Software Algorithms for Design of Symmetric Protein Complexes Applied to Cryo-Electron Microscopy Scaffolds and Antibody Nanoparticles

Ivan Vulovic

A dissertation submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

University of Washington

2020

Reading Committee:
David Baker, Chair
David Beck
Neil King
Roland Strong

Program Authorized to Offer Degree:

Molecular Engineering and Sciences

University of Washington

Abstract

Software Algorithms for Design of Symmetric Protein Complexes Applied to Cryo-Electron Microscopy Scaffolds and Antibody Nanoparticles

Ivan Vulovic

Chair of Supervisory Committee:

David Baker

Department of Biochemistry

Innovation in the symmetric assembly and protein material design space has the potential to – eventually – reinvent medicine and nanotechnology. One leading strategy for design of symmetric assemblies uses genetic fusion of protein homo-oligomer subunits via α-helical linkers. For the nearly two decades since its inception in 2001, this method has been applied as it was originally formulated. In this dissertation, I present the development and application of a tripartite fusion method that builds on this work and addresses two of its principal limitations: linker flexibility and a dearth of geometric solutions. The method is applied to investigate two proof-of-principle design concepts: 1) target capture and structure determination on symmetric cryo-EM scaffolds 2) antibody display and integration into nanoparticles. Designs are characterized by native mass spectrometry, small angle X-ray scattering, and electron microscopy. The experimental results in both of these areas showcase the viability and promise of this design strategy for further use.

Acknowledgements

The Institute for Protein Design (IPD) houses a tremendous assortment of experts and facilities for protein science. The combination of funding and freely available access to equipment gives IPD members a unique opportunity to learn and perform research, unimpeded by traditional academic concerns. I am lucky to have had this access. Surprisingly, what I learned about how institutions and people operate is as important as what I learned about science. The IPD has grown tremendously during my four and a half years here. During this time, I gained a totally new appreciation for the roles of leadership, culture, incentives, process, and reputation in defining how an organization evolves. I value the life lessons and new perspective.

Upon joining the lab, I spent a brief period playing to my strengths and developing software methods, before focusing exclusively on bench work for the rest of my tenure. While I am not yet an expert (and have no plans to become one) in any biophysical method, what I did learn is sufficient for me to operate effectively in the fields of structural biology and protein design. For this, I owe thanks to many people: George Ueda and Jorge Fallas introduced me to protein expression, purification, and homo-oligomer design. I learned crystallography basics and fluorescence assisted cell sorting in the context of binder design from Danny Sahtoe. Alexis Courbet taught me image processing for 3D reconstruction of electron micrographs and how to use the Talos transmission electron microscope. I am also grateful to Will Sheffler, Yang Hsia, Ariel Ben-Sasson, Tamuka Chidyausiku, Alena Khmelinskaia, Florian Praetorius, Ta-Yi Yu, TJ Brunnette, Matt Bick, Chris Bahl, Jorgen Nelson, Jason Klima, and Vikram Mulligan. In addition, Marc Lajoie, Daniel Silva, Umut Ulge, and Scott Boyken proved that sometimes science can pay off, which is refreshing to see. I am also indebted to collaborators Young-Jun Park and Qing Yao for having performed EM 3D reconstruction. I must also thank Valerie

Daggett for introducing me to the field and former lab members Alissa Bleem, Mathew Childers, Steve Rysavy, and Denny Bromley for working with me early on.

I would also like to thank the members of my committee: My introduction to programming structural biology software came by way of Dave Beck's molecular dynamics package, *in lucem* molecular mechanics (*il*mm). Tinkering with and modifying *il*mm gave me a grounding in molecular modeling concepts. Shortly thereafter, it was through a meeting with the always enthusiastic Neil King that I secured a rotation in David Baker's lab. Next, I am indebted to David Baker for having established and grown the IPD. Beyond the science, I am struck by his relentless drive and ability effectively switch between and focus on so many different areas. Finally, Roland Strong has given me good advice along the way and also served on the committee that awarded a Molecular Biophysics Training Grant, which provided learning opportunities and greater freedom in navigating my way than I otherwise would have had.

Table of Contents

Introduction

Nature has evolved proteins that span an enormous diversity of structure and function, which structural biologists have traditionally attempted to uncover. More recently, the design of new proteins with new functions has become feasible as well, thanks to growing databases of working examples from nature, increases in computing power, and advent of protein modeling and design software. One design concept is the creation of symmetric protein assemblies akin to those found in nature. Among possible applications, designed protein assemblies might multivalently scaffold antigens to act as vaccines, encapsulate cargo for therapeutic delivery, or simply serve as tools for basic research. Symmetry is a useful tool in this endeavor, abundant in nature for the same reasons that it is appealing as a design principle; symmetry enables the creation of large assemblies through repetition of one or more unique components, which reduces the genomic encoding size, reduces the total amount of each constituent protein that must be synthesized to reach a favorable assembly concentration, and dramatically reduces the number of unique self-assembly interfaces that must be produced by either evolution or design.

The existing body of academic work on symmetric protein design fits into two broad categories: fusion and interface design. Both approaches attempt to enforce relative orientations between protein subunits that drive the formation of the regular geometric patterns and compose the target architecture. This dissertation describes a new variation on the fusion category, one that joins three entities via fusion – an oligomer, a spacer subunit, and another oligomer – to obviate two problems that typically limit fusion-based approaches: a paucity of geometric solutions and flexibility at the fusion site that leads to heterogeneity. Researchers may eventually develop fully modular building blocks that allow arbitrary topologies to be created with ease, but until then, this open source software provides an accessible means to generate assemblies from existing components.

Section 1. Generation of ordered protein assemblies using a multi-domain fusion approach

Abstract

Protein nanomaterial design is an emerging discipline with applications in medicine and beyond. A longstanding design approach uses genetic fusion to join protein homo-oligomer subunits via α-helical linkers to form more complex symmetric assemblies, but this method is hampered by linker flexibility and a dearth of geometric solutions. Here, we describe a general computational method that performs rigid three-body fusion of homo-oligomer and spacer building blocks to generate user-defined architectures, while at the same time significantly increasing the number of geometric solutions. The fusion junctions are then optimized using Rosetta to minimize flexibility. We apply this method to design and test 92 dihedral protein assemblies from a set of designed homo-dimers and de-novo helical repeat protein building blocks. Experimental validation by native mass spectrometry, small angle X-ray scattering, and negative-stain single-particle electron microscopy confirms the assembly states for 10 designs. Next, we assess the use of these scaffolds for multivalent display and cryo-EM structure determination of target proteins held in place by DARPins (designed ankyrin repeat proteins evolved to bind targets). The scaffolds introduce a new dual-anchoring strategy, in which the DARPin is positioned by an α-helical fusion as well as a homo-dimer interface.

Background and Motivation

Protein assemblies found in nature have inspired the development of novel protein assemblies to solve modern challenges. Already, modular protein nanoparticle vaccine platforms and cryo-EM scaffolds have been created. The dominant protein assembly design paradigm uses symmetry and either a designed interface or genetic fusion to fix the orientation of symmetric homo-oligomeric building blocks within the overall assembly. The genetic fusion approach has been used to create a tetrahedral protein nanocage and fiber[1], two-dimensional layers[2], tetrahedra[3,4], octahedra[5,6], and icosahedra[7,8]. The fusion procedure is relatively straightforward, not inherently destabilizing, and has perfect specificity as the interaction partners are fused. Despite its success and relative simplicity, several aspects of the genetic fusion approach have limited its utility compared to methods that employ non-covalent protein-protein interface design[9,10,11,12]. Interface design produces vastly more geometric solutions than genetic fusion, wherein the available alignment geometries at helical termini are spatially discrete and finite in number. This reduces the number of possible structures accessible by fusion and increases the difficulty of building into an assembly any particular building block of interest for a given application. In contrast, the adjustable degrees of freedom (rotation and translation) accessible through noncovalent interface design have continuous ranges, so the set of valid geometric solutions is essentially unlimited. Other issues with the fusion approach are that the termini must be accessible and that flexibility is often introduced at the point of fusion, even with α-helical linkers and especially with disordered linkers. In the best cases, model-deviations are subtle[13,14], however varying levels of heterogeneous off-pathway assembly are also commonly observed [1,4-6].

Genetic fusion has been applied to the creation of cryo-EM scaffolds[15,16,17,18,19]: if a small target protein can be immobilized and rigidly bound onto a larger symmetric assembly, EM particle images can be more readily aligned and binned into class averages than those of the target protein alone. Structures that would normally be too small to analyze would then be amenable to structure determination. Yeates and colleagues demonstrated the potential of this approach by fusing a DARPin to the outside of a previously designed protein nanocage; the bound target was resolved at 3.8 Å resolution with only a single α-helical fusion anchoring the DARPin. We achieve a resolution of only 4.3 Å, but still demonstrate the first dual anchoring strategy where a DARPin is secured through both α-helical fusion and a secondary scaffold interface. Moreover, the tripartite protein assembly method we describe is a general method with applications beyond cryo-EM and is useful in finding geometric solutions where the traditional direct fusion method does not.

A computational method for rigid multi-domain symmetric fusion

We set out to develop a computational method for generating symmetric assemblies by gene fusion that explores vastly more combinations than previous methods and enforces rigid connections between the building blocks. Previous genetic fusion studies have focused on fusing symmetric oligomeric building blocks together at their N and C termini. We reasoned that a much larger set of possible configurations could be generated by (1) incorporating a variable length rigid monomeric protein spacer between the two oligomers, and (2) allowing fusions at internal residues (not just the termini). The number of accessible configurations increases from N(oligomer1) * N(oligomer2) for direct fusion at building block termini, to N(oligomer1) * N(oligomer2) * N(spacer) with the addition of spacer domains, to N(oligomer1)*N(fusion sites per oligomer1) * N(oligomer2)*N(fusion sites per oligomer2) * N(spacer)*N(fusion sites per spacer)*N(fusion sites per spacer) when internal fusion sites are allowed, a very considerable increase. To ensure rigid structurally coherent junctions between the building blocks, we only allow fusion via alignment and superposition of shared helices from both building blocks being fused and disallow α-helical extension. Most globular proteins are destabilized by truncation in the midst of secondary structure elements; to maintain stability, we use idealized repeat protein building block spacers (and oligomers, when possible), where every repeat unit is identical - such proteins are amenable to truncation or fragmentation without undermining folding and stability[20,21,22,23].

Geometric matches to the desired symmetry are identified for each (oligomer1, spacer, oligomer2) tuple by the following procedure, which is illustrated in Figure 1.1. First, all rigid body transforms (T1) on oligomer1 are enumerated that superimpose one of its helical segments onto one from the spacer. Rigid body transforms T2 (applied to oligomer2) are calculated identically, to identify possible fusions between oligomer2 and the spacer. Second, for each (T1, T2) instance combination, the transforms are applied to both oligomers and the arrangement of the axes of the two oligomers are tested for compatibility with the target architecture: they must intersect at 90 degrees (D2) or 60 degrees (D3). If the match is close, the symmetry is idealized by rotating each homo-oligomer about its corresponding shared-helix center-of-mass. In cases where the applied rotation (a measure of non-ideality) is less than the configured error tolerance (5 degrees), 3D models of the resulting assemblies are built by superposition of the repositioned structures. Models are discarded in which the fusions truncate homo-oligomer interface residues, as are models with backbone clashes; sidechain clashes are acceptable, since redesign can eliminate them.

In selecting helical pairs for fusion in the first step of the above procedure, we use several criteria to try to ensure fusion rigidity. First, we require superpositions with a low backbone R.M.S.D. (default 0.5 Å) over multiple residues (default 8) to ensure the shared helix is consistent with both building blocks. The stringency is controlled by user-configurable backbone R.M.S.D. and overlap-length parameters. Unlike linkers, overlapping fusions preserve proximity and sidechain packing of the fusion region with remaining secondary structures from the original building blocks, which reinforces its structure. For the sake of efficiency, this R.M.S.D. and overlap-length thresholding occurs during geometric-match identification, when alignments are initially computed. Second, we rank fusions according to a rigidity metric, which counts the number of putative sidechain contacts between secondary structures of previously separate building blocks. This ensures that significant cross-building-block contacts can be made to buttress the interactions along the shared helix (this step occurs prior to sidechain redesign, so Cα-Cß vectors are used in place of contact counts to measure designability). Third, we redesign with Rosetta[24] the regions adjacent to the new building block junctions to eliminate steric clashes and improve packing between newly joined regions. Fusion-related truncation often exposes hydrophobic residues that were previously buried, so these regions must be redesigned as well.

Design and characterization of D2 and D3 symmetric oligomers

As a proof of concept, we applied the multi-fusion method to design dihedral assemblies from de novo designed repeat protein monomers and oligomers. For D2 structures, two different designed C2 dimers were fused as described above with a repeat protein monomer such that the C2 axes were at 90 degrees and intersect at the origin (Figure 1.2A), while for D3 structures, the intersection angle was 60 degrees (Figure 1.2B). Two different design rounds were performed targeting D2 and D3 symmetries. Even in the second design round, where fewer building blocks and additional selection criteria were imposed, millions of tripartite alignment combinations were scanned per symmetry, nearly ten thousand fusions matched the target geometry within the angular error tolerance (5 degrees), and several hundred passed junction rigidity metrics. The sequences at the junction regions in these assemblies were then optimized using Rosetta, as described in the Methods. In brief, the full symmetric assemblies were generated from the single-chain asymmetric unit and a single round of sidechain redesign with the beta_nov16 score-function was performed on the junction residues, while disallowing the introduction of cysteines, prolines, or methionines. Additional redesign was performed sparingly on regions identified by manual inspection.

In a first design round, 28 D2 and 9 D3 assemblies which matched the geometric selection criteria and had low energy junctions were selected for experimental characterization and recombinantly produced in E. coli. Of these 37 designs, seven (three D2 and four D3) were both soluble and eluted as single monodisperse peaks by size exclusion chromatography (SEC) (Figure 2.1). Scattering profiles and radius of gyration (Rg) determined from solution X-ray scattering (SAXS) data were consistent with design models for the three D2 designs and for two of the four D3 designs (Figure 1.2 and 2.1). Of these designs, native mass spectrometry (native-MS) showed all three of the D2 designs had the total mass expected for the designed assembly, but for one of the two D3 designs (D3-1.5A2) lower order species were observed in addition to the full assembly, when run at moderate concentrations. However, at mid-micromolar concentration (Figure 1.2), full hexamer was the only species detected. Despite the limited dataset, clear trends emerged: for example, all five designs corroborated by SAXS data incorporated a C-terminal ankyrin homo-dimer, which was present in only 60% of the tested designs. In addition, three of four SAXS and native-MS corroborated designs incorporated the same N-terminal three-helix homo-dimer “rop20”, as did the marginal design D3-1.5A2, despite rop20 only being present in ~40% of tested designs. The set of 12 designs that combined both an N-terminal rop20 and C-terminal ankyrin dimer contained three of the four successes and the one marginal design as well. Meanwhile, no designs using hairpin helical bundle dimers were successful.

A second round of design was performed with the same procedure, this time using mostly the rop20 helical bundle at the N-terminus and any of the three similar ankyrin dimers at the C-terminus that had been successful in the first round. We also introduced two additional design constraints with the aim of improving the designs’ suitability as cryo-EM DARPin scaffolds. First, structures with fewer total secondary structures (helices) were selected, anticipating that they would be more rigid. Although the design method reinforces the point of fusion, the DHR and ankyrin building blocks have small cross-sectional areas, so each additional repeat likely adds flexibility. Second, we selected designs with the ankyrin binding groove facing away from the assembly center (Figure 2.3), to reduce the chance of steric hindrance in multivalent target binding (described in the next section). 31 D2 and 24 D3 designs were ordered, of which 15 and 12 respectively had high levels of soluble expression and single major peaks by SEC. SAXS and native-MS agreed with the design models of five D2 designs and five D3 designs (designs D2-21.29 and D2-21.30 have 85% sequence identity so they should not be considered independent successes).

With the exception of design D2-21.22, negative-stain EM 2D class averages and 3D reconstructions recapitulated the expected shape for all designs that passed both native-MS and SAXS screening. As shown in Figures 1.2 and 1.3, many designs have a pronounced central cavity that makes their top views readily identifiable in micrographs. An unusual structural aspect of our D3 designs is their subunit connectivity; in natural D3 architectures, subunits related through C3 rotations normally make direct contact. In almost all of our designs, they do not. Instead, C3-related subunits are positionally fixed relative to one another only indirectly by shared binding partners with which they share C2 axes; the C3 axis is an emergent property. This means that each subunit is anchored by only two neighbor subunits, as opposed to three or more in a typical natural D3 architecture. Another notable feature of our designs is that they were produced in large part from building blocks without crystal structures. Of the ten successful and sequence-independent designs, two use not a single crystal-verified building block (D2-1.4H and D3-19.24) and another seven use only one crystal-verified building block (Table 2.2).

