of the four cis hairpins resulted in tilts with similar amplitude but different directions relative to the membrane axis (supplementary text and fig. S4, B to G); we focused on the placement in which the four-residue register shift is in the middle of the β sheet.

We next investigated the structural consequences of the fact that the planes representing the cis and trans membrane boundaries must be roughly parallel to each other to keep the hydrophobic thickness constant. To achieve this, the offset Z between any two neighbor anchor residues on the cis face must be matched by a similar offset Z' between the anchor residues above it on the trans face. For barrel topologies of (n = 8, S = 10) spanning a membrane of 24 Å, an anchor residue on the cis side of strand N stacks along the main β-barrel axis with the anchor residue on the trans side of strand N+3 because of the staggered β strands (Fig. 1.1D and supplementary text). Hence, to maintain constant thickness, the register shift between strands N and N+1 on the cis side must be equal to the register shift between strands N+3 and N+4 on the trans side. To confirm this prediction of our geometric model, we set the cis side register shift between strands N and N+1 to four residues and ran Rosetta design simulations and transmembrane plane predictions on backbones with a matching four-residue register shift on the trans side between (i) strands N+3 and N+4 and (ii) strands N+5 and N+6. We averaged planes representing the membrane boundary in cis and trans and found, consistent with the model, parallel planes and constant hydrophobic thickness for the N+3 case (i) but a 3-Å change of thickness in the N+5 case [(ii), Fig. 1.1E, bottom].

We used this constant hydrophobic thickness constraint to guide the distribution of the register shifts around the β barrel. The cis hairpins were closed with short β-turns

associated with a β bulge, which match the local twist of the β strands [these are abundant in water-soluble (*5*) and transmembrane β barrels (fig. S2A)].

On the trans side, the strands were connected with canonical β-turn sequences with strong β-hairpin nucleating properties [3:5 type I β-turns + G1 bulge with canonical Ser-Asp-Gly (SDG) sequence (*21–23*)] in place of the long loops found in native TMBs; such turns were previously used to design water-soluble β barrels (fig. S2, E and G) (*5*). To relieve strain from high β-sheet curvature, we placed glycine kinks (*5*)—glycine residues in a fully extended conformation within β strands—into the blueprint such that (i) every Cβ strip pointing to the core of the barrel contains a glycine and (ii) no Cβ strip contains more than four nonglycine residues in a row (¼ of the average barrel circumference). The glycine kinks in the designed β-barrel blueprint stack along four vertical lines (Fig. 1.2, A and B) such that the resulting Rosetta models have four regions of strong β-sheet bending that delimit a lumen with a distinctive square-shaped cross-section not observed in naturally occurring β barrels (fig. S2F).

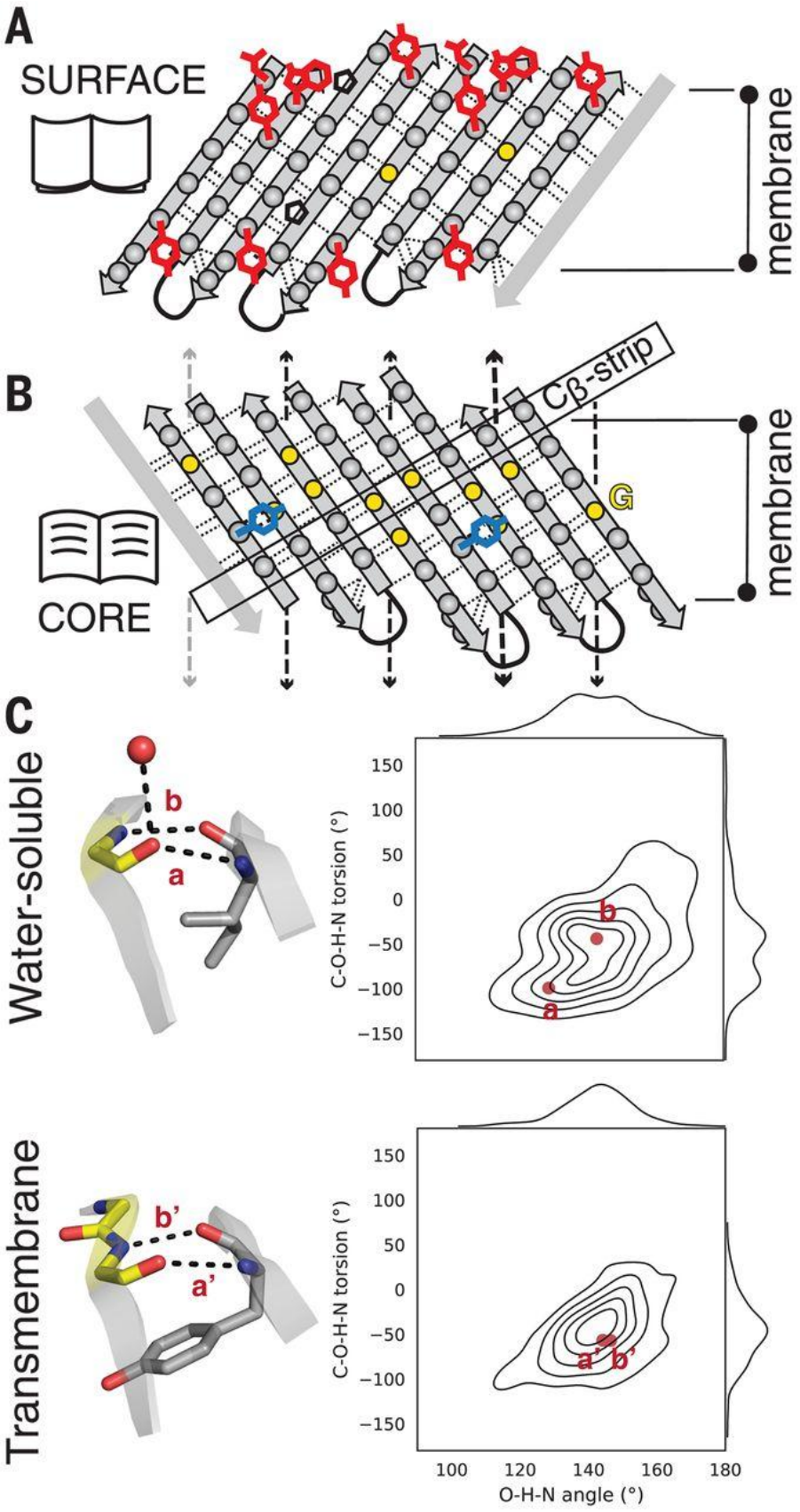

Fig. 1.2 Sequence features defining *de novo* TMB fold and shape. (A and B) Two-dimensional schematic representation of the connectivity (hydrogen bonds as dashed lines) between β strands in the TMB designs. Side chains are shown as gray spheres and glycine residues as yellow dots. Aromatic girdle motifs are shown in red, tyrosines of the mortise–tenon motifs in blue, and prolines as black pentagons. Glycine kinks were arranged to bend the β sheet into four corners (vertical arrows). (C) Hydrogen bond geometries between pairs of residues involving a glycine kink. Left: Examples from crystal structures of water-soluble (PDB ID: 6CZH) and transmembrane β barrels (PDB ID: 1BXW). Glycine residues are in yellow, and water molecules are shown as red dots. Right: Distributions of the C-O-H-N and O-H-N angle values describing the hydrogen bond geometry in β-barrel crystal structures.

Sequence design and initial experimental characterization

To delimit the upper and lower membrane boundaries, four tyrosine residues were placed two positions upstream of the anchor residues on the cis side, and alternating tyrosine and tyrosine-tryptophan motifs were placed at the trans boundary [Fig. 1.2A and supplementary text; such "aromatic girdles" are observed in native TMBs (*24*)]. To design the remainder of the sequence, we first experimented with the approach that we took for helical transmembrane proteins (*6*), using standard Rosetta design methods to design core residues (which results in largely hydrophobic interiors, as in helical transmembrane proteins) and resurfacing the outside with hydrophobic residues. However, this resulted in sequences that had strong amyloid propensity (fig. S5). To reduce the amyloid propensity and the hydrophobicity [native TMBs are usually less hydrophobic than α-helical membrane proteins (*25*)], we experimented with requiring all residues in the interior of the barrel (excluding the glycine kink positions) to be polar and the surface residues to be hydrophobic, resulting in the hydrophobic-polar sequence pattern characteristic of the β-sheet secondary structure but inside-out compared to water-soluble β barrels. To help define the register between β strands we placed tyrosine residues adopting (+60,90) rotamer angles to closely interact with the groove formed by a neighbor glycine kink (*26*) [the "mortise–tenon" motif (*27*, *28*)]. We placed two such motifs in the regions where defining the register shift seemed likely to be particularly important: Tyr69 (Y69) on strand 5, where the four-residue register shifts in cis and trans produce a larger vertical

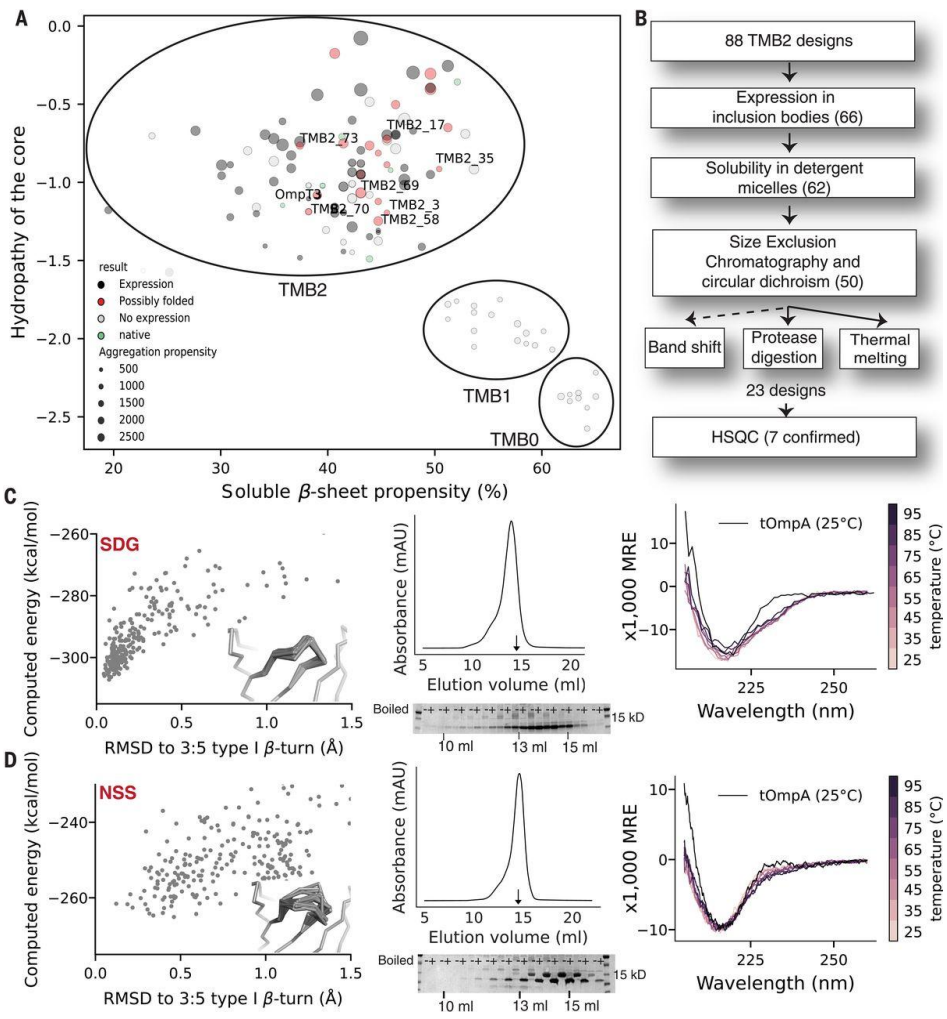

Fig. 1.3 Negative design is critical for *de novo* TMB folding.

(A) Successful design of TMBs requires reducing β-sheet propensity of the transmembrane β strands. *x* axis, β-sheet propensity of the transmembrane region [calculated with RaptroX (*59*)]; *y* axis, hydrophobicity of the core [GRAVY hydropathy index (*60*)]. Gray spheres, nonexpressing TMB designs (repeated twice); black circles, expressing designs that do not fold; red, TMB designs that pass biochemical folding screening (repeated in two different detergents)—labels indicate that the folded species was validated by HSQC; green, naturally occurring TMBs with eight strands. Circle size indicates aggregation propensity of the sequence predicted with TANGO (*61*). (B) Experimental workflow. The number of TMB2 designs satisfying each criterion is shown in parentheses. (C and D) Proper folding of tOmpA requires negative design against strong β-turn nucleating sequences on the trans side. Left: Rosetta energy landscapes of designs with canonical low energy (C) or suboptimal (D) sequences substituted in a 3:5 type I β-turn with a G1 β-bulge. Conformational perturbations were generated with kinematic loop closure (*62*); the inset shows the backbone conformations of the 25 lowest-energy models. Center: After refolding in 2× CMC DDM detergent, Omp*Trans*3 (bottom panel) elutes on SEC similarly to tOmpA (arrow, 14.62 ml for OmpTrans3 and 14.53 ml for tOmpA) and runs as a heat-modifiable species on SDS-PAGE characteristic of folded tOmpA, whereas (top panel) the OmpAAG peak elutes earlier (13.96 ml) and does not show a band shift (band shift assay repeated three times; SEC repeated two times). Right: The far-UV CD spectrum of Omp*Trans*3 (bottom panel), but not OmpAAG (top panel), is similar to that of tOmpA (repeated two times).

offset in the β sheet, and Tyr11 (Y11) on strand 1, where the β sheet closes on itself but lacks a register-defining β-turn between the first and last strands (Fig. 1.2B and fig. S6). Finally, we designed full β-barrel sequences using Rosetta combinatorial sequence design and the ref2015 energy function (*29*) with increased weight on the electrostatics term to favor side-chain–side-chain hydrogen bonds in the core of the β barrels. As expected, the secondary structure of the resulting designs was accurately recapitulated by secondary-structure prediction programs (Fig. 1.3A).

Folding of TMBs is chaperone-mediated and catalyzed in vivo by the β-barrel assembly machinery (BAM) complex in Gram-negative bacteria, the sorting and assembly machinery (SAM) complex in mitochondria, and the OEP80 insertase in the outer chloroplast membrane (*30*). Because it was unclear whether our TMB designs would be able to interact with the chaperone machinery to fold in the outer membrane of *Escherichia coli*, we expressed the designed sequences in the cytoplasm, anticipating that they would form inclusion bodies that could then be solubilized in urea–guanidinium chloride [both natural and engineered TMBs have been produced in this way (*31*)]. We obtained *E. coli* codon-optimized synthetic genes for nine designs (set TMB0, fig. S7), but no protein of the correct molecular weight was produced upon the induction of protein expression (data S2). Reasoning that the designed sequences may have had too much positive charge, which can impair translation (*32*), in a second round of 16 designs, we reduced the number of charged residues in the

core of the protein (set TMB1, fig. S7). Again, none was expressed in *E. coli* (data S2).

Because of the failures at the expression stage, experimental feedback to improve the design methodology could not be obtained. To gain insight, we instead compared our designs to sequences of natural eight-strand TMBs. We noted two differences: First, the natural TMBs often have long and disordered trans loops rather than short β-turns (*20*), and second, the secondary-structure propensity of their transmembrane β strands was lower than that of the designs we had tried to express (Fig. 1.3A). We hypothesized that the high β-turn and/or β-strand propensities of our designed sequences could result in rapid formation of off-target β-sheet structures when expressed in the cytoplasm, which could be cleared rapidly or hinder growth of expressing cells.

Role of trans β-turns

We first explored the role of the trans loops in TMB folding and expression by redesigning the native TMB of the protein OmpA (tOmpA), replacing its trans loops with the canonical SDG β-turn sequence used in our designs (fig. S8, A and B). The relooped tOmpA construct (OmpSDG) was expressed at high levels in *E. coli* (where it was found in inclusion bodies), but it could not be correctly refolded (Fig. 1.3C and fig. S10, C and D). To understand this observation, we carried out Rosetta energy

landscape calculations on short β-turns at the trans membrane boundary of natural TMB structures and observed that their sequences have relatively low propensity to form β-turn structures (supplementary text) compared to the β-turns of soluble β barrels and the SDG β-turns of OmpSDG. We constructed and tested four variants of tOmpA (Omp*Trans*1-4) that each contain two such 3:5 type I β-turns with suboptimal sequences [these designs are shorter than the shortest variant of tOmpA previously reported, which has trans connections of 5 to 18 residues (*33*)]. The proteins were again expressed at high levels in inclusion bodies (table S2), but this time all four of these sequences showed a heat-modifiable band [analyzed by cold SDS–polyacrylamide gel electrophoresis (PAGE)] when folded into n-dodecyl-β-D-maltoside DDM detergent micelles and large unilammelar vesicles (LUVs), characteristic of properly folded tOmpA (Fig. 1.3D and fig. S10, C and D). The best expressed design, Omp*Trans*3, was characterized in more detail; following refolding in detergent micelles, it had a retention time similar to that of native tOmpA on a size exclusion chromatography (SEC) column (Fig. 1.3D and fig. S12), a similar native mass spectrometry (nMS) profile (fig. S13), well-dispersed resonance peaks by ^{1}H-^{15}N–heteronuclear single-quantum coherence nuclear magnetic resonance (HSQC NMR) in Fos-choline-12 (DPC) detergent (fig. S10B) and a circular dichroism (CD) spectrum similar to that of tOmpA in both DDM micelles (Fig. 1.3D) and in 1,2-diundecanoyl-*sn*-glycero-3-phosphocholine (DUPC, *di*$C_{11:0}$PC) LUVs with distinctive peaks at ~220 and 231 nm (fig. S10A) (*34*). These data suggest that either

long loops or short suboptimal β-hairpin loops on the trans side are necessary to slow down nucleation of the trans β-hairpins and allow proper folding of TMBs in vitro. However, simply replacing the trans loops on four of the TMB0 designs with the extracellular loops of tOmpA or scrambled versions of these loops did not appreciably increase protein expression (fig. S9), suggesting that there was a further problem with the properties of the transmembrane β strands in the original designs.

Reducing β-sheet propensity

We sought to further use negative design (*35*, *36*) to disfavor off-target states and slow down folding, this time through reduction of the high secondary-structure propensity of the β strands in our designs. We increased the hydrophobicity of the side chains in the β-barrel lumen, thereby disrupting the strict alternation of polar and hydrophobic residues along the β strands. To do so, we experimented with the design of networks of hydrogen bonds surrounded with scattered hydrophobic patches. We extended the mortise–tenon motifs to include a hydrogen bond between the tyrosine and a negatively charged Asp or Glu residue to seed the design of hydrogen bond networks. Possible positions for the Asp or Glu were exhaustively searched using the Rosetta HBNet protocol (*37*) to design β-barrel backbones with preinstalled Tyr-Gly-Asp/Glu (YGD/E) motifs at one or both of the locations identified above. We used Rosetta

combinatorial sequence optimization to design the remainder of the positions facing the core of the barrel, allowing all 18 amino acids other than Cys and Pro.

To further lower the β-sheet propensity, we experimented with incorporation of glycine residues [which destabilize β strands (*38*)] on the hydrophobic outer surface of the β barrels. To guide placement of the glycines, we compared crystal structures of natural TMBs with those of water-soluble β barrels, which rarely have water-exposed surface glycine (supplementary text). We found that the extended backbone conformation of core glycine kinks in the water-soluble β barrels results in noncanonical out-of-plane backbone hydrogen bond geometry characteristic of a left-hand twist (O–H–N angle ~130°; C–O–H–N dihedral ~−100°; Fig. 1.2C, top, and fig. S14A, center), whereas the surface residues preceding the glycine kink have a more pronounced right-hand twist (C–O–H–N dihedral >0°, fig. S14A, right) than canonical in-plane backbone hydrogen bonds (O–H–N angle ~155°; C–O–H–N dihedral ~0°; fig. S14A, left) (*39*). The backbone carbonyls of glycine kinks involved in extremely out-of-plane hydrogen bonds (resulting in strongly bent β strands) in water-soluble β barrels are often exposed to solvent to interact with a water molecule or other hydrogen bond donor in crystal structures (fig. S15, E and F). Such exposed carbonyls are likely disfavored on the lipid-buried surface of TMBs because there are no water molecules in the bilayer to stabilize them, and TMBs indeed have a smaller population of glycine kinks and preglycine hydrogen bonds deviating from in-plane

geometry (Fig. 1.2C, bottom, and fig. S14B, center). We hypothesized that glycines in positions preceding glycine kinks could allow more canonical in-plane hydrogen bonds and hence reduce unfavorable surface exposure of the carbonyls to the apolar lipid environment, and we confirmed this with explicit Rosetta design calculations (supplementary text). In the subsequent sequence design calculations, we identified strongly bent glycine kinks in the β-barrel blueprint and placed glycines in surface-exposed positions directly preceding them.

Negative design of loops and strands enables *de novo* TMB design

We carried out three iterations of core and surface design according to the above principles using the suboptimal loops from Omp*Trans3* on the trans side, allowing the backbone to relax on the basis of the current sequence by gradient-based energy minimization at each step (glycine placement, in particular, allows local backbone rearrangement). The design calculations converged on 20 distinct core hydrogen bond network architectures with overall amino acid composition similar to that of natural eight-strand TMBs (fig. S7D). Codon-optimized synthetic genes were obtained for several representatives of each network architecture for a total of 88 designs (set TMB2). In sharp contrast with the lack of expression in our previous unsuccessful design rounds, 66 of these designs were well expressed in inclusion bodies, as intended. To test the influence of the trans loops on expression, we expressed variants

of 20 of these designs incorporating the extracellular loops of tOmpA. The same designs expressed or did not express with the short suboptimal β-turns or the long tOmpA loops (data S2), indicating that the transmembrane β strands—rather than the β-strand connections—carry the main sequence determinants of cytoplasmic expression (the turn sequence does matter for subsequent assembly and membrane insertion, as exemplified by the failure of OmpSDG to fold properly).

Characterization of folding, stability, and structure

To test the ability of the designs to stably fold to TMB structures in vitro, we followed procedures used to fold tOmpA and other natural TMBs (Fig. 1.3B) (*40*, *41*). Briefly, the inclusion bodies were dissolved in 8 M urea and rapidly diluted into DDM, DPC, or *n*-octyl-β-D-glucopyranoside (OG) detergents at twice the critical micelle concentration (2× CMC) (data S4). Out of the 66 expressing designs, 62 formed soluble species in such conditions. We purified the protein–detergent complexes by SEC and characterized the 50 designs that had a SEC retention volume expected for a monomeric TMB (similar to the eight-stranded tOmpA monomer and Omp*Trans*3) and a far-ultraviolet (UV) CD spectrum characteristic of a β-sheet protein. The band-shift assay used to monitor native TMB folding (*42*) was uninformative for the identification of folded *de novo*–designed TMBs. Instead, we found good agreement between the resistance of a design to protease digestion and thermostability up to

95°C of the characteristic β-sheet far-UV CD spectrum. In total, 23 designs satisfied the biochemical screening criteria, suggesting that they fold into a TMB structure. Eleven such designs were randomly selected for analysis by ^{1}H-^{15}N HSQC NMR in DPC detergent micelles, and seven had well-dispersed chemical-shift profiles, characteristic of a folded protein in this detergent (figs. S16 and S17, validated designs; fig. S18, designs that failed biochemical tests; fig. S19, designs that passed the biochemical tests but appear misfolded by NMR).

We selected two *de novo* designs, TMB2.17 (highest BLAST E-value to the nonredundant protein database: 0.10) and TMB2.3 (BLAST E-value: 0.035) and the Omp*Trans*3 construct for detailed biophysical characterization in a lipid bilayer to determine whether the proteins exhibit the expected properties for a membrane-spanning β barrel (using tOmpA as a control). After refolding into 100 nm DUPC LUVs, all four proteins had far-UV CD spectra characteristic of a β-sheet protein both in 0.24 and 2 M urea, and distinct from the spectra of both the fully unfolded proteins in 8 M urea and the proteins refolded in the absence of lipid (figs. S10A and S20). We next determined the stability of the folded proteins by monitoring their ability to fold into or unfold out of LUVs at increasing urea concentrations, monitored by the change of fluorescence intensity between water-exposed and lipid-embedded surface tryptophans (*43*). The designed TMB proteins are more thermodynamically stable [midpoint urea concentration for folding (C_{mF}) 5.7 and 7.2

M for TMB2.3 and TMB2.17, respectively; Fig. 1.4A) than tOmpA (Cm_F = 4.7 M), and Omp*Trans*3 is the most stable protein as it appears folded even in 9 M urea (fig. S21), in agreement with the far-UV CD data. It has been previously shown that the folding and unfolding transitions of many natural TMBs exhibit hysteresis due to the high kinetic barrier to unfolding and extraction from the membrane environment (*11*, *44*, *45*). Under the conditions tested here, this behavior was observed for tOmpA but not for the designs TMB2.3 and TMB2.17, which showed superimposable and reversible unfolding and folding transitions, suggesting reduced kinetic stability relative to tOmpA. These observations likely explain the lack of a band shift in SDS-PAGE: The *de novo* designs unfold during electrophoresis because of lower kinetic barriers to unfolding (*46*) (fig. S22). The equilibrium unfolding curves for TMB2.3 and TMB2.17 fitted well to a two-state transition (fig. S23) with unfolding free energies (ΔG^0_{UF}) of 38 and 56 kJ mol^{-1}. These ΔG^0_{UF} values fall within the range of natural TMBs [ΔG^0_{UF} 10 to 140 kJ mol^{-1} (*43*, *47*–*49*)].

The designed TMBs fold more than an order of magnitude more rapidly than tOmpA [(*50*); folding rate constant of 3×10^{-3} s^{-1} for tOmpA]—too rapid to allow accurate measurement of the folding rate constant (Fig. 1.4B). Tryptophan fluorescence emission spectra of the end point of the folding reactions confirmed that the TMBs were indeed fully folded (fig. S24). To confirm that the designs integrate into the lipid bilayer rather than folding on the lipid surface or in the absence of lipid, proteins dissolved in 8 M urea were diluted into 2 M urea without lipid or into LUVs

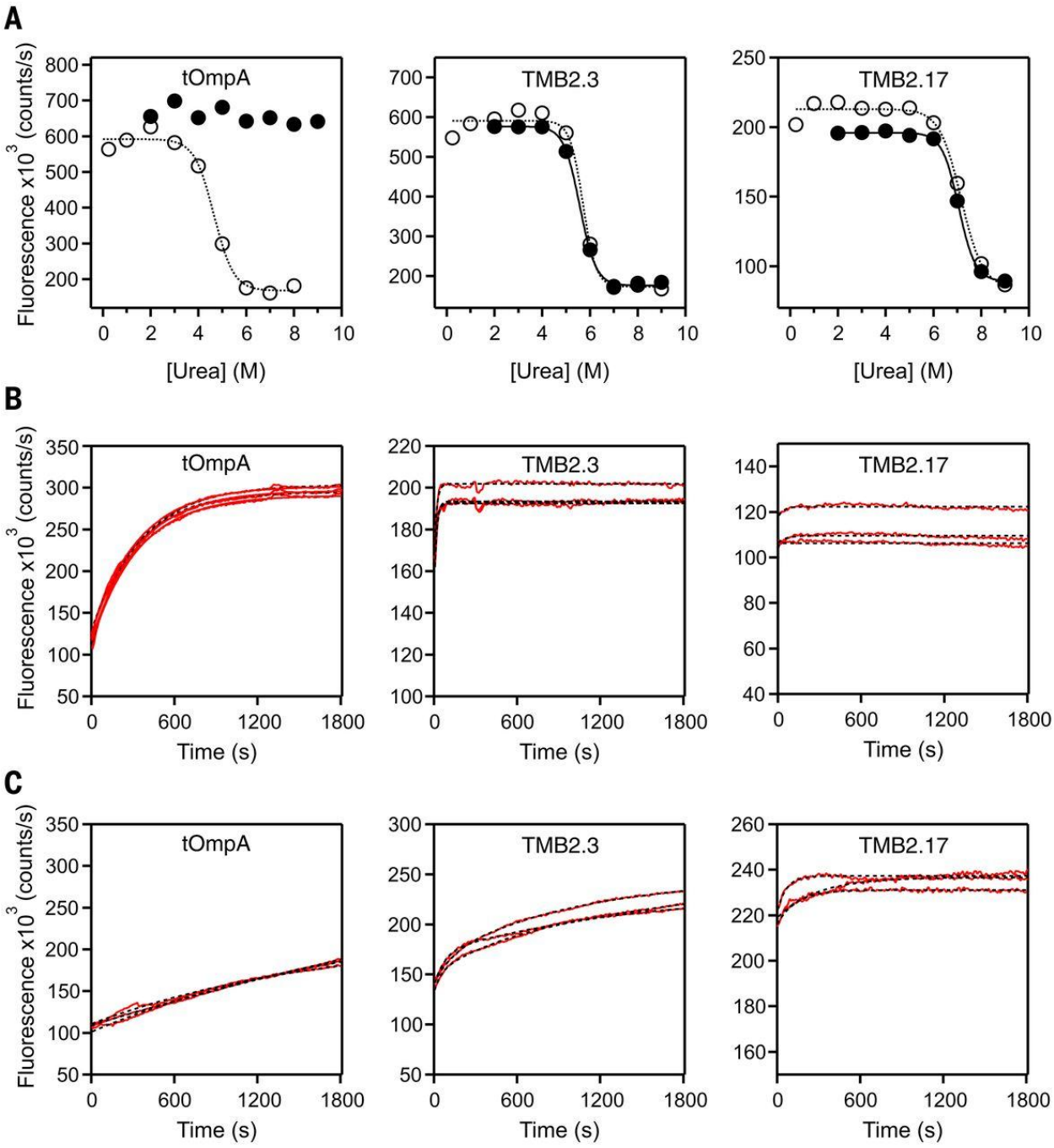

Fig. 1.4 Folding of *de novo*–designed TMB2.3 and TMB2.17 compared to tOmpA in synthetic lipid membranes. (A) Urea dependence of folding and unfolding in DUPC LUVs. The fluorescence intensity at 335 nm was plotted against urea concentration to determine the midpoint urea concentration for folding (CmF) (open circles, dashed line) and unfolding (CmUF) (filled circles, solid line). (B and C) Kinetics of folding into (B) DUPC and (C) DMPC LUVs at a lipid to protein ratio of 3200:1 (mol/mol) in 50 mM glycine-NaOH (pH 9.5), 2 M urea at 25°C monitored by tryptophan fluorescence at 335 nm over 30 min (red line). Data were fitted with a single exponential function to determine folding rate constants (black dashed line). Three technical replicates each.

composed of 1,2-dimyristoyl-*sn*-glycero-3-phosphocholine (DMPC, *di*C$_{14:0}$PC). Consistent with previous results showing that the folding rates of natural TMBs are inversely correlated with lipid chain length (*9*, *51*), the designed TMBs fold more slowly into lipids of longer acyl chain length (Fig. 1.1.4C) and do not fold in the absence of lipid (fig. S25B), confirming that they indeed integrate into the lipid bilayer.

The NMR structure ensemble generated on the basis of the chemical shifts and NOE information agrees closely with the design model [average backbone root mean square deviation (RMSD) of 2.2 Å, Fig. 1.5B]. We observed low-intensity additional resonance peaks for a subset of residues, indicating the presence of a (minor) secondary conformation.

The secondary signals strong enough for analysis were consistent with the secondary-structure assignment and NOEs of the main conformation, indicating that the secondary conformation does not involve modification of the β-barrel architecture. The residues with double peaks cluster in the cis region of strands 1, 2, and 8 (fig. S26B); these could result from close proximity to the flexible N terminus or transient dimeric interactions identified by native mass spectrometry in detergent micelles (figs. S28 and S29).

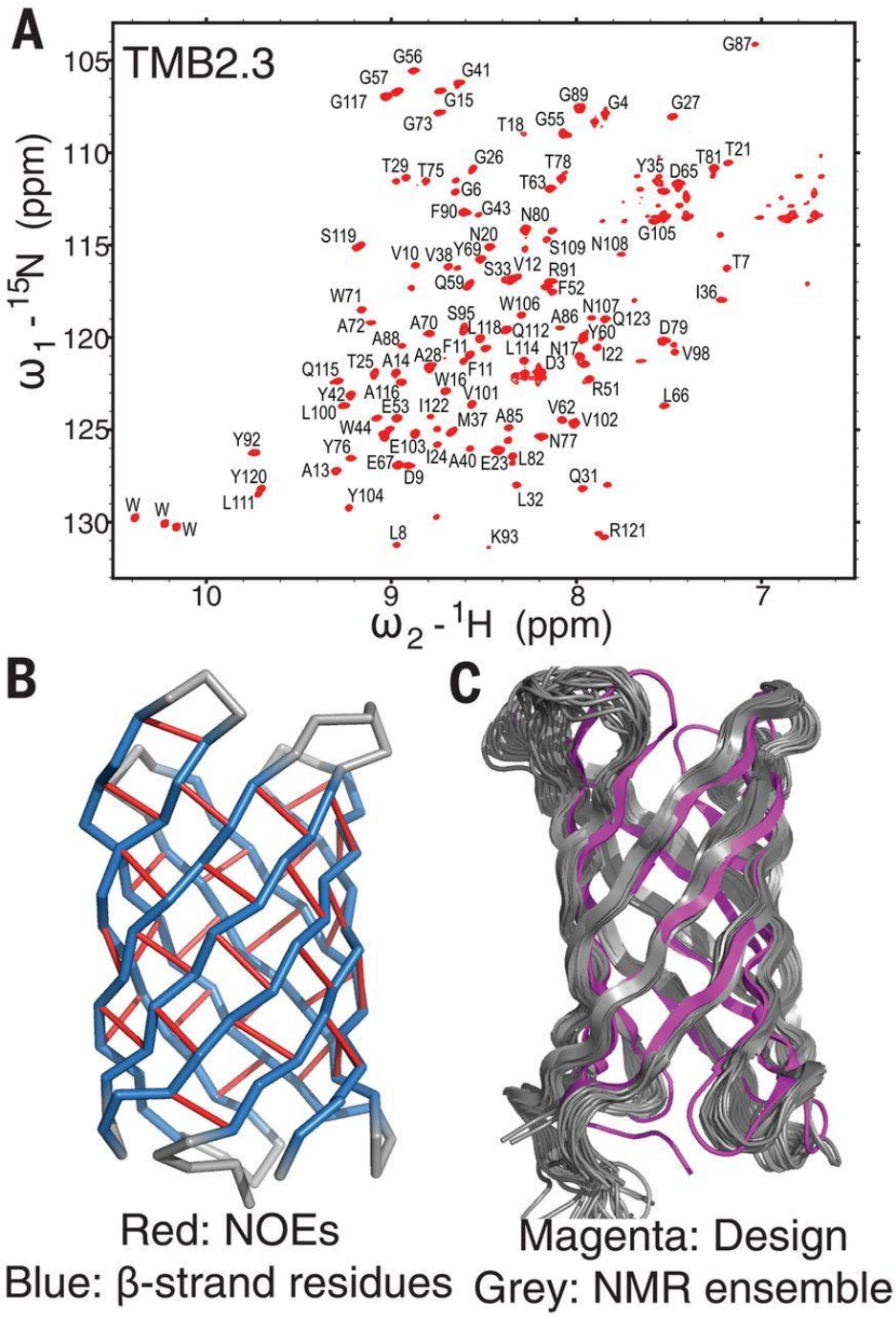

Fig. 1.5 NMR structure of TMB2.3 in DPC detergent micelles.(A) Assigned 15N-1H TROSY (transverse relaxation–optimized spectroscopy) spectrum of TMB2.3. (B) NMR constraints mapped on the TMB2.3 design model. Residues predicted to have β-sheet secondary structure are colored in blue. Collected interresidues NOEs are shown as red sticks. (C) TMB2.3 Rosetta design model (magenta) aligned to the 20 lowest-energy models generated with NMR constraints (gray

To determine structure at the atomic level, we crystallized TMB2.17 and solved the structure at 2.05-Å resolution (table S4). All but two residues located in one trans β-turn were resolved in the electron density map. The crystal structure of TMB2.17 closely matches the design model (1.1-Å backbone RMSD over all residues, Fig. 1.6A), and the β barrel has a wide lumen delimited by glycines in an extended conformation, which form kinks in the β strands as designed (Fig. 1.6, B and C). The two YGD/E interactions (Y69, Y11, G27, G89, D39, E103) belonging to the extended mortise–tenon motifs are present in the crystal structure, and the second shell of interactions, involving K71, E53, and Q29, is also properly recapitulated with additional interactions to water molecules (Fig. 1.6D); these extended side-chain hydrogen bond networks fill the lumen of the β barrel. Overall, the buried amino acid side-chain conformations and interactions in the design model are in very good agreement with the crystal structure (Fig. 1.6E; compare pink and gray).

