The Design of Symmetric Protein Nanomaterials Comes of Age in Theory and Practice

Todd O. Yeates^{1,2,3}, Yuxi Liu¹, Joshua Laniado³

¹UCLA Department of Chemistry and Biochemistry

²UCLA-DOE Institute for Genomics and Proteomics

³UCLA-Molecular Biology Institute

Contact Information:

Todd O. Yeates

UCLA Department of Chemistry and Biochemistry

611 Charles Young Dr. East

Los Angeles, CA 90095-1569

yeates@mbi.ucla.edu

(tel) 310-206-4866

(fax) 310-206-3914

Key words: protein design, protein assembly, symmetry, protein cages. bionanotechnology

Abstract

In nature, protein molecules have evolved as building blocks for the assembly of diverse and complex structures, many of which exhibit a high degree of symmetry. This observation has motivated a number of recent engineering efforts in which the advantages of symmetry have been exploited to design novel self-assembling protein structures of great size. Materials ranging from cages to extended two and three-dimensional arrays have been demonstrated. Especially for extended arrays, a vast number of geometrically different design types are possible. A table of geometric rules is provided for designing a universe of novel materials by combining two component symmetries.

INTRODUCTION

Building blocks that have self-complimentary interfaces can self-assemble into elaborate structures. Nature serves as a rich source of inspiring specimens. At the macromolecular scale, viral capsids are quintessential examples, but other equally extraordinary macromolecular assemblies abound in nature (reviewed in [1-3]). The beauty and functional utility of these assemblies have long-motivated engineering efforts to create comparable structures in the laboratory. Beginning in the 1980's Ned Seeman pioneered ideas for using DNA molecules as building blocks for nanostructures [4]. Over the years, those ideas and various strategic variations led to the creation of elaborate supramolecular architectures and design patterns based on nucleic acids (reviewed in [5]). In nature, protein molecules have been the choice for the evolution of large assemblies with diverse form and function. But the engineering path to following Nature's lead has been challenged by the complexity of the rules that govern protein folding and assembly. To overcome those challenges, special strategies are needed.

In developing a strategic approach for building with protein molecules, Nature provides a major clue. Symmetry prevails in naturally evolved protein assemblies. This is an empirical fact evident in the vast

database of known macromolecular structures [6,7], but the prevalence of high symmetry in large protein assemblies was anticipated at least as far back as 1956 when Crick and Watson emphasized that viral capsids were likely able to evolve more easily in symmetric forms because symmetric assemblies require the fewest number of distinct interfacial contacts between individual subunits [8]. That key observation applies as well to designed structures, and indeed the early history of designing protein assemblies is rich with cases of relatively simple symmetric structures such as dimers and helical filaments [9-12]. The push in recent years to create very large protein assemblies has been guided even more strongly by principles of symmetry.

SYMMETRY-BASED DESIGN STRATEGIES

The symmetry of an object is fully described by the set of spatial operations (e.g. rotations) that leave the entire object unchanged except for an undetectable exchange of identical subunits. Because the symmetry of an object obeys the properties of a mathematical group, each specific type of symmetry is often referred to as a symmetry group. The symmetry group of a structure can be used to understand how many structurally distinct contact types are required to hold all the subunits together in one connected object. Certain simple types of architecture can be created from a building block that touches itself in just one way; i.e. using a single contact type. The possible outcomes are limited to structures like cyclic rings of subunits, or head-to-tail filaments (Fig 1A). More complex architectures require building blocks with more than one distinct interface.

A relatively simple group theory analysis explains the minimum number of distinct contact types required to achieve a given target symmetry. This was articulated first in the context of three-dimensional crystals [13] and then in the context of designed protein assemblies by Padilla et al. [14]. For example, if all the elements of a symmetry group can be generated by repeated application of a single element of the group

(i.e. a rotational operation), then one contact type is sufficient. The cyclic or single filament architectures noted above are examples of this type. If two elements from the symmetry group must be chosen in order to obtain the full symmetry group by repeated operations, then two contact types are required, and so on. Surprisingly, it turns out that a great many types of symmetry – including finite cages and many extended two and three-dimensional arrays – can be generated using just two properly chosen symmetry elements in combination (Fig. 1A). This key point frames the problem of designing novel protein assemblies by prescribing the number of distinct contact types that must be built into a protein building block in order for it to assemble into the desired architecture. More than the minimum number of contact types can be present in a final assembly, but not fewer. The specific geometry of the interfaces is of course crucial for obtaining the desired result, and molecular strategies are required for creating these oriented interfaces.

Various strategies have been developed for building multiple distinct interfacial contact types into a protein molecule in order to generate elaborate supramolecular structures (Fig. 1B). Padilla et al. [14] laid out a first strategy at a time when prospects for designing de novo interfaces into protein molecules were still remote. By necessity, naturally evolved interfaces were exploited by using simple natural protein oligomers (e.g. dimers and trimers) as a starting point. To create a single molecular building block containing two distinct interface types, a method was developed for genetically fusing two naturally oligomeric protein domains. In order to control relative geometry, only oligomeric domains having terminal alpha helices were considered, so that directly fusing two such proteins might create a geometrically predictable outcome if a continuous alpha helix was preserved between the two domains. The diversity of architectures possible by the general approach was described, and a first demonstration was provided – a 12-subunit assembly in the form of a tetrahedral cage was designed from a dimer plus trimer fusion[14]. This protein assembled into geometric structures consistent with the design, among a

range of other polymorphic forms. Several years later, Lai et al [15,16] showed that introducing two or three amino acid mutations into the originally designed protein sequence was sufficient to produce 12-subunit assemblies in high yield, which could be crystallized and validated in atomic detail. Not surprisingly, some flexibility of the helix linker gave rise to assemblies that were flexed or deformed from perfect symmetry, but which otherwise conformed to the intended tetrahedral design [16]. A different, 24-subunit cubic cage in good agreement with its design, and closely obeying the intended octahedral symmetry, was subsequently demonstrated using the same helix fusion strategy [17].

