Engineering highly ordered two-dimensional arrays and cages mediated by noncovalent protein-protein interfaces

Shmuel Shane Gonen

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2015

Reading Committee:

David Baker, Chair

Frank DiMaio

Ning Zheng

Program Authorized to Offer Degree:

Biochemistry

University of Washington

Abstract

Engineering highly ordered two-dimensional arrays and cages mediated by noncovalent protein-protein interfaces

Shmuel Shane Gonen

Chair of Supervisory Committee:

David Baker

Professor, Department of Biochemistry

Natural proteins often assemble into higher order structures by symmetric assembly of many copies of the same protein subunit through weak, non-covalent interaction to perform their tasks. Some examples of higher order structures include cages used for exocytosis, fibers used for structural stability and channels used for the flow of ions and water in and out of the cell. Two-dimensional (2D) protein assemblies also occur in nature, either assembled in or on lipid membranes. These 2D assemblies are usually made for large-scale transport of ions or water or as a cellular barrier against antagonists. Assemblies in 2D have been a challenge to engineer through design in the past due to the complexity of the proteins and their propensity to misfold. Previous successful attempts used different avenues to create 2D arrays, for example using metal-mediated assembly

or by using fused proteins but there are no examples of 2D assemblies being designed to have non-covalent interaction in their assembly as similarly seen in nature. We aimed to design such protein assemblies in order to allow for new avenues in biosensing, atomic-scale repeat patterning, structure determination and drug delivery. My thesis describes the first successful design of 2D assemblies using non-covalent interactions in different types of 2D crystal space groups (layer groups) as well as of similarly designed tetrahedral and icosahedral protein cages.

Table of Contents

List of Figures **7**

List of Tables **11**

Acknowledgements **12**

Dedication **14**

Chapter I – Introduction and Methodology **15**

References to Chapter I **32**

Preface to Chapter II **35**

Chapter II – Design of 2D arrays **37**

Appendix I – 2D extended Methods and Results **48**

References to Chapter II and Appendix I **68**

Appendix II – Additional Images of 2D arrays **73**

Appendix III - Symmetry Definition Files, run and XML scripts **84**

Preface to Chapter III **103**

Chapter III – 2D Fusion Arrays **104**

Preface to Chapter IV **111**

Chapter IV – Two-component Tetrahedral Cages **113**

References to Chapter IV **141**

Appendix IV – Additional class averages **145**

Preface to Chapter V 146

Chapter V – One- and Two-component Icosahedral Cages **..... 147**

Appendix V - Biochemical Protocols, Reagents and Recipes **..... 170**

Covers and Posters..... 174

Vita 177

List of Figures

Chapter I - Introduction and Methodology

Fig. 1. Clathrin Cages **23**

Fig. 2. TEM of Microtubules with Dam1 Attachments **24**

Fig. 3. Freeze-fracture of AQP4 channels **25**

Fig. 4. S-layer protein arrays **26**

Fig. 5. Aquaporin 0 2D crystals **27**

Fig. 6. The 17 2D Crystal Space Groups **28**

Fig. 7. Copper grids for TEM **29**

Fig. 8. Cryo-EM grid preparation **30**

Chapter II - Design of 2D arrays

Fig. 1. Computational design strategy and experimental analysis of arrays **45**

Fig. 2. Cryo-EM analysis of design p3Z_42 ..... **47**

Appendix I - 2D extended Methods and Results

Fig. S1. Design p3Z_11 ..... **60**

Fig. S2. In-plane views of p3Z_42, p4Z_9 and p6_9H **61**

Fig. S3. SDS-PAGE gel of p3Z_42, p4Z_9, p6_9H and p3Z_11 expression **62**

Fig. S4. In-vitro array formation of p3Z_42, p4Z_9 and p6_9H designs **63**

Fig. S5. Mutagenesis of p6_9 (precursor to p6_9H) **65**

Chapter III - 2D Fusion Arrays

Fig. 1. Schematic for fused 2D arrays **107**

Fig. 2. Spycatcher fused onto p3Z_42 TEM..... **108**

Fig. 3. Low-Magnification micrograph of CyaY protein fusion **109**

Fig. 4. TDRD2 protein fused to p3Z_42..... **110**

Chapter IV - Two-component Tetrahedral Cages

Fig. 1. Overview of the computational design method..... **125**

Fig. 2. Experimental characterization of coassembly..... **126**

Fig. 3. Modeled interfaces of designed two-component protein nanomaterials..... **127**

Fig. 4. Electron micrographs of designed two-component protein nanomaterials..... **129**

Fig. 5. Crystal structures of designed two-component protein nanomaterials **131**

Extended Data Figure 1. Comparison of one-component and multi-component symmetric fold trees..... **133**

Extended Data Figure 2. Models of the 57 designs selected for experimental characterization..... **134**

Extended Data Figure 3. Native PAGE analysis of cleared cell lysates **136**

Extended Data Figure 4. Structural metrics for the computational design models..... **138**

Extended Data Figure 5. Electron Micrographs of in-vitro assembled T33-15 (unpurified) and T33-15A and T33-15B in isolation..... **139**

Chapter V - One- and Two-component Icosahedral Cages

Fig. 1. Negative-stain TEM of empty I3-01 cages **149**

Fig. 2. Initial Cryo-EM image of I3-01 empty cages **150**

Fig. 3. Empty I3-01 cages imaged by cryo-EM after optimized sample and grid preparation **151**

Fig. 4. Cryo-EM averages of empty I3-01..... **152**

Fig. 5. Initial density model of empty I3-01 cage..... **153**

Fig. 6. Montage of Cryo-EM and averaging results of empty I3-01 ..... **154**

Fig. 7. Negative-stain TEM of I3-01-superfolderGFP cages **155**

Fig. 8. Initial interpretation of GFP density compared to empty I3-01 ..... **156**

Fig. 9. Initial I3-01-superfolderGFP cryo-EM micrograph..... **157**

Fig. 10. Motion-corrected movie of superfolder-GFP fused I3-01 ..... **158**

Fig. 11. Class averages of superfolder-GFP fused I3-01..... **159**

Fig. 12. Icosahedral design I32-06 overview..... **161**

Fig. 13. Icosahedral design I32-28 overview..... **162**

Fig. 14. Icosahedral design I52-03 overview..... **163**

Fig. 15. Icosahedral design I52-33 overview..... **164**

Fig. 16. Icosahedral design I53-34 overview..... **165**

Fig. 17. Icosahedral design I53-40 overview..... **166**

Fig. 18. Icosahedral design I53-47 overview..... **167**

Fig. 19. Icosahedral design I53-50 overview..... **168**

Fig. 20. Overview of two-component Icosahedral cages **169**

List of Tables

Chapter I - Introduction and Methodology

Table 1. The 17 possible space groups in 2D crystals **31**

Appendix I - 2D extended Methods and Results

Table S1. Mutagenesis of p6_9 design (precursor to p6_9H)..... **66**

Table S2. Final rosettascripts filter scores table for p3Z_11, p3Z_42, p4Z_9 and p6_9H designs..... **67**

Chapter IV - Two-component Tetrahedral Cages

Extended Data Table 1. Root mean square deviations (r.m.s.d.) between crystal structures and design models **140**

Acknowledgements

I would like to give my sincerest thanks to my advisor, David Baker for his mentorship, his contagious excitement for science and for introducing me to an amazing place to learn protein design. Through this, I have been able to collaborate with many outstanding people and those experiences have only increased my love for science.

I would like to thank Frank DiMaio for not only his collaboration on the 2D design project, but also for always having his door open to questions and advice.

Both members of the Baker and Gonen groups have been fantastic and have offered valuable support throughout my years in graduate school. I want to thank Jacob Bale and Neil King for their advice during my rotation and their collaborations along with Yang Hsia. I also want to thank Dan Shi for his ever impressive TEM knowledge and help and Steve Reichow and Silvia Sanchez-Martinez for their help around the lab.

I would like to thank the people in the biochemistry department and those running the BMSD/BPSD program for taking care of my student life behind the scenes. I would like to also extend my thanks to HHMI and the Janelia Research Campus for providing valuable resources for my projects and to the people there for their enthusiasm and support, including Ulrike Heberlein and Ed McCleskey.

I would like to thank Gabriele Varani and Ning Zheng for their guidance and support during my rotations and to them and their respective lab members for welcoming me so kindly into their groups.

My thesis committee members: David Baker, Frank DiMaio, Ron Stenkamp, Carlos Catalano, Ning Zheng, Gabriele Varani, John Scott and Bill Zagotta have been tremendous in their enthusiasm for my projects. I would like to thank them for their advice, their guidance and for giving up their time to make my many meetings and exams.

I also want to thank my fellow Biochemistry and BMSD/BPSD students of my incoming year, including Sergey Ovchinnikov. It was easy and fun to be around them and they have made graduate school life in and out of the lab and classroom even more enjoyable for it.

I mostly want to thank my family, as I would not be graduating without their love and encouragement. My brother, Tamir, has guided my graduate studies beyond measure. He has helped with difficult situations, contributed highly through his vast scientific knowledge, listened to my many ridiculous science ideas and questions all while giving up his time and resources. My sister, Carmit, has always been supportive and has encouraged what I wanted to do. My parents, Aviva and Yehezkel are the reason I was able to get to this point in my life and it is to them that I dedicate this thesis.

Dedication

I dedicate this thesis to my loving parents, my mother Aviva and the real engineer in the family, my father Yehezkel.

Chapter I

Introduction and Methodology

Introduction

Proteins in nature are microscopic workhorses, responsible for all manner of cellular processes. Each protein is geared towards it's own specific task but some are involved in performing function as part of large machinery made up of many proteins. These protein assemblies are often formed from many copies of the same protein oligomer. They have been observed to form assemblies of vast cage-like structures (Fig.1) (*1*), large fibrous material (Fig.2) (*2*), porous channels (Fig. 3) (*3*), two-dimensional (2D) crystals (Fig. 4) (*4*) and many other kinds of nanomaterials. 2D protein arrays occur in nature either in or on a membrane bilayer. Aquaporins have been shown to make square arrays through tetrameric proteins and are used to shuttle water in and out of cells (Fig. 3) (*3*). S-layer proteins form different types of arrays and are found on the cell surface as a barrier to foreign material and for structural stability (Fig. 4) (*4*). Closer inspection of these kinds of nanomaterials usually reveals a simplified mechanism of assembly where the subunits assemble symmetrically, thus preserving the same structure and interacting residues of the monomeric subunit allowing for rapid expression, assembly and response. Their assembly usually occurs using many non-covalent van der Waals associations and hydrogen bonding. 2D crystals can be formed in-vitro by reconstituting membrane proteins in a lipid bilayer (Fig. 5) (*5*). This process is limited to membrane proteins and is very laborious and can take years. By harnessing the natural characteristics of protein materials, we aimed to use new methods in protein design to

create new nanomaterials that we hope will usher in new techniques and therapeutic strategies. Proteins are highly customizable and the successful design of both 2D arrays and cages will result in scaffolds onto which newly engineered functional groups can be added. Designed 2D arrays should enable new methods in protein structure determination, whereby unknown proteins are fused to the arrays. Other avenues for 2D material include patterning of fluorescent matter to be used as an atomic scale marker and for biosensing, amongst others. Protein cages could offer new therapeutic strategies for drug delivery whereby a drug is targeted to a specific location in the body and a triggered release into the cell results in delivery without exposing the body to the drug along the way. This method may even be extended to include the blood-brain barrier.

Methodology

Protein Design and Rosetta:

The field of protein design has matured over the last few years. Proteins not found in nature have been successfully designed (*6*), along with new drug binders (*7*) and recently, 3D crystalline nanomaterials (*8*). The design of protein assemblies has traditionally been a challenge due to the complicated chemistry of proteins and their propensity to misfold and aggregate when non-favorable changes are made. Structures have been designed previously using various methods, for example, using fusion partners (*9*), metal-mediated assembly (*10*) or by forgoing proteins and only using DNA (*11*), but two-dimensional protein arrays only having noncovalent interactions have never been designed.

Rosetta is a protein design and folding prediction software developed in the Baker lab and is being improved every year by people both in the Baker lab and in other groups around the world. It has been used to design proteins such as those described above and more recently to refine Cryo-EM electron density models (*12*) and for the projects described in this thesis in the creation of new macromolecular assemblies.

Rosetta uses many algorithms and libraries for the rapid sampling of protein backbone torsion angles, side-chain rotomeric conformations, hydrogen bond networks and the docking of separate protein chains while estimating buried and surface exposed residues, shape complementarity and the difference in energy between bound and unbound states, to name a few. These libraries are made possible in part to statistics derived from solved protein structures and the strains allowed on amino acid bonds (described by the Ramashandran plot).

The aim of my thesis study was to create and structurally characterize new protein nanomaterials, specifically two-dimensional arrays and cages, using Rosetta. The strength of Rosetta comes from its customizability whereby the general sampling of the protein backbone and side-chains described above can be tailored to a specific question using constraints. We extended this to both the design of 2D arrays and tetrahedral and icosahedral cages. There are only a certain number of possible ways for proteins to interact with one another in a symmetrical way to create these types of nanomaterials and those properties have been mathematically calculated. There are 17 possible ways to pattern proteins in 2D (Fig. 6 and Table. 1) (*13*) and tetrahedrons and icosahedrons can be formed resulting in different symmetrical axes.

Rosettascripts:

Rosettascripts is an XML-script based method to run rosetta design by using a user-customized XML with differing movers, filters and task operations tailored for a specific design problem. These include setting the sampling to be performed symmetrically. In the case of 2D arrays, constraints are given for all the rosetta sampling aligning with a particular 2D space group of choice, therefore any change that is made preserves the crystalline repeats. In the case of tetrahedral and icosahedral cages, all the sampling is constrained along each cage's symmetrical axes, again, preserving the properties of the cage.

The following chapters describe the methods involved in the design of specific 2D space groups and cages made of single and two-components using Rosettascripts. An example XML-script can be found in the Appendix III.

In-vivo assembly:

All the designs made in Rosettascripts that pass all the filters, refinement and manual design and inspection then need to be analyzed biochemically. Genes coding for the designed proteins are made and the proteins are expressed in the *E. coli* expression system. Since the proteins are designed to form these nanomaterials, this should occur in-vivo. In order to structurally validate these designs, I turned to a powerful microscopic method called Transmission Electron Microscopy (TEM).

Transmission Electron Microscopy:

TEM is a powerful microscopic tool whereby organic and/or inorganic matter can be imaged or diffracted by a beam of electrons under vacuum. Samples are frozen, adhered or fixed on small metallic grids (Fig. 7) and the image is relayed through a camera or direct detector. Proteins have been studied by TEM for decades and sometimes, high-resolution structural information has been obtained. This is possible by finding the optimal biochemical conditions, handling methods and grid preparation (Fig. 8) for the sample and by having proper alignment of the beam, microscope settings and data collection and analysis. TEM can be split into many different fields, for organic material these usually fall under four main disciplines, single-particle, electron crystallography, microed and tomography. This thesis will describe the use of both single particle and electron crystallography. Each of these techniques can be performed either by staining the sample – called negative-stain TEM, or without stain, usually under cryogenic temperatures by freezing the sample in liquid nitrogen or liquid ethane – called cryo-EM. The last few years have seen a proliferation of protein structures being determined to high resolution by TEM due to emerging camera technology and detectors allowing for better quality data to be obtained while minimizing the damage done to the sample due to radiation damage.

Negative-stain TEM:

There are many stains that are widely used for negative-stain experiments depending on availability and nature of the experiment. The stain grain size is an important factor in the way a sample will ultimately be seen by negative stain. The finer

the grain, the more features will be highlighted. Uranyl Acetate and Uranyl Formate provide the best grain size and are the most commonly used commercial stains. Uranyl Acetate, once made into a solution, lasts for a long time at room temperature without precipitating and has a slightly higher pH than Uranyl Formate. Uranyl Formate has the smallest grain size and is the stain used in all of the negative-stain experiments presented in this thesis. Copper grids are mostly used for these experiments and are coated with a thin layer of carbon to hold the protein material (Fig. 7). The sample is placed on these grids after the grids are charged in order to maximize the retention of the sample on the surface. After a few rounds of cleaning and blotting, the grid is finally stained and is ready to be viewed by TEM.

Negative-stain is a powerful method for screening samples as once the sample is stained, the grid can be viewed at anytime. The major benefit when screening is the additional contrast provided by the stain, allowing for fine details to be viewed without any need for data analysis and averaging. There are drawbacks to using negative-stain, mainly the loss of high-resolution information. As the stain envelops the sample, the resolution is limited by the grain size of the stain used. For Uranyl Formate, the highest resolution possible is not exactly known, but resolutions of ~14Å have been achieved in 2D projections of 2D crystals. Another drawback is flattening of the proteins by the stain. For most proteins this is a minor drawback but sometimes this can be more pronounced.

Cryo-EM:

Cryo-EM refers to the technique of freezing the sample and viewing without staining procedures. As stain is not used, Cryo-EM does not cause the loss of high-

resolution information. Usually, liquid ethane is used as that freezes the sample quickly resulting in vitreous, amorphous ice that has the least effect on the protein.

Electron Crystallography:

Electron crystallography is a technique whereby 2D protein crystals are either imaged or diffracted under either negative-stain or cryo-EM conditions. Either 2D projections or 3D structures can be obtained using this method, the latter of which requires the crystal to be imaged/diffracted multiple times at different angles.

Single-particle TEM:

Single-particle TEM is a technique whereby many single proteins from images are separated into classes and averaged in an attempt to create an overall projection image of the sample from various angles. These projection averages are then used to create a density model and can be refined into a 3D structure. Micrographs with well-dispersed particles are required to make class averages, so particles that are prone to aggregation are not useful. Good sample and grid prep resulting in grids with mono-dispersed particles is required. Once these grids are obtained, enough image data needs to be collected because many particles are required to be averaged into separate classes. Each class should ultimately have many particles in order to be considered as a good representation of the particle and many classes are required to sample the entirety of the protein. Therefore, enough data needs to be collected and in cryo-EM, the number of particles needs to be significantly higher due to the low contrast of cryo. For symmetrical particles, that number can be reduced. Classes that have only a few particles are not

representative of the protein and picked particles that are broken/aggregated or too close to other proteins will be averaged and will result in a hazy area of the final average that is not optimal. Once good averages are obtained, the angles of the protein are estimated from the 2D averages to generate 3D initial models. Initial models are then used with the particle sets to refine into a higher resolution structure.

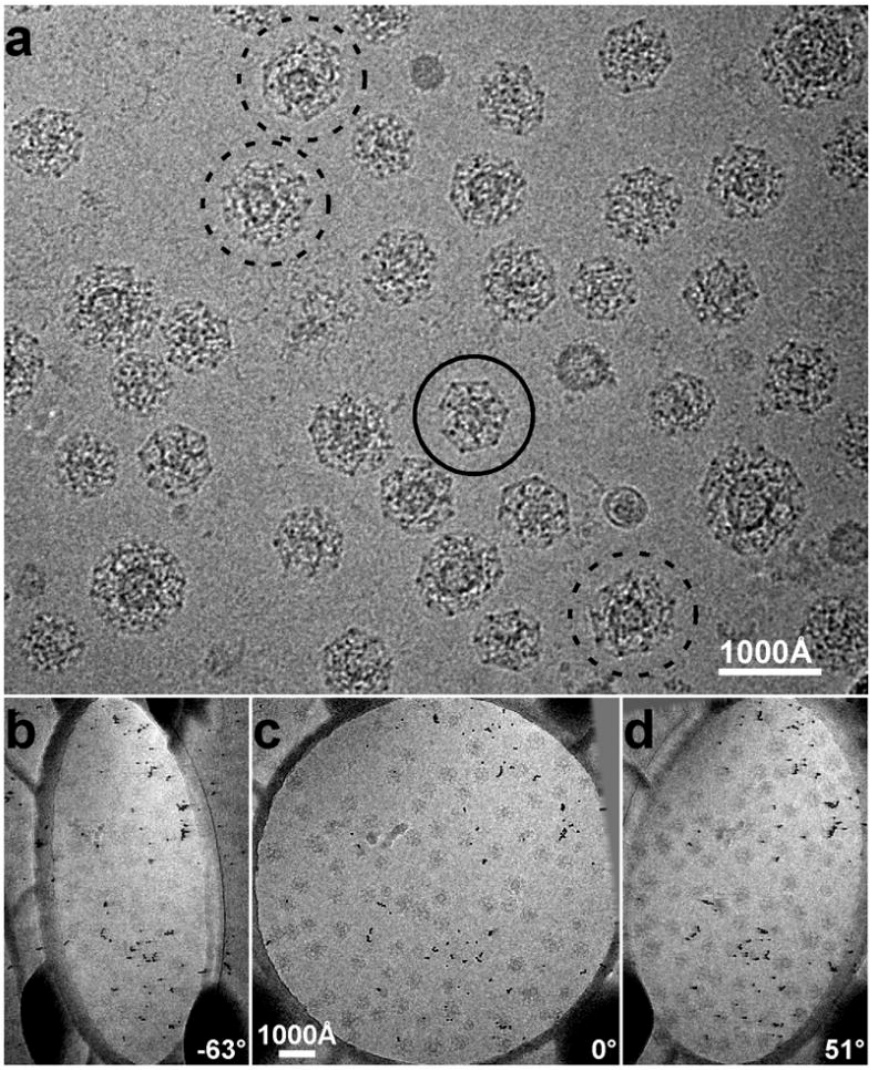


Figure 1. (Adapted from Ref. 1). Clathrin Cages.

(A-D) TEM micrographs of Clathrin cages showing their size and symmetry.

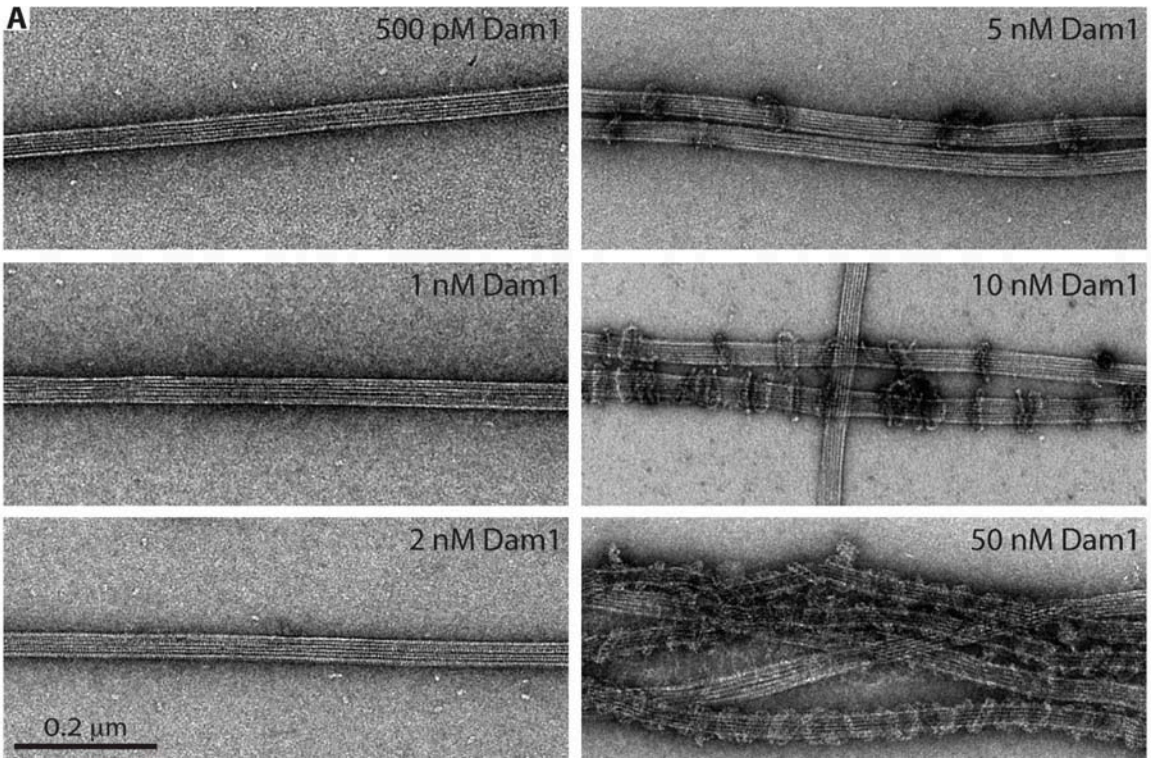


Figure 2. (Adapted from Ref. 2). TEM of Microtubules with Dam1 attachment. TEM micrographs showing taxol-stabilized microtubules and increasing concentrations of a protein involved in cell division, Dam1, forming rings.

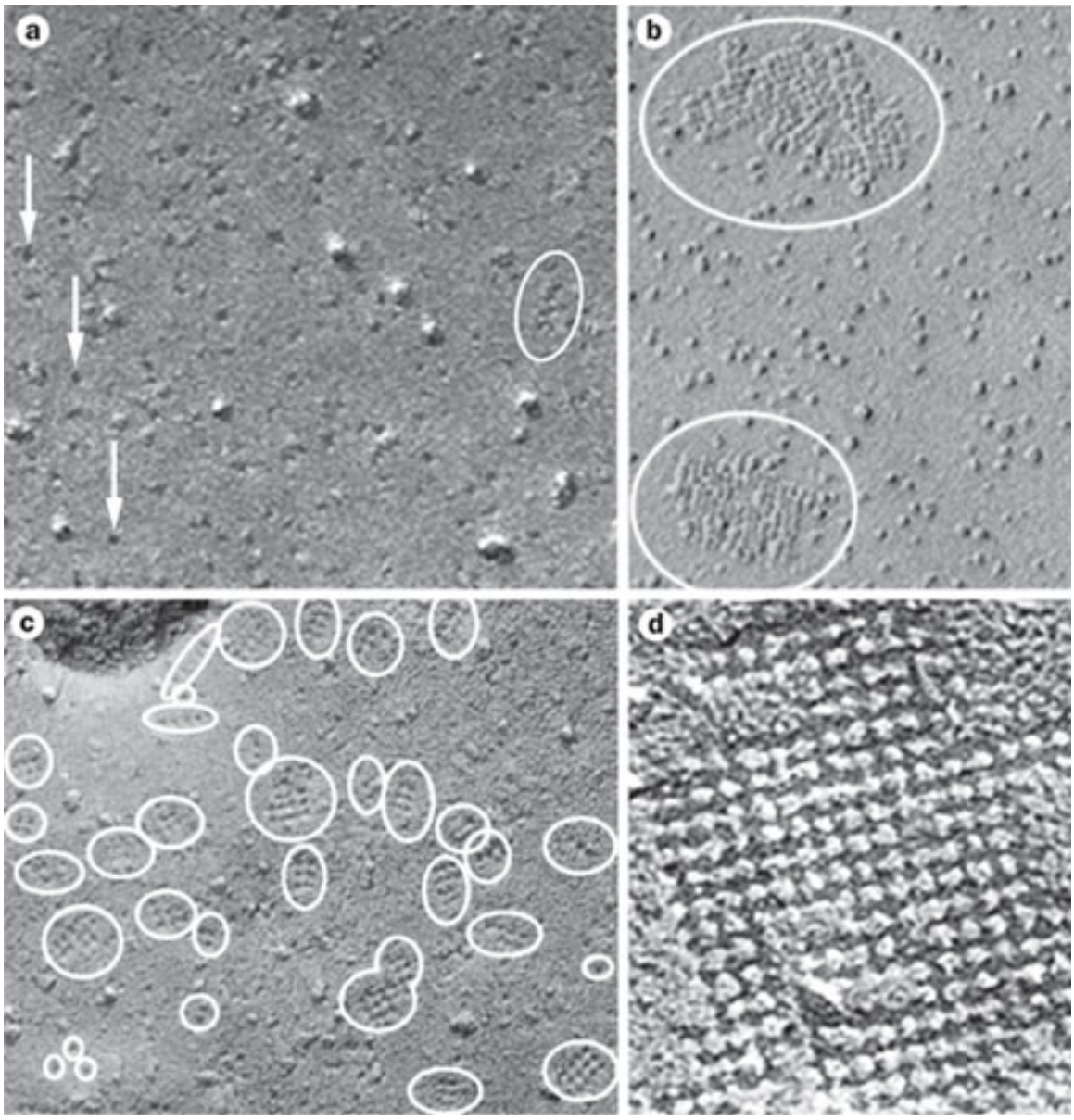


Figure 3. (Adapted from Ref. 4) Freeze-fracture of AQP4 channels.

(A-D) Freeze-fracture of Aquaporin-4 single channels (background and arrows) and crystalline patches of many channels (circled).

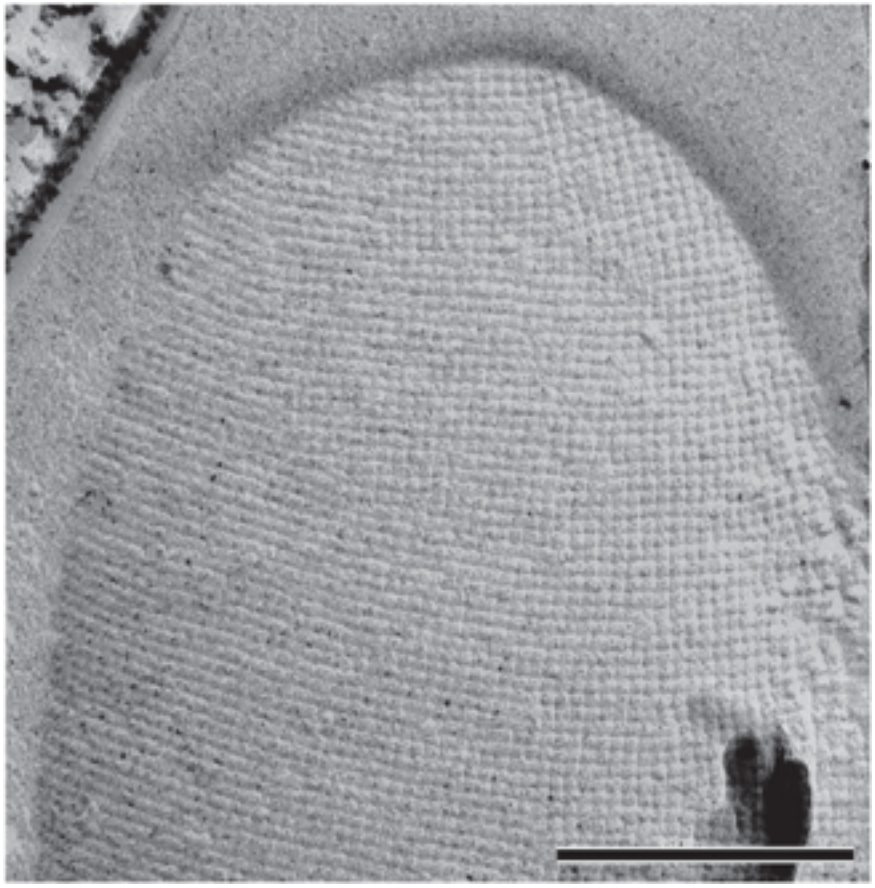


Figure 4. (Adapted from Ref. 4) S-layer protein array.

S-layer protein array can be seen on the surface of an *E. coli* cell showing a square pattern.

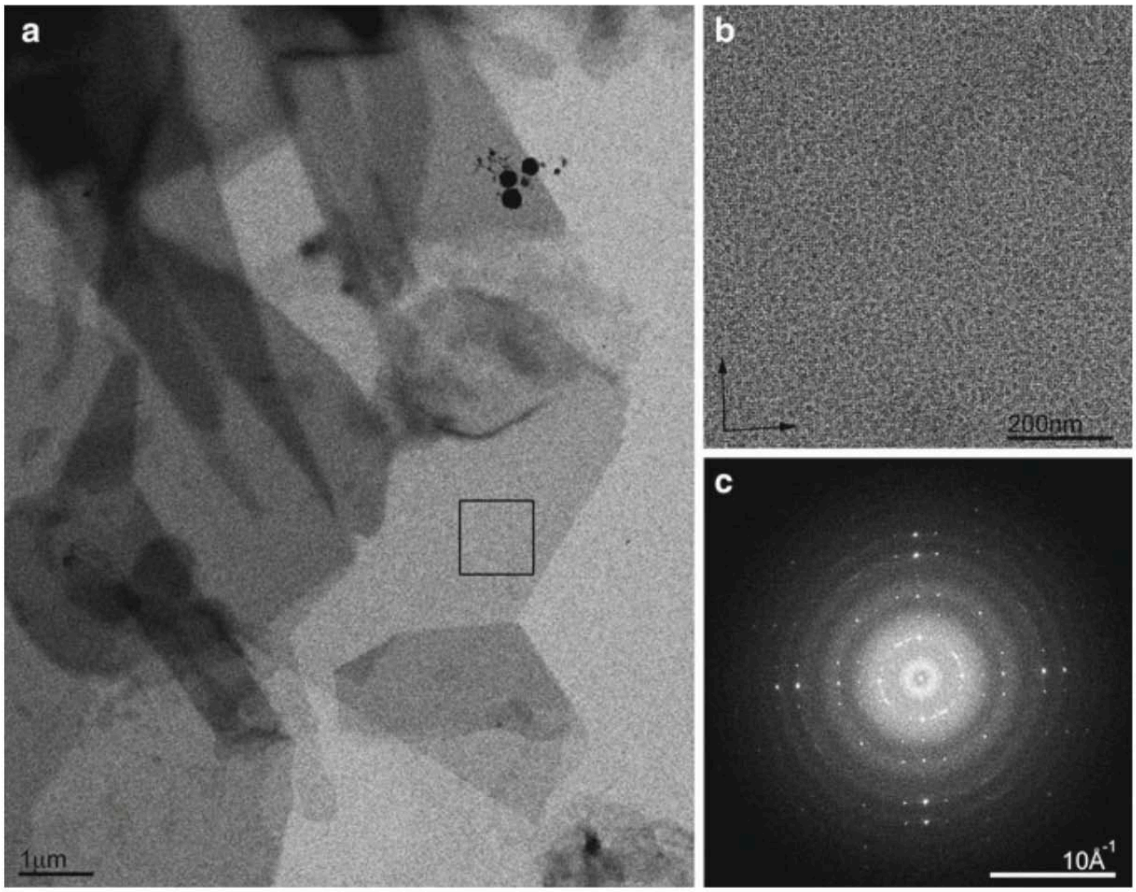


Figure 5. (Adapted from Ref. 5). Membrane proteins reconstituted in a lipid bilayer forming 2D crystals in-vitro.