Cryo-EM of coassembled DARPin and GFP

In prior studies, imaging scaffolds have been constructed through fusion of a DARPin onto existing symmetric protein assemblies via a shared helix or α-helical extension[18,19]. Our constructs differ in that they incorporate a designed ankyrin dimer in addition to fusion, so the ankyrin (or DARPin post interface-grafting) is positioned by two separate mechanisms that should both contribute to rigid placement. To benchmark our scaffolds against prior work, we grafted the surface binding residues from a GFP-binding DARPin[25] into six constructs: D2-1.1D, D2-1.4H, D3-1.5C, D2-21.7/8, D2-21.29, and D3-19.20, taking care not to alter the ankyrin homo-dimerization interface in the base construct (Figure 2.2). As ankyrins are repeat proteins, the DARPin alignment can be shifted up or down by one or more repeats, which we did to create variants. After SEC purification and cryo-EM of GFP-scaffold co-complexes, a majority of constructs exhibited high levels of preferred orientation or outright aggregation (Table S3), but D2-1.4H and D2-21.7/8 derivatives seemed promising. D2-1.4Hv1 is derived from a first round design scaffold D2-1.4H and was reconstructed at 4.3 Å resolution for the core scaffold, full complex, and GFP target (Figure 1.4).

Cryo-EM of coassembled DARPin and Human Serum Albumin (HSA)

We similarly incorporated an anti-HSA DARPIN into designed assemblies D2-21.8 and D2-21.29 with the intent of determining the DARPin binding mode and also to assess the feasibility of integrating DARPins in the absence of co-complex structural information, as existed for the GFP-binding DARPin. Since the HSA-DARPin co-complex structure was not known, the interface-grafted assembly was created with sequence alignment alone, as in Figure 2.3. Binding to HSA for the D2-21.8v1 based scaffold was confirmed by SDS-PAGE of elution fractions, which showed co-elution of HSA and scaffold, on a Superdex 200 5/150 GL Increase. Then, a sample for cryo-EM was purified on a Superdex 200 10/300 GL Increase, as it provides better size separation.

Co-complexes with varying numbers of HSA were observed, so cryo-EM reconstruction efforts focused on a single major species. A reconstruction was created at 5.5 Å resolution (Figure 1.5), sufficient to resolve the relative placement of core scaffold and bound HSA. The scaffold has four binding sites for HSA, but the reconstruction showed only two were occupied by HSA due to steric hindrance between HSA copies. The co-complex structure (Figure 1.5) also shows that the DARPin binds a hydrophobic region at the junction of HSA's DII and DIIIA domains. Since the DARPin binding occurs away from the HSA's FcRn binding site, it can be inferred that protein therapeutics may exhibit increased serum half-life through FcRn recycling if the HSA-binding DARPin is fused. The structure of HSA is known and it is also above the size resolution limit imposed by cryo-EM, so DARPin-HSA in complex could have been determined without the scaffold at all. Still, the effort here demonstrates success in interface grafting without structural information, which indicates possible success for structure determination with smaller structures.

Discussion

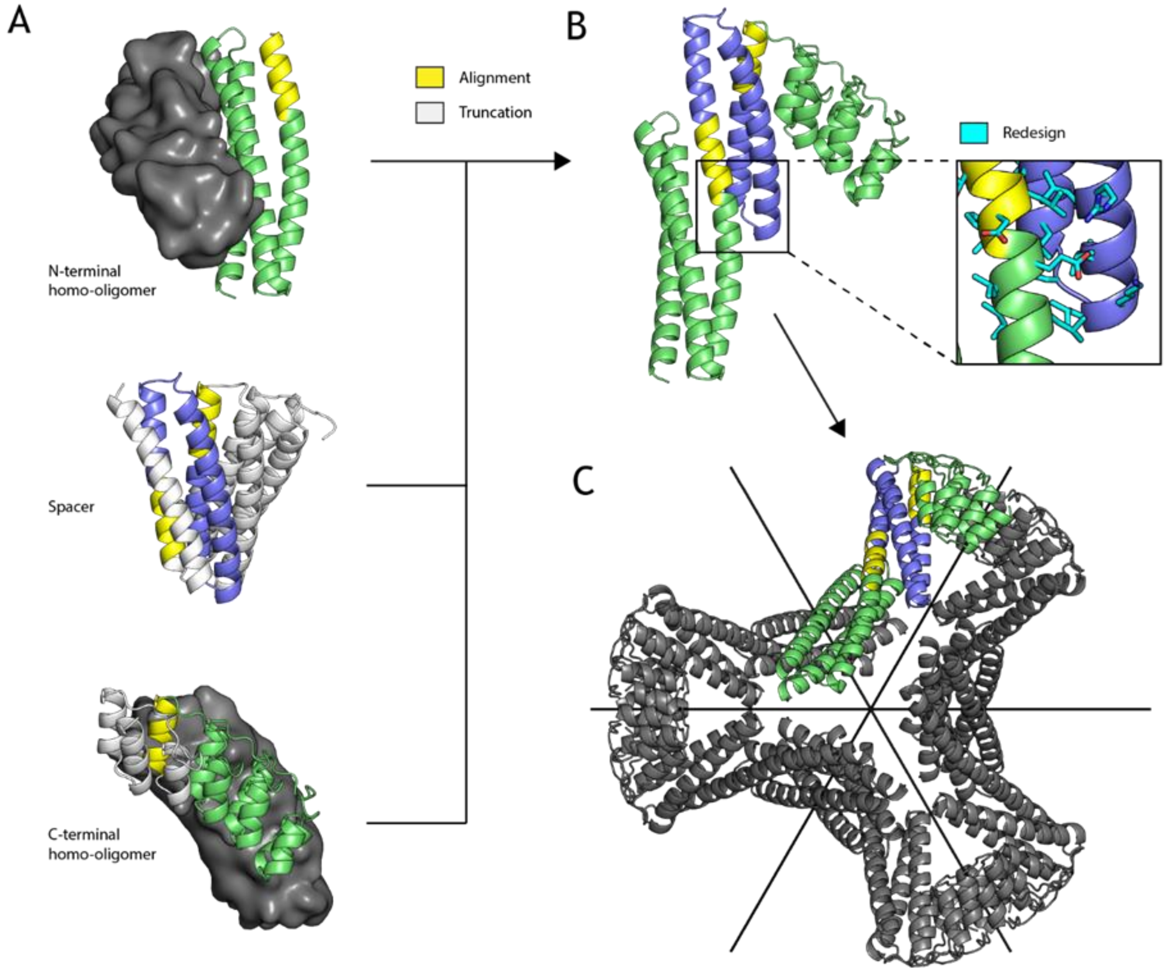
The multi-fusion method introduced here specifically remedies two of the long-standing drawbacks to fusion-based assembly construction: the low number of geometric solutions and linker flexibility. Our results demonstrate the feasibility of performing multiple fusions, ranking, and redesign in a single pass to produce intended architectures with low levels of off-pathway assembly. It remains to be seen how well the method performs for more complex and multi-component assembly architectures. The role of rigidity is also an open question: the cryo-EM application would clearly benefit from a more rigid overall structure, but the design method may require a measure of component flexibility to compensate for design inaccuracies around the redesigned fusion. It is notable that all successful designs in the present work incorporated at least one, usually two, and occasionally even all three design models lacking high resolution structural validation (SAXS validation only); it is possible that using x-ray crystal structures of the building blocks rather than computational design models might result in more working designs. On the other hand, the success using design models bodes well for future applications of this approach as far more plausible building block structures can be designed than can be solved by x-ray crystallography.

We had hypothesized that DARPins would be more rigidly embedded in our new designed assemblies than in the previous designs by Yeates et al, as they form structural components of the assembly and are tethered by either a fusion or designed interface at either end. However, the resolution we achieved was half an angstrom lower than in the Yeates study.

In that study, the core scaffold was resolved to 2.9 Å and the GFP target was resolved to 3.8 Å, indicating additional flexibility in the DARPin relative to the core scaffold. In this study, the core scaffold and GFP target were both resolved to the same 4.3 Å resolution, which we interpret as the dual-tethering strategy being effective in rigidifying the DARPin placement with respect to the rest of the scaffold. However, this benefit was neutralized by some combination of inherent flexibility in the core dihedral scaffold and lower order symmetry. We knew that rigidity of the core scaffold was important and expected that the shorter closure path selected for in the second-round designs would improve rigidity and produce better scaffolds than those from the first round of design. Instead, the second round designs showed a tendency toward preferred orientation and aggregation (Table S3), which may be rooted in our selection criteria; while the shorter closure paths may indeed increase rigidity, it also biased toward smaller designs that are likely more sensitive to destabilization when grafting in the hydrophobic DARPin interface. In line with this hypothesis, D2-1.4Hv1 was well behaved post-grafting and also contains the largest (in terms of cross-section width and total mass) DHR building block among designs tested by cryo-EM. While the scaffolds here fall short of recapitulating or improving upon the 3.8 Å GFP resolution achieved by Liu et al., the design methods are quite general, and secondary stabilization through interface design could still yield improved results if used in conjunction with bulkier building blocks and higher order symmetry. Beyond scaffolds for cryo-EM structure determination, our new method facilitates exploration of the vast protein nanomaterial design space.

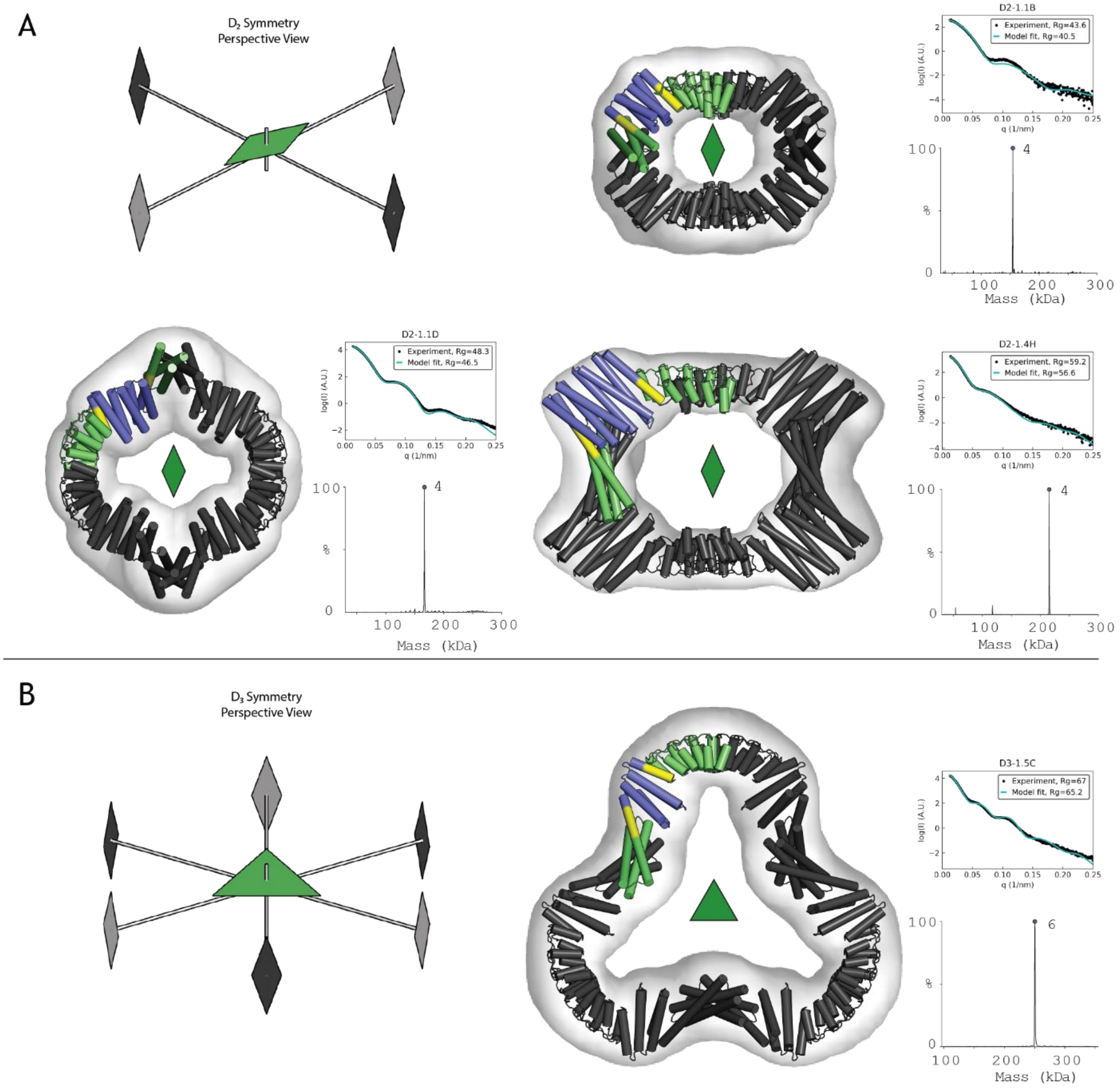
Figures

Figure 1.1



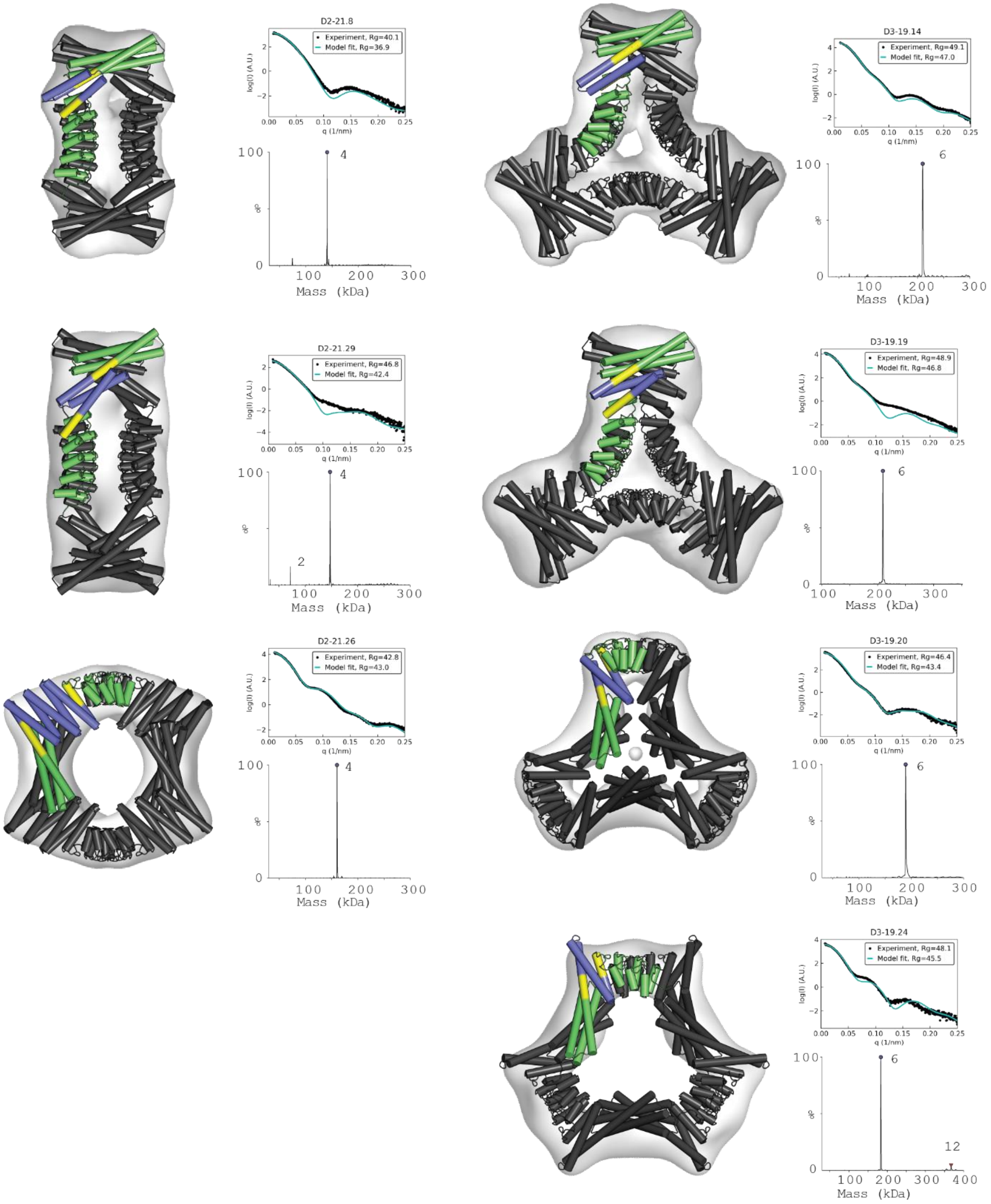
The tripartite design strategy for the D3 architecture is depicted. The final structure is composed of (A) two homo-dimers (top and bottom; the partner subunit is shown as a surface) and a de-novo helical repeat protein (middle). All possible non-clashing backbone alignments are geometrically analyzed and filtered to generate (B) a three-component fusion, which is redesigned to improve core packing and remove exposed hydrophobics. The result is (C) a D3 assembly, with symmetric C2 axes (drawn in black) that correspond to those of the original homo-dimers and a C3 axis orthogonal and through the center.