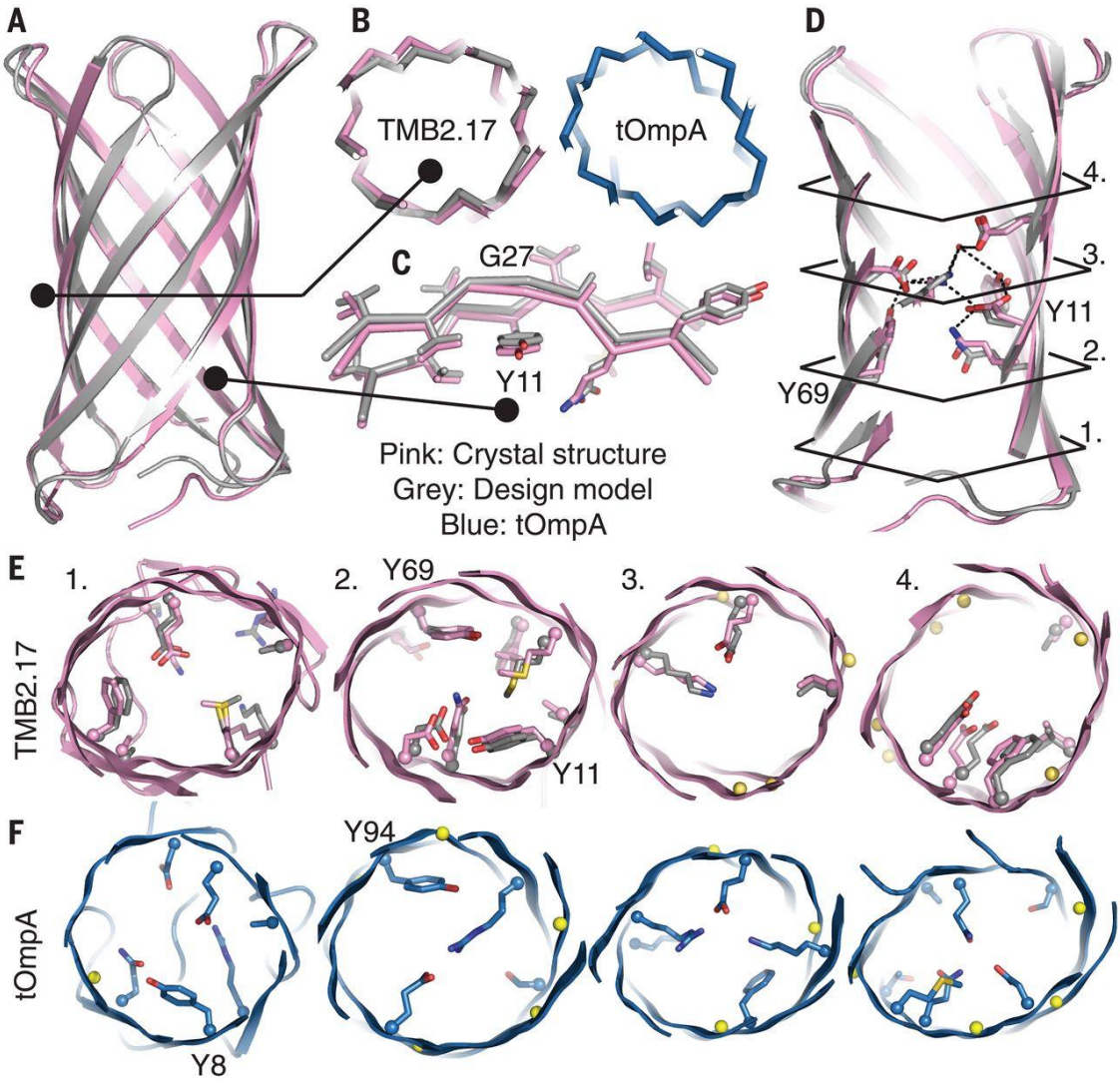

Fig. 1.6 Crystal structure of TMB2.17 is nearly identical to the design model. The crystal structure (pink) of TMB2.17 determined in DPC detergent, superimposed on the design model (gray), and compared to the crystal structure of the naturally occurring tOmpA (blue; PDB ID, 1BXW). (A) Full backbone superposition. (B) Comparison of transverse β-barrel cross-section geometries. (C) Superposition of the β strands around a mortise–tenon motif, showing the extended backbone conformation of the glycine kink (G27) and the rotamer of the tyrosine involved in the aromatic rescue interaction (Y11), which are nearly identical in crystal structure and design model. (D) Superposition of the side chains involved in the core network of polar interactions around the two mortise–tenon motifs. The black lines indicate the locations of the four transverse slices for which core packing is shown in (E) for the design model and crystal structure; the two are very similar. Corresponding slices in tOmpA (F) are quite different in both shape and amino acid composition and placement. Cα atoms are shown as spheres, and glycine kink residues are colored in yellow; the positions of the tyrosines in the mortise–tenon folding motifs are labeled.

We compared TMB2.17 to the transmembrane region of tOmpA, the only natural TMB sequence with a known structure returned by a BLAST search for sequences similar to TMB2.17 in the nonredundant sequence database (the E-value of 1.6 is in the range expected from random matches; alignment shown in fig. S30). The shape of the β-barrel lumen is quite different in the two proteins (Fig. 1.6B), as are the amino acid identities and packing arrangements of the core side chains (Fig. 1.6F).

Conclusions

Both the initial failures and the ultimate success of our hypothesis, design, and test approach to *de novo* TMB design inform our understanding of the sequence determinants of TMB folding and structure. The sequential approach previously used to build helical transmembrane proteins (*6*)—the design of proteins with hydrophobic cores compatible with folding of water-soluble proteins and subsequent hydrophobic residue resurfacing to convert them to membrane proteins—yielded sequences strongly predicted to form amyloid. Designs with more polar cores, which had high β-sheet propensity because of the perfect alternation of hydrophobic and polar residues, systematically failed to express in *E. coli*. Iterative improvement of the design protocol ultimately enabled the generation of a set of sequences, with more than 8% of sequences encoding proteins able to adopt a β-barrel fold (based on ^{1}H-^{15}N-HSQC NMR). The NMR structure and high-resolution crystal structure of two of these designs are very close to the design models. The key to this success was

introducing glycine kinks, β bulges, and register-defining side-chain interactions—also critical for the folding of water-soluble β barrels (*5*, *53*) and hence important for defining β-barrel architecture irrespective of the solvent environment—and balancing the hydrophobicity and β-sheet propensities of the sequences.

Our results suggest that, to enable TMB expression and folding, the β hairpins of outer membrane β barrels need to be sufficiently unstable in water that they do not form off-target β-sheet–containing species and become populated at high levels only in the context of the fully folded state in the hydrophobic environment of the membrane. Slowing down the folding and assembly of trans hairpins could also allow more time for passage of the mostly hydrophilic amino acids in these β-strand connections across the lipid membrane, which likely has a large activation barrier. The overall β-sheet propensity and hydrophobicity of our successful designs are in the range of those of naturally occurring TMB sequences, suggesting that the natural TMBs might be under a similar negative selection pressure against formation of non-native β-sheet structures in an aqueous environment (*54*). Further work is required to determine whether the design principles applied here enable TMB folding into biological membranes [whose properties present a formidable kinetic barrier to folding (*55*)]. In Gram-negative bacteria, the BAM complex is responsible for accelerating the assembly of natural TMB substrates into the outer membrane by

lowering the kinetic barrier to folding (*55*). Our design strategy incorporates neither signals for BAM complex association, such as the conserved β signal (*56*), nor evolution-conserved functional motifs and hence represents a "blank slate" for probing the trade-offs between TMB folding, stability, and function, as well as the evolutionary constraints on OMP trafficking and biogenesis.

Larger TMBs share similar sequence properties with the eight-strand TMBs considered in this study (supplementary text), suggesting that the general design principles and methods that we have described here should be applicable to the design of larger pore-containing β barrels, after the generalization of the β-barrel architecture definition rules (glycine kinks, β bulges, etc.) to different combinations of n and S. The extent to which essentially all of the key design features are recapitulated with atomic-level accuracy in the crystal structure of TMB2.17 suggests considerable control over TMB structure, which should enable custom design of transmembrane pores with geometric and chemical properties tailored for specific applications.

Chapter 2. Design of Symmetric Metal Binders

2.1 Introduction to Metalloenzymes

Metalloenzymes catalyse a large number of complex reactions using a variety of transition metals in their active site. The metal centers of these proteins play many different roles in these reactions, from acting as simple lewis acid catalysts, to performing complex multi-step redox reactions involving multiple charge and spin states to catalyse reactions such as carbon fixation or water splitting. Each transition metal has a preferred coordination state and binding geometry. For example, iron prefers to bind molecules at 6 points around it in octahedral geometry [36]. Each catalytic reaction usually proceeds through specific oxidation and spin states of the metal. Natural metalloenzymes have well optimized the coordination number of the metal, the binding geometry, and the type of amino acids to best bind the metal for the reaction at hand. Some of the most widely encountered metals in natural metalloenzymes are copper, zinc, and iron.

Copper sites are generally separated into classes based on spectroscopic properties. Copper complexes are redox active and are often involved in O_2 binding, oxidation, hydrolysis, and substrate activation [37,38]. The common categories of copper metalloproteins are type 1 blue copper, type 2 normal copper, and type 3 coupled binuclear copper centers[39]. Type 1 blue copper proteins are generally not enzymatically active and are involved in electron transfer pathways [40]. Type 2 copper enzymes consist of categories of oxidoreductases, including superoxide dismutase and diamine oxidase, as well as hydrolases, such as phenylalanine hydrolase [40]. Finally, type 3 binuclear

copper clusters have oxidoreductive activity, such as catechol oxidase, tyrosinase, and some hemocyanins with catalase activity [37,40]. In addition to these classes, there are trinuclear copper clusters, such as in laccases, and complexes such as those found in Heme-copper enzymes, which are involved in reducing O_2 [41], as well as NO_2^- and N_2O reduction [42].

Zinc is necessary for catalysis in enzymes spanning all of the original 6 Enzyme Commision categories [43]. Because of its fully occupied d-orbital, zinc can take any geometry with equal stability [36]. Despite this, it is common for zinc to adopt a distorted tetrahedral geometry [44] with protein coordinating three of the four points, with a single molecule of water coordinating the last coordination site [45]. A common motif found is a single zinc ion coordinated by three histidines, two coordinated via the N^{ε} atom and one via the N^{δ}, as seen in carbonic anhydrase [46].

Iron metalloenzymes come in three main categories: non-heme, heme, and iron-sulfur clusters. Non-heme iron metalloenzymes generally bind a single or dinuclear iron in octahedral geometry. These proteins often catalyse reactions using cofactors, such as O_2 and α-ketoglutarate [47,48]. A common motif in non-heme iron complexes is the 2-His 1-carboxylate motif, which catalyses reactions such as catechol dioxygenase, Rieske Dioxygenases, and other oxygenases [47]. Iron-heme enzymes are generally oxidoreductases, and can be seen partnered with copper ions in the electron transport chain, as mentioned above. Some examples of iron-heme enzymes include cytochrome P450, cytochrome c oxidase, and indolamine 2,3-dioxygenase [49,50]. Iron-sulfur clusters

are also incredibly redox active and they are often involved in electron transport, and are able to perform radical chemistry, such as in radical *S*-Adenosyl-l-methionine reactions [49,51].

While there are a huge variety of reactions catalysed by natural proteins, there is still a need for designer enzymes. In many cases, there is a demand for the reengineering of native enzymes to perform efficiently under the conditions required for new industrial or medical applications. An example of this is the thermostabilization of cytosine deaminase, which has potential uses in antitumor activity [25]. In other cases, the native enzyme is so complicated that its industrial application is not practical; photosynthesis, water splitting, and ATP generation are examples of such systems. Despite the diverse reactions performed by natural proteins, there are still many more reactions that are desired, such as synthesis of new molecules for medicine or breaking down harmful chemicals. While researchers have been working on generating novel enzymes to perform such reactions, creating new enzymes remains an outstanding challenge. The common approaches used to tackle this problem are evolving natural scaffolds [52] or rational design of metal binding sites into naturally occuring proteins [53,54].

However, natural proteins often work in a small range of temperatures or chemical environments and many of them cannot withstand extensive mutagenesis. In addition, natural proteins span only a small part of the protein structural and sequence space; the shapes that these proteins offer may not be suitable for reactions not observed in nature, or even for substrates that have structures which substantially differ from the

native substrate. Because natural proteins can only be tuned to a limited extent without affecting their stability or functionality, the role natural proteins can play in industrial settings is limited. In contrast, *de novo* scaffolds can be designed from any part of protein structural space to match the shape needed to bind substrates, to be stable under the required reaction conditions, and to perform the desired reaction.

Therefore, developing methods for designing metalloenzymes from scratch facilitates the generation of proteins that can catalyse any desired reactions across the spectrum of enzyme-catalysed reactions. The field of *de novo* designed metalloenzymes has seen great progress in the last few years. Many exciting advances have been made, including binding to a wide variety of metals in both mononuclear [6,55] and dinuclear sites [56]. However, most of these designed proteins involve binding metals inside a single 3- or 4- helix bundle [29,30]. The limitations of this shape restrict the functionalities possible for these enzymes. In order to design a wider range of enzymes, it is important to develop new methods that enable the design of metal binding proteins in different shapes and sizes, tailor-made for the specific reaction in mind. With recent advances in *de novo* design of proteins, it is now possible to generate a wide array of scaffolds [15,57–60] and create a diverse set of active site topologies to enable a variety of novel reactions. My goal is to create a *de novo* metalloenzyme by designing the entire scaffold around a desired active site, allowing for a generation of topologies custom designed for the reaction of interest, while still maintaining the stability that designed proteins can afford.

2.2 Modification of Pre-validated Protein

Initial designs were made by attempting to place a metal binding site into an existing de-novo designed protein [Figure 2.1a]. The original protein was created by taking an existing tertramer made from 2 helix bundles, 5L8HC4_6 [59], and adding repeat proteins to the bottom to stabilize the design as a tetramer [31] . In order to modify the backbone to accommodate a metal binding site, loops connecting each pair of helices were cut and reconnected using three different Rosetta methods: Rosetta Remodel [61], Rosetta ConnectChainsMover, or the Rosetta DirectSegmentLookupMover. These grew and reconnected the monomers, adding backbone diversity in the area in which we were trying to bind metals [Figure 2.1b]. Once the backbone was set, histidines were placed symmetrically in the protein using a script, called the symmetric matcher mover, that took in a set of backbones, a symmetry file, and a parameter file describing a histidine bound to a nickel ion in square pyramidal geometry [Figure 2.1 c and d]. The script mutates a given set of positions on chain A to the metal binding histidines, re-symmetrizes the protein and then outputs a set of designs where the metal coordination sites of the bound metal overlap with the generated symmetry mates, indicating a single coordination site. We then designed the amino acid sequence with FastDesign with restrictions that prevented modification or repacking of the histidines, and allowed design only at the modified portion of the backbone and a short sequence flanking it to either side. These designed proteins behaved well via SEC, and this seems to be a viable strategy for design of de-novo metal binders. This design strategy was abandoned for 2 reasons. First, because it was determined that the single protein had a limited set of feasible backbone modifications to create new metal binding sites,

so the diversity of designs was very low. Second, because protein was too large to be an ideal test case for validation the binding of a single metal ion using UV-Vis, as it would need a 500 μM concentration of protein, or 50 mg/mL concentration to be able to see the signal.

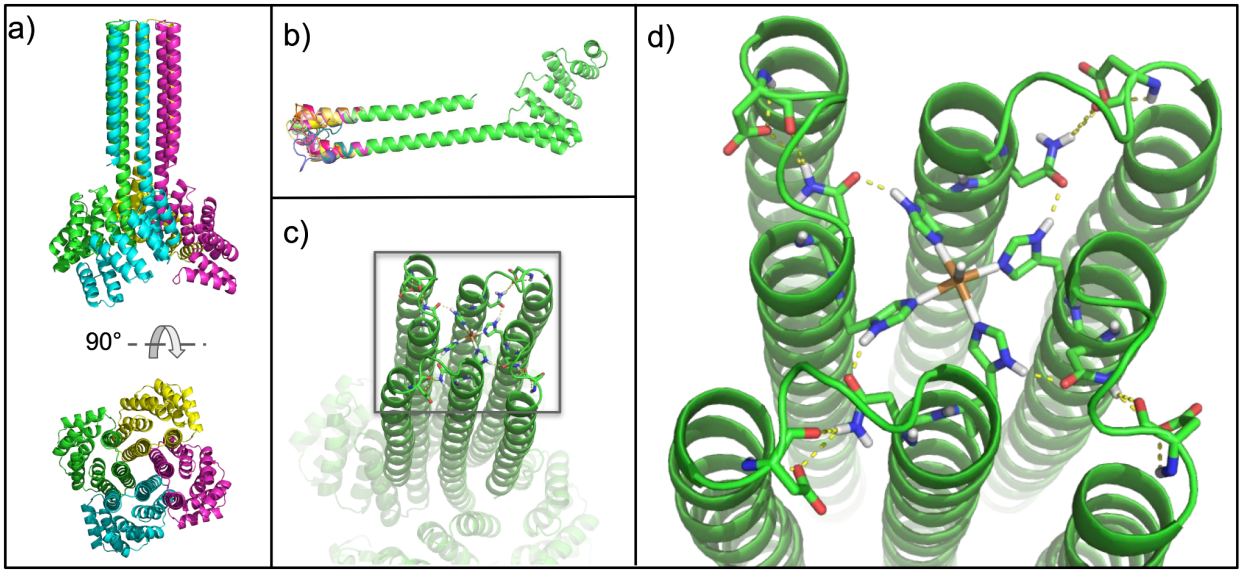

Figure 2.1: Example modifications of a pre-validated scaffold. a) shows the base homotetramer design model combining the stabilizing domains in (Lu et al 2018) [31] with the original soluble design helical protein 5L8HC4_6 (Boyken et al 2016) [59] b) shows a monomer subunit of the homotetramer(green) with loop backbone variations overlaid in order to show the available diversity. c) shows an example redesign in the context of the full protein, and d) shows a close up of the modified loops, showing a proposed binding site, including the 4 histidine binders interacting with a nickel atom(brown) including virtual atoms describing the desired coordination geometry (white)

2.3 Generating Protein Backbone Around a Binding Site

In order to address lack of backbone diversity, and reduce the size of our proteins in order to ease validation by UV-Vis, we set out to generate new scaffolds around a metal binding site. We continued to use helical bundles as the basis of the design, as there are robust published methods describing the design of *de novo* helical bundles[30,57,59,62], and helical bundle oligomers [63,64]. The goal of this project was to create a symmetric

scaffold with histidines binding a single metal ion. Three histidines binding a metal ion have been shown in nature to have enzymatic activity, as in the case of carbonic anhydrase [65]. If a metal is bound on the symmetry axis, the available coordination geometries are limited to the ones that match that symmetry. Given this, the use of a symmetric scaffold to bind a metal ion at the symmetry axis also improves the likelihood that a bound metal will be binding with the correct coordination geometry. This means that testing for oligomeric state and metal binding is enough to give us a strong initial signal of accurate binding rather than requiring more complicated or time consuming methods as initial assurances of proper structure.

Protein backbones were generated around a 3 histidine symmetric binding site. Backbone positions of the histidine residues were sampled by rotating histidine residues around the axis of coordination with the metal, allowing for a variety of different starting conformations [Figure 2.2a]. Short idealized helices were then grown from histidine backbone positions using Crick coiled coil equations [57] [Figure 2.2b]. A diverse set of 3-helix bundle homotrimeric scaffolds were then generated from each of these short helices using fragment insertion with Rosetta remodel [61] [Figure 2.2c]. Generated scaffolds were then checked to ensure they have solid core and interface structure, and that they have pockets above the coordination site which allow for the metal to access the binding residues. In order to pre-orient the coordinating histidines to bind a metal between them, I designed hydrogen bond acceptors in an orientation to bind the delta nitrogen of each of the histidines. These second-shell interactions are common in

metalloenzymes and are often very important for tuning the reactivity of the metal ion [66,67].

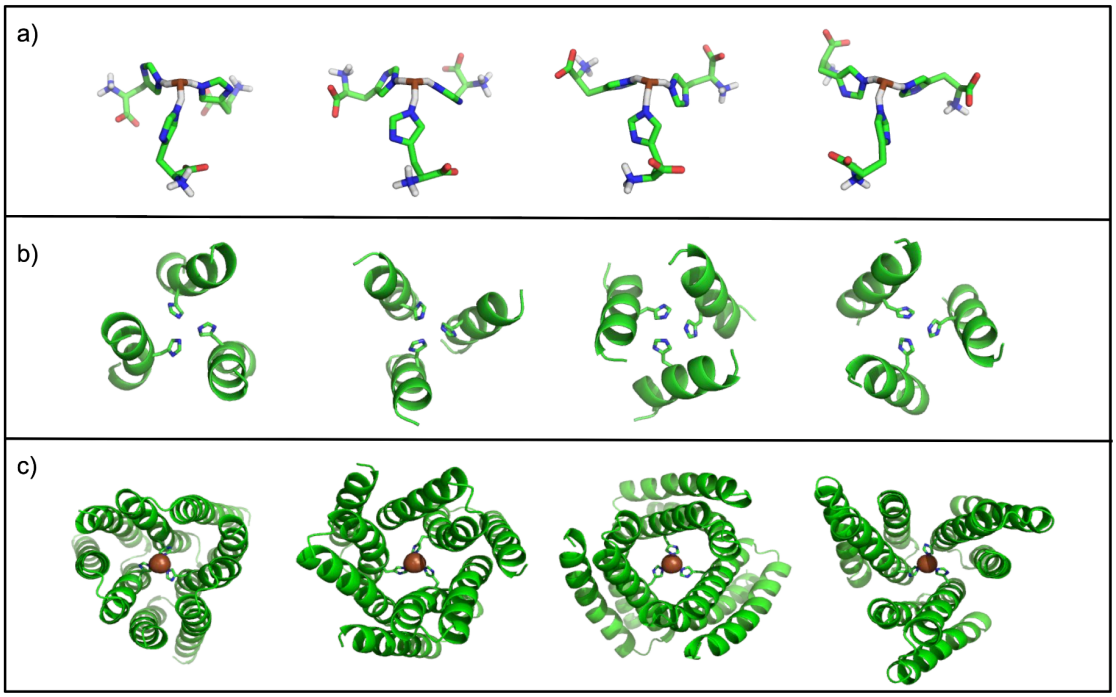

Figure 2.2 Backbone generation of new metal binding proteins a) proposed metal binding motif with histidine backbone positions in different locations b) Short helices grown out from histidine backbone position and were checked for clashing c) Example 3-helix trimers with fully grown backbones

2.4 Validation of Oligomeric State and Metal Binding

SEC was used to validate the initial state of designed proteins with soluble expression. Most of the proteins appeared to be monomers. Simultaneously, UV-Vis experiments were done to see if metal binding occurred. UV-Vis was carried out on a Jasco V-750 spectrophotometer using an 8-strip sample holder requiring 100µL of sample (Jasco model 660.236-QS). Cobalt was used for binding experiments because it is spectroscopically active upon binding, and can readily take tetrahedral and octahedral geometry [36]. Designs were measured between 300 and 800 nm with cobalt being added

to the protein sample and a blank in 100 μM increments from 0 μM to 1000 μM, and then 1500 μM and 2000 μM. The curves were then normalized by subtracting the 0μM cobalt sample and the appropriate cobalt concentration to obtain a graph of spectroscopic activity due to metal binding [Figure 2.3 c-e]. Some proteins appeared to be C3 by SEC using a Superdex 75 Increase 10/300 GL column. As an example, sg46 and sg47 have calculated oligomeric masses of 27.7 kDA and 28.4 kDa respectively. Both of these designs have major peaks appearing near 12.2 mL [Figure 2.3 a and b]. This coincides closely with the elution volume of carbonic anhydrase (MW 29kDA, elution volume 12.5 mL) when run as a control in the manual for the Superdex 75 Increase 10/300 GL column [68]. It also appeared that sg47 had metal binding activity, with an increase in absorbance around 540 nm upon the addition of cobalt [Figure 2.3 d] that was not seen in sg46 [Figure 2.3 c], an elimination of the absorbance when the designed metal binding histidine is changed to an alanine [Figure 2.3 e] and a clear visual purple shift in the vial after addition of 2000μM cobalt compared to other designed proteins with the same cobalt concentration added [Figure 2.3f]. However, the proteins do not have the correct oligomeric state according to SEC-MALS, where the multi-angle light scattering calculations found masses of 10.8 kDa and 10.0 kDA, respectively. Instead of associating in its designed trimeric state this indicates it was instead predominantly monomeric. The monomeric state of these designed proteins in turn indicates that any metal binding seen is not binding as designed.

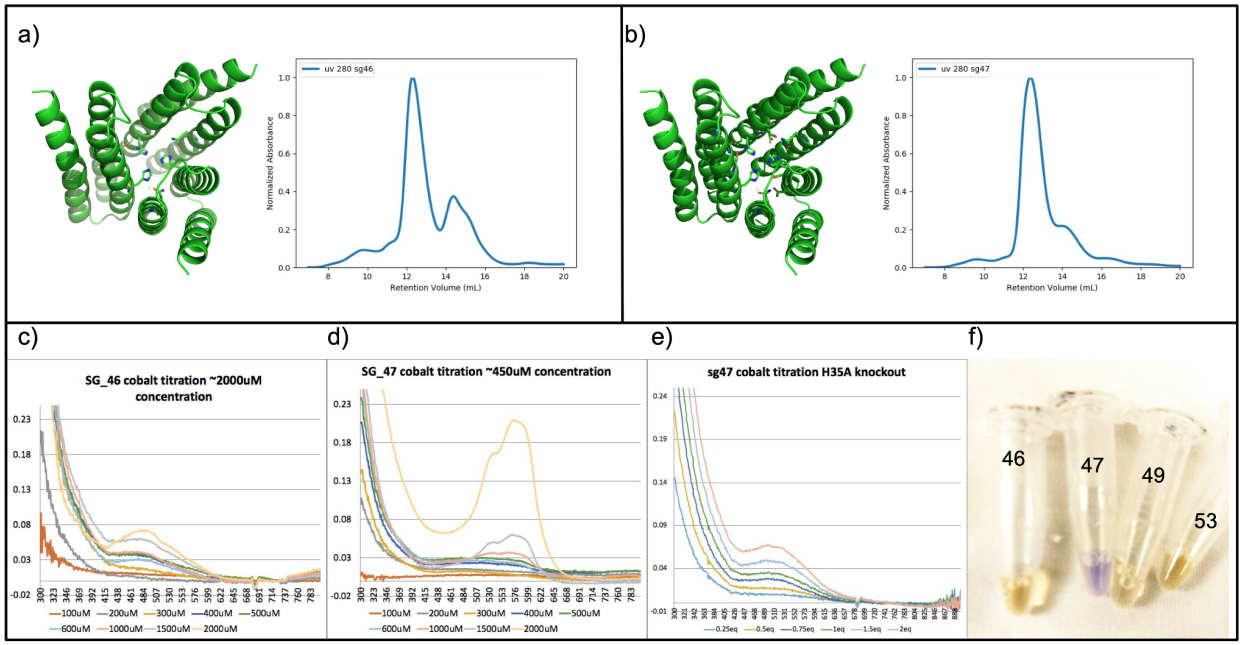

Figure 2.3: SEC traces and metal binding curves for two designed homo-oligomers a) sg46 designed structure and superdex 75 increase SEC trace b) sg47 designed structure and superdex 75 increase SEC trace c) sg46 metal binding curve, showing small increase in absorbance around 484 nm d) sg47 metal binding curve, showing large increase in absorbance around 540 nm e) metal binding curve sg47 knockout showing the knockout of designed binding histidine removes peak at 540 nm f) picture of vials with purified sg46, sg47, sg49, and sg53, with 2000μM cobalt added. sg47 has a visible blue shift, in contrast to other designed proteins.

2.5 Discussion and Future Directions

This work was incomplete and unpublished due to the very low rate of proteins with soluble expression that were well folded and had the correct oligomeric state. There are a few considerations that anyone wishing to pursue this research course should be aware of. First, the strategy of modifying existing scaffolds is viable, and has been done with natural proteins for a very long time [52–54].The initial piece of advice I would have to anyone trying this is to compile as many pre-validated homo-oligomeric scaffolds as possible with molecular weights < 50 kDa. The modified scaffolds from already validated

proteins have a much higher likelihood of expressing and forming the correct oligomer than newly generated scaffolds. Using a variety of initial scaffolds would mitigate issues of having very low output and variability of designs.

If pre-validated scaffolds remain insufficient, scaffold design should be done. The process that was described here for generating backbones is very inefficient. Less than 10% of the generated backbones pass the worst9mer filter leaving a very small starting sample for design. Given that the fragment insertion method is costly both in real and computational time, this is not a good strategy. A more efficient idea would be to either use pre-validated monomeric proteins, or generate new monomers entirely by using crick coiled-coil parameters, a method that can output thousands of backbones in a few minutes. The generated monomers should then be thoroughly screened down to a small set of backbones with optimal backbone parameters, such as omega dihedrals and backbone-backbone hydrogen bonding, and then designed and thoroughly filtered again. Ideally these monomeric designed proteins would then be ordered and validated, and only the well behaved ones would be used as a basis for the oligomers.

Cyclic docking should be done using Rosetta rpxDock to dock all of the validated monomers into your desired symmetry. All docks should be initially filtered by rpx score and the number of contacts at the interface. After that, docks should be filtered by using the symmetric matcher mover, described in section 2.2, with the caveat that the only positions checked by the mover should be positions at the “surface” or “boundary” of the protein. Because the sequence is going to be redesigned, these locations should be

calculated by having two or fewer neighboring residues, rather than solvent accessible surface area. Designs that have positions with the correct geometry to bind a metal ion can then be filtered for the accessibility of the metal binding site to the outside environment and be redesigned to have a proper interface.

Metal binding experiments should wait until after clear validation of at least oligomeric state, given that an incorrect oligomeric state necessarily means that any metal binding is not as designed. The third section of my thesis focuses on the creation and validation of new symmetric homo-oligomers, and though my focus was on creating pockets at the asymmetric interface, the method is highly applicable to creating new scaffolds for binding at the symmetric interface.

Chapter 3. Design of Homo-oligomers with Pockets at Asymmetric Interface

3.1 Introduction to Pocketed Scaffolds

Designing *de novo* proteins specifically for small molecule binding and enzymatic activity is an ongoing problem. Creating a new reaction or binder requires having a protein with a pocket of a size and shape that can accommodate the desired function. There are numerous successful cases of designing new functions into natural proteins [69–71].However, this method is inherently limited by the finite set of natural proteins with viable shapes for enzymatic activity and small molecule binding. Rather than screening through large numbers of natural proteins for one with specific desired properties, *de novo* design allows for the creation of new proteins with specific attributes. Previous work has been done on various scaffold types to create pockets for binders, such as mixed alpha-beta proteins with NTF2-like folds[14,15] and beta barrels [10,13]. While these methods have achieved success, the position of the pocket in the center of the monomeric fold naturally restricts the positions that can be mutated without destabilizing the core of the protein.

Looking to nature for examples of protein shapes that do not encounter this problem, we find a number of enzymatically active proteins with their active site at the asymmetric interface of a symmetric scaffold. Symmetry plays a vital role in the structure and function of many natural proteins [72]. Examples of this include proteobacterial carboxymuconolactone decarboxylase (PDB 2QEU) and human glutamine synthetase (PDB 2OJW). By designing similar oligomeric structures with pockets that are not

involved in stabilizing the protein conformation, we were able to generate a new set of modular scaffolds that were amenable to functionalization. In addition, we were able to create proteins with the ability to be functionalized on their own, or be easily incorporated into larger assemblies. These larger assemblies include *de novo* 2D materials [11] or *de novo* nanocages, which have a variety of functions and require 3-fold, 4-fold, or 5-fold symmetric components, depending on the geometry of the cage [9,73,74].

We set out to generate symmetric oligomers with binding pockets at asymmetric subunit interfaces. We took advantage of the wide body of literature regarding the generation and design of helical repeat proteins [60,62,75]. Helical repeat proteins have recently been used as base building blocks for the generation of protein filaments [76], nanocages [77,78], and a variety of other nanomaterials [79,80]. Sets of curved helical repeat proteins make these a desirable monomeric building block, both due well-defined rules for generation and because the geometry of these curved repeats allow for the formation of a wide variety of pockets [Figure 3.2 d]. By combining these curved helical repeat proteins with a recently developed method for in silico docking of cyclic oligomers [64] we can quickly generate large numbers of designs with topologies which allow for varying size cavities across the asymmetric oligomeric interface.

3.2 Monomer Generation and Oligomeric Docking

Monomers were generated computationally using Rosetta to generate designed helical repeat (DHR) proteins using the helical extension method described in “Perturbing the energy landscape for improved packing during computational protein design”[75]. The

method is similar to the SEWING protocol, and assembles helical fragments into helical bundles by aligning and combining overlapping helical and loop fragments together when it finds a consecutive pair of residues along the fragment with an RMSD of 0.3 Å or less, and then recombining generated helical bundles into a curved helical repeat protein. Monomers were separated into two groups, named bd4 and bd5, based on the type of twist. Bd4 monomers create curvature by shifting the overlap between helices such that the n terminus of helix i will interact with center residues on helix i+2 of the repeat protein [Figure 3.1a]. In contrast, bd5 monomers curve only around the axis down each helix, but keeping ends of the helices more aligned across the length of the monomer [Figure 3.1b].

Monomers were docked using Rosetta method rpxDock [64] to symmetrically sample ways for the monomers to interact in order to form an oligomer. 3-fold to 6-fold symmetries were generated using symmetry files located in the Rosetta database and called following the method described previously [64]. Docked models were then filtered by the number of contacts in the interface between subunits, and by the predicted quality of the interface interactions. 400 monomers were docked in 4 symmetries, and output the top 100 docks of each symmetry, resulting in 1.6 million docks.

The 1.6 million resulting docks were then filtered based on the two score terms outputted from rpxDock. The two metrics were residue-pair transform (rpx) - a total score describing the quality of the pairwise interactions of backbone N, Cα and C atoms of interacting residues and ‘ncontact’ - the number of residues in a docked protein

model involved in the interactions between two chains. These scores are further described in the paper which introduced this method [64].

Filter limits for ncontact were determined by binning and outputting designs in sets of 10 random designs with scores ranging from 40 to 140 contacts (i.e. 40-49, 50-59, 60-69, etc. These binned designs were then visually screened for docks that had at least one full helix-helix interaction between the interface, to create the lower bound on contacts. To create the upper bound of contacts, docks were visually screened for the presence of pockets, because a dock that was too interconnected would not have any space available to form the pocket. From this binning strategy we determined that an ncontact score below 80 would not allow for enough interface to form an oligomer [Figure 3.1c], and an ncontact above 120 would be too interconnected to form pockets [Figure 3.1d]. These boundary values will vary depending on your input monomers, with smaller monomers or shorter helices necessitating lower numbers for both the lower and upper cutoffs, and larger monomers allowing for higher cutoffs. Rpx scores were filtered in a similar binning manner, and we found that scores above 50 showed mostly parallel helical interactions, allowing for solid residue packing, and scores below 50 had a higher incidence of ‘V’ and ‘X’ packing [Figure 3.1 e and f]. ‘V’ packing describes helices that are angled away from each other [Figure 3.1e] in an orientation that would require very small residues near the closest section and progressively larger residues as the helices grow apart. ‘X’ packing has helices that cross one another [Figure 3.1f] in a way that reduces the size of the interface and potentially interferes with the interdigitation of residue packing that is easy to accomplish with more parallel helices. Both of these

features are non-ideal for residue packing and stabilization, and cannot be easily remedied through sequence design.

The two filter metrics, ‘rpx_score’ and ‘ncontact’ were highly correlated, with a Pearson coefficient of 0.60. We determined that an additional decoupled metric was needed that describes the rpx score per interface contact. Calculating ‘rpx_score’^{2}/’ncontact_score’ was sufficient to decouple these score terms - giving a Pearson coefficient of 0.09. This new ‘rpx_per_contact’ score was binned and filtered the same way as the total rpx score, and all designs that scored below 40 were filtered out.

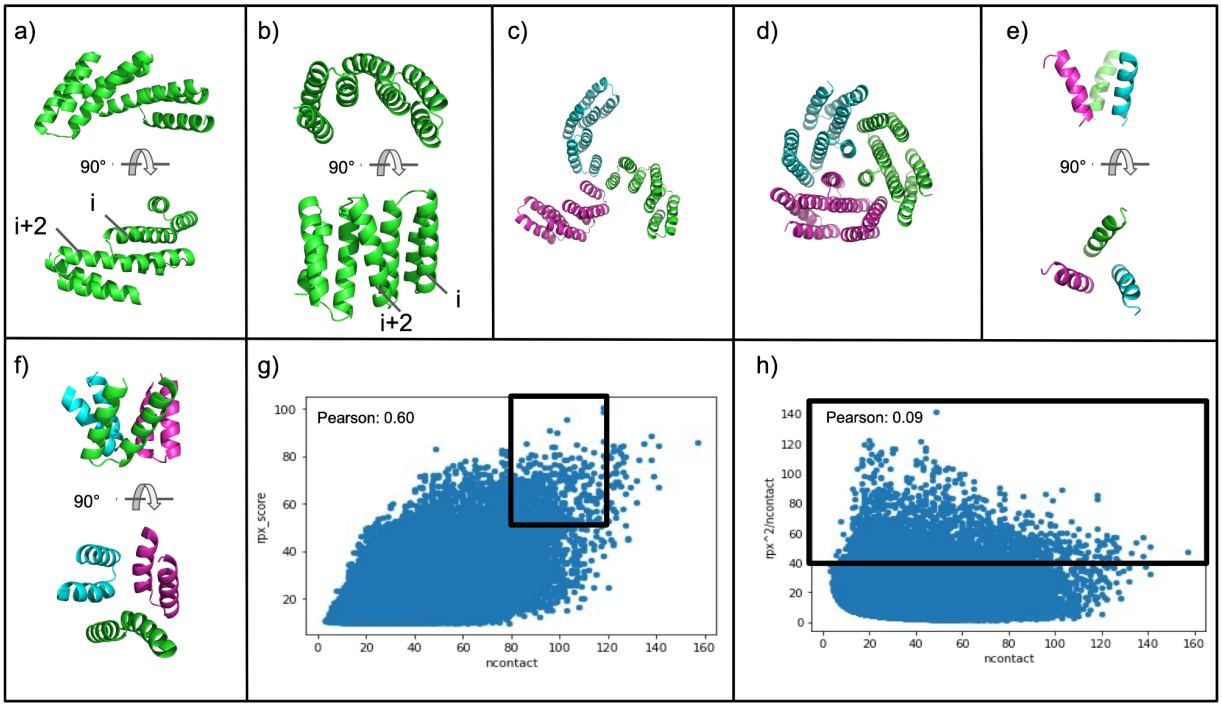

Figure 3.1: Creating and filtering homo-oligomers by rpxDock metrics a) Example input monomer from bd4 monomer set b) Example input monomer from bd5 monomer setc) Example oligomer dock with ncontact metrics at the lower bound of filter d) Example oligomer dock with ncontact metrics at the higher bound of filter e) Example of interface dock with V packing, scoring below rpx filter f) Example of interface dock with X packing scoring below rpx filter g)Graphs of ncontact and rpx score, h) graph of rpx squared with ncontact. The box in e-f) indicates the section of scores which pass filters.

3.3 Pocket Determination and Variety

Pockets were identified by a script that takes an input of a list of protein PDBs and outputs the 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 path to the design. Pockets were determined by converting chains A and B of each protein to a space-filling 3d model using voxels [Figure 3.2a], then calculating convex hull around protein (the smallest bounding box using triangular faces) and excluding all space outside of that hull, leaving only empty space within the protein to be considered. Each discrete empty space inside the protein is then filled with a different "color" indicating separate potential pockets. The volume of each “color” is then calculated, and the largest color is determined to be the pocket of interest [Figure 3.2b]. The moments of inertia were then calculated as though the pocket had mass of the spheres used to fill the space, allowing us to calculate the primary axes and the NPRs.

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 ratio of I_1/I_3 and I_2/I_3 were outputted 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.[81]. In addition, we modified the graph to also show the size of each pocket by the color of the point. Each volume range covers a wide area of the available shape space, with little overall correlation between the size of the pockets and the shapes that they may take [Figure 3.2 d]. Examples of protein docks with rod-like pockets [Figure 3.2e], disk-like pockets [Figure 3.2f], and spherical pockets [Figure 3.2g]

were located in the appropriate regions of the graph. The coverage of shape space dwindled near the rod-disk axis, which ran from completely rod-like shapes at (0,1) to completely disk-like shapes at (0.5,0.5). This was in contrast to many small molecules, which often tend towards geometric planarity [82]. Because the pocket volumes were calculated using only backbone and Cα positions, shape complementarity to a given small molecule could be further improved after filtering via subsequent design steps. This script allows future scientists to filter for the desired size and shape of the pockets using the same method they would use to describe the small molecules that one might desire to bind in them.