2D crystals of Aquaporin-0 were formed after purification of the protein and slow dialysis in the presence of lipids. (A) Image of the 2D crystals under cryo-EM. (B-C) Close up view and FFT of the lattice.

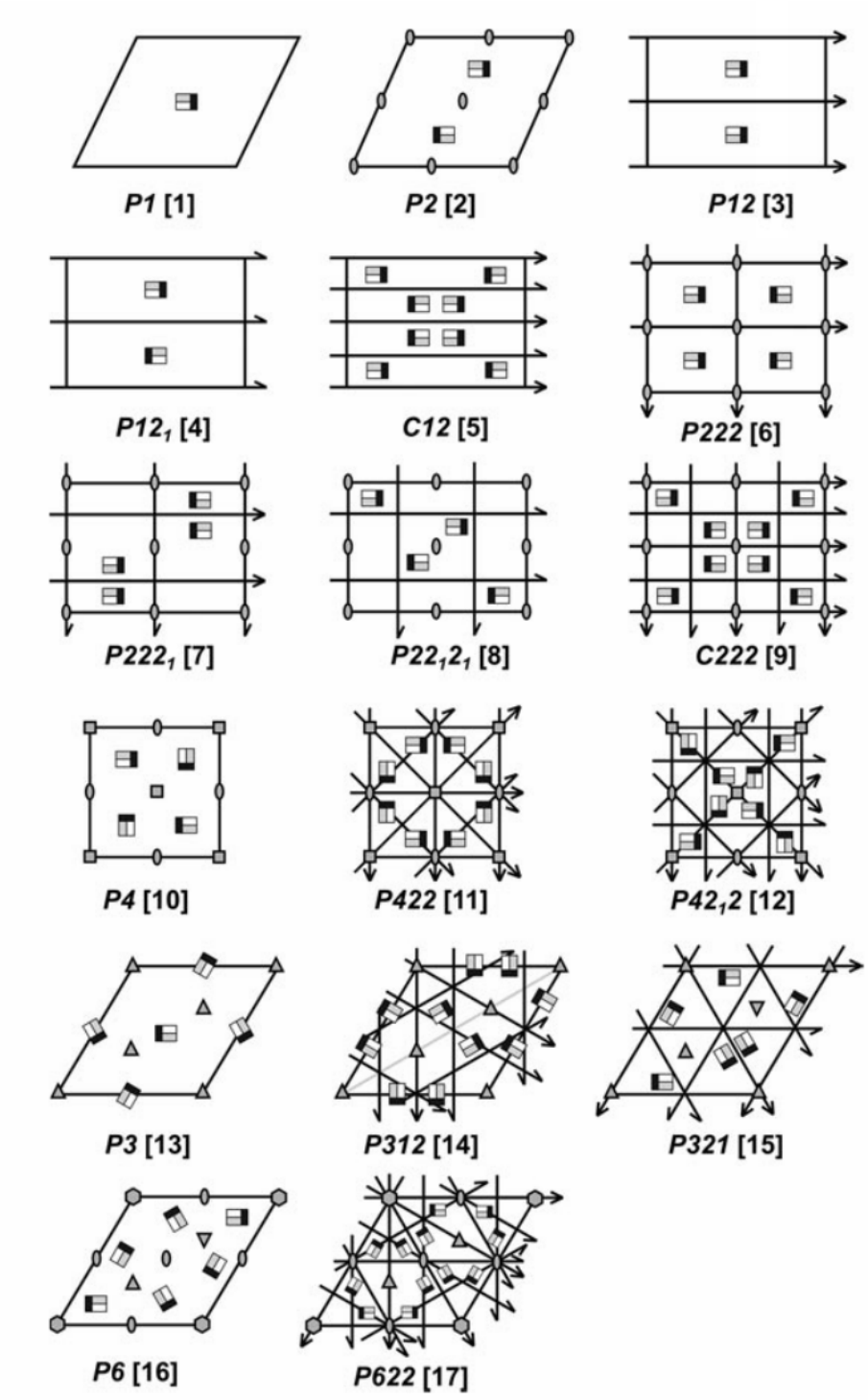


Figure 6. (Adapted from Ref. 13). The 17 2D crystal space groups (layer groups) and their symmetrical constraints.

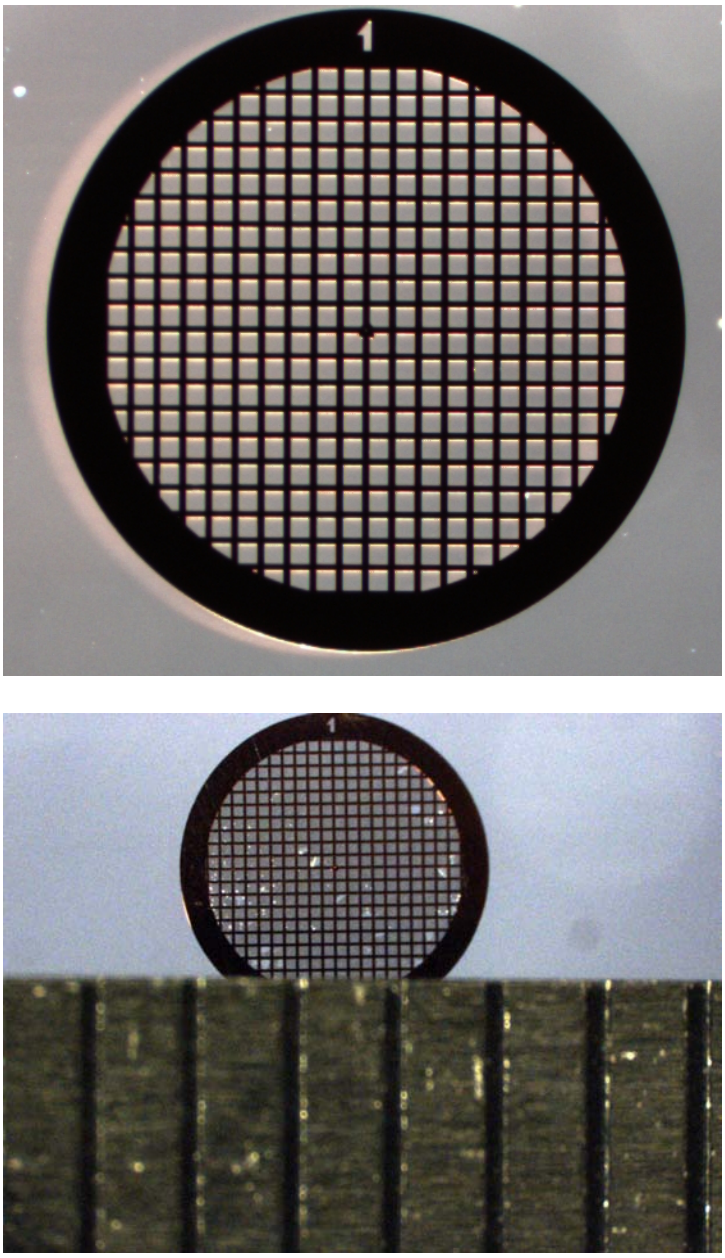


Figure 7. Copper grids for TEM

Copper grids (*Ted Pella Inc.*) with varying amounts of separated hollow windows (top) are coated with a thin layer of carbon (bottom) and measure approximately 3mm. The thin layer of carbon creates a surface where the sample is placed for TEM and "broken" windows can be seen where the carbon has peeled off the grid.

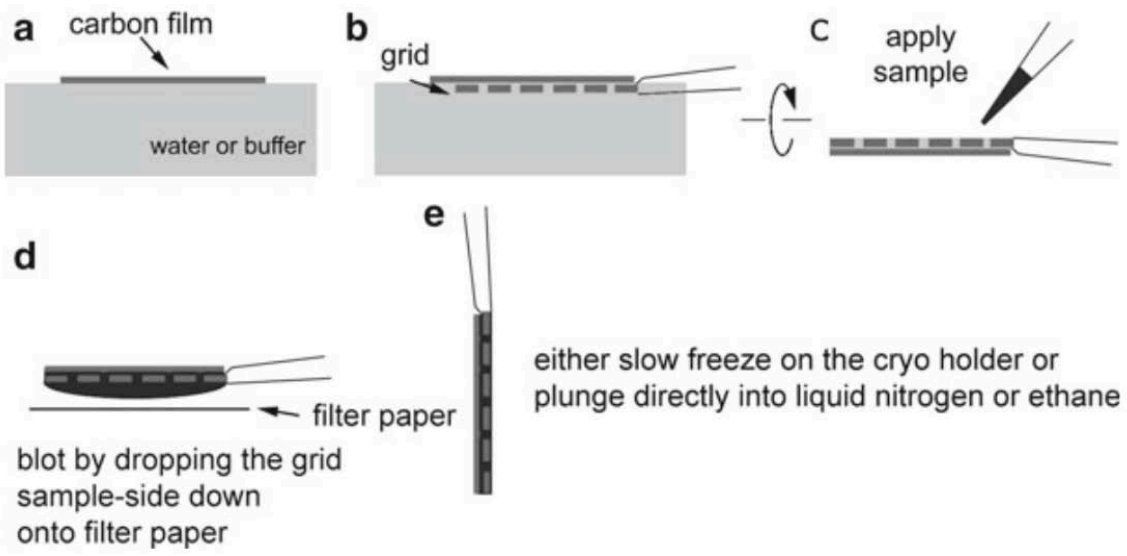


Figure 8. (Adapted from Ref. 14). Some grid preparation and blotting procedures for negative-stain and cryo-EM.

Table 1. The 17 possible space groups in 2D crystals.

Plane Group	2D space group	2D space group number	Crystal system
p1	P1	1	Oblique
p2	P2	2	Oblique
pm	P12	3	Rectangle
pg	$P12_1$	4	Rectangle
cm	C12	5	Rectangle
p2mm	P222	6	Rectangle
p2mg	$P222_1$	7	Rectangle
p2gg	$P22_12_1$	8	Rectangle
c2mm	C222	9	Rectangle
p4	P4	10	Square
p4mm	P422	11	Square
p4gm	$P42_12$	12	Square
p3	P3	13	Hexagonal
p3m1	P321	14	Hexagonal
p31m	P312	15	Hexagonal
p6	P6	16	Hexagonal
p6mm	P622	17	Hexagonal

(Adapted from Ref. 13)

References highlighted in Chapter I

1. Y. Cheng, W. Boll, T. Kirchhausen, S. C. Harrison, T. Walz, Cryo-electron tomography of clathrin-coated vesicles: structural implications for coat assembly., *J. Mol. Biol.* **365**, 892–9 (2007).

2. J. F. Tien *et al.*, Cooperation of the Dam1 and Ndc80 kinetochore complexes enhances microtubule coupling and is regulated by aurora B., *J. Cell Biol.***189**, 713–23 (2010).

3. S. Jarius, B. Wildemann, AQP4 antibodies in neuromyelitis optica: diagnostic and pathogenetic relevance., *Nat Rev Neurol* **6**, 383–92 (2010).

4. D. Pum, U. B. Sleytr, Reassembly of S-layer proteins., *Nanotechnology* **25**, 312001 (2014).

5. T. Gonen, The collection of high-resolution electron diffraction data., *Methods Mol. Biol.* **955**, 153–69 (2013).

6. N. Koga *et al.*, Principles for designing ideal protein structures., *Nature* **491**, 222–7 (2012).

7. E. Procko *et al.*, A computationally designed inhibitor of an Epstein-Barr viral Bcl-2 protein induces apoptosis in infected cells., *Cell* **157**, 1644–56 (2014).

8. C. J. Lanci *et al.*, Computational design of a protein crystal. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 7304-7309 (2012).

9. J. C. Sinclair, K. M. Davies, C. Venien-Bryan, M. E. Noble, Generation of protein lattices by fusing proteins with matching rotational symmetry. *Nature nanotechnology* **6**, 558-562 (2011).

10. A. Laganowsky *et al.*, An approach to crystallizing proteins by metal-mediated synthetic symmetrization., *Protein Sci.* **20**, 1876–90 (2011).

11. A. Rajendran, M. Endo, K. Hidaka, H. Sugiyama, Control of the two-dimensional crystallization of DNA origami with various loop arrangements.,*Chem. Commun. (Camb.)* **49**, 686–8 (2013).

12. R. Y. Wang *et al.*, De novo protein structure determination from near-atomic-resolution cryo-EM maps., *Nat. Methods* **12**, 335–8 (2015).

13. B. L. Nannenga, M. G. Iadanza, B. S. Vollmar, T. Gonen, Overview of electron crystallography of membrane proteins: crystallization and screening strategies using negative stain electron microscopy.,*Curr Protoc Protein Sci* **Chapter 17**, Unit17.15 (2013).

14. V. Cabra, M. Samsó, Do’s and don'ts of cryo-electron microscopy: a primer on sample preparation and high quality data collection for macromolecular 3D reconstruction., *J Vis Exp* , **52311** (2015).

Preface to Chapter II

Chapter II describes my main graduate project on the design and characterization of two-dimensional protein arrays. Protein crystal design is in it's infancy when compared to those developed with DNA. We wanted to develop highly customizable 2D lattices and proteins are the perfect choice. The work in this chapter describes the first case of designed 2D protein lattices held together to high order solely by noncovalent interactions, diffracting to high resolutions and not involving a lipid bilayer for support. Traditionally, 2D protein lattices are meticulously developed over (sometimes) several years from membrane proteins for high-resolution structural determination by electron crystallography. The main objective to the work described in this chapter was to see if we could use a designed array to coax proteins to crystallize by fusion to use for electron crystallography (described in chapter III). Other potential applications include patterning fluorescent proteins at an atomic level for use as atomic scale rulers, membrane binding and curvature and engulfing toxic cells.

Acknowledgements: We thank D. Shi and J. de la Cruz for help with electron microscopy, S. Sanchez-Martinez for help with protein expression, J. Bale and N. King for support and helpful discussions and W. Sheffler for rosetta code. We would also like to thank members of both the Baker and Gonen labs for scripts and useful discussions. We thank HHMI's Janelia Research Campus visitor program, the University of Washington's Biochemistry Department and Biological Physics, Structure and Design

program, DTRA, AFOSR and the Howard Hughes Medical Institute for Funding and Support.

The work described in Chapter II was published in the journal of *Science* as:

S. Gonen, F. DiMaio, T. Gonen, D. Baker. Design of ordered two-dimensional arrays mediated by noncovalent protein-protein interfaces. *Science* 348, 1365–8 (2015).

Chapter II

Design of ordered two-dimensional arrays mediated by noncovalent protein-protein interfaces

Abstract:

We describe a general approach to designing two-dimensional protein arrays mediated by non-covalent protein-protein interfaces. Protein homo-oligomers are placed into one of the seventeen 2D layer groups, the degrees of freedom of the lattice are sampled to identify configurations with shape complementary interacting surfaces, and the interaction energy is minimized using sequence design calculations. We use the method to design proteins that self assemble into layer groups P 3 2 1, P 4 2_1 2 and P 6. Projection maps of micron scale arrays assembled both in vitro and in vivo are consistent with the design models and display the target layer group symmetry. Such programmable 2D protein lattices should enable new approaches to structure determination, sensing, and nanomaterial engineering.

Introduction:

Programmed self-assembly provides a route to patterning matter at the atomic scale. DNA origami methods (*1, 2*) have been used to generate a wide variety of ordered structures, but progress in designing protein assemblies has been slower due to the greater complexity of protein-protein interactions. Biology provides a number of examples of ordered two dimensional protein arrays: bacterial S-layer proteins assemble into trigonal, square or hexagonal planar symmetry (*3*)*;* gap junction plaques abundant in muscle and

heart tissue display hexagonal planar symmetry (*4*); and water channels display square planar symmetry (*5*). While proteins which form ordered three-dimensional (3D) crystals have been designed (*6*), and two-dimensional lattices have been generated by genetically fusing or chemically crosslinking oligomers with appropriate point symmetric groups (*7-10*), there has been little success in designing self-assembling two-dimensional lattices with order sufficient to diffract electrons or x-rays below 15Å resolution (*5*). Naturally occurring two-dimensional arrays and assemblies are stabilized by extensive non-covalent interactions between protein subunits (*10, 11*), and this principle has been used to design self-assembling tetrahedral and octahedral cages (*12, 13*).

We sought to design ordered two-dimensional arrays mediated by designed protein-protein interfaces stabilized by extensive non-covalent interactions. We focused on symmetric arrays, as symmetry reduces the number of distinct protein interfaces required to stabilize the lattice (*14, 15*). There are seventeen distinct ways (*layer groups*) in which three-dimensional objects can come together to form periodic two-dimensional layers (*16*). In some layer groups there are only two unique interfaces between identical subunits, in others, three or four (*17*). To simplify the design challenge, we focused on the layer groups involving only two unique interfaces, and building blocks with internal point symmetry (which already contain one of the two required interfaces), leaving only one unique interface to be designed to form the two-dimensional array. Eleven of the seventeen layer groups have two unique interfaces; we focused here on six of these eleven groups involving cyclic rather than dihedral point groups because there are considerably more cyclic oligomers than dihedral oligomers in the Protein Data Bank (PDB) that can serve as building blocks. The six layer groups with two unique interfaces

that can be built from cyclic oligomers are P 2 2_1 2_1 (from C2 building blocks), P 3 and P 3 2 1 (from C3 building blocks), P 4 and P 4 2_1 2 (from C4 building blocks), and P 6 (from C6 building blocks). The different groups have different numbers of degrees of freedom describing the placement of an object with cyclic symmetry in the lattice, for example for P 3 2 1 (Fig. 1a) and P 4 2_1 2 (Fig. 1f), there are three degrees of freedom (A,θ and Z) while for P 6 (Fig. 1k) there are only two (A,θ).

Methods:

We used symmetric docking in Rosetta (*14, 18, 19*) to search for placements of cyclic oligomers into each of the six layer groups with shape complementary (*20*) interfaces between different oligomer copies. The docking scoring function consisted of a soft sphere model of steric interactions and a simple measure of the designable interface area -- the number of interface Cβs within 7Å. For each cyclic oligomer in each layer group, ~20 independent Monte Carlo docking trajectories were carried out starting from placements of 6-9 copies of the oligomer with its symmetry axis aligned with the corresponding symmetry axes of the layer group (for example, trimers were placed on the three-fold symmetry axes indicated by the triangles in Fig. 1a, tetramers on the four-fold symmetry axes indicated by squares in Fig. 1f, and hexamers on the six-fold symmetry axes indicated by hexagons in Fig. 1k). In the Monte Carlo docking simulations, the degrees of freedom sampled were those compatible with the layer group (Fig. 1a,f, and k right), and hence the layer group symmetry was preserved throughout the calculations.

We then selected the most shape complementary (largest number of contacting residues with fewest clashes) solutions from the trajectories and carried out Rosetta

sequence design calculations to generate well packed low energy interfaces between oligomers. Monte Carlo searches were carried out over all amino acid identities and side chain rotamer states for residues near the newly formed interface between oligomers optimizing the Rosetta all atom energy of the entire complex (*12, 13, 21*). Following this sequence design step, the energy was further minimized with respect to the side chain torsion angles of residues near the interface and the symmetric degrees of freedom of the layer group. Finally, the resulting lattice models were filtered based on the shape complementarity of the designed interface (>0.5), surface area of the designed interface (>400Å per monomer), buried unsatisfied hydrogen bonds introduced at the new interface (<4 using a 1.4Å solvent accessibility probe) (*22*), and predicted ΔΔG (*23*) of complex formation (<-10 Rosetta energy units per subunit). The filters were adjusted for each layer group such that approximately 200 designed sequences passed the filters (sample Rosettascripts files accompany the supplementary material). Following further sequence optimization (*13, 24*) models passing the filters were manually inspected, and 62 designs were selected for experimental characterization; 16 for P 2 2_1 2_1, 2 for P 3, 10 for P 3 2 1, 16 for P 4, 3 for P 4 2_1 2 and 15 for P 6.

Synthetic genes were obtained for the 62 designs, and the proteins were assayed for expression in the *E. coli* cytoplasm using a standard T7 based expression vector. 43 of the 62 designs expressed, of these 18 had protein in the supernatant after clearing the lysate at 12,000 x G for 30 minutes, while all 43 had protein in the pellet. To investigate the degree of order in the pelleted material, we examined negative stained samples by electron microscopy. Regular lattices were observed for four of the designs: one formed

only stacked 2D layers (Fig. S1) while three formed planar arrays. The latter are described in the following sections.

Results:

Design p3Z_42 is in layer group P 3 2 1. The rigid body arrangement of the constituent beta-helix trimers in the lattice was identified by Monte Carlo search over the three degrees of freedom of the lattice: the rotation of the trimer around its axis, the lattice spacing, and the z offset of the trimer from the lattice plane (Fig. 1a). In the lattice identified in the Monte Carlo docking calculations, the oligomeric building blocks pack into a dense array (Fig. 1b; the yellow and purple copies are inverted with respect to each other, side view Fig. S2a) stabilized by a large contact surface between adjacent copies with close complementary side chain packing (Fig. 1c) generated in the sequence design calculations.

p3Z_42 formed large and very well ordered 2D crystals (Fig. 1d). Very little of the protein produced in *E. coli* was found in the soluble fraction (Fig. S3), suggesting the vast majority of the expressed protein assembled into the crystalline arrays found in the pellet fraction. At low (16° Celsius) expression temperatures, 2D sheets were obtained (Fig. 1d), while at 37° Celsius, where larger amounts of proteins are produced, large 2D sheets mainly stacked into thick 3D crystals. Higher magnification (Fig. 1d, inset) showed a trigonal lattice similar to that of the design model (compare Fig. 1d (right) with 1b). Fourier transformation of the lattice (Fig. 1d (left)) yielded peaks out to 15Å resolution; the order in the unstained lattice is probably significantly higher as the negative stain likely limits the observed resolution. A 15Å projection map (Fig. 1e)

back-computed from the Fourier components followed the contour of the designed lattice (Fig. 1e (right)) (unit cell dimensions a=b=85Å, γ=120°). It is notable that planar crystals of such large size can grow without support within the confines (and with the many cellular obstacles) of an *E. coli* cell. Cell free expression of this design yielded large ordered 2D crystals similar to those formed in *E. coli* (Fig. S4a).

Design p4Z_9 is in layer group P 4 2_1 2. Search over the three degrees of freedom of the layer group (the rotation around the internal C4 axis, the lattice spacing, and the z offset between adjacent inverted tetramers (Fig. 1f)) yielded the close packed arrangement shown in Figure 1g (side view Fig. S2b). The designed interface is composed of hydrophobic residues nestled between two alpha helices surrounded by polar residues (Fig. 1h).

p4Z_9 formed crystals up to a micron in width (Fig. 1i) with little of the protein present in the soluble fraction (Fig. S3). Incubation of the pellet material with 6M guanidine and subsequent purification and refolding (by dialysis or fast dilution) yielded crystalline 2D arrays and fibers with the same square packing (Fig. S4b,c). Fourier transformation of the negatively stained large in vivo generated 2D lattices yielded peaks out to 14Å resolution (Fig. 1i (left)). The 14Å projection map produced by back transformation had distinctive rectangular voids in alternating directions closely matching the design model (Fig. 1j and 1j right) (unit cell dimensions a=b=56Å, γ=90°).

Design p6_9 is built from alpha helical hexamers in layer group P 6. In this case all oligomers are in the same orientation along the z-axis (perpendicular to the plane in Fig. 1k) and hence there are only two degrees of freedom -- the rotation around the six-fold axis and the lattice spacing (Fig. 1k (right)). The shape complementary docking

solution (Fig. 1l, side view Fig. S2c)) is composed of four closely associating alpha helices along the two-fold axis of the lattice (Fig. 1m) with two interacting phenylalanines. We also tested a variant, p6_9H, which introduces a hydrogen bond network across the interface (Fig. 1m).

Design p6_9 expressed in *E. coli* was found in both the supernatant and pellet (Fig. S3). EM investigation revealed that the pellet contained highly ordered single layer 2D hexagonal arrays while the supernatant did not. p6_9H formed even larger arrays (Fig 1n. Fig. S5 and Table S1). The 2D layers in the pellet were highly ordered with clearly evident hexagonal packing (Fig. 1n and 1n inset). Fourier transformation of the negatively stained arrays (Fig. 1n (left)) yielded peaks out to 14Å resolution; and the back-computed 14Å map was again closely consistent with the design model of the array (Fig. 1o and 1o right) (unit cell dimensions: a=b=120Å, γ=120°). Large arrays were also formed in vitro following concentration of soluble p6_9H purified from the supernatant after lysis of *E. coli* (Fig. S4d,e).

To achieve higher resolution than possible with negatively stained samples, we analyzed designs without stain by electron cryomicroscopy (cryo EM). Analysis of p3Z_42 crystals by cryo EM (Fig. 2a,b) and electron diffraction yielded data visible to 3.5Å resolution (Fig. 2c). The vast majority of crystals diffracted to this resolution in the cryo preparations indicating high long-range order. Movie micrographs of the resulting crystals were also collected, motion corrected and processed in 2dx (*25*) to yield a projection map at 4Å resolution in agreement with the design model (Fig. 2, compare panels d and e). To our knowledge, this is highest order observed for a designed macromolecular 2D lattice to date.

Discussion:

Our designed planar protein arrays form large planar 2D crystals both in vivo and in vitro that are closely consistent with the design models. Two of the three successes were with layer groups with adjacent building locks in opposite orientations along the z axis; these have the advantage that 1) there is an additional degree of freedom (the z offset) providing more possible packing arrangements for a given oligomeric building block, 2) the interfaces are antiparallel rather than parallel so that in the design calculations opposing residues can have different identities, and 3) inaccuracies in the design calculations that result in deviation from planarity effectively cancel out. On the other hand, designed "polar" arrays with all subunits orientated in the same direction--such as p6_9--have advantages for functionalization as the two sides are distinct and can be addressed separately.

It is notable that for all three designs extensive crystalline arrays form unsupported in *E. coli* and from purified protein in vitro. The arrays can extend up to a micron in length but are only 3-8nM thick by design (Fig. S2). We anticipate that even larger and perhaps more highly ordered crystals would form on a solid support, which will be useful for future nanotechnology applications. The ability to precisely design two dimensional arrays at the near atomic level should enable new approaches in structural biology (fusing proteins of unknown structure to array components for electron crystallography or using these as nucleation sites for 3D crystal growth for x-ray and MicroED (*26*) applications), new sensing modalities with the coupling of analyte binding domains to the arrays, and the organization of enzyme networks and light harvesting chromophores in two dimensions.

Figures

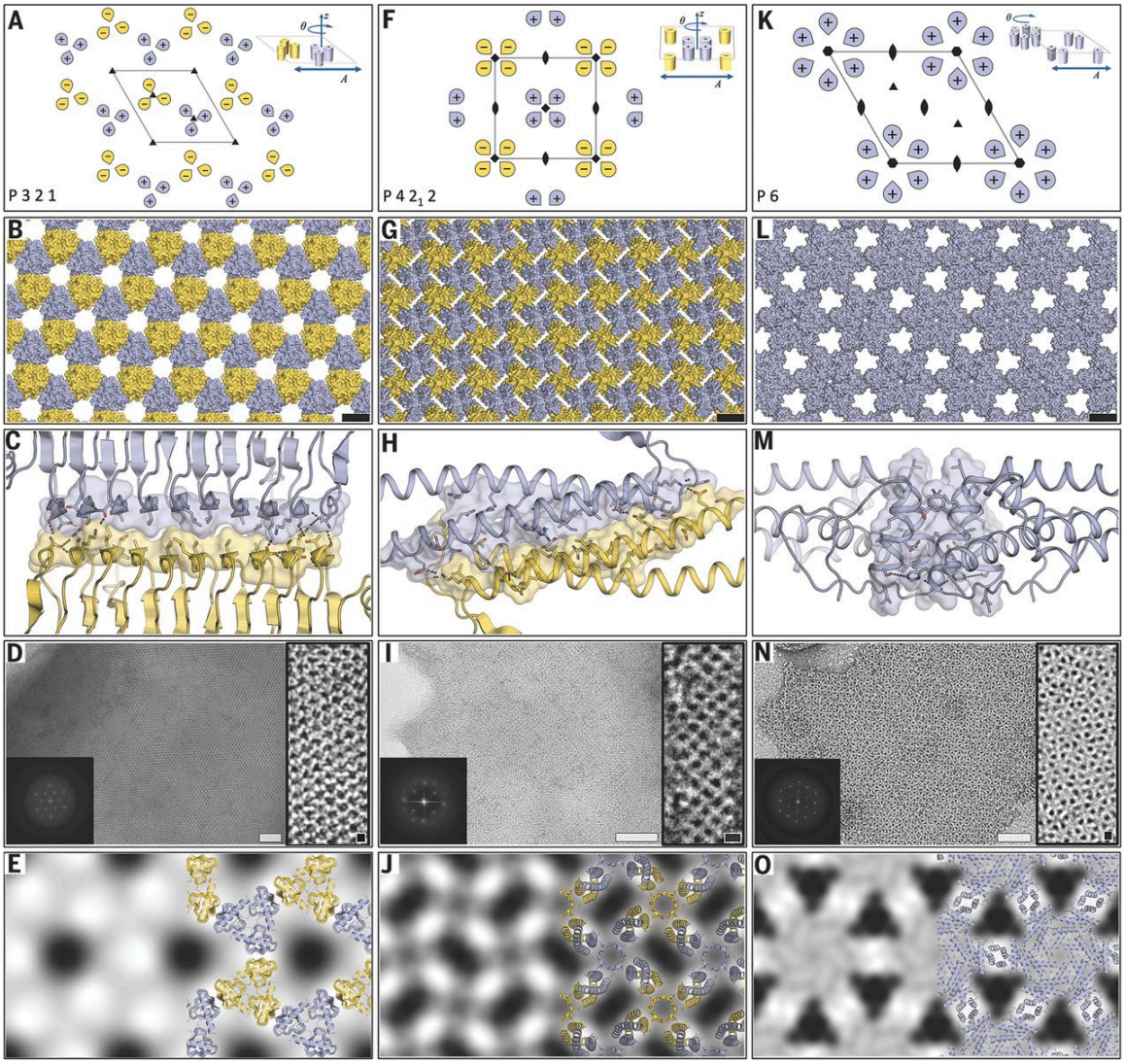


Fig. 1. Computational design strategy and experimental analysis of designed arrays. (**A**) The P 3 2 1 unit cell with three-fold axes represented by triangles. Yellow (-) and purple (+) C3 objects have opposite orientations along the z axis. Inset indicates the three degrees of freedom of the lattice. (**B**) p3Z_42 2D array. (**C**) p3Z_42 designed interface with "zipper-like" hydrophobic packing and peripheral hydrogen bonds. (**D**) Large (>1μm) *E. coli* grown array (middle), higher magnification view with lattice spacing as in (b) (right), and Fourier transform (amplitudes) of the large array (left). (**E**) Left: 15Å

projection map calculated from a large array. Right: overlay of the p3Z_42 design model on the projection map. (**F**) The P 4 2_1 2 lattice. Ovals represent two-fold axes and squares, four-fold axes. (**G**) p4Z_9 array. (**H**) p4Z_9 designed interface. (**I**) Negatively stained *E. coli* grown array (main panel), an in-vitro re-folded lattice at higher magnification (inset), and Fourier transform of the main panel (left) . (**J**) 14Å projection map calculated from an *E. coli* array as in (i) without (left) and with (right) p4Z_9 design model. (**K**) The P 6 lattice has two degrees of freedom (A, θ) (inset) available for sampling. Six-folds are represented by hexagons (**L**) p6_9H array. (**M**) p6_9H designed interface. (**N**) p6_9H lattice grown in vivo with Fourier transform at left and higher magnification view at right. (**O**) 14Å projection map of p6_9H from *E. coli* grown arrays as in (n) and cartoon overlay (right). All scale bars: Black = 5nm, White = 50nm.