Figure 1.2



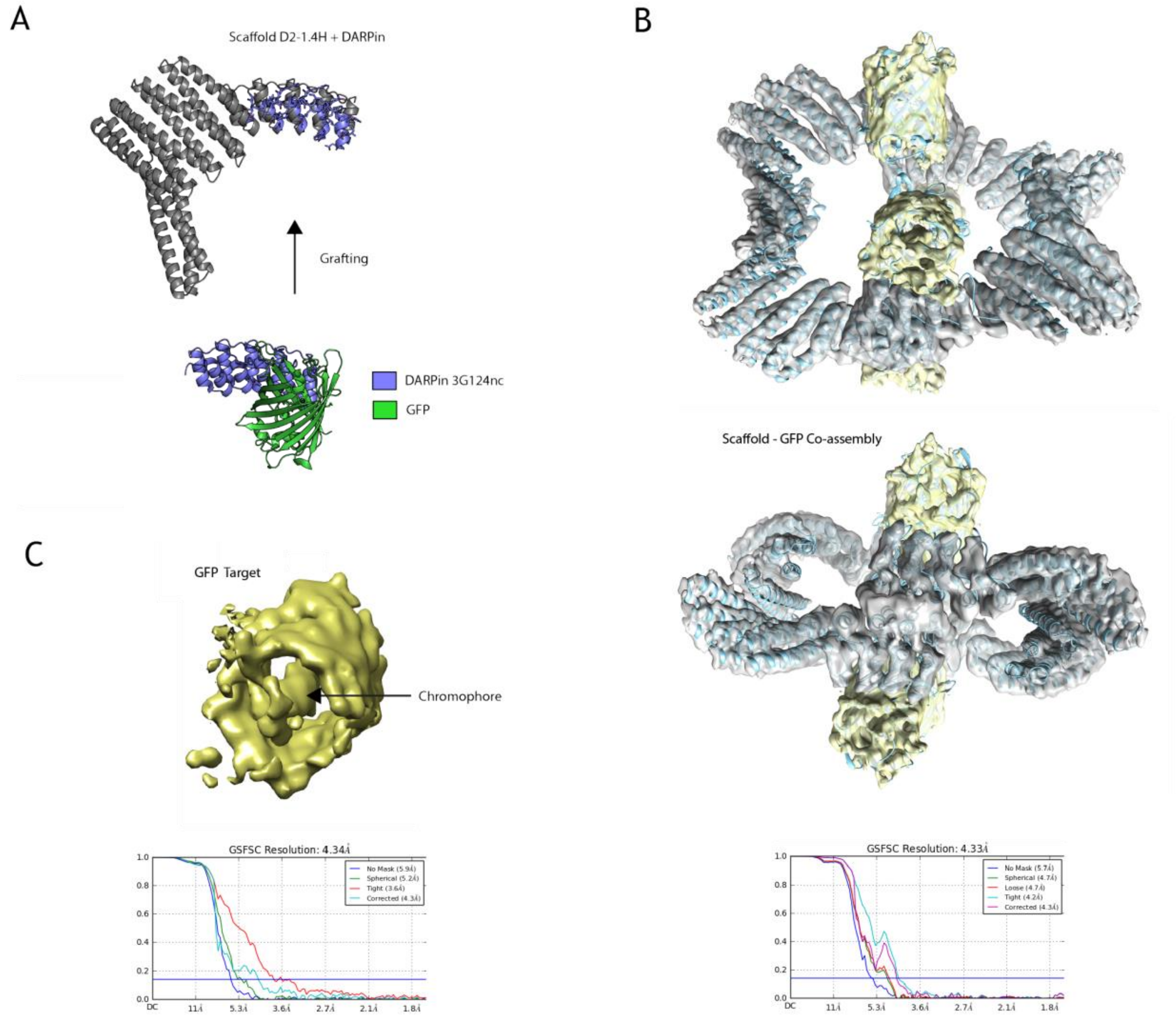
EM, native-MS, and SAXS experiments are consistent with the formation of intended architectures for four models from the first design round. Negative-stain 3D reconstructions are overlaid by design models, whose asymmetric unit is colored according to its constituent building blocks (oligomers, green; spacer, purple; shared alignment, yellow). Native-MS deconvolutions show the relative abundance of different molecular weight species, with the expected reading and oligomerization state noted. SAXS plots compare the theoretical (cyan) and experimental (black) scattering intensities (log scale) as a function of q, as well as radius of gyration (Rg).

Figure 1.3


D2-21.8
Experiment, Rg=40.1
Model fit, Rg=36.9
log(I) (A.U.)
q (1/nm)
4
D2-21.29
Experiment, Rg=46.8
Model fit, Rg=42.4
4
2
D2-21.26
Experiment, Rg=42.8
Model fit, Rg=43.0
4
D3-19.14
Experiment, Rg=49.1
Model fit, Rg=47.0
6
D3-19.19
Experiment, Rg=48.9
Model fit, Rg=46.8
6
D3-19.20
Experiment, Rg=46.4
Model fit, Rg=43.4
6
D3-19.24
Experiment, Rg=48.1
Model fit, Rg=45.5
6
12
100
%
0
Mass (kDa)
100
200
300
400

EM, native-MS, and SAXS experiments are consistent with the formation of intended architectures for seven models from the second design round. Design models are overlaid on negative-stain 3D reconstructions, with the model asymmetric unit colored according to its constituent building blocks (oligomers, green; spacer, purple; shared alignment, yellow). Native-MS deconvolutions show the relative abundance of different molecular weight species, with the expected reading and oligomerization state noted. SAXS plots compare the theoretical (cyan) and experimental (black) scattering intensities (log scale) as a function of q, as well as radius of gyration (Rg). Design D3-19.24 showed low levels of 12-mer that likely result from association between the designed hexamers.

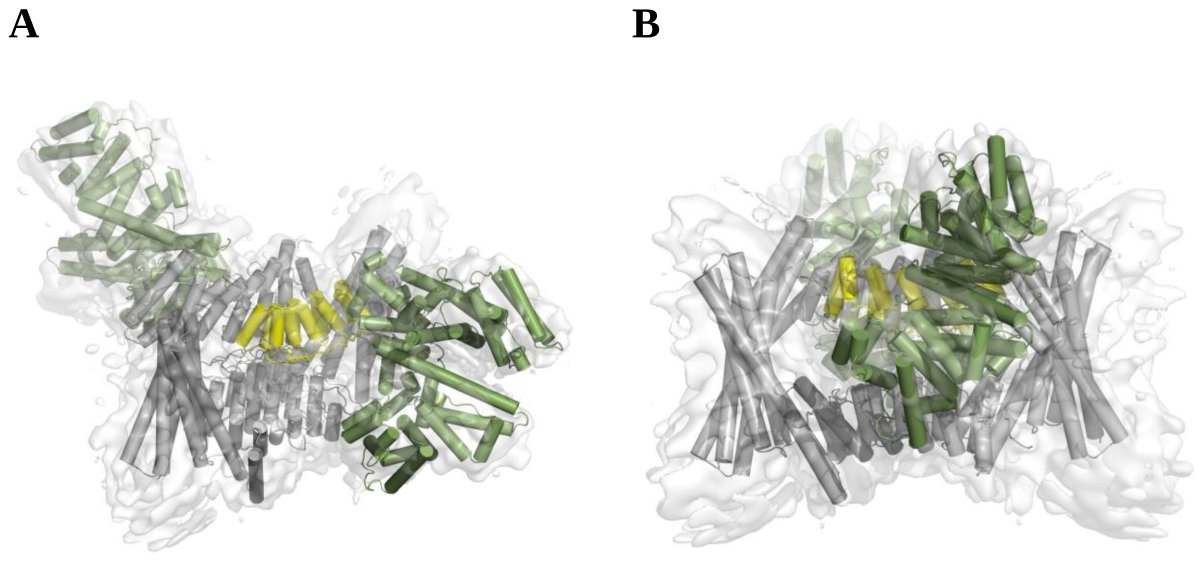
Figure 1.4



Cryo-EM data processing of co-assembled DARPin-grafted scaffold and GFP. (A) Surface residues of the GFP-binding DARPin “3G124nc” are grafted onto the assembly subunit of D2-1.4H, while preserving core residues and homo-oligomer interfaces, to form a hybrid that still self-assembles and also binds GFP. (B) Cryo-EM density of the co-assembled complex is inlaid with a cartoon model diagram. The Fourier shell correlation (FSC) curves are shown with a

0.143 cutoff indicated by a horizontal line. (C) GFP density with the core scaffold masked and corresponding FSC curves are shown.

Figure 1.5



A cryo-EM reconstruction is shown of the co-complex between HSA and HSA-binding scaffold variant of D2-21.8, with embedded cylinder models. Two copies of HSA (green) are bound to the scaffold (gray) and one of the four binding ankyrins/DARPins within the scaffold is highlighted (yellow). Panel (A) contains a perspective view and panel (B) shows a front view.

Section2. Supplementary information for the generation of ordered protein assemblies using a rigid multi-domain approach

Computational design

A custom software library was built with the .Net Framework, which includes functionality for PDB parsing, alignment, symmetry/patterning, clash and contact checking, structure editing, and running the multi-domain fusion algorithm. The several parameters that control fusion were assigned based only on manual curation of outputs during testing in-silico and are likely not optimal for all scenarios. In particular, an 8-residue minimum overlap-length was selected because the idealized ankyrins used in this study have short helices compared to those of DHRs and helical bundles, but longer overlaps might be desirable with other starting components. Likewise, a lower angular-error tolerance might increase the success rate of tested designs, but it was kept at a moderately high 5 degrees, because the lack of crystal structures for so many designs introduced uncertainty about the initial model accuracy, so a tight angle tolerance would have been somewhat arbitrary.

As the method creates a much larger solution space than direct fusion, optimizations were necessary to keep runtimes reasonable while still exhaustively enumerating geometries. The most impactful optimization eliminates redundant alignments by greedily expanding the alignment windows of any valid 8-residue alignment until the R.M.S.D. threshold is exceeded or either secondary structure element ends; all shorter alignment windows contained within the expanded alignment need not be examined. The result is that fewer alignment combinations are considered than if every 8-residue window were examined and nearly identical outputs are largely avoided. The protocol could be run on all building-block combinations in our set of de-novo scaffolds for a target geometry in less than 24 hours on a quad-core laptop; higher parallelism would produce a speedup accordingly.

The fusion output models were redesigned by Rosetta with a very simple RosettaScripts protocol (Text File S1), involving only two Movers (operators that modify a design model): SetupForSymmetry and SymPackRotamersMover. These Movers respectively recreate the full symmetric assembly from the input single-chain asymmetric unit and redesign those residue side chains that were identified by output files in the Resfile format. After the initial sidechain redesign pass, models deemed promising by a combination of total score and manual inspection were subjected to one or more additional redesign passes with the same protocol, but with user-generated Resfiles, to eliminate exposed hydrophobic residues, revert residues to their original wildtype identity when deemed appropriate, or mutate Rosetta-designed glycines to alanines within helices to improve helical propensity. The beta_nov16 score function was used throughout.

The input structure set consisted of 20 homo-dimer and 42 DHR spacer proteins already verified within the lab, with 5 homo-dimers and 15 DHRs having been previously published with solved crystal structures available in the Protein Data Bank[26,27,28] (Table S4). Two designed crystal structures were unintentionally omitted from the input set (2L4HC2_4 and 3L6HC2_2 from Boyken et al). Two-helix dimers were removed from the scaffold set in the second round of design, because better results were obtained from three-helix dimers. The third helix leads to a larger hydrophobic core than exists in the two-helix dimers, which we expect leads to a higher degree of degree of order even in the monomeric form and might help to avoid aggregation and misassembly. The other type of successfully incorporated dimer was based on ankyrins.

Although very similar, the minor binding orientation differences between the three ankyrin homo-dimers was sufficient to make all three useful in finding distinct geometric solutions.

Cryo-EM of coassembled DARPin and GFP

Electron microscopy grids were prepared at 4°C at 100% humidity using vitrobot (FEI). In brief, 3 µl of purified sample at 1.0 mg/ml was applied to glow discharged Quantifoil 200 mesh R1.2/1.3 grid, and was manually blotted with a filter paper (Whatman No. 4) for approximately 3 seconds before plunging into liquid ethane. The grids were screened on a Talos Arctica 200 kV with K3 direct electron detector for ice thickness and sample distribution. Micrographs of the screened grid were collected on a Titan Krios microscope (FEI) operating at 300 kV with energy filter (Gatan) and equipped with K2 Summit direct electron detector (Gatan), using data collection program SerialEM[29]. A nominal magnification of 165,000x was used for data collection, corresponding to a pixel size of 0.865Å at specimen level, with the defocus ranging from -1.0µm to -3.0µm. Movies were recorded in superresolution mode, with a total dose of -60 e-/Å2 and dose rate of 8.4 electron per pixel per second and fractioned into 40 frames. Movies were decompressed and gain-normalized using the program Clip in IMOD. Raw movies were corrected for beam-induced motion and binned by two using MotionCor2[30], and exposure filtered in accordance with relevant radiation damage curves[31]. The CTF estimation was performed with GCTF[32] on non-dose weighted micrographs. Micrographs with high CTF Figure of Merit scores and promising maximum resolution (better than 3.6 Å) were selected for further processing (total 1532 micrographs). Several rounds of autopicking using combinations of different references and manual picking were analyzed to determine optimal settings, and yielded similar results. These particles were subjected to iterative rounds of 2D classification, subset selection of high-quality classes, and re-extraction, yielding 138348 particles from 1023 micrographs, all in RELION 3.0[33]. The initial model was de novo generated and subsequent 3D heterogenous refinement was performed using cryoSPARC[34]. Particles from the best quality 3D class were selected for further processing. The UCSF PyEM package[35] was used to convert the cryoSPARC coordinates into RELION. The resulting particles were analyzed by 3D refinement, Bayesian Particle Polishing and CTF Refinement in RELION with C1 or D2 symmetry. All the reconstructions were analyzed using UCSF Chimera[36]. The coordinate model was built by breaking the initial design model into domains and rigidly docking these individual protein structures into the EM map using Chimera. Once the orientation was identified, the model was then fit and adjusted manually in Chimera and Coot[37]. The local resolution and final Fourier shell correlation were calculated using Resmap[38] and cryoSPARC. The core resolution was calculated using the validation function in cryoSPARC.

Native mass spectrometry

Sample purity and oligomeric state was analyzed by online buffer exchange MS[39] using a Vanquish UHPLC coupled to a Q Exactive Ultra-High Mass Range (UHMR) mass spectrometer (Thermo Fisher Scientific)[40,41] modified to allow for surface-induced dissociation (SID) similar to that previously described[42]. 1 µL of 50 µM protein in 25 mM Tris and 150 mM NaCl were injected and online buffer exchanged into 200 mM ammonium acetate, pH 6.8 by a self-packed buffer exchange column (P6 polyacrylamide gel, Bio-Rad Laboratories) at a flow rate of 100 µL per min. A heated electrospray ionization (HESI) source with a spray voltage of 4 kV was used for ionization. Mass spectra were recorded for 1000 – 20000 m/z at 3125 resolution as defined at 400 m/z. The injection time was set to 200 ms. Voltages applied to the transfer optics were optimized to allow for ion transmission while minimizing unintentional ion activation, and a higher-energy collisional dissociation (HCD) of 5 V was applied. Mass spectra were deconvoluted using UniDec V4.2.2[43]. Deconvolution settings included mass sampling every 10 Da, smooth charge states distributions, automatic peak width tool, point smooth width of 10, and beta of 50 (artifact suppression).

Protein Expression and purification

DNA sequences encoding proteins with 6xHis tags were codon optimized by Genscript and cloned into pET28b+ or pET29b+ vector under the control of a T7 promoter. Plasmids were transformed into BL21(DE3) E. Coli and plated on LB agar plates. On different occasions, either 50ml or 500ml expression cultures were used. 50ml expression cultures were directly inoculated from plate colonies and grown for 24 hours in Studier's autoinduction media[44] with shaking. Alternatively, 5ml starter cultures in TB were inoculated and grown for 9-12 hours before transfer to 500ml autoinduction media for 16-18 hours. All growth media was prepared with 100uM kanamycin as a selection antibiotic.

Expression cultures were spun down for 10 minutes at 4,000 rcf, resuspended in 40ml TBS (150mM NaCl, 25mM Tris) with Pierce protease inhibitor (Product No. A32963), and lysed by sonication. Lysate was spun at 25,000 rcf for 40 minutes to separate the insoluble fraction. The soluble fraction was purified by affinity chromatography over Ni-NTA Agarose (Qiagen) gravity columns. Eluate was concentrated and fractionated by SEC on a Superdex 200 Increase 10/300 GL.

Negative-stain EM

PELCO 300 mesh Copper grids with Carbon film (Product 01843-F) were glow-discharged and 3ul of sample in TBS was applied to the grid and blotted immediately. 3ul 2% uranyl formate stain was applied and blotted immediately, twice, and then allowed to dry.