Major leaps forward in design strategy were made by King et al. [18] working with globular protein domains and Lanci et al. [19] working with coiled-coil polypeptides; they foresaw that computational methods for introducing novel interfaces into protein surfaces by amino acid sequence design had matured to the point where they might allow large symmetric assemblies to be created. Grueninger et al., took an earlier step in this direction by designing double-ring assemblies from naturally cyclic structures [20]. Following the symmetry ideas discussed above, starting with simple oligomeric proteins (e.g. dimers and trimers) means that only one additional interface needs to be designed into the protein in order to create complex architectures (Fig. 1B). In King et al., the procedure was enabled by a special algorithm written for the Rosetta-Design program to preserve overall symmetry while sampling the rigid body degrees of freedom available to the component oligomers [21]. From 41 initial designs, two cubic cages were produced in high yield and could be validated by crystallography[18]. One was a tetrahedral cage built from four trimers situated at the vertices of a tetrahedron and contacting each other primarily via a designed interface with two-fold symmetry. The other was a cubic/octahedral cage built from eight trimers at the corners of a cube, again with the trimers interacting primarily via a designed interface with two-fold symmetry. The polypeptide design work by Lanci et al. [19] relied on a trimeric coil-coil motif as

the starting point; the introduction of lateral and vertical contacts gave a three dimensional crystalline material also validated by crystallography.

Within the bounds of symmetry-based methods of design, several strategic variations are possible beyond those noted above. Some of the possibilities are briefly described here (Fig. 1B). Sinclair et al. [22] introduced a variation on the oligomer fusion method that relaxed the requirement for a continuous alpha helical linker; it applies to certain extended two and three-dimensional assemblies where two oligomers can be fused in a way that preserves a rotational symmetry element they both share. A few cases of well-ordered layers were demonstrated with that approach. Sciore et al. created a 24-subunit octahedral cage by fusing a naturally trimeric protein to a sequence that forms a parallel tetrameric coiled coil. The geometry of the connection in that case was also uncontrolled, but the limited range of possible outcomes from combining those two symmetries favored assembly to a single type of structure. King et al. [23] introduced a two-component variation on the de novo interface design strategy. Two different natural oligomers comprise the starting materials, and computational sequence design is used to introduce a heterotypic interface between the two subunit types. By relying on two separate oligomeric components, the idea shares similarity with the helix fusion method. But the helix fusion is rendered unnecessary by the designed interface between the two oligomers. Furthermore, the non-covalent nature of the association between the two oligomeric components enables production and purification of separate components, with full assembly occurring upon mixing. In the first application of the two-component strategy, a series of approximately 60 designed cubic cages were tested experimentally, and five were validated in detail by crystallography [23,24]. A new application of the 2-component approach by Bale et al. has produced a whole suite of even larger cages with icosahedral symmetry, the largest of which is approximately 400 Å in diameter[25].

In other design approaches, metals or bivalent ligands have been used as a way to introduce a new interface or self-associating interaction between oligomeric components ([26-28], reviewed in [29]). The addition of metals promotes assembly when metal binding half-sites (e.g. two suitably disposed histidine residues) are designed into a protein surface. Without further computational design of protein-protein interactions beyond the metal site, the metal site approach tends to give rise to assemblies whose outcomes are hard to predict in detail owing to alternate possible orientations of the metal ligands. Interestingly, despite incomplete control over geometry, various reports [30-37] have shown that the approach can be used to create diverse and intriguing materials ranging from small oligomers to helical structures to dynamic layers and three-dimensional arrays. A particularly notable variation was demonstrated recently by the introduction of a Zn-binding site at the three-fold symmetry axis of the natural, cubically symmetric ferritin cage. Adding a bivalent, metal-binding organic ligand joins the cubic protein cages into a three-dimensional body centered crystal lattice in a predictable fashion [36]. In a report by Sakai et al., the authors used bivalent organic ligands to create an additional dimeric subunit association for linking D2 tetramers together, in an unspecified orientation in this case. Changing the spacer length between the binding moieties on the organic ligand resulted in protein molecules held together in layers, which grew into two different types of three-dimensional crystals [38]. The design of rigid polyvalent ligands offers prospects for connecting proteins together in geometrically controlled ways. Particularly in view of the large number of proteins that naturally bind rigid (e.g. multi-ring) ligands, such a strategy remains relatively untapped.

Other strategic variations for combining symmetry elements are possible but have not been deeply explored (Fig. 1B). DNA (or RNA) provides a facile route for directing molecular associations by complimentary base pairing. This can be exploited in combination with DNA binding proteins and either interface design or oligomeric fusion to create hybrid materials composed of proteins and nucleic acids.

A strategy along that line has been taken by Mou et al. in creating linear or helical filaments [39]. In another study, Brodin, et al. attached multiple copies of two different DNA molecules onto the surfaces of proteins to create two component types, which then assembled into cubically packed crystals upon addition of a complimentary bridging DNA molecule [40]. DNA and RNA both offer an easy route for introducing a symmetry element into designed assemblies using palindromic sequences. The construction of symmetric, hybrid protein-nucleic acid materials using such a strategy has not been tested yet.