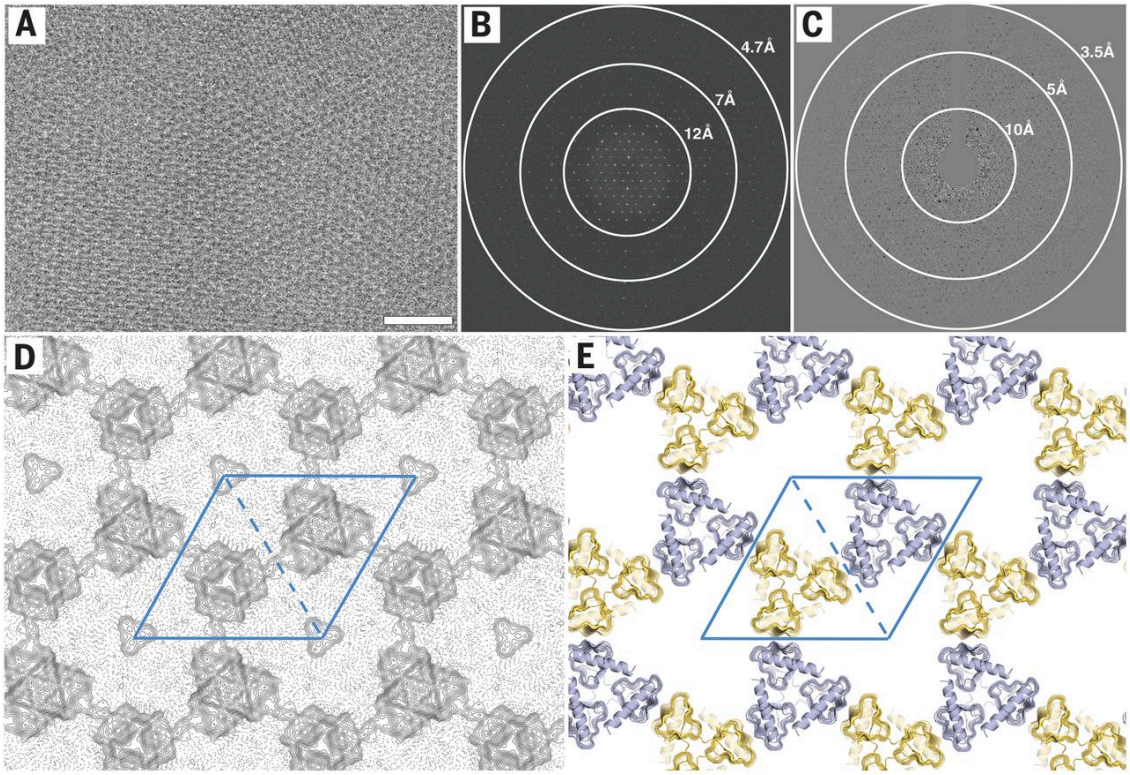


Fig. 2. Cryo-EM analysis of design p3Z_42. (**A**) Cryo-EM micrograph of *E. coli* grown p3Z_42 recorded from non-purified, re-suspended insoluble material. (**B**) Fourier transform calculated from motion-corrected movies taken from samples like in (a). (**C**) Electron-diffraction of a crystal as in (a) (**D**) 4Å projection map calculated from motion-corrected movies from material as in (a) showing a linked repeat protein arrangement similar to the p3Z_42 design model. The unit cell is shown in blue and contains two alternating trimeric units. Triangular density at the corners of the unit cell is likely an averaging artifact. (**E**) p3Z_42 design model in a similar view as in (d). Scale bar = 50nm.

Appendix I

Computational Design Methods:

We set out to design 2D layers that consisted of a native complex with cyclic symmetry, such that one designed interface would lead to self-assembling two-dimensional lattices. This leads to 7 possible layer groups: C 2 1 1 and P 2 2_1 2_1 (from C2 building blocks), P 3 and P 3 2 1 (from C3 building blocks), P 4 and P 4 2_1 2 (from C4 building blocks), and P 6 (from C6 building blocks). Additional layer groups (C 2 2 2, P 3 1 2, P 4 2 2, and P 6 2 2) are possible starting from native complexes with dihedral symmetry, but the relatively low availability of crystal structures of such complexes led us to focus on only starting structures with cyclic symmetry. The remaining six layer groups require the design of more than one interface starting from a point-symmetric building block.

Initially, we searched the PDB (http://www.rcsb.org/pdb/) for native complexes with the appropriate symmetry. Structures with a biological unit containing 2, 3, 4, or 6 chains with identical (or nearly identical) sequences that deviated from perfectly symmetric by less than 2Å RMSD were identified. The data was further limited to complexes with an asymmetric unit between 100 and 400 residues, and was trimmed to reduce redundancy by throwing out structures with >90% sequence identity; due to the large number of native C2 complexes, this was reduced to 30% for C2-symmetric building blocks. This resulted in 2929 native C2 complexes, 290 native C3 complexes, 74 native C4 complexes, and 26 native C6 complexes.

We then used symmetric docking in Rosetta (*14*) in order to find designable

configurations of each of the point-symmetric complexes into 2D layers. A symmetry definition file (*18*) was generated that modeled the inner point symmetric complex as well as the 6 or 8 complexes immediately surrounding it. During docking, the rigid-body perturbations were limited to those that maintained the configuration of the native point symmetric complexes. This led to only 2 (P 3, P 4 and P 6), 3 (P 3 2 1 and P 4 2_1 2), or 4 (P 2 2_1 2_1) rigid-body degrees of freedom that are allowed to optimize during each docking trajectory. During docking, a scoring function with only two terms was used: the first modeled sterics using a soft sphere model; the second provides a rough estimate of designable interface area by counting the number of interface Cβs within 7Å distance. For each starting model, ~20 independent Monte Carlo docking trajectories were carried out from each starting point (with more for C6 building blocks and fewer for C2 building blocks). Each resulting model was then designed.

The design methodology employed was similar to that used for the design of closed symmetric complexes in Rosetta (*12, 13*). All residues near to the interface and not part of the native interface had their residue identity and rotameric state changed in a Monte Carlo search optimizing the Rosetta energy of the entire complex (*21*). Each model then had side chain torsions as well as the symmetric degrees of freedom simultaneously minimized with respect to the energy function. Finally, these models were filtered using several different criteria: shape complementarity of the designed interface (>0.5), surface area of the designed interface (>400Å per monomer), buried unsatisfied hydrogen bonds (*22*) introduced at the new interface (<4 using a 1.4Å solvent accessibility probe size), and predicted ΔΔG (*23*) of complex formation (<-10 energy

units per subunit). The filters were adjusted for each layer group such that approximately 200 designed sequences passed the filters. Structures passing the filters were manually inspected, and then subject to additional automatic (*24*) and manual optimization. All designs were visualized in PyMOL (*27*). The filter scores for the four designs that yielded crystals are presented in Table S2.

All scripts and source code used in computational layer design has been included in Rosetta3 including source code, available at https://www.rosettacommons.org/. Any weekly release of Rosetta after May 1, 2015 can be used for the material in this study.

All the necessary inputs for replicating the calculations performed in this manuscript – including native PDB files, symmetry definition files, RosettaScripts inputs, and PDB files of the final designs of four crystals highlighted in this paper accompany the online version of this manuscript and/or in the appendix of this thesis. Sequence design also made use of previously published optimization scripts (*13*). *note* Scripts contain a %%nbblock%% flag – this is equivalent to the cyclic symmetry of the associated scaffold (e.g. 2 for C2, 3 for C3, 4 for C4 and 6 for C6) *note*

Finally, a perl script is available that allows the creation of symmetry definition files for any of the seven C-symmetry compatible layer groups described in the manuscript. The script handles symmetrization of nearly-symmetric inputs as well as generation of the inputs needed for Rosetta to construct the lattice. It can be found in the Rosetta directory path ‘apps/public/symmetry/make_Pn_tiling.pl’.

Design Sequences:

Genes were purchased from either Gen9 (http://www.gen9bio.com/) (including p6_9H) or Genescript (http://www.genscript.com/) (including p3Z_11, p3Z_42 and p4Z_9). Genes purchased from Gen9 were cloned into pet15 (Ampicillin/Carbenicillin resistant) expression vector. Genescript genes were purchased pre-inserted into pet29b (Kanamycin resistant) expression vector. A mutation (A29D) was introduced during gene synthesis to p6_9 and was retained in this study.

p3Z_11 Wildtype (PDB ID: 1NZA (*28*)) Sequence:

MEEVVLITVPSEEVARTIAKALVEERLAACVNIVPGLTSIYRWQGEVVEDQELLL
LVKTTTHAFPKLKERVKALHPYTVPEIVALPIAEGNREYLDWLRENTG

p3Z_11 Design Sequence:

MEEVVLITVPSESVARIIAKALVASRLAACVNIVPGLTSIYRWQGSVVEDQELLLL
VKTTTHAFPKLKHTVKIIHPYTVPEIVALPIAEGNREYLDWLRENTGLEHHHHHH

p3Z_42 Wildtype (PDB ID: 4AA7 (*29*)) Sequence:

MHNNRLQLSRLERVYQSEQAEKLLLAGVMLRDPARFDLRGTLTHGRDVEIDTN
VIIEGNVTLGHRVKIGTGCVIKNSVIGDDCEISPYTVVEDANLAAACTIGPFARLRP
GAELLEGAHVGNFVEMKKARLGKGSKAGHLTYLGDAEIGDNVNIGAGTITCNY
DGANKFKTIIGDDVFVGSDTQLVAPVTVGKGATIAAGTTVTRNVGENALAISRVP
QTQKEGWRRPVKKK

p3Z_42 Design Sequence:

MHNNRLQLSRLERVYQSEQAEKLLLAGVMLRDPARFDLRGSLTHGRDVEIDTN
VIIEGNVSLGNRVKIGTGCVIKNSAIGDDCEISPYTVVEDAVLAAACTIGPFARLRP
GAVLLEGAHVGNFVEMKKAVLGKGSKAGHLTYLGDAAIGDNVNIGAGTITCNY
DGANKFTTIIGDDVFVGSDTQLVAPVSVGKGATIAAGTTVTRNVGANALAISRVP
QTQKEGWRRPVKKKLEHHHHHH

p4Z_9 Wildtype (PDB ID: 3AEI (*30*)) Sequence:

MEAVRAYELQLELQQIRTLRQSLELKMKELEYAEGIITSLKSERRIYRAFSDLLVE
ITKDEAIEHIERSRLVYKREIEKLKKREKEIMEELSKLRAPLS

p4Z_9 Sequence:

MEAVRAYELQLELQQIRTLRQSLELKAKELEYAAGIITSLKSERRIYRAFSDLLVEI
TKLEAIEHIARSIIVYVREIAKLAKRETEIMEELSKLRAPLSLEHHHHHH

p6_9H Wildtype (PDB ID: 2DI4 (*31*)) Sequence:

FQGPLGSHMTISPKEKEKIAIHEAGHALMGLVSDDDDKVHKISIIPRGMALGVTQ
QLPIEDKHIYDKKDLYNKILVLLGGRAAEEVFFGKDGITTGAENDLQRATDLAYR
MVSMWGMSDKVGPIAIRRVANPFLGGMTTAVDTSPDLLREIDEEVKRIITEQYEK
AKAIVEEYKEPLKAVVKKLLEKETITCEEFVEVFKLYGIELKDKCKKEELFDKDR
KSEENKELKSEEVKEEVV

p6_9H Sequence:

MGFQGPLGSHMTISPKEKEKIAIHEAGHDLMGLVSDDDDKVHKISIIPRGMALGV
TQQLPIEDKHIYDKKDLYNKILVLLGGRAAEEVFFGKDGITTGAENDLQRATDLA
YRMVSMWGMSDKVGPIAIRRVANPFLGGMTTAVDTSPDLLREIDEEVKRIITEQ
YEKAKAIVEEYKLPLKFVVAALLHSETILCSLFAEVFKTFGIELKDKCKKEELFDK
DRKSEENKELKSEEVKEEVVHHHHHH

p6_9H_KDKCKXX Sequence:

MGFQGPLGSHMTISPKEKEKIAIHEAGHDLMGLVSDDDDKVHKISIIPRGMALGV
TQQLPIEDKHIYDKKDLYNKILVLLGGRAAEEVFFGKDGITTGAENDLQRATDLA
YRMVSMWGMSDKVGPIAIRRVANPFLGGMTTAVDTSPDLLREIDEEVKRIITEQ
YEKAKAIVEEYKLPLKFVVAALLHSETILCSLFAEVFKTFGIELKDKCK

Extended Biochemical Methods:

Mutagenesis (p6_9 and p6_9H)

Oligonucleotides containing the mutations required were ordered from IDT (http://www.idtdna.com/). Mutations were made by either the single stranded DNA "Kunkel Mutagenesis" method (*32*) or by quickchange mutagenesis using pFU Ultra II DNA polymerase (Agilent) and dNTP's (Thermo Scientific). Figure. S5 and Table S1 highlight the mutants made on design p6_9 (precursor to p6_9H). All mutated sequences were verified by either Genewiz (http://www.genewiz.com/) or internally at Janelia Research Center's molecular biology core.

Protein expression

All proteins were expressed by first transforming all purified plasmid DNA into BL21 (DE3) *E. coli* cells. Culture was grown in LB medium with the addition of either 50mg L^{-1} Kanamycin (Sigma) (p3Z_11, p3Z_42 and p4Z_9) or 100mg L^{-1} Ampicillin (Fisher Scientific) (p6_9H) until OD_{600} ~0.4 was reached at 37° Celsius. Expression was induced by the addition of 1mM IPTG (Sigma) and allowed to continue for 4 hours at 37° Celsius. For p3Z_42 cryo-EM sample, expression was induced with 0.1mM IPTG for ~19 hours at 16° Celsius after reaching OD_{600} ~0.2-0.4 at 37° Celsius. All culture was centrifuged to separate and remove the media from the cells and the cells frozen at -20° Celsius. Cells were re-suspended in Lysis buffer (25mM Tris pH 8.0, 150mM NaCl) with 1mM DTT (Acros) (p3Z_11, p3Z_42 and p6_9H) or without DTT (p4Z_9). Protein was recovered by the use of either a Sonicator (Fisher Scientific) or a Microfluidizer (microfluidics) after the addition of either 1mM PMSF (Fisher Scientific) or recommended amount of dissolved EDTA-free protease inhibitor tablet/s (Thermo Scientific). Soluble supernatant was separated from insoluble pellet material by ultracentrifugation at 12,000 x G using a Ti50.2 or Ti70 rotor (Beckman Coulter) at 4° Celsius for 30 minutes. Pellet material was re-suspended in lysis buffer and kept at 4° Celsius. All expressions were verified by SDS-PAGE (Biorad).

In-vitro expression (p3Z_42)

An Expressway (Invitrogen) cell-free protein expression kit was used as recommended with purified p3Z_42 plasmid DNA and left for the maximum time

recommended for expression (4 hours) at 37° Celsius. Negative-stain sample grids were made using the expression solution directly without purification or separation of material and visualized for crystal growth. Expression was also verified by SDS-PAGE as above.

Protein denaturing and refolding (p4Z_9)

Frozen cell pellets made from expressed p4Z_9 cells grown at 37° Celsius were re-suspended in lysis buffer (25mM Tris pH 8.0, 150mM NaCl) supplemented with EDTA-free protease inhibitor tablets (Thermo Scientific) and lysed by use of a Microfluidizer (Microfluidics). The resulting solution was spun in a Ti50.2 or Ti70 ultracentrifuge rotor (Beckman Coulter) for 30 minutes at 12,000 x g at 4° Celsius. Supernatant was discarded and pellet material was re-suspended in denaturing buffer (6M Guanidine HCL, 25mM Tris pH 8.0, 150mM NaCl) and the solution left in a 37° Celsius incubator for 1 hour. The solution was then filtered with 0.22μm filters (Millipore). Ni-NTA agarose (Qiagen) in denaturing buffer with 20mM Imidazole were added and the solution allowed to rotate slowly at 4° Celsius for two or more hours or overnight. The solution was then run on a gravity column and the beads washed twice with the same denaturing solution with 20mM Imidazole. p4Z_9 proteins were then eluted with denaturing buffer with 500mM Imidazole and concentrated using a 5K MWCO Vivaspin (Sartorius Stedim) column. The solution was then run through a Superdex 200 (10/300) column (GE Healthcare) on a (Biorad) FPLC, pre-equilibrated with denaturing buffer. Pure p4Z_9 was collected by fractionation. Fractions containing protein were pooled and concentrated again as above. Concentrations were verified by Nanodrop (Thermo

Scientific) or BCA assay (Thermo Scientific). Purity was verified by SDS-PAGE (Biorad).

Refolding of p4Z_9 was done using either fast dilution or dialysis. For dilution, the concentrated solution was added to varying amounts of lysis buffer (25mM Tris pH 8.0, 150mM NaCl) at 4° Celsius. The solution was then concentrated as above and analyzed by negative-stain EM (Fig. S4b). For dialysis, the denatured solution was injected into a wet dialysis cassette (Thermo Scientific) revolving in a bath of lysis buffer at room temperature and allowed to refold for 1 hour or overnight at 4° Celsius. Re-folded protein was extracted from the dialysis cassette and viewed by negative-stain EM (Fig. S4c).

Protein purification and in vitro assembly (p6_9H)

Supernatant p6_9H was separated from the pellet material and filtered with 0.22μm filters (Millipore). Ni-NTA agarose (Qiagen) in lysis buffer with 1mM DTT and 20mM Imidazole was added to the solution allowed to rotate slowly at 4° Celsius for 2 Hours or more. The solution was then run on a gravity column and beads washed twice with lysis buffer and 1mM DTT and 20mM Imidazole for the first wash and 1mM DTT and 40mM imidazole for the second. The protein was then eluted with lysis buffer with 1mM DTT and 500mM Imidazole. The solution was run on a pre-equilibrated Sephacryl S-300 (26/60) (GE Healthcare) column in a (biorad) FPLC and fractions collected. Fractions were then pooled and concentrated in a 10K MWCO Vivaspin (Sartorius

Stedim) column. The protein concentration was determined using a BCA assay (Thermo Scientific) and purity was verified by SDS-PAGE (Biorad) and flash frozen using liquid nitrogen and stored at -80° Celsius. Arrays were not seen at this point and the sample appeard as homogeneous single particles (Fig. S4d). The protein was concentrated to ~30mg/mL and extensive arrays were observed after 1 hour incubation at 37° Celsius (Fig. S4e).

Negative-stain Electron Microscopy

A drop of 2-3µL sample was applied on negatively glow discharged, carbon-coated 200-mesh copper grids (Ted Pella, Inc.), washed with Milli-Q Water and stained using 0.75% uranyl formate as described previously (*16*). Screening was performed on either a 120kV Tecnai Spirit T12 transmission electron microscope (FEI, Hillsboro, OR) or a 100kV Morgagni M268 transmission electron microscope (FEI, Hillsboro, OR). Images were recorded on a bottom mount Teitz CMOS 4k camera system. The contrast of the images was enhanced in Fiji (*33*) for clarity.

Projection maps

Micrographs of negatively stained preparations or of cryo preparations were processed in the MRC suite of programs (*25, 34, 35*) through the 2dx interface (*25*) using well-established protocols (*36*)

Cryo Electron Microscopy and motion corrected movies

An aliquot of 2µL of p3Z_42 sample was placed onto a holey carbon grid and plunged into liquid ethane using a FEI vitrobot and cryo transferred onto a cryo microscope under liquid nitrogen temperatures. Samples were viewed on either an FEI Technai F20 using a Teitz 4x4k camera or an FEI Titan Krios using a K2 camera to record super-resolution movies. All movies were motion corrected using software as described previously (*37*) with a bin of 1. Diffraction data were collected on the FEI Technai F20 operating in diffraction mode and recorded on a Teitz 2x2k camera and processed in XDP (*38*) using established procedures (*39*). The contrast of the images was enhanced in Fiji (*33*) for clarity.

All panels were made using PyMOL (*27*) , Fiji (*33*) and assembled in Adobe Photoshop CS5 (http://www.adobe.com/).

Extended Results:

p3Z_11 Design

Design p3Z_11 (P 3 2 1 symmetry) (Fig. S1) was found to make stacked 2D or 3D crystals in vivo. The interface is made up of six interlocking Isoleucine residues flanked by serine-histidine hydrogen bonds on two sides of the anti-parallel interface resulting from the flipped orientation of the trimeric building blocks. The z height

between subunits differs from the plane of the crystal by a substantial amount causing the entire 2D assembly to be in a zipper-like motif that is perhaps conducive to the formation of 3D crystals in the small, highly concentrated environment found in vivo.

p6_9H_KDKCKXX Construct

A new construct was made from p6_9H, where 33 C-terminal amino acid residues (including 6xHIS) not used at the protein-protein interface and not having structural information in the original WT crystal structure were removed in order to check protein stability, called p6_9H_KDKCKXX. This significant (~15% including 6xHis) removal of residues from the protein did not result in breaking the arrays (Table S1). Protein stability was reduced however with stacked 2D crystals viewed in a similar ratio as single layered sheets suggesting these residues are required for the original C6 scaffold stability.

Figures

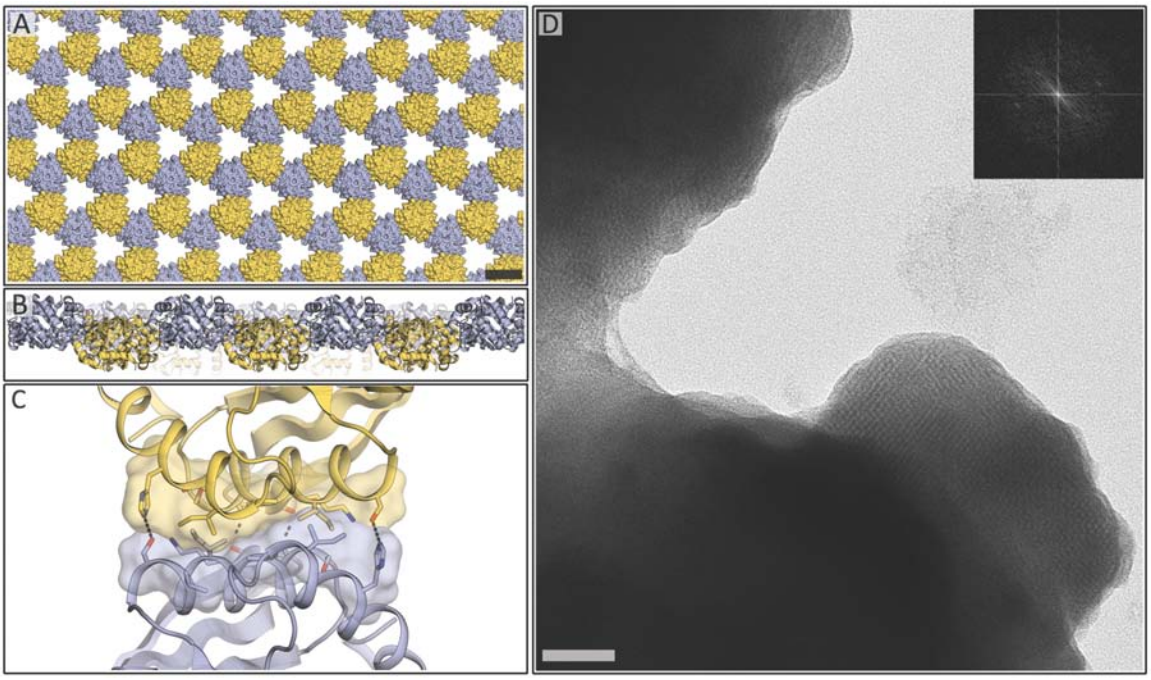


Fig. S1. Design p3Z_11 in P 3 2 1 symmetry (A) Design p3Z_11 shown in VDW space filled view with the purple and yellow proteins oriented 180° from each other on Z axis in P 3 2 1 symmetry, similar to p3Z_42 design. (B) In-plane view of the p3Z_11 design showing the change in z height between the trimeric subunits. Lattice thickness by design =~4nm (C) p3Z_11 design interface showing a large hydrophobic patch made of six isoleucines flanked by hydrogen bond networks. Transparent VDW interface area is also shown to highlight the lock-and-key docked design between trimeric subunits. (D) Negative-stain micrograph of p3Z_11 showing a large stacking of proteins in 2D to form 3D crystals. The edges of which contain an observable lattice giving spots on a Fourier transform (top right). Scale bars: Black = 5nm, white = 50nm.

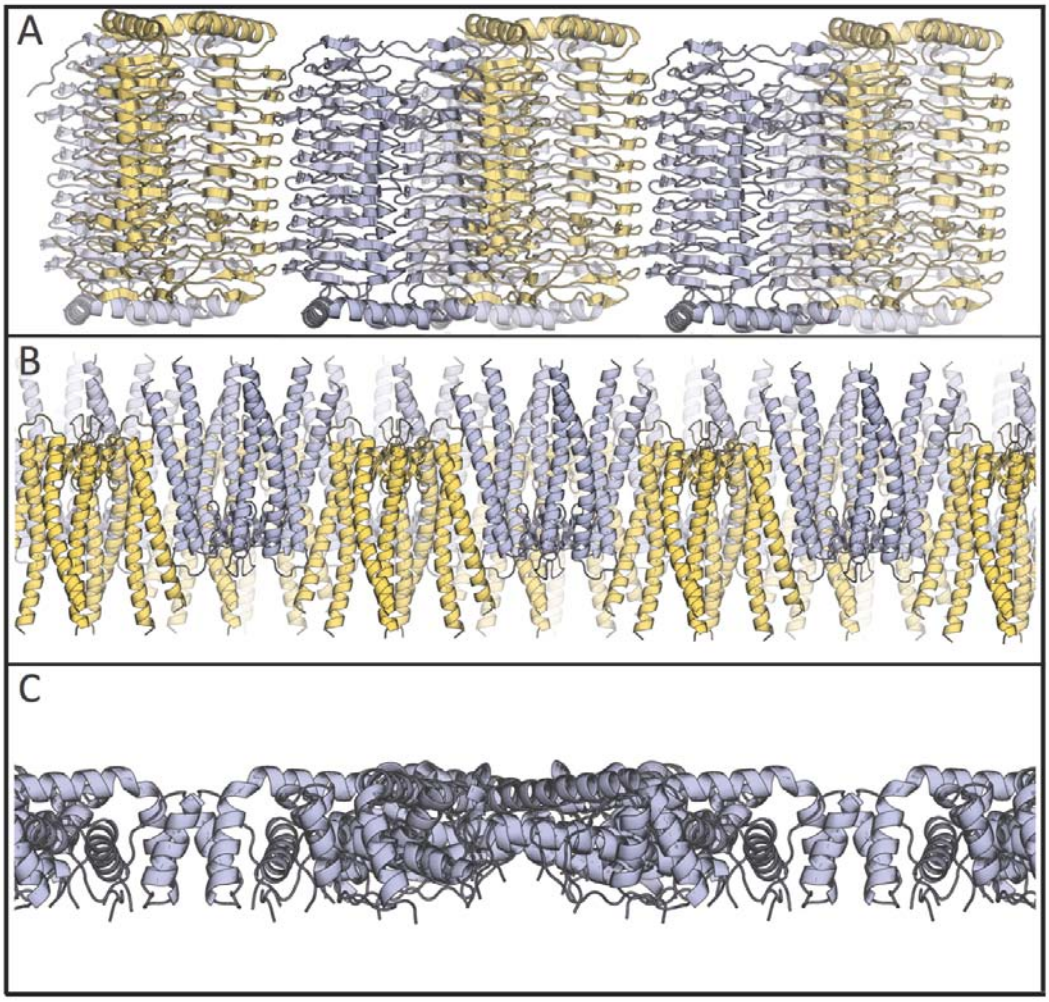


Fig. S2. In-plane views of p3Z_42, p4Z_9 and p6_9H. (**A**) p3Z_42 design in-plane view showing a slight difference in z height between neighboring trimers. Lattice design thickness =~7nm (**B**) p4Z_9 design in-plane view highlighting a great difference in z height between neighboring tetrameric proteins. Lattice design thickness =~8nm (**C**) p6_9H design in-plane view showing no difference in z height between neighboring hexameric proteins due to the lack of a z degree of freedom in P 6 symmetry. Lattice design thickness =~3nm.

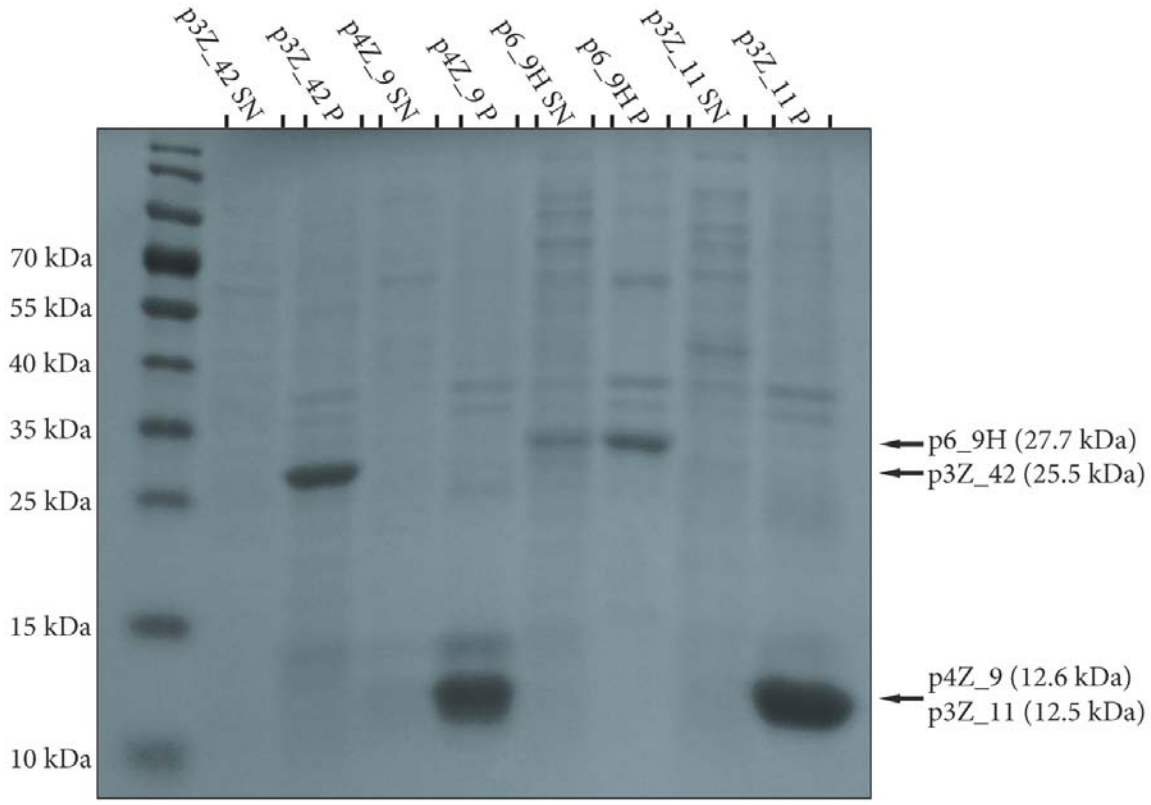


Fig. S3. SDS-PAGE gel of (from left to right) p3Z_42, p4Z_9, p6_9H and p3Z_11 protein expression. SN=soluble supernatant, P=insoluble pellet. Expression of p3Z_42, p4Z_9 and p3Z_11 protein is almost exclusively contained in the insoluble pellet material while design p6_9H proteins express mostly in the pellet while some proteins remain soluble.