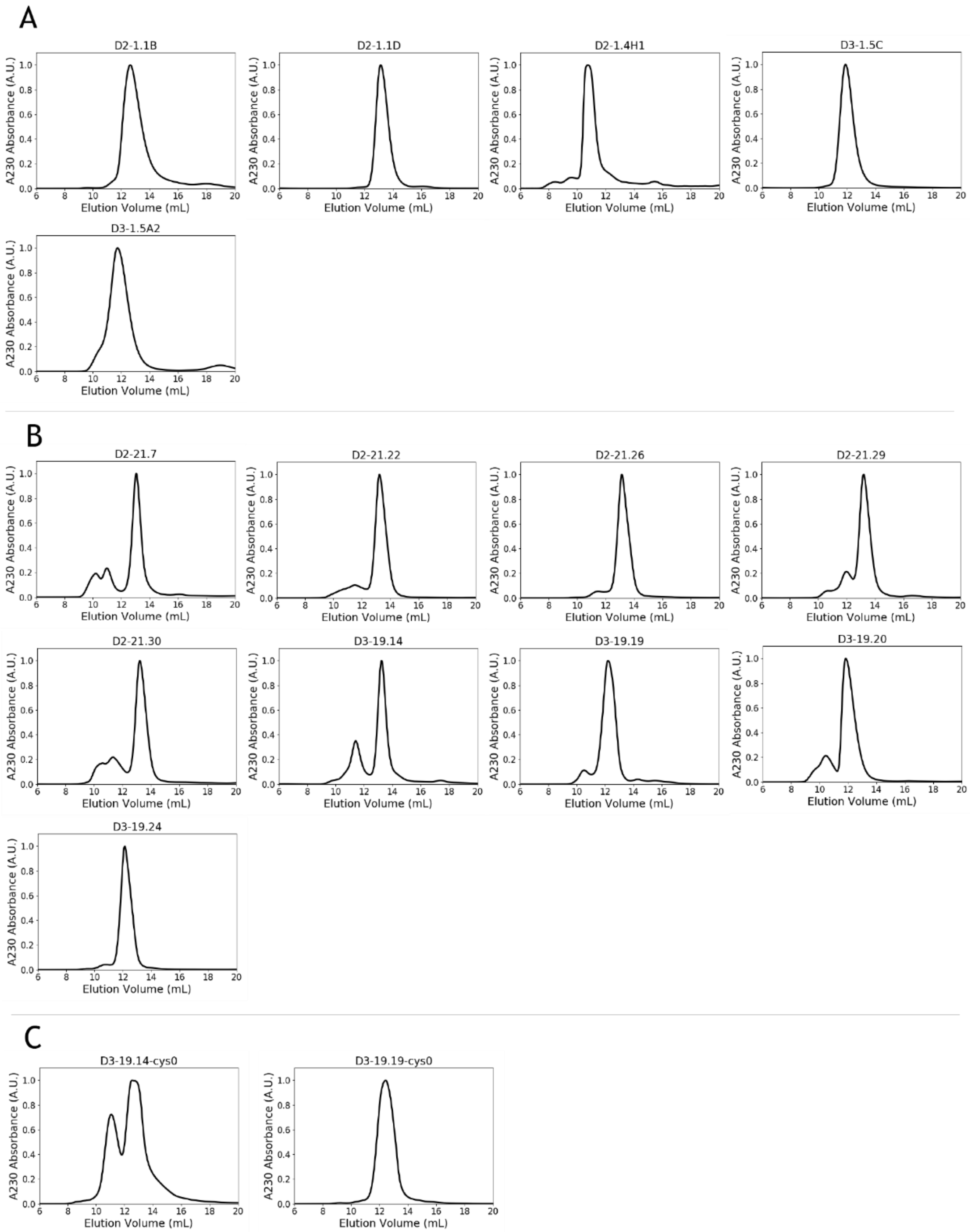
Approximately 50 micrographs per construct were recorded on a Thermo Scientific Talos transmission electron microscope with 200kV source. The known symmetry (D2 or D3) was applied during reconstruction, except for designs D3-19.14 and D3-19.19, for which C1 symmetry was applied (although the design model is D3). 3D reconstructions were generated in either Relion or cisTEM (v 1.0.0-beta).

SAXS analysis

SAXS data were collected at the SYBILS Beamline (Advanced Light Source in Berkeley, CA) via their Mail-In SAXS program. KNO3 was added to buffer solutions in the range of 2 to 5 mM to minimize radiation-damage induced aggregation. Samples were concentrated in Amicon Ultra 0.5ml centrifugal filters and flow-through was used as the background subtraction buffer. For each sample, the average scattering profile was computed, excluding data in the Guinier region for timepoints after radiation damage became observable. The Scatter software was used for analysis; model and experimental Rg values were determined from their respective Guinier region data. Combined datasets (model-vs-experiment) were generated with the FOXS web server[45,46] for plotting.

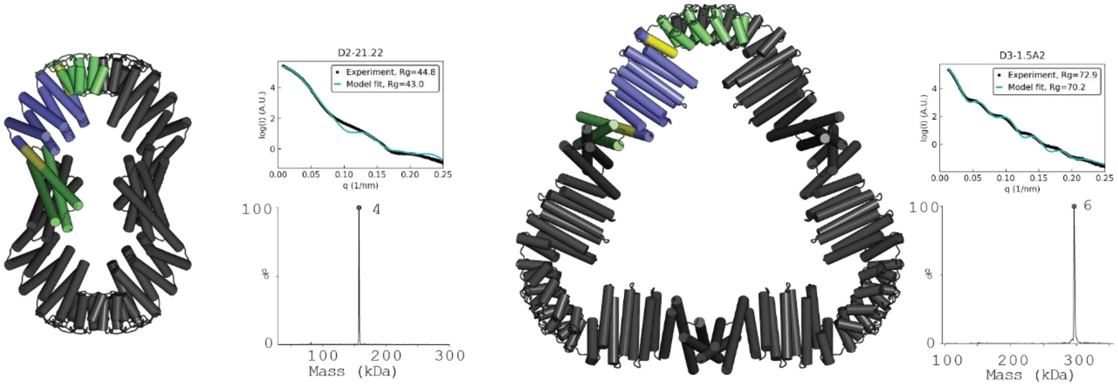
Figures

Figure 2.1



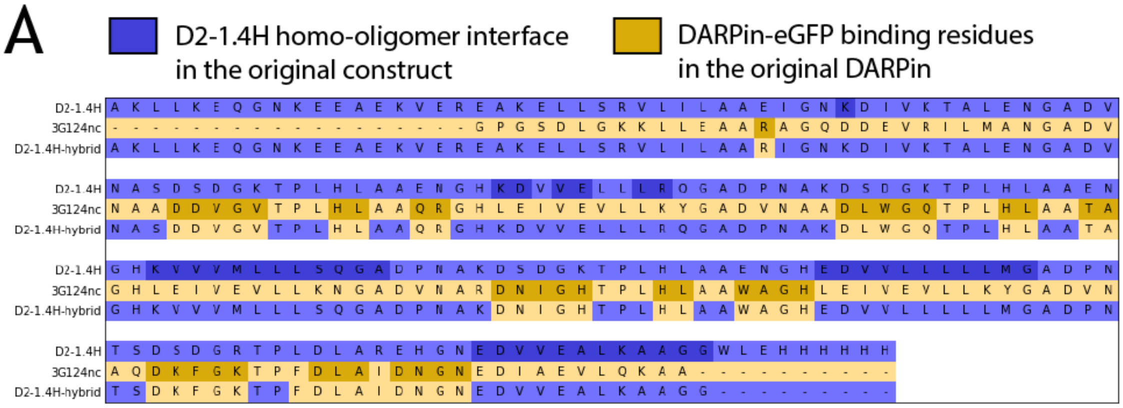
Size exclusion chromatography of post-IMAC eluate (not rechromatography) shows a dominant species for constructs from the (A) first and (B) second design rounds. The minor peak for D3-19.14 could not be resolved by native-MS, however in (C) a disulfide variant was created for which the minor peak was more prominent - native-MS confirmed the peak as octameric. This suggests the minor peak assembly in both the D3-19.14 original and disulfide variant D3-19.14-cys0 has D4 symmetry. In the case of D3-19.19, the same cysteine mutant was not disruptive. That the same disulfide across the ankyrin dimer has different effects in the two structures (D3-19.19 and D3-19.14) suggests that the interface geometries differ due to the context of the surrounding structures.

Figure 2.1



Certain designs showed promising results according to SAXS and native-MS, but appeared disordered by EM. It may be that the assemblies are sensitive to the low pH of uranyl formate stain or exhibit more concentration-dependent oligomerization than those that were validated by EM.

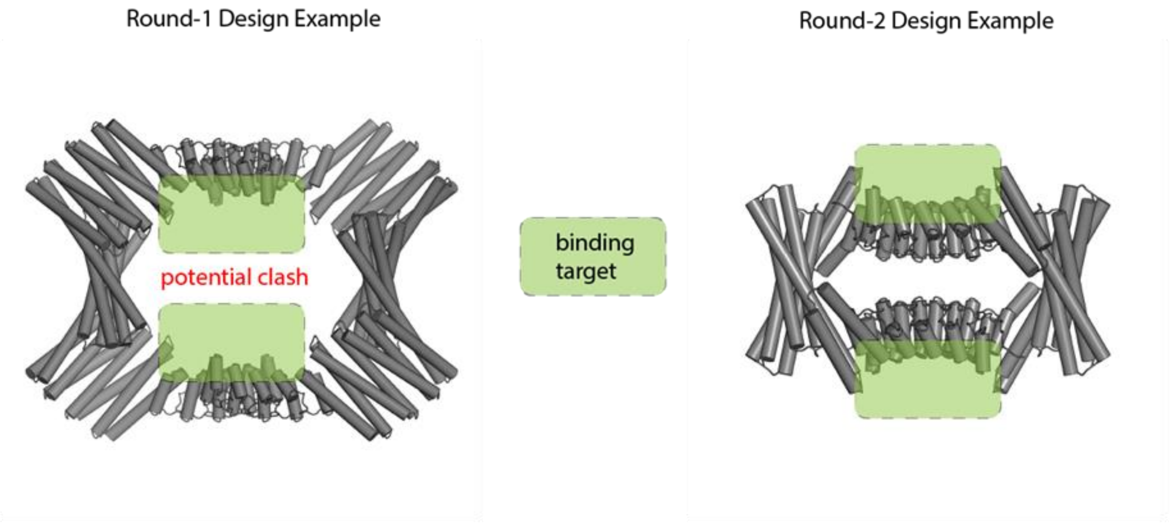
Figure 2.2



```
D2-1.4H         A K L L K E Q G N K E E A E K V E R E A K E L L S R V L I L A A E I G N K D I V K T A L E N G A D V
3G124nc         - - - - - - - - - - - - - - - - - - G P G S D L G K K L L E A A R A G Q D D E V R I L M A N G A D V
D2-1.4H-hybrid  A K L L K E Q G N K E E A E K V E R E A K E L L S R V L I L A A R I G N K D I V K T A L E N G A D V

D2-1.4H         N A S D S D G K T P L H L A A E N G H K D V V E L L L R Q G A D P N A K D S D G K T P L H L A A E N
3G124nc         N A A D D V G V T P L H L A A Q R G H L E I V E V L L K Y G A D V N A A D L W G Q T P L H L A A T A
D2-1.4H-hybrid  N A S D D V G V T P L H L A A Q R G H K D V V E L L L R Q G A D P N A K D L W G Q T P L H L A A T A

D2-1.4H         G H K V V V M L L L S Q G A D P N A K D S D G K T P L H L A A E N G H E D V V L L L L L M G A D P N
3G124nc         G H L E I V E V L L K N G A D V N A R D N I G H T P L H L A A W A G H L E I V E V L L K Y G A D V N
D2-1.4H-hybrid  G H K V V V M L L L S Q G A D P N A K D N I G H T P L H L A A W A G H E D V V L L L L L M G A D P N

D2-1.4H         T S D S D G R T P L D L A R E H G N E D V V E A L K A A G G W L E H H H H H H
3G124nc         A Q D K F G K T P F D L A I D N G N E D I A E V L Q K A A - - - - - - - - - -
D2-1.4H-hybrid  T S D K F G K T P F D L A I D N G N E D V V E A L K A A G G - - - - - - - - -
```

Alignment-based construction of a hybrid DARPin that binds eGFP. An alignment is performed during DARPin grafting to ensure that residues responsible for homo-oligomerization in the base construct (D2-1.4H) are preserved after hybridization with the DARPin. Binding residues (D2-1.4H homo-oligomerization and the DARPin-GFP interface) in the source constructs are darkened and the hybrid construct is colored according to whichever source construct the sequence was based on.

Figure 2.3



DARPin orientation criteria: in the first design round, DARPin orientation was not considered and might be expected to cause steric clashes between multiple bound targets. In round 2, designs were selected such that the DARPin binding groove (the region between helices and loops responsible for target-binding) was expected to orient bound targets away from one another to prevent clashes. In the general case of target structure determination by cryo-EM, the target binding mode is unknown to begin with, but it is expected that the DARPin orientation in round 2 will be better for most DARPin-target complexes.

Tables

Table 2.1 –Design Sequences

Round-1 Design Sequences:

>D2-1.1B

MASDYLRLATEHNKLATEAASLAAELAASAVTLTVTDPSKTAQEHTELASRLLEMMSQ

FLKAAQELTREAIRKEGRNEESEKTLRKSKSSYKALKALLKAIKAIEKGDVETAVRAAQ

EAVRLASEAGNNNVLRAVAEVALAIAKVAEEQGNVEVAVKAAQVAVSAALNAGDED

VLKKVAEQASRISKEAEKQGNQEVSKKALSVSLIAAAASGDKDLVKDLLESGADVNAS

SSDGKTPLHVAAENGHAKVVLLLLEQGADPNAKDSDGKTPLHLAAENGHAVVVALLL

MHGADPNAKDSDGKTPLHLAAENGHEEVVILLLAMGADPNTSDSDGRTPLDLAREHGN

EEVVKVLEDHGGWLEHHHHHH

>D2-1.1D

MASEKARIAVENLEAALRLNKAAIEMAKSAIKITRDNSSDEKATRYSLLTAKVLVMSLE

LLTASLELAEKALREEGSDDSAEKVRKEAEEILSKAVEEAVRVMQEMVTIMKRTGSNDS

LREVAELALRVAKAAEKAGNVEVAVQAARVAVEAAKQAGDNDVLRKVAEQALRIAK

EAEKQGNVEVAVKAAKVAVEAAKQAGDEDVLKKVAEQASRIASEASKQGNKEVASK

ALIVAAQAGSKEAVKKAIESGADVNASDSDGRTPLHHAAENGHAEVVALLIEKGADVN

AKDSDGRTPLHHAAENGHDEVVLILLLKGADVNAKDSDGRTPLHHAAENGHKRVVLV

LILAGADVNTSDSDGRTPLDLAREHGNEEVVKALEKQGGWLEHHHHHH

>D2-1.4H

MGSEKARIAVENLEAALRLNRAAAEMQKSAIKIMDDNSDDEKALRYLRLTTKVLRMSV

ELLRASLELAEKALREEGSDDSAEKVRKEAEEILKESTAILKLADAATKVADIKHDIKKA

KEQQEQGNKEEAEKTLREATEKIKRVTEELEKIAKNSKTPEIALKAAEALVKLIKLLIEIA

KLLQEQGNKEEAEKVLREATELIKRVTELLEKIAKNSDTPELALRAAELLVRLIKLLIEIA

KLLQEQGNKEEAEKVLREATELIKRVTELLEKIAKNSDTPELASRAAELLVRLIKLLQEIA

KLLKEQGNKEEAEKVEREAKELLSRVLILAAEIGNKDIVKTALENGADVNASDSDGKTP

LHLAAENGHKDVVELLLRQGADPNAKDSDGKTPLHLAAENGHKVVVMLLLSQGADPN

AKDSDGKTPLHLAAENGHEDVVLLLLLMGADPNTSDSDGRTPLDLAREHGNEDVVEAL

KAAGGWLEHHHHHH

>D3-1.5C

MGSEKARIAVENLEAALRLNRAAAEMQKSAAKIVADNASDEKALRYLRLTTKVLRMS

VELLRASLELAEKALREEGSDDSAEKVRKEAEEILKESTEILKEADQITEVADLAFELAN

KATDEELRKEISKCARLALELASRSTNDELIKQILEVAKLAFELASKATDEELIKLILKCC

QAAFERASRSTNDEEIKKILEVAKRAFETASKATDEEEIKSILLICAAALGNKDAVKSAIE

NGADVNASDSDGRTPLHHAAENGNAEVVALLIEKGADVNAKDSDGRTPLHHAAENGH

DEVVLILLLKGADVNAKDSDGRTPLHHAAENGHKRVVLVLILAGADVNTSDSDGRTPL

DLAREHGNEEVVKALEKQGGWLEHHHHHH

>D3-1.5A2

MGSEKARIAVENLEAALRLNRAAAEMQKSAAKIVADNASDEKALRYLRLTTKVLRMS

VELLRASLELAEKALREEGSDDSAEKVRKEAEEILKESTEILKLADQITEVADLAFELAN

KATDEELRKEIVKCAKLALELASRSTNDELKKQILEVAKLAFELASKATDEELIKEILKCC

QLAFELASRSTNDELIKQILEVAKLAFELASKATDEELIKEILKCCQLAFELASRSTNDELI

KLILEVAKAAFERASKATDEEEIKEILKVCQEAFEEASRSTNDEEIASILLVAAALLGNKD
AVKDAIENGADVNASDSDGRTPLHHAAENGNAEVVALLIEKGADVNAKDSDGRTPLH
HAAENGHDEVVLILLLKGADVNAKDSDGRTPLHHAAENGHKRVVLVLILAGADVNTS
DSDGRTPLDLAREHGNEEVVKALEKQGGWLEHHHHHH

Round-2 Design Sequences (D2 symmetry):