DESIGNED PROTEIN CAGES

Designed assemblies of defined, finite size can take the form of shells or cages or more compact clusters following one of the three possible cubic symmetries in three dimensions. These are based on the symmetries of the five Platonic solids; the cube and the octahedron are duals of each other while the icosahedron and the dodecahedron are duals, so together with the tetrahedron (which is its own dual), there are three possible symmetries: T (order 12), O (order 24), and I (order 60). The symmetry rules for generating these have been articulated [14,41] (Fig. 2A). For each symmetry type, nearly every combination of two *intersecting* symmetry axes is a possible choice for generating the final structure, but there are some exceptions (Fig. 2 legend).

Using designed protein molecules, cages or shells belonging to all three of the cubic symmetry types (T, O, and I) have been successfully produced. Those that have been validated in atomic detail by X-ray crystallography are shown in Figure 2B. Each of the target symmetries has been obtained using multiple different symmetry combinations. Symmetry T has been obtained by combinations of two components or interfaces having 2-fold plus 3-fold symmetry [14,15,18,23] and 3-fold plus 3-fold symmetry [23]. Symmetry O has been obtained by 3-fold plus 2-fold symmetries [17,18,23] and by 4-fold plus 3-fold symmetries [42]. Finally, icosahedral symmetry has been obtained in new studies, using all three possible

symmetry combinations (2 plus 3, 2 plus 5, and 3 plus 5)[25,43]. Together, these designed architectures range in number of subunits from 12 to 120, with diameters from 11 to 40 nm and masses from 276 kDa to 2.8 MDa. They cover most of the available symmetry options for creating cage and shell structures. The numerous possible applications for designed protein cages have been reviewed elsewhere [44-47] and are not elaborated here other than to summarize that practical uses are likely to include both interior and exterior capabilities: (i) encapsulation, delivery and release of molecular cargo [48-53], and (ii) polyvalent display of motifs for signaling or antigenic effects, as in synthetic vaccines [54,55].

DESIGN RULES FOR BUILDING EXTENDED MATERIALS IN TWO AND THREE DIMENSIONS

In contrast to the finite cage designs that arise from two intersecting symmetry axes as discussed above, when two component symmetries are combined in an arrangement where any of their axes are not intersecting, the result cannot be finite and must instead be an extended or unbounded material. Filaments are one possible outcome, arising from two non-intersecting 2-fold axes of symmetry [14]. But more complex outcomes are obtained by combining higher symmetries. There, unbounded materials that extend as either 2-dimensional layers or 3-dimensional arrays (i.e. crystals) are possible. The geometric rules for a few possibilities of this type were laid out earlier based on the combination of 2-fold and 3-fold axes of symmetry [14]. Beyond those, a vast range of possibilities arise from combinations of higher component symmetries. A few designs within that scope have been demonstrated in recent work [56,57], but a complete set of geometric rules for generating two and three-dimensional materials has not been articulated previously. In order to promote further studies, we provide a list of the allowable symmetry combinations here (Table 1).

Two-dimensional layers can be of two different classes depending on their sidedness, or lack thereof. When two symmetry axes that are both perpendicular to the layer are combined, and at least one of those

axes is of higher order than a 2-fold, the result is a layer with distinguishable sides (i.e. a distinct top and bottom). In that sense such layers are oriented. The allowable symmetry combinations for oriented layers are 2+3, 2+4, 2+6, 3+3, 3+6, and 4+4 (Table 1). Layered structures of the other class, where the top and bottom of the layer are indistinguishable, arise whenever one or both of the two symmetries being combined carries a 2-fold axis of symmetry in the plane of the layer. A total of 33 layer designs are possible (Table 1).

Much of the design space for designed layers is unexplored, but a few recent successes have been reported. Small two-dimensional assemblies with limited range order were described in early work by Ringler et al. by doubly biotinylating the subunits of aldolase (a C₄ tetramer) and then assembling those tetramers using streptavidin (a D₂ tetramer with a biotin binding site in each subunit) [27]. If long range order had been achieved in this case, the result would have corresponded to layer symmetry p4₂2. Longer range order in designed layers was demonstrated by Sinclair et al. [22]. There the best case was obtained by combining the D₂ tetrameric streptavidin with a D₄ octameric protein that had been biotinylated. The relative orientations of the components in that case were not specifically designed and could have produced other results, but a two-dimensional layer with p4₂2 symmetry was obtained (Fig. 3). More recently, computational interface design was used to create specifically defined protein layers of a few different types [56]. A 2-fold interface was designed between natural C₆ hexameric units to give layer symmetry p6, between natural C₄ tetrameric units to give layer symmetry p4₂1₂, and between natural C₃ trimeric units to give layer symmetry p3₂1 (Fig. 3). In another variation, two copies of a C₆ hexameric protein were fused in tandem in such a way that the short linker between them, along with a designed 2-fold interface, led to a pseudo-p6 layer [57]. In a most recent study, C₄ symmetric units were connected through 2-fold interactions based on engineered cysteine disulfide bonds or engineered metal binding

sites[37]. Multiple kinds of symmetric outcomes are possible from such a combination, and 2-D layers belonging to $p4$ and $p4_212$ symmetry types were obtained.

For designing extended three-dimensional crystalline materials, the possibilities are even more expansive, and the design rules are not so obvious. Defining exactly what symmetry type would be generated from a combination of two separate symmetries relies on two considerations. The overall rotational symmetry of the resulting material is given by the (group) product of the rotational symmetries of the two components; this is a relatively straightforward issue. A somewhat more complex problem is discerning the correct outcome among a set of candidate space groups once the underlying rotational symmetry is known. The correct choice can generally be ascertained from the standard tables of crystallographic space groups by identifying the one having Wyckoff positions with symmetries corresponding to those of the two components being combined. In total, more than 70 distinct types of 3D materials are possible within the scheme of combining two types of rotation (point group) symmetries. The majority of them have underlying cubic (T or O) symmetries and are therefore isotropic, while the others are dihedral.