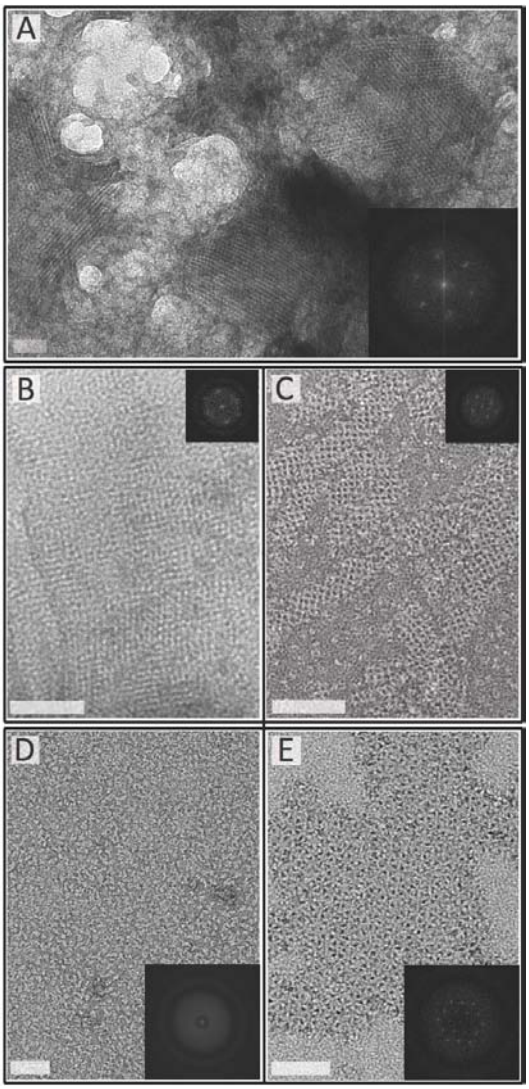


Fig. S4. In-vitro array formation of p3Z_42, p4Z_9 and p6_9H designs. (**A**) Design p3Z_42 expressed using an in-vitro expression kit. This negative-stain micrograph was made 4 hours after adding pure plasmid DNA of p3Z_42 to the kit components without purification. A Fourier transform is shown from a crystal in the micrograph showing the same P 3 2 1 lattice as visualized in p3Z_42 *E. coli* expression. (**B**) Fast dilution re-folded p4Z_9 protein. Large arrays form analogous to those seen from *E. coli* expressed protein. A Fourier transform is shown highlighting the square lattice. (**C**) Dialysis re-folded p4Z_9 protein. Large fibrous structures form with the same square array pattern as in E.

coli expressed proteins. Fourier transform is shown highlighting the square repeat pattern. (**D**) Purified and concentrated protein from p6_9H soluble fractions. Arrays were not visualized at this point. Fourier transform of the image reveals no P 6 repeat pattern. (**E**) p6_9H array formation from material as in (d). These arrays formed after further concentration of protein as in (d) and heat application in a water bath. The EM grid was prepared by a 50-fold dilution of the concentrated array product, suggesting that once formed, the arrays are very stable in solution. Fourier transform is shown with the same P 6 arrangement seen in the pellet sample. Scale bars = 50nm

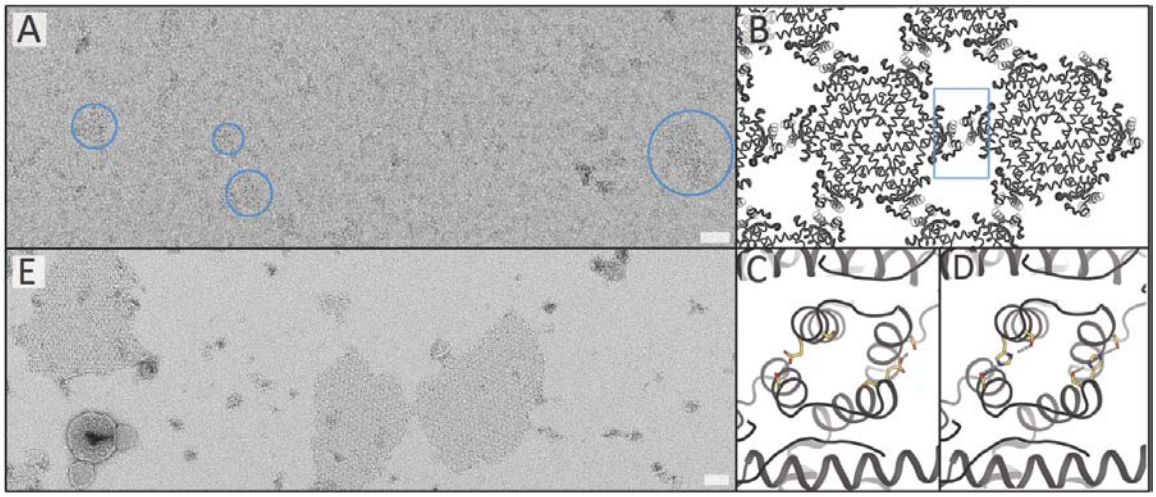


Fig. S5. Mutagenesis of p6_9 (precursor to p6_9H). (**A**). Micrograph of negatively-stained p6_9 pellet. Small patches of single-layer, 2D hexamers could be clearly observed. (**B**) p6_9 protein design highlighting the repeat interface area (blue). (**C**) Zoom-in view of the p6_9 interface showing E188. (**D**) Zoom-in view of the p6_9H interface highlighting the E188H mutation made to stabilize the design by forming a hydrogen bond network with neighboring serines on both the same hexamer and the P 6 related hexamer. (**E**) Micrograph of negatively-stained p6_9H pellet. Larger, more stable 2D arrays could be readily observed in sharp contrast to p6_9. Scale bars = 50nm

Table S1. Mutagenesis of p6_9 design (pre-cursor of p6_9H)

Mutation/s	**Sizes of crystals observed in the pellet**
Original Design p6_9 (Control)	+
A184S	+
T203V	+
E188R	+
E199L	+
E188H (p6_9H)	+++
F181R	**None observed**
L193T	+
L193T, A198V	+
L193T, S189K	+
L193T, A198V, S189K	++
L193T, A198V, S189K, L177E	**None observed**
L193T, A198V, S189K, cut 6xHis	++
E188H, V200M (p6_9HM)	+++
E188H, F218Y	+++
E188H, D29A	+++
E188H, L193T, A198V	+++
E188H, cut 6xHis	+++
E188H, short construct (p6_9H_KDKCKXX)	++

Table S2. Final Rosettascripts filter scores for p3Z_11, p3Z_42, p4Z_9 and p6_9H

Design	ΔΔG	Mutations	Shape Complementarity	Unsatisfied Polar Residues
p3Z_11	-13.34	9	0.682	1
p3Z_42	-20.8	11	0.634	2
p4Z_9	-16.12	10	0.648	2
p6_9H	-15.83	12*	0.73	0

* An additional mutation (A29D) was introduced during gene synthesis

References for Chapter II and Appendix I

1. N. C. Seeman, Nanomaterials based on DNA. *Annual review of biochemistry* **79**, 65-87 (2010).
2. P. W. Rothemund, Folding DNA to create nanoscale shapes and patterns. *Nature* **440**, 297-302 (2006).
3. D. Pum, J. L. Toca-Herrera, U. B. Sleytr, S-layer protein self-assembly. *International journal of molecular sciences* **14**, 2484-2501 (2013).
4. D. L. Caspar, D. A. Goodenough, L. Makowski, W. C. Phillips, Gap junction structures. I. Correlated electron microscopy and x-ray diffraction. *The Journal of cell biology* **74**, 605-628 (1977).
5. J. E. Rash, K. G. Davidson, T. Yasumura, C. S. Furman, Freeze-fracture and immunogold analysis of aquaporin-4 (AQP4) square arrays, with models of AQP4 lattice assembly. *Neuroscience* **129**, 915-934 (2004).
6. C. J. Lanci *et al.*, Computational design of a protein crystal. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 7304-7309 (2012).
7. J. C. Sinclair, K. M. Davies, C. Venien-Bryan, M. E. Noble, Generation of protein lattices by fusing proteins with matching rotational symmetry. *Nature nanotechnology* **6**, 558-562 (2011).
8. J. Zhang, F. Zheng, G. Grigoryan, Design and designability of protein-based assemblies. *Current opinion in structural biology* **27**, 79-86 (2014).

9. J. D. Brodin *et al.*, Metal-directed, chemically tunable assembly of one-, two- and three-dimensional crystalline protein arrays. *Nature chemistry* **4**, 375-382 (2012).

10. F. Baneyx, J. F. Matthaei, Self-assembled two-dimensional protein arrays in bionanotechnology: from S-layers to designed lattices. *Current opinion in biotechnology* **28**, 39-45 (2014).

11. E. D. Levy, E. Boeri Erba, C. V. Robinson, S. A. Teichmann, Assembly reflects evolution of protein complexes. *Nature* **453**, 1262-1265 (2008).

12. N. P. King *et al.*, Computational design of self-assembling protein nanomaterials with atomic level accuracy. *Science* **336**, 1171-1174 (2012).

13. N. P. King *et al.*, Accurate design of co-assembling multi-component protein nanomaterials. *Nature* **510**, 103-108 (2014).

14. I. Andre, P. Bradley, C. Wang, D. Baker, Prediction of the structure of symmetrical protein assemblies. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 17656-17661 (2007).

15. I. Andre, C. E. Strauss, D. B. Kaplan, P. Bradley, D. Baker, Emergence of symmetry in homooligomeric biological assemblies. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 16148-16152 (2008).

16. B. L. Nannenga, M. G. Iadanza, B. S. Vollmar, T. Gonen, Overview of electron crystallography of membrane proteins: crystallization and screening strategies using negative stain electron microscopy. *Current protocols in protein science / editorial board, John E. Coligan ... [et al.]* **Chapter 17**, Unit17 15 (2013).

17. S. W. Wukovitz, T. O. Yeates, Why protein crystals favour some space-groups over others. *Nature structural biology* **2**, 1062-1067 (1995).

18. F. DiMaio, A. Leaver-Fay, P. Bradley, D. Baker, I. Andre, Modeling symmetric macromolecular structures in Rosetta3. *PloS one* **6**, e20450 (2011).

19. S. J. Fleishman *et al.*, RosettaScripts: a scripting language interface to the Rosetta macromolecular modeling suite. *PloS one* **6**, e20161 (2011).

20. M. C. Lawrence, P. M. Colman, Shape complementarity at protein/protein interfaces. *Journal of molecular biology* **234**, 946-950 (1993).

21. B. Kuhlman *et al.*, Design of a novel globular protein fold with atomic-level accuracy. *Science* **302**, 1364-1368 (2003).

22. Z. S. Hendsch, T. Jonsson, R. T. Sauer, B. Tidor, Protein stabilization by removal of unsatisfied polar groups: computational approaches and experimental tests. *Biochemistry* **35**, 7621-7625 (1996).

23. E. H. Kellogg, O. F. Lange, D. Baker, Evaluation and optimization of discrete state models of protein folding. *The journal of physical chemistry. B* **116**, 11405-11413 (2012).

24. L. G. Nivon, R. Moretti, D. Baker, A Pareto-optimal refinement method for protein design scaffolds. *PloS one* **8**, e59004 (2013).

25. B. Gipson, X. Zeng, Z. Y. Zhang, H. Stahlberg, 2dx--user-friendly image processing for 2D crystals. *Journal of structural biology* **157**, 64-72 (2007).

26. D. Shi, B. L. Nannenga, M. G. Iadanza, T. Gonen, Three-dimensional electron crystallography of protein microcrystals. *eLife* **2**, e01345 (2013).

27. The PyMOL Molecular Graphics System, Version 1.7.2, Schrödinger, LLC (http://www.pymol.org).

28. B. Bagautdinov, The structures of the CutA1 proteins from Thermus thermophilus and Pyrococcus horikoshii: characterization of metal-binding sites and metal-induced assembly. *Acta crystallographica. Section F, Structural biology communications* **70**, 404-413 (2014).

29. O. M. Green *et al.*, Inhibitors of acetyltransferase domain of N-acetylglucosamine-1-phosphate-uridyltransferase/glucosamine-1-phosphate-acetylt ransferase (GlmU). Part 1: Hit to lead evaluation of a novel arylsulfonamide series. *Bioorganic & medicinal chemistry letters* **22**, 1510-1519 (2012).

30. M. Sahlan *et al.*, Thermodynamic characterization of the interaction between prefoldin and group II chaperonin. *Journal of molecular biology* **399**, 628-636 (2010).

31. R. Suno *et al.*, Structure of the whole cytosolic region of ATP-dependent protease FtsH. *Molecular cell* **22**, 575-585 (2006).

32. R. Huang, P. Fang, B. K. Kay, Improvements to the Kunkel mutagenesis protocol for constructing primary and secondary phage-display libraries. *Methods* **58**, 10-17 (2012).

33. J. Schindelin *et al.*, Fiji: an open-source platform for biological-image analysis. *Nature methods* **9**, 676-682 (2012).

34. B. Gipson, X. Zeng, H. Stahlberg, 2dx_merge: data management and merging for 2D crystal images. *Journal of structural biology* **160**, 375-384 (2007).

35. R. A. Crowther, R. Henderson, J. M. Smith, MRC image processing programs. *Journal of structural biology* **116**, 9-16 (1996).

36. K. Mitsuoka *et al.*, The structure of bacteriorhodopsin at 3.0 A resolution based on electron crystallography: implication of the charge distribution. *Journal of molecular biology* **286**, 861-882 (1999).

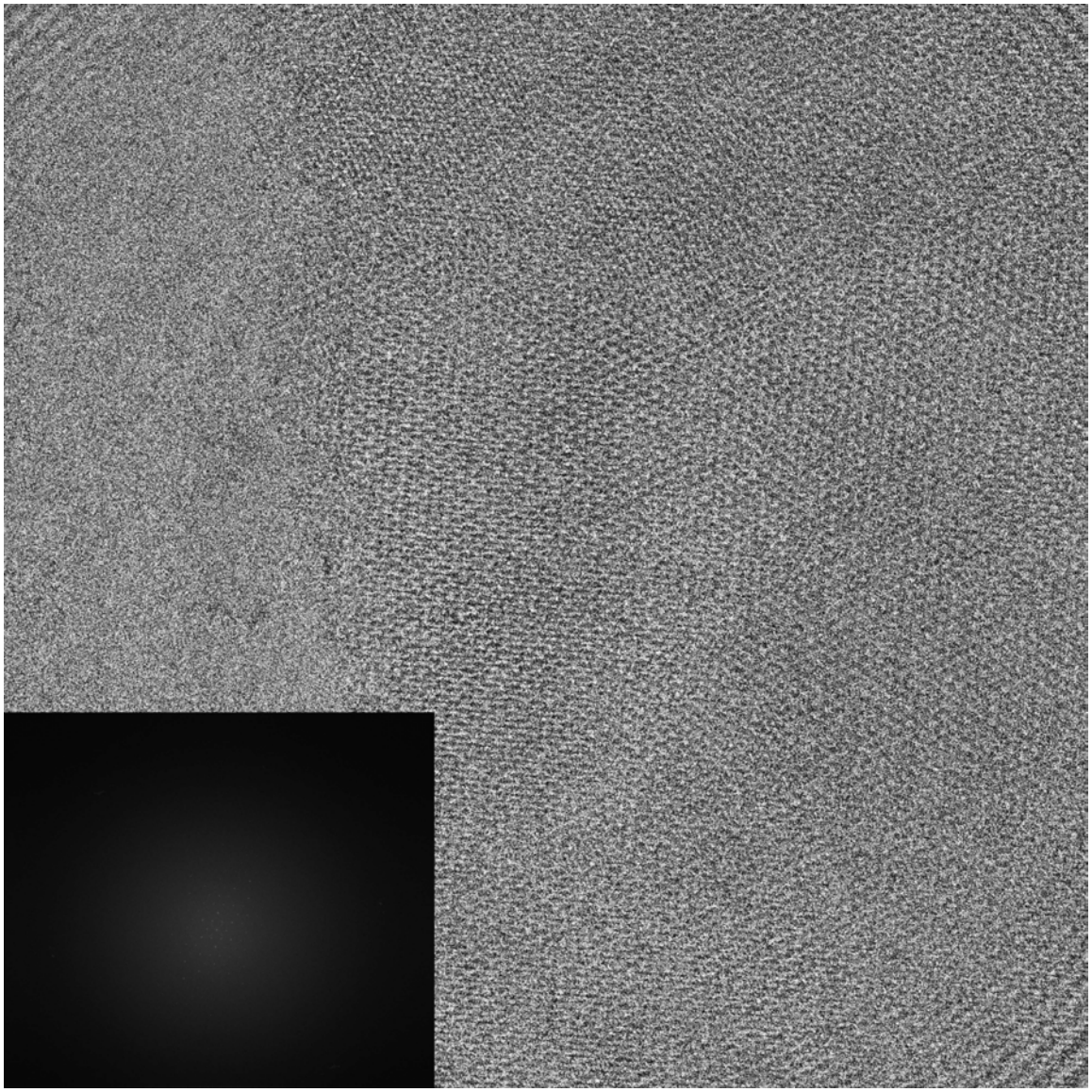
37. X. Li *et al.*, Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nature methods* **10**, 584-590 (2013).

38. K. Mitsuoka, Processing of electron diffraction patterns with the XDP program. *Methods in molecular biology* **955**, 539-550 (2013).

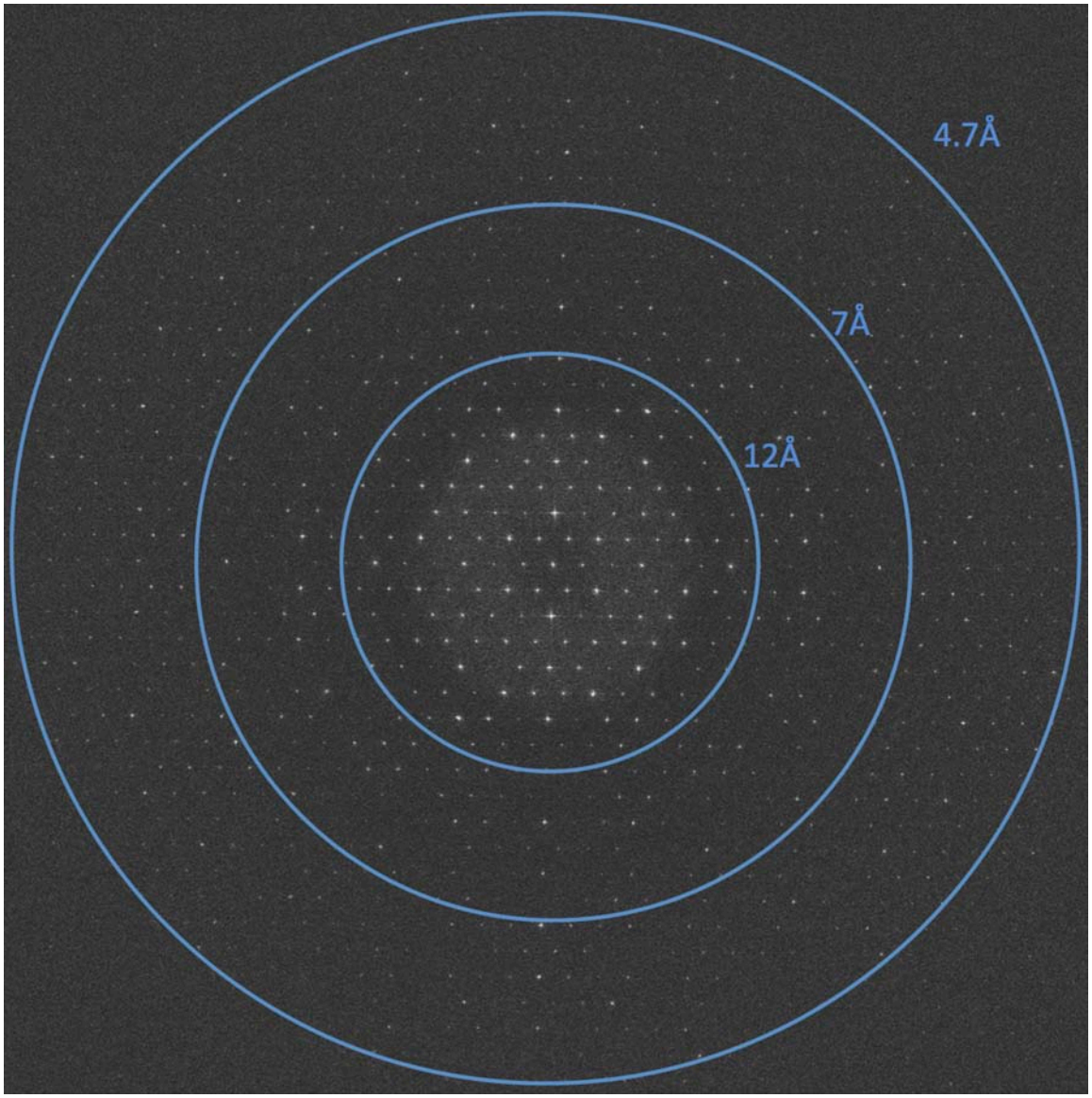
39. T. Gonen *et al.*, Lipid-protein interactions in double-layered two-dimensional AQP0 crystals. *Nature* **438**, 633-638 (2005).

Appendix II

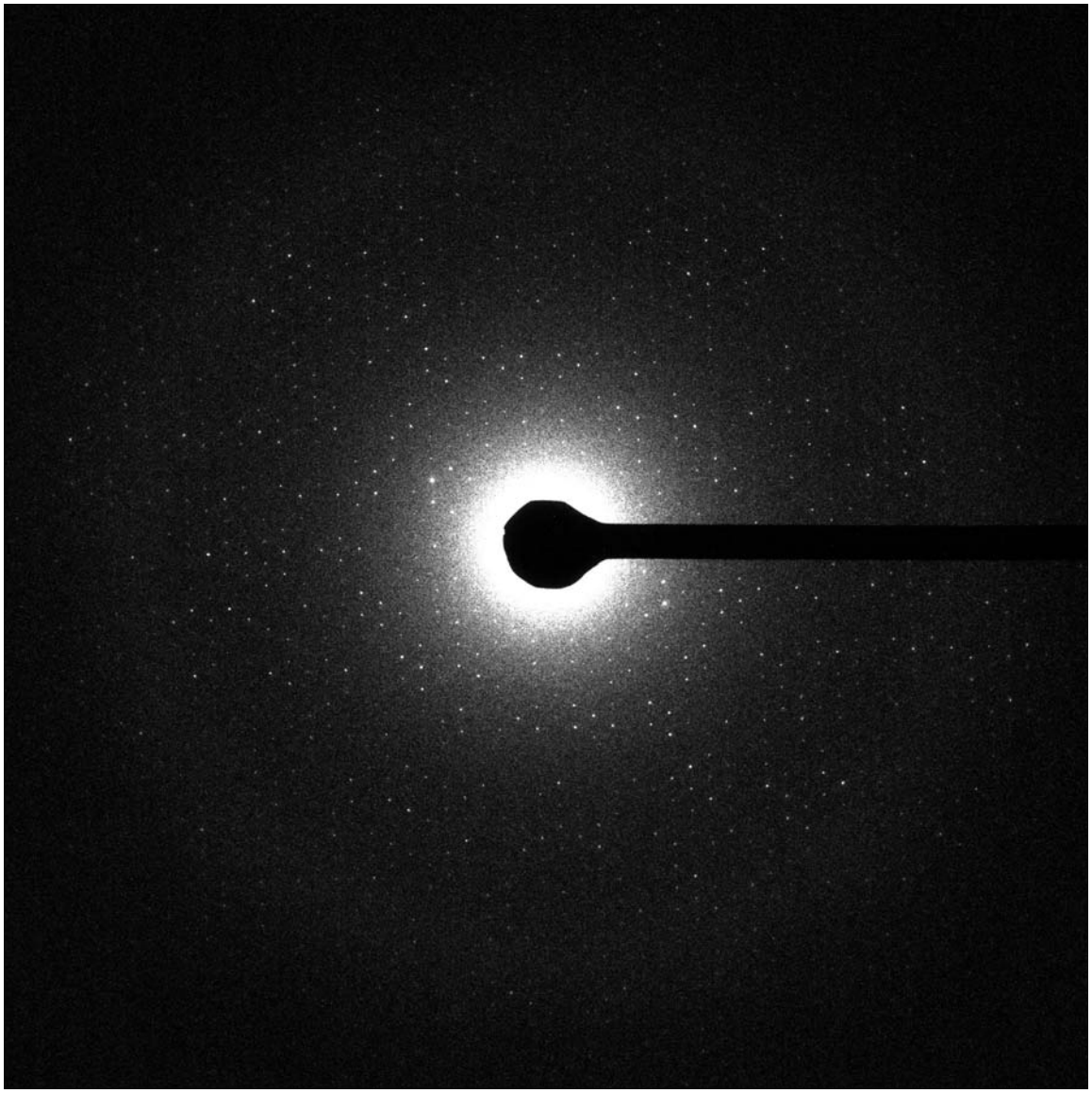
Additional Images, Micrographs and Results



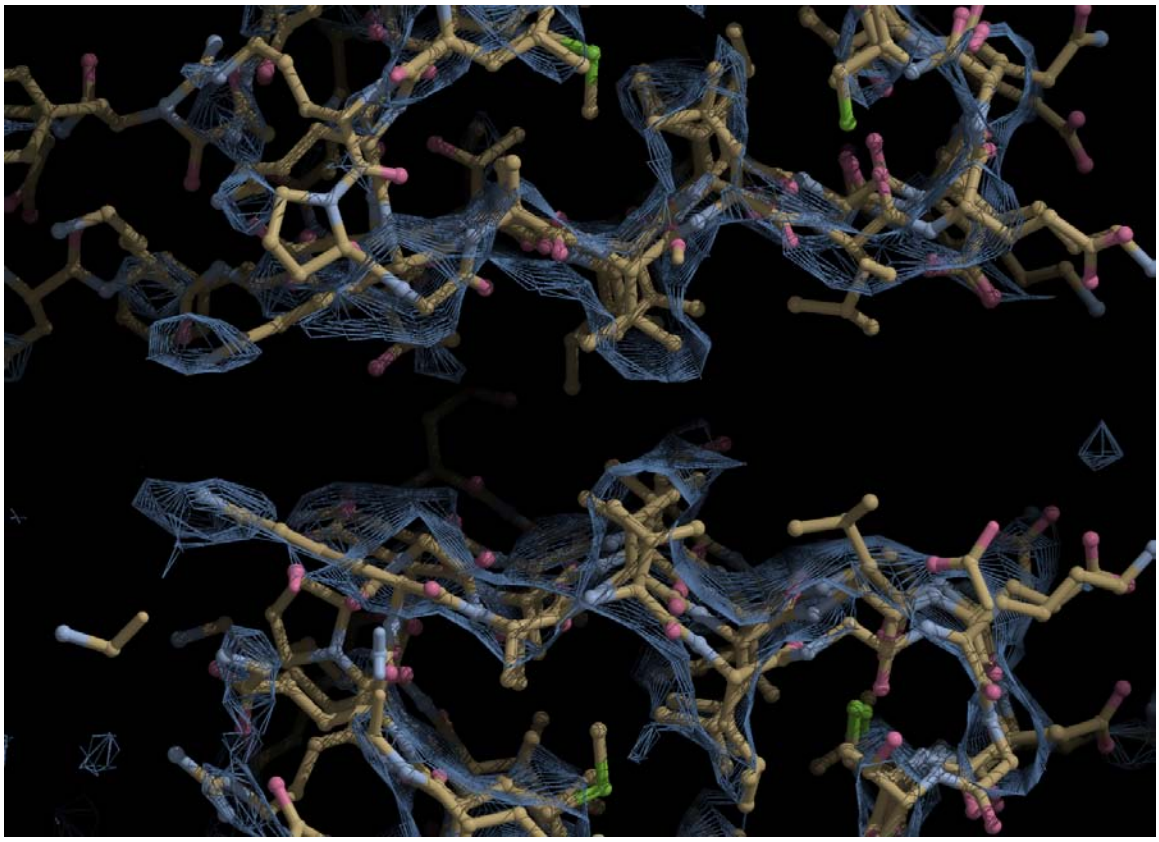
p3Z-42 crystal from pellet material under cryo-EM showing high-resolution and order.



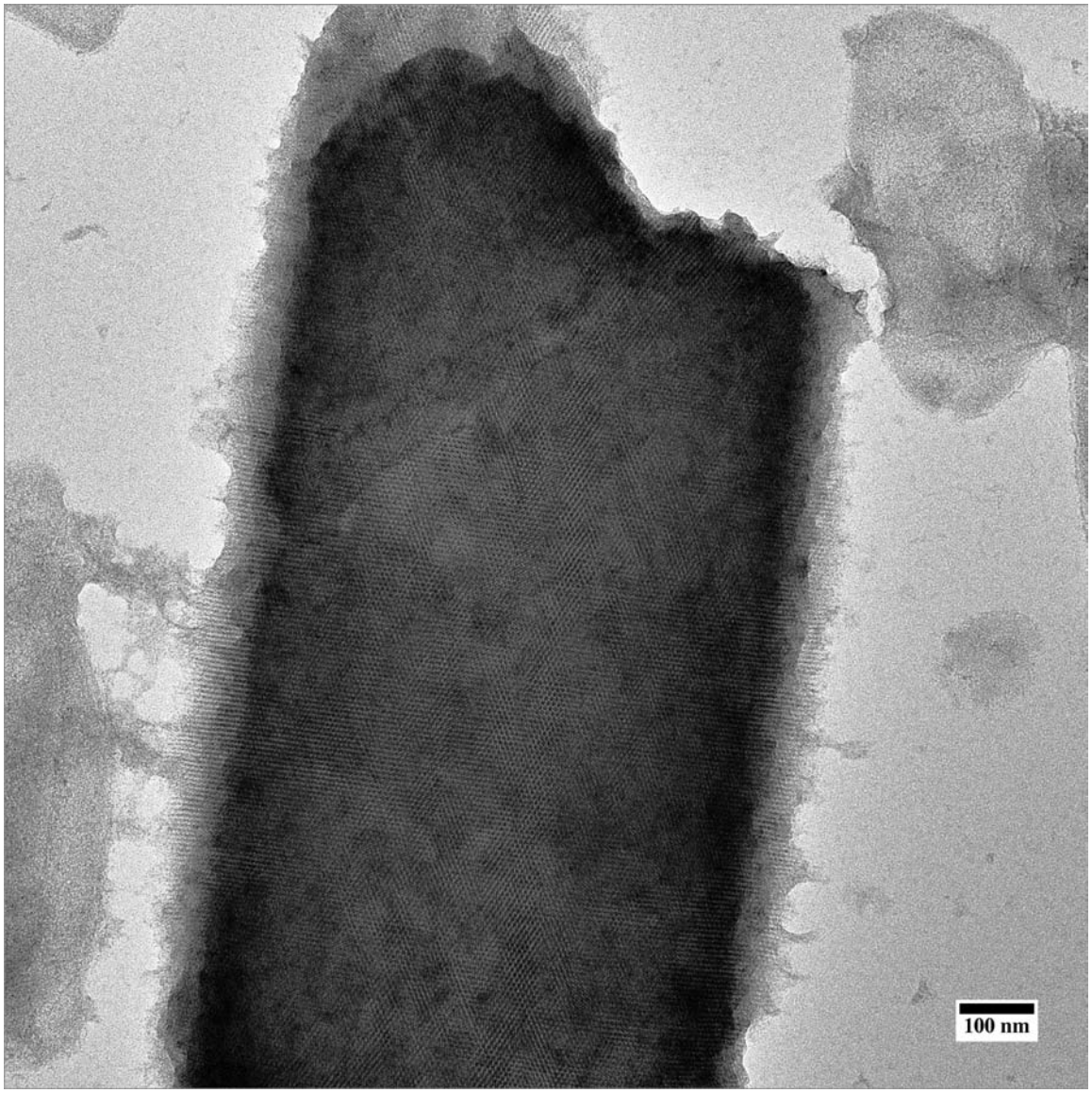
FFT from a micrograph of p3Z-42 crystal in cryo-EM.



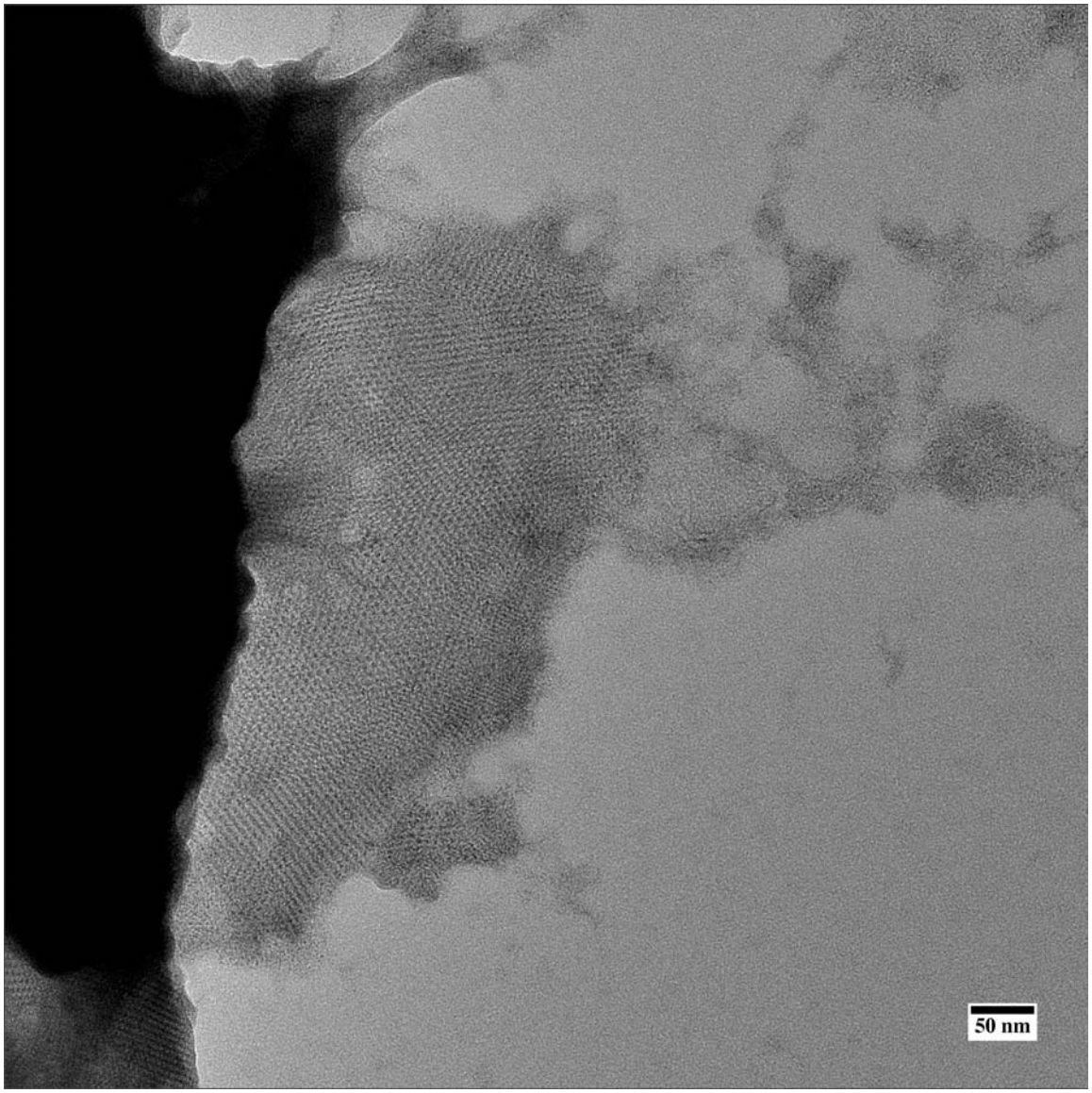
Raw, unprocessed diffraction from a p3Z-42 crystal by cryo-EM.



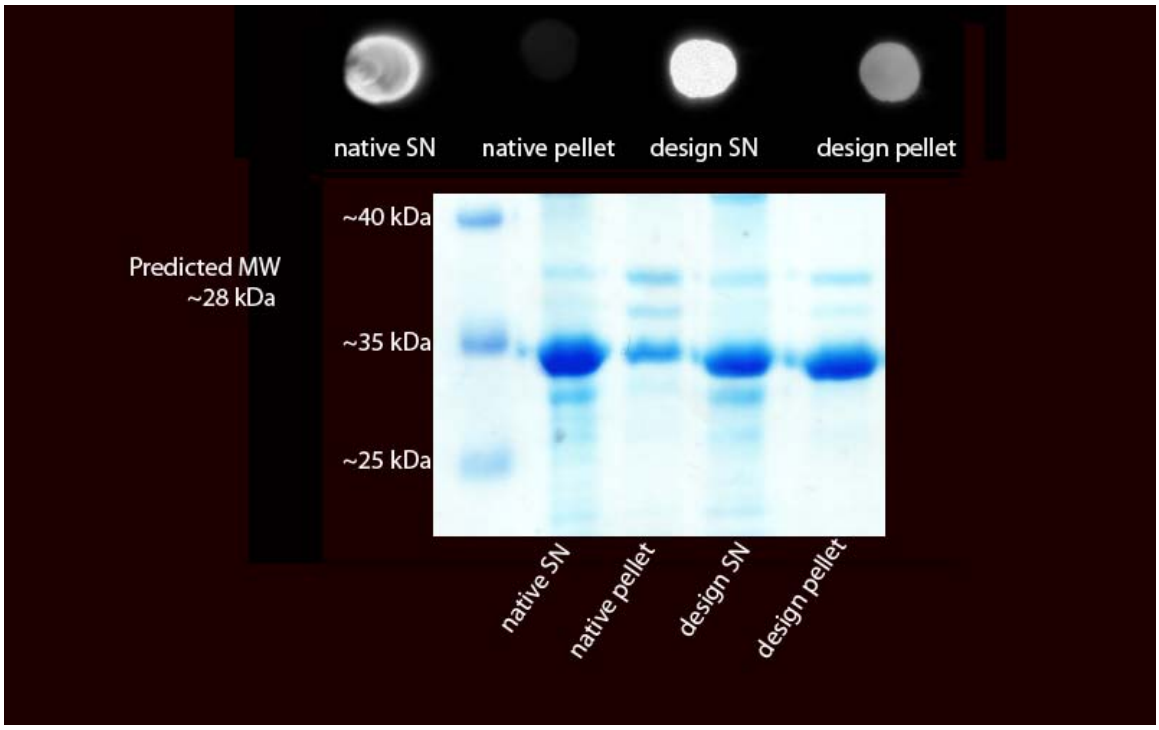
Rosetta model fit to the density map obtained by electron diffraction of p3Z-42. Because the sheets bound to cellular debris during expression, tilting of the p3Z-42 crystal resulted in the loss of information as the crystals were not flat, and a full 3D density map could not be obtained. The above map is only ~20% complete, mostly by projection. Zoom in view of the interface between p3Z-42 trimers resulting in the crystal formation is shown above, matching the model.