>D2-21.6

MHHHHHHGSGSALEKIAKLIIEAARLSAELARRAARASAEMARLAIEAVSKERGSSELL
KIVADLIVEAQEAVVRLIIESQQIAAKLAEDLIRAAKEAASDESKMEEVAKEVQERAERA
ARDIEEKLLKVLIELIKKLARSIGDEERLKATKLAEEAIRVAREVGDSLLERIALEAALKG
DSRAAKAVLKAAELAREAAERGDEEKVKAAALIAAAAAGDKDAVKDLIENGADVNGR
DSDGRTPLHHAAENGNAEVVALLIAKGADVNAKDSDGRTPLHHAAENGHEEVVLILLL
KGADVNAKDSDGRTPLHHAAENGHKRVVLVLILAGADVNTSDSDGRTPLDLAREHGN
EEVVKALEKQ

>D2-21.8

MHHHHHHGSGSEKARIAVENLEAALRLNRAAAEMQKSAIKIMRDNSSDEKAFRYLLLT
TKVLKMSVELLRASLELAEKALREEGSDDSAEKVRKEAEEILKESTEILKRAELETLKAA
VRVAAEAAARNATDEEERKRIEEELKKAEERANRSTNEEEIKKILEEALARFLIILAEKGA
KEAVKLALEAGADVNGKDSDGKTPLHLAAENGHAKVVLLLLEQGADPNAKDSDGKTP
LHLAAENGHAVVVALLLMHGADPNAKDSDGKTPLHLAAENGHEEVVILLLAMGADPN
TSDSDGRTPLDLAREHGNEEVVKVLEDHG

>D2-21.22

MHHHHHHGSGSEKARIAVENLEAALRLNRAAAEMQLSAAKIAADNFSDKKAAEYTRL
TTKVLEMSVELLRASLELAEKALREEGSDDSAEKVRKEAEEILKESKEILEAAEALTRIA
HLARKAAESTDPEEALKIAKEAIEIALKTVKENPSELALQAVLAAVILASAVAKRVTDPD
KALKIAKLVIELALEAVKEDPSTDALRAVLEAVRLAEEVARRVTDPIKALKIAALVIQLA
AEAAKEDPSEEAQRALKLAAELAAEALERGADVNYHDEDGRTPLHHAAEAGADEAVLI
LLLKGADVNAKDSDGRTPLHHAAENGNKRVVLVLILAGADVNTSDSDGRTPLDHAREH
GNEEVVKALEKQ

>D2-21.26

MHHHHHHGSGSEKARIAVENLEAALRLNRAAAEMQASAAKIVADNTSDEKAYRYLEL
TTKVLLMSVELLRASLELAEKALREEGSDDSAEKVRKEAEEILKESTLALEAAEELTKAA
KAALRAREASERGDEEEFRKAAEEALEAAKRVVERAKKAGIPELVAAAAAVALAIAEL
AAKNGDKEVFKKAAESALEVAKRLVEVASKEGDPELVLEAAKVALRVAELARKNGDK
EVLKKAALSAAEVALRLAEVAKKEGDPDLVREAAEILADALEKGLDVNIHDEDGRTPL
HHAAELGADEAVLILLLAGADVNAKDSDGRTPLHHAAENGHKRVVLVLILAGADVNTS
DSDGRTPLDHAREHGNEEVVKALEKQ

>D2-21.29

MHHHHHHGSGSEKARIAVENLEAALRLNRAAAEMQKSAIKIADDNRSDEKALRYALLT
TKVLEMSVELLRASLELAEKALREEGSDDSAEKVRKEAEEILEKSSRILAEAFVITARLAT
ELARLLQEKAKKTGDAKELREAKRALKEAAEYVEKALKINKDDDEARELLERIEEELKK
VEKLLEEILIKAAARGDKDLVKLALKAGADVNASDSDGKTPLHKAAENGHAKVVLLLL
EQGADPNAKDSDGKTPLHLAAENGHAVVVALLLMHGADPNAKDSDGKTPLHLAAENG
HEEVVILLLAMGADPNTSDSDGRTPLDLAREHGNEEVVKVLEDHG

>D2-21.30

MHHHHHHGSGSEKARIAVENLEAALRLNRAAAEMQKSAIKIAEDNSSDEKAIRYTLLTT
RVLEMSFELLRASLELAEKALREEGSDDSAEKVRKEAEEILEESLRILAEAFVRTARFLKE
LAERLQERAKKTGDPELLAEAYEALREAVEFVKKAEKINPDNERAKKTLEELKEELRKV
EELLKELLIRAAERGDKDTVRRALEAGADVNAKDSDGKTPLHLAAENGHAKVVLLLLE
QGADPNAKDSDGKTPLHLAAENGHAVVVALLLMHGADPNAKDSDGKTPLHLAAENG
HEEVVILLLAMGADPNTSDSDGRTPLDLAREHGNEEVVKVLEDHG

Round-2 Design Sequences (D3 symmetry):

>D3-19.14

MHHHHHHGSSEKARIAVENLEAALRLNRAAAEMQKSAIKIARDNRSDKKALLYLLLAT
YVLEMSLELLRASLELAEKALREEGSDDSAEKVRKEAEEILKESKEIFLRAALETAKAAA
EYVEEAAREAERRGNPELRDAAKALRKYLEEANEEAAKQGNAEKILRVALAALLIAAA
ALGDKDLVKDLIEMGADVNGHDLDGRTPLHLAAAAGHAEVVALLIEKGADVNAKDSD
GRTPLHHAAENGHDEVVLILLLKGADVNAKDSDGRTPLHHAAENGHKRVVLVLILAGA
DVNTSDSDGRTPLDLARENGNEEVVKVLEKA

>D3-19.19

MHHHHHHGSSEKARIAVENLEAALRLNRAAAEMQKSAIKIVDDNSSDVRAIEYLALTSA
VLAESLLLLLASLELAEKALREEGSDDSAEKVRKEAEEILEESARIAAEAAEESLRAAEE
AIELARKTGDSDALRAAAEALKAARAAVRAAIAANPDDDKAEEIAKRLEEALNRVLHE
AAERGDQDAVKLVIEAGGDVNARDSDGRTPLHHAAENGHAEVVALLIRKGADVNAKD
SDGRTPLHHAAENGHDEVVLILLLKGADVNAKDSDGRTPLHHAENGHKRVVLVLILA
GADVNTSDSDGRTPLDHARENGNEKVVKALQEQ

>D3-19.20

MHHHHHHGSSEKARIAVENLEAALRLNRAAAEMQKSAIKIALDNSSDEKAIRYARLTTK
VLKMSVELLRASLELAEKALREEGSDDSAEKVRKEAEEILKESTLILEAADLATALLDLL
QKVRKVEKEIKSNKDNEEAVETAARLAIELARVAKRLEELAKKLGDGFLKKLAEKAIKI
AARALEVALEAGYDVNAKDSDGRTVLHHAAENGALEVVLLALLNGADVNAKDSDGR
TPLHHAAENGNKRVVLVLILAGADVNTSDSDGRTPLDLARENGNEEVVKALERR

>D3-19.24

MHHHHHHGSSEKARIAVENLEAALRLNRAAAEMQFLAIKIMLLNSSDEKAARFLRLTTK
VLKMSVELLRASLELAEKALREEGSDDSAEKVRKEAEEILKESTEILEATEEATKLLELLE
EARKVEEAIKSNPDNDEAVETAKRIAEEARKVALKLFEYASKLGIPLLAKAAAEALAVA
LKAGADPNAKDSDGKTPLHHAAEAGVKLAVMLLLSHGADPNAKDSDGKTPLHLAAEN
GHEDVVLLLLLMGADPNTSDSDGRTPLDLAREHGNEDVVKALKAAG

>D3-19.24-C

MSEKARIAVENLEAALRLNRAAAEMQFLAIKIMLLNSSDEKAARFLRLTTKVLKMSVEL
LRASLELAEKALREEGSDDSAEKVRKEAEEILKESTEILEATEEATKLLELLEEARKVEEA
IKSNPDNDEAVETAKRIAEEARKVALKLFEYASKLGIPLLAKAAAEALAVALKAGADPN
AKDSDGKTPLHHAAEAGVKLAVMLLLSHGADPNAKDSDGKTPLHLAAENGHEDVVLL
LLLMGADPNTSDSDGRTPLDLAREHGNEDVVKALKAAGSWLEHHHHHH

Table 2.2 – Listing of building-block validation levels

Design ID	N-terminal oligomer	Spacer	C-terminal oligomer
D2-1.1B	rop4	DHR62	ank3C21
D2-1.1D	rop20	DHR62	ank1C2G3
D2-1.4H	rop20	DHR68	ank3C22
D3-1.5C	rop20	DHR15	ank1C2G3
D3-1.5A2 (marginal)	rop20	DHR15	ank1C2G3
D2-21.8	rop20	DHR15	ank3C21
D2-21.22 (marginal)	rop20	DHR57	ank1C2G3
D2-21.29 and D2-21.30	rop20	DHR82	ank3C21
D2-21.26	rop20	DHR71	ank1C2G3
D3-19.19	rop20	DHR82	ank1C2G3
D3-19.14	rop20	DHR76	ank1C2G3
D3-19.20	rop20	DHR82	ank1C2G3
D3-19.24	rop20	DHR82	ank3C22

Constituent building blocks for successful and marginal designs are shown. Those building blocks that have been SAXS-verified, but not crystallized, are colored green; most successful designs are composed with two of three components having only SAXS verification. Two successful designs were created without any building block crystal verification whatsoever.

Table 2.3 - GFP-binding scaffold sequences

>D2-1.1Dv1: Aggregated

SEKARIAVENLEAALRLNKAAIEMAKSAIKITRDNSSDEKATRYSLLTAKVLVMSLELLT
ASLELAEKALREEGSDDSAEKVRKEAEEILSKAVEEAVRVMQEMVTIMKRTGSNDSLRE
VAELALRVAKAAEKAGNVEVAVQAARVAVEAAKQAGDNDVLRKVAEQALRIAKEAE
KQGNVEVAVKAAKVAVEAAKQAGDEDVLKKVAEQASRIASEASKQGNKEVASKALIV
AAQAGSKEAVKKAIESGADVNASDSDGRTPLHHAAENGHAEVVALLIEKGADVNAKD
SNGHTPLLHAARNGHDEVVLILLLKGADVNAKDDVGVTPLHLAAQRGHKRVVLVLILA
GADVNTADLWGQTPLHLAATAGHLEVVKALLKQGADVNARDNIGHTPLHLAAWAGH
LEIVEVLLKYGADVNAQDKFGKTPFDLAIDNGNEDIAEVLQKA

>D2-1.4Hv1: 4.3 Å resolution

SEKARIAVENLEAALRLNRAAAEMQKSAIKIMDDNSDDEKALRYLRLTTKVLRMSVEL
LRASLELAEKALREEGSDDSAEKVRKEAEEILKESTAILKLADAATKVADIKHDIKKAKE
QQEQGNKEEAEKTLREATEKIKRVTEELEKIAKNSKTPEIALKAAEALVKLIKLLIEIAKL
LQEQGNKEEAEKVLREATELIKRVTELLEKIAKNSDTPELALRAAELLVRLIKLLIEIAKL
LQEQGNKEEAEKVLREATELIKRVTELLEKIAKNSDTPELASRAAELLVRLIKLLQEIAKL
LKEQGNKEEAEKVEREAKELLSRVLILAARIGNKDIVKTALENGADVNASDDVGVTPLH
LAAQRGHKDVVELLLRQGADPNAKDLWGQTPLHLAATAGHKVVVMLLLSQGADPNA
KDNIGHTPLHLAAWAGHEDVVLLLLLMGADPNTSDKFGKTPFDLAIDNGNEDVVEALK
AAGG

>D2-1.5Cv1: Disordered

SEKARIAVENLEAALRLNRAAAEMQKSAAKIVADNASDEKALRYLRLTTKVLRMSVELLRASLELAEKALREEGSDDSAEKVRKEAEEILKESTEILKEADQITEVADLAFELANKATDEELRKEISKCARLALELASRSTNDELIKQILEVAKLAFELASKATDEELIKLILKCCQAAFERASRSTNDEEIKKILEVAKRAFETASKATDEEEIKSILLICAAALGNKDAVKSAIENGADVNASDSDGRTPLHHAAENGNAEVVALLIEKGADVNAKDSDGHTPLLHAARNGHDEVVLILLLKGADVNAKDDVGVTPLHLAAQRGHKRVVLVLILAGADVNTADLWGQTPLHLAATAGHEEVVKALIKQGADVNARDNIGHTPLHLAAWAGHLEIVEVLLKYGADVNAQDKFGKTPFDLAIDNGNEDIAEVLQKA

>D2-21.8v1: Slight aggregation

MHHHHHHGSGSEKARIAVENLEAALRLNRAAAEMQKSAIKIMRDNSSDEKAFRYLLLTTKVLKMSVELLRASLELAEKALREEGSDDSAEKVRKEAEEILKESTEILKRAELETLKAAVRVAAEAAARNATDEEERKRIEEELKKAEERANRSTNEEEQKKILEEALGRFLIILARKGAKEAVKLALEAGADVNAADDVGVTPLHLAAQRGHAKVVLLLLEYGADPNAADLWGQTPLHLAATAGHAVVVALLLMHGADPNARDNIGHTPLHLAAWAGHEEVVILLLAMGADPNAQDKFGKTPLDLARDNGNEEVVKVLEDHAA

>D2-21.8v2: 5 Å overall resolution and minor preferred orientation

MHHHHHHGSGSEKARIAVENLEAALRLNRAAAEMQKSAIKIMRDNSSDEKAFRYLLLTTKVLKMSVELLRASLELAEKALREEGSDDSAEKVRKEAEEILKESTEILKRAELETLKAAVRVAAEAAARNATDEEERKRIEEELKKAEERANRSTNEEEQKKILEEALGRFLIILARKGAKEAVKLALEAGADVNAADDVGVTPLHLAAQRGHAKIVLLLLEYGADPNAADLWGQTPLHLAATAGHAVIVALLLMHGADPNARDNIGHTPLHLAAWAGHEEIVILLLAMGADPNAQDKFGKTPLDLARDNGNEEVVKVLEDHAA

>D2-21.8v3: Preferred orientation

MHHHHHHGSGSEKARIAVENLEAALRLNRAAAEMQKSAIKIMRDNSSDEKAFRYLLLT
TKVLKMSVELLRASLELAEKALREEGSDDSAEKVRKEAEEILKESTEILKRAELETLKAA
VRVAAEAAARNATDEEERKRIEEELKKAEERANRSTNEEEQKKILEEALGRFLIILARKG
AKEAVKLALEAGADVNAADDVGVTPLHLAAQRGHAKVVLLLLEQGADPNAADLWGQ
TPLHLAATAGHAVVVALLLMHGADPNARDNIGHTPLHLAAWAGHEEVVILLLAMGAD
PNAQDKFGKTPLDLARDNGNEEVVKVLEDHAA

>D2-21.29v1: Severe aggregation

MHHHHHHGSGSEKARIAVENLEAALRLNRAAAEMQKSAIKIADDNRSDEKALRYALLT
TKVLEMSVELLRASLELAEKALREEGSDDSAEKVRKEAEEILEKSSRILAEAFVITARLAT
ELARLLQEKAKKTGDAKELREAKRALKEAAEYVEKALKINKDDDEARELLERIEEELKK
VEDLLGKILLEAARAGDKDLVKLALKAGADVNAADDVGVTPLHLAAQRGHAKVVLLL
LEYGADPNAADLWGQTPLHLAATAGHAVVVALLLMHGADPNARDNIGHTPLHLAAW
AGHEEVVILLLAMGADPNAQDKFGKTPLDLARDNGNEEVVKVLEDHAA

>D2-21.29v2: Preferred orientation and aggregated

>MHHHHHHGSGSEKARIAVENLEAALRLNRAAAEMQKSAIKIADDNRSDEKALRYALL
TTKVLEMSVELLRASLELAEKALREEGSDDSAEKVRKEAEEILEKSSRILAEAFVITARLA
TELARLLQEKAKKTGDAKELREAKRALKEAAEYVEKALKINKDDDEARELLERIEEELK
KVEKLLGEILLEAARAGDKDLVKLALKAGADVNAADDVGVTPLHLAAQRGHAKIVLL
LLEYGADPNAADLWGQTPLHLAATAGHAVIVALLLMHGADPNARDNIGHTPLHLAAW
AGHEEIVILLLAMGADPNAQDKFGKTPLDLARDNGNEEIVKVLEDHAA

>D2-19.20v1: Aggregated

MHHHHHHGSGSEKARIAVENLEAALRLNRAAAEMQKSAIKIALDNSSDEKAIRYARLTT
KVLKMSVELLRASLELAEKALREEGSDDSAEKVRKEAEEILKESTLILEAADLATALLDL
LQKVRKVEKEIKSNKDNEEAVETAARLAIELARVAKRLEELAKKLGDGFLKKLAEKAIK
IAARALEVALEAGYDVNAKDSDGATVLHHAARNGALEVVLLALLNGADVNAADDVG
VTPLHLAAQRGNKRVVLVLILAGADVNAADLWGQTPLHLAATAGHLEVVKALLKRGA
DVNARDNIGHTPLHLAAWAGHLEIVEVLLKYGADVNAQDKFGKTPFDLAIDNGNEDIA
EVLQKA

The GFP-binding variants tested are listed, along with the corresponding (scaffold + GFP) resolution achieved - or the observed failure mode, where applicable. Design identifiers are comprised of the underlying scaffold identifier plus the suffix v<number>, to separate variants of the same scaffold.