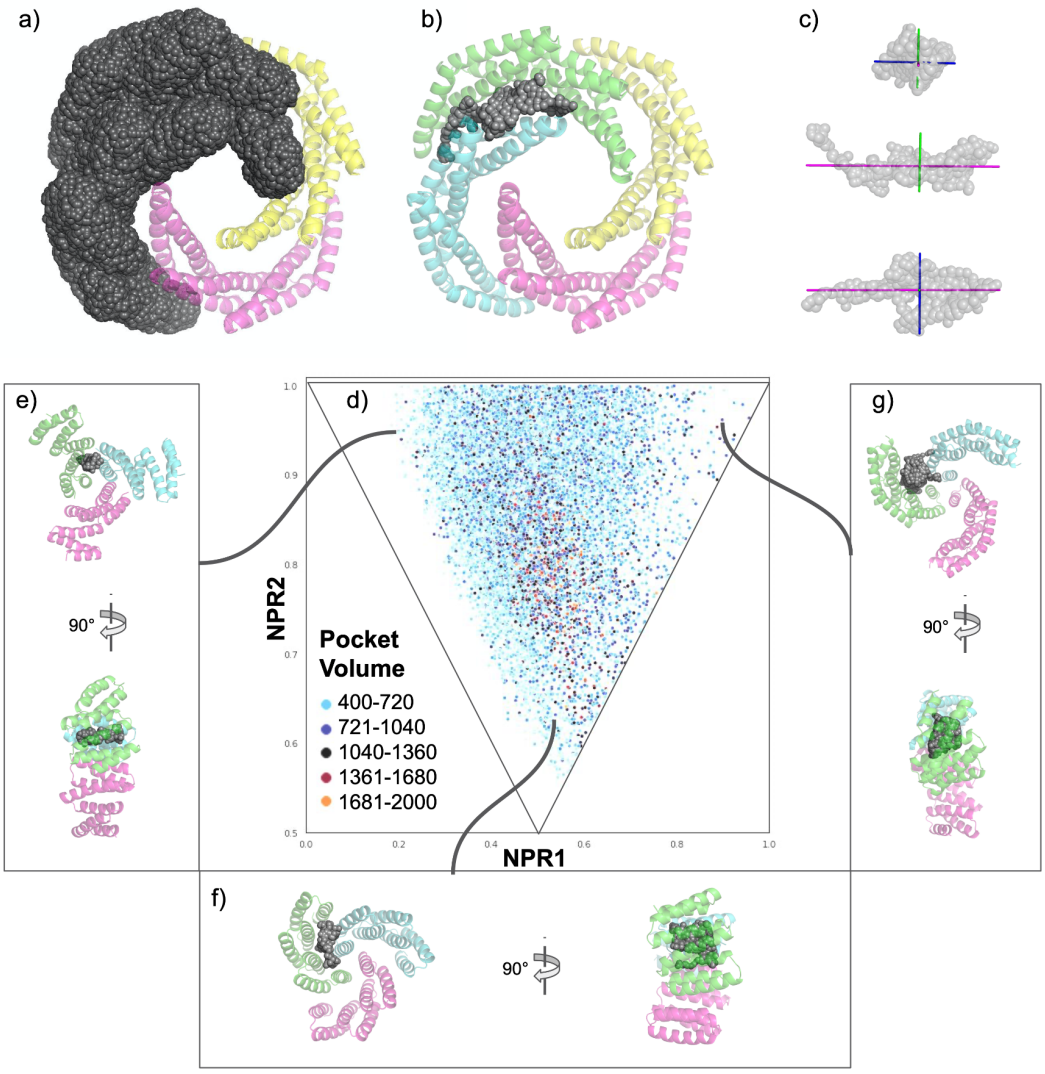

Figure 3.2: Pocket determination at asymmetric subunit interfaces. a) sg135 spherical representation of every N, C, O, Cα atom in chains A and B b) sg135 pocket c) pocket x y and z debug images from script d) pocket determination graph **d)** Designs span a wide range of pocket shapes and volumes e) Example rod-like pocket shaped dock, NPR1 0.228, NPR2 0.964, pocket volume 520 $Å^3$ f) Example disk-like pocket shaped dock, NPR1 0.554, NPR2 0.639, pocket volume 932 $Å^3$ g) Example sphere-like pocket shaped dock, NPR1 0.925, NPR2 0.945, pocket volume 823 $Å^3$

3.4 Sequence Design

Two RosettaScripts [83,84] 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 Appendix 4. Interface residues were defined as any residue

in chain A within 10 Å of chain B, and any residue in chain B within 10 Å of chain A, with the symmetry designator allowing for selected residues being modified on chains C, D, E, and F in 3-fold to 6-fold symmetries. 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, i.e buried helical residues could only be designed with the 12 amino acids A F I L V W Y D N S T and H [Appendix 4 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 a number of relevant score terms, including the total score, probability of amino acid phi psi angles (to insure residues were not placed in positions where their preferred phi and psi angle were disfavored), the presence of unsatisfied buried charged atoms (because we did allow HIS, ASP, and ASN in core positions, to allow for hydrogen bonds and salt bridges at the interface), the comparison of the quality of packing and residue interactions as a monomer vs as a complex (to ensure that the protein preferred to be an oligomer, rather than a monomer).

Initially the interface was the only round of design done, as the initial monomers were pre-designed, though unvalidated. However, this was found to have a high rate of insoluble expression. DNA for 32 designs were ordered. Of those, only 25 out of the 32 ordered plasmids were able to be transformed into an expression strain of *E. coli*. Once the *E. coli* was grown and lysed, we found that 16 of the proteins were either insoluble or didn't express. Nine proteins had soluble expression, but of those 3 proteins had such low expression rates that there was not enough protein purified from a 50mL

expression to analyze by either SEC-MALS or Mass Spectrometry(MS) (<15μg), leaving only 6 proteins that had soluble expression with high enough expression to analyse with SEC-MALS and MS. This gives a total usable soluble expression rate of 6/32, or <19%.

We then added a second round of design, which redesigned the surface of the protein to reduce surface hydrophobicity, which likely caused the aggregation and insoluble expression. An example script for redesigning the surface is given in Appendix 5. This script used only a single round of FastDesign and FastRelax, as this was the third time that the oligomer or its components were being designed, and the overall structural change were expected to be minimal. These proteins were designed using 'beta_genpot' weights with an overall net charge score term to even out 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. 2012) [85]. The surface was designed with the same DesignRestrictions settings as the interface design script run previously. Additional filters were added, including 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.

After addition of the surface redesign script, cDNA for a total of 138 independent designs was ordered. Of these, there was only a single instance of insoluble

expression, 8 had soluble expression at levels too low to characterize, and 58 had soluble expression at high enough levels for characterization, for a total useable soluble expression rate of 56/138, or 41%, which is more than double the previous soluble expression rate. Of the 106 designed proteins with plasmids that transformed into the protein expression strain of *E. coli*, 53% of them had soluble expression.

3.5 Characterization through Size Exclusion Chromatography and Small Angle X-ray Scattering

Genes encoding each of the 138 designs were synthesized and cloned into a vector with a T7 promoter system and either an N- or C-terminal (His) tag. Plasmids were transformed, and proteins were expressed, and purified as described in Appendix 1 Expression. The elution from purification by binding the His-tagged proteins to Ni-NTA resin 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.

Designed proteins with an oligomeric peak near the expected elution volume via SEC had all fractions from the peak collected. These fractions were then pooled together and concentrated or diluted as needed to a final concentration of 2 mg/mL. 100 uL of each sample was then run through a high-performance liquid chromatography system (Agilent) using a Superdex 200 10/300 GL column. These fractionation runs were coupled to a Wyatt multi-angle light scattering detector in order to determine the

absolute molecular weights for each designed protein following the method described previously [64].

Designed proteins with a measured SEC-MALS oligomeric state with <=13% discrepancy from design were considered to have validated oligomeric state via SEC-MALS. Of the 138 designs ordered, 27 were found to be within this range by SEC-MALS, for a successful oligomerization rate of 20% of all initial designed proteins, and 48% of all proteins which had soluble expression. An additional 10 proteins had a discrepancy between 13% and 24%. This range likely indicates a mix of oligomeric states, however, there were 4 designed proteins (sg165_C5, sg191_C3, sg192_C3, and sg222_C3) that fall within this range that have SAXS curves which match very well ($X^2 \leq 3.1$), indicating that they may still be forming the correct overall structure [Figure 3.3].

Proteins found to have a near-correct oligomeric state by SEC-MALS (<25% discrepancy) were sent to SYBLYS for Small-Angle X-ray Scattering analysis at the Advanced Light Source (ALS) [86–89]. Purified protein was concentrated to 1 mg/mL and 5mg/mL and concentrator flow through was collected and sent along with samples to be screened and used for buffer subtraction to normalize the SAXS curves. Data files were averaged using SAXS Frameslice Version 1.4.13. Data was compared to design models using the FOXS web server with a fixed hydration layer of 1.0 [90,91]. Of the 37 proteins calculated by SEC-MALS to have the correct or near-correct oligomeric state, 16 were determined to have accurate SAXS curves (Chi squared of <18.1). This cutoff was determined by the highest chi squared of a design with an oligomeric state validated by

Electron Microscopy [Figure 3.4 b and e]. Designed proteins sg205_C5 and sg206_C5 were not included in the successful percentages, as they were helix cap modifications of sg151_C5. However, they also show monodisperse peaks and accurate SAXS curves [Figure 3.3]. This gives a success rate of 12% of all designed proteins ordered, 29% of all soluble expressions, and 43% of all proteins determined to be the correct or near-correct oligomeric state by SEC-MALS.

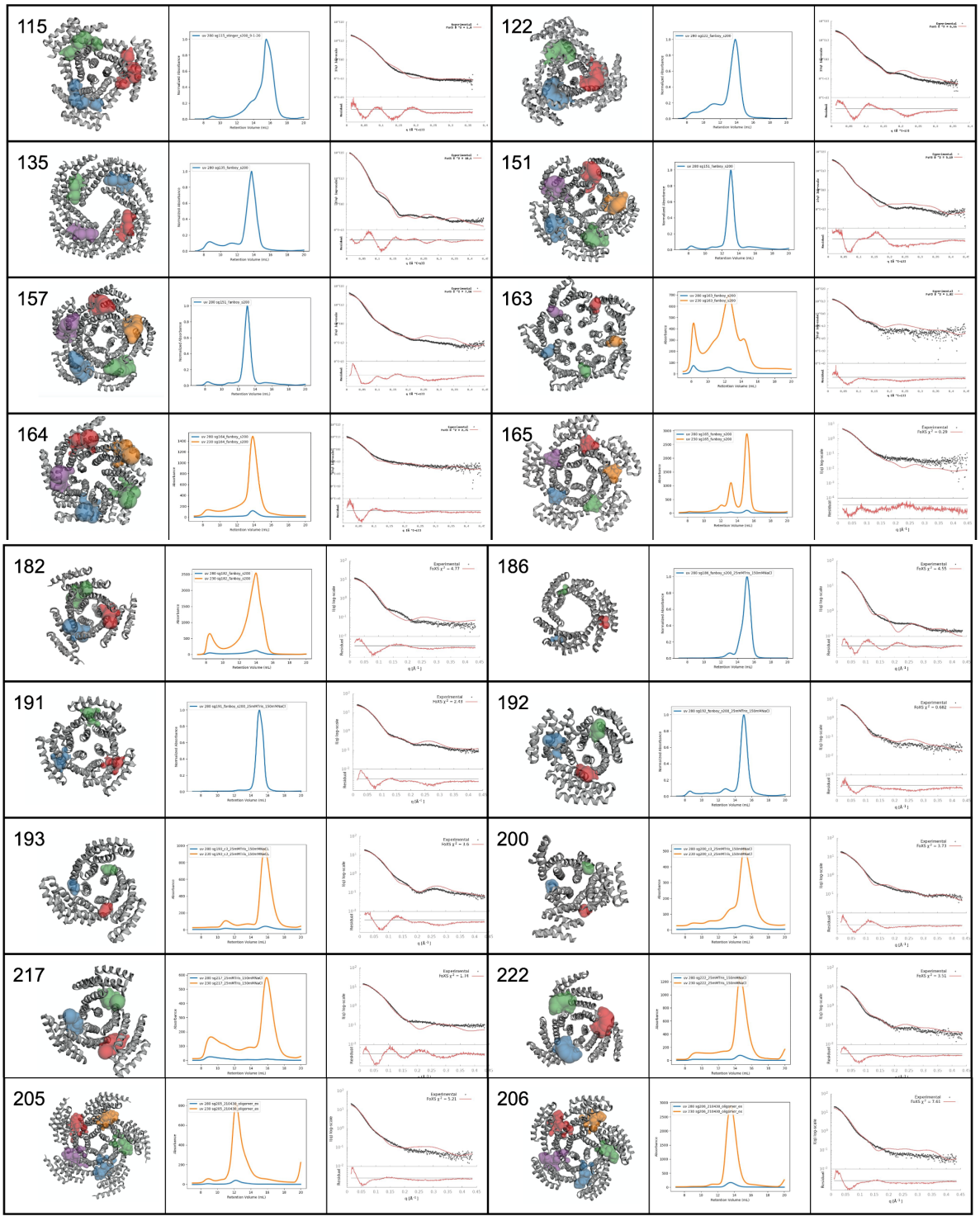

Figure 3.3: SAXS validated designed proteins. Each block shows designed model pockets calculated using CASTp web server(Tian et. al 2018).(right) Unlike the initial filtering step, these pockets were calculated using all designed amino acids, and not just the backbone positions. Each block shows the SEC curve from initial purification of design on a Superdex S200 Increase 10/300 GL column(middle), as well as the experimental saxs curve (right, black dots) and calculated saxs curve from the design (right, red line)

3.6 Characterization through Electron Microscopy(EM) and X-ray Crystallography

To further validate these designs, proteins with monodisperse SEC peaks and MALS data indicating a correct or near-correct oligomeric state were also screened through negative-stain Electron Microscopy (nsEM). The fraction containing the peak of the elution volume was collected and diluted in a range between 0.02-0.1 mg/mL. Four of my proteins had clear and dispersed particles in the uranyl formate stain used to contrast with the proteins [Figure 3.4 a-d]. We were able to collect datasets on these 4 designed proteins, and create 2D class averages for each of them to show the oligomeric state of these proteins. All 4 designed proteins appeared to have a preferred orientation along the symmetry axis. This is likely due to the overall shape of these proteins being similar to cyclical disks. The preferred orientation of the designed proteins made it difficult to get an accurate volume reconstruction of our proteins, it does clearly demonstrate the oligomeric state of 3 of the proteins, with sg122_C3 [Figure 3.4a] and sg193_C3 [Figure 3.4d] showing visible trimeric averages, and sg151_C5 showing clear pentamers [Figure 3.4d]. There was enough diversity in the views of sg122 to get a very broad 3d reconstruction (>20 Å). This density appears to show that the arms of the protein are slightly more open than in the design model [Figure 3.4j]. The 2D class averages from sg135_C4 did not show as clear an oligomeric state as sg122_c3, sg151_C5, or sg193_C3. However, the pore running through the center of the protein was clearly visible[Figure 3.4c]. In addition, sg135_C4 is large enough (101.5kDa) to progress to cryogenic Electron Microscopy(cryo-EM). We were able to get far higher resolution 2D class averages of sg135_C4 which show

multiple angles including a set directly down the symmetry axis clearly showing the ciclic 4-fold symmetry. For ease of comparison, each class average is paired with the design model in matching orientation[Figure 3.4e].

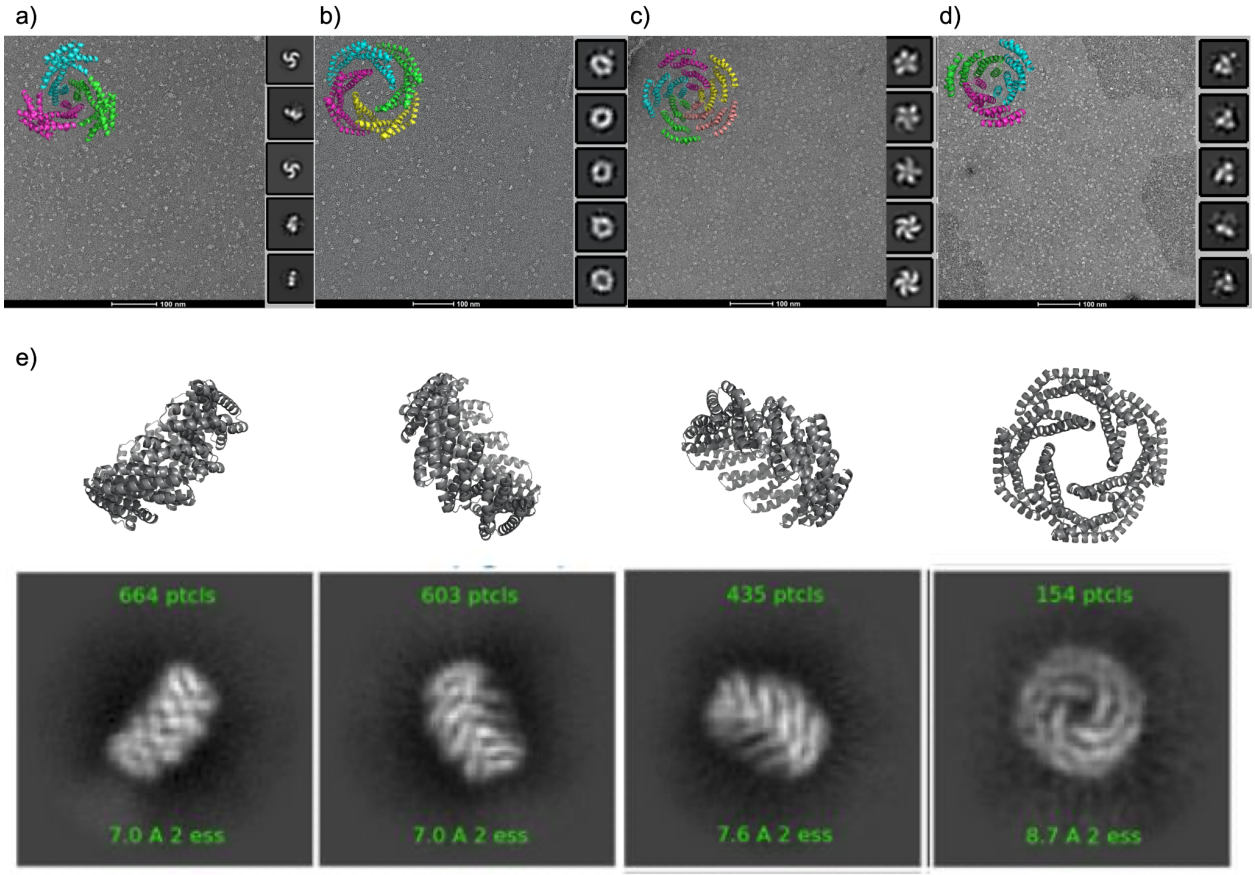

Figure 3.4: Structural validation of oligomeric state using Electron Microscopy. a-d)Negative stain images including representative micrograph with designed model in top left corner, and 2d class averages to the right of the micrograph. a) sg122_c3, b) sg135_c4, c) sg151_c5, d) sg193_c3 e) 2D class averages of sg135_c4 from cryo-EM (bottom) paired with designed model in matching orientation (top)

Crystal trays were set up for all 37 proteins which showed a correct or near-correct oligomeric state by SEC-MALS. Proteins were concentrated to a final concentration between 20 mg/ml and 63 mg/ml, depending on the solubility of proteins at high concentrations. Screening conditions were at 4 °C or 20 °C and depending on the quantity of sample, were screened with a 1:2 ratio 1:1 ratio and 2:1 ratio of concentrated

protein to buffer in Index, JCSG+, Morpheus, Classsics1, Classics2, CrystalScreen, PEG/Ion, and PGA. Crystal hits were optimized using additive screens and varying protein concentration. Data from crystals were collected on an ALS beamline. A crystal of sg122_C4 diffracted to 4.2 Å, allowing us to see the overall backbone structure of the oligomer. This structure shows that the interface is very similar to the design model [Figure 3.5b rmsd 0.84] and that there is a small extension of the curve of each monomer [Figure 3.5c, rmsd 1.7]. Due to this extension, the overall structure shows a slight opening of the arms which form the pockets in the oligomer [Figure 3.5a, rmsd 3.4]. The crystal structure was then fit into the ~20 Å negative stain density map which shows a similar widening[Figure 3.5f]. Taken together, these data indicate that the widening seen in the crystal structure is likely not a crystallization artifact.

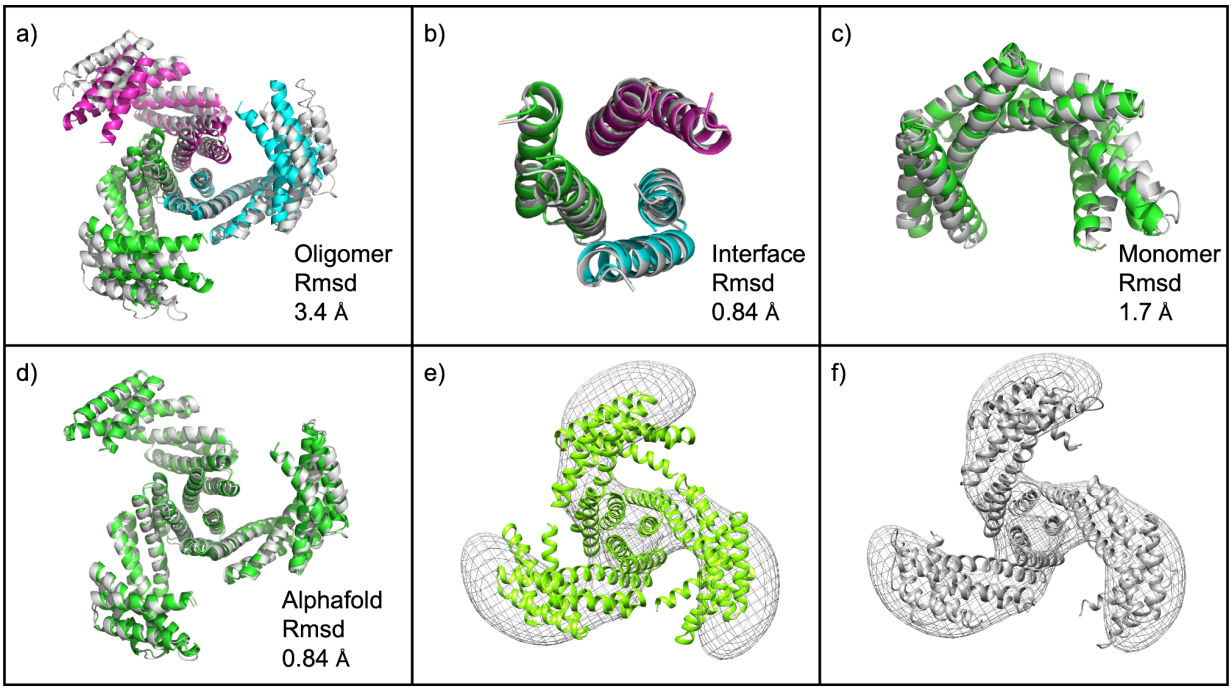

Figure 3.5: Crystal structure of sg122_C3. a) Alignment of designed model (green/cyan/magenta by chain) to crystal structure (white) b) alignment of designed interface (green/cyan/magenta by chain) to crystal structure (white) c) alignment of chain a of design (green) to chain a of crystal structure (white) d) alignment of alphafold 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)

3.7 Comparison of Designed Models to Alphafold Predictions

Alphafold has been shown to have a very high accuracy in predicting protein structures[92,93]. Given it's published accuracy, as well as the accuracy of finding the oligomer with matched crystal structure of sg122_C3 [Figure3.5 d], we decided to compare alphafold predictions to our design models. To do this, we used alphafold to generate oligomers of all of the proteins that were validated by SEC-MALS or SAXS. Alphafold is able to output up to 5 different models with different parameters used to generate them. When generating our oligomer models we output all 5 models and filtered for models that found an interface. From the subset of models that found an interface, the model with the best lddt score was selected and backbone rmsd was compared to the design model [Figure 3.6]. Seventeen of the 37 runs found oligomers. Of those oligomers generated, most found interfaces very similar to the design model, with 12 of the 17 models having less than 3 Å rmsd from the design model.

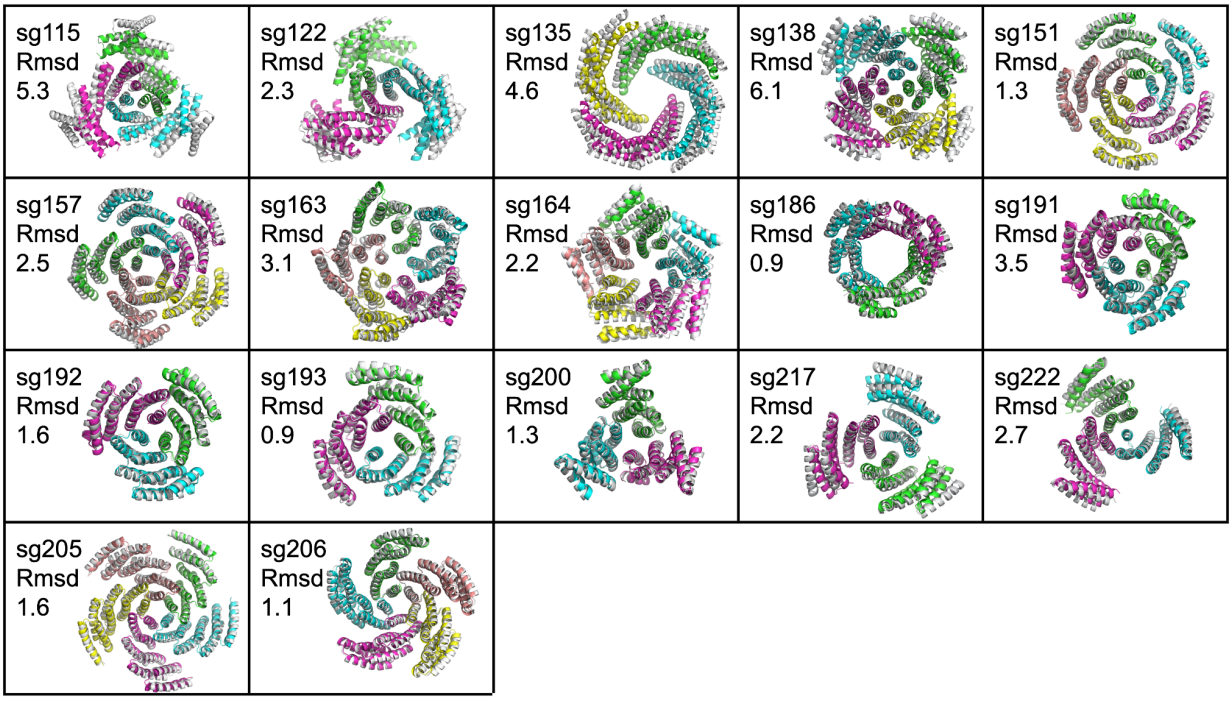

Figure 3.6: Alphafold generated oligomers aligned to designed oligomers. Alphafold oligomer prediction is shown in white in each square, the design model is shown colored by chain. Each square has the design number and rmsd between the alphafold predicted oligomer and the computationally designed oligomeric model.

We ran alphafold predictions on the asymmetric unit (chain A) of all proteins that either had accurate SAXS data or output a symmetric oligomeric state in the initial mulit-chain model predictions. These monomers were even more similar to the design model, with 15 of the 19 models having <2 Å rmsd from the design monomer [Figure 3.7]. Given the ability of alphafold to find oligomers similar to design, the alphafold prediction should be considered when designing validated scaffolds for future functionalization.

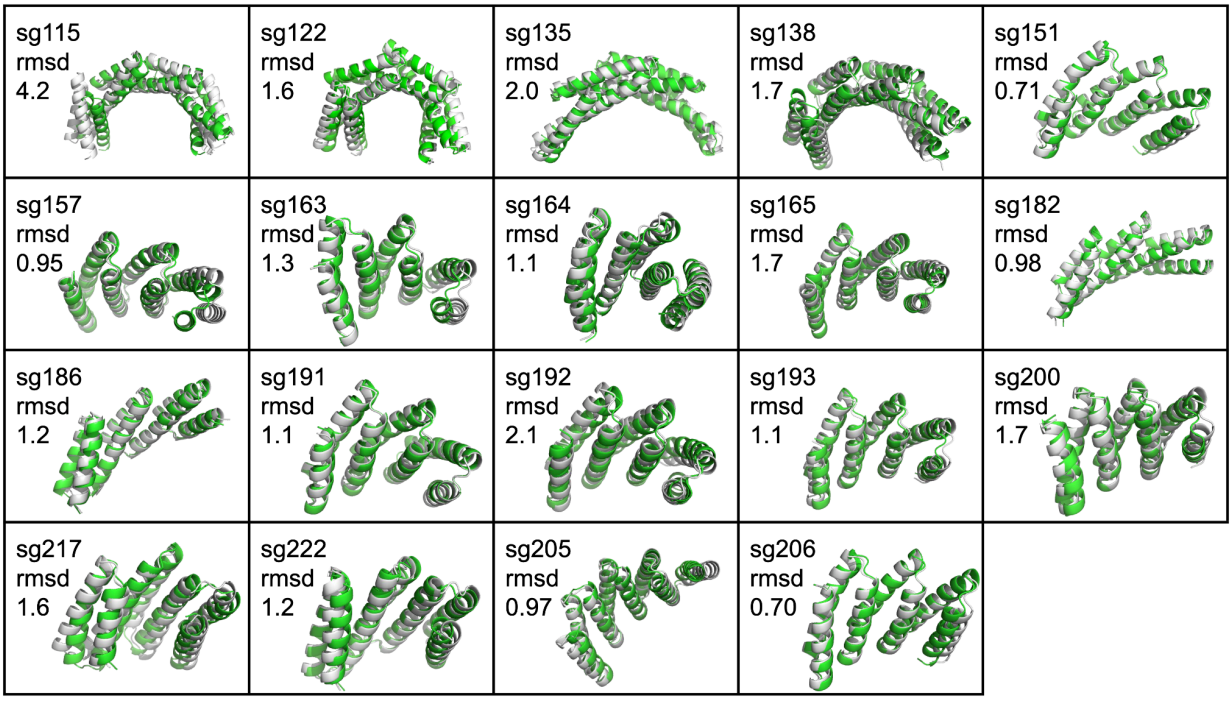

Figure 3.7: Alphafold generated monomers aligned to designed monomers. Alphafold monomer prediction is shown in white in each square, the design model is shown in green. Each square has the design number and rmsd between the alphafold predicted monomer and the computationally designed monomer model.

3.8 Discussion and Future Directions

These cyclically symmetric proteins have diverse shapes and sizes to accommodate a wide variety of ligands and substrates. This method illustrates a way to generate a large set of proteins with a variety of pockets which cover a wide area of possible shape space [Figure 3.2d]. Due to the fact that the residues forming the pocket were distinct from the residues in the core and interface of the protein, these pockets should be able to be mutated for functionality with minimal disruption of structurally important features. In addition, the location of the pockets at the asymmetric interface removed the necessity of linking the chains of the oligomer into a larger single/monomeric protein in order to begin to design functionality into the pockets.

Designed proteins were shown to be highly structurally accurate to both the monomer and interface, with a successful oligomerization rate of 20%, with 12% validated by SEC-MALS.The methods used here use only unvalidated helical repeat proteins as a base monomer for docking an oligomer design. Higher accuracy could likely be achieved by only using pre-validated monomers. In addition, an even wider variety of pocket features could be created by using a wider set of monomers as the basis, such as curved alpha-beta proteins.

Validation efforts show that designs proteins were highly accurate at the interface, with any deviations propagating out from the center along the monomeric repeats. Given the structural validation we have, functionalization efforts should prioritise binding using surface residues in the neighborhood of the interface, rather than farther out along the repeats for highest accuracy.

Appendix:

Appendix 1: Methods

Expression

Genes encoding each of the 138 designs were synthesized and cloned into a vector with a T7 promoter system and 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 50mL cultures in 250 mL flasks for initial validation in and then 0.5-liter cultures in 2-liter flasks if more protein was needed for crystallography. Proteins were expressed in Studiers M2 autoinduction media with 50 ug/mL kanamycin. Pre-cultures were grown at 37°C for ~14 hrs, and cultures were inoculated with 1mL of preculture per 100mL of media. Cultures were grown at 37°C for ~14 hours. Cells were pelleted at 4,000g for 10 minutes, after which the supernatant was discarded. Pellets were resuspended in 30 ml lysis buffer (25 mM Tris HCl pH 8, 150 mM NaCl, 30 mM imidazole, 1mM PMSF, 1 mM DNase, 10mM Lysozyme). Cell suspensions were lysed by sonication, and the lysate was clarified at 14,000g for 30 minutes. 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 flowthrough 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.

Size Exclusion Chromatography and Multi-Angle Light Scattering

Designs were with a single peak near the expected elution volume via SEC had the relevant fractions collected. These fractions were then pooled and concentrated to a 2 mg/mL. 100 uL 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 multi-angle light scattering detector in order to determine the absolute molecular weights for each designed protein following the method described in (Fallas et al. 2017)(Hicks et al.

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 1mg/mL and 5mg/mL and concentrator flow through was collected for buffer subtraction. Proteins were sent in a 96 well plate in 4-well blocks of [buffer, 1mg/mL protein, 5mg/mL protein, buffer]. Data files were averaged using SAXS Frameslice Version 1.4.13. Data was compared to design models using the FOXS web server with a fixed hydration layer of 1.0[90,91].

Negative-Stain EM Sample Preparation

Samples were diluted to a range of concentrations between 0.02-0.1 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.[94]

Negative-Stain EM data collection and processing (3 to 6-mer)

Data were collected on an Talos L120C 120kV electron microscope equipped with a CETA camera. A total of 150-300 images were collected per sample by using a random defocus range of -1.3 to -2.3 µm, with a total exposure between 4000-5000 e−/A2, and a pixel size of 1.51 Å/pixel. Data were automatically acquired using EPU (ThermoFisher Scientific). All subsequent data processing was performed using cisTEM and Relion. The parameters of the contrast transfer function (CTF) were estimated using CTFFIND4 (Mindell and Grigorieff, 2003), and particles were picked in a reference-free manner in cisTEM[95]. Particles were extracted after correcting for the effect of the CTF for each micrograph with cisTEM. [96] Particle stacks were exported and pre-processed in RELION/3.1) [97–99] with a binning factor of 4 applied, resulting in a final pixel size of 6.16 Å/pixel. Resulting particles were sorted by reference-free 2D classification over 25 iterations. 3D ab initio was performed in cisTEM the presence of C1 symmetry, with the subsequent homogenous refinement steps performed using C3 symmetry. 3D maps were visualized with rigid-body-docked models using ChimeraX (PMID: 32881101).

Cryogenic Electron Microscopy

sg135_C4 was diluted to a concentration of 1mg/mL in 25mM Tris pH 8.0 150mM NaCl using. It was frozen using a Vitrobot Mk.IV with blot times of 6.5, 7.5, and 8.5 seconds. Data was collected on a Titan Krios and processed using RELION and cryoSPARC.

Crystallography

All crystallization trials were carried out at 20°C in 96-well format using the sitting-drop method. Diffraction quality crystals appeared in 25mM Tris 150mM NaCl Index well B11 with 2.1M 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 crystal sg122_C3 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 [100] and merged/scaled using Pointless/Aimless in the CCP4 program suite [101]. Low resolution structure determination and refinement starting phases were obtained by molecular replacement using Phaser[102] using the designed model for the structures. Following molecular replacement, the models were improved using phenix.autobuild Structures were refined in Phenix [103]. Model building was performed using COOT [104]. The final model was evaluated using MolProbity (6)(Williams et al. 2018).

ACKNOWLEDGMENTS
This work was conducted at the Advanced Photon Source and Advanced Light Source. Advanced Photon Source (APS) Northeastern Collaborative Access Team beamlines,

which 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), 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.

Crystallographic data collection and refinement statistics

	SG122_C3
Data collection	
Space group	I 2 3
Cell dimensions	
a, b, c (Å)	234.09, 234.09, 234.09
α, β, γ (°)	90, 90, 90
Resolution (Å)	100 - 4.26 (4.76 - 4.26)*
Rmeas	0.143 (4.145)
I / σI	22.1 (1.2)
Completeness (%)	99.8 (99.5)
Unique reflections	15064 (4200)
Redundancy	40 (40)
CC1/2	1.000 (0.504)
Refinement	

Resolution (Å)	95.57 - 4.27
No. reflections	14976 (2957)
Rwork / Rfree	0.39/0.43
No. atoms	5619
Protein	5619
Ligand/ion	0
Water	0
B-factors (Å2)	
Protein	220
Ramachandran Plot (%)	
Favoured/Allowed Outliers	93.48/ 5.65 0.87
Bond lengths (Å)	0.005
Bond angles (°)	1.158

*Data collected from a single crystal. *Values in parentheses are for the highest-resolution shell.