On the experimental side, the space of designed protein crystals is mostly unexplored, but there are a few early examples. As discussed above, Lanci et al. [19] designed a coiled-coil peptide to form $P6$ space group symmetry, and Sontz et al. [36] combined a ferritin, into which a metal site had been engineered, with a bivalent bridging compound to form a crystal whose space group symmetry was pseudo- $I432$. Sinclair, et al. used their fusion method to form three-dimensional solid materials, but without sufficient order to confirm the intended crystalline packing by X-ray diffraction [22]. Beyond these few examples, the area of designed protein crystals remains open. The possible applications for such materials are diverse: creating materials with a very high density of reactive/catalytic groups or recognition motifs, and

conferring specific physical properties on target proteins, including spacing, dimensionality, porosity, and solid-phase separability from solution components.

VARIATIONS, CHALLENGES AND FUTURE DIRECTIONS

The ideas and rules formulated here are somewhat narrowly constructed. They represent the simplest design routes (i.e. the minimum requirements) for construction by symmetric assembly. Broader outcomes are possible if interfaces between components are designed in different ways. For example, King's single-component designed interface method [18] allows for symmetries not accounted for in Table 1 if a new contact between like oligomers creates a screw axis of symmetry instead of a pure rotation. Ultimately, in the absence of limits on designing protein-protein interfaces, any symmetric architecture could be designed, including those that require larger numbers of distinct interaction types. The three-dimensional crystal in space group P6 designed by Lanci et al. [19] is a case in point using relatively simple building blocks. In addition, other approaches for designing large protein and peptide-based structures have been developed that rely less strictly, or in different ways, on symmetry. Fletcher et al. [58] combined a homotrimeric coiled-coil and a heterodimeric coiled-coil that interact with each other. This resulted in unilamellar spheres approximately 100 nanometers in diameter, with overall structures that were not exactly symmetric though assembly was driven by local symmetry. Doll et al. [59] also combined coiled-coil sequences with different symmetric properties (5-fold and 3-fold) to produce roughly spherical clusters having sizes consistent with icosahedral assembly. In a distinctly different line of attack, Gradišar et al. demonstrated the construction of a tetrahedral architecture on the basis of asymmetric interactions between coiled coil motifs in a long, designed protein molecule whose folding pattern traverses the edges of the entire polyhedron twice [60].

The different design strategies discussed here present their own advantages and challenges. The initial design strategy of helical fusions between oligomers presents a relatively low barrier in the sense of not needing to create de novo interfaces, but flexibility creates an obvious challenge. This sometimes leads to alternate assembly outcomes [17,61]. Strategies involving more rigid linkers could improve the reliability of this method. A recent report demonstrated some success in rigidifying a continuous alpha helical linker between two protein components by a specific chemical cross-link between cysteines at positions i and $i+11$ [62]. Another approach to rigidification would be to use a coiled-coil linker as the motif joining two separate oligomeric units, with a single helix of a hetero coiled-coil motif extending from each of the oligomeric components. These are avenues of ongoing study.

The main challenges with methods based on de novo interface design relate to misfolding or unintended assembly – often insoluble aggregation – most likely caused by introduction of new regions of hydrophobicity in a protein surface. The success rates for geometrically specific interface design in the context of symmetric assemblies is currently in the range of about 10%. One case has been reported where a failed design could be rescued by increasing the charge on the protein molecule [24]. This suggests that optimizing certain design parameters and selection criteria might substantially increase the success rates. Another notable trend is the increasing success in designing novel protein folds and assemblies by focusing on building blocks that are all (or mainly) alpha helical[63-66]. New strategies for high throughput selection or screening of designs for correct assembly could also be impactful.

As design strategies continue to improve, and with construction rules in hand for building wide-ranging types of symmetric architectures, the coming years should bring a rich diversity of new protein based materials with useful applications.

[annotated references confined to recent publications]