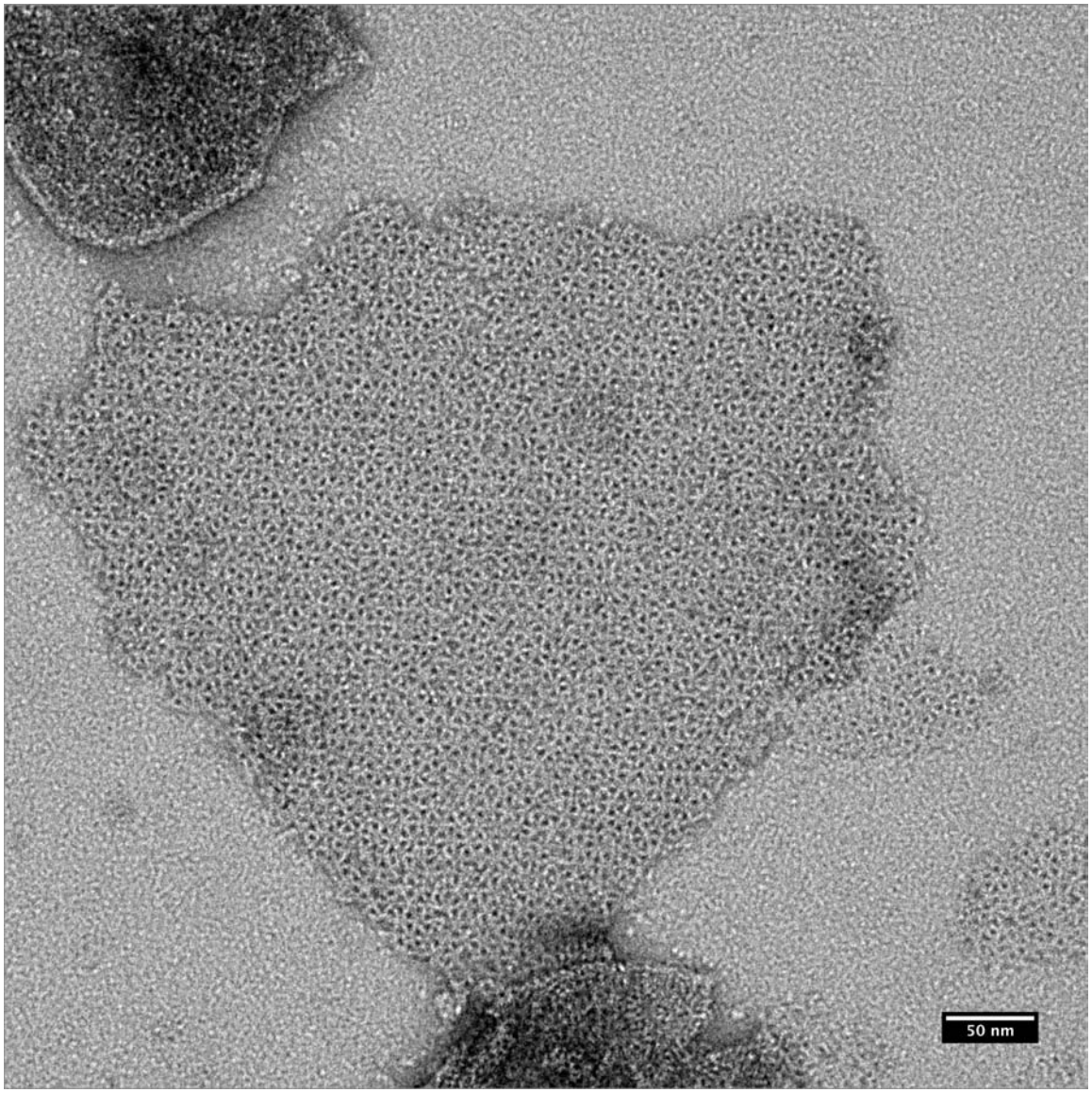
Negative-stain TEM micrograph of 16^{o}C-expressed p3Z-42 crystal showing a large 2D crystal with cellular debris underneath from the *E. coli* cell.



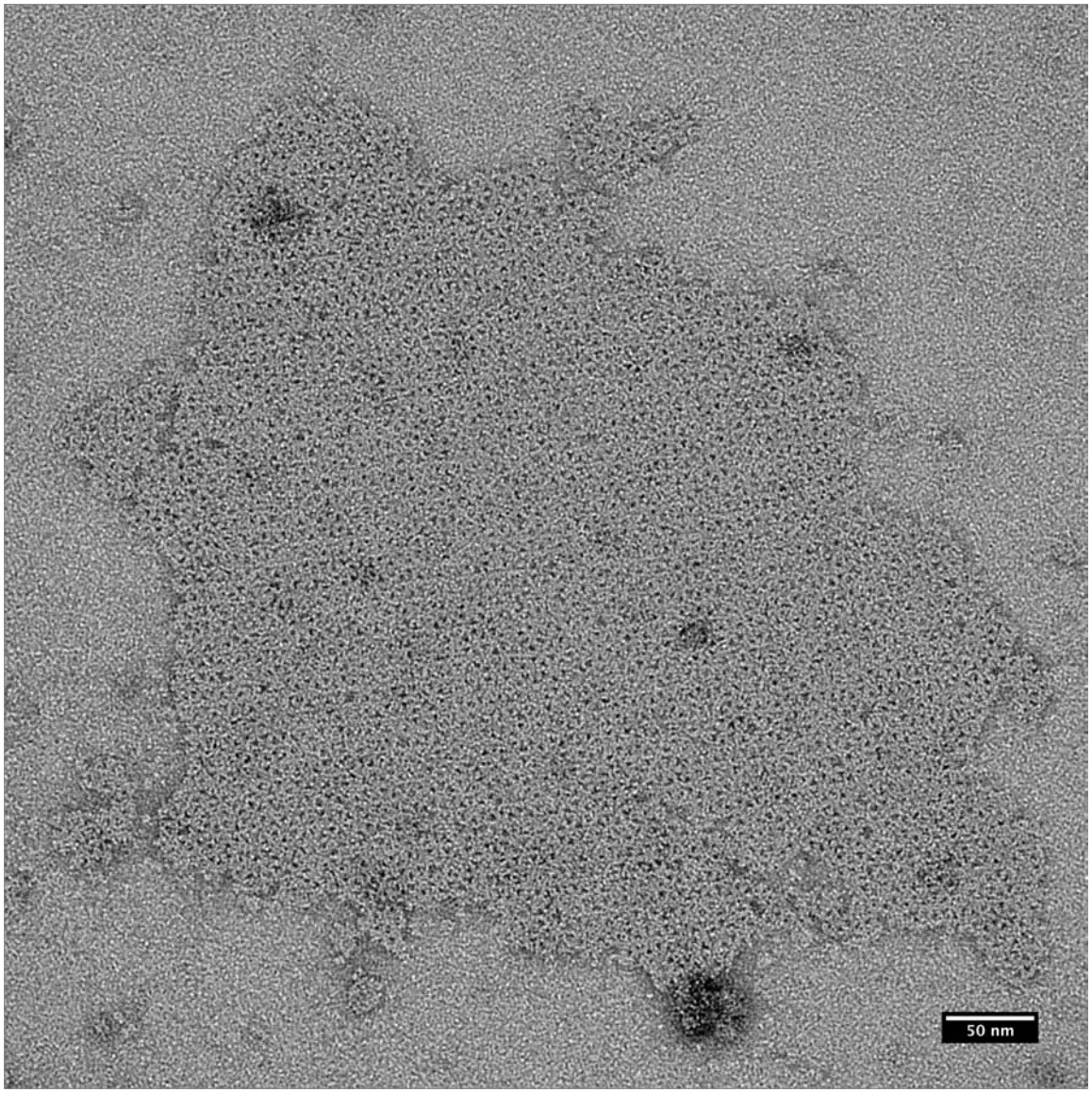
Negative-stain TEM micrograph of 37^{o}C-expressed p3Z-42 crystals showing a 2D crystal (right) next to stacked 2D crystals (black-left). The stacked crystals were readily seen at this expression temperature.



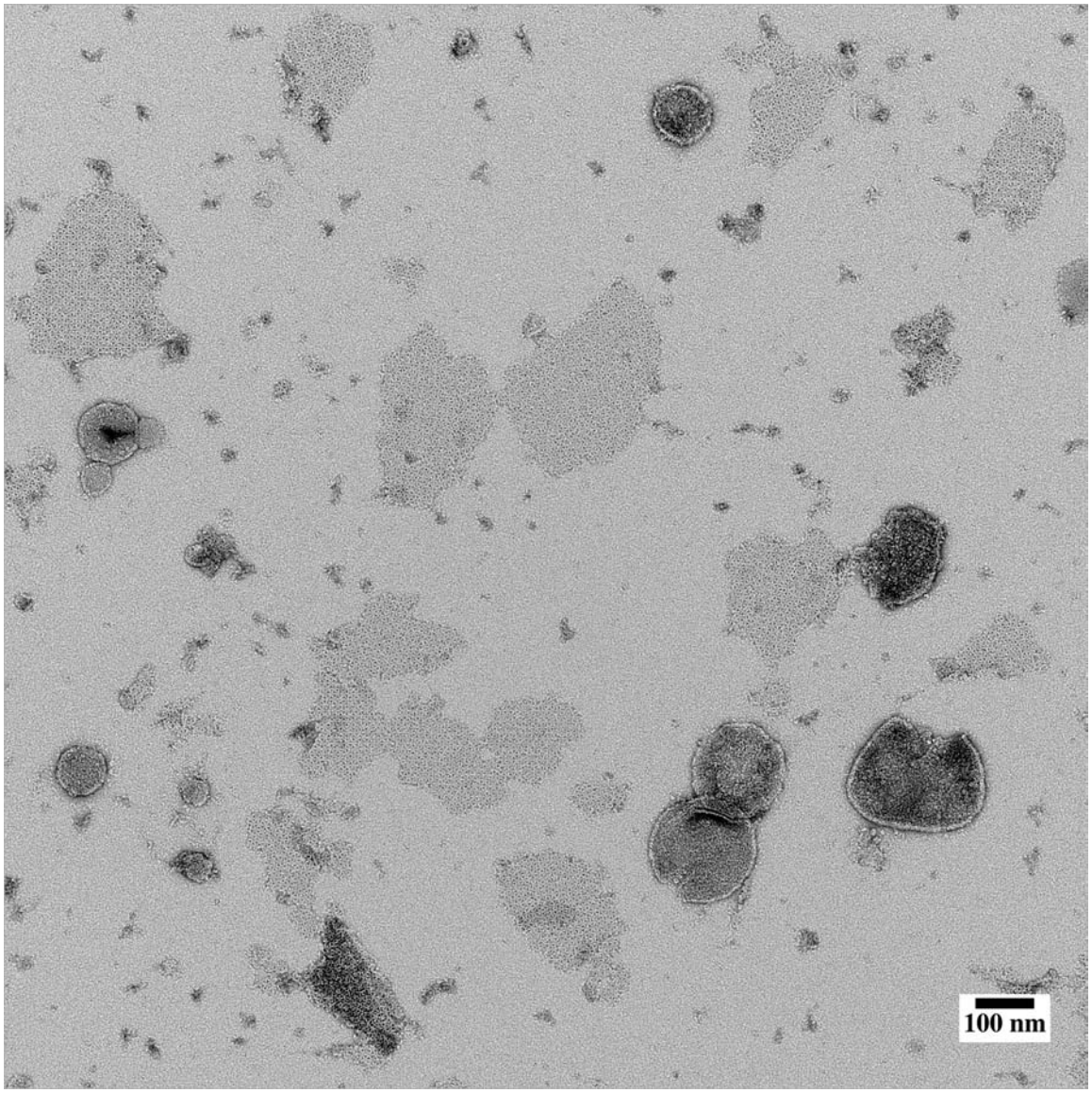
Dot blot (above) and SDS-Gel (below) showing the original p6 scaffold protein 2DI4's supernatant and pellet expression compared to the original p6-9 design supernatant and pellet. The dot blot was made as a probe against a 6xHIS tag on nitrocellulose membrane. This shows the design expressing and going mostly to the pellet material, as would be expected from large 2D crystals. These preparations were viewed by negative-stain TEM as described in Chapter II and Appendix I.



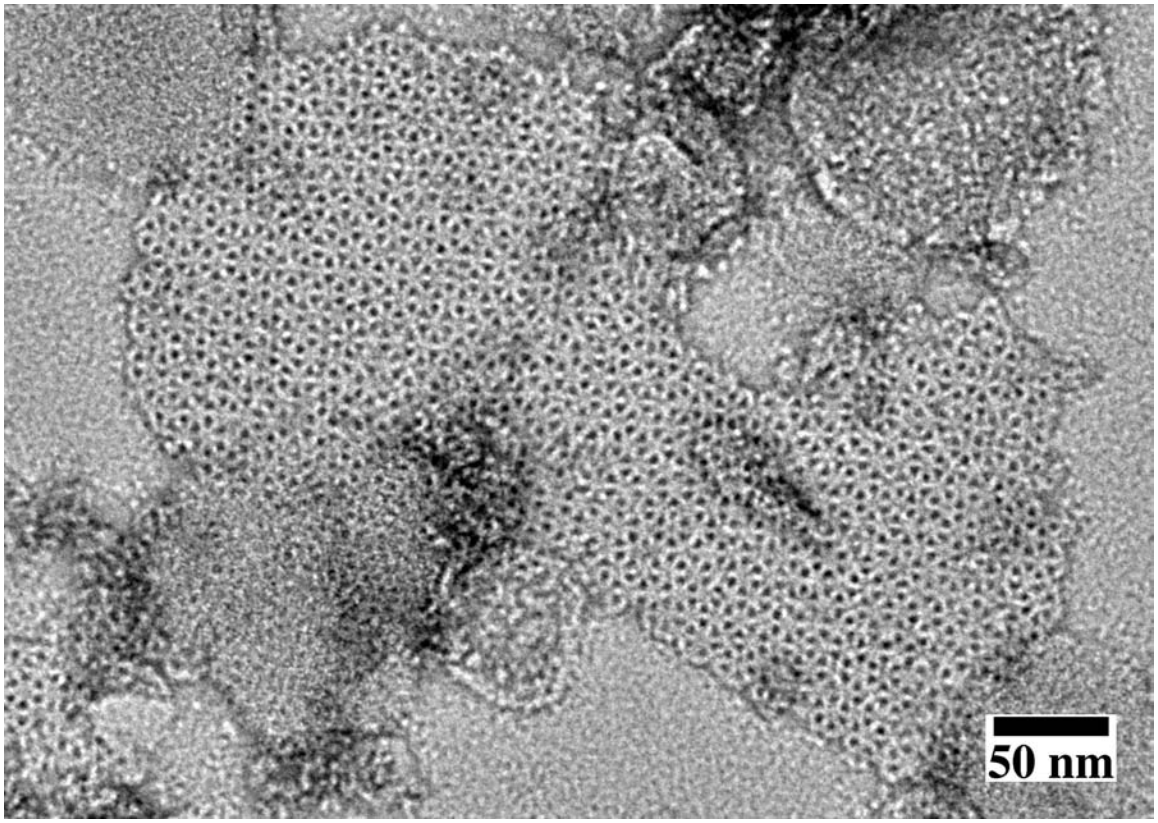
p6-9H Crystal



p6-9H crystal



Low-magnification micrograph of p6-9H crystals from a pellet sample.



p6-9H_KDKCKXX construct described in Appendix I. Clearer p6 binding could be observed but by a less stable protein construct.

Appendix III

Symmetry Definition Files, Run script and XML-Script Examples

p3Z-11

```
symmetry_name 1NZAA_C3__p31m
E = 2*VRT0_1_1 + 1*(VRT0_1_1:VRT0_2_1) + 1*(VRT0_1_1:VRT3_2_1) +
1*(VRT0_1_1:VRT2_2_1) + 1*(VRT0_1_1:VRT1_3_1) + 1*(VRT0_1_1:VRT0_3_1) +
1*(VRT0_1_1:VRT2_1_1) + 1*(VRT0_1_1:VRT3_3_1) + 1*(VRT0_1_1:VRT1_2_1) +
1*(VRT0_1_1:VRT3_1_1) + 1*(VRT0_1_1:VRT2_3_1) + 1*(VRT0_1_1:VRT1_1_1)
anchor_residue COM
virtual_coordinates_start
xyz VRT0  1.000000,0.000000,0.000000  0.000000,1.000000,0.000000
56.703625,55.703625,27.299375
xyz VRT0_ctrl  -0.229714,0.793402,0.563688  0.783517,-0.192820,0.590696
55.703625,55.703625,27.299375
xyz VRT0_1_1  -0.229714,0.793402,0.563688  0.783517,-0.192820,0.590696
55.703625,55.703625,27.299375
xyz VRT0_2_1  0.793402,-0.563688,0.229714  -0.192820,-0.590696,-0.783517
55.703625,55.703625,27.299375
xyz VRT0_3_1  -0.563688,-0.229714,-0.793402  -0.590696,0.783517,0.192820
55.703625,55.703625,27.299375
xyz VRT1_outer  0.793402,-0.563688,0.229714  -0.192820,-0.590696,-0.783517
55.703625,55.703625,27.299375
xyz VRT1_redir  0.793402,-0.563688,0.229714  -0.192820,-0.590696,-0.783517  -
23.636594,112.072456,4.327988
xyz VRT1_ctrl  -0.793402,0.563688,-0.229714  -0.192820,-0.590696,-0.783517  -
23.636594,112.072456,4.327988
xyz VRT2_outer  -0.563688,-0.229714,-0.793402  -0.590696,0.783517,0.192820
55.703625,55.703625,27.299375
xyz VRT2_redir  -0.563688,-0.229714,-0.793402  -0.590696,0.783517,0.192820
112.072456,78.675012,106.639594
xyz VRT2_ctrl  -0.793402,0.563688,-0.229714  -0.192820,-0.590696,-0.783517
112.072456,78.675012,106.639594
xyz VRT3_outer  -0.229714,0.793402,0.563688  0.783517,-0.192820,0.590696
55.703625,55.703625,27.299375
xyz VRT3_redir  -0.229714,0.793402,0.563688  0.783517,-0.192820,0.590696
78.675012,-23.636594,-29.069456
xyz VRT3_ctrl  -0.793402,0.563688,-0.229714  -0.192820,-0.590696,-0.783517
78.675012,-23.636594,-29.069456
xyz VRT1_1_1  -0.793402,0.563688,-0.229714  -0.192820,-0.590696,-0.783517  -
23.636594,112.072456,4.327988
xyz VRT1_2_1  0.563688,0.229714,0.793402  -0.590696,0.783517,0.192820  -
23.636594,112.072456,4.327988
```

xyz VRT1_3_1 0.229714,-0.793402,-0.563688 0.783517,-0.192820,0.590696 -
23.636594,112.072456,4.327988
xyz VRT2_1_1 -0.793402,0.563688,-0.229714 -0.192820,-0.590696,-0.783517
112.072456,78.675012,106.639594
xyz VRT2_2_1 0.563688,0.229714,0.793402 -0.590696,0.783517,0.192820
112.072456,78.675012,106.639594
xyz VRT2_3_1 0.229714,-0.793402,-0.563688 0.783517,-0.192820,0.590696
112.072456,78.675012,106.639594
xyz VRT3_1_1 -0.793402,0.563688,-0.229714 -0.192820,-0.590696,-0.783517
78.675012,-23.636594,-29.069456
xyz VRT3_2_1 0.563688,0.229714,0.793402 -0.590696,0.783517,0.192820
78.675012,-23.636594,-29.069456
xyz VRT3_3_1 0.229714,-0.793402,-0.563688 0.783517,-0.192820,0.590696
78.675012,-23.636594,-29.069456
virtual_coordinates_stop
connect_virtual JUMP0 VRT0 VRT0_ctrl
connect_virtual JUMP0_1_1 VRT0_ctrl VRT0_1_1
connect_virtual JUMP0_1_1_to_subunit VRT0_1_1 SUBUNIT
connect_virtual JUMP0_2_1 VRT0_1_1 VRT0_2_1
connect_virtual JUMP0_2_1_to_subunit VRT0_2_1 SUBUNIT
connect_virtual JUMP0_3_1 VRT0_1_1 VRT0_3_1
connect_virtual JUMP0_3_1_to_subunit VRT0_3_1 SUBUNIT
connect_virtual JUMP1_to_outer VRT0_ctrl VRT1_outer
connect_virtual JUMP1_to_redir VRT1_outer VRT1_redir
connect_virtual JUMP1_to_ctrl VRT1_redir VRT1_ctrl
connect_virtual JUMP1_1_1 VRT1_ctrl VRT1_1_1
connect_virtual JUMP1_1_1_to_subunit VRT1_1_1 SUBUNIT
connect_virtual JUMP1_2_1 VRT1_1_1 VRT1_2_1
connect_virtual JUMP1_2_1_to_subunit VRT1_2_1 SUBUNIT
connect_virtual JUMP1_3_1 VRT1_1_1 VRT1_3_1
connect_virtual JUMP1_3_1_to_subunit VRT1_3_1 SUBUNIT
connect_virtual JUMP2_to_outer VRT0_ctrl VRT2_outer
connect_virtual JUMP2_to_redir VRT2_outer VRT2_redir
connect_virtual JUMP2_to_ctrl VRT2_redir VRT2_ctrl
connect_virtual JUMP2_1_1 VRT2_ctrl VRT2_1_1
connect_virtual JUMP2_1_1_to_subunit VRT2_1_1 SUBUNIT
connect_virtual JUMP2_2_1 VRT2_1_1 VRT2_2_1
connect_virtual JUMP2_2_1_to_subunit VRT2_2_1 SUBUNIT
connect_virtual JUMP2_3_1 VRT2_1_1 VRT2_3_1
connect_virtual JUMP2_3_1_to_subunit VRT2_3_1 SUBUNIT
connect_virtual JUMP3_to_outer VRT0_ctrl VRT3_outer
connect_virtual JUMP3_to_redir VRT3_outer VRT3_redir
connect_virtual JUMP3_to_ctrl VRT3_redir VRT3_ctrl
connect_virtual JUMP3_1_1 VRT3_ctrl VRT3_1_1
connect_virtual JUMP3_1_1_to_subunit VRT3_1_1 SUBUNIT
connect_virtual JUMP3_2_1 VRT3_1_1 VRT3_2_1

connect_virtual JUMP3_2_1_to_subunit VRT3_2_1 SUBUNIT
connect_virtual JUMP3_3_1 VRT3_1_1 VRT3_3_1
connect_virtual JUMP3_3_1_to_subunit VRT3_3_1 SUBUNIT
set_dof JUMP0_1_1 angle_z(0:360) z
set_dof JUMP1_to_redir x(100)
set_jump_group JUMPGROUP1 JUMP1_to_redir JUMP2_to_redir JUMP3_to_redir
set_jump_group JUMPGROUP2 JUMP0_1_1 JUMP1_1_1 JUMP2_1_1 JUMP3_1_1
set_jump_group JUMPGROUP3 JUMP0_1_1_to_subunit JUMP0_2_1_to_subunit
JUMP0_3_1_to_subunit JUMP1_1_1_to_subunit JUMP1_2_1_to_subunit
JUMP1_3_1_to_subunit JUMP2_1_1_to_subunit JUMP2_2_1_to_subunit
JUMP2_3_1_to_subunit JUMP3_1_1_to_subunit JUMP3_2_1_to_subunit
JUMP3_3_1_to_subunit
slide_type ORDERED_SEQUENTIAL
slide_order JUMP1_to_redir

p3Z-42

symmetry_name 4AA7A_C3__p31m
E = 2*VRT0_1_1 + 1*(VRT0_1_1:VRT0_2_1) + 1*(VRT0_1_1:VRT3_2_1) +
1*(VRT0_1_1:VRT2_2_1) + 1*(VRT0_1_1:VRT1_3_1) + 1*(VRT0_1_1:VRT0_3_1) +
1*(VRT0_1_1:VRT2_1_1) + 1*(VRT0_1_1:VRT3_3_1) + 1*(VRT0_1_1:VRT1_2_1) +
1*(VRT0_1_1:VRT3_1_1) + 1*(VRT0_1_1:VRT2_3_1) + 1*(VRT0_1_1:VRT1_1_1)
anchor_residue COM
virtual_coordinates_start
xyz VRT0 1.000000,0.000000,0.000000 0.000000,1.000000,0.000000
41.356535,23.299797,-0.726986
xyz VRT0_ctrl 0.432380,-0.901692,0.000000 0.901692,0.432380,-0.000000
40.356535,23.299797,-0.726986
xyz VRT0_1_1 0.432380,-0.901692,0.000000 0.901692,0.432380,-0.000000
40.356535,23.299797,-0.726986
xyz VRT0_2_1 0.564698,0.825298,-0.000000 -0.825298,0.564698,-0.000000
40.356535,23.299797,-0.726986
xyz VRT0_3_1 -0.997078,0.076394,-0.000000 -0.076394,-0.997078,0.000000
40.356535,23.299797,-0.726986
xyz VRT1_outer 0.564698,0.825298,-0.000000 -0.825298,0.564698,-0.000000
40.356535,23.299797,-0.726986
xyz VRT1_redir 0.564698,0.825298,-0.000000 -0.825298,0.564698,-0.000000 -
16.113250,-59.229975,-0.726955
xyz VRT1_ctrl -0.564698,-0.825298,0.000000 -0.825298,0.564698,-0.000000 -
16.113250,-59.229975,-0.726955
xyz VRT2_outer -0.997078,0.076394,-0.000000 -0.076394,-0.997078,0.000000
40.356535,23.299797,-0.726986
xyz VRT2_redir -0.997078,0.076394,-0.000000 -0.076394,-0.997078,0.000000
140.064308,15.660415,-0.726971

```
xyz VRT2_ctrl  -0.564698,-0.825298,0.000000  -0.825298,0.564698,-0.000000
140.064308,15.660415,-0.726971
xyz VRT3_outer  0.432380,-0.901692,0.000000  0.901692,0.432380,-0.000000
40.356535,23.299797,-0.726986
xyz VRT3_redir  0.432380,-0.901692,0.000000  0.901692,0.432380,-0.000000  -
2.881452,113.468952,-0.727031
xyz VRT3_ctrl  -0.564698,-0.825298,0.000000  -0.825298,0.564698,-0.000000  -
2.881452,113.468952,-0.727031
xyz VRT1_1_1  -0.564698,-0.825298,0.000000  -0.825298,0.564698,-0.000000  -
16.113250,-59.229975,-0.726955
xyz VRT1_2_1  0.997078,-0.076394,0.000000  -0.076394,-0.997078,0.000000  -
16.113250,-59.229975,-0.726955
xyz VRT1_3_1  -0.432380,0.901692,-0.000000  0.901692,0.432380,-0.000000  -
16.113250,-59.229975,-0.726955
xyz VRT2_1_1  -0.564698,-0.825298,0.000000  -0.825298,0.564698,-0.000000
140.064308,15.660415,-0.726971
xyz VRT2_2_1  0.997078,-0.076394,0.000000  -0.076394,-0.997078,0.000000
140.064308,15.660415,-0.726971
xyz VRT2_3_1  -0.432380,0.901692,-0.000000  0.901692,0.432380,-0.000000
140.064308,15.660415,-0.726971
xyz VRT3_1_1  -0.564698,-0.825298,0.000000  -0.825298,0.564698,-0.000000  -
2.881452,113.468952,-0.727031
xyz VRT3_2_1  0.997078,-0.076394,0.000000  -0.076394,-0.997078,0.000000  -
2.881452,113.468952,-0.727031
xyz VRT3_3_1  -0.432380,0.901692,-0.000000  0.901692,0.432380,-0.000000  -
2.881452,113.468952,-0.727031
virtual_coordinates_stop
connect_virtual JUMP0 VRT0 VRT0_ctrl
connect_virtual JUMP0_1_1 VRT0_ctrl VRT0_1_1
connect_virtual JUMP0_1_1_to_subunit VRT0_1_1 SUBUNIT
connect_virtual JUMP0_2_1 VRT0_1_1 VRT0_2_1
connect_virtual JUMP0_2_1_to_subunit VRT0_2_1 SUBUNIT
connect_virtual JUMP0_3_1 VRT0_1_1 VRT0_3_1
connect_virtual JUMP0_3_1_to_subunit VRT0_3_1 SUBUNIT
connect_virtual JUMP1_to_outer  VRT0_ctrl VRT1_outer
connect_virtual JUMP1_to_redir VRT1_outer VRT1_redir
connect_virtual JUMP1_to_ctrl VRT1_redir VRT1_ctrl
connect_virtual JUMP1_1_1 VRT1_ctrl VRT1_1_1
connect_virtual JUMP1_1_1_to_subunit VRT1_1_1 SUBUNIT
connect_virtual JUMP1_2_1 VRT1_1_1 VRT1_2_1
connect_virtual JUMP1_2_1_to_subunit VRT1_2_1 SUBUNIT
connect_virtual JUMP1_3_1 VRT1_1_1 VRT1_3_1
connect_virtual JUMP1_3_1_to_subunit VRT1_3_1 SUBUNIT
connect_virtual JUMP2_to_outer  VRT0_ctrl VRT2_outer
connect_virtual JUMP2_to_redir VRT2_outer VRT2_redir
connect_virtual JUMP2_to_ctrl VRT2_redir VRT2_ctrl
```

connect_virtual JUMP2_1_1 VRT2_ctrl VRT2_1_1
connect_virtual JUMP2_1_1_to_subunit VRT2_1_1 SUBUNIT
connect_virtual JUMP2_2_1 VRT2_1_1 VRT2_2_1
connect_virtual JUMP2_2_1_to_subunit VRT2_2_1 SUBUNIT
connect_virtual JUMP2_3_1 VRT2_1_1 VRT2_3_1
connect_virtual JUMP2_3_1_to_subunit VRT2_3_1 SUBUNIT
connect_virtual JUMP3_to_outer VRT0_ctrl VRT3_outer
connect_virtual JUMP3_to_redir VRT3_outer VRT3_redir
connect_virtual JUMP3_to_ctrl VRT3_redir VRT3_ctrl
connect_virtual JUMP3_1_1 VRT3_ctrl VRT3_1_1
connect_virtual JUMP3_1_1_to_subunit VRT3_1_1 SUBUNIT
connect_virtual JUMP3_2_1 VRT3_1_1 VRT3_2_1
connect_virtual JUMP3_2_1_to_subunit VRT3_2_1 SUBUNIT
connect_virtual JUMP3_3_1 VRT3_1_1 VRT3_3_1
connect_virtual JUMP3_3_1_to_subunit VRT3_3_1 SUBUNIT
set_dof JUMP0_1_1 angle_z(0:360) z
set_dof JUMP1_to_redir x(100)
set_jump_group JUMPGROUP1 JUMP1_to_redir JUMP2_to_redir JUMP3_to_redir
set_jump_group JUMPGROUP2 JUMP0_1_1 JUMP1_1_1 JUMP2_1_1 JUMP3_1_1
set_jump_group JUMPGROUP3 JUMP0_1_1_to_subunit JUMP0_2_1_to_subunit
JUMP0_3_1_to_subunit JUMP1_1_1_to_subunit JUMP1_2_1_to_subunit
JUMP1_3_1_to_subunit JUMP2_1_1_to_subunit JUMP2_2_1_to_subunit
JUMP2_3_1_to_subunit JUMP3_1_1_to_subunit JUMP3_2_1_to_subunit
JUMP3_3_1_to_subunit
slide_type ORDERED_SEQUENTIAL
slide_order JUMP1_to_redir

p4Z-9

symmetry_name 3AEIA_C4__p4g
E = 2*VRT0_1_1 + 1*(VRT0_1_1:VRT0_2_1) + 1*(VRT0_1_1:VRT3_2_1) +
1*(VRT0_1_1:VRT0_4_1) + 1*(VRT0_1_1:VRT2_2_1) + 1*(VRT0_1_1:VRT1_3_1) +
1*(VRT0_1_1:VRT0_3_1) + 1*(VRT0_1_1:VRT4_4_1) + 1*(VRT0_1_1:VRT1_2_1) +
1*(VRT0_1_1:VRT3_1_1) + 1*(VRT0_1_1:VRT1_4_1) + 1*(VRT0_1_1:VRT2_3_1) +
1*(VRT0_1_1:VRT2_1_1) + 1*(VRT0_1_1:VRT3_3_1) + 1*(VRT0_1_1:VRT4_3_1) +
1*(VRT0_1_1:VRT4_1_1) + 1*(VRT0_1_1:VRT4_2_1) + 1*(VRT0_1_1:VRT3_4_1) +
1*(VRT0_1_1:VRT2_4_1) + 1*(VRT0_1_1:VRT1_1_1)
anchor_residue COM
virtual_coordinates_start
xyz VRT0 1.000000,0.000000,0.000000 0.000000,1.000000,0.000000
34.657441,0.000090,8.516077
xyz VRT0_ctrl -0.806805,0.590817,0.000013 -0.590817,-0.806805,-0.000018
33.657441,0.000090,8.516077
xyz VRT0_1_1 -0.806805,0.590817,0.000013 -0.590817,-0.806805,-0.000018
33.657441,0.000090,8.516077