Appendix 2: Design Sequences

Modified from pre-validated scaffold

>allHID_3_4_21_bp_42
SKDTEDSRKIWEDIRRLLEEARKNSEEIWRLIKESPEEENHELLSEQLLELAEMLVRIAELLSRQT
EQVAESIRRDVSALAYVMLGLLLSLLNRLSLAAEAYKKAIELDPNDALAWLLLGSVLEKLKRLDEA
AEAYKKAIELKPNDASAWKELGKVLEKLGRLDEAAEAYKKAIELDPEDAEAWKELGKVLEKLGR
LDEAAEAYKKAIELDPND
>allHID_9_3_7_bp_46
SKDTEDSRKIWEDIRRLLEEARKNSEEIWRESSKCENQSQQHKLLSEQLLELAEMLVRIAELLSR
QTEQVAESIRRDVSALAYVMLGLLLSLLNRLSLAAEAYKKAIELDPNDALAWLLLGSVLEKLKRLD
EAAEAYKKAIELKPNDASAWKELGKVLEKLGRLDEAAEAYKKAIELDPEDAEAWKELGKVLEKL
GRLDEAAEAYKKAIELDPND
>allHID_4_7_20_bp_49
SKDTEDSRKIWEDIRRLLEEARKNSEEIWKQQEEDPTNVQKNHRRLSEQLLELAEMLVRIAELL
SRQTEQVAESIRRDVSALAYVMLGLLLSLLNRLSLAAEAYKKAIELDPNDALAWLLLGSVLEKLK
RLDEAAEAYKKAIELKPNDASAWKELGKVLEKLGRLDEAAEAYKKAIELDPEDAEAWKELGKVL
EKLGRLDEAAEAYKKAIELDPND

>allHID_10_5_25_bp_51
SKDTEDSRKIWEDIRRLLEEARKNSEEIWKRIEEESKRGSSPQNSHRLLSEQLLELAEMLVRIAE
LLSRQTEQVAESIRRDVSALAYVMLGLLLSLLNRLSLAAEAYKKAIELDPNDALAWLLLGSVLEKL
KRLDEAAEAYKKAIELKPNDASAWKELGKVLEKLGRLDEAAEAYKKAIELDPEDAEAWKELGKV
LEKLGRLDEAAEAYKKAIELDPND
>allHID_12_4_21_bp_47
SKDTEDSRKIWEDIRRLLEEARKNSEEIWRLSEESQKKVPESRQHNSVSKLLSEQLLELAEMLV
RIAELLSRQTEQVAESIRRDVSALAYVMLGLLLSLLNRLSLAAEAYKKAIELDPNDALAWLLLGSV
LEKLKRLDEAAEAYKKAIELKPNDASAWKELGKVLEKLGRLDEAAEAYKKAIELDPEDAEAWKEL
GKVLEKLGRLDEAAEAYKKAIELDPND
>allHID_13_2_23_bp_48
SKDTEDSRKIWEDIRRLLEEARKNSEEIWKLIELSTNKTYSHKLLSEQLLELAEMLVRIAELLSRQ
TEQVAESIRRDVSALAYVMLGLLLSLLNRLSLAAEAYKKAIELDPNDALAWLLLGSVLEKLKRLDE
AAEAYKKAIELKPNDASAWKELGKVLEKLGRLDEAAEAYKKAIELDPEDAEAWKELGKVLEKLG
RLDEAAEAYKKAIELDPND

Newly generated scaffolds for metal binding

>sg21_quick_pack_20190311124955_0001
GEELSRVAKRLSDLALFILLQALQLNDEEVAELVDWVVHLVQVMLKLANQGDEDDVNRVWQLV
KRLTKEVRQLLKRWKKI
>sg22_quick_pack_20190311133938_0001
SSLDKLVDAVKLVVDYVKERVRELGSEEALLRLWIVVNLVLHVALDLLKSSDEDVSQKARRVMQ
KAQELLIELLKRILNT
>sg23_quick_pack_20190313184241_0001_mod
GKELVRYLILVLKVEVISIKQVIQNNDWQSALLWLAHNLLVLLLAKELDPDDQQVRKLSEESRELL
DIAKQLMNDG
>sg24_quick_pack_20190313183900_0001
SSVKELVERILEQAKRWSDKLRKLGDLSDIASMVHRVLLLLVQVLILWLQNPDDEDVREALKLAQ
EAFELLQKVIKDT
>sg25_quick_pack_20190311125444_0001
TDEVSQIIRLEAQMVDLVLRALEEVGDFLLLLVIVLHQSLQVDILLRNTDSREVRKKAKEWLEKIK
KILRQVQRLLDQT
>sg26_quick_pack_20190313183822_0001
SDDYSRKASRLVRELKELVQRFKESPSAEIWIQVHLLLASVVALWLSMKENGDEDASKLISRLLN
DAVKLEQRVQQVLVKL
>sg27_quick_pack_20190312182231_0001_dislike
GDEANRVLRIVKLVQDNVQILKELGNKSMLAILAIHESQLVKFLSQQVKQVLLRNPDSEVKNVLR
NVNRLLEILNELLRRAQKYV
>sg28_quick_pack_20190313192840_0001
GDDVSRRAKRLARKLKELVQQVKESPSISLWVLVHMLLILVVRLWFDLSQNGDEEASELVQRWL
RQAVKLEQQVEKILVSV
>sg29_quick_pack_20190311144804_0001

DELLKQVLKWAAIVDFLANNLVRLNNNLSWQMEWHLRVAKWVFSADQFLRLGDFEELRKVIQK
VRELSEELKRLLRKVQQG
>sg30_quick_pack_20190311154549_0001
DRESAKLVRLVVRAVSVIVKEMSKSNNEDQVYQVLFHAILALQIVNENGDEDVSRDVKKWVKEI
RQWARDVSNT
>sg31_quick_pack_20190311190916_0001
GKQQAKILTQVANAIVVQVSELSKSNNENSVLEVFLHVFLAWQLVQEMGDEDLREIVKRVLDLV
RQWLKFVKRR
>sg32_quick_pack_20190312181620_0001
SKEVQQLQQLVRQLVEKVRKLASKLGDWRLSVLAIHVFQAMQRLLQSANQSGNEEESSQWLK
QVTQEVQWILRVVRL
>sg33_quick_pack_20190312184325_0001
DLLRLVLRWILTLVRVLIQNFRDSGDERFALKVLSVLDNLLHLVELLGDEKLKELAKELLELVAKVA
APMLSRLALK
>sg34_quick_pack_20190312184634_0001_0005
TEVVLQALLTVTLLVALVLILVEQSGDWDMAREVNDLQHKVSGQLVQSTDKEELKQLSQEANRL
VKRLLRDVRK
>sg35_quick_pack_20190313185717_0001
DDLSKDVSIESQVVKELWRAAKDAGDESVLLVALVLILHFLLVIKRLSDEMGSERVRKLVQRLSKE
VKEFSDTLRQLVQRL
>sg36_quick_pack_20190313190305_0001
SDELSRKASRLMRQLRDLVQQVNENVDLKLVVFVHMLLALIIRLLWEAKQNGDEEATELIRRVVN
KAVKLEQLVLKLSVKV
>sg37_X_h24_l2_h10_t2_D37C_t3_h4_l5_h24_16_0001_0003_0001_6.56_0001_0010_6.52_
mod
SSAQRVSNIDKKVAEIISEAVKLANDESISQIAEIDVLHLEANMESETQSGNLDKKLSKQIKEQLRK
IEREIEKLSRK
>sg38_X_h24_l2_h10_t2_D42C_t3_h5_l5_h28_1_0001_0002_0001_5.54_0001_0001_6.54_m
od
SSWIEEQRRLIQLVAELTAKSISSGTEESARQLQSLILEAHQKSEVQSLEGPDLDKDKLAKSDTKL
MEKSVKKIEKTVKENEK
>sg39_X_h24_l2_h4_t2_D21C_t2_h11_l3_h28_18_0001_0004_0001_4.66_0001_0002_6.6_m
od
SKTENKIAENSAKSDYNQEEASESGNIEQKISVHITQNLILRKVVEKTKEEIPNEDETEKELERQV
QRIAERQSERSKKRQK
>sg40_X_h24_l2_h4_t2_D21C_t2_h11_l4_h24_18_0001_0002_0001_5.6_0001_0015_6.77_00
04_6.68_6.68
DKYVNNVSNNARRLTRQQNELSKSGNESNTIIIHFQLNIQISEYAYQEASRQTKDEENLENINNSI
SINRLASRVSNSG
>sg41_X_h24_l2_h4_t2_D21C_t3_h9_l3_h24_0_0001_0005_0001_4.73_0001_0008_6.73_000
4_6.38
SEVLQNIISNASNLYEDNQSLRNSGNESQWEIIHFQANLRISLLVANQSGNESVSRIVSLQSKIAS
SISKNSKKE

>sg42_X_h24_l2_h4_t2_D21C_t3_h9_l3_h24_0_0001_0005_0001_4.73_0001_0016_6.77_0002_6.49
SSVAKEIIKRAKQAYRNEEKLSNSGNEVSLIITHIQLNLQLNSLIASESGNESELNSNSRLEKIVSEIRNSSEKI
>sg43_X_h24_l2_h4_t2_D21C_t3_h9_l3_h24_0_0001_0005_0001_4.73_0001_0038_6.75_0002_6.46_mod
SSVLSQIASNASNYYEKEEELRRSGNEIQYIITHIQANLSINQLIAKQSGNEKSYELYELNQEVVSKISKYTSKR
>sg44_X_h24_l2_h4_t2_D35C_t2_h11_l5_h24_6_0001_0001_5.18_0001_0014_6.56_mod
SRSEEWKRTLERLRKELKKKIEKTNNEDFENKNRHNEESLKLASLNVNSDEGPENEDLDKKLSISSAQLASNNASNVSR
>sg45_X_h24_l2_h5_t3_D21C_t2_h9_l5_h24_5_0001_0003_0001_4.9_0001_0011_6.81_mod
DEESELARERSKESSKEAEKIRRLGNLSQLLQWHARVSRIIKKIEDSTELGLVDERASEEVNKEQESDNSQAMRLIIS
>sg46_X_h24_l2_h5_t3_D35C_t2_h9_l4_h28_0_0001_0003_0001_4.96_0001_0035_6.8
DEEKESEKLSEEVIRRSKSSSNKSGNLSEQAQSVHNTLQVEIIEKRQKRGEQDERDSIESLQKQLSAMKLQANAVNSNKNK
>sg47_X_h24_l2_h5_t3_D35C_t2_h9_l4_h28_0_0001_0004_0001_5.39_0001_0015_6.8
DESKQLEKLVNRTIKQSEELTEKLGNEQAKANSKHDTLSSNQVYKNEREGRQDERDAIESQEVQLQVVKNSMRALEKIEKR
>sg48_X_h24_l2_h7_t2_D43C_t1_h6_l3_h28_15_0001_0001_0001_5.6_0001_0079_6.56
TDEVKESIKDEKEIEKLIEYAKRQGNEVAISLIQLSSNHSYQSVQKLAEESNDEDSLKKSKKRQKKIRDLAEKIRKELEER
>sg49_X_h24_l2_h7_t2_D47C_t1_h7_l2_h24_9_0001_0001_5.34_0001_0011_6.56_mod
SDEARQASNSLKNISKSEDQISKEGNKEALTASQVVNQHSKIQLEILIEEAGSEKDKRELEKQMNESKKRAEENSRR
>sg50_X_h24_l2_h8_t4_D32C_t1_h4_l2_h28_0_0001_0012_0001_5.23_0001_0005_6.54
SKESSRNSKKENKSIRTAEDILRQGNLEDILRILVELHNLSSNAENSGSEKTAKKSQEVETKISEKSSEWLKEQ
>sg51_X_h24_l2_h8_t4_D32C_t1_h7_l3_h28_5_0001_0005_0001_5.09_0001_0074_6.56_0003_6.83
ESEIRELERQANSLTKNLKKVSKSGNIQEVLEILLNIHNINNNISSYANNQRDESAKNSLKKIENSEKQAVEITAQLIQKE
>sg52_X_h24_l2_h8_t4_D32C_t1_h7_l3_h28_5_0001_0010_0001_6.78_0001_0014_6.87_0002_6.45_mod
SSEVSNTYKRASSITKNIKKVQNSGDINSAVSIWTQLHNLLNQIEELYNEQKDESASRYLKKLSSYQEKVANILASLLEEQ
>sg53_X_h24_l2_h8_t4_D32C_t1_h7_l3_h28_5_0001_0011_0001_10.16_0001_0029_6.76_mod
REDASRLSKQQSEVIRDIKKITEQGNISNIIELLNNSHEISNQQESIAEKQNDEEVSKSVNKANKQQRSILKISADLNDRR
>sg54_X_h24_l3_h10_t2_D47C_t4_h4_l4_h24_9_0001_0006_0001_9.65_0001_0013_6.43_2hNNQSSLAVQIAIKNVLLHISWSIQESDNEKQVKNQRQSDSDMQSNVDKVQKQ

>sg55_X_h24_l3_h5_t3_D21C_t2_h8_l5_h24_3_0001_0001_6.77_0001_0013_6.71_0003_6.74
GSRVSNTVISYATKLQSEANELSKQGDEKISSEIHNQLNSIYELAEKLSRSTSNEEEIIKRISNSVNESSNRMKNNSK
>sg56_X_h24_l3_h5_t3_D21C_t2_h8_l5_h24_3_0001_0001_6.77_0001_0013_6.71_0005_6.71
GSSIEKKAAEIAKQIQSRQNELSNQGNERIQIELHVILNQIIEEAQNLSRSTTDESSVNKNLKNSSKKLEQSVENQSN
>sg57_X_h24_l3_h5_t3_D35C_t2_h10_l2_h28_1_0001_0001_0001_6.6_0001_0011_6.55_mod
SEDVIKQLDKLKELAEKAIEKAERNNQTSLNMLLRHSQSSIENVSREVKKSNDEKSSKEANSQYKIESSILRLIEKAIENG
>sg58_X_h24_l3_h5_t3_D35C_t3_h10_l4_h28_3_0001_0001_5.37_0001_0024_6.81_mod
GDEISKLSKSVKRKIEEAREEIKRSRDETLKANLLHSELSIERSEIQLNENPESESLTQIQSEQSKLESQIIEAIKKSLKKG
>sg59_X_h24_l3_h5_t3_D21C_t2_h8_l5_h24_3_0001_0001_6.77_0001_0013_6.71_0003_6.74_2
GSRVSNTVISYLVKLQSEANELSKQGDEKISIEIHLQLNSIYELAEKLSRSTSNEEEIIKRISNSVNESSNRMKNNSK
>sg60_2_quick_pack_20190516163419_0001_mod
DEARKLIEKYEKLIEKIARLVLRANDEKLWELMLRLHVRIALTSLELSARGIDPRQVAEQITKLVEQLSDQIERSVN
>sg61_2_quick_pack_20190516164620_0001
SEASENIQKLEKEINQLEEKLQKTGNIASSVAILVASHMIAIKLQIEIQKLNNESSRREIEQWDSRIVELVSKLASR
>sg62_2_quick_pack_20190516170056_0001_mod
SDWARDILRAAEKEAQLLEIVVQLGNEQANSILLQLHAEQVLYQMKLESSRLQGKIDEKDLQSIRSQILEKIVQKAEKISNQVQS
>sg63_X_h24_l2_h5_t2_D48C_t1_h7_l4_h24_2_0001_0009_0001_mod
SRENQLSAQLAEINIEIVELVVQNGNPQWSLLLLKSHARYVAIQLRLQANGLSEQNLEQILQQLVQEADQLLQQVES
>sg64_X_h24_l2_h5_t2_D48C_t1_h10_l5_h28_2_0001_0004_0001
SSYIQLIAQWAQLQVQIIQQVVQLGSPQAISYILIAHAESTAIDVQLQVQIQQGNIDQQQLQQIRVKVEQKIVKQLEKIMDKVES
>sg65_X_h24_l2_h5_t2_D43C_t1_h8_l4_h24_0_0001_0001_mod
SQELEQVAQLEQRNAQLLLQIIQLGDPQSLAQALLAHFNVIVSQLRWQSVGVDPQSLLQNLIQILQQILQQIEQQVF
>sg66_X_h24_l2_h5_t1_D30C_t3_h10_l3_h24_10_0001_0001_0001_mod
DRESELIEEQVQVSNELLKILDRYGNETAVLIIQIHSLVQQFARVYIQVRQLNGEQDIEQWQEQAVQLISQIAQQTEL
>sg67_X_h24_l2_h5_t1_D40C_t4_h8_l2_h28_5_0001_0003_0001_mod
SELQQQLQQIQQEAEQWVRQANQQLDQDKIILLQHAKVIILQNLAQDVGSEQISQKLRQENLQVAKQILEIIVSR
>sg68_X_h24_l2_h5_t1_D40C_t4_h8_l2_h28_5_0001_0006_0001

TQEAQELQQIQQRVQQATQQASQNYDESQTILIDHLSVAILSDLSQVVGSEQISQRLTQQNVQK
AEQALKRWGSS
>sg69_X_h24_l2_h5_t2_D43C_t1_h10_l3_h24_4_0001_0009_0001_mod
DEVLQIQEILIQLNKQLVQIVSQLGNPQDIAQVLEDHIDAAQAILQLQKAQRLGNQQTKQQAQQL
LQQIQQRMQQLSNR
>sg70_X_h24_l2_h5_t1_D47C_t1_h10_l2_h28_2_0001_0002_0001
SLVDQWTQFLEQTSQEIEQRAQQMGNSSIRIAALLLNHLSLSAKQISKQLKQSGNQDLEQLAQI
DAQIVQKIKQNVQEVEQIG
>sg71_X_h24_l2_h5_t1_D47C_t1_h9_l3_h24_3_0001_0002_0001
SQIQQLNQQLEQILNQIQQIAEQTGSLSSLVAVILILHNRISLQIISQRYQEQNNPRESQIAQEDVQ
DLVQLLQKAVQR
>sg72_X_h24_l2_h5_t1_D47C_t4_h10_l3_h24_3_0001_0001_0001_mod
SQARQQLQQLEQLIEQKIQQWKQSGNESSQIANLVALHAIAIYLQIQIKRQGQQQSQQEITELDL
RIVQKISQEVND
>sg73_X_h24_l2_h5_t1_D47C_t4_h10_l3_h24_3_0001_0002_0001
SQAQQQLQQWEQEIEQIIQQLRQSGNQSSTVANLIALHMVAIDLQLQIKRNGNQQSEETINELNI
QVVRKITQEASS
>sg74_X_h24_l2_h5_t1_D47C_t4_h10_l3_h24_3_0001_0003_0001_2h_zero
GNQSSTIASLIIEHAQIIYQIIEASELGDQQSQRQLIKINIQIVQRLQQQAQQ
>sg75_X_h24_l2_h5_t1_D47C_t4_h10_l3_h24_3_0001_0003_0001
SELSQNVQQWEQEIQQNEQQYQQSGNQSATIAIQIALHAIIIYIEIQAKRLNRQQSQQQVLEINLR
IVQQIVQKARQ
>sg76_X_h24_l2_h5_t1_D48C_t3_h9_l3_h28_2_0001_0001_mod
SQIEQQAQQLEQQAQQKIQQIEQTGSEKQVSILLLELHLEVIELTYQARQNSNPRAQEWAQEVSI
KLVQQIRRILQKIIQS
>sg77_X_h24_l2_h5_t2_D21C_t1_h9_l3_h24_8_0001_0003_0001_mod
SQLQQAAQELEQLAQQLEQLVQQLNDNDSRKIVQHLQQAAILIEVVLLMKQSGNNQQRQQILQ
QAQQKLQQITQLEKD
>sg78_X_h24_l2_h5_t2_D21C_t1_h9_l3_h24_8_0001_0003_0001
RQQAQEIQQLAQELEQLVQQLNDNDSQKIVQHLQQAAILSEIVLELKQSGNNQQRQQIKQQAQ
QKLQQITQLEKS
>sg79_X_h24_l2_h5_t2_D21C_t1_h9_l3_h24_8_0001_0004_0001
KQNAQQLKQVVQILKQLVQKRNSEDSQKILQHAQQAAIIQQAILEVTQSGNEQQQQQWLEQVE
QELQQISQEQKS
>sg80_X_h24_l2_h5_t2_D43C_t1_h10_l3_h24_4_0001_0001_0001_mod
DQILQIQEQDIQLLEQLIQRAAQLGNEQTIAQVLQLHVRSIQVLLNIQKAQKNGNQDIQQQAQQQI
QQVLQDIQQNINQ
>sg81_X_h24_l2_h5_t2_D43C_t1_h10_l3_h24_4_0001_0006_0001_mod
DQNERWTEIIVQLNEQAVQRARQNGNQQLVVELLQNHIQLAKAKLQYLRAQKNGDNQTQQQA
QQEAEQINQQIQQFSNS

Cyclic homo-oligomeric scaffolds: no surface redesign

>sg83_top21_21__X_h16_l2_h26_l3_05678loop_3_C30001_A
EVLERITAIYLALLALQFAKQAGDPEIYETVQELLKRVSRAKSEEEIKELSLRILAILLALRALQYAK
QAGDPEIYETVQELLKRVSRAKSEEEIKELSLRIKAILLALRALQFAKQLGDPQIFQTVVELLKRVS
RAKSEEEIRELSIEIQLIYLYLRVT
>sg84_top43_43__X_h16_l2_h26_l3_05678loop_3_C30001_A
EEQYRYYAIVQALRALQRAKQAGDPEIYETVQELLKRVSRAKSEEEIKELILRIKAILLALRALQFA
KQAGDPEIYETVQELLKRVSRAKSEEEIKELSLRILAILLALRALQLAKQKGDPEIVYTVQELLKRV
SRAKSEEEIFRLLLLIAIIYLLLLLL
>sg85_top61_61__X_h17_l2_h29_l4_041062_loop_92_C30001_A
WEQLLALLLVFLEELQQNLRRTGDETYQELIELVKEAIKQGDEELMRFVTTLAELAKFLEELQQN
LRRTGDETYQELIELVKEAIKQGDEELMRFVTLLAELAKFLEELQQNLRRTGDLTYLRLIQLVKEAI
KQGDRVLMVYVIAQAAVAKTQERVQ
>sg86_top03_03__X_h16_l3_h17_l2_13278loop_9_C40001_A
PEEFKTLVQLLFYLVLAGDEEAQQKALEIIQEAEKQFRRTGDPEYFKTLVVLLLVLVVAGDEEAQQ
KALEIIQEAEKQFRRTGDPEWFKTLVQLLVVLVLAGDEEAQQKAYRIVSVALVVYITTGDPEWFQ
TLVLLLQALQL
>sg87_top14_14__X_h17_l2_h29_l4_041062_loop_92_C40001_A
LVEYLLELIKFLEELQQNLRRTGDETYQELIELVKEAIKQGDVELMRFVTTLAELAKFLEELQQNL
RRTGDDTYQELIELVKEAIKQGDEELMQFVIILAILAVQLELLQQNLRRIGDRSYYDLIRFVKIAIKR
GDERQMLIVFVVAVLVVLFERLL
>sg88_top21_21__X_h17_l2_h29_l4_041062_loop_92_C40001_A
DVTLLLELQKFLEELQQNLRRTGDETYQELIELVKEAIKQGDRRLMILVTTLAELAKFLEELQQNL
RRTGDETYQELIELVKEAIKQGDLRLMVLVQLLARLAKFLEELQQNLRRTGDVYYQDLIYLVKVAI
FLGDFSVMVVVLILASQAVQRERLQ
>sg89_top23_23__X_h17_l2_h29_l4_041062_loop_92_C4A
DETYIRLEYLVFLALNQGDEELMQFVNLLYLLAILLLLLIRNLLITGDETYQELIELVKEAIQVGDVVL
MRYVTTLAELAQFLEELQRNLRRTGDETYQELIELVKEAIKQGDEELMRLVTTLAELAQLLVQLQ
VYLRLVGDETYQELIELVKEAIKQG
>sg90_top35_35__X_h17_l2_h29_l4_041062_loop_92_C4A
AYRDALARLLELLQQNLRRTGDETYQELIELVKEAIKQGDLELMVIVYTLAVLAVFLEELQQNLRR
IGDEIYQELIELVKEAIKQGDEELMRFVTTLAELAKFLYELVQNLRRTGDQTYVRLVFHVYIAILVG
DEELMRLVTTLAILAKLIEIQL
>sg92_top02_02__X_h16_l4_h27_l2_080062_loop_79_C30001_A
REVQIVIAVLLVLLILIYARQAGDPEAWQTAVSLLVLILQARGEEDIRRLLLIAQAVLLVMLILYYAKEA
GDPEAYETARELLQKIQQARGEEDIRRLLLIAFAVLLVMLILYLAKVAGDPEAYETARELLQKIQQA
RGEEDIRRLLLIARVVLLVMTLLNQNK
>sg93_top37_37__X_h24_l4_h16_l4_033492_loop_12_C3A
AEQAAILALIFLSRVGYSPVQVLTSAKELAEIVQKLQEQLRRLGMSEKALQAAILAAMFLLVQGYS
PREVFKSAKELAEIVQKLQEQLRRLGMSEKALQAAIQAAMWLLVNGYSPEEVFKSAKELAEIVQ
KLQEQLRRLGFSERALQRAINAAMVNLVNG
>sg94_top42_42__X_h20_l2_h26_l2_076702_loop_31_C3A

LTDLILYVILAYLMLIEVKKAGDEEQYRLVKELFLLAVSALFQGDLETLLLLYYVLLALEMLIEVKKA
GDEEQFRLVKELFERALRALQEGDLETLQLLEYVLLALFMLILVKKRGDEEQFRVVKELFERALR
ALQEGDLETLQLLLNVLTALANIVIVK
>sg95_top43_43__X_h16_l2_h26_l3_05678loop_3_C30001_A
EEQYRYYAIVQALRALQRAKQAGDPEIYETVQELLKRVSRAKSEEEIKELILRIKAILLALRALQFA
KQAGDPEIYETVQELLKRVSRAKSEEEIKELSLRILAILLALRALQLAKQKGDPEIVYTVQELLKRV
SRAKSEEEIFRLLLLIAIIYLLLLLL
>sg96_top52_52__X_h17_l2_h29_l4_041062_loop_92_C3A
WVITLLLLLIFLQHLQQNLRRTGDETYYILIVLVYVAILQGDEELMRRVTTLAELAKFLEELQQNLR
RTGDETYQELIELVKEAIKQGDEELMRFVTTLAELAKFLEILQVQLRRTGDETYQELIELVKEAIKQ
GDEELMRLVTTLAYLAVLLYARQ
>sg97_top67_67__X_h20_l2_h26_l2_076702_loop_31_C30001_A
LQNLQYYVLLALVMQVLVQLLGDEEQFQLVVRLFLLAQIALREGDLETLQLLEYVLLALYMLIEVK
RAGDEEQFRLVKELFERALRALQEGDVETLQLLEYVLLALEMLIVVKLAGDEEQFRVVKELFER
ALRALQEGDLETLQLLEVVLLALLRVVLVK
>sg98_top75_75__X_h16_l2_h26_l3_05678loop_3_C3A
ELILRILAILLAYRALVFAQQAGDPEIVQTVSVLYVRVVRAKSEEEIKELILRILAILLALRALQYAKQ
AGDPEIYETVQELLKRVSRAKSEEEIKELSLRIKAILFALRALQRAKQRGDPEIYETVQELLKRVS
RAKSEEEIKELLQRIVEILVQLRSL
>sg99_top82_82__X_h16_l3_h26_l3_016702_loop_68_C3A
FELQSRLVLILAALILVMAKELGDEELTENARQILRDFKNARSEEELEFIFTSAQVLSLAALILIMAK
ELGDEELTENARQIVRDFKNARSEEELEFIFTSAAVLSLAAIILILAKELGDQLLTQLALKIVQDFKN
ARSEEELEVIFTSAIYLVIAAINLRTR
>sg100_top89_89__X_h17_l2_h29_l4_041062_loop_92_C3A
WIETLVVLQRLLEELQQNLRRTGDETYQELIELVKIAIRVGDEQLMRYVITLAILAKFLELLQQNLR
RTGDETYQELIELVKEAIKQGDEELMRLVTTLAQLAVFLVLLQQNLRRTGDETYQELIELVKEAIK
QGDEELMRLVLTLAQRQLYLYVYQ
>sg101_top00_00__X_h17_l3_h20_l3_170362_loop_81_C3ABC_000000008_2_orient_0001_A
LLEVVALAVWGVWAAIVGDVLALLWLFALAQALLEIAKKRGDPQVAVAALWILVILAAKTGDEEKL
RELFRQAQELLEIAKKRGDPQVAVMALANLVLLAAKTGDEEKLRELFRQAQELLEIAKKRGDPAV
ALGALWALIVLLKK
>sg102_top28_28__X_h17_l3_h20_l3_170362_loop_32_C3ABC_000000012_orient_0001_A
PEVEAALLLLWALWAALLGIEEALVALFHQAQVILRKAETPAVMAIALAVLAWLVAKLGDEEKLRQI
FEQAQEILRKAETPAVMVLALLILALLVAKLGDEEKLRQIYEQAQEILRKAETPAVQQWARAILELL
VKKLGD
>sg103_top29_29__X_h16_l3_h17_l2_13278loop_9_C3ABC_000000008_2_orient_0001_A
PEEFKTVVQILSVLVLLGDEEAQQKALEIIQEAEKQFRRTGDPEYFKTLVQLLSILVLAGDEEAQQ
KALEIAWEAIRHWLRTGDPEWWKTAVQLLSILVLAGDEEAQRMAIVVAAAALIYFYVTGDPEYLK
ALLQLWFIWAL
>sg104_top39_39__X_h14_l4_h24_l4_029792_loop_8_C3ABC_000000019_2_orient_0001_A
IELLVAIDAWVTFIRTARKDPEKASEKVKEIIKRAGVSEEIMEYVIRVAVAWLAAATFARTARKDPEK
ASEKLKEIIKRAGVSEEDMEAVITAAVAALAAATFIRTARKDWWKAAVKVWEIIKRAGVSEEDVW
RAALLAIWEAAAQ

>sg105_top55_55__X_h17_l3_h20_l3_170362_loop_109_C3ABC_000000011_2_orient_0001_A
MEELVTALAVLVAVAIALGDEEKLRQLFEQIQEAARKAKERGNMEILVAALLLLAALAIALGDEEKMRQLFEQAQEAARKAAERGNAEIAVWALIALVWLAAWLGDEEKLRQLFEQIFMAVAWAAWLGNFEIWVAAVLMALQLALS
>sg106_top62_62__X_h17_l3_h20_l3_170362_loop_32_C3ABC_000000011_2_orient_0001_A
PEVEAFWIALAAVLWAVAGDELALQAAFVVAQEILRKAETPRVMVVALVALAFLVAKLGDEEKLRQIFEQAQEILRKAETPEIMALALLILAALVAALGDEEKLRQIYEQAQEILRKAETPNVQQWALAILIALVKKLGD
>sg107_top65_65__X_h21_l2_h17_l3_01483loop_17_C3ABC_000000009_2_orient_0001_A
EVEMELISTVLESAQKQGDPEKLRQALELLKRLSEEAQKNGNLEVLMELISTLLESAQKQGDPEKLRQALELLWWALIEAFLNGNLEVMMELISTMLESAQKQGDPEKAWWALLLLLLAAMLAQANGNVEVLLELASTFKEWLRKL
>sg108_top75_75__X_h17_l3_h20_l3_170362_loop_114_C3ABC_000000000_2_orient_0001_A
NEILALAIWAIMVALYGAEHLLVWIFEAAQETIEQARQLGNLEVLIMALIVLTLLVLALGDEEKLRKIFEQAQETIEQARQRGNLIVLIWALWLLMALVSKLGDEEKLRKIYEQAQETIEQARQRGNLRVLQMALEILVQLVSK
>sg109_top83_83__X_h17_l3_h20_l3_170362_loop_81_C3ABC_000000007_2_orient_0001_A
QEEVLALWILMLLAAATGDEEKLRELFRQAQELLEIAKKRGDPQVAVAALMALVALAAWTGDEEKLRELFRQAQELLEIAKKRGDPQVAVLALLALMFLALRTGDEEKLWELLLQAAWLLMIAAKRGDPAVALAAMIIMAVLMAR
>sg110_top85_85__X_h21_l2_h17_l3_01483loop_17_C3ABC_000000016_2_orient_0001_A
EVEMELISTVLESAQKQGDPEKLRQALELLKRLSEEAQKNGNLEVLMELISTLLESAQKQGDPEKLRQALELLLWAAFLAWLNGNLEVMMELISTILESAQKQGDPEKAWVALVLLALLYVLAAINGNVEVALELLSTMLEALQKM
>sg111_top87_87__X_h17_l3_h20_l3_170362_loop_32_C3ABC_000000012_2_orient_0001_A
PEVEVWLLMILAIIVIIQGDEEKLRQIFEQAQEILRKAETPAVMAWALAVLMWCVAKLGDEEKLRQIFEQAQEILRKAETPAVMAVALWILAILVAALGDREKLVQIALQALEIMRKAETPQVAAAAWRIIIWLALK
>sg112_top88_88__X_h17_l3_h20_l3_170362_loop_81_C3ABC_000000017_2_orient_0001_A
LALVALAAAWMLAAAVAGDEEALRLLFLFAQALLEIAKKRGDALVAIVALLILAELAAKTGDEEKLRELFRQAQELLEIAKKRGDPLIAVWALWILMALAKKTGDEEKLRELFRQAQELLEIAKKRGDPLVAQLALAILLALLKK
>sg113_top90_90__X_h14_l4_h24_l4_029792_loop_8_C3ABC_000000017_2_orient_0001_A
IAEVVIEIAAWEFFRTARKDPEKASEKLKEIIKRAGVSEEDMELAIFLAVAALAFWTFARTARKDPEKASEKLKEIIKRAGVSEEAMEAVITAAVAALAAATATATARKDPEKAIKKVAEIIMRAGVSWEWVWWVIAAAAAAAASI
>sg114_top91_91__X_h17_l3_h20_l3_170362_loop_32_C3ABC_000000018_2_orient_0001_A
PEELALLLLVVALLALALGDEEKLRQIFEQAQEILRKAETPEVMAAALWILALCVLQLGDEEKLRQIFEQAQEILRKAETPEVMAAALEILALLVIALGDLEKLWEAALLAMEIFRKAETPRVQLAAAAILLALMLAAGF

Cyclic homo-oligomeric scaffolds: with surface redesign

>sg115_rpxdock_top11_11__X_h27_l2_h18_l3_14564_1_loop_15_0001_one_capped_optimized_C3_1_0001
HHHHHHGSENLYFQGSGDEKLEYNQELLKRVMKEARKQGDEELEKIAKKATENLKNGDHEEA
KKLLKLVRSILQLNKVMKIAKEQGDEELYEIARKATENLKNGNLEEAKKLLYLVLLILLLNEVMEAA
KRMGDEDLYETARKATENLKNGNEEEASHLFSLVRAILLLNEVQKEAEKLGDEELYEVAEFATQL
LKDGDLEVALVNLVWVLILLHKL
>sg116_rpxdock_top17_17__X_h15_l3_h26_l2_10485_1_2_loop_45_0001_one_capped_optimized_C3_1_0001
HHHHHHGSENLYFQGSGDSEKLIAEIILEQIEQLALSPEIKKTVRKAKESLKNGDEEKAERMLEIA
VLLYLLIFVFRFAKDPEIKKTVQEAEESLKNGDFEKAEKMIKIAQLLLELIEIYERAQDPLIQQIVKEA
LESLKNGDFERAQEMIKIAELLLELIEIFRKATNPQIKYTVFQALLSLAQGDFEKAREMIEKARKLL
EEL
>sg117_rpxdock_top27_27__X_h15_l3_h26_l2_10485_1_2_loop_45_0001_one_capped_optimized_C3_1_0001
HHHHHHGSENLYFQGSGDEEEERAKKLLRKIEEEALSPEIQKTVREAKESLKNGDHERAEEMIK
IAKLLLQLLRIFERAQDPTIQETVQKALESLKNGDFERAREMVQIALLLLLLIEIFKKAESPEILKTV
EKALESLKNGNFEEAVKMLQIAVLLLILIRIKKEAQDPEIQRTVEIALELLSLGDVQIALLYVLYAVILL
TQK
>sg118_rpxdock_top37_37__X_h19_l3_h30_l2_14039_1_loop_26_0001_one_capped_optimized_C3_1_0001
HHHHHHGSENLYFQGSGTEELVKALAIANTVVLLALIAGDPKVQRVAIKIYEAIVEMVEKGDEDRL
QRILKLARALSTAIKVLLLALDAGDPLVKKVAKEILERILEMIERGDEEKLKEILELAKALSTAIEVLLL
ALEAGDPQVEEVARRILELILRMIQKGDKDKLQEILELAEALSTAIKVLLLARKAGDPEIEKEAEEIL
KLILEMIEKGDREKLREILKLAKELSEKIKKQL
>sg119_rpxdock_top10_10__X_h18_l3_h29_l4_05684_1_2_loop_17_0001_one_capped_optimized_C3_1_0001
HHHHHHGSENLYFQGSGSEELQKEALVLAVQVVLIALIAIIQGDEKLYRLARYALELAIKALQEGD
DAQVRRAILFMQAVVLIQIVKEIAEKAKKQGDEELFQLARWAIELAEKAIEKGDEESVKRAILFML
AVILIQLVKEIAKKAKEEGDDELFKLAEWAIKLAKEAIEKGDEESVKRAILFMEAVILIQKVKEIAKKA
KEKGDEELFRKAQEAIEDAKEAIKQGDEEKVKEAIEKMKRVLKEFNEK
>sg120_rpxdock_top10_10__X_h29_l3_h23_l4_18471_1_loop_38_0001_one_capped_b_optimized_C3_1_0001
HHHHHHGSENLYFQGSGNLQEQILRNNIWLMRRLVEEVQKLAREEGDEEIERIAVLAYVILLLAL
QLLVESTDPETQKTIEALILIAEMLLLVAEVLLEAKREGDEEIQKIAQLAYEILREALELLKKSLDPET
LKTIEALIKIAEMLLLVARVLLLAKRRGDEEISRIAQEAYKILKEALELLQKSLDPETQKRIEELIRKA
LRLLTEALRKLYQK
>sg121_rpxdock_top15_15__X_h27_l2_h15_l3_16437_3_loop_79_0001_one_capped_optimized_C3_1_0001
HHHHHHGSENLYFQGSGDEVLRETLEQALEIALQLAQKQGDEDAERKIKELKELAKKASTPEEL
EKIFKQVQLLYILLRALELAQRQGDELAEEKIKELLKLAAHAQSPEELKKIEKQVKLLISLLKALELA
QRQGDEVAERHIKRLLELAAKASSPEELEKIMLQVRLLILLLQALLRAQEKGDEEAEFLIHILLQLA
IEARTPEELKKIEKIVRELLLQLLLK