REFERENCES

1. Goodsell DS, Olson AJ: **Structural symmetry and protein function**. *Annual Review of Biophysics and Biomolecular Structure* 2000, **29**:105-153.
2. Marsh JA, Teichmann SA: **Structure, Dynamics, Assembly, and Evolution of Protein Complexes**. *Annual Review of Biochemistry* 2015, **84**:551-575.
3. Yeates TO, Thompson MC, Bobik TA: **The protein shells of bacterial microcompartment organelles**. *Current Opinion in Structural Biology* 2011, **21**:223-231.
4. Seeman NC: **Nucleic acid junctions and lattices**. *Journal of Theoretical Biology* 1982, **99**:237-247.
5. Jones MR, Seeman NC, Mirkin CA: **Nanomaterials. Programmable materials and the nature of the DNA bond**. *Science (New York, N.Y.)* 2015, **347**:1260901.
6. Ispolatov I, Yuryev A, Mazo I, Maslov S: **Binding properties and evolution of homodimers in protein-protein interaction networks**. *Nucleic Acids Research* 2005, **33**:3629-3635.
7. Pereira-Leal JB, Levy ED, Kamp C, Teichmann SA: **Evolution of protein complexes by duplication of homomeric interactions**. *Genome Biology* 2007, **8**:R51.
8. Crick FH, Watson JD: **Structure of small viruses**. *Nature* 1956, **177**:473-475.
9. Pandya MJ, Spooner GM, Sunde M, Thorpe JR, Rodger A, Woolfson DN: **Sticky-End Assembly of a Designed Peptide Fiber Provides Insight into Protein Fibrillogenesis**. *Biochemistry* 2000, **39**:8728-8734.
10. Ogihara NL, Ghirlanda G, Bryson JW, Gingery M, DeGrado WF, Eisenberg D: **Design of three-dimensional domain-swapped dimers and fibrous oligomers**. *Proceedings of the National Academy of Sciences* 2001, **98**:1404-1409.
11. Robertson DE, Farid RS, Moser CC, Urbauer JL, Mulholland SE, Pidikiti R, Lear JD, Wand AJ, DeGrado WF, Dutton PL: **Design and synthesis of multi-haem proteins**. *Nature* 1994, **368**:425-432.
12. Kuhlman B, O'Neill JW, Kim DE, Zhang KY, Baker D: **Conversion of monomeric protein L to an obligate dimer by computational protein design**. *Proceedings of the National Academy of Sciences of the United States of America* 2001, **98**:10687-10691.
13. Wukovitz SW, Yeates TO: **Why protein crystals favour some space-groups over others**. *Nature Structural Biology* 1995, **2**:1062-1067.
14. Padilla JE, Colovos C, Yeates TO: **Nanohedra: Using symmetry to design self assembling protein cages, layers, crystals, and filaments**. *Proceedings of the National Academy of Sciences* 2001, **98**:2217-2221.
15. Lai Y-T, Cascio D, Yeates TO: **Structure of a 16-nm Cage Designed by Using Protein Oligomers**. *Science* 2012, **336**:1129-1129.
16. Lai Y-T, Tsai K-L, Sawaya MR, Asturias FJ, Yeates TO: **Structure and Flexibility of Nanoscale Protein Cages Designed by Symmetric Self-Assembly**. *Journal of the American Chemical Society* 2013, **135**:7738-7743.
17. Lai Y-T, Reading E, Hura GL, Tsai K-L, Laganowsky A, Asturias FJ, Tainer JA, Robinson CV, Yeates TO: **Structure of a designed protein cage that self-assembles into a highly porous cube**. *Nature Chemistry* 2014, **6**:1065-1071.

- Using the alpha helical oligomer fusion strategy, the authors reported the design and crystallographic validation of a highly porous 24-subunit cubic assembly with a large central cavity, demonstrating the

feasibility of controlling geometry over long distance while providing porosity. Helix flexibility admitted alternative assembly forms.

18. King NP, Sheffler W, Sawaya MR, Vollmar BS, Sumida JP, André I, Gonen T, Yeates TO, Baker D: **Computational Design of Self-Assembling Protein Nanomaterials with Atomic Level Accuracy.** *Science* 2012, **336**:1171-1174.
19. Lanci CJ, MacDermaid CM, Kang S-g, Acharya R, North B, Yang X, Qiu XJ, DeGrado WF, Saven JG: **Computational design of a protein crystal.** *Proceedings of the National Academy of Sciences of the United States of America* 2012, **109**:7304-7309.
20. Grueninger D, Treiber N, Ziegler MOP, Koetter JWA, Schulze M-S, Schulz GE: **Designed protein-protein association.** *Science (New York, N.Y.)* 2008, **319**:206-209.
21. DiMaio F, Leaver-Fay A, Bradley P, Baker D, André I: **Modeling Symmetric Macromolecular Structures in Rosetta3.** *PLoS ONE* 2011, **6**:e20450.
22. Sinclair JC, Davies KM, Vénien-Bryan C, Noble MEM: **Generation of protein lattices by fusing proteins with matching rotational symmetry.** *Nature Nanotechnology* 2011, **6**:558-562.
23. King NP, Bale JB, Sheffler W, McNamara DE, Gonen S, Gonen T, Yeates TO, Baker D: **Accurate design of co-assembling multi-component protein nanomaterials.** *Nature* 2014, **510**:103-108.

- Using improved algorithms in the Rosetta software suite, the authors illustrate the design and validation of a series of novel protein cages wherein each cage assembles from two distinct oligomeric components. A de novo interface between the two types of oligomers drives the symmetric assembly.

24. Bale JB, Park RU, Liu Y, Gonen S, Gonen T, Cascio D, King NP, Yeates TO, Baker D: **Structure of a designed tetrahedral protein assembly variant engineered to have improved soluble expression.** *Protein Science* 2015, **24**:1695-1701.
25. Bale JB, Gonen S, Liu Y, Sheffler W, Ellis D, Thomas C, Cascio D, Yeates TO, Gonen T, King NP, et al.: **Accurate design of megadalton-scale two-component icosahedral protein complexes.** *Science* 2016, (in press).

- This study demonstrates the design and atomic level validation of the largest geometrically specific protein cages to date using a two-component strategy and interface design. Controlled encapsulation of cargo proteins by charge complementarity is demonstrated.

26. Dotan N, Arad D, Frolow F, Freeman A: **Self-Assembly of a Tetrahedral Lectin into Predesigned Diamondlike Protein Crystals.** *Angewandte Chemie (International Ed. in English)* 1999, **38**:2363-2366.
27. Ringler P, Schulz GE: **Self-Assembly of Proteins into Designed Networks.** *Science* 2003, **302**:106-109.
28. Salgado EN, Ambroggio XI, Brodin JD, Lewis RA, Kuhlman B, Tezcan FA: **Metal templated design of protein interfaces.** *Proceedings of the National Academy of Sciences of the United States of America* 2010, **107**:1827-1832.
29. Salgado EN, Radford RJ, Tezcan FA: **Metal-Directed Protein Self-Assembly.** *Accounts of Chemical Research* 2010, **43**:661-672.