Table 2.4 – Listing of building blocks validated by x-ray crystallography

DHR spacers: 5CWB (DHR4), 5CWD (DHR7), 5CWF (DHR8), 5CWG (DHR10), 5CWH (DHR14), 5CWI (DHR18), 5CWJ (DHR49), 5CWK (DHR53), 5CWL (DHR54), 5CWM (DHR64), 5CWN (DHR71), 5CWO (DHR76), 5CWP (DHR79) and 5CWQ (DHR81).
C2 homo-dimers: 5KBA (Ank1C2), 5HRY (Ank3C2_1), 5J73 (2L4HC2_9), 5J0K (2L4HC2_23), 5J10 (2L4HC2_24)

The building blocks used in this study that have been solved by X-ray crystallography are listed. Although solved by crystallography, DHR5 was not included in the set because homo-oligomerization was detected in the original study.

Rosetta files

File 2.1 - RosettaScripts design script and command-line

```
<ROSETTASCRIPTS>
        <SCOREFXNS>
                <ScoreFunction name="sfx_hard_symm"
weights="beta.wts" symmetric="1"  >
                </ScoreFunction>
        </SCOREFXNS>
        <TASKOPERATIONS>
                <InitializeFromCommandline name="init" />
                <RestrictIdentities name="nomutate_VIRTUAL"
identities="XXX" prevent_repacking="1" />
                <DisallowIfNonnative name="disallow_nonnative"
disallow_aas="CPM" />
                <ReadResfile name="resfile_designable"
filename="%%resfile%%" />
        </TASKOPERATIONS>
        <MOVERS>
        <SetupForSymmetry name="symmetry_setup"
definition="%%symdef%%"></SetupForSymmetry>
                <SymPackRotamersMover
name="design_rotamers_resfile" scorefxn="sfx_hard_symm"
task_operations="init,nomutate_VIRTUAL,resfile_designable,disall
ow_nonnative"></SymPackRotamersMover>
        </MOVERS>
        <PROTOCOLS>
        <Add mover_name="symmetry_setup" />
                <Add mover_name="design_rotamers_resfile" />
        </PROTOCOLS>
</ROSETTASCRIPTS>
```

```
bash$: <rosetta_scripts_path> -ignore_zero_occupancy false -
database <rosetta_database_path> -linmem_ig 10 -lazy_ig true -
parser:protocol <rosettascripts_xml_path> -s <pdb_path> -native
<pdb_path> -nstruct 1 -parser:script_vars
symdef=<D2_or_D3_symmetry_definition> resfile=<resfile> -ex1 -
ex2 -unmute all -out:pdb_gz true -out:path:all ./ -beta -
overwrite -scorefile <scorefile_name>.sc
```

A simple design script and command-line example applies symmetry and designs sidechains with whatever score-function is in beta at the time of use (beta_nov16 during this work). Symmetry definition files are provided in Text Files 2.2 and 2.3.

File 2.2 – Symmetry definition file (D_2)

```
symmetry_name d2
subunits 4
number_of_interfaces  3
E = 4*VRT0001 + 2*(VRT0001:VRT0002) + 2*(VRT0001:VRT0003) +
2*(VRT0001:VRT0004)
anchor_residue COM
virtual_transforms_start
start -1,0,0 0,1,0 0,0,0
rot Rz_angle 180.0
rot Rx_angle 180.0
rot Rz_angle 180.0
virtual_transforms_stop
connect_virtual JUMP1 VRT0001 VRT0002
connect_virtual JUMP2 VRT0002 VRT0003
connect_virtual JUMP3 VRT0003 VRT0004
set_dof BASEJUMP x(50) angle_x(0:360) angle_y(0:360)
angle_z(0:360)
set_dof JUMP2 z(50) angle_z(0:90.0)
```

File 2.3 – Symmetry definition file (D_3)

```
symmetry_name d3
subunits 6
number_of_interfaces  4
E = 6*VRT0001 + 6*(VRT0001:VRT0002) + 3*(VRT0001:VRT0004) +
3*(VRT0001:VRT0005) + 3*(VRT0001:VRT0006)
anchor_residue COM
virtual_transforms_start
start -1,0,0 0,1,0 0,0,0
rot Rz_angle 120.0
rot Rz_angle 120.0
rot Rx_angle 180.0
rot Rz_angle 120.0
rot Rz_angle 120.0
virtual_transforms_stop
connect_virtual JUMP1 VRT0001 VRT0002
connect_virtual JUMP2 VRT0002 VRT0003
connect_virtual JUMP3 VRT0003 VRT0004
connect_virtual JUMP4 VRT0004 VRT0005
```

```
connect_virtual JUMP5 VRT0005 VRT0006
set_dof BASEJUMP x(50) angle_x(0:360) angle_y(0:360)
angle_z(0:360)
set_dof JUMP3 z(50) angle_z(0:60.0)
```

Section 3 – Antibody nanoparticle proof-of-concept

Abstract

Antibodies are an important class of therapeutics and also lie at the heart of many diagnostic tests. Methods have matured for the discovery and testing of recombinant monoclonal antibodies and now more exotic antibody formats are being explored, such as bispecific antibodies and antibody-drug conjugates. We sought to design a new format in which antibodies co-assemble with a designed oligomer to form protein nanoparticles with increased antibody valency and precisely defined structure. Antibody nanoparticles would offer different capabilities from those traditional antibodies, bispecifics, or antibody-drug conjugates; they might allow for targeted delivery of encapsulated cargo, receptor clustering *in vivo* or otherwise, or avidity-based binding improvements. Five designs were created by symmetric fusion of a homo-tetramer to the Fc-binder *Staphylococcal* protein A, in an orientation predicted to facilitate co-assembly of six designed tetramers and twelve antibodies into a nanoparticle with octahedral symmetry. One of the designs proved capable of co-assembly with antibodies into particles resembling those designed. This work demonstrated the feasibility the antibody nanocage concept.

Background on antibody therapeutics

In 2017, five of the top ten drugs by sales were monoclonal antibodies and an additional two were antibody fusions consisting of a cell surface receptor fused to Fc[47]. FDA-approved antibody therapeutics largely fall within three categories: monoclonal antibodies, bispecific antibodies, and antibody-drug conjugates, with monoclonal antibodies being the most common. Monoclonal antibodies can kill cells that display specific surface markers, as in the case of rituximab, which targets the CD20 receptor and is prescribed for a variety of B-cell lymphomas.

Binding kills targeted cells by inducing apoptosis, complement-dependent cytotoxicity, or antibody-dependent cellular cytotoxicity. mAbs can also block oligomerization of receptors that require oligomerization for downstream signaling, such as HER2 and HER. Alternatively, they can act to remove a soluble target from circulation, as in the case of adalimumab and infliximab – these antibodies reduce inflammation by binding the cytokine TNF-α, thereby preventing downstream signaling.

Bispecific antibodies (bsAbs) are a newer engineered antibody format in which the two Fabs differ, allowing a single antibody to target two distinct antigens. In early patient studies, the most prevalent mechanism of action was to target an overexpressed cell surface marker (e.g. HER2, EGFR) with one Fab and to then recruit and activate T cells with a second anti-CD3 Fab[48] for cancer treatment. Unfortunately, therapeutic antibodies that target cell-surface markers cells can have serious off-target effects, since usually the targeted antigen does not uniquely identify the class of desired target cells. This shortcoming provides a second motivation for the development of bsAbs: preferential targeting of cells that express both antigens over those that express only one. Still, targeting with "AND logic" remains an open problem as non-avid binding still occurs to cells that express either marker alone; for now, bsAbs are only useful for indications where a degree of targeting by "OR logic" is acceptable or where multiple targets exist, but no combinatorial logic is required. Antibody based targeting modalities that are strictly governed by "AND logic" have yet to be developed.

Antibody drug conjugates (ADCs) are another class of engineered antibodies that have fallen short of early expectations. At present, only four ADCs are approved by the FDA[49,50,51,52], despite decades of research. ADCs are produced through conjugation of cytotoxic drugs to antibodies. The toxic cargo is released upon endocytosis, through the use of linkers that are sensitive to cues from the intracellular environment. This helps to direct toxicity at the targeted cells. For lysine-conjugated drugs, the average drug to antibody ratio (DAR) is typically 3 to 4, but conjugation products are heterogeneous due to the pool of approximately 30 lysines to which conjugation stochastically occurs[53]. Conjugation to all or most of the fewer cysteines results in less heterogeneity, so most ADCs under development at this time use maleimide linkers. A balance must be struck between drug loading on the one hand and stability and binding on the other; higher drug loading is desirable, but only to the extent that it can be achieved without interfering with binding or inducing aggregation.

Antibody nanocage design concept and motivation

We aimed to create a new antibody-based nanoparticle format whose characteristics might be advantageous in certain contexts. The envisioned protein nanoparticles would stably incorporate antibodies in a binding-competent orientation and be formed through co-assembly of mAbs, bsAbs, or ADCs with a computationally designed protein oligomer (Figure 3.1). Two-component systems have previously been designed[9,10], but in this case, one component is an antibody. All platonic solids (tetrahedra, octahedra, icosahedra) and dihedral symmetries are theoretically possible, because they all incorporate C_2 symmetric axes for placement of an antibody. These assemblies could allow for ADC-like drug targeting, but with higher-drug loading through the encapsulation of hydrophobic drugs by the nanocage container. Nanoparticle formation with a mixture of antibodies, mixed in desired ratios prior to assembly, could confer bsAb-like multi-target affinity. Activation of T-cells is one possibility, as is co-localization of molecules, surface receptors, or even cells. Finally, the high avidity could be leveraged to make the overall nanoparticle binding highly sensitive to the target molecule concentration, by engineering weakened interactions into the antibodies.

Apart from functionality that derives from the incorporation of antibodies, these nanocages should retain the advantages that previously designed protein nanocages hold over traditional nanoparticle systems. First, computationally designed protein nanocages have the potential to be highly uniform, as demonstrated by a subset of existing cubic viruses, naturally occurring bacterial microcompartments[54], and previously designed nanocages. In contrast, non-protein-based nanoparticle production methods result in assembly products with a distribution of sizes. Second, these assemblies can be endowed with behaviors. Nature – viruses especially – provides a starting guide for nanoparticle functionalization possibilities: cargo encapsulation, cell-type specific delivery, endosomal escape, and nuclear entry are a few examples.

A variety of approaches exist to build such a system. To produce a flexible platform that allows facile substitution of any human IgG, no engineering or interface design was deemed allowable on the antibody component. Furthermore, the previously described fusion-based strategy using Staphylococcal "protein A" (SpA) was selected (Figure 3.2), because it would allow for reuse of the well-studied and high affinity interface between SpA and antibody Fc. The binding mode for SpA is known. Although no crystal structure exists of Fc in complex with the natural three-helix bundle version of SpA, a crystal structure does exist of Fc in complex with a redesigned minimal fragment of SpA (PDB code 1L6X)[55] and is termed the "Z domain". The Z domain comprises two helices stapled with a disulfide bond. To produce a design model of protein A in complex with Fc that could be used as a starting point for computational design, the structure of protein A was taken from the crystal structure (Graille, 2000) (PDB code 1DEE) and aligned to the Z domain in complex with Fc.

Design of an octahedral antibody nanocages by two-component fusion

I sought to fuse SpA directly to previously designed de-novo C_4 oligomers, to create octahedra with O42 symmetry. O42 symmetry was thought to be the most promising complex symmetry, because it includes C_4 building blocks, which cannot misassemble to form either icosahedra or tetrahedra - related but off-target symmetries without a C_4 axis. Thus, the use of C_4 oligomers is a form of negative selection against unwanted states. C_5 oligomers have a similar property – only icosahedra have five-fold symmetry among the platonic solids. However, at the time of this work, we had no de-novo pentamers with N-termini available for fusion (any fusion of SpA must occur at its C-terminus to avoid removal or disruption of the Fc-binding region, which exists at SpA's N-terminus). C_3 building blocks were also avoided, because all cage symmetries contain C_3 symmetry, so selecting against unwanted states might be more difficult.

Among two-component fusion options for the formation of octahedra, only one C_4 building block and single alignment option was found even roughly consistent with the intended final geometry. Even in this case, the fusion of protein A onto the C_4 tetramer was nearly 15° from that required to properly place the C_2 antibody and C_4 tetramer relative to one another. Although the protocol repositions components to idealize the geometry, the new position must be stabilized through sequence design of the fused region. For reference, the successful D2 ring designs had angular errors of less than 5°.

A total of five fusion proteins were ordered, all based on the same C_4 tetramer and alignment between protein A and tetramer, but with different mutations produced by the design protocol. The five sequences explored options for hydrophobic packing of the core, as well as charged and polar residues at solvent-exposed positions. The native SpA sequence begins with FNK, which was preserved in all designed proteins, despite its not being resolved in the 1DEE crystal structure. The rationale for its conservation was that a highly similar FNM sequence is resolved in the engineered Z domain co-crystal with Fc.

Four out of five designs (designated 7A through 7E) were partially soluble, according to denaturing gels of the soluble and insoluble fractions post lysis. Designs 7A, 7C, and 7D tend to gradually precipitate, while 7B is fairly stable in solution. The Fc binding interface of SpA contains a highly exposed tyrosine and phenylalanine, so poor solubility was an expected failure mode. The four proteins were further purified by SEC, where the elution profiles for all designs showed a single peak consistent with the expected size and in the same location.

Next, the tendency of these proteins to form a complex with IgG1 was evaluated with dynamic light scattering (Figure 3.3). An anti-CMV in-house stock of IgG1 was mixed with each of the four designs in the necessary stoichiometric ratio (as estimated by 280nm wavelength absorption). Of these, design 7B gave the clearest indication of binding and assembly formation near the expected size.

From the DLS experiment it was clear that design 7B binds antibody, but whether in the desired manner was unclear. To help probe this, Fc rather than full antibody was used for subsequent experiments. The Fc reagent was purchased from the Scripps Research Institute, where IgG had been purified from human serum and cleaved with papain to remove Fabs. Fc and 7B were mixed at micromolar concentration and purified by SEC (Figure 3.4). The assembled products eluted across a broad range of times, suggesting heterogeneous assembly. Uneven SEC profiles remained across a variety of buffers and assembly incubation times.

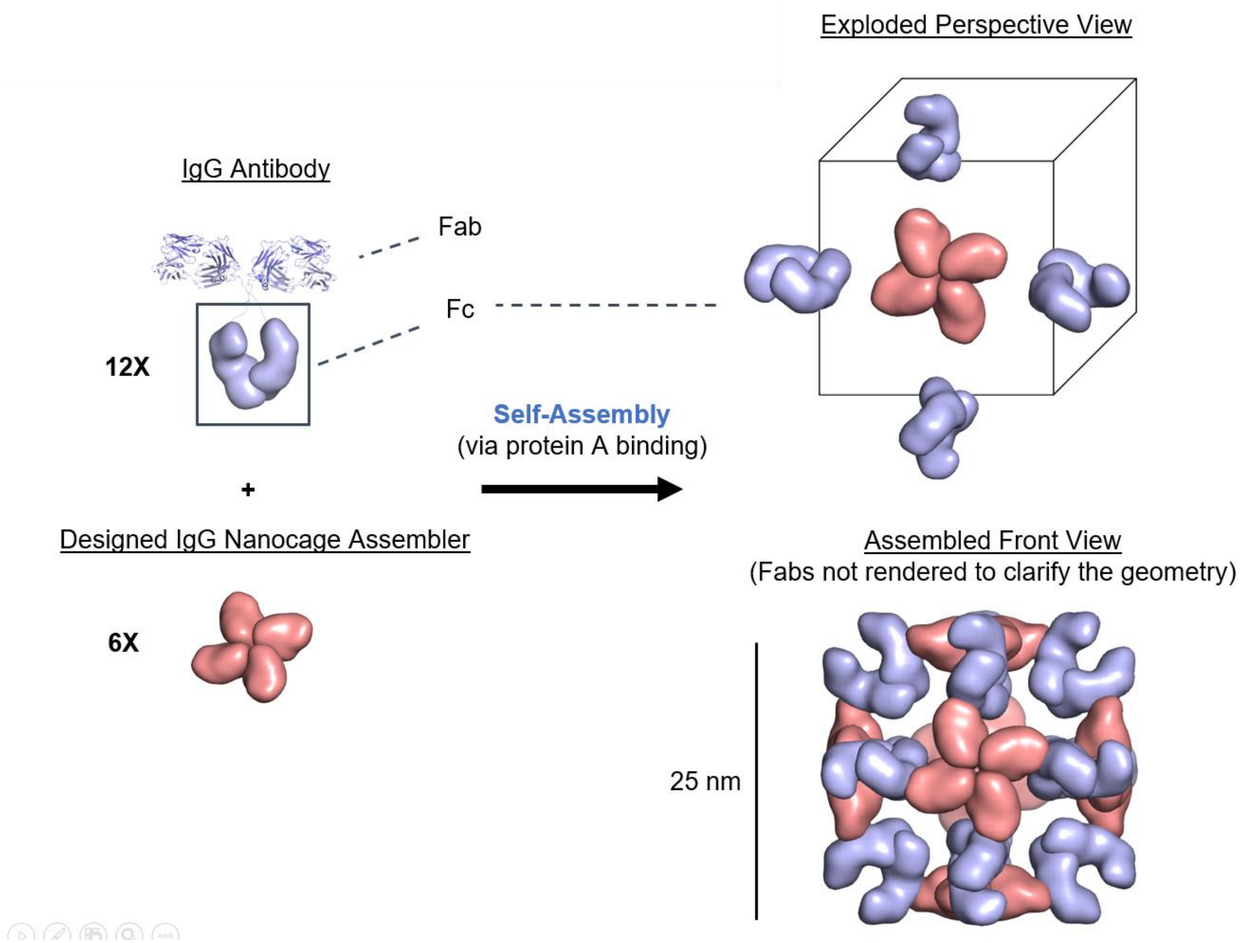
Negative stain electron microscopy was performed on the larger assembly products, after separation by size with SEC. Most regions of the carbon grid contained aggregates, but a few promising cage-shaped particles were observed (Figure 3.5), as well as what appear to be misassembled aggregates or flocculated particles. The particles are not totally uniform in size, indicating some misassembly. The extent of misassembly is unclear, because the projection varies with orientation. A second EM session was performed to image assemblies containing produced with either Fc or antibody. A greater number of nanoparticles was observed, but at higher resolution, it is clear that the nanoparticles exhibited some heterogeneity (Figure 3.6), which is consistent with the broad SEC elution profiles.