>sg122_rpxdock_top31_31__X_h18_l3_h29_l4_05684_1_2_loop_17_0001_one_capped_optimized_C3_1_0001
HHHHHHGSENLYFQGSGSEELEKQALVYAVQVLLIALRAILQGDEKLFKLAEWAVELAIKALQNGDERQVQRAILFMQAVILIQIVKEIAEEARKQGDEKLFELARWAIELAEKAIERGDEESVKRAILFMQAVILIQIVKKIAEKAKRQGDDKLFKLAEWAIELAERAIEEGDEEKVKRAILFMEAVILIQLVKEIAKKAKEQGDEELFQRAEEAIKDAEKAIEEGDEEKVKEAIKKMKEVIRIQERK
>sg123_rpxdock_top00_00__X_h20_l2_h26_l3_00455_1_loop_12_0001_one_capped_optimized_C4_1_0001
SESQELQLERLAELVEKYAEQAKTPEQEELARRLVELAKEALKKGDEEKIRKIIHMLDLIILAIRVLQYAEQAKDPEQKELAEKLVELALEALKKGDDEKIQHIQKMLRLIILAIRVLQYAEEATTPEQKELAERLVELALEALKKGDEELIQKIAIMLQLIILAIRVNKYALHAKDPYIREVAENLVKLALVALRIGDILVILQIIVQLVKIIALIEKL
>sg124_rpxdock_top05_05__X_h23_l4_h30_l2_06341_2_loop_21_0001_one_capped_optimized_C4_1_0001
TDERLESYEKLVHYLLIIIEILSQYAETPEQQEWLFRLLRFVLQTVVEAVKYIYQDGDKERAELLLRLAFLILQVMILQWEIEELLKEAETPEQKERLKKLLEFVKETIERALRYTQEEGDLEKAEILLRLAELILRVMKLQIRIEKILKKAETPEQQEWLKKLLKFVQETIELALKYTKEVGDLEKAELLLRLALLILQVMELQWRIEEILKHAQTPEQKEKLKKLLEEVKRTIERALRYTKEKGDLERAEELLKRAFEILKKMVEEWRRE
>sg125_rpxdock_top11_11__X_h26_l3_h30_l4_02060_1_loop_18_0001_one_capped_optimized_C4_1_0001
DQVIKRIKKRLKKLSEEIKRLAEKLKKLFEDPKLEELIKLIEEIMKLAQETVKEAEEKGDEKLARIALIALQLAKLSLEILLLAEKLKKLFSDPKLEELIKLIEQIMKEALETVKEAIEKGDEKLAKIALIALQLALLSLQILELAERLKQEFEDPKLEELIKLIEKIMKKALETVREAVEKGDEELARIALIALRLALLSLQILQLAEELKKKFTDPLLELIIEYIEVQMELALLLVERAVEEGNRELALKALLLLRAYIILYELEQKQ
>sg126_rpxdock_top15_15__X_h26_l4_h16_l3_00599_3_loop_42_0001_one_capped_optimized_C4_1_0001
NETELEEVAKELEETMEKSKREGDEETYQILRELSELAKEALKRGDEEKVTKLLLLAKLILVAVELKETMKESKKRGDEETYQILRELSELALQAIKEGDEELVKKLLLLAKLILVAVELKKTMKKSKRLGDDETYEILKRLSELALEAIKEGDEEKVRKLLLLAKLILVAVELKEWMEESRKKGDLETYYILQILSYLALIAIKKGDQEKVYELLALAELILKQLKKK
>sg127_rpxdock_top18_18__X_h19_l3_h29_l3_15220_1_2_loop_85_0001_one_capped_optimized_C4_1_0001
DEKRKQAELYAQVAIVASQLAQDPELREMLVRLVQLLLQALVEGDEEKLEEILRLIKAAIYAVNALRLTERAKDPELREMLKLLVELLLEALERGDEKVLEHILRLIKAAEYAVEALELTERAKDPELREMLKLLVKLLLQALKEGDEEKLKHILKLIKAAKYAVIALELTERAKDPRLRELLERLVKLLLQALKEGDEEKLERILKLIEEAKEIVIKQ
>sg128_rpxdock_top18_18__X_h28_l3_h19_l3_02324_3_2_loop_22_0001_one_capped_optimized_C4_1_0001
DLKLLVVLSTLLLVLLFSMELARRLGDEELAKILEEIVKRILRVLIRGDEHRVERLITLAELAISALELLQSMELARRLGDEELAKILEKIVKRIIEVIKLGNEREVRKLLTLAELALSALRLLESREEAERSGDEELAKILEKIVKLIIEVIEKGDEEKVRHLLTLAKLALSALQLLQSMKIAKERGDEELAKELKKIVELIIEVIEKGDEEKVRKLLEEAERANQKL

>sg129_rpxdock_top18_18__X_h29_l2_h20_l2_14874_2_2_loop_30_0001_one_capped_b_optimized_C4_1_0001
SLEDLLKALVLYQVIVELMREIEKKGDEEAYRAARTALWLLQLAVHVDKEQFEKLTQLVKALQLYI
EIVDLMREIEKKGDEEALEFAKEALKLLQKALRVSEEQFKKLIKRVEALQLFIEIVELMREIKKKGD
EEAFRFAQRALELLKKAHHVDEEQFEKLTEEVKELQLFILIFLTS
>sg130_rpxdock_top22_22__X_h19_l3_h26_l3_01639_2_loop_25_0001_one_capped_optimized_C4_1_0001
DEKLEELARKVKELLKKAKEIAQKLGDPRYEQLVNLVEQFLRYAELHNDEKLVRLILLAAKVLELLI
KALEVAQKLGDPKYQKSVELVLKLLLEALENNDEELVKLILLAAQVLELLIKAWELAQKLGDPEYE
KSVELVLQLLKEALENKDEELVHLILIAAQVLVLLIQAFELAQKLGDPEYEKAVRKVLELLKEALKN
KDEELVKKILQLALQLLVQITRAL
>sg131_rpxdock_top24_24__X_h26_l4_h18_l4_13842_3_loop_16_0001_one_capped_optimized_C4_1_0001
TEEQKFQVLVLVLILQTLVLWSGDEELYQIALFALEVALEALQKGDPELLEQAKLLLQVVILMFKLL
QLLKQSGDEELREIAEIALKFASEAIKEGDPKALKHAQLLLEVVILMFELLELLRQSGDEELKEIAEI
ALKFASEAIRKGDPELLKKAKLLLEVVLLMFKLLELLRRVGDEELKKIAEKALKFASRAIREGDPEL
LKEAKRLLEEVLRLFKQK
>sg132_rpxdock_top26_26__X_h29_l3_h20_l2_13757_1_loop_75_0001_one_capped_optimized_C4_1_0001
SEERLQAEVLVLQALYLAIINGNYEFLLIVLRLLEEIIRAIEQGDDEKIRRLIKQIRALYHVLVAFILAQ
SAGQEEFLKIVERLSEEILQAIELGDDRAIEKLMEQIQALIHVLRAWFLAKQAGDEEFLKIVEELSE
RILKAIRKGDDERIRRLMEQIRALIHVLIAWFLAKRAGNEEFLKRVEELSELILRAIQEGDDEKIREL
AKEIRELIEKLRKL
>sg133_rpxdock_top34_34__X_h26_l4_h20_l3_07365_1_loop_40_0001_one_capped_optimized_C4_1_0001
DEEIERLAYEAQLNSLVALFLSVLVNLAGFTQDTYKILQIMQLVQRSIKEALKKGDEKTIRIAYLAAQ
ALALSLAANVLILFVRRAGFTEETELIQRIMQLVLESIEEAIKKGDEQLIRIAHLAAQALWLSLAANQ
LIEEVRRAGFTEETEIIEEIMREVLKSIEEAIKKGDEQLIRIAHLAASALWLSLLANKLIKEVREAGFT
EETEEIQKLMREVLKKIKEAIKKGDEELIKEAHKIAQELYEILRENEKQK
>sg134_rpxdock_top46_46__X_h26_l3_h30_l4_12739_2_loop_16_0001_one_capped_optimized_C4_1_0001
SDRIQRRAIIYLEIAWLLAELIKEQAKSPIVEKIVQIILEAIIEARQLILQVDEEQLKQVLILAQALVYLA
QALLLALRILENATTPLVEEIVKIIMEAIEEAAELILHVDERLLRQVLILAQALYYLASALLLALQIEKE
ATSPEVEEIVREIMEAIKEAAELIRNVDEEQLKQVLILARALYYLASALLLALLIEKNAKSPEVEKIV
EEIMRLIRKAAQLIRNVDEEQLRKVLELAQRLFEKARKLLRQLKHE
>sg135_rpxdock_top48_48__X_h28_l3_h20_l2_10040_2_loop_19_0001_one_capped_a_optimized_C4_1_0001
DREIKEEARKLIREAIELLQKGDPRAKEILRQAILILLAIRLLEEMEENIEKAEKLGNEELSELAKRAI
KLVREALELLKEGDPRAEEILKLALKIIKAILLLLEMYENIKQAEELGDEDLSELAKIAIRLVRQALKL
LQEGDPRAEEILEIALRIIKLILQLLFLKQRIEEAKKKGDQQFVFEAEEKIRRIVEELFKLLEG
>sg136_rpxdock_top50_50__X_h19_l3_h28_l4_13252_2_loop_10_0001_one_capped_optimized_C4_1_0001
SEEQERELKKLVEEIEETLEETKKKNPELMKIAKKLAEFAKEVLKLGDQRLVYEAFLLLELILLVIRI
FEALEETRKKDPKLMKIAKELAEWALEVLKEGDIEKVIKALILLKLILLVIEIKEALEETEKYDPRLM

RIAKELAEFALEVLQRGDEELVEKAIYLLRLILLVIKIKKALEETQQKDPKLMQEAELLAQLAQIVLR
EGDLFRVLEAFIRLLRILEKIEKALKS
>sg137_rpxdock_top56_56__X_h19_l2_h30_l2_13262_4_loop_3_0001_one_capped_optimize
d_C4_1_0001
DEDSEKQVEEIREIIEKAKKQGDEEVAKLLEKLLEKIEKAIKILKETGDPRAQELVVEMLKLAQYLIF
VFRILEYIEKAEKRGDEEVAKLLKKLLEKILRAIKILEETGDPRARELVEEMTQLARYLIFVFRILEIIR
RAQQQGDEKVAKLLEELLEKIKKAIKLLEKTGDPRAKELVREMVKLATYLVFVFEILEIIKEAEEQG
DLVVAILLQELLELILRAIEILEKTGDPRARELVEKMLFQALQLVVYFRVV
>sg138_rpxdock_top56_56__X_h20_l2_h27_l2_14278_2_loop_58_0001_one_capped_optimiz
ed_C4_1_0001
DSQVLNVAAAFVVSVTVLASQTGDQELLEIALEILKRMLHVDEEQQHKLLRLLEALLTVVLVARVA
QKTGDEELLQAAKLIFQEMQHVDEEQQEHLRNLLQALLTVVQVAELAQQTGDEELLKAAKEIFK
LMQHVSLEQQERLLRLLEALLTVVKVAYLAQKTGDEELLKKAKEIFKKMQHVDKEQQEELLKEL
RKLLKEVQERL
>sg139_rpxdock_top58_58__X_h22_l3_h30_l2_15151_1_2_loop_25_0001_one_capped_b_opt
imized_C4_1_0001
DEEQLRRTEELLRRLEKEAKEKGDEELEKEIRKLQETLQLMEILLRILEFAEKFFKRTGDPKAKLL
AETVKELIRLLFEALQKGDEELLKEIKKLVQTLALMFLLLRILLYAEQVYKRTGDPLARLLAQTVEE
LIKLLYEALEKGDEELLREIFELVETLIKMYLLLLRLESAEQVLKKTGDPQARAQAEEVKALILLLYI
KLTLG
>sg140_C5_rpxdock_top0_0__JHR_bd4_01289_0001_0001_A
HHHHHHGSENLYFQGSGGLESVILLVIALITARYIGDPQLLRIIFQILLLLARELLKRGDVQAAVIAV
LFLVHVAQHLGDPRLLRELYELLRRIFEQALKDGDIEVAVLAIEILLTIAELLGDPELLKELKRLLEEL
EREAREQGDEEVLKLIEELLEKLG
>sg141_C5_rpxdock_top0_0__JHR_bd4_01295_0001_0001_A
HHHHHHGSENLYFQGSGSLESITLLLLTLVLILYLGDPDLLRVLNLLVRQLLERLLKNGDLHAVVILI
DVAIILLRHLGDPELLKELFEFIKKVLEEAQKQGNLEVVVELLRLLLELAEILGDPELEKEVEERIER
VLEELERRGNERLYEELKRLLEELRKG
>sg142_C5_rpxdock_top0_0__JHR_bd4_01541_0001_0001_A
HHHHHHGSENLYFQGSGDFEQAIQLLAILVLVQYLGDPDVIRILIILVLHTLEEVVKKGDLDVALVL
AHILILLAELLGDPELLKELFELLTELLKLALRRGDLRVILELIRLLLRIAELVGDPELFRRLLELVEEA
RELAREEGDERVLKELEEIYRRLKKG
>sg143_C5_rpxdock_top0_0__JHR_bd5_16455_0001_0001_A
HHHHHHGSENLYFQGSGTGEQLVQIILTILHAKEKGDETSLRVLLHILLRLLEEARKRGDPHLALL
IISLAVELARTLGDRRVLRRLLEILLELVEELHKRGDPELLILAIELAAILARELGDEEALEFLRRILE
ELRKRIEEEGDPDLLKDVEELLRELREG
>sg144_C5_rpxdock_top1_1__JHR_bd4_00469_0001_0001_A
HHHHHHGSENLYFQGSGDEEELEHLRRLLRRLRREGDPRLLEEILKLIKELVKRARERGDPQLL
LFLLFVLVRLLLITGDPRVLELALEIIKVLLEIAEKRGDPDLLATLLLILLLLLHTTGDPRIRELIFEILLL
LLYLADRDGDPHLLRFVLQIISSLNQG
>sg145_C5_rpxdock_top1_1__JHR_bd4_01440_0001_0001_A
HHHHHHGSENLYFQGSGGFEQLLLLFLTVVLLRHNGDEHLLRLLLQILLLLLEELLREGNLKLIVIL
LQLVITLAEELGDERVLRNLLELLLEALELARREGDLEALVILLELLVKLAQKLGDEEVLEEVRRIIK
EVLKKAEEEGDEDAVKRLRELLKTVG

>sg146_C5_rpxdock_top1_1__JHR_bd4_01682_0001_0001_A
HHHHHHGSENLYFQGSGDSLDLVTVMILALQRGDPDLLRILLLIAINIAEELIRRGDLKEAVHVLLI
ALLVALRLGDPRLLRQLLRLLLQVVERAHERGDVHTAVLALEVLVRIATELGDPELKRRIEELAEKI
YEEARKQGNEEVLEQVRELLEELRRG
>sg147_C5_rpxdock_top2_2__JHR_bd4_00964_0001_0001_A
HHHHHHGSENLYFQGSGDEELLHFVIQLLLLYLHVGDPTVLRLALYLLLKLSRRAEEKGDLELLIK
LVVLLLVLFKLLGDPQVLHLAKEILRVAAEIARRRGDLDSLLIIVQLLLLLYELSGDPEVLELARRVV
EHAREIAEKQGDEEIKRLLEELKREIEEG
>sg148_C5_rpxdock_top2_2__JHR_bd4_01067_0001_0001_A
HHHHHHGSENLYFQGSGDDQDILLVILIVLTASHLGDEDVLLILLFLILRLIREAAERGNLELVLVAV
QLAYQIARLLGDEQVKKELLEIILELLEQAVERGDLELAIRLLELALRIAVELGDEEVLKRLEELVER
LIELAERRGDEQLLENLRRLLEKIRKG
>sg149_C5_rpxdock_top2_2__JHR_bd4_01295_0001_0001_A
HHHHHHGSENLYFQGSGSTETIILLLLVLLTIQHLGDPDFLEILNIIVENLLEELLKNGDLEAVVELIA
VAILLLHLLGDPQLLKELFERIKKVLEEAEKQGNLHVVVRLLRLLLKLAEILGDPELEKEVEEKIER
VLEELERQGNEELYEELKQLLEELRKG
>sg150_C5_rpxdock_top2_2__JHR_bd4_01480_0001_0001_A
HHHHHHGSENLYFQGSGGEEELKELKELLERIEKEGDPELVKRVLELGRKLLEEATRRGDLETL
LTVLKILLELLRLVGDPRIVELILRIGERAVEVALRQGDLRVLLQLLILLLLLLNHVGDPRIREVIDRV
GLRVLLLAIEEGDQTILLVLIVLLLSLLHL
>sg151_C5_rpxdock_top2_2__JHR_bd4_01853_0001_0001_A
HHHHHHGSENLYFQGSGTLEEVEELIKEAEEKGDPELLKKAFELLEELVREARERGDLEVLLRIV
ELLIEIAKLLGDPEVLKLALEILEDLTRTARERGDLRILLNIVILLVLLAELRGDPEVLKLAKEILFFVL
YRAIERGDQFVVNVVNVLLTLIQNG
>sg152_C5_rpxdock_top3_3__JHR_bd4_02173_0001_0001_A
HHHHHHGSENLYFQGSGDFEKVTQLINLLLLALYTGDPNILEVLIFIAFKLAERLLKNGDPELAALI
LIVMLHIAQQLGDPELLHRILKLALEVVRVAIEQGDLETALILIKFIVEVATRLGDPEVLRLARELLQK
LLETARKRGDEEIVRRLEEVIRELEEG
>sg153_C5_rpxdock_top4_4__JHR_bd4_00654_0001_0001_A
HHHHHHGSENLYFQGSGGEEQAISLFNALVIAVVHGDEHLLRTLLLIALVILEILLEQGDVETAVFL
VIVVLSVARRLGDEEVLRLVLKIALELLRQLLEEGDDELLVLLLIVVLEVAKELGDEEILREVLKIAR
ELLERLRKQGDEDLLRLLEEAVKRVEG
>sg154__C5_rpxdock_top4_4__JHR_bd4_01668_0001_0001_A
HHHHHHGSENLYFQGSGGEEQLLLLIILVLTASYIGDPHILRLLVILLLVLLRRLKENGDVETVAVAL
FLMVNIARLLGDPDLLRRLVEQIEEVIKRALKNGDVEVAAELLELLVKIAKQLGDPELEERIRELLE
RVLKEARKQGNEDVVRLLEEIREKLEK
>sg155_C5_rpxdock_top4_4__JHR_bd4_02108_0001_0001_A
HHHHHHGSENLYFQGSGDRELALLVLLIVLNAHNNGDERILRILLVAVLLLVREAIKNGDLEAVLIL
LLVLVTIAQFLGDEDILRDVLRLILEALKRAIERGDLLVAVYLLELALEVARELGDEEALEELRRLIEE
VLKLAEEQGDETIKKRVREILEKIR
>sg156_C5_rpxdock_top4_4__JHR_bd4_02110_0001_0001_A
HHHHHHGSENLYFQGSGDEEIVITLVQLLLHAINHGDPKFVELLLRFAVYLIKQLHRKGNLEFAVL
LLVIALHVALRLGDPLILKQLLELALELLEELIKQGNLDLAVLLVIVALEVAHELGDPELLRLLEELIEE
LLEQLERRGDEDNLKDLERVLREIKE

>sg157_C5_rpxdock_top5_5__JHR_bd4_00591_0001_0001_A
HHHHHHGSENLYFQGSGDEEDLEKLEELLKELEEKGDPRVFEEILKLVEELLRRALERGDLELIL
RAIRVLLELLRLTGDPRVFELILEALERLLRKLLREGDVHKILTVIHLLLILLIITGDPRVRKLILEALLE
VTLIAAQRGDEKQLLAILVFISLLQLG
>sg158_C5_rpxdock_top5_5__JHR_bd4_01303_0001_0001_A
HHHHHHGSENLYFQGSGDLEEIERLLEEARKKGDPELLREVVELAKEVFERAHKRGDLETLVKL
IRLILTVAKLLGDPELLRTVVELAKRVLEQAKKQGDLRILVLLIFLILSAAFNLGDPELYQTVLELLLE
VLQIAIEQGDDRILLLLNILVLVVTSG
>sg159_C5_rpxdock_top5_5__JHR_bd4_01781_0001_0001_A
HHHHHHGSENLYFQGSGGEQEIIQILQVILDLRHLGDEHLLRVVLFLLVIVIKKLEEEGNLELLLVV
LHIAIQLARELGDERVLQFLLEALKRVLKLAHERGDLELIVLLAELAARLARDQGDEELLRHVREV
LREALKLAEERGDETVFEELRKVLEELERG
>sg160_C5_rpxdock_top5_5__JHR_bd5_14749_0001_0001_A
HHHHHHGSENLYFQGSGDLEELYREAQQVLRNGNPEEALKIVEEALRIAKEKGDEERLVELLYL
SAEILLDRGDPEKAREVVEEALRTLRQQGDDELLLKLLLLLVVILLRLGNPKEALKVLLEALKLAF
QLGDHLVLQVINSLILLVSSG
>sg161_C5_rpxdock_top6_6__JHR_bd4_00020_0001_0001_A
HHHHHHGSENLYFQGSGDELERLLERVHREGDPELARRALELIREDVEEARKRGDVELLLKLL
EILVRLFRVIGDPEILELALEVAKLVSELAEEKGDLRLLLELLLLLLELLLLTGDPKIRKLLNRVLLLAL
QIASRRGDTKVINSLILILLLAQNG
>sg162_C5_rpxdock_top7_7__JHR_bd4_00349_0001_0001_A
HHHHHHGSENLYFQGSGGFELLVLAAILLSHAQENGDPRLLQVLAIILERILKEARENGDQELVV
VAALLLVTVAELLGDPKLLRRVAELLETVIRTAREQGDLDLLLLAVELLLIVAKLLGDPELEKLVEEA
LEHLEKLARERGDEHILKIVERLQEELKRG
>sg163_C5_rpxdock_top7_7__JHR_bd5_03365_0001_0001_A
HHHHHHGSENLYFQGSGDLEELLRRAEQLLLNGDPEKAEEVLREAIRLAKEKGDSELLLRALLL
LAFVLLNLGDPEKAREVLEEVLRRAIEQGDEHILLLALLLLVRVLLELGDPERAREVIRLLLFLALEI
GDTIVINVLLNLHEDVERG
>sg164_C5_rpxdock_top7_7__JHR_bd5_15772_0001_0001_A
HHHHHHGSENLYFQGSGSLEEIEEVIRLARELGDPELLRILEELLRRFREAERRGDLELLLIIILLV
DIIHVLLVARREGDPEIVTLLEILLRRARELFEKGDLRRLFILAVIATVLRILLEARKRGDPRIVHQLE
DILQALLIAYHSGDLQVAIFLLLYISRLYVELQG
>sg165_C5_rpxdock_top7_7__JHR_bd5_16455_0001_0001_A
HHHHHHGSENLYFQGSGTGEQLVQIVLTILHAVEKGDEDTLRLLLEILQQLLEEAVKRGDPRHAL
LIIFLAVRLARELGDERVLRKLLEQLLELVERLHKRGDPELLILAIELAAILARELGDEEALEFLRRIL
EELRKRIEEEGDPFLLRQVEELLRELREG
>sg166_C5_rpxdock_top8_8__JHR_bd4_00997_0001_0001_A
HHHHHHGSENLYFQGSGDEELLVLLIQSLSDAQRKGDEDSLRVLLQLLFLLIERLVKKGDVQLAL
IALELIILIARREGDEELLRRVLELLLEVAEIALKQGDVEILLQALELAIKLAERLGDEEVVRILKELLE
KAKELAEKQGDEEALRRIEELLEKLQRG
>sg167_C5_rpxdock_top9_9__JHR_bd4_00236_0001_0001_A
HHHHHHGSENLYFQGSGDEDLLAILLLIALHLGNLEVALRIVRLLSRFLREGDPRLVETALNLVLLV
ILVRLARKLGDEDVLRYIEELLRRFIREGDPRLVEEAIRAVTLAELLRTVEEQGDEEMFRRIEELLK
RFLREGDPRLVEEVRKLLEKAEG

>sg168_C5_rpxdock_top9_9__JHR_bd4_00844_0001_0001_A
HHHHHHGSENLYFQGSGGHEHLLILLLLLTLAYDKGDPDTLKILVQLLLLLAKELLENGDPHLAVI
VLLIALTVAQRLGDPRLLKHLIKLLLEIAEELIKQGDPELALIALEIAAIFATRLGDPELLKEVRRLLEE
LKRRLEENGDPELAKRVEEVLKRVKEG
>sg169_C6_rpxdock_top0_0__JHR_bd4_00057_0001_0001_A
GLEEVRILLDLLRIATERGDEEFLEALLIHILHQVEEALKRGDLTLAVLLLILAAQAVEELGNKSLIKK
LLRLLLELLRLALKHGDLDTLVTVLKLVIRLARHLGDEEVLEELRRLIKRLIEHARREGDEDIVELLE
ELLEEVEKG
>sg170_C6_rpxdock_top0_0__JHR_bd4_01230_0001_0001_A
GEESVLHLVYVLIELLQNGDRENVRIIAFFLLILAHRLIEEGDLVSVVLALSILSEVAEFLGDDLLLQ
KVLELLLEVIEEALKEGNLLVAVLALELVVKVARKLGDEEALRRVRELIKEAEELAKKEGDEVSLKL
LERILEEVEKG
>sg171_C6_rpxdock_top1_1__JHR_bd4_00017_0001_0001_A
GLEHVIQLVYVIHHALREGDEQILEVLLNLILVLLEKLLKEGDPFTALAAAFLASHLAENLGDEEAL
RRILELLLKLAELALRDGDPRLALRILKLAIKLAEKLGDEEVLKRIREILEELLEEAKRQGDPELYRK
IEELLRKLG
>sg172_C6_rpxdock_top1_1__JHR_bd4_00943_0001_0001_A
DRESFKTVLQALILALLLGDPQIHRLLRDVLEFLERIGDPHAEELIEVLYLAAEALLNAREVGDPEV
FRRLQELLRRYLHEGDPRYLELIRLLLLILRAILIARKLGDPELLRRVEELLEKFLREGDPRYAEEV
EEILKELERRRG
>sg173_C6_rpxdock_top3_3__JHR_bd4_00017_0001_0001_A
GSELVVQLVYVILFALKEGDEQILKVLLLVILSLLEELLKRGDPSTALAAAFLASNLAENLGDEEAL
REILELLLKLAELALRQGDPRLALRILELAIKLAEKLGDEEVLKRIREILEELLERAKRNGDPELYRK
IEELLRKLG
>sg174_C6_rpxdock_top4_4__JHR_bd4_00021_0001_0001_A
DLERLLQIILIALHAREVGDPEILRILVLILRDLVKELVERGDIESALRVLLLLAIVAHNLGDPRVLKYV
LEILRELIRLAHERGDWRLVLRVAELLLHVARLLGDPEVIKELREILRRLLEEAEKRGDEETVREV
REVLRLVEEG
>sg175_C6_rpxdock_top5_5__JHR_bd4_02220_0001_0001_A
DVDELIELLIELLSKGDPKSLRIVAQIALILAIDLRHEGNPEEASKLLLILLHVALHLGDPQLLHILFEF
ALRLVEELLKNGNPTQAAVLLLVLLRVAKELGDPEKLKRLEEIAREVIRELKENGNPILAQLLEIAL
ERVKKG
>sg176_C6_rpxdock_top6_6__JHR_bd4_00238_0001_0001_A
DEDLLRLVLHLLSLYIYQGDPRALEIALQLLEYLVLRALTRGDLELLLHLALLLLLLYELLGDPRVLE
LVLELLKLAVEDATKNGDLEVMLILILLLLQLYKLLGDPRVRELIRELLESIEKLAKERGDKTVERLL
ERLKEQLRRG
>sg177_C6_rpxdock_top6_6__JHR_bd4_01338_0001_0001_A
GIEQLINLLLVVVNARHIGDPDVLRVLVDLVLRILDEAVRKGDLELALVILSLLYLLAKLLGDPELLR
EIVEILLELIEEAIRRGNLEIAIRLIKLALEIAKDLGDPELFKELERLVEELEEIAKKRGDERLLEELKR
LTEEIRRG
>sg178_C6_rpxdock_top8_8__JHR_bd4_01289_0001_0001_A
GTEDVLTIILLLLHAQHNGDPETLRSLFIILLLLALRLLQRGDVHAAVIAVLVLVNVARHLGDPELLK
NLYDLLKRIFETALKRGDILVAVFAIEILLRIAELLGDPELRKELKRLLEELERRAREEGDEIVLQLIE
QLLEKLG

>sg179_C6_rpxdock_top9_9__JHR_bd4_00221_0001_0001_A
GEELLIHLFYLLVYLQERGDPELLRLVLNLLIITLRRALERGDLQLVLIIAQVLLYFLRRLGDPRLLET
VLRLLLEALELAIKRGDLELALEIAETLLRVLLHDGDPELLKRVLKLLEEVLRLARERGNERLVREI
EEVLEEARRG
>sg180_C6_rpxdock_top9_9__JHR_bd4_00411_0001_0001_A
DHDLFIHLLLEALTRGDPRLLKILLELALYVVRLLLIKGDLESAFTLLVVIAHAAHHLGDPEVAKKVL
RLALELVERLIKEGNLELAVLVLEIAVKLARELGDPEVIKELRKLAEELERLLREKGDERLLRELEEI
LRELRG
>sg181_C6_rpxdock_top9_9__JHR_bd4_00476_0001_0001_A
GEFTVLQLITNVLNALRRGDEQVLRLLMRLMLTLAFILRRNGDVESAQLLLQVLLHLAEHLGDEE
VLHLILQLLLEALELALKEGDLVVAVFLVELVLKLAEKLGDERVLRRIEELLEELLKRLEEEGNENLV
RIVRELLRKLKEG
>sg182_C3_rpxdock_top0_0__JHR_bd4_01527_0001_0001_A
DKHLHRTLLLLIAVLLSLLGNEELLRKLEELIKRFREEGDPRLVELAERALILALVLHLLRRLGNEEA
VRALEELIKLFLREGDPRVVELAELLLKMARLLKEARELGDEEVVRRIEELIRRFIEEGDPRYVEE
VEELLKEAREKKKGSHHHHHH
>sg183_C3_rpxdock_top0_0__JHR_bd4_02036_0001_0001_A
HHHHHHGSENLYFQGSDHIDRVLIFAIYVLLELQREGDPELFEELSRLLEELYLIGDPEIAKRLQA
VALAALAILSARRHGNPELAELVEELLRLLLRHGDPRLVELLRVVAHALLVILRLREEGNPEKAEE
LLRLLEELVRKGDPRLVEELERLARKLLEELG
>sg184_C3_rpxdock_top0_0__JHR_bd4_02151_0001_0001_A
HHHHHHGSENLYFQGSGEEEVRELNRELIEALEEGDPEKLRELIERANERLEELLRRGDVETAL
RLLLILLLIALQLGDPELLRKLIEVAEKLLERAYRLGNVEVVVLLISVLVLVVLHLGDPELLKRLVEIL
QRLFHLAIRLGNDHVLSLIIILILLLRQG
>sg185_C3_rpxdock_top0_0__JHR_bd5_15506_0001_0001_A
HHHHHHGSENLYFQGSTVEELIERARELLKNGNPEEAEKLLREALKLAKERGNPELVIEVLLLLA
QVLVRNGNPEEARQVLQEALYLIVKLGDPHLLVELLIVLAEVLLRQGNPKEAALVLIIAASTAINLG
NPELLRLILEVVEEVTRG
>sg186_C3_rpxdock_top1_1__JHR_bd4_01200_0001_0001_A
HHHHHHGSENLYFQGSDERELLYEVLQLLLRALQVGSRELIERIYELARRLHEEGDPRYVEEVR
LLAELAELLLLAREVGNEEVVRAIEELIRRFIKEGDPEYVRRAQLLVLIAQLLKLARQAGDEEAER
RIRELVERLIKEGDPRYLLEAFRLVIEAYG
>sg187_C3_rpxdock_top1_1__JHR_bd4_01216_0001_0001_A
SFVLKHARELLEEGDEEQLFQYLTLILILIIILINRGNVERAVEAVKVALEVALRLGDDALLEIVLYLVL
ELLRLLLEDGNVELAVELVILALEVAEELGDEEVLRDLLRIAEELLKRLREEGNEELLRKLEEVVKK
IKGSHHHHHH
>sg188_C3_rpxdock_top2_2__JHR_bd4_01593_0001_0001_A
DDKLVIVLLLLHQSLGDPLVAELAFHILEQLVKEAEKKGDLELLLTLIVLVLVLFHDLGDPRLREIAV
RAIKELARLVHERGNLELILKVVFLLIVLAETTGDPELKELAKRLVKHVIKLAREEGDEELLRIAEHL
LERLKEGSHHHHHH
>sg189_C3_rpxdock_top3_3__JHR_bd4_00257_0001_0001_A
GERELLHLALLLILTLNLGNEEFLKDLLEELLKLIEKLEREGNPELVVSALFIILAVADKLGDEEILKR
ALELLEELIERLIKEGDPELALRAIQILLQILKRRGDEEKLKRVLELLKELIELLERDGDPELLKRAR
ELLKRVEEGSHHHHHH

>sg190_C3_rpxdock_top4_4__JHR_bd4_00145_0001_0001_A
DLDLLLVLIDIARTRGDPSYREFIERLIRHIYERLREEGNVELLLTLLALAILALELLGDPRIRELLVEII
ELLFELLRKEGNVELLLILLLLAITVLELLGDPRVRELLERILKLLKRLIKEEGDEELRRLLEELEEQ
VKRGSHHHHHH
>sg191_C3_rpxdock_top4_4__JHR_bd4_01216_0001_0001_A
HHHHHHGSENLYFQGSTWEELTRRALELLHEGDEEKLREILERILELIEELTRRGNVEEAVRAVL
VALVVALTLGDEELLRLVLELVRELLKELLENGNVELAVHLVILALLVAHRLGDEEVLRDLLRIAHHL
LVDLLKEGNEHLVTLLIIVVVLIEG
>sg192_C3_rpxdock_top6_6__JHR_bd4_01348_0001_0001_A
HHHHHHGSENLYFQGSDVEELRRRVIEALKKGDEELLKRLLEELRKLAKRLIEEGDLERAVEVLH
VAVLVVIYLGDEEVMHEILRLLLELLEDALRRGNLEVAVKIVFVIILLAVTLGDKEILERLRELVFKLLI
DLLEEGNEELLNLLILLYALVHNG
>sg193_C3_rpxdock_top7_7__JHR_bd4_00145_0001_0001_A
DLDLLLSLILLARERGDPNLRDLIEQLIRRIYERLREEGNVEQLLNLLVLAILALRVLGDPRIRELLV
QIIELLFELLRKEGNVQLLLILLVLAINVLELLGDPRVRELLEEILKLLKRLIKEEGDEELREILERLEE
DVKRGSHHHHHH
>sg194_C3_rpxdock_top7_7__JHR_bd4_00591_0001_0001_A
DERELLRLILLLLLLHLLGDPNVFEQILKEVEELLRRALERGDLELILNAILVLLVLLHLTGDPRVFEL
ILEALETLLRRLLEEGDVEKILKVIHLLLLLLHITGDPRVRELIEEALRHVEEIARRRGDEELLKKIEE
HKKRLERGSHHHHHH
>sg195_C3_rpxdock_top7_7__JHR_bd4_00609_0001_0001_A
HHHHHHGSENLYFQGSTPEEILRKLQEALRNGDEELLKRLLEELLRLVEELLRRGNPEEAVRLLI
LAALVALNLGDEELQKELLKLLLEAIELLIKLGNPRLAVVVATVAVILALLLGDKEVLKELNRLLAELL
IEVLERGDPYLISLVLLLLYLLHQG
>sg196_C3_rpxdock_top8_8__JHR_bd4_01092_0001_0001_A
HHHHHHGSENLYFQGSTLREIIEKIIEAIRKGDEELLKKLIEEIIELIKELERQGDPEKVFEAALLLLL
VATSLGDEEALKVLIELLEELVRELLERGNPERALNLLIILVVVVLTLGDEEARERVERLILYVIVLLI
HLGNPELVLTALLLLAIILS
>sg197_C3_rpxdock_top8_8__JHR_bd4_01102_0001_0001_A
HHHHHHGSENLYFQGSTERLEEIVEKILRELREEGNEELIRFVEELLRRLRREGDPRYVQLAILA
YVVLKLLKQAEELGDEEVVRKIRELIERFLREGDPRYAQLVILLVVVLNIERLVKEQGDEEVLREV
REALEALIEIGDPSFVLYAINLLLELLTRRG
>sg198_C3_rpxdock_top8_8__JHR_bd4_01243_0001_0001_A
DPEIILALLILYTNLGDPHVLELAIEIVEDLAEEAKERGDPKLLLTIVVLLLVLFHLSGDPRIIRLALELL
KRVIELAEREGDPKLLLRVLYLLLILYNLLGDPRVLELAKELVRRVEELIKKDGDPELQKRIEELKR
QVEEGSHHHHHH
>sg199_C3_rpxdock_top8_8__JHR_bd4_01602_0001_0001_A
HHHHHHGSENLYFQGSDRQELFERVAEALEKGDRELAKELVERLIEVVERLTKNGNLEEAVFVLI
IALEAAELIGDEELLRLLLELAEELLERLLKNGNLELAVSLIVIALLIALDLGDEEILKTLRRLALTLLL
RLLQLGNHDLYVLVLLVLYEVRKG
>sg200_C3_rpxdock_top8_8__JHR_bd4_01689_0001_0001_A
SELERLLKEAEEEGNREELQRLIERATIAVLILVNLGNHEQAVELLKLALEIAVKLGDRSLVEVLVLL
VLLIVKELIELGNVELAVELLKVALEVAHHIGDKRLIRELLELAEELIKRLRELGDDETAEELEKLYE
ELKRGSHHHHHH

>sg201_C3_rpxdock_top9_9__JHR_bd4_01899_0001_0001_A
DHEKVITLLILLLLALNHGDPNLLQALLRILREVLERALRNGDLEVAVIAVLIILDVLKQNGDPEKLR
EAVEIALELLERALREGDLKLALLLVIILVNVVLHLGDPELIRRVLELLEELLEIAERQGDEITVKIAR
RLEERLRKGSHHHHHH
>sg235_C3_rpxdock_top0_0__JHR_bd4_00726_0001_0001
GIEDVLVLYLLLKSLEEEGDEELLRLILQLLEQIVKRLLKNGDPLQALYVLYLMIELAEKLGDEEVLK
KLLELLLEVVELALKQGDPLTALIAVFLALRLAKKLGDPEVRKRLLELLREIRELARKQGDPVLERL
VETLIERVERG
>sg236_C3_rpxdock_top0_0__JHR_bd4_02111_0001_0001
SEAVLLLTEVHRLGDPLAIRVVIELLQRYLRTGDPRYLEEIRLILLALRLLKTLEEAGDPKLYEHAR
ELLERLLREGDPRYLELIARLVFLVNALLIARKAGDPEIYRRIEELLERFLKEGDPRYVEEVAELVR
RVLERHG
>sg237_C3_rpxdock_top2_2__JHR_bd4_00175_0001_0001
DYDILELIEEAHKLGNPSVFHFLVELLLRFLEKGDPRIRELAHQLAQVVEALRKVKELGNPEVYRF
IEELLEKFIREGDPRYLKLAILAHLILFLLRIAKELGDPEVYRLIEELLEKFLREGDPKYVKKVLRLLL
EVLERKG
>sg238_C3_rpxdock_top6_6__JHR_bd4_00864_0001_0001
DVYVLFLVIELAKERGDPENLKLLLELALELLEELLKRGDLERALFLAIVVINLALHLGDPEVLRRVI
ELLLEIIRLLIERGDILSAVLAILVVLEAIEELGDPELLRELEELVEELLELVRKQGDETVLRLVERLLE
RIRKG
>sg239_C3_rpxdock_top6_6__JHR_bd4_00960_0001_0001
SDIVLLLEQAHRLGDPDAFHLIVELFFRFQREGDPRYKELVRLIVNLAQLLVFVRELGDPDVFRLI
EELLKRFYEEGDPKYVRQAELITLALILLEQVRQLGDPEAYRKIEELIKRFIEEGDPRYLEKIYELTE
EARG
>sg240_C3_rpxdock_top9_9__JHR_bd4_00258_0001_0001
SAEFLLILAHQVRERGDPRLREVILLLVKNLLQRARRKGDDRELIELLLVAANIARLLGDPELRELV
LRLAEEVLRLAQERGDVELIIHIAVIVAFIARTLGDPELLKLAREAAKEAIERARREGDEELVRLLE
QLLKVIEIL
>sg241_C3_rpxdock_top9_9__JHR_bd4_01893_0001_0001
SEDLVILMVVALIKGDEELVRQLIKIAELLVELLERLGNPEEALKLLIVLLAVAVTLGDEELFKKLVELI
ERLLEELQKQGNPRVVASVLILLLILARTFGDEELVRKLREQIKELIERLKKEGDPDLARFLEEVLK
TLEG
>sg242_C4_rpxdock_top0_0__JHR_bd5_12274_0001_0001
SSLEFAERLIRNGNPESAIHLLILALILANTRGDPERLAKILLVLAKLLEELGNPEQAREVLEEAIEL
AKEAGDPEVLLKILLRLAELLHKQGNPEEARKILREAEEIARRTGDPELLRELEELLERIEG
>sg243_C4_rpxdock_top4_4__JHR_bd5_01922_0001_0001
SLLYLVEELIKNGNPEEALKVLALVEIVVLYRGDPHHLIRVLLLLVEVLLRKGNPEDAEEILRELERL
VREEGNPELLARVLYLLVQLLRLKGNPEEAKKVLEELEKLVKEIGNPELERLLEELKEQRG
>sg244_C4_rpxdock_top4_4__JHR_bd5_08237_0001_0001
SLLEVARKLLENGNPEEAVQYLAIALLLADLQGDPELLLRILVLLAEILLELGNPEDAERLLRKAEEI
ARKRGNPELLVEILILLAEVLQRNGNPEEARKLLEEAEELVKRIGDPELERRVKELLEEIKKG
>sg245_C4_rpxdock_top4_4__JHR_bd5_11012_0001_0001