xyz VRT0_2_1 -0.590817,-0.806805,-0.000018 0.806805,-0.590817,-0.000013
33.657441,0.000090,8.516077
xyz VRT0_3_1 0.806805,-0.590817,-0.000013 0.590817,0.806805,0.000018
33.657441,0.000090,8.516077
xyz VRT0_4_1 0.590817,0.806805,0.000018 -0.806805,0.590817,0.000013
33.657441,0.000090,8.516077
xyz VRT1_outer 0.152726,-0.988269,-0.000022 0.988269,0.152726,0.000004
33.657441,0.000090,8.516077
xyz VRT1_redir 0.152726,-0.988269,-0.000022 0.988269,0.152726,0.000004
18.384799,98.826941,8.518265
xyz VRT1_ctrl 0.590817,0.806805,0.000018 0.806805,-0.590817,-0.000013
18.384799,98.826941,8.518265
xyz VRT2_outer 0.988269,0.152726,0.000004 -0.152726,0.988269,0.000022
33.657441,0.000090,8.516077
xyz VRT2_redir 0.988269,0.152726,0.000004 -0.152726,0.988269,0.000022 -
65.169409,-15.272552,8.515722
xyz VRT2_ctrl 0.590817,0.806805,0.000018 0.806805,-0.590817,-0.000013 -
65.169409,-15.272552,8.515722
xyz VRT3_outer -0.152726,0.988269,0.000022 -0.988269,-0.152726,-0.000004
33.657441,0.000090,8.516077
xyz VRT3_redir -0.152726,0.988269,0.000022 -0.988269,-0.152726,-0.000004
48.930084,-98.826760,8.513889
xyz VRT3_ctrl 0.590817,0.806805,0.000018 0.806805,-0.590817,-0.000013
48.930084,-98.826760,8.513889
xyz VRT4_outer -0.988269,-0.152726,-0.000004 0.152726,-0.988269,-0.000022
33.657441,0.000090,8.516077
xyz VRT4_redir -0.988269,-0.152726,-0.000004 0.152726,-0.988269,-0.000022
132.484292,15.272733,8.516432
xyz VRT4_ctrl 0.590817,0.806805,0.000018 0.806805,-0.590817,-0.000013
132.484292,15.272733,8.516432
xyz VRT1_1_1 0.590817,0.806805,0.000018 0.806805,-0.590817,-0.000013
18.384799,98.826941,8.518265
xyz VRT1_2_1 -0.806805,0.590817,0.000013 0.590817,0.806805,0.000018
18.384799,98.826941,8.518265
xyz VRT1_3_1 -0.590817,-0.806805,-0.000018 -0.806805,0.590817,0.000013
18.384799,98.826941,8.518265
xyz VRT1_4_1 0.806805,-0.590817,-0.000013 -0.590817,-0.806805,-0.000018
18.384799,98.826941,8.518265
xyz VRT2_1_1 0.590817,0.806805,0.000018 0.806805,-0.590817,-0.000013 -
65.169409,-15.272552,8.515722
xyz VRT2_2_1 -0.806805,0.590817,0.000013 0.590817,0.806805,0.000018 -
65.169409,-15.272552,8.515722
xyz VRT2_3_1 -0.590817,-0.806805,-0.000018 -0.806805,0.590817,0.000013 -
65.169409,-15.272552,8.515722
xyz VRT2_4_1 0.806805,-0.590817,-0.000013 -0.590817,-0.806805,-0.000018 -
65.169409,-15.272552,8.515722

xyz VRT3_1_1 0.590817,0.806805,0.000018 0.806805,-0.590817,-0.000013
48.930084,-98.826760,8.513889
xyz VRT3_2_1 -0.806805,0.590817,0.000013 0.590817,0.806805,0.000018
48.930084,-98.826760,8.513889
xyz VRT3_3_1 -0.590817,-0.806805,-0.000018 -0.806805,0.590817,0.000013
48.930084,-98.826760,8.513889
xyz VRT3_4_1 0.806805,-0.590817,-0.000013 -0.590817,-0.806805,-0.000018
48.930084,-98.826760,8.513889
xyz VRT4_1_1 0.590817,0.806805,0.000018 0.806805,-0.590817,-0.000013
132.484292,15.272733,8.516432
xyz VRT4_2_1 -0.806805,0.590817,0.000013 0.590817,0.806805,0.000018
132.484292,15.272733,8.516432
xyz VRT4_3_1 -0.590817,-0.806805,-0.000018 -0.806805,0.590817,0.000013
132.484292,15.272733,8.516432
xyz VRT4_4_1 0.806805,-0.590817,-0.000013 -0.590817,-0.806805,-0.000018
132.484292,15.272733,8.516432
virtual_coordinates_stop
connect_virtual JUMP0 VRT0 VRT0_ctrl
connect_virtual JUMP0_1_1 VRT0_ctrl VRT0_1_1
connect_virtual JUMP0_1_1_to_subunit VRT0_1_1 SUBUNIT
connect_virtual JUMP0_2_1 VRT0_1_1 VRT0_2_1
connect_virtual JUMP0_2_1_to_subunit VRT0_2_1 SUBUNIT
connect_virtual JUMP0_3_1 VRT0_1_1 VRT0_3_1
connect_virtual JUMP0_3_1_to_subunit VRT0_3_1 SUBUNIT
connect_virtual JUMP0_4_1 VRT0_1_1 VRT0_4_1
connect_virtual JUMP0_4_1_to_subunit VRT0_4_1 SUBUNIT
connect_virtual JUMP1_to_outer VRT0_ctrl VRT1_outer
connect_virtual JUMP1_to_redir VRT1_outer VRT1_redir
connect_virtual JUMP1_to_ctrl VRT1_redir VRT1_ctrl
connect_virtual JUMP1_1_1 VRT1_ctrl VRT1_1_1
connect_virtual JUMP1_1_1_to_subunit VRT1_1_1 SUBUNIT
connect_virtual JUMP1_2_1 VRT1_1_1 VRT1_2_1
connect_virtual JUMP1_2_1_to_subunit VRT1_2_1 SUBUNIT
connect_virtual JUMP1_3_1 VRT1_1_1 VRT1_3_1
connect_virtual JUMP1_3_1_to_subunit VRT1_3_1 SUBUNIT
connect_virtual JUMP1_4_1 VRT1_1_1 VRT1_4_1
connect_virtual JUMP1_4_1_to_subunit VRT1_4_1 SUBUNIT
connect_virtual JUMP2_to_outer VRT0_ctrl VRT2_outer
connect_virtual JUMP2_to_redir VRT2_outer VRT2_redir
connect_virtual JUMP2_to_ctrl VRT2_redir VRT2_ctrl
connect_virtual JUMP2_1_1 VRT2_ctrl VRT2_1_1
connect_virtual JUMP2_1_1_to_subunit VRT2_1_1 SUBUNIT
connect_virtual JUMP2_2_1 VRT2_1_1 VRT2_2_1
connect_virtual JUMP2_2_1_to_subunit VRT2_2_1 SUBUNIT
connect_virtual JUMP2_3_1 VRT2_1_1 VRT2_3_1
connect_virtual JUMP2_3_1_to_subunit VRT2_3_1 SUBUNIT

```
connect_virtual JUMP2_4_1 VRT2_1_1 VRT2_4_1
connect_virtual JUMP2_4_1_to_subunit VRT2_4_1 SUBUNIT
connect_virtual JUMP3_to_outer  VRT0_ctrl VRT3_outer
connect_virtual JUMP3_to_redir VRT3_outer VRT3_redir
connect_virtual JUMP3_to_ctrl VRT3_redir VRT3_ctrl
connect_virtual JUMP3_1_1 VRT3_ctrl VRT3_1_1
connect_virtual JUMP3_1_1_to_subunit VRT3_1_1 SUBUNIT
connect_virtual JUMP3_2_1 VRT3_1_1 VRT3_2_1
connect_virtual JUMP3_2_1_to_subunit VRT3_2_1 SUBUNIT
connect_virtual JUMP3_3_1 VRT3_1_1 VRT3_3_1
connect_virtual JUMP3_3_1_to_subunit VRT3_3_1 SUBUNIT
connect_virtual JUMP3_4_1 VRT3_1_1 VRT3_4_1
connect_virtual JUMP3_4_1_to_subunit VRT3_4_1 SUBUNIT
connect_virtual JUMP4_to_outer  VRT0_ctrl VRT4_outer
connect_virtual JUMP4_to_redir VRT4_outer VRT4_redir
connect_virtual JUMP4_to_ctrl VRT4_redir VRT4_ctrl
connect_virtual JUMP4_1_1 VRT4_ctrl VRT4_1_1
connect_virtual JUMP4_1_1_to_subunit VRT4_1_1 SUBUNIT
connect_virtual JUMP4_2_1 VRT4_1_1 VRT4_2_1
connect_virtual JUMP4_2_1_to_subunit VRT4_2_1 SUBUNIT
connect_virtual JUMP4_3_1 VRT4_1_1 VRT4_3_1
connect_virtual JUMP4_3_1_to_subunit VRT4_3_1 SUBUNIT
connect_virtual JUMP4_4_1 VRT4_1_1 VRT4_4_1
connect_virtual JUMP4_4_1_to_subunit VRT4_4_1 SUBUNIT
set_dof JUMP0_1_1 angle_z(0:360) z[0;-5:5]
set_dof JUMP1_to_redir x(100)
set_jump_group JUMPGROUP1 JUMP1_to_redir JUMP2_to_redir JUMP3_to_redir
JUMP4_to_redir
set_jump_group JUMPGROUP2 JUMP0_1_1 JUMP1_1_1 JUMP2_1_1 JUMP3_1_1
JUMP4_1_1
set_jump_group JUMPGROUP3 JUMP0_1_1_to_subunit JUMP0_2_1_to_subunit
JUMP0_3_1_to_subunit JUMP0_4_1_to_subunit JUMP1_1_1_to_subunit
JUMP1_2_1_to_subunit JUMP1_3_1_to_subunit JUMP1_4_1_to_subunit
JUMP2_1_1_to_subunit JUMP2_2_1_to_subunit JUMP2_3_1_to_subunit
JUMP2_4_1_to_subunit JUMP3_1_1_to_subunit JUMP3_2_1_to_subunit
JUMP3_3_1_to_subunit JUMP3_4_1_to_subunit JUMP4_1_1_to_subunit
JUMP4_2_1_to_subunit JUMP4_3_1_to_subunit JUMP4_4_1_to_subunit
slide_type ORDERED_SEQUENTIAL
slide_order JUMP1_to_redir
```

p6-9

symmetry_name 2DI4A_C6__p6
E = 2*VRT0_1_1 + 1*(VRT0_1_1:VRT0_4_1) + 1*(VRT0_1_1:VRT1_3_1) +
1*(VRT0_1_1:VRT0_3_1) + 1*(VRT0_1_1:VRT0_5_1) + 1*(VRT0_1_1:VRT4_4_1) +
1*(VRT0_1_1:VRT3_1_1) + 1*(VRT0_1_1:VRT1_2_1) + 1*(VRT0_1_1:VRT1_4_1) +
1*(VRT0_1_1:VRT3_6_1) + 1*(VRT0_1_1:VRT6_3_1) + 1*(VRT0_1_1:VRT6_6_1) +
1*(VRT0_1_1:VRT6_5_1) + 1*(VRT0_1_1:VRT5_6_1) + 1*(VRT0_1_1:VRT4_5_1) +
1*(VRT0_1_1:VRT3_5_1) + 1*(VRT0_1_1:VRT5_1_1) + 1*(VRT0_1_1:VRT2_1_1) +
1*(VRT0_1_1:VRT1_6_1) + 1*(VRT0_1_1:VRT4_6_1) + 1*(VRT0_1_1:VRT4_1_1) +
1*(VRT0_1_1:VRT4_3_1) + 1*(VRT0_1_1:VRT3_4_1) + 1*(VRT0_1_1:VRT2_5_1) +
1*(VRT0_1_1:VRT2_4_1) + 1*(VRT0_1_1:VRT1_1_1) + 1*(VRT0_1_1:VRT5_5_1) +
1*(VRT0_1_1:VRT0_2_1) + 1*(VRT0_1_1:VRT3_2_1) + 1*(VRT0_1_1:VRT0_6_1) +
1*(VRT0_1_1:VRT2_2_1) + 1*(VRT0_1_1:VRT1_5_1) + 1*(VRT0_1_1:VRT5_3_1) +
1*(VRT0_1_1:VRT2_3_1) + 1*(VRT0_1_1:VRT6_4_1) + 1*(VRT0_1_1:VRT5_4_1) +
1*(VRT0_1_1:VRT3_3_1) + 1*(VRT0_1_1:VRT5_2_1) + 1*(VRT0_1_1:VRT6_1_1) +
1*(VRT0_1_1:VRT6_2_1) + 1*(VRT0_1_1:VRT2_6_1) + 1*(VRT0_1_1:VRT4_2_1)
anchor_residue COM
virtual_coordinates_start
xyz VRT0 1.000000,0.000000,0.000000 0.000000,1.000000,0.000000
59.399671,33.811829,37.431372
xyz VRT0_ctrl 0.113873,0.993495,-0.000647 -0.993448,0.113861,-0.009857
58.399671,33.811829,37.431372
xyz VRT0_1_1 0.113873,0.993495,-0.000647 -0.993448,0.113861,-0.009857
58.399671,33.811829,37.431372
xyz VRT0_2_1 0.917287,0.398141,0.008213 -0.398107,0.917323,-0.005489
58.399671,33.811829,37.431372
xyz VRT0_3_1 0.803415,-0.595354,0.008860 0.595341,0.803461,0.004368
58.399671,33.811829,37.431372
xyz VRT0_4_1 -0.113873,-0.993495,0.000647 0.993448,-0.113861,0.009857
58.399671,33.811829,37.431372
xyz VRT0_5_1 -0.917287,-0.398141,-0.008213 0.398107,-0.917323,0.005489
58.399671,33.811829,37.431372
xyz VRT0_6_1 -0.803415,0.595354,-0.008860 -0.595341,-0.803461,-0.004368
58.399671,33.811829,37.431372
xyz VRT1_outer -0.398107,0.917323,-0.005489 -0.917287,-0.398141,-0.008213
58.399671,33.811829,37.431372
xyz VRT1_redir -0.398107,0.917323,-0.005489 -0.917287,-0.398141,-0.008213
98.210376,-57.920425,37.980232
xyz VRT1_ctrl 0.113873,0.993495,-0.000647 -0.993448,0.113861,-0.009857
98.210376,-57.920425,37.980232
xyz VRT2_outer -0.993448,0.113861,-0.009857 -0.113873,-0.993495,0.000647
58.399671,33.811829,37.431372
xyz VRT2_redir -0.993448,0.113861,-0.009857 -0.113873,-0.993495,0.000647
157.744449,22.425722,38.417068

xyz VRT2_ctrl 0.113873,0.993495,-0.000647 -0.993448,0.113861,-0.009857
157.744449,22.425722,38.417068
xyz VRT3_outer -0.595341,-0.803461,-0.004368 0.803415,-0.595354,0.008860
58.399671,33.811829,37.431372
xyz VRT3_redir -0.595341,-0.803461,-0.004368 0.803415,-0.595354,0.008860
117.933743,114.157976,37.868209
xyz VRT3_ctrl 0.113873,0.993495,-0.000647 -0.993448,0.113861,-0.009857
117.933743,114.157976,37.868209
xyz VRT4_outer 0.398107,-0.917323,0.005489 0.917287,0.398141,0.008213
58.399671,33.811829,37.431372
xyz VRT4_redir 0.398107,-0.917323,0.005489 0.917287,0.398141,0.008213
18.588965,125.544083,36.882513
xyz VRT4_ctrl 0.113873,0.993495,-0.000647 -0.993448,0.113861,-0.009857
18.588965,125.544083,36.882513
xyz VRT5_outer 0.993448,-0.113861,0.009857 0.113873,0.993495,-0.000647
58.399671,33.811829,37.431372
xyz VRT5_redir 0.993448,-0.113861,0.009857 0.113873,0.993495,-0.000647 -
40.945108,45.197936,36.445676
xyz VRT5_ctrl 0.113873,0.993495,-0.000647 -0.993448,0.113861,-0.009857 -
40.945108,45.197936,36.445676
xyz VRT6_outer 0.595341,0.803461,0.004368 -0.803415,0.595354,-0.008860
58.399671,33.811829,37.431372
xyz VRT6_redir 0.595341,0.803461,0.004368 -0.803415,0.595354,-0.008860 -
1.134402,-46.534318,36.994535
xyz VRT6_ctrl 0.113873,0.993495,-0.000647 -0.993448,0.113861,-0.009857 -
1.134402,-46.534318,36.994535
xyz VRT1_1_1 0.113873,0.993495,-0.000647 -0.993448,0.113861,-0.009857
98.210376,-57.920425,37.980232
xyz VRT1_2_1 0.917287,0.398141,0.008213 -0.398107,0.917323,-0.005489
98.210376,-57.920425,37.980232
xyz VRT1_3_1 0.803415,-0.595354,0.008860 0.595341,0.803461,0.004368
98.210376,-57.920425,37.980232
xyz VRT1_4_1 -0.113873,-0.993495,0.000647 0.993448,-0.113861,0.009857
98.210376,-57.920425,37.980232
xyz VRT1_5_1 -0.917287,-0.398141,-0.008213 0.398107,-0.917323,0.005489
98.210376,-57.920425,37.980232
xyz VRT1_6_1 -0.803415,0.595354,-0.008860 -0.595341,-0.803461,-0.004368
98.210376,-57.920425,37.980232
xyz VRT2_1_1 0.113873,0.993495,-0.000647 -0.993448,0.113861,-0.009857
157.744449,22.425722,38.417068
xyz VRT2_2_1 0.917287,0.398141,0.008213 -0.398107,0.917323,-0.005489
157.744449,22.425722,38.417068
xyz VRT2_3_1 0.803415,-0.595354,0.008860 0.595341,0.803461,0.004368
157.744449,22.425722,38.417068
xyz VRT2_4_1 -0.113873,-0.993495,0.000647 0.993448,-0.113861,0.009857
157.744449,22.425722,38.417068

xyz VRT2_5_1 -0.917287,-0.398141,-0.008213 0.398107,-0.917323,0.005489
157.744449,22.425722,38.417068
xyz VRT2_6_1 -0.803415,0.595354,-0.008860 -0.595341,-0.803461,-0.004368
157.744449,22.425722,38.417068
xyz VRT3_1_1 0.113873,0.993495,-0.000647 -0.993448,0.113861,-0.009857
117.933743,114.157976,37.868209
xyz VRT3_2_1 0.917287,0.398141,0.008213 -0.398107,0.917323,-0.005489
117.933743,114.157976,37.868209
xyz VRT3_3_1 0.803415,-0.595354,0.008860 0.595341,0.803461,0.004368
117.933743,114.157976,37.868209
xyz VRT3_4_1 -0.113873,-0.993495,0.000647 0.993448,-0.113861,0.009857
117.933743,114.157976,37.868209
xyz VRT3_5_1 -0.917287,-0.398141,-0.008213 0.398107,-0.917323,0.005489
117.933743,114.157976,37.868209
xyz VRT3_6_1 -0.803415,0.595354,-0.008860 -0.595341,-0.803461,-0.004368
117.933743,114.157976,37.868209
xyz VRT4_1_1 0.113873,0.993495,-0.000647 -0.993448,0.113861,-0.009857
18.588965,125.544083,36.882513
xyz VRT4_2_1 0.917287,0.398141,0.008213 -0.398107,0.917323,-0.005489
18.588965,125.544083,36.882513
xyz VRT4_3_1 0.803415,-0.595354,0.008860 0.595341,0.803461,0.004368
18.588965,125.544083,36.882513
xyz VRT4_4_1 -0.113873,-0.993495,0.000647 0.993448,-0.113861,0.009857
18.588965,125.544083,36.882513
xyz VRT4_5_1 -0.917287,-0.398141,-0.008213 0.398107,-0.917323,0.005489
18.588965,125.544083,36.882513
xyz VRT4_6_1 -0.803415,0.595354,-0.008860 -0.595341,-0.803461,-0.004368
18.588965,125.544083,36.882513
xyz VRT5_1_1 0.113873,0.993495,-0.000647 -0.993448,0.113861,-0.009857 -
40.945108,45.197936,36.445676
xyz VRT5_2_1 0.917287,0.398141,0.008213 -0.398107,0.917323,-0.005489 -
40.945108,45.197936,36.445676
xyz VRT5_3_1 0.803415,-0.595354,0.008860 0.595341,0.803461,0.004368 -
40.945108,45.197936,36.445676
xyz VRT5_4_1 -0.113873,-0.993495,0.000647 0.993448,-0.113861,0.009857 -
40.945108,45.197936,36.445676
xyz VRT5_5_1 -0.917287,-0.398141,-0.008213 0.398107,-0.917323,0.005489 -
40.945108,45.197936,36.445676
xyz VRT5_6_1 -0.803415,0.595354,-0.008860 -0.595341,-0.803461,-0.004368 -
40.945108,45.197936,36.445676
xyz VRT6_1_1 0.113873,0.993495,-0.000647 -0.993448,0.113861,-0.009857 -
1.134402,-46.534318,36.994535
xyz VRT6_2_1 0.917287,0.398141,0.008213 -0.398107,0.917323,-0.005489 -
1.134402,-46.534318,36.994535
xyz VRT6_3_1 0.803415,-0.595354,0.008860 0.595341,0.803461,0.004368 -
1.134402,-46.534318,36.994535

xyz VRT6_4_1 -0.113873,-0.993495,0.000647 0.993448,-0.113861,0.009857 -1.134402,-46.534318,36.994535
xyz VRT6_5_1 -0.917287,-0.398141,-0.008213 0.398107,-0.917323,0.005489 -1.134402,-46.534318,36.994535
xyz VRT6_6_1 -0.803415,0.595354,-0.008860 -0.595341,-0.803461,-0.004368 -1.134402,-46.534318,36.994535
virtual_coordinates_stop
connect_virtual JUMP0 VRT0 VRT0_ctrl
connect_virtual JUMP0_1_1 VRT0_ctrl VRT0_1_1
connect_virtual JUMP0_1_1_to_subunit VRT0_1_1 SUBUNIT
connect_virtual JUMP0_2_1 VRT0_1_1 VRT0_2_1
connect_virtual JUMP0_2_1_to_subunit VRT0_2_1 SUBUNIT
connect_virtual JUMP0_3_1 VRT0_1_1 VRT0_3_1
connect_virtual JUMP0_3_1_to_subunit VRT0_3_1 SUBUNIT
connect_virtual JUMP0_4_1 VRT0_1_1 VRT0_4_1
connect_virtual JUMP0_4_1_to_subunit VRT0_4_1 SUBUNIT
connect_virtual JUMP0_5_1 VRT0_1_1 VRT0_5_1
connect_virtual JUMP0_5_1_to_subunit VRT0_5_1 SUBUNIT
connect_virtual JUMP0_6_1 VRT0_1_1 VRT0_6_1
connect_virtual JUMP0_6_1_to_subunit VRT0_6_1 SUBUNIT
connect_virtual JUMP1_to_outer VRT0_ctrl VRT1_outer
connect_virtual JUMP1_to_redir VRT1_outer VRT1_redir
connect_virtual JUMP1_to_ctrl VRT1_redir VRT1_ctrl
connect_virtual JUMP1_1_1 VRT1_ctrl VRT1_1_1
connect_virtual JUMP1_1_1_to_subunit VRT1_1_1 SUBUNIT
connect_virtual JUMP1_2_1 VRT1_1_1 VRT1_2_1
connect_virtual JUMP1_2_1_to_subunit VRT1_2_1 SUBUNIT
connect_virtual JUMP1_3_1 VRT1_1_1 VRT1_3_1
connect_virtual JUMP1_3_1_to_subunit VRT1_3_1 SUBUNIT
connect_virtual JUMP1_4_1 VRT1_1_1 VRT1_4_1
connect_virtual JUMP1_4_1_to_subunit VRT1_4_1 SUBUNIT
connect_virtual JUMP1_5_1 VRT1_1_1 VRT1_5_1
connect_virtual JUMP1_5_1_to_subunit VRT1_5_1 SUBUNIT
connect_virtual JUMP1_6_1 VRT1_1_1 VRT1_6_1
connect_virtual JUMP1_6_1_to_subunit VRT1_6_1 SUBUNIT
connect_virtual JUMP2_to_outer VRT0_ctrl VRT2_outer
connect_virtual JUMP2_to_redir VRT2_outer VRT2_redir
connect_virtual JUMP2_to_ctrl VRT2_redir VRT2_ctrl
connect_virtual JUMP2_1_1 VRT2_ctrl VRT2_1_1
connect_virtual JUMP2_1_1_to_subunit VRT2_1_1 SUBUNIT
connect_virtual JUMP2_2_1 VRT2_1_1 VRT2_2_1
connect_virtual JUMP2_2_1_to_subunit VRT2_2_1 SUBUNIT
connect_virtual JUMP2_3_1 VRT2_1_1 VRT2_3_1
connect_virtual JUMP2_3_1_to_subunit VRT2_3_1 SUBUNIT
connect_virtual JUMP2_4_1 VRT2_1_1 VRT2_4_1
connect_virtual JUMP2_4_1_to_subunit VRT2_4_1 SUBUNIT

```
connect_virtual JUMP2_5_1 VRT2_1_1 VRT2_5_1
connect_virtual JUMP2_5_1_to_subunit VRT2_5_1 SUBUNIT
connect_virtual JUMP2_6_1 VRT2_1_1 VRT2_6_1
connect_virtual JUMP2_6_1_to_subunit VRT2_6_1 SUBUNIT
connect_virtual JUMP3_to_outer  VRT0_ctrl VRT3_outer
connect_virtual JUMP3_to_redir VRT3_outer VRT3_redir
connect_virtual JUMP3_to_ctrl VRT3_redir VRT3_ctrl
connect_virtual JUMP3_1_1 VRT3_ctrl VRT3_1_1
connect_virtual JUMP3_1_1_to_subunit VRT3_1_1 SUBUNIT
connect_virtual JUMP3_2_1 VRT3_1_1 VRT3_2_1
connect_virtual JUMP3_2_1_to_subunit VRT3_2_1 SUBUNIT
connect_virtual JUMP3_3_1 VRT3_1_1 VRT3_3_1
connect_virtual JUMP3_3_1_to_subunit VRT3_3_1 SUBUNIT
connect_virtual JUMP3_4_1 VRT3_1_1 VRT3_4_1
connect_virtual JUMP3_4_1_to_subunit VRT3_4_1 SUBUNIT
connect_virtual JUMP3_5_1 VRT3_1_1 VRT3_5_1
connect_virtual JUMP3_5_1_to_subunit VRT3_5_1 SUBUNIT
connect_virtual JUMP3_6_1 VRT3_1_1 VRT3_6_1
connect_virtual JUMP3_6_1_to_subunit VRT3_6_1 SUBUNIT
connect_virtual JUMP4_to_outer  VRT0_ctrl VRT4_outer
connect_virtual JUMP4_to_redir VRT4_outer VRT4_redir
connect_virtual JUMP4_to_ctrl VRT4_redir VRT4_ctrl
connect_virtual JUMP4_1_1 VRT4_ctrl VRT4_1_1
connect_virtual JUMP4_1_1_to_subunit VRT4_1_1 SUBUNIT
connect_virtual JUMP4_2_1 VRT4_1_1 VRT4_2_1
connect_virtual JUMP4_2_1_to_subunit VRT4_2_1 SUBUNIT
connect_virtual JUMP4_3_1 VRT4_1_1 VRT4_3_1
connect_virtual JUMP4_3_1_to_subunit VRT4_3_1 SUBUNIT
connect_virtual JUMP4_4_1 VRT4_1_1 VRT4_4_1
connect_virtual JUMP4_4_1_to_subunit VRT4_4_1 SUBUNIT
connect_virtual JUMP4_5_1 VRT4_1_1 VRT4_5_1
connect_virtual JUMP4_5_1_to_subunit VRT4_5_1 SUBUNIT
connect_virtual JUMP4_6_1 VRT4_1_1 VRT4_6_1
connect_virtual JUMP4_6_1_to_subunit VRT4_6_1 SUBUNIT
connect_virtual JUMP5_to_outer  VRT0_ctrl VRT5_outer
connect_virtual JUMP5_to_redir VRT5_outer VRT5_redir
connect_virtual JUMP5_to_ctrl VRT5_redir VRT5_ctrl
connect_virtual JUMP5_1_1 VRT5_ctrl VRT5_1_1
connect_virtual JUMP5_1_1_to_subunit VRT5_1_1 SUBUNIT
connect_virtual JUMP5_2_1 VRT5_1_1 VRT5_2_1
connect_virtual JUMP5_2_1_to_subunit VRT5_2_1 SUBUNIT
connect_virtual JUMP5_3_1 VRT5_1_1 VRT5_3_1
connect_virtual JUMP5_3_1_to_subunit VRT5_3_1 SUBUNIT
connect_virtual JUMP5_4_1 VRT5_1_1 VRT5_4_1
connect_virtual JUMP5_4_1_to_subunit VRT5_4_1 SUBUNIT
connect_virtual JUMP5_5_1 VRT5_1_1 VRT5_5_1
```

connect_virtual JUMP5_5_1_to_subunit VRT5_5_1 SUBUNIT
connect_virtual JUMP5_6_1 VRT5_1_1 VRT5_6_1
connect_virtual JUMP5_6_1_to_subunit VRT5_6_1 SUBUNIT
connect_virtual JUMP6_to_outer VRT0_ctrl VRT6_outer
connect_virtual JUMP6_to_redir VRT6_outer VRT6_redir
connect_virtual JUMP6_to_ctrl VRT6_redir VRT6_ctrl
connect_virtual JUMP6_1_1 VRT6_ctrl VRT6_1_1
connect_virtual JUMP6_1_1_to_subunit VRT6_1_1 SUBUNIT
connect_virtual JUMP6_2_1 VRT6_1_1 VRT6_2_1
connect_virtual JUMP6_2_1_to_subunit VRT6_2_1 SUBUNIT
connect_virtual JUMP6_3_1 VRT6_1_1 VRT6_3_1
connect_virtual JUMP6_3_1_to_subunit VRT6_3_1 SUBUNIT
connect_virtual JUMP6_4_1 VRT6_1_1 VRT6_4_1
connect_virtual JUMP6_4_1_to_subunit VRT6_4_1 SUBUNIT
connect_virtual JUMP6_5_1 VRT6_1_1 VRT6_5_1
connect_virtual JUMP6_5_1_to_subunit VRT6_5_1 SUBUNIT
connect_virtual JUMP6_6_1 VRT6_1_1 VRT6_6_1
connect_virtual JUMP6_6_1_to_subunit VRT6_6_1 SUBUNIT
set_dof JUMP0_1_1 angle_z(0:360)
set_dof JUMP1_to_redir x(100)
set_jump_group JUMPGROUP1 JUMP1_to_redir JUMP2_to_redir JUMP3_to_redir
JUMP4_to_redir JUMP5_to_redir JUMP6_to_redir
set_jump_group JUMPGROUP2 JUMP0_1_1 JUMP1_1_1 JUMP2_1_1 JUMP3_1_1
JUMP4_1_1 JUMP5_1_1 JUMP6_1_1
set_jump_group JUMPGROUP3 JUMP0_1_1_to_subunit JUMP0_2_1_to_subunit
JUMP0_3_1_to_subunit JUMP0_4_1_to_subunit JUMP0_5_1_to_subunit
JUMP0_6_1_to_subunit JUMP1_1_1_to_subunit JUMP1_2_1_to_subunit
JUMP1_3_1_to_subunit JUMP1_4_1_to_subunit JUMP1_5_1_to_subunit
JUMP1_6_1_to_subunit JUMP2_1_1_to_subunit JUMP2_2_1_to_subunit
JUMP2_3_1_to_subunit JUMP2_4_1_to_subunit JUMP2_5_1_to_subunit
JUMP2_6_1_to_subunit JUMP3_1_1_to_subunit JUMP3_2_1_to_subunit
JUMP3_3_1_to_subunit JUMP3_4_1_to_subunit JUMP3_5_1_to_subunit
JUMP3_6_1_to_subunit JUMP4_1_1_to_subunit JUMP4_2_1_to_subunit
JUMP4_3_1_to_subunit JUMP4_4_1_to_subunit JUMP4_5_1_to_subunit
JUMP4_6_1_to_subunit JUMP5_1_1_to_subunit JUMP5_2_1_to_subunit
JUMP5_3_1_to_subunit JUMP5_4_1_to_subunit JUMP5_5_1_to_subunit
JUMP5_6_1_to_subunit JUMP6_1_1_to_subunit JUMP6_2_1_to_subunit
JUMP6_3_1_to_subunit JUMP6_4_1_to_subunit JUMP6_5_1_to_subunit
JUMP6_6_1_to_subunit