30. Salgado EN, Lewis RA, Mossin S, Rheingold AL, Tezcan FA: **Control of Protein Oligomerization Symmetry by Metal Coordination – C2 and C3 Symmetrical Assemblies through Cu(II) and Ni(II) Coordination.** *Inorganic chemistry* 2009, **48**:2726-2728.
31. Brodin JD, Ambroggio XI, Tang C, Parent KN, Baker TS, Tezcan FA: **Metal-directed, chemically tunable assembly of one-, two- and three-dimensional crystalline protein arrays.** *Nature Chemistry* 2012, **4**:375-382.
32. Salgado EN, Lewis RA, Faraone-Mennella J, Tezcan FA: **Metal-Mediated Self-Assembly of Protein Superstructures: Influence of Secondary Interactions on Protein Oligomerization and Aggregation.** *Journal of the American Chemical Society* 2008, **130**:6082-6084.
33. Salgado EN, Faraone-Mennella J, Tezcan FA: **Controlling Protein–Protein Interactions through Metal Coordination: Assembly of a 16-Helix Bundle Protein.** *Journal of the American Chemical Society* 2007, **129**:13374-13375.
34. Laganowsky A, Zhao M, Soriaga AB, Sawaya MR, Cascio D, Yeates TO: **An approach to crystallizing proteins by metal-mediated synthetic symmetrization.** *Protein Science* 2011, **20**:1876-1890.
35. Leibly DJ, Arbing MA, Pashkov I, DeVore N, Waldo GS, Terwilliger TC, Yeates TO: **A Suite of Engineered GFP Molecules for Oligomeric Scaffolding.** *Structure* 2015, **23**:1754-1768.
36. Sontz PA, Bailey JB, Ahn S, Tezcan FA: **A Metal Organic Framework with Spherical Protein Nodes: Rational Chemical Design of 3D Protein Crystals.** *Journal of the American Chemical Society* 2015, **137**:11598-11601.

- The authors constructed a three-dimensional protein crystal with a prescribed lattice by introducing metal–organic linker interactions between adjacent ferritin cages, which are naturally cubic/octahedral. This is the first report of a designed metal-organic mediated 3-D protein crystal. Its designed structure was shown to be accurate by x-ray diffraction at atomic resolution.

37. Suzuki Y, Cardone G, Restrepo D, Zavattieri PD, Baker TS, Tezcan FA: **Self-assembly of coherently dynamic, auxetic, two-dimensional protein crystals.** *Nature* 2016, **533**:369-373.

- The authors obtain multiple well-ordered two-dimensional arrays from a cyclic tetrameric protein engineered to make 2-fold contacts with itself via disulfide or metal binding motifs. The layers exhibit shrinking/expansion transitions.

38. Sakai F, Yang G, Weiss MS, Liu Y, Chen G, Jiang M: **Protein crystalline frameworks with controllable interpenetration directed by dual supramolecular interactions.** *Nature Communications* 2014, **5**:4634.
39. Mou Y, Yu J-Y, Wannier TM, Guo C-L, Mayo SL: **Computational design of co-assembling protein-DNA nanowires.** *Nature* 2015, **525**:230-233.
40. Brodin JD, Auyeung E, Mirkin CA: **DNA-mediated engineering of multicomponent enzyme crystals.** *Proc Natl Acad Sci U S A* 2015, **112**:4564-4569.

- The authors create hybrid protein-DNA crystalline materials by chemically coating proteins with DNA and then adding bridging DNA molecules to drive association between two different types of protein particles.

41. Lai Y-T, King NP, Yeates TO: **Principles for designing ordered protein assemblies.** *Trends in Cell Biology* 2012, **22**:653-661.
 42. Sciore A, Su M, Koldewey P, Eschweiler JD, Diffley KA, Linhares BM, Ruotolo BT, Bardwell JCA, Skiniotis G, Marsh EN: **A flexible, symmetry-directed approach to assembling protein cages.** *Proceedings of the National Academy of Sciences of the United States of America* 2016, (in press).
- The authors create a 24-subunit assembly by fusing a trimeric protein to a tetrameric coiled-coil motif.
43. Hsia Y, Bale JB, Gonen S, Shi D, Sheffler W, Fong KK, Nattermann U, Xu C, Huang PS, Ravichandran R, et al.: **Design of a hyperstable 60-subunit protein icosahedron.** *Nature* 2016, 10.1038/nature18010.
- The authors create a 60-subunit icosahedral cage using interface design and a one-component strategy.
44. Yeates TO, Padilla JE: **Designing supramolecular protein assemblies.** *Current Opinion in Structural Biology* 2002, **12**:464-470.
 45. Doll TAPF, Raman S, Dey R, Burkhard P: **Nanoscale assemblies and their biomedical applications.** *Journal of the Royal Society Interface* 2013, **10**.
 46. López-Sagaseta J, Malito E, Rappuoli R, Bottomley MJ: **Self-assembling protein nanoparticles in the design of vaccines.** *Computational and Structural Biotechnology Journal* 2015, **14**:58-68.
 47. Howorka S: **Rationally engineering natural protein assemblies in nanobiotechnology.** *Current Opinion in Biotechnology* 2011, **22**:485-491.
 48. Wörsdörfer B, Woycechowsky KJ, Hilvert D: **Directed evolution of a protein container.** *Science (New York, N.Y.)* 2011, **331**:589-592.
 49. Patterson DP, Schwarz B, El-Boubbou K, Oost Jvd, Prevelige PE, Douglas T: **Virus-like particle nanoreactors: programmed encapsulation of the thermostable CelB glycosidase inside the P22 capsid.** *Soft Matter* 2012, **8**:10158-10166.
 50. Champion CI, Kickhoefer VA, Liu G, Moniz RJ, Freed AS, Bergmann LL, Vaccari D, Raval-Fernandes S, Chan AM, Rome LH, et al.: **A Vault Nanoparticle Vaccine Induces Protective Mucosal Immunity.** *PLOS ONE* 2009, **4**:e5409.
 51. Han M, Kickhoefer VA, Nemerow GR, Rome LH: **Targeted Vault Nanoparticles Engineered with an Endosomolytic Peptide Deliver Biomolecules to the Cytoplasm.** *ACS Nano* 2011, **5**:6128-6137.
 52. Kar UK, Jiang J, Champion CI, Salehi S, Srivastava M, Sharma S, Rabizadeh S, Niazi K, Kickhoefer V, Rome LH, et al.: **Vault Nanocapsules as Adjuvants Favor Cell-Mediated over Antibody-Mediated Immune Responses following Immunization of Mice.** *PLOS ONE* 2012, **7**:e38553.
 53. Zschoche R, Hilvert D: **Diffusion-Limited Cargo Loading of an Engineered Protein Container.** *Journal of the American Chemical Society* 2015, **137**:16121-16132.
 54. Correia BE, Bates JT, Loomis RJ, Baneyx G, Carrico C, Jardine JG, Rupert P, Correnti C, Kalyuzhnyi O, Vittal V, et al.: **Proof of principle for epitope-focused vaccine design.** *Nature* 2014, **507**:201-206.
 55. Sliepen K, Ozorowski G, Burger JA, van Montfort T, Stunnenberg M, LaBranche C, Montefiori DC, Moore JP, Ward AB, Sanders RW: **Presenting native-like HIV-1 envelope trimers on ferritin nanoparticles improves their immunogenicity.** *Retrovirology* 2015, **12**:82.
 56. Gonen S, DiMaio F, Gonen T, Baker D: **Design of ordered two-dimensional arrays mediated by noncovalent protein-protein interfaces.** *Science* 2015, **348**:1365-1368.