DEEIITVLAVIVLILLNNGDPESALENIIIAVLLAEQLGDPRLLVFLLFLLAEALLRNGNPEEARKHLE
KAEELAEELGDPFLLISLLLLLAEVLLNNGNPEEAEKVLERAEKLARELGDPLLEQLVKRVQERV
RKG
>sg246_C4_rpxdock_top5_5__JHR_bd5_00048_0001_0001
DIVAKHLHERGNPEFALVVLDVLLTLAKHHGDPEQVVRVLYLLAEILLELGNPEDALKLAREALEV
ARREGDPELLVEILVLMAEILLELGNPEEARELLKEARRLAEEEGDPELERRVKEAREKIERG
>sg247_C4_rpxdock_top6_6__JHR_bd5_10218_0001_0001
SLELLVRTIEELLERGDPETLTLLLALLYLVARLQGDPKAELLLLVLIIRLLINQGDPELVEKYLEELR
RKAEERGDPEAEILLLFLRAQLLIQEGDPEKVKELIKRMRELAEKLGDPEARKRIEELERQRG
>sg248_C4_rpxdock_top7_7__JHR_bd5_08929_0001_0001
SVQHLVEQLQRRGNPEQAIQLLVLAYLVAHNNGDPHRLIRVLLLLVEVLLQLGNPEAAKQVLEDI
RERAEKDGDPKLQLQLLVLRVRLLRLNGNPEEAQKVLEELKRQLEEVGDPEIEEQFKELQKELG
>sg249_C4_rpxdock_top7_7__JHR_bd5_12421_0001_0001
SLLELAERLLRNGNPEEAVRDVFLALILAYVKGNPEDVVKILILLAEVLLELGDPHTAKQVLEEAEK
LARELGDPELLVRILILLAELFLRRGDPEKAKELVKEAIKLLEELGNPELLRRARELLKEIEES
>sg250_C4_rpxdock_top8_8__JHR_bd5_04995_0001_0001
DKRLVEESIRELLERGNPESALHLIVLLSILALNLGDPKLRVAVILLEVELAIRRGNPEIAEELLEEL
REHIEKEGDPEARLRLLILEVRLLLLKGNPEEAEKVLRELEEEAKKQGDPEVEKEVEKLKEEVER
G
>sg251_C4_rpxdock_top8_8__JHR_bd5_07085_0001_0001
SIDVLALIVLLLHNGNPIIAILILQHLERLARERGDPELLVEVLLLLVQILVENGNPEDAEEVLKEAER
LIREQGNPRLEVLLLLVEAQLLLDNGNPEEAREVLKEAERLIREQGDPVLKIILDHIRRQVEEG
>sg252_C5_rpxdock_top0_0__JHR_bd5_06420_0001_0001
DTVTIHLIETLLNQGDHHLIVLVIVLLDVLATLTGNPHAQLVAIFLLLQLLLQQGDREVFKKLLEELE
RQARERGDPHILLLVLVLEIQLALQEGDEELVKELLRRLEKEARERGDPEIERLIEELKEELEEG
>sg253_C5_rpxdock_top2_2__JHR_bd5_13392_0001_0001
SSVANVALRRGNPELAVQLALLALLFARHDGDPELLVEILLLLARILIELGNPEDAERDLREALKLV
RELGNPELEIKILILLAEILLRKGNPEEARKVLEEAERLARELGDPELEREIEELLEEIERG
>sg254_C5_rpxdock_top3_3__JHR_bd5_12734_0001_0001
PDVLHIVEHLIKRGNPRDALKVLVQVILTAELQGDPKDLVEILLRVARLILENGDPELARKIVEKALR
IARELGDPELLVEVLLLLAEILLELGNPEEAEKLVEEAERLAREIGDPELRRLVEELLRKVREG
>sg255_C5_rpxdock_top4_4__JHR_bd5_05100_0001_0001
VDVLLAVAKTLLHRGNPEDALKIAAQAVNVAQDRGNPELVVKALLLLAEILLRLGDPEVAKEVLEL
ALRLAREAGNPELVVQALLLLVRVLQQQGNPEEARKVLEEAERLVKELGDPELEKEVEELKRRV
G
>sg256_C5_rpxdock_top7_7__JHR_bd5_02319_0001_0001
NVETVLHSVKRILENGNPEDAIKLLSHVILIVLYEGDPRLQVLLLLLQVELLLDEGNPELAKRVLEE
ARRLAEELGDPHLLVLVLLVLIRLLLEQGNPEEAKKALEEAKRLAEEEGDPNLKRELEEAEKRIRE
G
>sg257_C5_rpxdock_top7_7__JHR_bd5_17422_0001_0001
TPDTLLSIAIQLLTEGNPELARFVAQLALILAKREGDPDLLVVILILLARILLRSGDPENAREIIEEALR
IARELGDPDLIIIVLLVLVRILLRNGNPEEAKKVLEEAERLARENGNPFLLREIEELKEEVEKG
>sg258_C5_rpxdock_top8_8__JHR_bd5_11357_0001_0001

SEDAAFIALLQGDPELALNIVQLALNIAKRRGDPLRILELLILQAQILQRLGNPEEAEELIREAIERA
EKEGDPNIVIELLLLLAKILNQLGNPEEAEKLIRKALEEARKRGNPQLEREAKRLLEKG
>sg259_C6_rpxdock_top3_3__JHR_bd5_04250_0001_0001
DEVVELLLKRGDPEEAEHLARLALELALQDGDPEQIIRALLLLAKVLLRLGNPTEAYLYAYFALRN
ARKRGNPELIVDALLLLAEILLHLGNPESAKRLVEEAERLAKEEGDPELERRVREVKEKVERG
>sg260_C6_rpxdock_top3_3__JHR_bd5_04549_0001_0001
VSLLIVLAQRLSRNGNPEEAVRVIEEALYNAEEQGDPELLLRALLVIVEILMLLGNPQTAEFVLRRL
EERARKRGNPQVLLEVLLLLALVLLRQGNPEEAKKVLEEAEEEAKKRGDPEIERKIEELLRRVEE
G
>sg261_C6_rpxdock_top5_5__JHR_bd5_13792_0001_0001
VHTLIVLAQHLQREGNPQEAARVVNQALREAREDGNEHLLVYILILAAEILAELGNPRLAHQLLKE
ALKIAREQGNEELLVLVLVAIAHVLRLLGDPEEARRVLEEARKIAEELGDEELRKRVEKVLEKVER
G
>sg262_C6_rpxdock_top7_7__JHR_bd5_16640_0001_0001
DVLLLYVLAQRNEKNGNPEEAVKVLVEALQLAEERGDPELLVQLLLLAAKVALNLGNPENARKLV
EKALEIARENGNPELLVKLLLLLAEILLELGNPEEARKVIEEARRLAEEIGNPELLRKLEELLKRIER
G
>sg263_C6_rpxdock_top7_7__JHR_bd5_17141_0001_0001
DVIYLLSLAIEHLNKGNPEVAHELIRLALFLARHKGDPTLQVLVLILLAEVLLQLGDPHVALRVVEE
ALRLAKEAGNPELVVRALLLAAKIALELGDPERARELVEEAIRLAEEQGDPDLLNRARELLEEIKR
G
>sg264_C6_rpxdock_top8_8__JHR_bd5_15916_0001_0001
VSALFVLALHLVRNGNPREAVRDLEEALKNAEEQGDPHLLALLLILQAEILIRLGNPENAERVLRR
AEELIKEQGDPRLRVQVLLVLARILHKNGNPEEAEKYLKEAKEIVKELGDPELKRLLEELEEQIKN
>sg265_C6_rpxdock_top9_9__JHR_bd4_00970_0001_0001
DEQLLFRVTEALTKGDPELLRYVLHILKLLLLEARHRGDSERLVQLVLLLAQVALQRGDPEYLHY
VANLLEDLLRRLREEGNLELLVLVVYMAVRVALDLGDPELLKLAERLLEELERLARERGDKLLLE
HLEEAKERLKRG

Cyclic homo-oligomer extension and cap redesign

>sg202_Reverted_sg151_with_interface_5repeats_B1_0002_0001_0001_A_A1_0001_truncN4
_A
DPEELREAFEELEERVKEAKERGDLEVLLRIVEELIEIAERLGDPEVLKLALEILEDLTRTARERGD
LRILLNIVILLVLLAELRGDPEVLKLAKEILFFVLYRAIERGDQFVVNVVNVLLTLIQNG
>sg203_Reverted_sg151_with_interface_5repeats_B1_0002_0001_0001_A_A1_0001_truncN5
_A
DLEVDLEEVERLIEEAKEKGDPEVLKKALEILEDLTKTAEERGDLRILLNIVILLVLLAELRGDPEVL
KLAKEILFFVLYRAIERGDQFVVNVVNVLLTLIQNG
>sg204_Reverted_sg151_with_interface_5repeats_B1_0002_0002_0001_A_A7_0020_truncN1
_A

DLEVRLEEVEELIKEAEEEGDPEKLKEASETLEELVKEAQKKGDLEVLLKIVELLILIAEIEGDPELL
LEAFELLEELVREARERGDLEVLLRIVELLIEIAKLLGDPEVLKLALEILEDLTRTARERGDLRILLNI
VILLVLLAELRGDPEVLKLAKEILFFVLYRAIERGDQFVVNVVNVLLTLIQNG
>sg205_Reverted_sg151_with_interface_5repeats_B1_0002_0002_0001_A_A7_0020_truncN2
_A
DPEELKKAEEKLEEEVKEAQERGDLEVLLRIVERLIEIAEEEGDPELLLEAFELLEELVREARERG
DLEVLLRIVELLIEIAKLLGDPEVLKLALEILEDLTRTARERGDLRILLNIVILLVLLAELRGDPEVLKL
AKEILFFVLYRAIERGDQFVVNVVNVLLTLIQNG
>sg206_Reverted_sg151_with_interface_5repeats_B1_0002_0002_0001_A_A7_0020_truncN3
_A
DLEEQLRQVEELIKRAEKEGDPEELEEAFETLEELVREARERGDLEVLLRIVELLIEIAKLLGDPE
VLKLALEILEDLTRTARERGDLRILLNIVILLVLLAELRGDPEVLKLAKEILFFVLYRAIERGDQFVVN
VVNVLLTLIQNG
>sg207_Reverted_sg157_with_interface_7repeats_B4_0003_0001_0001_A_A8_0002_truncN3
_A
DEETELERLERQSEKLEHEGDPRVREEILEQLESITRTAQERGDFKTLLQILQLLSQLLEIVGDPR
VRKLIKEVLEHLTKEAEDRGDEKQLLAILKHISELLEILGDPRVRKLILEALEFVTRIALRRGDEELIL
RAIRVLLELLRLTGDPRVFELILEALERLLRKLLREGDVHKILTVIHLLLLILLIITGDPRVRKLILEALLE
VTLIAAQRGDEKQLLAILVFISLLQLG
>sg208_Reverted_sg157_with_interface_7repeats_B4_0003_0001_0001_A_A8_0002_truncN5
_A
DSETTLQELQQESQKLEKHGDPRVEEKIKEKLEKLTEEAEDEGDEKQLLAILKHISELLEILGDPR
VRKLILEALEFVTRIALRRGDEELILRAIRVLLELLRLTGDPRVFELILEALERLLRKLLREGDVHKIL
TVIHLLLLILLIITGDPRVRKLILEALLEVTLIAAQRGDEKQLLAILVFISLLQLG
>sg209_Reverted_sg157_with_interface_7repeats_B4_0003_0001_0001_A_A8_0002_truncN8
_A
DPRKDEKNLEALERETEEALEKGDEETILEAIRSLLELLEKTGDPRVFELILEALERLLRKLLREG
DVHKILTVIHLLLLILLIITGDPRVRKLILEALLEVTLIAAQRGDEKQLLAILVFISLLQLG
>sg210_Reverted_sg157_with_interface_7repeats_B4_0003_0002_0001_A_A3_0007_truncN1
_A
DLEEELERLHEQSEELHREGDPRVFEEILEALERITREASKKGDIKKLLEILRHLSQLLKIVGDPR
VFKLILEALEHATRQARKRGDRKTLLEILKHLSQLLRIVGDPRVRKLILEALHTVTERARDEGDEK
ALLAILKFISELLSILGDPRVRKLIHEALETVTQIALKRGDEELILRAIRVLLELLRLTGDPRVFELILE
ALERLLRKLLREGDVHKILTVIHLLLLILLIITGDPRVRKLILEALLEVTLIAAQRGDEKQLLAILVFISLL
QLG
>sg211_Reverted_sg157_with_interface_7repeats_B4_0003_0002_0001_A_A3_0007_truncN7
_A
DEEAELAELEKISERLSREGDPRVREKIEEALETVTQEALEKGDEELILRAIRVLLELLRLTGDPR
VFELILEALERLLRKLLREGDVHKILTVIHLLLLILLIITGDPRVRKLILEALLEVTLIAAQRGDEKQLLAI
LVFISLLQLG
>sg212_Reverted_sg165_with_interface_7_5repeats_E4_0003_0002_0001_A_A8_0003_trunc
C6_A

TGEQLVQIVLTILHAVEKGDEDTLRLLLEILQQLLEEAVKRGDPRHALLIIFLAVRLARELGDERVL
RKLLEQLLELVERLHKRGDPELLILAIELAAILARELGDEEALEELRRILEELIERLWKEGDPSLLID
AIELAARLARHLGDEEALEKLLKTLEELEERLEKEGDPELLRQARELAAKLARELG
>sg213_Reverted_sg165_with_interface_7_5repeats_E4_0003_0003_0001_A_A6_0001_trunc
C4_A
TGEQLVQIVLTILHAVEKGDEDTLRLLLEILQQLLEEAVKRGDPRHALLIIFLAVRLARELGDERVL
RKLLEQLLELVERLHKRGDPELLILAIELAAILARELGDEEALEELRRILEELIERLLKEGDPSLLIRA
IELAAELARHLGDEEALERLLRVLLELIERLHKKGDPELLIQAIELAAELAAQLGDEEALEELLRVL
EELRERLEKRGDPELLREARELAAQLARELG
>sg214_Reverted_sg165_with_interface_8repeats_minus8_B1_0003_0003_0001_A_A2_0004
_truncC2_A
TGEQLVQIVLTILHAVEKGDEDTLRLLLEILQQLLEEAVKRGDPRHALLIIFLAVRLARELGDERVL
RKLLEQLLELVERLHKRGDPELLILAIELAAILARELGDEEALEFLRRILEELIERLKKQGDPYLLIR
AIELAARLARTLGDEEALERLLRTLLELLEELHKRGDPELLIQAIELAAKLAQELGDEEALEELRRI
LLELIEELHKRGDPELLIQAIELAARLAESLGDEEALEELLRVLLELTEKLHKEGDPELLQRAIELA
AELAHILGDEEAEEELKKVRKELEEHRKKEG

Appendix 3: Validation metrics of cyclic homo-oligomers

sg115_C3	yes	soluble		24900	74700	46110	1.9	0.38
sg116_C3	yes	low						
sg117_C3	yes	low						
sg118_C3	yes	soluble		26518.37	79555.11	115900	4.4	0.46
sg119_C3	yes	none						
sg120_C3	yes	none-low	s200-13.75	25142	75426			
sg121_C3	yes	none-low	s200-12					
sg122_C3	yes	soluble	monodisperse	28549.12	85647.36	87330	3.1	0.02
sg123_C4	yes	low yield	s200-agg	27112.8	108451.2			
sg124_C4	no			34883	139532			
sg125_C4	yes	no-yes		33222	132888			
sg126_C4	yes	soluble	s200-11.75	28403	113612			

sg127_C4	yes	soluble	s200-agg, 12.75ml	27214.98	108859.92	97000	3.6	0.11
sg128_C4	no			28110	112440			
sg129_C4	yes	insol		22944	91776			
sg130_C4	no			27587	110348			
sg131_C4	yes	low yield	too low	27035	108140			
sg132_C4	yes	soluble	s200-14.3 ml	27601	110404	48000	1.7	0.57
sg133_C4	yes	soluble	s200-agg, 13, 14ml	31019.8	124079.2	56000	1.8	0.55
sg134_C4	yes	no		31207				
sg135_C4	yes	soluble	s200-13.5ml	25389	101556	97000	3.8	0.04
sg136_C4	yes	soluble	aggregate					
sg137_C4	no							
sg138_C4	yes	soluble	s200-13.5, 14.75 ml	25427	101708	82000	3.2	0.19
sg139_C4	no							
sg140_C5	yes	none						
sg141_C5	yes	none						
sg142_C5	yes	none						
sg143_C5	yes	soluble		18927	94635	95940	5.1	0.01
sg144_C5	yes	soluble		19119	95595	87400	4.6	0.09
sg145_C5	yes	none			0			
sg146_C5	yes	soluble		18430	92150	92530	5.0	0.00
sg147_C5	yes	soluble		18900	94500	89590	4.7	0.05
sg148_C5	yes	n			0			
sg149_C5	yes	n			0			
sg150_C5	yes	n			0			
sg151_C5	yes	soluble		18439	92195	80390	4.4	0.13
sg152_C5	yes	n			0			
sg153_C5	yes	n			0			
sg154_C5	yes	n			0			
sg155_C5	no				0			
sg156_C5	yes	soluble			0			
sg157_C5	yes	soluble		18850.23	94251.15	95150	5.0	0.01

sg158_C5	no				0			
sg159_C5	no				0			
sg160_C5	no				0			
sg161_C5	yes	soluble		18236	91180	82910	4.5	0.09
sg162_C5	yes	soluble		18485	92425	79590	4.3	0.14
sg163_C5	yes	soluble		17370	86850	91850	5.3	0.06
sg164_C5	yes	soluble		19817	99085	90940	4.6	0.08
sg165_C5	yes	soluble		18934.88	94674.4	73160, 34370	3.9	0.23
sg166_C5	yes	soluble		18636	93180	93800	5.0	0.01
sg167_C5	yes	no			0			
sg168_C5	yes				0			
sg169_C6	yes	low	11.47	18767.778	112606.668			
sg170_C6	yes							
sg171_C6	yes							
sg172_C6	yes							
sg173_C6	yes	soluble	13.31	18264.214	109585.284	102500	5.6	0.06
sg174_C6	yes							
sg175_C6	no							
sg176_C6	yes							
sg177_C6	yes	low						
sg178_C6	yes							
sg179_C6	yes							
sg180_C6	yes							
sg181_C3	yes	no						
sg182_C3	yes	soluble		18319.382	54958.146	50800	2.8	0.08
sg183_C3	yes	soluble	14.9	18940.875	56822.625	52950	2.8	0.07
sg184_C3	no							
sg185_C3	yes	soluble	15.13	16748.407	50245.221	47290	2.8	0.06
sg186_C3	yes	soluble	15.15	18907.59	56722.77	49890	2.6	0.12
sg187_C3	no							
sg188_C3	no	no						
sg189_C3	no							

sg190_C3	yes	no						
sg191_C3	yes	soluble	15.01	18098.987	54296.961	46700	2.6	0.14
sg192_C3	yes	soluble	15.05	18434.578	55303.734	45900	2.5	0.17
sg193_C3	yes	soluble	15.6	17669.821	53009.463	52840, 47770, 44370	2.99, 2.7, 2.5	0.00
sg194_C3	yes	none						
sg195_C3	yes	soluble	15.13,16.08	18033.264	54099.792	52300	2.9	0.03
sg196_C3	yes	soluble	15.01	17985.165	53955.495	61720, 56710, 50120	3.43, 3.15, 2.7	0.05
sg197_C3	yes	soluble	16.1	19163.025	57489.075	34960	1.8	0.39
sg198_C3	yes	low	15.28	17314.372	51943.116			
sg199_C3	yes	soluble	14.14	18146.119	54438.357	76930	4.2	0.41
sg200_C3	yes	soluble	14.96	16960.7	50882.1	46970	2.8	0.08
sg201_C3	yes	none						
sg215_C3	yes							
sg216_C3	yes							
sg217_C3	yes	soluble		19199.52	57598.56	56850	3.0	0.01
sg218_C3	yes				0			
sg219_C3	yes	soluble		19679.07	59037.21	68180	3.5	0.15
sg220_C3	yes	soluble	14.75	20158.84	60476.52			
sg221_C3	yes	soluble	12.19,13.36	19582.93	58748.79	94890, 83540	4.85, 4.27	0.62
sg222_C3	yes			19754.07	59262.21	73060	3.7	0.23
sg223_C3	yes	low	15.71	19714.12	59142.36			
sg224_C4	no			19115.98	76463.92			
sg225_C4	yes	low	15.74	19188.18	76752.72			
sg226_C5	yes	soluble	16.11	16008.33	80041.65	63000	3.9	0.21
sg227_C6	yes	low	16	20077.43	120464.58			
sg228_C6	yes	soluble	15.23	16055.31	96331.86	40840	2.5	0.58
sg229_C6	no							
sg230_C6	yes	soluble	14.36,15.36,16.86	16681.2	100087.2	16290	1.0	0.84
sg231_C6	no							

sg232_C6	yes	soluble	13.67	20080.12	120480.72	92540	4.6	0.23
sg233_C6	yes	soluble	16.12	16290.77	97744.62			
sg234_C5	no							
sg235_C3	yes	none						
sg236_C3	yes	soluble	15.86	17649.55	52948.65	27540	1.6	0.48
sg237_C3	yes	none						
sg238_C3	no							
sg239_C3	no							
sg240_C3	yes	soluble	15.94	17417.44	52252.32	48270	2.8	0.08
sg241_C4	no							
sg242_C4	y	soluble	14.74	15699.05	62796.2	53760	3.4	0.14
sg243_C4	no							
sg244_C4	yes	soluble	14.57	15889.42	63557.68	56130	3.5	0.12
sg245_C4	no							
sg246_C4	yes	soluble	15.19	15711	62844	46090	2.9	0.27
sg247_C4	no							
sg248_C4	yes	soluble	14.83					
sg249_C4	yes	n						
sg250_C4	yes	soluble	14.5	16437.88	65751.52	48220	2.9	0.27
sg251_C4	no							
sg252_C5	no							
sg253_C5	yes	low	16.03	15613.98	78069.9			
sg254_C5	yes	n						
sg255_C5	no							
sg256_C5	no							
sg257_C6	no							
sg258_C6	yes	soluble	15.53	15345.72	76728.6	31860	2.1	0.58
sg259_C6	yes	soluble	14.65	15687	94122	104800	6.7	0.11
sg260_C6	no							
sg261_C6	yes	soluble	16.14	16095.51	96573.06	33890	2.1	0.65
sg262_C6	yes	soluble	13.16,15.3	16084.68	96508.08	96240	6.0	0.00
sg263_C6	yes	n						
sg264_C6	no							
sg265_C6	yes	n						

Appendix 4: Interface design script

Example command:
/software/Rosetta/latest/bin/Rosetta_scripts.hdf5.linuxgccrelease -parser:protocol design_cyclic_oligomers_interface.xml -beta -parser:script_vars sym=${2} -out:prefix ${2}_ -holes:dalphaball DAlphaBall.gcc -multithreading:total_threads 1 -s ${1}

Example of design_cyclic_oligomers_interface.xml:

```
<ROSETTASCRIPTS>
   <SCOREFXNS>
         <ScoreFunction name="sfxn" weights="beta_cart" >
         <Reweight scoretype="cart_bonded" weight="1.5" />
      </ScoreFunction>
      <ScoreFunction name="sfxn_fa_atr" weights="empty" >
         <Reweight scoretype="fa_atr" weight="1" />
      </ScoreFunction>
      <ScoreFunction name="sfxn_design" weights="beta_cart" >
         Reweight scoretype="res_type_constraint" weight="0.5" />
         <Reweight scoretype="cart_bonded" weight="1.5" />
         <Reweight scoretype="aa_composition" weight="1.0" />
         <Reweight scoretype="approximate_buried_unsat_penalty" weight="5.0" />
         <Set approximate_buried_unsat_penalty_assume_const_backbone="true" />
            <Set approximate_buried_unsat_penalty_natural_corrections1="true" />
            <Set approximate_buried_unsat_penalty_hbond_energy_threshold="-1.0" />
            <Set approximate_buried_unsat_penalty_hbond_bonus_cross_chain="-2.5" />
            <Set approximate_buried_unsat_penalty_hbond_bonus_ser_to_helix_bb="0.0" />
      </ScoreFunction>
   </SCOREFXNS>
   <TASKOPERATIONS>
      <SelectBySASA name="PR_monomer_core_sel" mode="sc" state="monomer"
probe_radius="2.2" core_asa="15" surface_asa="15" core="0" boundary="1" surface="1"
verbose="0" />
   </TASKOPERATIONS>
   <RESIDUE_SELECTORS>
      <Layer name="surface" select_core="false" select_boundary="false" select_surface="true"
use_sidechain_neighbors="true"/>
      <Layer name="boundary" select_core="false" select_boundary="true"
select_surface="false" use_sidechain_neighbors="true"/>
         <Layer name="core" select_core="true" select_boundary="false" select_surface="false"
use_sidechain_neighbors="true"/>

         <SecondaryStructure name="entire_loop_DRH" overlap="0" minH="3" minE="2"
include_terminal_loops="false" use_dssp="true" ss="L"/>
```

```
        <Chain name="chainA" chains="1"/>
        <Not name="chainB" selector="chainA"/>
        <Neighborhood name="interface_chA" selector="chainB" distance="10.0" />
        <Neighborhood name="interface_chB" selector="chainA" distance="10.0" />
        <And name="AB_interface" selectors="interface_chA,interface_chB" />
        <Not name="Not_interface" selector="AB_interface" />
        <And name="actual_interface_chA" selectors="AB_interface,chainA" />
        <And name="actual_interface_chB" selectors="AB_interface,chainB" />
        <And name="chainB_not_interface" selectors="Not_interface,chainB" />

        <ResidueName name="pro_and_gly_positions" residue_name3="PRO,GLY" />

        <InterfaceByVector name="interface_by_vector" cb_dist_cut="10" nearby_atom_cut="5.5"
vector_angle_cut="75" vector_dist_cut="9.0" grp1_selector="actual_interface_chA"
            grp2_selector="actual_interface_chB"/>

        <Task name="all_cores" fixed="true" task_operations="PR_monomer_core_sel"
packable="false" designable="false"/>
        <Layer name="all_surface" select_core="false" select_boundary="false"
select_surface="true" use_sidechain_neighbors="true"/>
        <And name="for_polar_boundary" selectors="interface_by_vector">
           <Not selector="surface" />
                <Not selector="entire_loop_DRH" />
                <Not selector="core" />
        </And>
        <And name="for_polar_core" selectors="interface_by_vector">
           <Not selector="surface" />
             <Not selector="entire_loop_DRH" />
             <Not selector="boundary" />
        </And>

        <ResidueName name="vrt_res" residue_name3="VRT" />
    </RESIDUE_SELECTORS>
    <RESIDUE_SELECTORS>

    </RESIDUE_SELECTORS>


    <RESIDUE_SELECTORS>
        <!-- Layer Design -->
        <SecondaryStructure name="sheet" overlap="0" minH="3" minE="2"
include_terminal_loops="false" use_dssp="true" ss="E"/>
        <SecondaryStructure name="entire_loop" overlap="0" minH="3" minE="2"
include_terminal_loops="true" use_dssp="true" ss="L"/>
```

```
      <SecondaryStructure name="entire_helix" overlap="0" minH="3" minE="2"
include_terminal_loops="false" use_dssp="true" ss="H"/>
      <And name="helix_cap" selectors="entire_loop">
         <PrimarySequenceNeighborhood lower="1" upper="0" selector="entire_helix"/>
      </And>
      <And name="helix_start" selectors="entire_helix">
         <PrimarySequenceNeighborhood lower="0" upper="1" selector="helix_cap"/>
      </And>
      <And name="helix" selectors="entire_helix">
         <Not selector="helix_start"/>
      </And>
      <And name="loop" selectors="entire_loop">
         <Not selector="helix_cap"/>
      </And>

   </RESIDUE_SELECTORS>

   <TASKOPERATIONS>
            <DesignRestrictions name="layer_design">
         <Action selector_logic="surface AND helix_start"  aas="DEHKPQR"/>
         <Action selector_logic="surface AND helix"        aas="EHKQR"/>
         <Action selector_logic="surface AND sheet"        aas="EHKNQRST"/>
         <Action selector_logic="surface AND loop"         aas="DEGHKNPQRST"/>
         <Action selector_logic="boundary AND helix_start" aas="ADEFHIKLNPQRSTVWY"/>
            <Action selector_logic="boundary AND helix"       aas="ADEFHIKLNQRSTVWY"/>
         <Action selector_logic="boundary AND sheet"       aas="ADEFHIKLNQRSTVWY"/>
         <Action selector_logic="boundary AND loop"        aas="ADEFGHIKLNPQRSTVWY"/>
         <Action selector_logic="core AND helix_start"     aas="AFILPVWYDNSTH"/>
            <Action selector_logic="core AND helix"           aas="AFILVWYDNSTH"/>
            <Action selector_logic="core AND sheet"           aas="AFILVWYDNSTH"/>
         <Action selector_logic="core AND loop"            aas="AFGILPVWYDNSTH"/>
         <Action selector_logic="helix_cap"                aas="DGNPST"/>
      </DesignRestrictions>
   </TASKOPERATIONS>
   <MOVERS>
      <AddCompositionConstraintMover name="5trp" >
         <Comp entry="PENALTY_DEFINITION;TYPE TRP;ABSOLUTE 0;PENALTIES 0
5;DELTA_START 0;DELTA_END 1;BEFORE_FUNCTION CONSTANT;AFTER_FUNCTION
LINEAR;END_PENALTY_DEFINITION;" />
      </AddCompositionConstraintMover>
      <AddCompositionConstraintMover name="2ala" >
         <Comp entry="PENALTY_DEFINITION;TYPE ALA;ABSOLUTE 0;PENALTIES 0
2;DELTA_START 0;DELTA_END 1;BEFORE_FUNCTION CONSTANT;AFTER_FUNCTION
LINEAR;END_PENALTY_DEFINITION;" />
```

```
    </AddCompositionConstraintMover>


    <AddCompositionConstraintMover name="50_percent_polar_boundary_max"
selector="for_polar_boundary">
      <Comp entry="PENALTY_DEFINITION;TYPE ASP HIS ASN GLN SER THR
TYR;FRACT_DELTA_START 0;FRACT_DELTA_END 0.01;PENALTIES 0 0.1 ;FRACTION
0.50;BEFORE_FUNCTION CONSTANT;AFTER_FUNCTION
QUADRATIC;END_PENALTY_DEFINITION" />
    </AddCompositionConstraintMover>
    <AddCompositionConstraintMover name="30_percent_polar_core_max"
selector="for_polar_core">
      <Comp entry="PENALTY_DEFINITION;TYPE ASP HIS ASN GLN SER THR
TYR;FRACT_DELTA_START 0;FRACT_DELTA_END 0.01;PENALTIES 0 0.1 ;FRACTION
0.30;BEFORE_FUNCTION CONSTANT;AFTER_FUNCTION
QUADRATIC;END_PENALTY_DEFINITION" />
    </AddCompositionConstraintMover>

    <AddCompositionConstraintMover name="30_percent_polar_boundary_min"
selector="for_polar_boundary">
      <Comp entry="PENALTY_DEFINITION;TYPE ASP HIS ASN GLN SER THR
TYR;FRACT_DELTA_START -0.01;FRACT_DELTA_END 0;PENALTIES 0.1 0 ;FRACTION
0.30;BEFORE_FUNCTION QUADRATIC;AFTER_FUNCTION
CONSTANT;END_PENALTY_DEFINITION" />
    </AddCompositionConstraintMover>
    <AddCompositionConstraintMover name="20_percent_polar_core_min"
selector="for_polar_core">
      <Comp entry="PENALTY_DEFINITION;TYPE ASP HIS ASN GLN SER THR
TYR;FRACT_DELTA_START -0.01;FRACT_DELTA_END 0;PENALTIES 0.1 0 ;FRACTION
0.20;BEFORE_FUNCTION QUADRATIC;AFTER_FUNCTION
CONSTANT;END_PENALTY_DEFINITION" />
    </AddCompositionConstraintMover>
    <DeleteRegionMover name="delete_vrt" residue_selector="vrt_res"/>
  </MOVERS>
  <TASKOPERATIONS>
    <IncludeCurrent name="current" />
    <LimitAromaChi2 name="limitchi2" chi2max="110" chi2min="70" include_trp="True" />
    <ExtraRotamersGeneric name="ex1_ex2" ex1="1" ex2aro="1" />

    <OperateOnResidueSubset name="restrict_to_interface" selector="Not_interface">
      <PreventRepackingRLT/>
    </OperateOnResidueSubset>

    <DisallowIfNonnative name="disallow_GLY" resnum="0" disallow_aas="G" />
```

```
      <DisallowIfNonnative name="disallow_PRO" resnum="0" disallow_aas="P" />

         <OperateOnResidueSubset name="restrict_PRO_GLY"
selector="pro_and_gly_positions">
                  <RestrictToRepackingRLT/>
         </OperateOnResidueSubset>

      <SelectBySASA name="PR_monomer_core" mode="sc" state="monomer"
probe_radius="2.2" core_asa="10" surface_asa="10" core="0" boundary="1" surface="1"
verbose="0" />

      <OperateOnResidueSubset name="restrict_loops2repacking_DRH"
selector="entire_loop_DRH">
      <RestrictToRepackingRLT/>
      </OperateOnResidueSubset>
   </TASKOPERATIONS>
   <MOVERS>

         StructProfileMover name="genProfile" add_csts_to_pose="1"
consider_topN_frags="100" eliminate_background="0"
         ignore_terminal_residue="1" only_loops="0" burialWt="0" RMSthreshold="0.6"
residue_selector="chainA" />

      <ClearConstraintsMover name="clear_constraints" />

      <SavePoseMover name="save_start" restore_pose="0" reference_name="pose_start" />

      <DeleteRegionMover name="chain1onlypre" residue_selector="chainB"/>
      <ScoreMover name="scorepose" scorefxn="sfxn" verbose="false" />
      <ParsedProtocol name="chain1only">
         <Add mover="chain1onlypre" />
         <Add mover="scorepose" />
      </ParsedProtocol>
   </MOVERS>
   <FILTERS>

      <Sasa name="interface_buried_sasa" confidence="0" />

      <BuriedUnsatHbonds name="vbuns5.5_heavy_ball_1.1" use_reporter_behavior="true"
report_all_heavy_atom_unsats="true" scorefxn="sfxn" residue_selector="AB_interface"
         ignore_surface_res="false" print_out_info_to_pdb="true" confidence="0"
use_ddG_style="false" dalphaball_sasa="true" probe_radius="1.1"
         atomic_depth_selection="5.5" burial_cutoff="1000" burial_cutoff_apo="0.2" />
```

```
      <BuriedUnsatHbonds name="sbuns5.5_heavy_ball_1.1" use_reporter_behavior="true"
report_all_heavy_atom_unsats="true" scorefxn="sfxn" residue_selector="AB_interface"
          ignore_surface_res="false" print_out_info_to_pdb="true" confidence="0"
use_ddG_style="false" burial_cutoff="0.01" dalphaball_sasa="true" probe_radius="1.1"
          atomic_depth_selection="5.5" atomic_depth_deeper_than="false" />

      <ResidueCount name="AlaCount_MBF" residue_types="ALA" max_residue_count="6"
confidence="0"/>
      <MoveBeforeFilter name="AlaCount" mover="chain1only" filter="AlaCount_MBF"
confidence="0" />

      <ScoreType name="total_score_MBF" scorefxn="sfxn" score_type="total_score"
threshold="0" confidence="0" />
      <MoveBeforeFilter name="total_score_monomer" mover="chain1only"
filter="total_score_MBF" confidence="0" />
      <ScoreType name="p_aa_pp_MBF" scorefxn="sfxn" score_type="p_aa_pp" threshold="0"
confidence="0" />
      <MoveBeforeFilter name="p_aa_pp_monomer" mover="chain1only" filter="p_aa_pp_MBF"
confidence="0" />
      <ResidueCount name="res_count_MBF" max_residue_count="9999" confidence="0"/>
      <MoveBeforeFilter name="res_count_monomer" mover="chain1only"
filter="res_count_MBF" confidence="0" />


      <CalculatorFilter name="score_per_res_monomer" equation="total_score_monomer / res"
threshold="-3.5" confidence="0">
          <Var name="total_score_monomer" filter="total_score_monomer"/>
          <Var name="res" filter="res_count_monomer"/>
          </CalculatorFilter>

      <CalculatorFilter name="score_per_res_complex" equation="total_score_monomer / res"
threshold="-3.5" confidence="0">
          <Var name="total_score_monomer" filter="total_score_MBF"/>
          <Var name="res" filter="res_count_MBF"/>
      </CalculatorFilter>

      <CalculatorFilter name="p_aa_pp_per_res" equation="p_aa_pp_monomer / res"
threshold="-3.5" confidence="0">
          <Var name="p_aa_pp_monomer" filter="p_aa_pp_monomer"/>
          <Var name="res" filter="res_count_monomer"/>
      </CalculatorFilter>

      <ScoreType name="fa_atr" scorefxn="sfxn" threshold="0" score_type="fa_atr"
confidence="0"/>
```

```
      <CalculatorFilter name="fa_atr_per_res" equation="fa_atr_score / res" threshold="-5.2"
confidence="0">
         <Var name="fa_atr_score" filter="fa_atr"/>
         <Var name="res" filter="res_count_MBF"/>
      </CalculatorFilter>

      <Rmsd name="rmsd_to_start" reference_name="pose_start" confidence="0" />

         <ShapeComplementarity name="interface_sc" min_sc="0.5" min_interface="0"
verbose="0" quick="0" residue_selector1="chainA" residue_selector2="chainB"
         write_int_area="1" write_median_dist="1" max_median_dist="1000" confidence="0"/>
      <SSShapeComplementarity name="ss_sc_complex" verbose="0" confidence="0" />
      <MoveBeforeFilter name="ss_sc_monomer" mover="chain1only" filter="ss_sc_complex"
confidence="0" />

      <PackStat name="packstat_complex" repeats="20" confidence="0" />
      <MoveBeforeFilter name="packstat_monomer" mover="chain1only"
filter="packstat_complex" confidence="0" />

      <CavityVolume name="cavity_complex" confidence="0"/>
      <MoveBeforeFilter name="cavity_monomer" mover="chain1only" filter="cavity_complex"
confidence="0" />

      <Ddg name="ddg_norepack"  threshold="-10" chain_num="2" repeats="1" repack="0"
confidence="0" scorefxn="sfxn" extreme_value_removal="0" />

      <SSPrediction name="mismatch_probability" confidence="0"
cmd="/software/psipred4/runpsipred_single" use_probability="1" mismatch_probability="1"
use_svm="0" />

      <ContactMolecularSurface name="contact_molecular_surface" distance_weight="0.5"
target_selector="chainA" binder_selector="chainB" confidence="0" />

      <BuriedSurfaceArea name="buried_npsa_FAMILYVW" select_only_FAMILYVW="True"
atom_mode="hydrophobic_atoms" confidence="0.0" />
      <MoveBeforeFilter name="buried_npsa_FAMILYVW_monomer" mover="chain1only"
filter="buried_npsa_FAMILYVW" confidence="0" />

      <Time name="timed"/>


   </FILTERS>
   <MOVERS>
      <SwitchChainOrder name="monomer_ASU" chain_order="1" />
```

```
        <SetupForSymmetry name="setup_sym"
definition="/software/Rosetta/main/database/symmetry/cyclic/%%sym%%_Z.sym"
preserve_datacache="false" />
        <ExtractAsymmetricPose name="extract_asp" clear_sym_def="true"/>

        <AddConstraintsToCurrentConformationMover name="CAcsts" cst_weight="1.0"
use_distance_cst="False" coord_dev="1.0" bound_width="0.5" CA_only="False"
bb_only="True"/>
        <VirtualRoot name="add_vrt_root" removable="true" remove="false" />
        <VirtualRoot name="rm_vrt_root" removable="true" remove="true" />

        <FastDesign name="FastDesign" scorefxn="sfxn_design" repeats="1" batch="false"
ramp_down_constraints="false"
            cartesian="false" bondangle="true" bondlength="true"
min_type="lbfgs_armijo_nonmonotone" relaxscript="InterfaceDesign2019"

task_operations="current,limitchi2,ex1_ex2,restrict_to_interface,disallow_GLY,disallow_PRO,PR
_monomer_core,restrict_PRO_GLY,restrict_loops2repacking_DRH,layer_design">
            <MoveMap name="MM" >
               <ResidueSelector selector="Not_interface" chi="false" bb="false" bondangle="false"
bondlength="false"/>
            </MoveMap>
        </FastDesign>

        <FastRelax name="FastRelax1" scorefxn="sfxn_design" repeats="1" batch="false"
ramp_down_constraints="false"
            cartesian="false" bondangle="true" bondlength="true"
min_type="lbfgs_armijo_nonmonotone" relaxscript="InterfaceRelax2019"
            task_operations="ex1_ex2,limitchi2,restrict_to_interface">
        </FastRelax>
        <FastRelax name="FastRelax2" scorefxn="sfxn" repeats="1" batch="false"
ramp_down_constraints="false"
            cartesian="false" bondangle="true" bondlength="true"
min_type="lbfgs_armijo_nonmonotone" relaxscript="InterfaceRelax2019"
            task_operations="ex1_ex2,limitchi2">
        </FastRelax>
    </MOVERS>
    <APPLY_TO_POSE>
    </APPLY_TO_POSE>
    <PROTOCOLS>
            <Add filter="timed" />

            <Add mover="monomer_ASU"/>
            <Add mover="setup_sym"/>
```

```
    <Add mover="save_start" />

  Add mover="genProfile" />
    <Add mover="5trp" />
    <Add mover="2ala"/>
    <Add mover="30_percent_polar_boundary_min" />
    <Add mover="50_percent_polar_boundary_max" />
    <Add mover="20_percent_polar_core_min" />
    <Add mover="30_percent_polar_core_max" />
    Add mover="CAcsts"/>
    Add mover="add_vrt_root"/>
    <Add mover="FastDesign" />
    <Add mover="FastRelax1" />
    <Add mover="FastDesign" />
    <Add mover="FastRelax2" />
    Add mover="rm_vrt_root" />

    <Add mover="extract_asp" />
    <Add mover="delete_vrt"/>

  <Add mover="clear_constraints" />

    <Add filter_name="interface_buried_sasa" />
    <Add filter="vbuns5.5_heavy_ball_1.1" />
    <Add filter="sbuns5.5_heavy_ball_1.1" />

    <Add filter="score_per_res_monomer" />
    <Add filter="score_per_res_complex" />
    <Add filter="AlaCount" />
    <Add filter="res_count_monomer" />
    <Add filter="rmsd_to_start" />
    <Add filter="p_aa_pp_per_res" />
    <Add filter="fa_atr_per_res" />
    <Add filter="ss_sc_monomer" />
    <Add filter="ss_sc_complex" />
    <Add filter="interface_sc" />
    <Add filter="packstat_monomer" />
    <Add filter="packstat_complex" />
    <Add filter="cavity_monomer" />
    <Add filter="cavity_complex" />

    <Add filter_name="ddg_norepack" />
    <Add filter_name="mismatch_probability" />
```

```
        <Add filter_name="contact_molecular_surface" />
        <Add filter="buried_npsa_FAMILYVW_monomer" />

        <Add filter="timed" />
    </PROTOCOLS>
    <OUTPUT />
</ROSETTASCRIPTS>
```