Rosettascripts Run Command Example

```
#!/bin/sh

$ROSETTA3/source/bin/rosetta_scripts.default.linuxgccrelease \
        -database $ROSETTA3/database \
        -parser:protocol dockdesign.xml \
        -s $1 \
        -in::file::native $1 \
        -nstruct 5 \
        -mute core.util.switchresiduetypeset \
        -parser::script_vars symmdef=$2 nbblock=$3 \
        -matdes::num_subs_building_block $3 \
        -initialize_rigid_body_dofs \
        -docking::dock_lowres_filter 100 10 1 \
        -out:file:silent $1.silent \
        -out:file:silent_struct_type binary \
        -chemical:exclude_patches LowerDNA  UpperDNA Cterm_amidation
SpecialRotamer  VirtualBB ShoveBB VirtualDNAPhosphate VirtualNTerm
CTermConnect sc_orbitals pro_hydroxylated_case1 pro_hydroxylated_case2
ser_phosphorylated thr_phosphorylated  tyr_phosphorylated tyr_sulfated
lys_dimethylated lys_monomethylated  lys_trimethylated lys_acetylated
glu_carboxylated cys_acetylated tyr_diiodinated N_acetylated C_methylamidated
MethylatedProteinCterm \
        -ignore_unrecognized_res
```

Rosettascripts Example Docking and Design Script

```
<ROSETTASCRIPTS>

        <SCOREFXNS>
                <talaris_symm weights=talaris2013 symmetric=1>
                </talaris_symm>
                <talaris_symm_cst weights=talaris2013 symmetric=1>
                        <Reweight scoretype=res_type_constraint weight=1.0 />
                </talaris_symm_cst>
                <soft_rep_symm_cst weights=soft_rep_talaris2013 symmetric=1>
                        <Reweight scoretype=res_type_constraint weight=1.0 />
                </soft_rep_symm_cst>
        </SCOREFXNS>

        <TASKOPERATIONS>
                <InitializeFromCommandline name=init />
                <IncludeCurrent name=ic />
```

```
		<DisallowIfNonnative name=disallow_aa
disallow_aas=CGPHWFMQERKDY />
		<BuildingBlockInterface name=bbi nsub_bblock=%%nbblock%%
contact_dist=12.0 sym_dof_names=JUMP1_to_ctrl />
		<SelectBySASA name=rtns mode="sc" state="monomer"
probe_radius=2.2 core_asa=0 surface_asa=30 core=0 boundary=1 surface=1 verbose=1
/>
		<RestrictIdentities name=pgr identities=PRO,GLY
prevent_repacking=0/>
		<LimitAromaChi2 name=limitaro chi2max=110 chi2min=70 />
		<RetrieveStoredTask name=matdes_design_task
task_name="design_task" />
		<RetrieveStoredTask name=revert_task task_name="revert_task" />
		<JointSequence name=joint use_current=1 use_native=1 />
		<RestrictToRepacking name=repack_only />
	</TASKOPERATIONS>

	<FILTERS>
		<SaveResfileToDisk name=save_resfile1
task_operations=matdes_design_task designable_only=0
resfile_general_property="NATRO" />

		<ResidueIE name=wie scorefxn=talaris_symm energy_cutoff=-6
restype3=TRP interface=1 whole_pose=0 interface_distance_cutoff=10.0/>
		<ResidueIE name=yie scorefxn=talaris_symm energy_cutoff=-6
restype3=TYR interface=1 whole_pose=0 interface_distance_cutoff=10.0/>
		<ResidueIE name=fie scorefxn=talaris_symm energy_cutoff=-5
restype3=PHE interface=1 whole_pose=0 interface_distance_cutoff=10.0/>

		<ShapeComplementarity name=sc jump=1 verbose=1 min_sc=0.5 />
write_int_area=1
		<OligomericAverageDegree name=avg_deg threshold=0
distance_threshold=10.0 task_operations=matdes_design_task /> write2pdb=1
		<Ddg name=ddG_filt scorefxn=talaris_symm jump=1 repack=1 repeats=2
threshold=-10.0 />
		<SymUnsatHbonds name=unsat_pols jump=1 cutoff=20 verbose=1
write2pdb=1 />

		<DesignableResidues name=des_pos
task_operations=matdes_design_task designable=1 packable=1 />
		<Sasa name=sasa_int_area threshold=600
upper_threshold=1000000000000000 hydrophobic=0 polar=0 jump=1 />

		<InterfacePacking name=packing lower_cutoff=-5 upper_cutoff=5
distance_cutoff=9.0 sym_dof_names=JUMP1_to_ctrl/>
```

```
		<TaskAwareScoreType name=air_energy
task_operations=matdes_design_task scorefxn=talaris_symm threshold=-1 bb_bb=0 />
		<TaskAwareScoreType name=air_fa_atr
task_operations=matdes_design_task scorefxn=talaris_symm score_type=fa_atr
threshold=1000 bb_bb=0 />
		<TaskAwareScoreType name=air_fa_rep
task_operations=matdes_design_task scorefxn=talaris_symm score_type=fa_rep
threshold=1000 bb_bb=0 />
		<TaskAwareScoreType name=air_fa_dun
task_operations=matdes_design_task scorefxn=talaris_symm score_type=fa_dun
threshold=1000 bb_bb=0 />

		<SequenceRecovery name=mutations rate_threshold=0.0
mutation_threshold=100 report_mutations=1 verbose=1
task_operations=matdes_design_task /> write2pdb=1

		<AtomicContactCount name=cc_jump partition=jump
sym_dof_name=JUMP0_1_1 ss_only=1 normalize_by_sasa=0 distance=7 confidence=0
/>
		<Range name=cc_jump_valid filter=cc_jump lower_bound=10
upper_bound=9999/>
		<ScoreType name=constraints scorefxn=talaris_symm_cst
score_type=res_type_constraint threshold=0 confidence=0/>
		<ScoreType name=total_score scorefxn=talaris_symm
score_type=total_score threshold=0 confidence=0/>
	</FILTERS>

	<MOVERS>
		<SetupForSymmetry name=setup_symm definition="%%symmdef%%"
/>

		<StoreTaskMover name=store_design_task task_name="design_task"
task_operations=init,ic,bbi,rtns,pgr,limitaro,disallow_aa />
		<StoreTaskMover name=store_revert_task task_name="revert_task"
task_operations=init,ic,bbi,rtns,pgr,limitaro />

		<SymPackRotamersMover name=design scorefxn=talaris_symm_cst
task_operations=matdes_design_task />
		<SymPackRotamersMover name=design_soft_rep
scorefxn=soft_rep_symm_cst task_operations=matdes_design_task />
		<SymPackRotamersMover name=repack scorefxn=talaris_symm
task_operations=matdes_design_task,repack_only />

		<TaskAwareSymMinMover name=min scorefxn=talaris_symm bb=0
chi=1 rb=1 task_operations=revert_task />
```

```
		<FavorSequenceProfile name=fsp scaling=prob use_native=1
matrix=IDENTITY weight=1.0 scorefxns=talaris_symm_cst />

		<SymPackRotamersMover name=repack_all scorefxn=talaris_symm
task_operations=ic,repack_only />
		<SymMinMover name=min_all scorefxn=talaris_symm bb=0 chi=1
jump=ALL />
		<SymMinMover name=min_rb scorefxn=talaris_symm bb=0 chi=0
jump=ALL />

		<ParsedProtocol name=design_process>
			<Add mover_name=fsp />
			<Add mover_name=design_soft_rep />
			<Add mover_name=min />
			<Add mover_name=design />
			<Add mover_name=min />
		</ParsedProtocol>

		<MakePolyX name="makepolya" aa="ALA" keep_pro=1  keep_gly=1
keep_disulfide_cys=1 />
		<SaveAndRetrieveSidechains name=sidechains allsc=1/>
		<Docking name=dock fullatom=0 local_refine=0 symmetry=1 />
		<SwitchResidueTypeSetMover name="tocen" set=centroid/>
		<SwitchResidueTypeSetMover name="tofa" set=fa_standard/>
		<GenericMonteCarlo name=mc_dock trials=10 mover_name=dock
filter_name=cc_jump preapply=1 drift=0 sample_type=high/>
	</MOVERS>

	<APPLY_TO_POSE>
	</APPLY_TO_POSE>

	<PROTOCOLS>
		Docking with Poly-A pose
		<Add mover_name=makepolya/>
		<Add mover_name=setup_symm/>
		<Add mover_name=mc_dock/>
		<Add filter_name=cc_jump_valid/>
		<Add mover_name=sidechains/>

		Fine grid search
		<Add mover_name=store_design_task />
		<Add mover_name=store_revert_task />
		<Add mover_name=design_process/>

		<Add filter_name=sc />
		<Add filter_name=ddG_filt />
```

```
            <Add filter_name=avg_deg />
            <Add filter_name=unsat_pols />
            <Add filter_name=sasa_int_area />
            <Add filter_name=air_energy />
            <Add filter_name=air_fa_atr />
            <Add filter_name=air_fa_rep />
            <Add filter_name=air_fa_dun />
            <Add filter_name=packing />
            <Add filter_name=mutations />
        </PROTOCOLS>

</ROSETTASCRIPTS>
```

Preface to Chapter III

As a follow up to the work described in Chapter II, the forced crystallization of proteins by fusion to the designed two-dimensional arrays is described in the following chapter. By taking different proteins and fusing them to the 2D arrays, the aim is to utilize the 2D arrays and crystallize the fused protein when the array is formed in-vivo for structure studies by electron crystallography and also for many different avenues in science where atomic level patterning of matter may be useful. For example, by making fluorescent proteins with known repeat patterns for use as atomic scale rulers for emerging microscopy techniques.

The work described in Chapter III is ongoing and has not yet been published.

Chapter III

Formation of protein crystals in-vivo by genetic fusion to designed self-assembling two-dimensional arrays

Summary:

A major highlight of the 2D arrays designed in Chapter II is the ability to create a crystal in-vivo that are themselves customizable. The three successfully designed arrays have free termini where other proteins can be fused with the hypothesis that the designed residues will continue to drive the assembly of the 2D array while taking the fused protein along with them, creating brand new 2D arrays of 3 layers (p3Z-42 and p4Z-9 designs) or 2 layers (p6-9H design).

Methods:

Fusion proteins were selected based on molecular weight and functional properties, for example:

1. Fluorescent proteins – Green Fluorescent Protein (GFP) and mutated versions of GFP were selected for the patterning of fluorescent matter at an atomic level and for their use in new microscopic techniques such as PALM microscopy.

2. Spycatcher – This protein is able to covalently bind other proteins using a small peptide sequence. Arrays formed using Spycatcher may allow for the binding (and subsequent crystallization) of other proteins in-vitro without previously being covalently bound to the array during expression.

3. Calmodulin – This protein regulates many cellular processes, including water channel activity. The activity of Calmodulin is regulated by Calcium and it's binding to

an array could enable us to monitor the effects and structural characteristics of the protein in-vivo.

Many other proteins were chosen based on molecular weight (larger proteins were not always avoided but smaller proteins were preferred) and other binding properties (for example, fatty acid binding) in order to get a good idea of the spectrum of different proteins that can bind to the arrays and keep the 2D assembly intact.

Results:

Fusions were created either on the C-terminus or the N-terminus of the proteins with appropriate linkers (example schematic in Fig. 1) and the proteins expressed as normal to those without fusions. Many fused proteins could be seen by negative-stain EM to still create a 2D lattice. At 37^{o}C, where p3Z-42 design would normally clump into 3D crystals, purely 2D crystals could be observed suggesting that the interactions required for the 3D crystal growth were deterred. The lattices between each fused protein also appeared different while retaining a similar FFT pattern as in the original 2D arrays. Figures 2-4 highlight some of the crystals obtained in by this scheme and Table 1 shows the fused proteins where lattices could still be observed by negative-stain TEM.

All fusion proteins were expressed and harvested using the same method as the native design and each other for clear comparison. Crystals observed ranged in sizes (some rivaling the original design).

One of the successful crystals obtained in this manner is Spycatcher and initial results (based on SDS-PAGE) indicate that it still maintains the ability to covalently bind other proteins using it's spytag peptide.

Table 1. List of fusion proteins (to p3Z-42) with observed 2D lattices by negative-stain TEM.

1. Spycatcher
2. Integrin Binder Design Av6_3
3. Ferrodoxin
4. Calmodulin
5. Human Glutaredoxin
6. Human Heart Fatty-Acid Binding Protein
7. T1 Domain of Kv1.3 Potassium Channel
8. Chemokine Receptor CXCR2
9. Human Acylphosphatase
10. CyaY Protein
11. DFFA -Like Effector C Protein
12. TDRD2

Potential hits needing further analysis:

13. Superfolder-GFP
14. mEos 3.2
15. MDM2 Protein
16. Chicken T-Cadherin
17. Thioredoxin
18. Truncated Hemoglobin

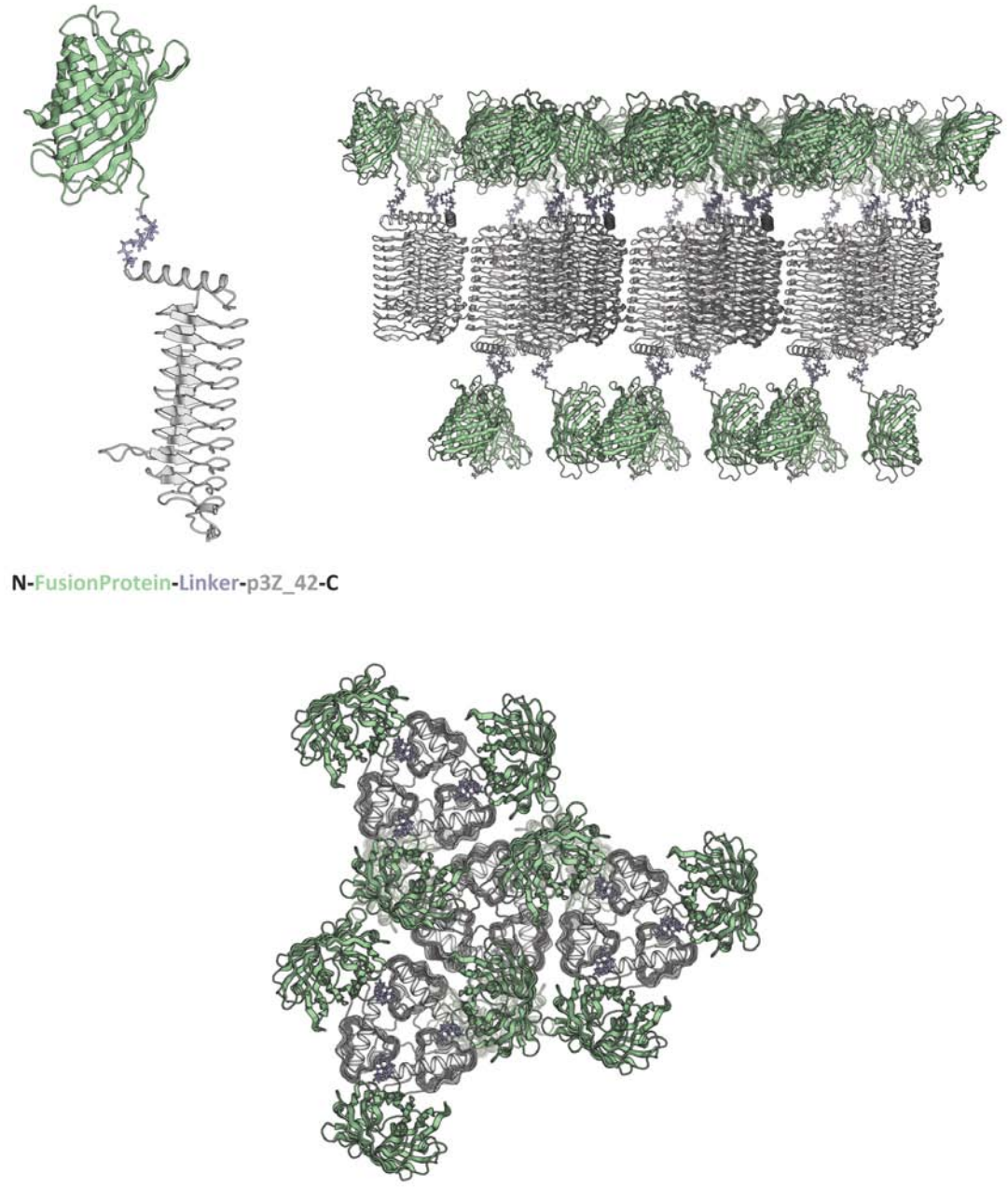


Figure 1. Schematic for fused 2D arrays.

Proteins of interest would be fused to either terminus of the original 2D designed monomer by a short linker (top left). (top right and below) – Side and Top views, respectively, of such a fusion to p3Z-42 design whereby a 3-layer 2D crystal forms with the original crystal in the middle.

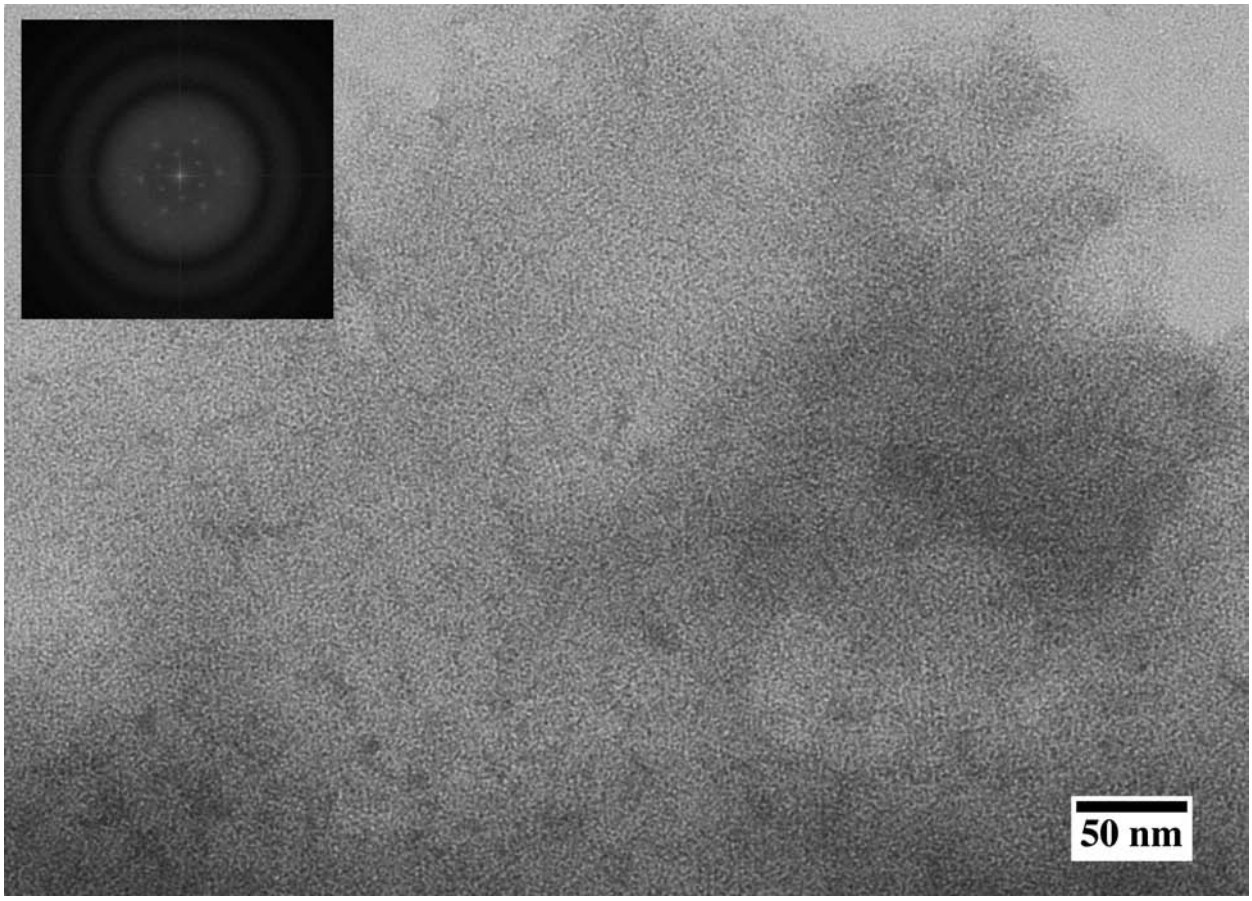


Figure 2. Spycatcher fused onto p3Z-42 protein showing 2D lattice and a similar FFT pattern observed in wildtype p3Z-42.

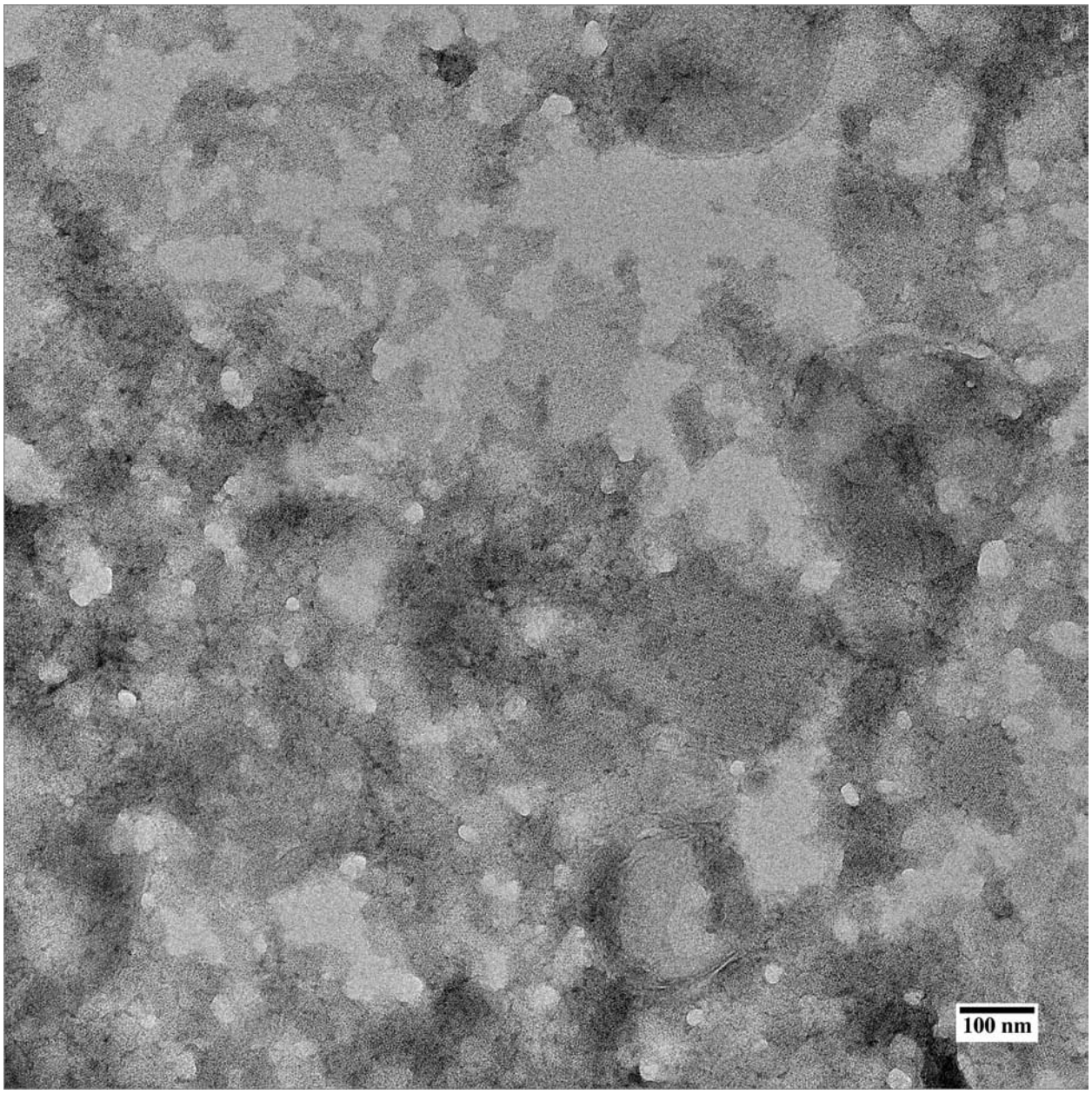


Figure 3. Low-Magnification micrography of CyaY Protein fused to p3Z-42 showing many small-large 2D crystals.

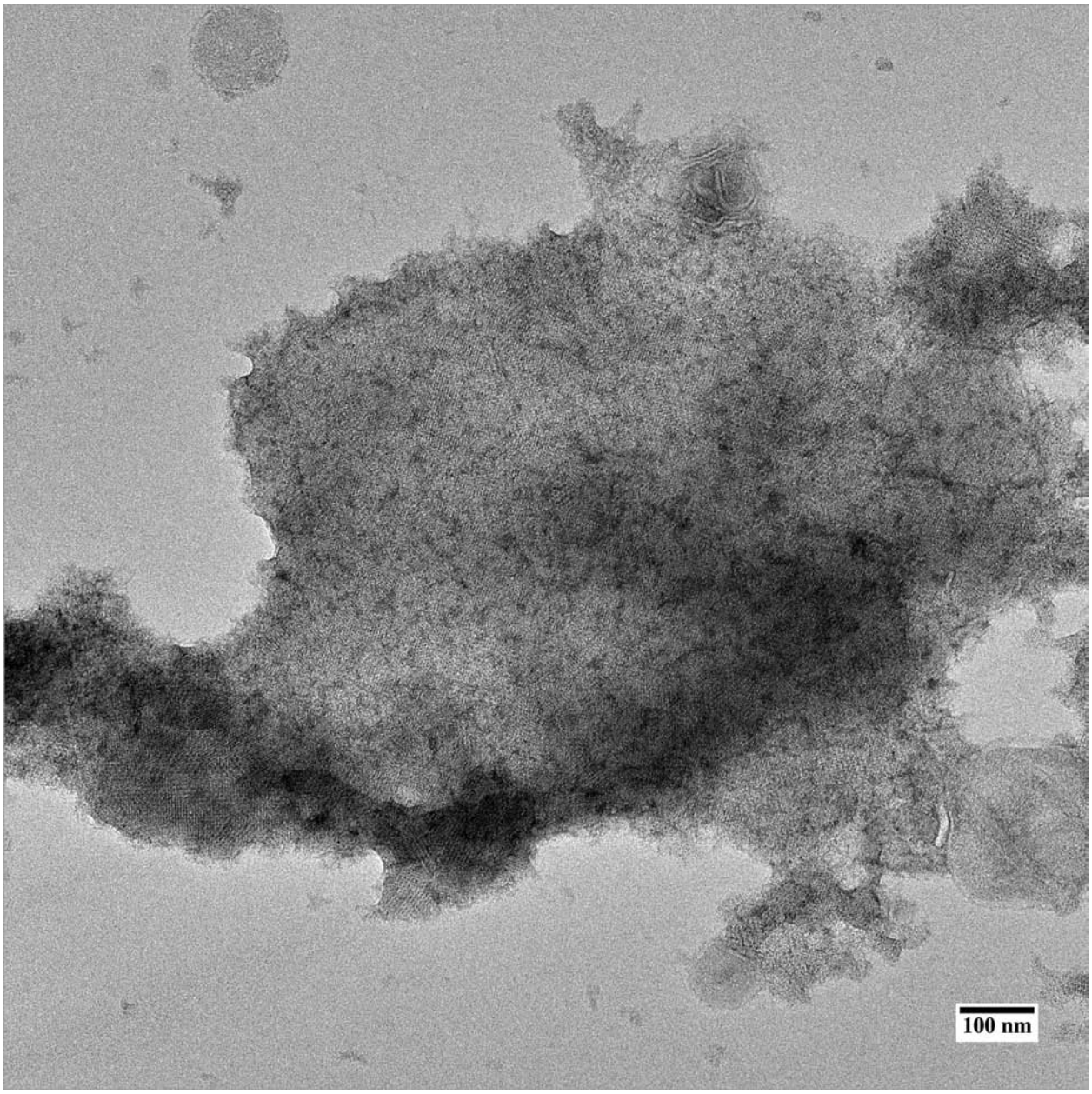


Figure 4. TDRD2 protein fused to p3Z-42 showing a very large 2D lattice.

Preface to Chapter IV

As a follow up to the proof-of-principle study on the design of single-component tetrahedral and octahedral cages (King NP *et al.* 2012), the goal was to design both larger cages and those made up of more than one-component to increase the customizability during real-world therapeutic use of the cages. The following chapter details the follow-up study of this work detailing the design of tetrahedral protein cages made of two separate components and combinations (dimers and trimers and trimers and trimers) where my collaboration on the project involved screening for optimization and the structural characterization of the cages by TEM.

Acknowledgements:

We would like to thank Dan Shi and Brent Nannenga (JFRC) for help with electron microscopy, Frank DiMaio and Rocco Moretti for assistance with software development, Per Greisen for scripts used to compare side chain conformations, Jasmine Gallaher for technical assistance, Michael Collazo for help with preliminary crystallization screening, Duilio Cascio and Michael Sawaya for help with crystallographic experiments, and Malcolm Capel, Jonathan Schuermann, and Igor Kourinov at NE-CAT beamline 24-ID-C for help with data collection. This work was supported by the Howard Hughes Medical Institute (TG and DB) and the JFRC visitor program (SG), the National Science Foundation under CHE-1332907 (DB and TOY), grants from the International AIDS Vaccine Initiative, DTRA (N00024-10-D-6318/0024), AFOSR (FA950-12-10112), and DOE (DE-SC0005155) to DB, an NIH Biotechnology Training Program award to DEM (GM067555), and an NSF graduate

research fellowship to JBB (DGE-0718124). TOY and DEM also acknowledge support from the BER program of the DOE Office of Science.

The work described in Chapter IV was published in the journal of *Nature* as:

N. P. King, J.B. Bale, W. Sheffler, D.E. McNamara, S. Gonen, T. Gonen, T.O. Yeates, D. Baker. Accurate design of co-assembling multi-component protein nanomaterials. *Nature* 510, 103-8 (2014).

Chapter IV

Accurate design of coassembling multi-component protein nanomaterials

Abstract:

The self-assembly of proteins into highly ordered nanoscale architectures is a hallmark of biological systems. The sophisticated functions of these molecular machines inspire the development of methods to engineer novel self-assembling protein structures. Although there has been exciting recent progress in this area, designing multi-component protein nanomaterials with high accuracy remains an outstanding challenge. Here we address this challenge by developing a general computational method for designing protein nanomaterials in which two distinct types of subunits coassemble to a target symmetric architecture. We use the method to design five novel 24-subunit cage-like protein nanomaterials in two distinct symmetric architectures, and experimentally demonstrate that the structures of the materials are in close agreement with the computational design models. The accuracy of the method and the universe of two-component materials that it makes accessible pave the way for the construction of functional protein nanomaterials tailored to specific applications.

Introduction:

The unique functional opportunities afforded by protein self-assembly range from the dynamic cellular scaffolding provided by cytoskeletal proteins to the encapsulation, protection, and delivery of the viral genome to a new host cell by virus capsids. Although natural assemblies can be repurposed to perform new functions[1, 2], this strategy is limited to the structures of existing proteins, which may not be suited to a given

application. To overcome this limitation, methods for designing novel self-assembling proteins are of considerable interest[3-6]. The central challenge in designing self-assembling proteins is to encode the information necessary to direct assembly in the structures of the protein building blocks. Although the complexity and irregularity of protein structures resulted in slow initial progress in this area, advances in computational protein design algorithms and new approaches such as metal-mediated assembly have recently yielded exciting results[6-16]. Despite these advances, the self-assembling protein structures designed to date have been relatively simple, and continued improvements in design strategies are needed in order to enable the practical design of functional materials.

The level of structural complexity available to self-assembled nanomaterials generally increases with the number of unique molecular components used to construct the material. This is illustrated by DNA nanotechnology, in which specific and directional interactions between hundreds of distinct DNA strands allow the construction of nanoscale objects with essentially arbitrary structures[17-20]. In contrast, designing well-ordered multi-component protein nanomaterials has remained a significant challenge. Multiple distinct intermolecular contacts are necessary to drive the assembly of such materials[3, 4, 8, 11, 21], and programming new, geometrically precise interactions between proteins is generally difficult. Compared to homooligomers, multi-component protein nanomaterials offer several advantages: a wider range of possible structures due to their combinatorial nature, greater control over the timing of assembly, and enhanced modularity through independently addressable building blocks. Although multi-

component protein assemblies have recently been generated using disulfide bonds[14, 22], flexible genetic linkers[11, 15, 22], or stereotyped coiled-coil interactions to drive assembly[14, 15], the flexibility of these relatively minimal linkages has generally resulted in materials that are somewhat polydisperse. Most natural protein assemblies, on the other hand, are constructed from protein-protein interfaces involving many contacts distributed over large interaction surfaces that serve to precisely define the positions of the subunits relative to each other[23, 24]. Advances in computational protein modeling and design algorithms have recently made it possible to design such interfaces[25-29] and thereby direct the formation of novel self-assembling protein nanomaterials with atomic-level accuracy[7, 9, 10], but the methods reported to date have been limited to the design of materials comprising only a single type of molecular building block. Here we expand the structural and functional range of designed protein materials with a general computational method for designing two-component coassembling protein nanomaterials with high accuracy.

Computational design method:

Our method centers on encoding the information necessary to direct assembly in designed protein-protein interfaces. In addition to providing the energetic driving force for assembly, the designed interfaces also precisely define the relative orientations of the building blocks. We illustrate the method in Figure 1 using the dual tetrahedral architecture (designated here as T33) as an example. In this architecture, four copies each of two distinct, naturally trimeric building blocks are aligned at opposite poles of the three-fold symmetry axes of a tetrahedron (Figure 1a). This places one set of building

blocks at the vertices of the tetrahedron and the other at the center of the faces, totaling twelve subunits of each protein. Each trimeric building block is allowed to rotate around and translate along its three-fold symmetry axis (Figure 1b); other rigid body moves are disallowed because they would lead to asymmetry. These four degrees of freedom are systematically explored during docking to identify configurations with symmetrically repeated instances of a novel inter-building block interface that is suitable for design (Figure 1c). The docking score function maximizes the number of inter-building block neighbors per residue and favors residues in highly anchored regions of the protein structure that are less likely to change conformation upon mutation of surface side chains (Figure 1d). RosettaDesign[30, 31] is then used to sample the identities and configurations of the side chains near the inter-building block interface, generating interfaces with features resembling those found in natural protein assemblies such as well-packed hydrophobic cores surrounded by polar rims[24] (Figure 1e). The end result is a pair of new amino acid sequences, one for each building block, predicted to stabilize the modeled interface and thereby spontaneously drive assembly to the specific target configuration.