Discussion

Electron micrographs of designed construct 7B in complex with IgG show that the design concept has potential. In an attempt to reduce flexibility and thereby create robust scaffolds, I limited the approach to two-component fusion in lieu of the three-component fusion method described earlier. However, this limits the geometric diversity and the closest match that could be achieved with only two components was still off by an angle of 15°. I expect this geometric non-ideality contributed a major portion of the observed misassembly, but a second cause may be flexibility around the Fc binding site; the interface area of SpA to Fc is low, so the binding interface may act as a semi-flexible joint or pivot point between Fc and the designed assembler. Although follow-up work in our lab generated constructs using the three-component fusion approach that I describe in Section 1, I expect that improved homogeneity and particle robustness would result from the design of assemblers whose monomeric subunits are shorter and thus less flexible. Such constructs could be created via more traditional protein design methods, such as fragment assembly and interface design.

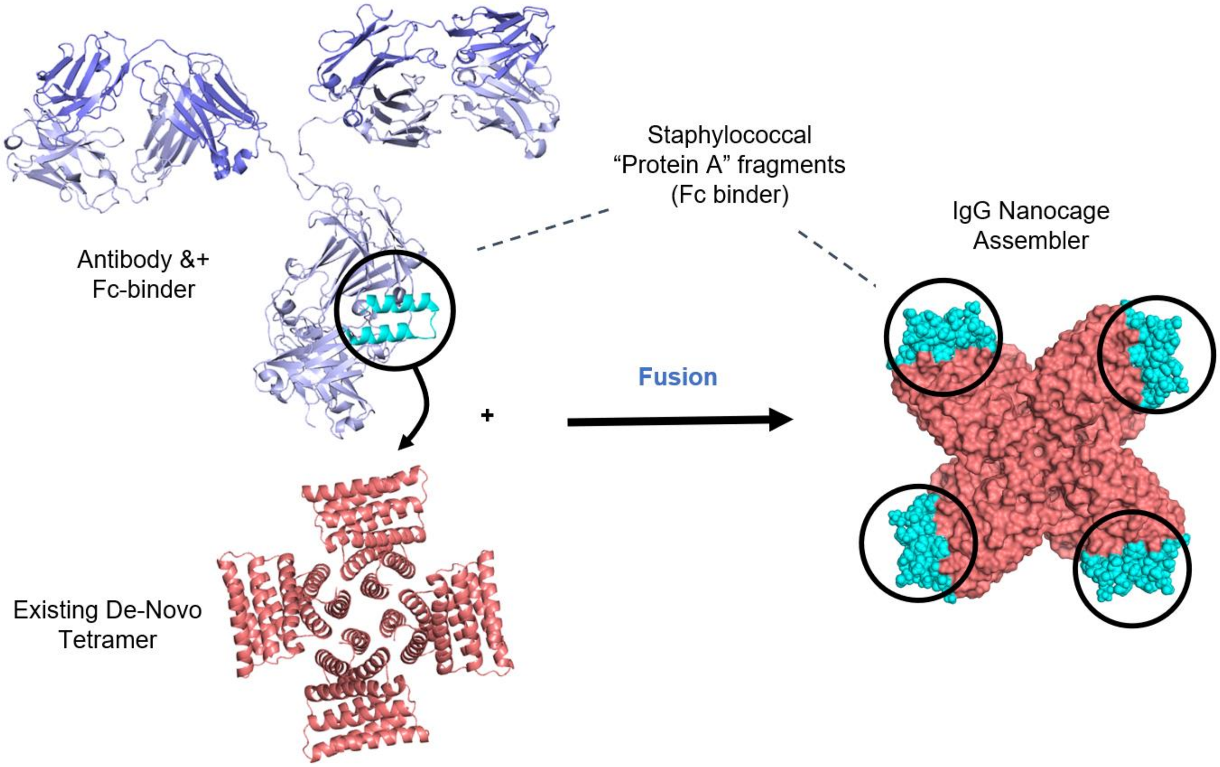
Figures

Figure 3.1



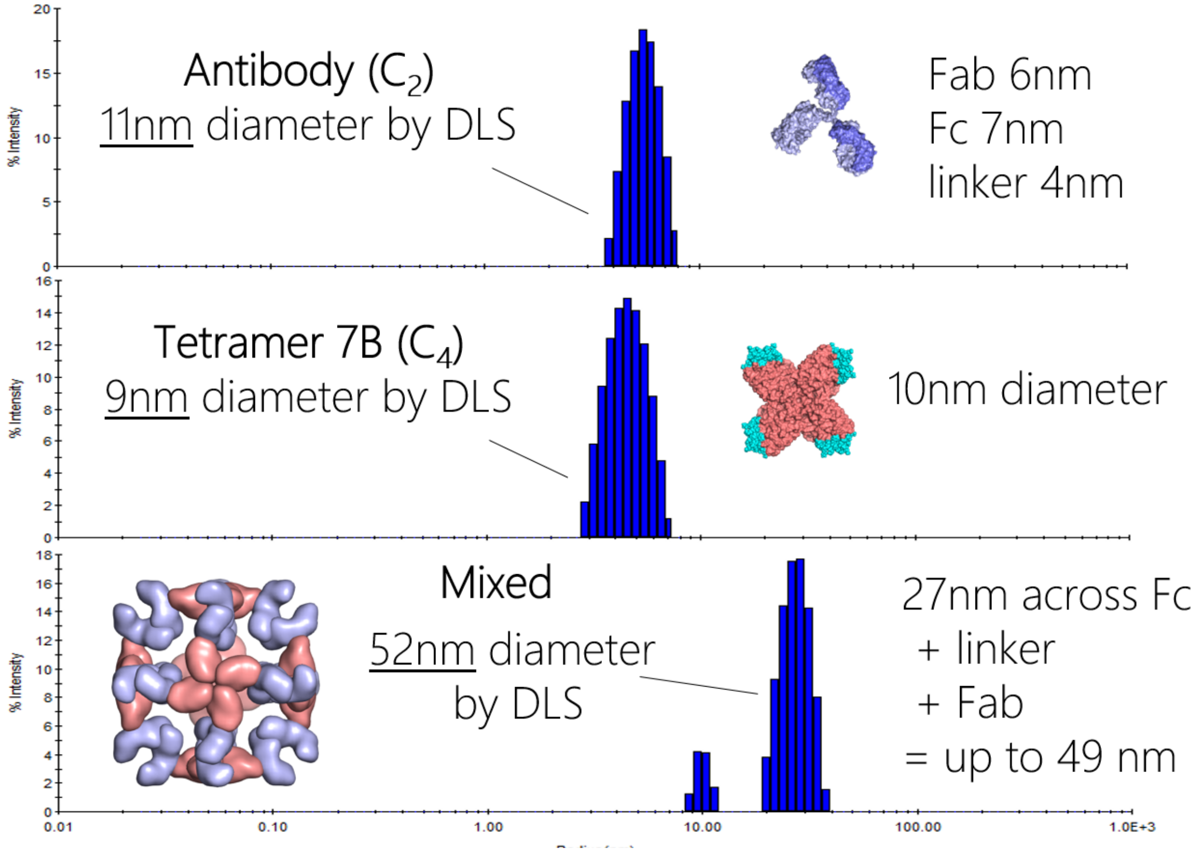
The design concept is illustrated. It consists of co-assembly between a designed oligomeric protein “assembler” (tetrameric in this example) with an antibody to form a nanoparticle with multiple copies of each. As with protein nanoparticle design in general, all point, plane, or crystal symmetries with C_2 axes are theoretically possible.

Figure 3.2



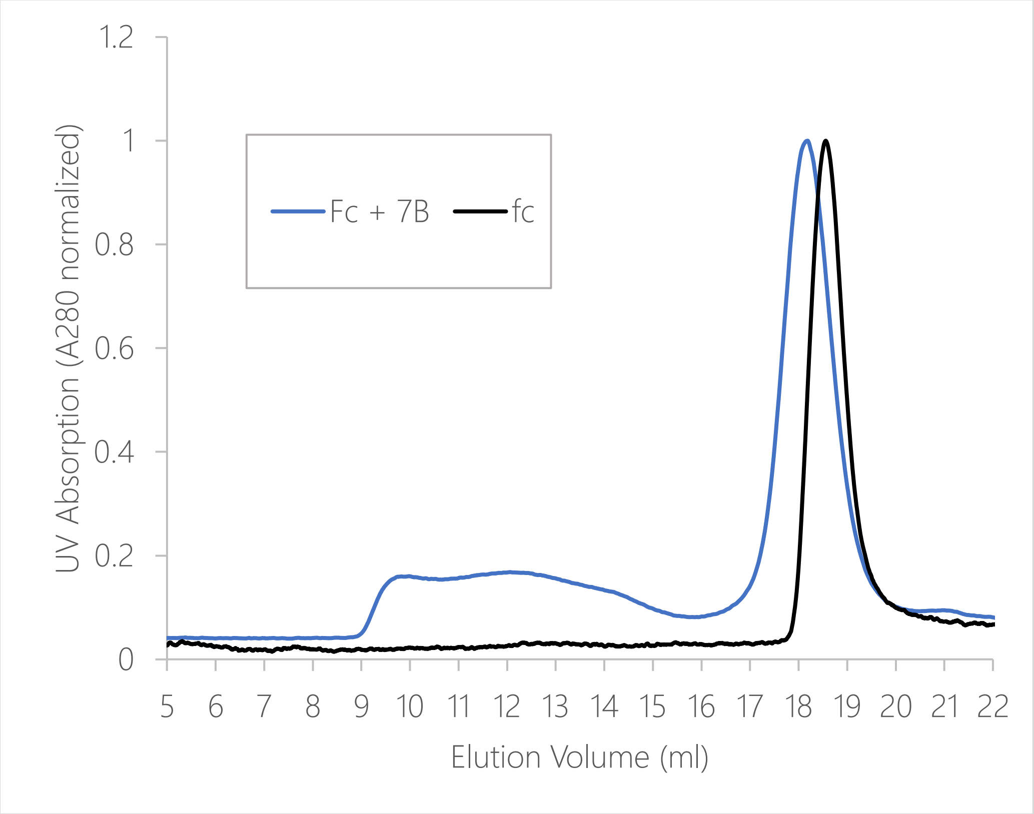
The protein sources for the designed tetrameric assembler are shown; the Fc-binding Staphylococcal protein A (SpA) is rigidly fused to an existing de-novo tetramer to form a new design. The fusion region is redesigned to maintain rigidity, without altering the binding interface between SpA and the antibody.

Figure 3.3



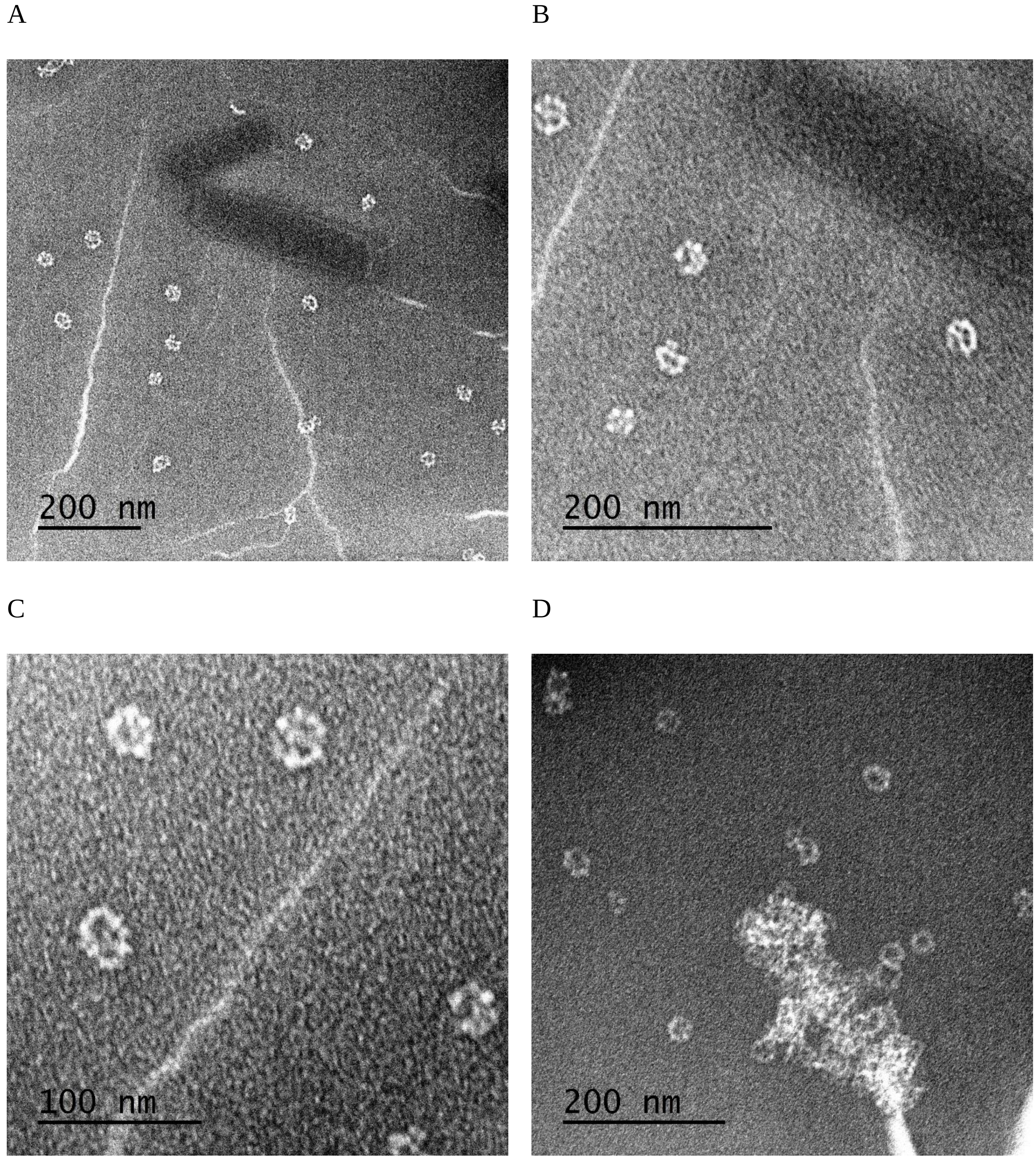
Three panels show dynamic light scattering size measurements for full antibody alone (top), designed tetramer 7B alone (middle), and assembly products when the two are mixed (bottom). The image in the last panel shows the intended assembly with only Fc, for visual clarity, but full antibody was used in the experiment. In the center panel (tetramer), the IgG binding portion that includes Staphylococcal protein A is colored cyan.