Appendix 5: Surface redesign script

Example command:
/software/Rosetta/latest/bin/Rosetta_scripts.hdf5.linuxgccrelease -database /software/Rosetta/main/database/ -parser:protocol design_cyclic_oligomers_surface.xml -indexed_structure_store:fragment_store ss_grouped_vall_all.h5 -parser:script_vars sym=${2} -holes:dalphaball DAlphaBall.gcc -multithreading:total_threads 1 -s ${1} -out:path ${3}

Example of design_cyclic_oligomers_surface.xml:

```
<ROSETTASCRIPTS>
    <SCOREFXNS>
        <ScoreFunction name="sfxn" weights="beta_cart" >
        <Reweight scoretype="cart_bonded" weight="1.5" />
      </ScoreFunction>
      <ScoreFunction name="sfxn_fa_atr" weights="empty" >
        <Reweight scoretype="fa_atr" weight="1" />
      </ScoreFunction>
        <ScoreFunction name="sfxn_design" weights="beta_cart" >
        <Reweight scoretype="netcharge" weight="1.0" />
          <Reweight scoretype="res_type_constraint" weight="0.5" />
        <Reweight scoretype="cart_bonded" weight="1.5" />
        <Reweight scoretype="approximate_buried_unsat_penalty" weight="5.0" />
        <Set approximate_buried_unsat_penalty_assume_const_backbone="true" />
          <Set approximate_buried_unsat_penalty_natural_corrections1="true" />
          <Set approximate_buried_unsat_penalty_hbond_energy_threshold="-1.0" />
          <Set approximate_buried_unsat_penalty_hbond_bonus_cross_chain="-2.5" />
          <Set approximate_buried_unsat_penalty_hbond_bonus_ser_to_helix_bb="0.0" />
          <Reweight scoretype="sap_constraint" weight="1.0" />
      </ScoreFunction>
    </SCOREFXNS>
    <TASKOPERATIONS>
      <SelectBySASA name="PR_monomer_core_sel" mode="sc" state="monomer"
probe_radius="2.2" core_asa="15" surface_asa="15" core="0" boundary="1" surface="1"
verbose="0" />
```

```
   </TASKOPERATIONS>
   <RESIDUE_SELECTORS>
      <Layer name="surface" select_core="false" select_boundary="false" select_surface="true"
use_sidechain_neighbors="true"/>
      <Layer name="boundary" select_core="false" select_boundary="true"
select_surface="false" use_sidechain_neighbors="true"/>
         <Layer name="core" select_core="true" select_boundary="false" select_surface="false"
use_sidechain_neighbors="true"/>
      <Layer name="surface_sasa" select_core="false" select_boundary="false"
select_surface="true"
         ball_radius="2.0" use_sidechain_neighbors="false" core_cutoff="20.0"
surface_cutoff="30.0" />
         <Layer name="core_sasa" select_core="true" select_boundary="false"
select_surface="false"
         ball_radius="2.0" use_sidechain_neighbors="false" core_cutoff="20.0"
surface_cutoff="30.0" />

         <Or name="surface_all" selectors="surface,surface_sasa"/>
         <Not name="not_surface" selector="surface_all"/>

      <Or name="core_all" selectors="core,core_sasa"/>

         <SecondaryStructure name="entire_loop_DRH" overlap="0" minH="3" minE="2"
include_terminal_loops="false" use_dssp="true" ss="L"/>

      <Chain name="chainA" chains="1"/>
      <Not name="chainB" selector="chainA"/>
      <Neighborhood name="interface_chA" selector="chainB" distance="10.0" />
      <Neighborhood name="interface_chB" selector="chainA" distance="10.0" />
      <And name="AB_interface" selectors="interface_chA,interface_chB" />
      <Not name="Not_interface" selector="AB_interface" />
      <And name="actual_interface_chA" selectors="AB_interface,chainA" />
      <And name="actual_interface_chB" selectors="AB_interface,chainB" />
      <And name="chainB_not_interface" selectors="Not_interface,chainB" />

      <ResidueName name="pro_and_gly_positions" residue_name3="PRO,GLY" />

      <InterfaceByVector name="interface_by_vector" cb_dist_cut="10" nearby_atom_cut="5.5"
vector_angle_cut="75" vector_dist_cut="9.0" grp1_selector="actual_interface_chA"
         grp2_selector="actual_interface_chB"/>

      <And name="for_polar_boundary" selectors="interface_by_vector">
         <Not selector="surface" />
            <Not selector="entire_loop_DRH" />
```

```
            <Not selector="core" />
        </And>
        <And name="for_polar_core" selectors="interface_by_vector">
            <Not selector="surface" />
              <Not selector="entire_loop_DRH" />
              <Not selector="boundary" />
        </And>

            <ResidueName name="vrt_res" residue_name3="VRT" />

            <StoredResidueSubset name="stored_bad_sap" subset_name="bad_sap_stored" />
        <Not name="not_stored_bad_sap" selector="stored_bad_sap" />
    </RESIDUE_SELECTORS>

    <TASKOPERATIONS>

        <OperateOnResidueSubset name="restrict_not_bad_sap"
selector="not_stored_bad_sap">
            <RestrictToRepackingRLT/>
        </OperateOnResidueSubset>

    </TASKOPERATIONS>

    <MOVERS>

        <AddSapConstraintMover name="add_sap" speed="lightning" sap_goal="0"
penalty_per_sap="10" score_selector="chainA" sap_calculate_selector="chainA"
sap_lb_goal="0" packing_correction="0" />

    </MOVERS>
    <SIMPLE_METRICS>

        <SapScoreMetric name="my_sap_score" />

        <PerResidueSapScoreMetric name="my_per_res_sap" />

    </SIMPLE_METRICS>
    <RESIDUE_SELECTORS>

        <SimpleMetricSelector name="bad_sap" metric="my_per_res_sap" lower_bound="1.4" />

    </RESIDUE_SELECTORS>
    <MOVERS>
```

```
      <StoreResidueSubset name="store_bad_sap" subset_name="bad_sap_stored"
residue_selector="bad_sap" overwrite="1" />

   </MOVERS>
   <FILTERS>

      <TrueFilter name="my_true_filter" />

   </FILTERS>

   <RESIDUE_SELECTORS>
      <!-- Layer Design -->
      <SecondaryStructure name="sheet" overlap="0" minH="3" minE="2"
include_terminal_loops="false" use_dssp="true" ss="E"/>
      <SecondaryStructure name="entire_loop" overlap="0" minH="3" minE="2"
include_terminal_loops="true" use_dssp="true" ss="L"/>
      <SecondaryStructure name="entire_helix" overlap="0" minH="3" minE="2"
include_terminal_loops="false" use_dssp="true" ss="H"/>
      <And name="helix_cap" selectors="entire_loop">
         <PrimarySequenceNeighborhood lower="1" upper="0" selector="entire_helix"/>
      </And>
      <And name="helix_start" selectors="entire_helix">
         <PrimarySequenceNeighborhood lower="0" upper="1" selector="helix_cap"/>
      </And>
      <And name="helix" selectors="entire_helix">
         <Not selector="helix_start"/>
      </And>
      <And name="loop" selectors="entire_loop">
         <Not selector="helix_cap"/>
      </And>

   </RESIDUE_SELECTORS>

   <TASKOPERATIONS>
            <DesignRestrictions name="layer_design">
         <Action selector_logic="surface AND helix_start"  aas="DEHKPQR"/>
         <Action selector_logic="surface AND helix"        aas="EHKQR"/>
         <Action selector_logic="surface AND sheet"        aas="EHKNQRST"/>
         <Action selector_logic="surface AND loop"         aas="DEGHKNPQRST"/>
         <Action selector_logic="boundary AND helix_start" aas="ADEFHIKLNPQRSTVWY"/>
            <Action selector_logic="boundary AND helix"       aas="ADEFHIKLNQRSTVWY"/>
         <Action selector_logic="boundary AND sheet"       aas="ADEFHIKLNQRSTVWY"/>
         <Action selector_logic="boundary AND loop"        aas="ADEFGHIKLNPQRSTVWY"/>
         <Action selector_logic="core AND helix_start"     aas="AFILPVWYDNSTH"/>
```

```
            <Action selector_logic="core AND helix"          aas="AFILVWYDNSTH"/>
            <Action selector_logic="core AND sheet"          aas="AFILVWYDNSTH"/>
          <Action selector_logic="core AND loop"          aas="AFGILPVWYDNSTH"/>
          <Action selector_logic="helix_cap"              aas="DGNPST"/>
      </DesignRestrictions>
   </TASKOPERATIONS>
   <MOVERS>
      <AddCompositionConstraintMover name="5trp" >
         <Comp entry="PENALTY_DEFINITION;TYPE TRP;ABSOLUTE 0;PENALTIES 0
5;DELTA_START 0;DELTA_END 1;BEFORE_FUNCTION CONSTANT;AFTER_FUNCTION
LINEAR;END_PENALTY_DEFINITION;" />
      </AddCompositionConstraintMover>
      <AddCompositionConstraintMover name="2ala" >
         <Comp entry="PENALTY_DEFINITION;TYPE ALA;ABSOLUTE 0;PENALTIES 0
2;DELTA_START 0;DELTA_END 1;BEFORE_FUNCTION CONSTANT;AFTER_FUNCTION
LINEAR;END_PENALTY_DEFINITION;" />
      </AddCompositionConstraintMover>

      <DeleteRegionMover name="delete_vrt" residue_selector="vrt_res"/>
   </MOVERS>
   <TASKOPERATIONS>
      <IncludeCurrent name="current" />
      <LimitAromaChi2 name="limitchi2" chi2max="110" chi2min="70" include_trp="True" />
      <ExtraRotamersGeneric name="ex1_ex2" ex1="1" ex2aro="1" />

      <OperateOnResidueSubset name="restrict_to_interface" selector="Not_interface">
         <PreventRepackingRLT/>
      </OperateOnResidueSubset>

      <DisallowIfNonnative name="disallow_GLY" resnum="0" disallow_aas="G" />
      <DisallowIfNonnative name="disallow_PRO" resnum="0" disallow_aas="P" />

         <OperateOnResidueSubset name="restrict_PRO_GLY"
selector="pro_and_gly_positions">
         <RestrictToRepackingRLT/>
         </OperateOnResidueSubset>

      <SelectBySASA name="PR_monomer_core" mode="sc" state="monomer"
probe_radius="2.2" core_asa="10" surface_asa="10" core="0" boundary="1" surface="1"
verbose="0" />

      <OperateOnResidueSubset name="restrict_loops2repacking_DRH"
selector="entire_loop_DRH">
         <RestrictToRepackingRLT/>
```

```
      </OperateOnResidueSubset>

      <OperateOnResidueSubset name="only_surface" selector="not_surface">
         <RestrictToRepackingRLT/>
      </OperateOnResidueSubset>
      <OperateOnResidueSubset name="not_core" selector="core_all">
         <PreventRepackingRLT/>
      </OperateOnResidueSubset>

   </TASKOPERATIONS>
   <MOVERS>

         StructProfileMover name="genProfile" add_csts_to_pose="1"
consider_topN_frags="100" eliminate_background="0"
         ignore_terminal_residue="1" only_loops="0" burialWt="0" RMSthreshold="0.6"
residue_selector="chainA" />

      <ClearConstraintsMover name="clear_constraints" />

      <SavePoseMover name="save_start" restore_pose="0" reference_name="pose_start" />

      <DeleteRegionMover name="chain1onlypre" residue_selector="chainB"/>
      <ScoreMover name="scorepose" scorefxn="sfxn" verbose="false" />
      <ParsedProtocol name="chain1only">
         <Add mover="chain1onlypre" />
         <Add mover="scorepose" />
      </ParsedProtocol>
   </MOVERS>
   <FILTERS>

      <Sasa name="interface_buried_sasa" confidence="0" />

      <BuriedUnsatHbonds name="vbuns5.5_heavy_ball_1.1" use_reporter_behavior="true"
report_all_heavy_atom_unsats="true" scorefxn="sfxn" residue_selector="AB_interface"
         ignore_surface_res="false" print_out_info_to_pdb="true" confidence="0"
use_ddG_style="false" dalphaball_sasa="true" probe_radius="1.1"
         atomic_depth_selection="5.5" burial_cutoff="1000" burial_cutoff_apo="0.2" />
      <BuriedUnsatHbonds name="sbuns5.5_heavy_ball_1.1" use_reporter_behavior="true"
report_all_heavy_atom_unsats="true" scorefxn="sfxn" residue_selector="AB_interface"
         ignore_surface_res="false" print_out_info_to_pdb="true" confidence="0"
use_ddG_style="false" burial_cutoff="0.01" dalphaball_sasa="true" probe_radius="1.1"
         atomic_depth_selection="5.5" atomic_depth_deeper_than="false" />
```

```
    <ResidueCount name="AlaCount_MBF" residue_types="ALA" max_residue_count="6"
confidence="0"/>
    <MoveBeforeFilter name="AlaCount" mover="chain1only" filter="AlaCount_MBF"
confidence="0" />

    <ScoreType name="total_score_MBF" scorefxn="sfxn" score_type="total_score"
threshold="0" confidence="0" />
    <MoveBeforeFilter name="total_score_monomer" mover="chain1only"
filter="total_score_MBF" confidence="0" />
    <ScoreType name="p_aa_pp_MBF" scorefxn="sfxn" score_type="p_aa_pp" threshold="0"
confidence="0" />
    <MoveBeforeFilter name="p_aa_pp_monomer" mover="chain1only" filter="p_aa_pp_MBF"
confidence="0" />
    <ResidueCount name="res_count_MBF" max_residue_count="9999" confidence="0"/>
    <MoveBeforeFilter name="res_count_monomer" mover="chain1only"
filter="res_count_MBF" confidence="0" />


    <CalculatorFilter name="score_per_res_monomer" equation="total_score_monomer / res"
threshold="-3.5" confidence="0">
      <Var name="total_score_monomer" filter="total_score_monomer"/>
      <Var name="res" filter="res_count_monomer"/>
      </CalculatorFilter>

    <CalculatorFilter name="score_per_res_complex" equation="total_score_monomer / res"
threshold="-3.5" confidence="0">
      <Var name="total_score_monomer" filter="total_score_MBF"/>
      <Var name="res" filter="res_count_MBF"/>
    </CalculatorFilter>

    <CalculatorFilter name="p_aa_pp_per_res" equation="p_aa_pp_monomer / res"
threshold="-3.5" confidence="0">
      <Var name="p_aa_pp_monomer" filter="p_aa_pp_monomer"/>
      <Var name="res" filter="res_count_monomer"/>
    </CalculatorFilter>

    <ScoreType name="fa_atr" scorefxn="sfxn" threshold="0" score_type="fa_atr"
confidence="0"/>
    <CalculatorFilter name="fa_atr_per_res" equation="fa_atr_score / res" threshold="-5.2"
confidence="0">
      <Var name="fa_atr_score" filter="fa_atr"/>
      <Var name="res" filter="res_count_MBF"/>
    </CalculatorFilter>
```

```
        <Rmsd name="rmsd_to_start" reference_name="pose_start" confidence="0" />

          <ShapeComplementarity name="interface_sc" min_sc="0.5" min_interface="0"
verbose="0" quick="0" residue_selector1="chainA" residue_selector2="chainB"
          write_int_area="1" write_median_dist="1" max_median_dist="1000" confidence="0"/>
        <SSShapeComplementarity name="ss_sc_complex" verbose="0" confidence="0" />
        <MoveBeforeFilter name="ss_sc_monomer" mover="chain1only" filter="ss_sc_complex"
confidence="0" />

        <PackStat name="packstat_complex" repeats="20" confidence="0" />
        <MoveBeforeFilter name="packstat_monomer" mover="chain1only"
filter="packstat_complex" confidence="0" />

        <CavityVolume name="cavity_complex" confidence="0"/>
        <MoveBeforeFilter name="cavity_monomer" mover="chain1only" filter="cavity_complex"
confidence="0" />

        <Ddg name="ddg_norepack"  threshold="-10" chain_num="2" repeats="1" repack="0"
confidence="0" scorefxn="sfxn" extreme_value_removal="0" />

        <SSPrediction name="mismatch_probability" confidence="0"
cmd="/software/psipred4/runpsipred_single" use_probability="1" mismatch_probability="1"
use_svm="0" />

        <ContactMolecularSurface name="contact_molecular_surface" distance_weight="0.5"
target_selector="chainA" binder_selector="chainB" confidence="0" />

        <BuriedSurfaceArea name="buried_npsa_FAMILYVW" select_only_FAMILYVW="True"
atom_mode="hydrophobic_atoms" confidence="0.0" />
        <MoveBeforeFilter name="buried_npsa_FAMILYVW_monomer" mover="chain1only"
filter="buried_npsa_FAMILYVW" confidence="0" />

        <Time name="timed"/>

        <ExposedHydrophobics name="exposed_hydrophobics_20" sasa_cutoff="20"
threshold="-1" confidence="0"/> #for full oligomer. only for surface as defined by sasa.
        <ExposedHydrophobics name="exposed_hydrophobics_25" sasa_cutoff="25"
threshold="-1" confidence="0"/>
        <ExposedHydrophobics name="exposed_hydrophobics_30" sasa_cutoff="30"
threshold="-1" confidence="0"/>
        <ExposedHydrophobics name="exposed_hydrophobics_35" sasa_cutoff="35"
threshold="-1" confidence="0"/>
        <ExposedHydrophobics name="exposed_hydrophobics_40" sasa_cutoff="40"
threshold="-1" confidence="0"/>
```

```
      <LongestContinuousApolarSegment name="longest_hydrophobic_stretch"
exclude_chain_termini="false" confidence="0" /> #for long hydrophobic helices

         <ResidueCount name="interface_hydrophobic_percent"
residue_types="ALA,ILE,LEU,VAL,PHE,TRP,MET,TYR" residue_selector="AB_interface"
count_as_percentage="1" confidence="0"/> #for interface hydrophobics

      <SecretionPredictionFilter name="transmembrane_helix" threshold="0" confidence="0" />
#calculates the likleyhood of a membrane helix

   </FILTERS>
   <MOVERS>
      <SwitchChainOrder name="monomer_ASU" chain_order="1" />
      <SetupForSymmetry name="setup_sym"
definition="/software/Rosetta/main/database/symmetry/cyclic/%%sym%%_Z.sym"
preserve_datacache="false" />
      <ExtractAsymmetricPose name="extract_asp" clear_sym_def="true"/>

      <AddConstraintsToCurrentConformationMover name="CAcsts" cst_weight="1.0"
use_distance_cst="False" coord_dev="1.0" bound_width="0.5" CA_only="False"
bb_only="True"/>
      <VirtualRoot name="add_vrt_root" removable="true" remove="false" />
      <VirtualRoot name="rm_vrt_root" removable="true" remove="true" />

      <FastDesign name="FastDesign" scorefxn="sfxn_design" repeats="1" batch="false"
ramp_down_constraints="false"
         cartesian="false" bondangle="true" bondlength="true"
min_type="lbfgs_armijo_nonmonotone" relaxscript="Rosettacon2018"

task_operations="limitchi2,ex1_ex2,only_surface,disallow_GLY,disallow_PRO,PR_monomer_co
re,not_core,restrict_PRO_GLY,layer_design,restrict_not_bad_sap">
      </FastDesign>

      <FastRelax name="FastRelax" scorefxn="sfxn" repeats="1" batch="false"
ramp_down_constraints="false"
         cartesian="false" bondangle="true" bondlength="true"
min_type="lbfgs_armijo_nonmonotone" relaxscript="Rosettacon2018"
         task_operations="ex1_ex2,limitchi2">
      </FastRelax>
      #InterfaceDesign2019
      <AddNetChargeConstraintMover name="netcharge_cst"
filename="/home/drhicks1/for_/for_isabelle/surface_design/netcharge" selector="chainA" />
   </MOVERS>
```

```
<APPLY_TO_POSE>
</APPLY_TO_POSE>
<PROTOCOLS>
      <Add filter="timed" />

      <Add mover="monomer_ASU"/>
      <Add mover="setup_sym"/>
      <Add mover="save_start" />

      <Add metrics="my_sap_score" labels="starting_sap" />

   <Add mover="store_bad_sap" />

   <Add mover="add_sap" />

      Add mover="genProfile" />
      Add mover="5trp" />
      Add mover="2ala"/>
      <Add mover="netcharge_cst"/>
      <Add mover="FastDesign" />
      <Add mover="FastRelax" />

      <Add mover="extract_asp" />
      <Add mover="delete_vrt"/>

   <Add mover="clear_constraints" />

      <Add filter_name="interface_buried_sasa" />
      <Add filter="vbuns5.5_heavy_ball_1.1" />
      <Add filter="sbuns5.5_heavy_ball_1.1" />

      <Add filter="score_per_res_monomer" />
      <Add filter="score_per_res_complex" />
      <Add filter="AlaCount" />
   <Add filter="res_count_MBF" />
      <Add filter="res_count_monomer" />
      <Add filter="rmsd_to_start" />
      <Add filter="p_aa_pp_per_res" />
      <Add filter="fa_atr_per_res" />
      <Add filter="ss_sc_monomer" />
      <Add filter="ss_sc_complex" />
      <Add filter="interface_sc" />
      <Add filter="packstat_monomer" />
      <Add filter="packstat_complex" />
```

```
        <Add filter="cavity_monomer" />
        <Add filter="cavity_complex" />

        <Add filter_name="ddg_norepack" />
        <Add filter_name="mismatch_probability" />
        <Add filter_name="contact_molecular_surface" />
        <Add filter="buried_npsa_FAMILYVW_monomer" />

      <Add filter="exposed_hydrophobics_20"/>
      <Add filter="exposed_hydrophobics_25"/>
      <Add filter="exposed_hydrophobics_30"/>
      <Add filter="exposed_hydrophobics_35"/>
      <Add filter="exposed_hydrophobics_40"/>
            <Add filter="longest_hydrophobic_stretch"/>
      <Add filter="interface_hydrophobic_percent"/>
            <Add filter="transmembrane_helix"/>

        <Add metrics="my_sap_score" />

        <Add filter="timed" />
  </PROTOCOLS>
  <OUTPUT />
</ROSETTASCRIPTS>
```

Appendix 6: Pocket Determination Script

```
#import time
import argparse
import os
import sys
import numpy as np


from numba import njit
import voxel_array
import npose_util as nu #pip install npose if this does not work

from scipy.spatial import ConvexHull
from scipy.spatial import Delaunay

import atomic_depth


parser = argparse.ArgumentParser(description='arguments')
parser.add_argument('--rolling_ball', type=bool, default=False,
```

```
    help='bool; Use the rolling_ball instead of convex hull for tigher surface.')
parser.add_argument('--rolling_ball_radius', type=bool, default=20.0,
    help='float; radius for rolling ball method')
parser.add_argument('--debug_dumping', type=bool, default=False,
    help='bool; Dump some debugging pdbs')
parser.add_argument('--pocket_features', type=bool, default=True,
    help='bool; Write pocket features to scorefile? They are calculated regardless.')
parser.add_argument('--pocket_residues', type=bool, default=False,
    help='bool; Do you want do write out pocket residues? Might be a major slow down...')
parser.add_argument('--oligomer_mode', type=bool, default=False,
    help='bool; For cyclic oligomer asymmetric pockets.')
parser.add_argument('--resl', type=float, default=0.5,
    help='float; Resolution of 0.5 is suggested; Use 1.0 for more speed.')
parser.add_argument('--pdbs',nargs='+', default=None,
    help='list; list of pdbs ie *pdb')
parser.add_argument('--pdbs_list_file',default=None,
    help='str; list file containing pdb paths')
parser.add_argument('--score_name', type=str, default="pocket_volume.sc",
    help='str; Name for pocket features scorefile.')

args = parser.parse_args()


def outer_product(a, b):
    return np.einsum('ij,ik->ijk',a,b)


@njit(cache=True, fastmath=True)
def indices_box(lb_idx, ub_idx ):
    sizes = ub_idx - lb_idx + 1
    size = sizes[0]*sizes[1]*sizes[2]

    indices = np.zeros((size, 3), np.int_)

    count = 0
    for k in range(lb_idx[2], ub_idx[2]+1):
        for j in range(lb_idx[1], ub_idx[1]+1):
            for i in range(lb_idx[0], ub_idx[0]+1):
                indices[count, 0] = i
                indices[count, 1] = j
                indices[count, 2] = k
                count = count + 1

    return indices
```

```
@njit(cache=True, fastmath=True)
def points_above_plane(pts, normal, pt_plane, padding=0):
    if ( padding != 0 ):
        normal = normal / np.linalg.norm(normal)
        pt_plane = pt_plane - normal*padding

    # to_pts = pts - pt_plane
    # dots = np.dot(to_pts, normal)
    # return dots > 0

    result = np.zeros(len(pts), np.bool_)

    for i in range(len(pts)):
        a = (pts[i, 0] - pt_plane[0]) * normal[0]
        b = (pts[i, 1] - pt_plane[1]) * normal[1]
        c = (pts[i, 2] - pt_plane[2]) * normal[2]

        result[i] = (a + b + c) > 0

    return result


@njit(cache=True, fastmath=True)
def triangular_prism_inner(fill_indices, fill_centers, to_other_face, triangle, upper_triangle,
padding):

    # when looking from inside the shape, this triangle runs clockwise
    cw_triangle = triangle.copy()
    # make upper counter clockwise so it's not super confusing
    ccw_upper = upper_triangle.copy()

    if ( np.dot( to_other_face, nu.cross(triangle[0]-triangle[1],
                                         triangle[2]-triangle[1])) < 0 ):
        cw_triangle[1] = triangle[2]
        cw_triangle[2] = triangle[1]
        ccw_upper[1] = upper_triangle[2]
        ccw_upper[2] = upper_triangle[1]


    normal = nu.cross(cw_triangle[0]-cw_triangle[1],
                                 cw_triangle[2]-cw_triangle[1])
```

```
    # original triangle
    mask = points_above_plane(fill_centers,
                          nu.cross(cw_triangle[0]-cw_triangle[1],
                                cw_triangle[2]-cw_triangle[1]),
                          cw_triangle[0],
                          padding
                          )
    fill_indices = fill_indices[mask]
    fill_centers = fill_centers[mask]


    # upper triangle

    mask = points_above_plane(fill_centers,
                          -nu.cross(ccw_upper[0]-ccw_upper[1],
                                 ccw_upper[2]-ccw_upper[1]),
                          ccw_upper[0],
                          padding
                          )
    fill_indices = fill_indices[mask]
    fill_centers = fill_centers[mask]

    # walls
    for iwall in range(3):
        # we are going to pretend triangle is the floor and we're inside
        #  the box
        bottom_right = cw_triangle[iwall]
        top_right = ccw_upper[iwall]
        bottom_left = cw_triangle[(iwall-1)%3]

        mask = points_above_plane(fill_centers,
                              nu.cross(top_right - bottom_right,
                                    bottom_left - bottom_right),
                              bottom_right,
                              padding
                              )
        fill_indices = fill_indices[mask]
        fill_centers = fill_centers[mask]

    return fill_indices


def triangular_prism_indices(vx, triangle, to_other_face, padding=0):
```

```
    upper_triangle = triangle + to_other_face

    # just going to assume it's cubic
    resl = vx.cs[0]

    lb = np.min( np.r_[upper_triangle, triangle], axis=0) - padding - resl
    ub = np.max( np.r_[upper_triangle, triangle], axis=0) + padding + resl

    lb_idx, ub_idx = vx.floats_to_indices(np.array([lb, ub]))

    fill_indices = indices_box(lb_idx, ub_idx)

    fill_centers = vx.indices_to_centers(fill_indices)

    return triangular_prism_inner(fill_indices, fill_centers, to_other_face, triangle, upper_triangle,
padding)


@njit(cache=True, fastmath=True)
def fill_pockets_inner(color, arr, lb, ub, cs, shape):
  for x in range(2, shape[0] -2):
      for y in range(2, shape[1] -2):
          for z in range(2, shape[2] -2):
              if ( arr[x, y, z] == 0 ):
                  arr[x, y, z] = color
                  voxel_array.numba_flood_fill_3d_from_here(color, 0, np.array([x, y, z], np.int_), arr,
lb, ub, cs, shape)
                  color += 1
  return color


def fill_pockets(vox_arr, color=4):
   color = fill_pockets_inner(color, vox_arr.arr, vox_arr.lb, vox_arr.ub, vox_arr.cs,
vox_arr.arr.shape)
   return color



def make_surf_grid(verts, vert_normals, surfgrid):

      atom_size = 2.0

      surfgrid.add_to_clashgrid(verts-vert_normals*-1.5, atom_size, -1)
      surfgrid.add_to_clashgrid(verts-vert_normals*-0.5, atom_size, -1)
```

```
        surfgrid.add_to_clashgrid(verts-vert_normals*0.5, atom_size, -1)
        surfgrid.add_to_clashgrid(verts-vert_normals*1.5, atom_size, -1)



first = True

pdbs = []
if args.pdbs:
    pdbs = args.pdbs
elif args.pdbs_list_file:
    with open(args.pdbs_list_file, "r") as f:
        pdbs = f.readlines()
        pdbs = [x.strip() for x in pdbs]

if len(pdbs) == 0:
    print("no input pdbs")
    sys.exit()

#start = time.time()
for pdb_name in pdbs:
    #print(f"working on {pdb_name}")

    if args.oligomer_mode:
        npose_in, symm = nu.npose_from_file_fast(pdb_name, chains=True)
        symm = len(''.join(set(symm)))
        asu_size = int(nu.nsize(npose_in)/symm)
        npose = npose_in[:nu.R*asu_size*2]
    else:
        npose = nu.npose_from_file_fast(pdb_name)


    atom_size = 2.2
    shell_thickness = 0.01
    padding = 7.0
    resl = args.resl


    clash_grid = nu.ca_clashgrid_from_npose(npose, atom_size, resl, padding=3)

    vx = voxel_array.VoxelArray(clash_grid.lb[:3], clash_grid.ub[:3], clash_grid.cs[:3],
dtype=np.bool)
```

```
if args.rolling_ball:


    cas = nu.extract_atoms(npose, [nu.CA])[:,:3]

    probe_radius = args.rolling_ball_radius

    radii = np.repeat(2, len(cas))
    surf = atomic_depth.AtomicDepth(cas[:,:3].reshape(-1), radii, probe_radius, 1.0, True, 1)

    verts = surf.get_surface_vertex_bases().reshape(-1, 3)

    if args.debug_dumping:
        nu.dump_pts(verts, "vertices.pdb")

    vert_normals = surf.get_surface_vertex_normals().reshape(-1, 3)
    face_centers = surf.get_surface_face_centers().reshape(-1, 3)
    face_normals = surf.get_surface_face_normals().reshape(-1, 3)

    if args.debug_dumping:
        nu.dump_pts(verts-vert_normals*1.45, "vert_vert_normals.pdb")

    make_surf_grid(verts, vert_normals, vx)


else:
    hull = ConvexHull(npose[:,:3])

    vertices = hull.points[hull.vertices]

    if args.debug_dumping:
        nu.dump_pts(vertices, "vertices.pdb")

    cluster_resl = 5
    centers, _ = nu.cluster_points(vertices, cluster_resl, find_centers=True)

    hull = ConvexHull(vertices[centers])
    delaunay_hull = Delaunay(vertices[centers])

    for simplex_idxs in hull.simplices:

        simplex_vertices = hull.points[simplex_idxs]

        normal = np.cross( simplex_vertices[0] - simplex_vertices[1],
```

```
                    simplex_vertices[0] - simplex_vertices[2]
                    )
        normal /= np.linalg.norm(normal)
        center = np.mean(simplex_vertices, axis=0)

        test_pt = center + normal*0.01

        test_inside = delaunay_hull.find_simplex(test_pt) >= 0

        if ( test_inside ):
            normal *= -1

        fill = triangular_prism_indices(vx, simplex_vertices, normal*shell_thickness, padding)

        vx.arr[ tuple(fill.T) ] = True

    if args.debug_dumping:
        nu.dump_pts(simplex_vertices, "triangle.pdb")
        vx.dump_mask_true("face.pdb", vx.arr, fraction=0.01)


number_grid = voxel_array.VoxelArray(vx.lb, vx.ub, vx.cs, np.int)

number_grid.arr.flat[vx.arr.reshape(-1)] = 1
number_grid.arr.flat[clash_grid.arr.reshape(-1)] = 2


# technically the border is supposed to be 0
# so go to 1, 1, 1
number_grid.arr[1, 1, 1] = 3

voxel_array.numba_flood_fill_3d_from_here(3, 0, np.array([1, 1, 1], np.int_), number_grid.arr,
number_grid.lb, number_grid.ub, number_grid.cs, number_grid.arr.shape)
color = fill_pockets(number_grid)

num_colors = color
"""
bounding box = 0
convex hull = 1
clash grid = 2
outside hull = 3
pockfill = 4+
"""
```

```
    if args.debug_dumping:
        vx.dump_mask_true("polygon.pdb", number_grid.arr == 1, fraction=1)
        vx.dump_mask_true("clash_grid.pdb", number_grid.arr == 2, fraction=0.1)
        vx.dump_mask_true("outside_hull.pdb", number_grid.arr == 3, fraction=0.01)

    size_by_color = [0, 0, 0, 0]

    for color in range(4, num_colors+1):
        volume = np.sum(number_grid.arr == color) * number_grid.cs[0] * number_grid.cs[1] *
number_grid.cs[2]
        size_by_color.append(volume)

    largest_color = np.argmax(size_by_color)
    pocket_volume = size_by_color[largest_color]

    if args.debug_dumping:
        colors_to_dump = [largest_color]
        #uncomment if you reeeaaalllly want to dump all
        #colors_to_dump = range(0, num_colors+1)
        for color in colors_to_dump:
            number_grid.dump_mask_true("pocket_{}.pdb".format(color), number_grid.arr == color,
fraction=0.1)


    #get 3 main dimensions of the pocket
    all_mass_idx = np.array(list(np.where(number_grid.arr == largest_color))).T
    all_mass = number_grid.indices_to_centers(all_mass_idx)

    mass_com = nu.center_of_mass(all_mass)

    # https://en.wikipedia.org/wiki/Moment_of_inertia#Inertia_tensor
    r = all_mass - mass_com
    r_norm2 = np.sum( np.square(r), axis=-1 )

    r_outer_prod = outer_product( r, r )

    dx_dy_dz = np.prod( number_grid.cs )

    inertia_tensor = np.sum( r_norm2[:,None,None] * np.identity(3) - r_outer_prod, axis=0 ) *
dx_dy_dz

    # https://ccrma.stanford.edu/~jos/pasp/Principal_Axes_Rotation.html
    eigen_values, eigen_vectors = np.linalg.eig(inertia_tensor)
    eigen_vectors = eigen_vectors.T
```

```
    argsort = np.argsort(-eigen_values)
    dimensions = []
    for i, eig_i in enumerate(argsort):

        #print("Eigenvalue %i: %8.3f  -- sqrt() %8.3f"%(i, eigen_values[eig_i],
np.sqrt(eigen_values[eig_i])))

        vector = eigen_vectors[eig_i]

        projections = np.dot(r, vector)
        size_in_this_dim = np.max(projections) - np.min(projections)
        dimensions.append(size_in_this_dim)
        #print("Eigenvalue %i length = %8.3f"%(i, dst))

        if args.debug_dumping:
            nu.dump_lines([mass_com - vector * (size_in_this_dim/2)], [vector], size_in_this_dim,
"eig%i.pdb"%i)

    dimensions.sort()
    pmoi = []
    for i in range(len(dimensions)):
        #for each principal moment of inertia, add the other 2 axes squared
        #technically this should also be multiplied by the mass of the object
        #but the mass cancels out in the ratios I1/I3 and I2/I3, so i didn't include it
        #i.e. Ixx =(sum of the mass) * (y^2 + x^2)
        #For info on the primary moment of inertia graph, see (Sauer and Schwarz 2003)
        pmoi.append((dimensions[i]*dimensions[i]) + (dimensions[(i+1)%3] * dimensions[(i+1)%3]))
    pmoi.sort()

    score_dict = {
    "pocket_volume": pocket_volume,
    "dimension_1": dimensions[0],
    "dimension_2": dimensions[1],
    "dimension_3": dimensions[2],
    "I1": pmoi[0],
    "I2": pmoi[1],
    "I3": pmoi[2],
    "I1/I3": pmoi[0]/pmoi[2],
    "I2/I3": pmoi[1]/pmoi[2],
    }

    if args.pocket_features:
```

```
        nu.add_to_score_file(pdb_name, args.score_name, first, score_dict)

        first = False


    if args.pocket_residues:

        """
        try to get the residues lining the pocket now
        """

        # clash grid for pocket
        pocket_clash_grid = nu.clashgrid_from_points(all_mass, 1.1, 0.5, padding=100)

        #protein Cbs
        cbs = nu.extract_atoms(npose, [nu.CB])[:,:3]

        pocket_res = []
        for i, cb in enumerate(cbs):
            clash = np.any( pocket_clash_grid.arr[ tuple( pocket_clash_grid.indices_within_x_of(
2.5, cb ).T) ] )
            if clash:
                pocket_res.append(i+1)
        '''
        if args.oligomer_mode:
            pocket_res_1 = []
            for res in pocket_res:
                if res <= asu_size:
                    pocket_res_1.append(res)
                else:
                    if res - asu_size not in pocket_res_1:
                        pocket_res_1.append(res - asu_size)
            pocket_res = pocket_res_1
        '''
        pocket_res_str = [str(x) for x in pocket_res]
        #for pymol selection...
        #print(f"select resi {'+'.join(pocket_res_str)}")

        base = os.path.basename(pdb_name)
        outname = os.path.splitext(base)[0] + ".pos"

        with open(outname, "w") as f:
            for resi in pocket_res_str:
                f.write(resi)
```

```
            f.write("\n")

#print(f"run time = {time.time() - start}")
#
```