- The authors report on three examples of symmetric protein layers designed by introducing a de novo interface between symmetry-related copies of a natural cyclic oligomer. The layers showed long range order.

57. Mattheaei JF, DiMaio F, Richards JJ, Pozzo LD, Baker D, Baneyx F: **Designing Two-Dimensional Protein Arrays through Fusion of Multimers and Interface Mutations.** *Nano Letters* 2015, 10.1021/acs.nanolett.5b01499.

- The authors create an ordered 2-D array by genetically fusing two monomers of a cyclic hexameric protein in an orientation that drive extended rather than closed assembly.

58. Fletcher JM, Harniman RL, Barnes FRH, Boyle AL, Collins A, Mantell J, Sharp TH, Antognozzi M, Booth PJ, Linden N, et al.: **Self-Assembling Cages from Coiled-Coil Peptide Modules.** *Science* 2013, **340**:595-599.

- The authors show that a self-assembling material can be designed by linking trimeric and dimeric coiled-coils through disulfide bonds. Hexagonal network form, which then close up to form roughly spherical nanoparticles with diameters in the 100 nm range.

59. Doll TAPF, Dey R, Burkhard P: **Design and optimization of peptide nanoparticles.** *Journal of Nanobiotechnology* 2015, **13**:73.

60. Gradišar H, Božič S, Doles T, Vengust D, Hafner-Bratkovič I, Mertelj A, Webb B, Šali A, Klavžar S, Jerala R: **Design of a single-chain polypeptide tetrahedron assembled from coiled-coil segments.** *Nature chemical biology* 2013, **9**:362-366.

61. Lai Y-T, Jiang L, Chen W, Yeates TO: **On the predictability of the orientation of protein domains joined by a spanning alpha-helical linker.** *Protein Engineering Design and Selection* 2015, 10.1093/protein/gzv035:gzv035.

62. Jeong WH, Lee H, Song DH, Eom J-H, Kim SC, Lee H-S, Lee H, Lee J-O: **Connecting two proteins using a fusion alpha helix stabilized by a chemical cross linker.** *Nature Communications* 2016, **7**:11031.

63. Boyken SE, Chen Z, Groves B, Langan RA, Oberdorfer G, Ford A, Gilmore JM, Xu C, DiMaio F, Pereira JH, et al.: **De novo design of protein homo-oligomers with modular hydrogen-bond network-mediated specificity.** *Science* 2016, **352**:680-687.

64. Brunette TJ, Parmeggiani F, Huang PS, Bhabha G, Ekiert DC, Tsutakawa SE, Hura GL, Tainer JA, Baker D: **Exploring the repeat protein universe through computational protein design.** *Nature* 2015, **528**:580-584.

65. Huang PS, Oberdorfer G, Xu C, Pei XY, Nannenga BL, Rogers JM, DiMaio F, Gonen T, Luisi B, Baker D: **High thermodynamic stability of parametrically designed helical bundles.** *Science* 2014, **346**:481-485.

66. Thomson AR, Wood CW, Burton AJ, Bartlett GJ, Sessions RB, Brady RL, Woolfson DN: **Computational design of water-soluble alpha-helical barrels.** *Science* 2014, **346**:485-488.

Acknowledgements

This work was supported by NSF grant CHE-1332907. The authors thank Dan McNamara, Yen-Ting Lai, Kevin Cannon and other members of the Yeates lab for their ideas, and members of the David Baker laboratory, including Jacob Bale and Neil King, for access to their designs ahead of publication. We thank Frank DiMaio and Will Sheffler for discussions on symmetry rules and Dek Woolfson for discussions related to coiled-coil polypeptides.