Figure 3.4



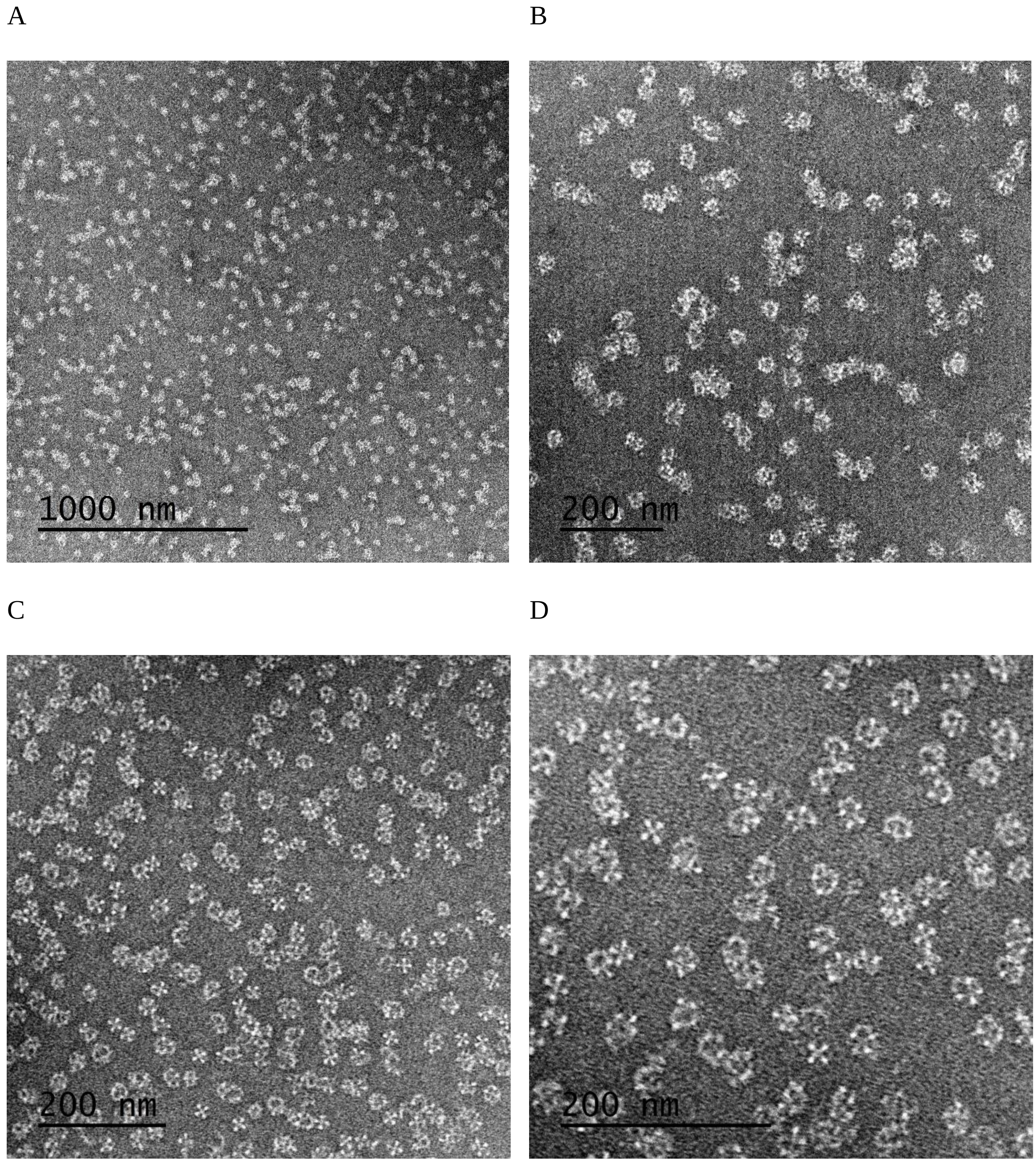
SEC of Fc and Fc in combination with design 7B on a Superose 6 10/300 GL column. In the Fc+7B case, design 7B is in excess, accounting for the peak shift.

Figure 3.5



Initial EM of designed tetramer 7B after co-assembly with Fc. Panels A-C show assemblies of tetramer 7B with Fc from human serum at increasing magnification with negative-stain electron microscopy. White lines are cracks in the deposited stain. Panel D shows ill-defined aggregates that covered most of the grid: misassembled particles, flocculation, or some combination.

Figure 3.6



Additional EM of designed tetramer 7B after co-assembly with antibody or Fc. Panels A-B show assemblies of tetramer 7B with full antibody and panels C-D are with Fc only.

References

---

1. Padilla, J. E., Colovos, C., & Yeates, T. O. (2001). Nanohedra: Using symmetry to design self assembling protein cages, layers, crystals, and filaments. Proceedings of the National Academy of Sciences, 98(5), 2217–2221. doi: 10.1073/pnas.041614998

2. Sinclair, J. C., Davies, K. M., Vénien-Bryan, C., & Noble, M. E. M. (2011). Generation of protein lattices by fusing proteins with matching rotational symmetry. Nature Nanotechnology, 6(9), 558–562. doi: 10.1038/nnano.2011.122

3. Lai, Y. T., Cascio, D., & Yeates, T. O. (2012). Structure of a 16-nm cage designed by using protein oligomers. Science, 336(6085), 1129. doi: 10.1126/science.1219351

4. Badieyan, S., Sciore, A., Eschweiler, J. D., Koldewey, P., Cristie-David, A. S., Ruotolo, B. T., … Marsh, E. N. G. (2017). Symmetry-Directed Self-Assembly of a Tetrahedral Protein Cage Mediated by de Novo-Designed Coiled Coils. ChemBioChem, 18(19), 1888–1892. doi: 10.1002/cbic.201700406

5. Lai, Y. T., Reading, E., Hura, G. L., Tsai, K. L., Laganowsky, A., Asturias, F. J., … Yeates, T. O. (2014). Structure of a designed protein cage that self-assembles into a highly porous cube. Nature Chemistry, 6(12), 1065–1071. doi: 10.1038/nchem.2107

6. Sciore, A., Su, M., Koldewey, P., Eschweiler, J. D., Diffley, K. A., Linhares, B. M., … Marsh, E. N. G. (2016). Flexible, symmetry-directed approach to assembling protein cages. Proceedings of the National Academy of Sciences of the United States of America, 113(31), 8681–8686. doi: 10.1073/pnas.1606013113

7. Cristie-David, A. S., Chen, J., Nowak, D. B., Bondy, A. L., Sun, K., Park, S. I., … Marsh, E. N. G. (2019). Coiled-Coil-Mediated Assembly of an Icosahedral Protein Cage with Extremely

High Thermal and Chemical Stability. Journal of the American Chemical Society, 141(23), 9207–9216. doi: 10.1021/jacs.8b13604

8. Cannon, K. A., Nguyen, V. N., Morgan, C., & Yeates, T. O. (2020). Design and Characterization of an Icosahedral Protein Cage Formed by a Double-Fusion Protein Containing Three Distinct Symmetry Elements. ACS Synthetic Biology. doi: 10.1021/acssynbio.9b00392

9. King, N. P., Sheffler, W., Sawaya, M. R., Vollmar, B. S., Sumida, J. P., André, I., … Baker, D. (2012). Computational design of self-assembling protein nanomaterials with atomic level accuracy. Science, 336(6085), 1171–1174. doi: 10.1126/science.1219364

10. King, N. P., Bale, J. B., Sheffler, W., McNamara, D. E., Gonen, S., Gonen, T., … Baker, D. (2014). Accurate design of co-assembling multi-component protein nanomaterials. Nature, 510(7503), 103–108. doi: 10.1038/nature13404

11. Gonen, S., DiMaio, F., Gonen, T., & Baker, D. (2015). Design of ordered two-dimensional arrays mediated by noncovalent protein-protein interfaces. Science. doi: 10.1126/science.aaa9897

12. Bale, J. B., Gonen, S., Liu, Y., Sheffler, W., Ellis, D., Chantz, T., … Baker, D. (2016). Accurate design of megadalton-scale two-component icosahedral protein complexes. Science, 353(6297), 389–395. doi: 10.5061/dryad.8c65s

13. Lai, Y. T., King, N. P., & Yeates, T. O. (2012). Principles for designing ordered protein assemblies. Trends in Cell Biology. doi: 10.1016/j.tcb.2012.08.004

14. Lai, Y. T., Tsai, K. L., Sawaya, M. R., Asturias, F. J., & Yeates, T. O. (2013). Structure and flexibility of nanoscale protein cages designed by symmetric self-assembly. Journal of the American Chemical Society. doi: 10.1021/ja402277f

15. Kratz, P. A., Böttcher, B., & Nassal, M. (1999). Native display of complete foreign protein domains on the surface of hepatitis B virus capsids. Proceedings of the National Academy of Sciences of the United States of America. doi: 10.1073/pnas.96.5.1915

16. Coscia, F., Estrozi, L., Hans, F. et al. Fusion to a homo-oligomeric scaffold allows cryo-EM analysis of a small protein. Sci Rep 6, 30909 (2016). doi: 10.1038/srep30909

17. Liu, Y., Gonen, S., Gonen, T., & Yeates, T. O. (2018). Near-atomic cryo-EM imaging of a small protein displayed on a designed scaffolding system. Proceedings of the National Academy of Sciences, 201718825. doi: 10.1073/pnas.1718825115

18. Liu, Y., Huynh, D. T., & Yeates, T. O. (2019). A 3.8 Å resolution cryo-EM structure of a small protein bound to an imaging scaffold. Nature Communications, 10(1), 1–7. doi: 10.1038/s41467-019-09836-0

19. Yao, Q., Weaver, S. J., Mock, J. Y., & Jensen, G. J. (2019). Fusion of DARPin to Aldolase Enables Visualization of Small Protein by Cryo-EM. Structure, 27(7), 1148-1155.e3. doi: 10.1016/j.str.2019.04.003

20. Watson, R. P., Christen, M. T., Ewald, C., Bumbak, F., Reichen, C., Mihajlovic, M., … Zerbe, O. (2014). Spontaneous self-assembly of engineered armadillo repeat protein fragments into a folded structure. Structure. doi: 10.1016/j.str.2014.05.002

21. Doyle, L., Hallinan, J., Bolduc, J., Parmeggiani, F., Baker, D., Stoddard, B. L., & Bradley, P. (2015). Rational design of α-helical tandem repeat proteins with closed architectures. Nature. doi: 10.1038/nature16191

22. Michel, E., Plückthun, A., & Zerbe, O. (2018). Peptide-Guided Assembly of Repeat Protein Fragments. Angewandte Chemie - International Edition. https://doi.org/10.1002/anie.201713377

23. Correnti, C. E., Hallinan, J. P., Doyle, L. A., Ruff, R. O., Jaeger-Ruckstuhl, C. A., Xu, Y., … Bradley, P. (2020). Engineering and functionalization of large circular tandem repeat protein nanoparticles. Nature Structural & Molecular Biology, 1–9. doi: 10.1038/s41594-020-0397-5

24. Leaver-Fay, A. Tyka, M., Lewis, S.M., Lange, O.F., Thompson, J., Jacak, R. … Bradley, P. (2011) ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules. Methods Enzymol. 487, 545–574

25. Hansen, S., Stüber, J. C., Ernst, P., Koch, A., Bojar, D., Batyuk, A., & Plückthun, A. (2017). Design and applications of a clamp for Green Fluorescent Protein with picomolar affinity. Scientific Reports, 7(1), 1–16. doi: 10.1038/s41598-017-15711-z

26. Brunette, T. J., Parmeggiani, F., Huang, P. S., Bhabha, G., Ekiert, D. C., Tsutakawa, S. E., … Baker, D. (2015). Exploring the repeat protein universe through computational protein design. Nature, 528(7583), 580–584. doi: 10.1038/nature16162

27. Boyken, S. E., Chen, Z., Groves, B., Langan, R. A., Oberdorfer, G., Ford, A., … Baker, D. (2016). De novo design of protein homo-oligomers with modular hydrogen-bond network-mediated specificity. Science, 352(6286), 680–687. doi: 10.1126/science.aad8865

28. Fallas, J. A., Ueda, G., Sheffler, W., Nguyen, V., McNamara, D. E., Sankaran, B., … Baker, D. (2017). Computational design of self-assembling cyclic protein homo-oligomers. Nature Chemistry. doi: 10.1038/nchem.2673

29. Mastronarde, D. N. (2005). Automated electron microscope tomography using robust prediction of specimen movements. Journal of Structural Biology. doi: 10.1016/j.jsb.2005.07.007

30. Li, X., Mooney, P., Zheng, S., Booth, C. R., Braunfeld, M. B., Gubbens, S., … Cheng, Y. (2013). Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. Nature Methods. doi: 10.1038/nmeth.2472

31. Grant, T., & Grigorieff, N. (2015). Measuring the optimal exposure for single particle cryo-EM using a 2.6 Å reconstruction of rotavirus VP6. ELife. https://doi.org/10.7554/eLife.06980

32. Zhang, K. (2016). Gctf: Real-time CTF determination and correction. Journal of Structural Biology. doi: 10.1016/j.jsb.2015.11.003

33. Zivanov, J., Nakane, T., Forsberg, B. O., Kimanius, D., Hagen, W. J. H., Lindahl, E., & Scheres, S. H. W. (2018). New tools for automated high-resolution cryo-EM structure determination in RELION-3. ELife. doi: 10.7554/eLife.42166

34. Punjani, A., Rubinstein, J. L., Fleet, D. J., & Brubaker, M. A. (2017). CryoSPARC: Algorithms for rapid unsupervised cryo-EM structure determination. Nature Methods. doi: doi.org/10.1038/nmeth.4169

35. Daniel Asarnow, Eugene Palovcak, & Yifan Cheng. (2019, December 16). asarnow/pyem: UCSF pyem v0.5 (Version v0.5). Zenodo. http://doi.org/10.5281/zenodo.3576630

36. Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004). UCSF Chimera - A visualization system for exploratory research and analysis. Journal of Computational Chemistry. doi: 10.1002/jcc.20084

37. Emsley, P., & Cowtan, K. (2004). Coot: Model-building tools for molecular graphics. Acta Crystallographica Section D: Biological Crystallography. doi: 10.1107/S0907444904019158

38. Kucukelbir, A., Sigworth, F. J., & Tagare, H. D. (2014). Quantifying the local resolution of cryo-EM density maps. Nature Methods. doi: 10.1038/nmeth.2727

39. VanAernum ZL, Busch F, Jones BJ, et al. Rapid online buffer exchange for screening of proteins, protein complexes and cell lysates by native mass spectrometry. Nat Protoc. 2020;15(3):1132-1157. doi:10.1038/s41596-019-0281-0

40. van de Waterbeemd M, Fort KL, Boll D, et al. High-fidelity mass analysis unveils heterogeneity in intact ribosomal particles. Nat Methods. 2017;14(3):283-286. doi:10.1038/nmeth.4147

41. Fort KL, van de Waterbeemd M, Boll D, et al. Expanding the structural analysis capabilities on an Orbitrap-based mass spectrometer for large macromolecular complexes. Analyst. 2017;143(1):100-105. doi:10.1039/c7an01629h

42. VanAernum ZL, Gilbert JD, Belov ME, Makarov AA, Horning SR, Wysocki VH. Surface-Induced Dissociation of Noncovalent Protein Complexes in an Extended Mass Range Orbitrap Mass Spectrometer. Anal Chem. 2019;91(5):3611-3618. doi:10.1021/acs.analchem.8b05605

43. Marty MT, Baldwin AJ, Marklund EG, Hochberg GK, Benesch JL, Robinson CV. Bayesian deconvolution of mass and ion mobility spectra: from binary interactions to polydisperse ensembles. Anal Chem. 2015;87(8):4370-4376. doi:10.1021/acs.analchem.5b00140

44. F.W. Studier
Protein production by auto-induction in high-density shaking cultures
Protein Expr. Purif., 41 (2005), pp. 207-234, doi: 10.1016/J.PEP.2005.01.016

45. Schneidman-Duhovny, D., Hammel, M., Tainer, J. A., & Sali, A. (2013). Accurate SAXS profile computation and its assessment by contrast variation experiments. Biophysical Journal, 105(4), 962–974. doi: 10.1016/j.bpj.2013.07.020

46. Schneidman-Duhovny, D., Hammel, M., Tainer, J. A., & Sali, A. (2016). FoXS, FoXSDock and MultiFoXS: Single-state and multi-state structural modeling of proteins and their complexes based on SAXS profiles. Nucleic Acids Research. doi: 10.1093/nar/gkw389

47. Genetic Engineering & Biotechnology News. (2018) The Top 15 Best-Selling Drugs of 2017. Retrieved from https://www.genengnews.com/a-lists/the-top-15-best-selling-drugs-of-2017/

48. Fan, G., Wang, Z., Hao, M., & Li, J. (2015). Bispecific antibodies and their applications. J Hematol Oncol, 8, 130. doi: 10.1186/s13045-015-0227-0

49. U.S. Food and Drug Administnration (2017) FDA Approves Gemtuzumab Ozogamicin for CD33-positive AML. Retrieved from https://www.fda.gov/Drugs/InformationOnDrugs/ApprovedDrugs/ucm574518.htm

50. U.S. Food and Drug Administration (2017) FDA approves inotuzumab ozogamicin for relapsed or refractory B-cell precursor ALL. Retrieved from https://www.fda.gov/Drugs/InformationOnDrugs/ApprovedDrugs/ucm572133.htm

51. ADC Review (2016) Brentuximab Vedotin (SGN35) Drug Description. Retrieved from https://adcreview.com/brentuximab-vedotin-sgn35

52. Genentech (2013) FDA Approves Genentech's Kadcyla (Ado-Trastuzumab Emtansine), the First Antibody-Drug Conjugate for Treating Her2-Positive Metastatic Breast Cancer. Retrieved from https://www.gene.com/media/press-releases/14347/2013-02-22/fda-approves-genentechs-kadcyla-ado-tras

53. Jain, N., Smith, S. W., Ghone, S., & Tomczuk, B. (2015). Current ADC Linker Chemistry. Pharmaceutical Research. doi: 10.1007/s11095-015-1657-7

54. Corchero JL, Cedano J. Self-assembling, protein-based intracellular bacterial organelles: emerging vehicles for encapsulating, targeting and delivering therapeutical cargoes. Microb Cell Fact. 2011;10:92. Published 2011 Nov 3. doi:10.1186/1475-2859-10-92

55. Idusogie, E. E., Presta, L. G., Gazzano-Santoro, H., Totpal, K., Wong, P. Y., Ultsch, M., … Mulkerrin, M. G. (2000). Mapping of the C1q Binding Site on Rituxan, a Chimeric Antibody

---

with a Human IgG1 Fc. The Journal of Immunology, 164(8), 4178–4184. doi:

10.4049/jimmunol.164.8.4178