References

Thesis References:

1. Walsh, S. T., Cheng, H., Bryson, J. W., Roder, H. & DeGrado, W. F. Solution structure and dynamics of a de novo designed three-helix bundle protein. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 5486–5491 (1999).
2. Harbury, P. B., Plecs, J. J., Tidor, B., Alber, T. & Kim, P. S. High-resolution protein design with backbone freedom. *Science* **282**, 1462–1467 (1998).
3. Langan, R. A. *et al.* De novo design of bioactive protein switches. *Nature* **572**, 205–210 (2019).
4. Kirkpatrick, R. L. *et al.* Conditional Recruitment to a DNA-Bound CRISPR-Cas Complex Using a Colocalization-Dependent Protein Switch. *ACS Synth. Biol.* **9**, 2316–2323 (2020).
5. Lajoie, M. J. *et al.* Designed protein logic to target cells with precise combinations of surface antigens. *Science* **369**, 1637–1643 (2020).
6. Der, B. S., Edwards, D. R. & Kuhlman, B. Catalysis by a de novo zinc-mediated protein interface: implications for natural enzyme evolution and rational enzyme engineering. *Biochemistry* **51**, 3933–3940 (2012).
7. Chino, M. *et al.* Artificial Diiron Enzymes with a De Novo Designed Four-Helix Bundle Structure. *Eur. J. Inorg. Chem.* **2015**, 3371–3390 (2015).

8. Cangelosi, V. M., Deb, A., Penner-Hahn, J. E. & Pecoraro, V. L. A DE Novo designed metalloenzyme for the hydration of CO2. *Angew. Chem. Weinheim Bergstr. Ger.* **126**, 8034–8037 (2014).
9. Divine, R. *et al.* Designed proteins assemble antibodies into modular nanocages. *Science* **372**, (2021).
10. Klima, J. C. *et al.* Incorporation of sensing modalities into de novo designed fluorescence-activating proteins. *Nat. Commun.* **12**, 856 (2021).
11. Ben-Sasson, A. J. *et al.* Design of biologically active binary protein 2D materials. *Nature* **589**, 468–473 (2021).
12. Cohen, C. & Parry, D. A. D. α-Helical coiled coils — a widespread motif in proteins. *Trends Biochem. Sci.* **11**, 245–248 (1986).
13. Dou, J. *et al.* De novo design of a fluorescence-activating β-barrel. *Nature* **561**, 485–491 (2018).
14. Basanta, B. *et al.* An enumerative algorithm for de novo design of proteins with diverse pocket structures. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 22135–22145 (2020).
15. Marcos, E. *et al.* Principles for designing proteins with cavities formed by curved β sheets. *Science* **355**, 201–206 (2017).
16. Fairman, J. W., Noinaj, N. & Buchanan, S. K. The structural biology of β-barrel membrane proteins: a summary of recent reports. *Curr. Opin. Struct. Biol.* **21**, 523–531 (2011).
17. Tokuda, H. Biogenesis of outer membranes in Gram-negative bacteria. *Biosci. Biotechnol. Biochem.* **73**, 465–473 (2009).
18. Kleinschmidt, J. H. & Tamm, L. K. Secondary and tertiary structure formation of the

beta-barrel membrane protein OmpA is synchronized and depends on membrane thickness. *J. Mol. Biol.* **324**, 319–330 (2002).

19. Kleinschmidt, J. H., den Blaauwen, T., Driessen, A. J. & Tamm, L. K. Outer membrane protein A of Escherichia coli inserts and folds into lipid bilayers by a concerted mechanism. *Biochemistry* **38**, 5006–5016 (1999).
20. Butler, T. Z., Pavlenok, M., Derrington, I. M., Niederweis, M. & Gundlach, J. H. Single-molecule DNA detection with an engineered MspA protein nanopore. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 20647–20652 (2008).
21. Guan, X., Gu, L.-Q., Cheley, S., Braha, O. & Bayley, H. Stochastic sensing of TNT with a genetically engineered pore. *Chembiochem* **6**, 1875–1881 (2005).
22. Haque, F., Lunn, J., Fang, H., Smithrud, D. & Guo, P. Real-time sensing and discrimination of single chemicals using the channel of phi29 DNA packaging nanomotor. *ACS Nano* **6**, 3251–3261 (2012).
23. Tu, Y.-M. *et al.* Rapid fabrication of precise high-throughput filters from membrane protein nanosheets. *Nat. Mater.* **19**, 347–354 (2020).
24. Vorobieva, A. A. *et al.* De novo design of transmembrane β barrels. *Science* **371**, (2021).
25. Korkegian, A., Black, M. E., Baker, D. & Stoddard, B. L. Computational thermostabilization of an enzyme. *Science* **308**, 857–860 (2005).
26. Kiss, G., Çelebi-Ölçüm, N., Moretti, R., Baker, D. & Houk, K. N. Computational enzyme design. *Angew. Chem. Int. Ed Engl.* **52**, 5700–5725 (2013).
27. Zanghellini, A. de novo computational enzyme design. *Curr. Opin. Biotechnol.* **29**, 132–138 (2014).

28. Chowdhury, R. & Maranas, C. D. From directed evolution to computational enzyme engineering—A review. *AIChE J.* **66**, (2020).
29. Peacock, A. F. A. Incorporating metals into de novo proteins. *Curr. Opin. Chem. Biol.* **17**, 934–939 (2013).
30. Zhang, S.-Q. *et al.* De Novo Design of Tetranuclear Transition Metal Clusters Stabilized by Hydrogen-Bonded Networks in Helical Bundles. *J. Am. Chem. Soc.* **140**, 1294–1304 (2018).
31. Lu, P. *et al.* Accurate computational design of multipass transmembrane proteins. *Science* **359**, 1042–1046 (2018).
32. Joh, N. H., Grigoryan, G., Wu, Y. & DeGrado, W. F. Design of self-assembling transmembrane helical bundles to elucidate principles required for membrane protein folding and ion transport. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **372**, (2017).
33. Danoff, E. J. & Fleming, K. G. Novel Kinetic Intermediates Populated along the Folding Pathway of the Transmembrane β-Barrel OmpA. *Biochemistry* **56**, 47–60 (2017).
34. Moon, C. P., Kwon, S. & Fleming, K. G. Overcoming hysteresis to attain reversible equilibrium folding for outer membrane phospholipase A in phospholipid bilayers. *J. Mol. Biol.* **413**, 484–494 (2011).
35. Sibanda, B. L., Blundell, T. L. & Thornton, J. M. Conformation of beta-hairpins in protein structures. A systematic classification with applications to modelling by homology, electron density fitting and protein engineering. *J. Mol. Biol.* **206**, 759–777 (1989).

36. Holm, R. H., Kennepohl, P. & Solomon, E. I. Structural and Functional Aspects of Metal Sites in Biology. *Chem. Rev.* **96**, 2239–2314 (1996).
37. Solomon, E. I. & Hadt, R. G. Recent advances in understanding blue copper proteins. *Coord. Chem. Rev.* **255**, 774–789 (2011).
38. Solomon, E. I. *et al.* Copper active sites in biology. *Chem. Rev.* **114**, 3659–3853 (2014).
39. Solomon, E. I., Sundaram, U. M. & Machonkin, T. E. Multicopper Oxidases and Oxygenases. *Chem. Rev.* **96**, 2563–2606 (1996).
40. Solomon, E. I., Baldwin, M. J. & Lowery, M. D. Electronic structures of active sites in copper proteins: contributions to reactivity. *Chem. Rev.* **92**, 521–542 (1992).
41. Sousa, F. L. *et al.* The superfamily of heme-copper oxygen reductases: types and evolutionary considerations. *Biochim. Biophys. Acta* **1817**, 629–637 (2012).
42. Zumft, W. G. Nitric oxide reductases of prokaryotes with emphasis on the respiratory, heme-copper oxidase type. *J. Inorg. Biochem.* **99**, 194–215 (2005).
43. Coleman, J. E. Zinc proteins: enzymes, storage proteins, transcription factors, and replication proteins. *Annu. Rev. Biochem.* **61**, 897–946 (1992).
44. McCall, K. A., Huang, C. & Fierke, C. A. Function and mechanism of zinc metalloenzymes. *J. Nutr.* **130**, 1437S–46S (2000).
45. Argos, P., Garavito, R. M., Eventoff, W., Rossmann, M. G. & Brändén, C. I. Similarities in active center geometries of zinc-containing enzymes, proteases and dehydrogenases. *J. Mol. Biol.* **126**, 141–158 (1978).
46. Winum, J.-Y. & Supuran, C. T. Recent advances in the discovery of zinc-binding motifs for the development of carbonic anhydrase inhibitors. *J. Enzyme Inhib. Med.*

Chem. **30**, 321–324 (2015).

47. Koehntop, K. D., Emerson, J. P. & Que, L., Jr. The 2-His-1-carboxylate facial triad: a versatile platform for dioxygen activation by mononuclear non-heme iron(II) enzymes. *J. Biol. Inorg. Chem.* **10**, 87–93 (2005).
48. Solomon, E. I. *et al.* Geometric and electronic structure/function correlations in non-heme iron enzymes. *Chem. Rev.* **100**, 235–350 (2000).
49. Liu, J. *et al.* Metalloproteins containing cytochrome, iron-sulfur, or copper redox centers. *Chem. Rev.* **114**, 4366–4469 (2014).
50. Sugimoto, H. *et al.* Crystal structure of human indoleamine 2,3-dioxygenase: catalytic mechanism of O2 incorporation by a heme-containing dioxygenase. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 2611–2616 (2006).
51. Mutti, F. G. Alkene cleavage catalysed by heme and nonheme enzymes: reaction mechanisms and biocatalytic applications. *Bioinorg. Chem. Appl.* **2012**, 626909 (2012).
52. Ilie, A. & Reetz, M. T. Directed evolution of artificial metalloenzymes. *Isr. J. Chem.* **55**, 51–60 (2015).
53. Lin, Y.-W. Rational design of metalloenzymes: From single to multiple active sites. *Coord. Chem. Rev.* **336**, 1–27 (2017).
54. Du, J., Sono, M. & Dawson, J. H. The H93G Myoglobin Cavity Mutant as a Versatile Scaffold for Modeling Heme Iron Coordination Structures in Protein Active Sites and Their Characterization with Magnetic Circular Dichroism Spectroscopy. *Coord. Chem. Rev.* **255**, 700–716 (2011).
55. Tegoni, M., Yu, F., Bersellini, M., Penner-Hahn, J. E. & Pecoraro, V. L. Designing a

functional type 2 copper center that has nitrite reductase activity within α-helical coiled coils. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 21234–21239 (2012).

56. Lombardi, A., Pirro, F., Maglio, O., Chino, M. & DeGrado, W. F. De Novo Design of Four-Helix Bundle Metalloproteins: One Scaffold, Diverse Reactivities. *Acc. Chem. Res.* **52**, 1148–1159 (2019).
57. Huang, P.-S. *et al.* High thermodynamic stability of parametrically designed helical bundles. *Science* **346**, 481–485 (2014).
58. Huang, P.-S. *et al.* De novo design of a four-fold symmetric TIM-barrel protein with atomic-level accuracy. *Nat. Chem. Biol.* **12**, 29–34 (2016).
59. Boyken, S. E. *et al.* De novo design of protein homo-oligomers with modular hydrogen-bond network-mediated specificity. *Science* **352**, 680–687 (2016).
60. Doyle, L. *et al.* Rational design of α-helical tandem repeat proteins with closed architectures. *Nature* **528**, 585–588 (2015).
61. Huang, P.-S. *et al.* RosettaRemodel: a generalized framework for flexible backbone protein design. *PLoS One* **6**, e24109 (2011).
62. Brunette, T. J. *et al.* Exploring the repeat protein universe through computational protein design. *Nature* **528**, 580–584 (2015).
63. Mills, J. H. *et al.* Computational design of a homotrimeric metalloprotein with a trisbipyridyl core. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 15012–15017 (2016).
64. Fallas, J. A. *et al.* Computational design of self-assembling cyclic protein homo-oligomers. *Nat. Chem.* **9**, 353–360 (2017).
65. Lindskog, S. Structure and mechanism of carbonic anhydrase. *Pharmacol. Ther.* **74**, 1–20 (1997).

66. Dudev, T., Lin, Y.-L., Dudev, M. & Lim, C. First-second shell interactions in metal binding sites in proteins: a PDB survey and DFT/CDM calculations. *J. Am. Chem. Soc.* **125**, 3168–3180 (2003).
67. Marshall, N. M. *et al.* Rationally tuning the reduction potential of a single cupredoxin beyond the natural range. *Nature* **462**, 113–116 (2009).
68. Superdex%2075%20Increase.pdf.
69. Feldmeier, K. & Höcker, B. Computational protein design of ligand binding and catalysis. *Curr. Opin. Chem. Biol.* **17**, 929–933 (2013).
70. Yang, W. *et al.* Rational design of a calcium-binding protein. *J. Am. Chem. Soc.* **125**, 6165–6171 (2003).
71. Regan, L. Protein design: novel metal-binding sites. *Trends Biochem. Sci.* **20**, 280–285 (1995).
72. Goodsell, D. S. & Olson, A. J. Structural symmetry and protein function. *Annu. Rev. Biophys. Biomol. Struct.* **29**, 105–153 (2000).
73. Cannon, K. A. *et al.* Design and structure of two new protein cages illustrate successes and ongoing challenges in protein engineering. *Protein Sci.* **29**, 919–929 (2020).
74. Hsia, Y. *et al.* Design of a hyperstable 60-subunit protein dodecahedron. [corrected]. *Nature* **535**, 136–139 (2016).
75. Maguire, J. B. *et al.* Perturbing the energy landscape for improved packing during computational protein design. *Proteins* **89**, 436–449 (2021).
76. Shen, H. *et al.* De novo design of self-assembling helical protein filaments. *Science* **362**, 705–709 (2018).

77. Ueda, G. *et al.* Tailored design of protein nanoparticle scaffolds for multivalent presentation of viral glycoprotein antigens. *Elife* **9**, (2020).
78. Antanasijevic, A. *et al.* Structural and functional evaluation of de novo-designed, two-component nanoparticle carriers for HIV Env trimer immunogens. *PLoS Pathog.* **16**, e1008665 (2020).
79. Foight, G. W. *et al.* Multi-input chemical control of protein dimerization for programming graded cellular responses. *Nat. Biotechnol.* **37**, 1209–1216 (2019).
80. Pyles, H., Zhang, S., De Yoreo, J. J. & Baker, D. Controlling protein assembly on inorganic crystals through designed protein interfaces. *Nature* **571**, 251–256 (2019).
81. Sauer, W. H. B. & Schwarz, M. K. Molecular shape diversity of combinatorial libraries: a prerequisite for broad bioactivity. *J. Chem. Inf. Comput. Sci.* **43**, 987–1003 (2003).
82. Meyers, J., Carter, M., Mok, N. Y. & Brown, N. On the origins of three-dimensionality in drug-like molecules. *Future Med. Chem.* **8**, 1753–1767 (2016).
83. DiMaio, F., Leaver-Fay, A., Bradley, P., Baker, D. & André, I. Modeling symmetric macromolecular structures in Rosetta3. *PLoS One* **6**, e20450 (2011).
84. Fleishman, S. J. *et al.* RosettaScripts: a scripting language interface to the Rosetta macromolecular modeling suite. *PLoS One* **6**, e20161 (2011).
85. Lauer, T. M. *et al.* Developability index: a rapid in silico tool for the screening of antibody aggregation propensity. *J. Pharm. Sci.* **101**, 102–115 (2012).
86. Dyer, K. N. *et al.* High-Throughput SAXS for the Characterization of Biomolecules

in Solution: A Practical Approach. in *Structural Genomics: General Applications* (ed. Chen, Y. W.) 245–258 (Humana Press, 2014).

87. Hura, G. L. *et al.* Robust, high-throughput solution structural analyses by small angle X-ray scattering (SAXS). *Nat. Methods* **6**, 606–612 (2009).
88. Classen, S. *et al.* 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.* **46**, 1–13 (2013).
89. 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.* **40**, 191–285 (2007).
90. Schneidman-Duhovny, D., Hammel, M., Tainer, J. A. & Sali, A. Accurate SAXS profile computation and its assessment by contrast variation experiments. *Biophys. J.* **105**, 962–974 (2013).
91. 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.* **44**, W424–9 (2016).
92. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
93. AlQuraishi, M. AlphaFold at CASP13. *Bioinformatics* **35**, 4862–4865 (2019).
94. Veesler, D. *et al.* Single-particle EM reveals plasticity of interactions between the adenovirus penton base and integrin αVβ3. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 8815–8819 (2014).

95. Grant, T., Rohou, A. & Grigorieff, N. cisTEM, user-friendly software for single-particle image processing. *Elife* **7**, (2018).
96. Grigorieff, N., Grant, T., Rohou, A. & IUCr. cisTEM: user-friendly software for single-particle image processing. *Acta Crystallographica Section A: Foundations and Advances* **73**, C1368–C1368 (2017).
97. Zivanov, J. *et al.* New tools for automated high-resolution cryo-EM structure determination in RELION-3. *Elife* **7**, (2018).
98. Scheres, S. H. W. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012).
99. Kimanius, D., Forsberg, B. O., Scheres, S. H. & Lindahl, E. Accelerated cryo-EM structure determination with parallelisation using GPUs in RELION-2. *Elife* **5**, (2016).
100. Kabsch, W. XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132 (2010).
101. Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 235–242 (2011).
102. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
103. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221 (2010).
104. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 (2004).

Vorobieva et al references:

1. R. A. Langan et al., De novo design of bioactive protein switches. Nature 572, 205–210 (2019). doi: 10.1038/s41586- 019-1432-8; pmid: 31341284
2. A. H. Ng et al., Modular and tunable biological feedback control using a de novo protein switch. Nature 572, 265–269 (2019). doi: 10.1038/s41586-019-1425-7; pmid: 31341280
3. D.-A. Silva et al., De novo design of potent and selective mimics of IL-2 and IL-15. Nature 565, 186–191 (2019). doi: 10.1038/s41586-018-0830-7; pmid: 30626941
4. E. Marcos et al., De novo design of a non-local b-sheet protein with high stability and accuracy. Nat. Struct. Mol. Biol. 25, 1028–1034 (2018). doi: 10.1038/s41594-018-0141-6; pmid: 30374087
5. J. Dou et al., De novo design of a fluorescence-activating b-barrel. Nature 561, 485–491 (2018). doi: 10.1038/s41586- 018-0509-0; pmid: 30209393
6. P. Lu et al., Accurate computational design of multipass transmembrane proteins. Science 359, 1042–1046 (2018). doi: 10.1126/science.aaq1739; pmid: 29496880
7. N. H. Joh, G. Grigoryan, Y. Wu, W. F. DeGrado, Design of self-assembling transmembrane helical bundles to elucidate principles required for membrane protein folding and ion transport. Philos. Trans. R. Soc. Lond. B Biol. Sci. 372, 20160214 (2017). doi: 10.1098/rstb.2016.0214; pmid: 28630154
8. J. H. Kleinschmidt, T. den Blaauwen, A. J. Driessen, L. K. Tamm, Outer membrane protein A of Escherichia coli inserts and folds into lipid bilayers by a concerted mechanism. Biochemistry 38, 5006–5016 (1999). doi: 10.1021/bi982465w; pmid: 10213603
9. J. H. Kleinschmidt, L. K. Tamm, Secondary and tertiary structure formation of the b-barrel membrane protein OmpA is synchronized and depends on membrane thickness. J. Mol. Biol. 324, 319–330 (2002). doi: 10.1016/S0022-2836 (02)01071-9; pmid: 12441110
10. E. J. Danoff, K. G. Fleming, Novel Kinetic Intermediates Populated along the Folding Pathway of the Transmembrane b-Barrel OmpA. Biochemistry 56, 47–60 (2017). doi: 10.1021/acs.biochem.6b00809; pmid: 28001375
11. C. P. Moon, S. Kwon, K. G. Fleming, Overcoming hysteresis to attain reversible equilibrium folding for outer membrane phospholipase A in phospholipid bilayers. J. Mol. Biol. 413, 484–494 (2011). doi: 10.1016/j.jmb.2011.08.041; pmid: 21888919
12. D. Chaturvedi, R. Mahalakshmi, Transmembrane b-barrels: Evolution, folding and energetics. Biochim. Biophys. Acta Biomembr. 1859, 2467–2482 (2017). doi: 10.1016/j.bbamem.2017.09.020; pmid: 28943271
13. T. Z. Butler, M. Pavlenok, I. M. Derrington, M. Niederweis, J. H. Gundlach, Single-molecule DNA detection with an engineered MspA protein nanopore. Proc. Natl. Acad. Sci. U.S.A. 105, 20647–20652 (2008). doi: 10.1073/ pnas.0807514106; pmid: 19098105
14. X. Guan, L.-Q. Gu, S. Cheley, O. Braha, H. Bayley, Stochastic sensing of TNT with a genetically engineered pore. ChemBioChem 6, 1875–1881 (2005). doi: 10.1002/cbic.200500064; pmid: 16118820
15. F. Haque, J. Lunn, H. Fang, D. Smithrud, P. Guo, Real-time sensing and discrimination of single chemicals using the channel of phi29 DNA packaging nanomotor. ACS Nano 6, 3251–3261 (2012). doi: 10.1021/nn3001615; pmid: 22458779
16. Y.-M. Tu et al., Rapid fabrication of precise high-throughput filters from membrane protein nanosheets. Nat. Mater. 19, 347–354 (2020). doi: 10.1038/s41563-019-0577-z; pmid: 31988513

17. T. Surrey, F. Jähnig, Refolding and oriented insertion of a membrane protein into a lipid bilayer. Proc. Natl. Acad. Sci. U.S.A. 89, 7457–7461 (1992). doi: 10.1073/pnas.89.16.7457; pmid: 1502158
18. A. D. McLachlan, Gene duplications in the structural evolution of chymotrypsin. J. Mol. Biol. 128, 49–79 (1979). doi: 10.1016/ 0022-2836(79)90308-5; pmid: 430571
19. A. G. Murzin, A. M. Lesk, C. Chothia, Principles determining the structure of beta-sheet barrels in proteins. I. A theoretical analysis. J. Mol. Biol. 236, 1369–1381 (1994). doi: 10.1016/ 0022-2836(94)90064-7; pmid: 8126726
20. M. W. Franklin, J. S. G. Slusky, Tight Turns of Outer Membrane Proteins: An Analysis of Sequence, Structure, and Hydrogen Bonding. J. Mol. Biol. 430 (18 Pt B), 3251–3265 (2018). doi: 10.1016/j.jmb.2018.06.013; pmid: 29944853
21. E. de Alba, M. A. Jiménez, M. Rico, J. L. Nieto, Conformational investigation of designed short linear peptides able to fold into b-hairpin structures in aqueous solution. Fold. Des. 1, 133–144 (1996). doi: 10.1016/S1359-0278(96)00022-3; pmid: 9079373
22. T. Blandl, A. G. Cochran, N. J. Skelton, Turn stability in b-hairpin peptides: Investigation of peptides containing 3:5 type I G1 bulge turns. Protein Sci. 12, 237–247 (2003). doi: 10.1110/ps.0228603; pmid: 12538887
23. J. S. Richardson, E. D. Getzoff, D. C. Richardson, The beta bulge: A common small unit of nonrepetitive protein structure. Proc. Natl. Acad. Sci. U.S.A. 75, 2574–2578 (1978). doi: 10.1073/pnas.75.6.2574; pmid: 275827
24. W. C. Wimley, Toward genomic identification of b-barrel membrane proteins: Composition and architecture of known structures. Protein Sci. 11, 301–312 (2002). doi: 10.1110/ ps.29402; pmid: 11790840
25. L. K. Tamm, H. Hong, B. Liang, Folding and assembly of b-barrel membrane proteins. Biochimica et Biophysica Acta Biomembr. 1666, 250–263 (2004). doi: 10.1016/ j.bbamem.2004.06.011
26. J. S. Merkel, L. Regan, Aromatic rescue of glycine in b sheets. Fold. Des. 3, 449–455 (1998). doi: 10.1016/S1359-0278(98) 00062-5; pmid: 9889161
27. D. L. Leyton et al., A mortise-tenon joint in the transmembrane domain modulates autotransporter assembly into bacterial outer membranes. Nat. Commun. 5, 4239 (2014). doi: 10.1038/ncomms5239; pmid: 24967730
28. M. Michalik et al., An evolutionarily conserved glycine-tyrosine motif forms a folding core in outer membrane proteins. PLOS ONE 12, e0182016 (2017). doi: 10.1371/journal. pone.0182016; pmid: 28771529
29. H. Park et al., Simultaneous Optimization of Biomolecular Energy Functions on Features from Small Molecules and Macromolecules. J. Chem. Theory Comput. 12, 6201–6212 (2016). doi: 10.1021/acs.jctc.6b00819; pmid: 27766851
30. R. Misra, Assembly of the b-Barrel Outer Membrane Proteins in Gram-Negative Bacteria, Mitochondria, and Chloroplasts. ISRN Mol. Biol. 2012, 708203 (2012). doi: 10.5402/2012/ 708203; pmid: 27335668
31. M. Fioroni, T. Dworeck, F. Rodriguez-Ropero, ß-barrel Channel Proteins as Tools in Nanotechnology: Biology, Basic Science and Advanced Applications (Springer Science & Business Media, 2013).

32. R. D. Requião et al., Protein charge distribution in proteomes and its impact on translation. PLOS Comput. Biol. 13, e1005549 (2017). doi: 10.1371/journal.pcbi.1005549; pmid: 28531225
33. R. Koebnik, Structural and functional roles of the surfaceexposed loops of the b-barrel membrane protein OmpA from Escherichia coli. J. Bacteriol. 181, 3688–3694 (1999). doi: 10.1128/JB.181.12.3688-3694.1999; pmid: 10368142
34. E. J. Danoff, K. G. Fleming, The soluble, periplasmic domain of OmpA folds as an independent unit and displays chaperone activity by reducing the self-association propensity of the unfolded OmpA transmembrane b-barrel. Biophys. Chem. 159, 194–204 (2011). doi: 10.1016/j.bpc.2011.06.013; pmid: 21782315
35. S. J. Fleishman, D. Baker, Role of the biomolecular energy gap in protein design, structure, and evolution. Cell 149, 262–273 (2012). doi: 10.1016/j.cell.2012.03.016; pmid: 22500796
36. J. S. Richardson, D. C. Richardson, Natural beta-sheet proteins use negative design to avoid edge-to-edge aggregation. Proc. Natl. Acad. Sci. U.S.A. 99, 2754–2759 (2002). doi: 10.1073/ pnas.052706099; pmid: 11880627
37. S. E. Boyken et al., De novo design of protein homo-oligomers with modular hydrogen-bond network-mediated specificity. Science 352, 680–687 (2016). doi: 10.1126/science.aad8865; pmid: 27151862
38. D. L. Minor Jr., P. S. Kim, Measurement of the b-sheet-forming propensities of amino acids. Nature 367, 660–663 (1994). doi: 10.1038/367660a0; pmid: 8107853
39. T. Kortemme, A. V. Morozov, D. Baker, An orientationdependent hydrogen bonding potential improves prediction of specificity and structure for proteins and protein-protein complexes. J. Mol. Biol. 326, 1239–1259 (2003). doi: 10.1016/ S0022-2836(03)00021-4; pmid: 12589766
40. A. Ebie Tan, N. K. Burgess, D. S. DeAndrade, J. D. Marold, K. G. Fleming, Self-association of unfolded outer membrane proteins. Macromol. Biosci. 10, 763–767 (2010). doi: 10.1002/ mabi.200900479; pmid: 20491126
41. J.-L. Popot, Folding membrane proteins in vitro: A table and some comments. Arch. Biochem. Biophys. 564, 314–326 (2014). doi: 10.1016/j.abb.2014.06.029; pmid: 24997361
42. A. Schüßler, S. Herwig, J. H. Kleinschmidt, Kinetics of Insertion and Folding of Outer Membrane Proteins by Gel Electrophoresis. Methods Mol. Biol. 2003, 145–162 (2019). doi: 10.1007/978-1-4939-9512-7_7; pmid: 31218617
43. H. Hong, L. K. Tamm, Elastic coupling of integral membrane protein stability to lipid bilayer forces. Proc. Natl. Acad. Sci. U.S.A. 101, 4065–4070 (2004). doi: 10.1073/ pnas.0400358101; pmid: 14990786
44. G. H. M. Huysmans, S. A. Baldwin, D. J. Brockwell, S. E. Radford, The transition state for folding of an outer membrane protein. Proc. Natl. Acad. Sci. U.S.A. 107, 4099–4104 (2010). doi: 10.1073/ pnas.0911904107; pmid: 20133664
45. C. L. Pocanschi, G. J. Patel, D. Marsh, J. H. Kleinschmidt, Curvature elasticity and refolding of OmpA in large unilamellar vesicles. Biophys. J. 91, L75–L77 (2006). doi: 10.1529/ biophysj.106.091439; pmid: 16891370
46. S. Ohnishi, K. Kameyama, Escherichia coli OmpA retains a folded structure in the presence of sodium dodecyl sulfate due to a high kinetic barrier to unfolding. Biochim. Biophys. Acta 1515, 159–166 (2001). doi: 10.1016/S0005-2736(01)00410-2; pmid: 11718671

47. C. P. Moon, N. R. Zaccai, P. J. Fleming, D. Gessmann, K. G. Fleming, Membrane protein thermodynamic stability may serve as the energy sink for sorting in the periplasm. Proc. Natl. Acad. Sci. U.S.A. 110, 4285–4290 (2013). doi: 10.1073/ pnas.1212527110; pmid: 23440211
48. C. P. Moon, K. G. Fleming, Side-chain hydrophobicity scale derived from transmembrane protein folding into lipid bilayers. Proc. Natl. Acad. Sci. U.S.A. 108, 10174–10177 (2011). doi: 10.1073/pnas.1103979108; pmid: 21606332
49. H. Hong, S. Park, R. H. Flores Jiménez, D. Rinehart, L. K. Tamm, Role of aromatic side chains in the folding and thermodynamic stability of integral membrane proteins. J. Am. Chem. Soc. 129, 8320–8327 (2007). doi: 10.1021/ ja068849o; pmid: 17564441
50. J. H. Kleinschmidt, L. K. Tamm, Folding intermediates of a b-barrel membrane protein. Kinetic evidence for a multi-step membrane insertion mechanism. Biochemistry 35, 12993–13000 (1996). doi: 10.1021/bi961478b; pmid: 8855933
51. N. K. Burgess, T. P. Dao, A. M. Stanley, K. G. Fleming, Beta-barrel proteins that reside in the Escherichia coli outer membrane in vivo demonstrate varied folding behavior in vitro. J. Biol. Chem. 283, 26748–26758 (2008). doi: 10.1074/ jbc.M802754200; pmid: 18641391
52. Y. Shen, F. Delaglio, G. Cornilescu, A. Bax, TALOS+: A hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. J. Biomol. NMR 44, 213–223 (2009). doi: 10.1007/s10858-009-9333-z; pmid: 19548092
53. J. M. Hemmingsen, K. M. Gernert, J. S. Richardson, D. C. Richardson, The tyrosine corner: A feature of most Greek key b-barrel proteins. Protein Sci. 3, 1927–1937 (1994). doi: 10.1002/pro.5560031104; pmid: 7703839
54. C. M. Bishop, W. F. Walkenhorst, W. C. Wimley, Folding of b-sheets in membranes: Specificity and promiscuity in peptide model systems. J. Mol. Biol. 309, 975–988 (2001). doi: 10.1006/jmbi.2001.4715; pmid: 11399073
55. J. E. Horne, D. J. Brockwell, S. E. Radford, Role of the lipid bilayer in outer membrane protein folding in Gram-negative bacteria. J. Biol. Chem. 295, 10340–10367 (2020). doi: 10.1074/jbc.REV120.011473; pmid: 32499369
56. B. Schiffrin, D. J. Brockwell, S. E. Radford, Outer membrane protein folding from an energy landscape perspective. BMC Biol. 15, 123 (2017). doi: 10.1186/s12915-017-0464-5; pmid: 29268734
57. N. Koga et al., Principles for designing ideal protein structures. Nature 491, 222–227 (2012). doi: 10.1038/nature11600; pmid: 23135467
58. P. D. Adams et al., PHENIX: A comprehensive Python-based system for macromolecular structure solution. International Tables for Crystallography (2012), pp. 539–547.
59. M. Källberg, G. Margaryan, S. Wang, J. Ma, J. Xu, RaptorX server: A resource for template-based protein structure modeling. Methods Mol. Biol. 1137, 17–27 (2014). doi: 10.1007/ 978-1-4939-0366-5_2; pmid: 24573471
60. J. Kyte, R. F. Doolittle, A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. 157, 105–132 (1982). doi: 10.1016/0022-2836(82)90515-0; pmid: 7108955
61. A.-M. Fernandez-Escamilla, F. Rousseau, J. Schymkowitz, L. Serrano, Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins. Nat. Biotechnol. 22, 1302–1306 (2004). doi: 10.1038/ nbt1012; pmid: 15361882
62. A. Stein, T. Kortemme, Improvements to robotics-inspired conformational sampling in Rosetta. PLOS ONE 8, e63090 (2013). doi: 10.1371/journal.pone.0063090; pmid: 23704889

63. A. Vorobieva, vorobieva/TransmembraneBBarrels: TMB_manuscript_supplementary_data (version bbarrel_ manuscript), Zenodo (2020); http://doi.org/10.5281/ zenodo.4068108.