Figure Legends

Figure 1. Assembly consequences and strategies for introducing multiple contact types into protein building blocks. A) Illustration of varied symmetric architectural forms and the number of distinct contact types required for connectivity between molecular building blocks. Two contact types are sufficient to create diverse assemblies. B) Different molecular strategies for creating a building block having two distinct contact types in a defined orientation. Left to right (top): alpha helical fusion; 1-component interface design; 2-component interface design. Left to right (bottom): metal or ligand bridging; coiled-coil helical fusions; designed symmetrization of DNA binding proteins.

Figure 2. Design and validation of self-assembling protein cages with high symmetry. A) The three types of cubic symmetry (T, O, and I) are illustrated on the framework of the Platonic solids. Angles between pairs of rotational symmetry axes that can be combined to create a self-assembling building block with the target symmetry are listed. B) Engineered protein cages obeying all three possible cubic symmetries have been produced; a subset of structures that have been validated by X-ray crystallography are shown to scale. Left panel: a tetrahedral cage (PDB 4IQ4) [16] and an octahedral cage (PDB 4QCC) [17] designed using the alpha helical fusion strategy [14]. Natural trimers are in blue; natural dimers are in yellow; the alpha-helical linkers are in red. Middle panel: tetrahedral cages and an octahedral cage engineered by de novo interface design with either one or two components. Top row are one-component designs (PDB 4EGG & 3VCD, left to right) [16], with each trimer shown in a different color for clarity. The middle and bottom rows (PDB 4NWR, 4NWP, 4NWO, 4NWN, & 4ZK7, left to right and top to bottom) are two-component designs [23,24]. Natural dimers are in yellow. Natural trimers are in blue and orange to differentiate the different trimers in the same design. Right panel: crystal structures of 2-component icosahedra obtained by all three possible symmetry combinations (PDB 5IM5 (I:5+3), 5IM4 (I:5+2), and 5IM6 (I:3+2)) [25]. Letters on the top left corner of the structure indicates the symmetry type (T:

tetrahedral, O: octahedral, I: icosahedral). The numbers in the annotation indicate the component symmetry types. Where present, parenthetical values indicate the symmetry of the main designed interface.

Figure 3. Electron micrographs of protein layers designed to assemble with high symmetry and showing long-range order. Symmetry diagrams are shown under each micrograph, accompanied in some cases by enlarged images. Each symmetry diagram shows the repeating unit cell within which one instance of each of the component symmetry elements is indicated using standard symbols: black arrows, 2-fold symmetry axes in the plane of the layer; black ovals, black triangles, black squares, and black hexagons indicate 2-, 3-, 4-, and 6- fold axes perpendicular to the plane of the layer, respectively. A) a p422 layer formed by combining D4 and D2 symmetry components; C) a p321 layer formed by combining C3 trimers with a 2-fold de novo interface; I) a p42₁2 layer formed by combining C4 tetramers with a 2-fold de novo interface; J) a p6 layer formed by combining C6 hexamers and a 2-fold de novo interface; L) a pseudo p6 layer formed by combining C6 hexamers using a covalent fusion and a pseudo 2-fold de novo interface. E, H, K: enlarged images of C, F, I, respectively. Scale bars: A – 20 nm; C – 50 nm; F – 50 nm; I – 50 nm; L – 20 nm. Images reproduced with permission from: panel A – Sinclair, J. C., et al. (2011) [22]; panels C, E, F, H, I, and K – Gonen, S., et al. (2015)[56]; panel L –Matthaei, J. F., et al. (2015) [57].

Table 1. Multiplication table for designing self-assembling protein materials from combinations of two simpler symmetric components or interfaces[#].

x	C2	C3	C4	C6	D2	D3	D4	D6	T	O
C2	‡	D3, T, O, I p6, p321 I213, P4132	D4, O p4, p4212 I432	D6 p6, p622	c222, p422, p622 I4122, P6222, I432, I4132	p312, p622 R32, P6322, F4132, I4132, I432, P4132	p422 I422, P432, I432	p622 P622	P23, F23, F4132	P432, F432, I432
C3		T p3 P213	O F432	p6	p622 P23, F432, I4132	p321, p312 P4132	P432	p622	F23	F432
C4			p4 P432		p422, p4212 I432, F432	I432	p422 P432		F432	P432
C6					p622	p622				
D2					p222, p622 F222, P4222, P6222, P4232, I4132	p622 P622, P4232, I4132	p422 P422, I422, I432	p622 P622	P23, F432, P4232	F432, I432
D3						p321 P312, P6322, P4232, P4132	I432	p622 P622	F4132	I432
D4							p422 P422, P432			P432
D6										
T									F23	F432
O										P432, F432

Finite assemblies (point group symmetries) are indicated in the blue font. 2-D layers are indicated in red, 3-D crystalline arrays in purple. In many cases, two component symmetries can be combined in different geometries giving rise to distinct symmetry types.

Gray boxes indicate symmetry combinations that are disallowed mathematically. A few symmetry combinations are not formally disallowed but are not amenable to design using compact building blocks.

Whenever a chiral space group appears (e.g. P4₁32), its enantiomer (e.g. P4₃21) is also possible but is not listed here for brevity.

[#] Additional possibilities exist but are not listed here for arrangements where more than two component symmetries are combined, or where one of the component symmetries is a screw axis of rotation.

[‡] Non-intersecting 2-fold axes give rise to linear or helical filaments. Linear assemblies, including rod groups, are not included here.

Figure 1A

A)

single contact type

linear or helical filaments

cyclic rings

two contact types

finite cages

extended materials in two and three dimensions

Figure 1B

Figure 2

A)

B)

Figure 3