These docking and design procedures were implemented by extending the Rosetta software[31, 32] to enable the simultaneous modeling of multiple distinct symmetrically arranged protein components. The new protocol allows the different components to be arranged and moved independently according to distinct sets of symmetry operators (Extended Data Figure 1). This enables the design strategy described above to be generalized to a wide variety of symmetric architectures in which multiple symmetric

building blocks are combined in geometrically specific ways[3, 4, 21]. Combining even two types of symmetry elements (as in the present study) can give rise to a large number of distinct symmetric architectures with a range of possible morphologies, including those with dihedral and cubic point group symmetries, as well as helical, layer, and space group symmetries (ref. 21 and T.O.Y., manuscript in preparation).

In this study we targeted two distinct tetrahedral architectures: the T33 architecture described above and the T32 architecture shown in Figure 1f, in which the materials are formed from four trimeric and six dimeric building blocks aligned along the three-fold and two-fold tetrahedral symmetry axes. We docked all pairwise combinations of a set of 1,161 dimeric and 200 trimeric protein building blocks of known structure in the T32 and T33 architectures (Supplementary Methods). This resulted in a large set of potential novel nanomaterials: 232,200 and 19,900 docked protein pairs, respectively, with a given pair often yielding several distinct promising docked configurations. Interface sequence design calculations were carried out on the 1,000 highest scoring docked configurations in each architecture, and the designs were evaluated based on the predicted binding energy, shape complementarity[33], and size of the designed interfaces, as well as the number of buried unsatisfied hydrogen bonding groups (Supplementary Methods). After filtering on these criteria, 30 T32 and 27 T33 materials were selected for experimental characterization (Extended Data Figure 2). The 57 designs were derived from 39 distinct trimeric and 19 dimeric proteins, and contained an average of 19 amino acid mutations per pair of subunits compared to the native sequences. The designed

interfaces resided mostly on elements of secondary structure, both α-helices and β-strands, with nearby loops often making minor contributions.

Screening and characterization of assembly state:

Synthetic genes encoding each designed pair of proteins were cloned in tandem in a single expression vector to allow inducible co-expression in *E. coli* (Supplementary Methods). Polyacrylamide gel electrophoresis (PAGE) under denaturing and non-denaturing (native) conditions was used to rapidly screen the level of soluble expression and assembly state of the designed proteins in clarified cell lysates. For most of the designs, either one or both of the designed proteins was not detectable in the soluble fraction, suggesting that insoluble expression is a common failure mode for the designed materials. Given that the majority of the mutations introduced by our method are polar to hydrophobic surface mutations at the designed interfaces, it is likely that the insolubility of these designs is due to either misfolding or nonspecific aggregation of the designed protein subunits. Nevertheless, several designed protein pairs yielded single bands under non-denaturing conditions that migrated more slowly than the wild-type proteins from which they were derived, suggesting assembly to higher-order species (Extended Data Figure 3). These proteins were subcloned to introduce a hexahistidine tag at the C terminus of one of the two subunits and purified by nickel affinity chromatography and size exclusion chromatography (SEC). Five pairs of designed proteins, one T32 design (T32-28) and four T33 designs (T33-09, T33-15, T33-21, and T33-28), co-purified off of the nickel column and yielded dominant peaks at the expected size of approximately 24 subunits when analyzed by SEC (Figure 2a and Supplementary Table 1).

We tested the ability of each of the five materials to assemble *in vitro* by expressing the two components in separate *E. coli* cultures and mixing them at various points after cell lysis (Extended Data Figure 3). Native PAGE revealed that in two cases (T33-15 and T32-28) the two separately expressed components efficiently assembled to the designed materials *in vitro* when equal volumes of cell lysates were mixed (Figure 2b, Extended Data Figure 3a, c). Adjusting the volume of each lysate in the mixture to account for differences in the level of soluble expression of the two components allowed for more quantitative assembly. In the case of T33-15, the two components of the material could also be purified independently: T33-15A and T33-15B each eluted from the SEC column as trimers in isolation. After mixing the two purified components in a 1:1 molar ratio and a two hour incubation at room temperature, the mixture eluted from the SEC column as predominantly the 24mer assembly, with small amounts of residual trimeric building blocks remaining (Figure 2a). It is thus possible to control the assembly of our designed materials by simply mixing the two independently produced components.

The details of the designed interfaces for the five materials are presented in Figure 3. Qualitatively, the interfaces are similar to those in the other designs that were experimentally characterized and reflect the hypothesis underlying the design protocol: they feature well-packed and highly complementary cores of hydrophobic side chains residing mostly in elements of secondary structure, surrounded by polar side chains lining the periphery of the hydrophobic cores. The successful designs are also quantitatively similar to the other designs according to the interface metrics used to select designs for experimental characterization (predicted binding energy, shape complementarity,

interface size, and number of buried unsatisfied hydrogen bonding groups; Extended Data Figure 4). The similarity of the successful and unsuccessful designs according to these structural metrics, combined with the observed insolubility of many of the designs, suggests that focusing on improving the level of soluble expression of the designed proteins could substantially improve the success rate of our approach in the future.

Structural characterization of the designed materials:

Negative stain electron microscopy of the five designed materials confirmed that they assemble specifically to the target architectures (Figure 4). For each material, fields of monodisperse particles of the expected size and symmetry were observed, confirming the homogeneity of the materials suggested by SEC. Particle averaging yielded images that recapitulate features of the computational design models at low resolution. For example, class averages of T33-09 revealed roughly square or triangle-shaped structures with well-defined internal cavities that closely resemble projections calculated from the computational design model along its two-fold and three-fold axes (Figure 4, T33-09 inset). Micrographs of T33-15 assembled *in vitro* as described above were indistinguishable from those of co-expressed T33-15 (Figure 4 and Extended Data Figure 5), demonstrating that the same material is obtained using both methods.

We solved X-ray crystal structures of four of the designed materials (T32-28, T33-15, T33-21, and T33-28) to resolutions ranging from 2.1 to 4.5 Å (Figure 5 and Supplementary Tables 2 and 3). In all cases, the structures reveal that the inter-building block interfaces were designed with high accuracy: comparing a pair of chains from each

structure to the computationally designed model yields backbone root mean square deviations (r.m.s.d.) between 0.5 and 1.2 Å (Figure 5 right and Extended Data Table 1). In the structures with resolutions that permit detailed analysis of side chain configurations (T33-15 and two independent crystal forms of T33-21), 87 of 113 side chains at the designed interfaces adopt the predicted conformations (Supplementary Tables 5 and 6). As intended, the designed interfaces drive the assembly of cage-like nanomaterials that closely match the computational design models: the backbone r.m.s.d. over all 24 subunits in each material range from 1.0 to 2.6 Å (Figure 5 left and Extended Data Table 1). The precise control over interface geometry offered by our method thus enables the design of two-component protein nanomaterials with diverse nanoscale features such as surfaces, pores, and internal volumes with high accuracy.

Discussion:

Due to the unique functional capabilities of self-assembling proteins, there is intense interest in engineering protein nanomaterials for applications in various fields. Most efforts to date have focused on repurposing naturally occurring protein assemblies, a strategy that is ultimately limited by the structures available and their tolerances for modification. Similarly, while directed evolution is a powerful method for protein engineering[34, 35] and can be used to improve, for example, the packaging capability of existing protein nanocontainers[36, 37], it is difficult to envision how it could accurately generate new protein nanomaterials with target structures defined at the atomic level. Our results demonstrate that computational protein design provides a general route for designing novel two-component self-assembling protein nanomaterials with high

accuracy. The combinatorial nature of two-component materials greatly expands the number and variety of potential nanomaterials that can be designed. For example, in this study we used 1,361 protein building blocks to dock over 250,000 distinct protein pairs among two target architectures with tetrahedral point group symmetry, resulting in a very large set of potential nanomaterials exhibiting a variety of sizes, shapes, and arrangements of chemically and genetically addressable functional groups, loops, and termini. With continued effort to increase the success rate of protein-protein interface design and reduce the rate of designed proteins that express insolubly, it should become possible to simultaneously design multiple novel interfaces in a single material, which would enable the construction of increasingly complex materials built from more than two components.

The conceptual framework that underlies our method—symmetric docking followed by protein-protein interface design—can be generally applied to a wide variety of symmetric architectures, including those capable of forming repetitive protein arrays that extend in one, two, or three dimensions. Multi-component materials are advantageous in these extended architectures because the uncontrolled self-assembly of a single-component material inside the cell can complicate biological production[5, 11, 21]. We have shown that the two components of the designed materials T32-28 and T33-15 can be produced separately and mixed *in vitro* to initiate assembly of the designed structure. With new symmetric modeling algorithms capable of handling the additional degrees of freedom associated with these architectures, the accurate computational design

and controllable assembly of complex, multi-component protein fibers, layers, and crystals should also be possible.

The capability to design highly homogeneous protein nanostructures with atomic-level accuracy and controllable assembly should open up new opportunities in targeted drug delivery, vaccine design, plasmonics, and other applications that can benefit from the precise patterning of matter on the sub-nanometer to hundred nanometer scale. Extending beyond static structure design, methods for incorporating the kinds of dynamic and functional behaviors observed in natural protein assemblies should make possible the design of novel protein-based molecular machines with programmable structures, dynamics, and functions.

Methods Summary:

The symmetric modeling framework in Rosetta[31,32] was updated to enable the modeling of multi-component symmetrical structures. A new application, tcdock, docks pairs of protein scaffolds in higher order symmetries, scoring each docked configuration according to its suitability for interface design. tcdock was used to dock all possible pairwise combinations of 200 trimeric scaffold proteins and all possible pairwise combinations of the same trimers and 1,161 dimeric proteins in the T33 and T32 symmetric architectures, respectively. New two-component protein-protein interface design protocols were used to design new amino acid sequences predicted to stabilize selected docked configurations. During the sequence design protocols, the symmetric rigid body degrees of freedom and the identities and conformations of the side chains at

the inter-building block interfaces were optimized to identify low-energy sequence-structure combinations. 30 T32 and 27 T33 designs were selected for experimental characterization.

The assembly states of the designed pairs of proteins were assessed by native PAGE, and those that migrated more slowly than the wild-type scaffolds were subjected to affinity purification and SEC. The ability of the materials to assemble *in vitro* was investigated by independently producing the two components, mixing them at various points after cell lysis, and analyzing the mixtures by native PAGE and SEC. The materials were structurally characterized by negative stain electron microscopy including particle averaging, and at high resolution by X-ray crystallography.

The crystal structures and structure factors for the designed materials have been deposited in the RCSB Protein Data Bank (http://www.rcsb.org/) under the accession codes 4NWN (T32-28), 4NWO (T33-15), 4NWP (T33-21, R32 crystal form), 4NWQ (T33-21, $F4_132$ crystal form), and 4NWR (T33-28).

Figures

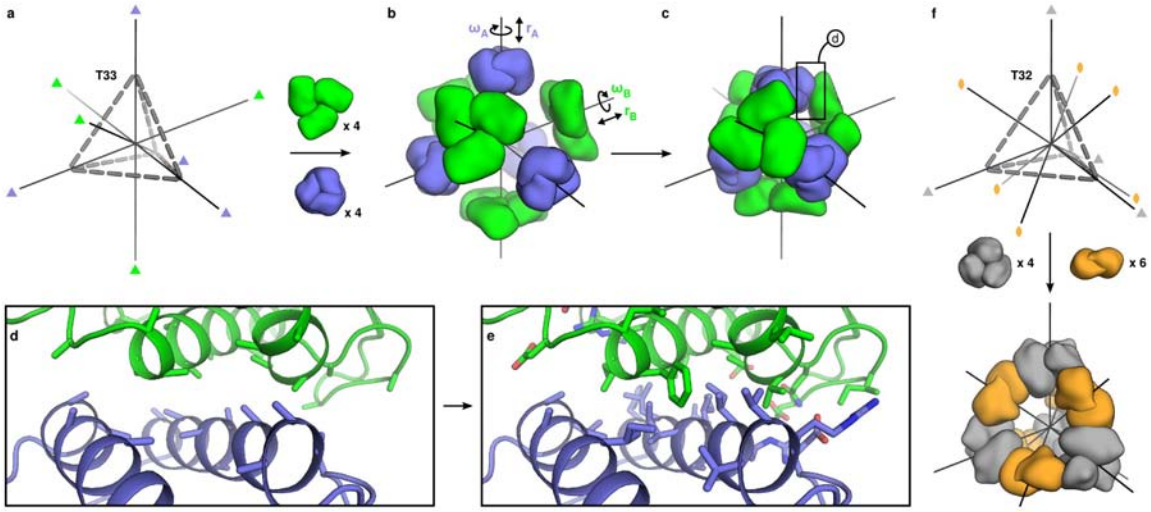


Figure 1 | Overview of the computational design method. **a,** The T33 architecture comprises four copies each of two distinct trimeric building blocks (green and blue) arranged with tetrahedral point group symmetry (24 total subunits; triangles indicate three-fold symmetry axes). **b,** Each building block has two rigid body degrees of freedom, one translational (r) and one rotational (ω), that are systematically explored during docking. **c–d,** The docking procedure, which is independent of the amino acid sequence of the building blocks, identifies large interfaces with high densities of contacting residues formed by well-anchored regions of the protein structure. **e,** Amino acid sequences are designed at the new interface to stabilize the modeled configuration and drive coassembly of the two components. **f,** In the T32 architecture, four trimeric (grey) and six dimeric (orange) building blocks are aligned along the three-fold and two-fold symmetry axes passing through the vertices and edges of a tetrahedron, respectively.

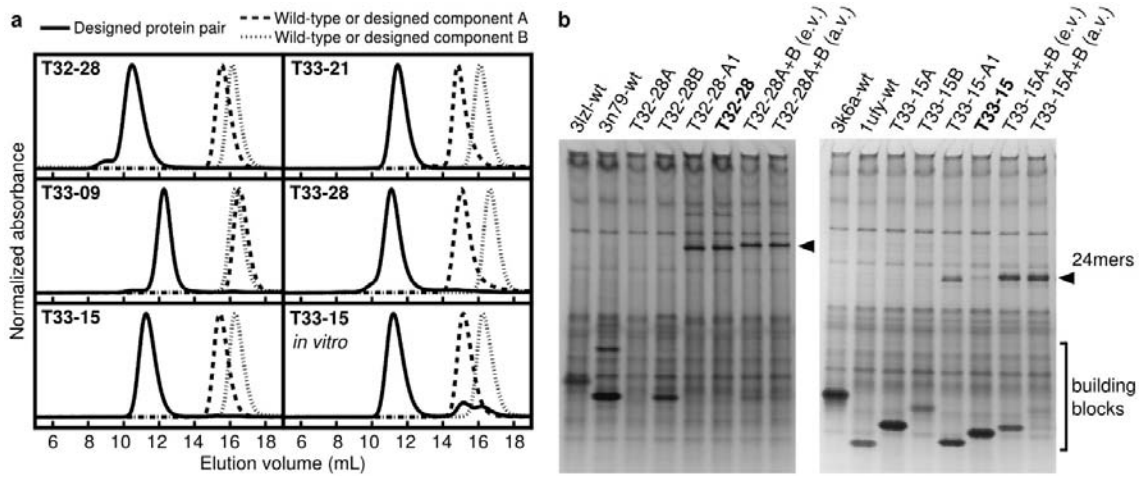


Figure 2 | Experimental characterization of coassembly. a, SEC chromatograms of the designed pairs of proteins (solid lines) and the wild-type oligomeric proteins from which they were derived (dashed and dotted lines). The co-expressed designed proteins elute at the volumes expected for the target 24-subunit nanomaterials, while the wild-type proteins elute as dimers or trimers. The T33-15 *in vitro* panel shows chromatograms for the individually produced and purified designed components (T33-15A and T33-15B) as well as a stoichiometric mixture of the two components. **b,** Native PAGE analysis of *in vitro*-assembled T32-28 (left panel) and T33-15 (right panel) in cell lysates. Lysates containing the co-expressed design components (lanes 5–6) contain slowly migrating species (arrows) not present in lysates containing the wild-type and individually expressed components (lanes 1–4). Mixing equal volumes (e.v.) of crude lysates containing the individual designed components yields the same assemblies (lane 7), although some unassembled building blocks remain due to unequal levels of expression (particularly for T33-15). When the differences in expression levels are accounted for by mixing adjusted volumes of lysates (a.v.), more efficient assembly is observed (lane 8).

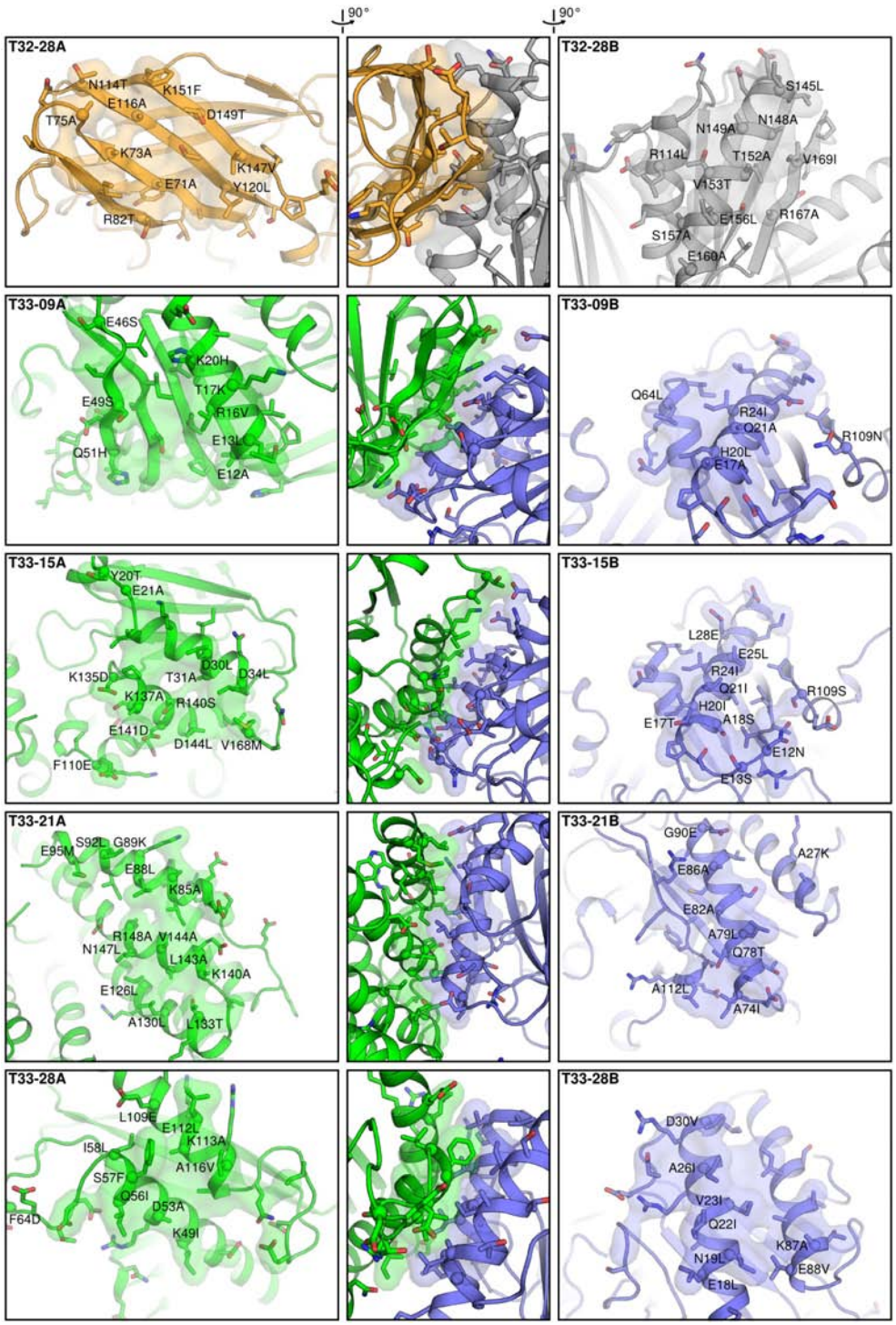


Figure 3 | Modeled interfaces of designed two-component protein nanomaterials.

The models of the designed interfaces in each component of T32-28, T33-09, T33-15, T33-21, and T33-28 are shown at left or right, and side views of each interface as a whole are shown at center. Each image is oriented such that a vector originating at the center of the tetrahedral material and passing through the center of mass of the designed interface would pass vertically through the center of the image. The side chains of all amino acids allowed to repack and minimize during the interface design procedure are shown in stick representation. The alpha carbon atoms of positions that were mutated during design are shown as spheres, and the mutations are labeled. To highlight the morphologies of the contacting surfaces, atoms within 5 Å of the opposite building block are shown in semi-transparent surface representation. Oxygen atoms are red; nitrogen, blue; and sulfur, orange.

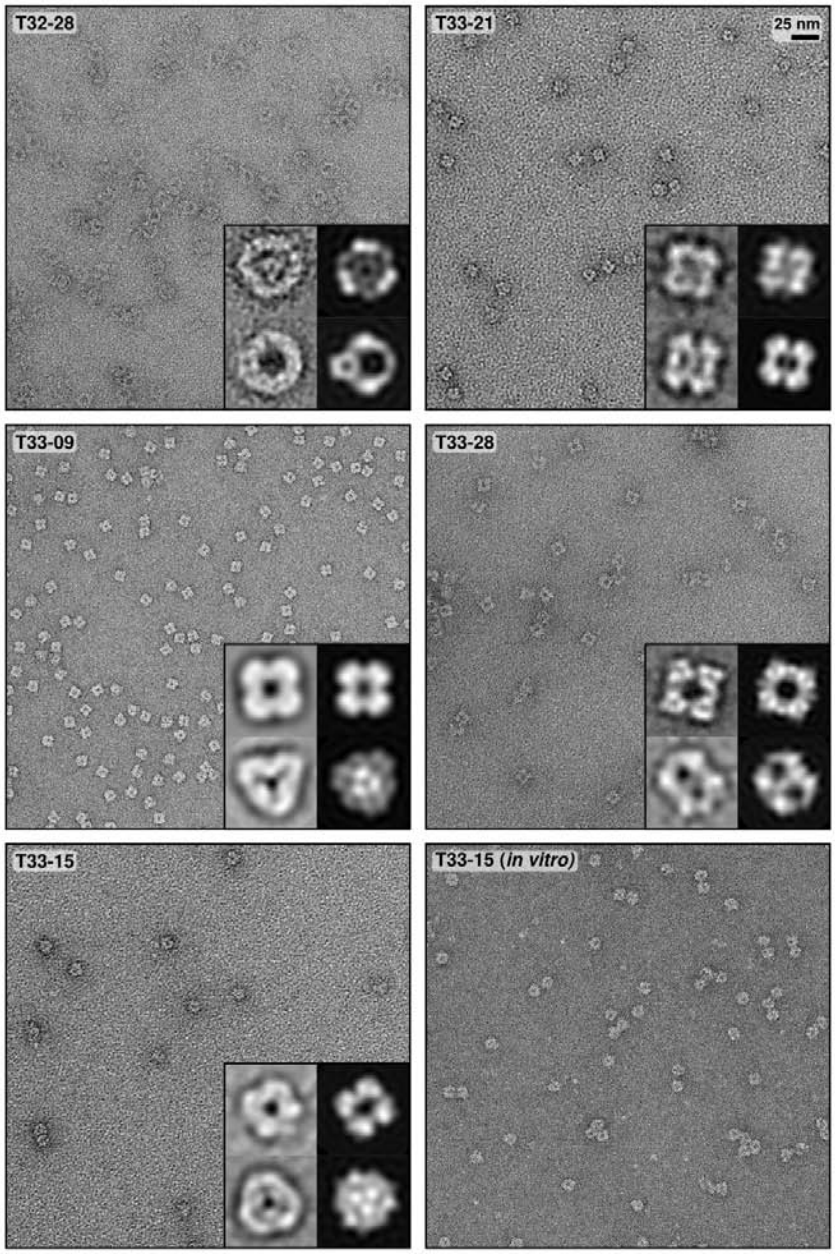


Figure 4 | Electron micrographs of designed two-component protein nanomaterials. Negative stain electron micrographs for five designed materials are shown to scale (scale bar: 25 nm). The T33-15 *in vitro* sample was prepared by stoichiometrically mixing the independently purified components (T33-15A and T33-15B) *in vitro* and purifying the assembled material by SEC (see Figure 2). Micrographs of unpurified, *in vitro*-assembled T33-15 as well as T33-15A and T33-15B in isolation are shown in Extended

Data Figure 5. For each material, two different class averages of the particles are shown in the insets (left) alongside back projections calculated from the computational design models (right).

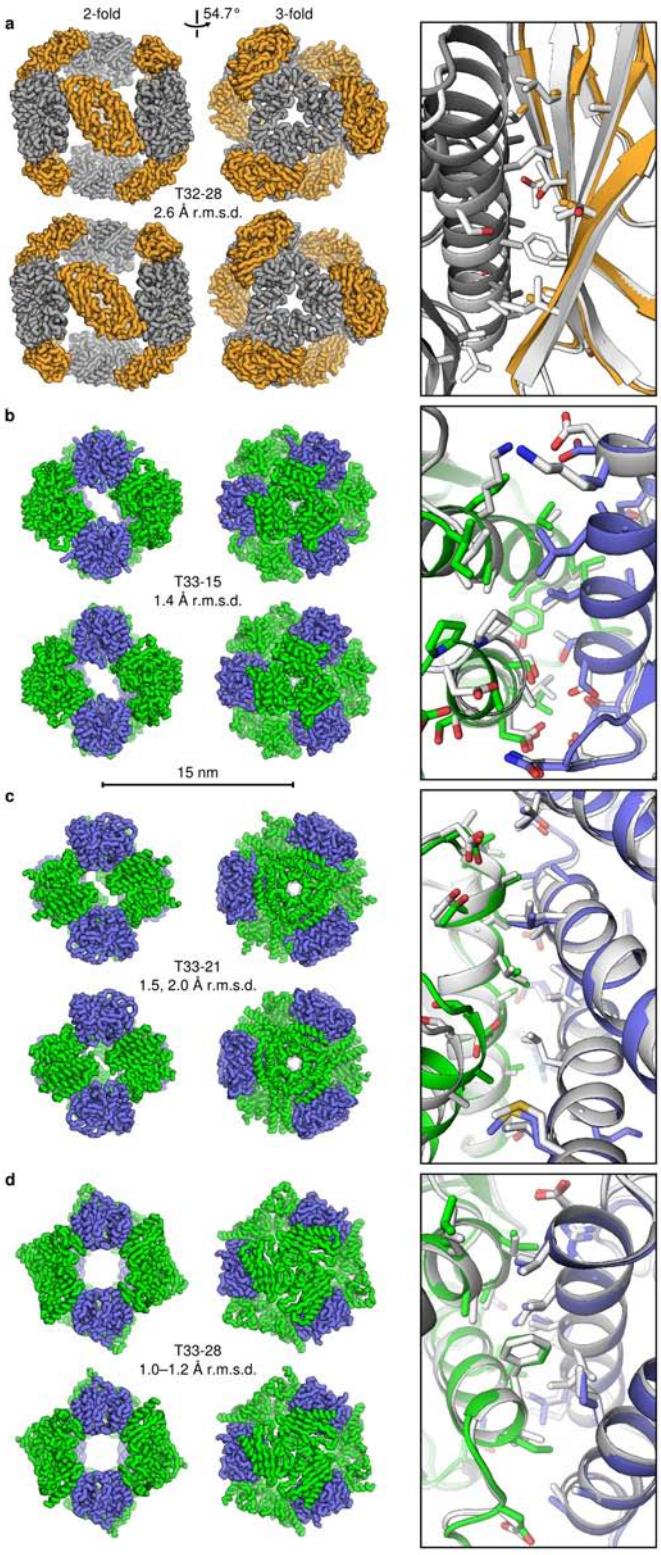
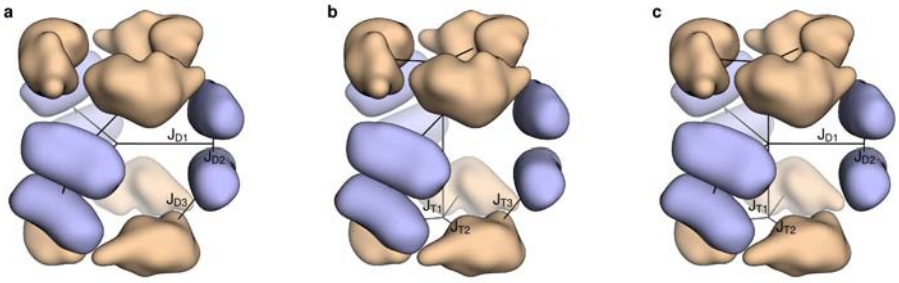
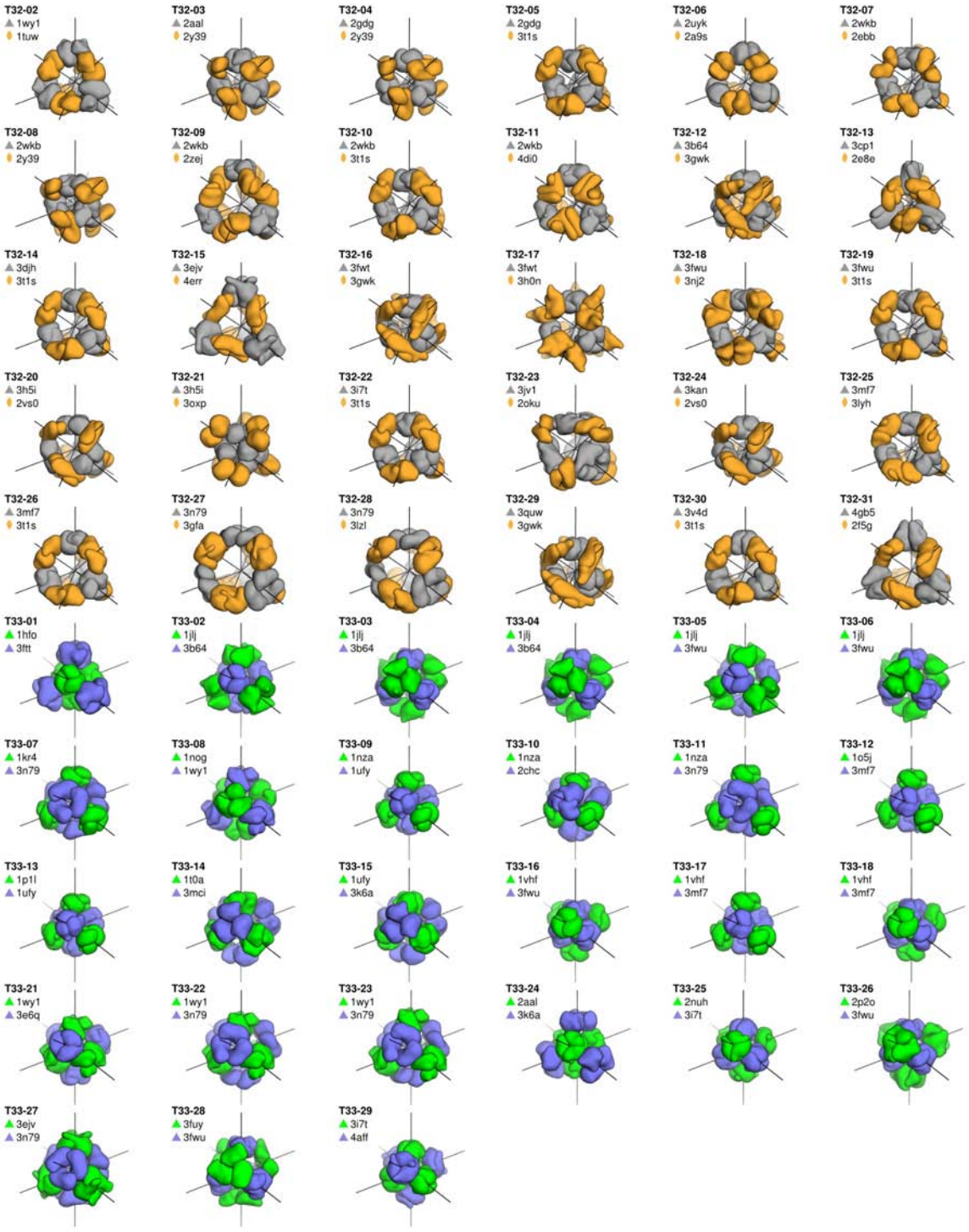


Figure 5 | Crystal structures of designed two-component protein nanomaterials.

The computational design models (top) and X-ray crystal structures (bottom) are shown at left for **a,** T32-28, **b,** T33-15, **c,** T33-21, and **d,** T33-28. Views of each material are shown to scale along the 2-fold and 3-fold tetrahedral symmetry axes (scale bar: 15 nm). The r.m.s.d. values given are those between the backbone atoms in all 24 chains of the design models and crystal structures. For T33-21, r.m.s.d. values are shown for both crystal forms (images are shown for the higher resolution crystal form with backbone r.m.s.d. 2.0 Å), while the r.m.s.d. range for T33-28 derives from the four copies of the fully assembled material in the crystallographic asymmetric unit. At right, overlays of the designed interfaces in the design models (white) and crystal structures (grey, orange, green, and blue) are shown. Due to the limited resolution of the T32-28 structure, the amino acid side chains were not modeled beyond the beta carbon. For the interface overlays, the crystal structures were aligned to the design models using the backbone atoms of two subunits, one of each component.

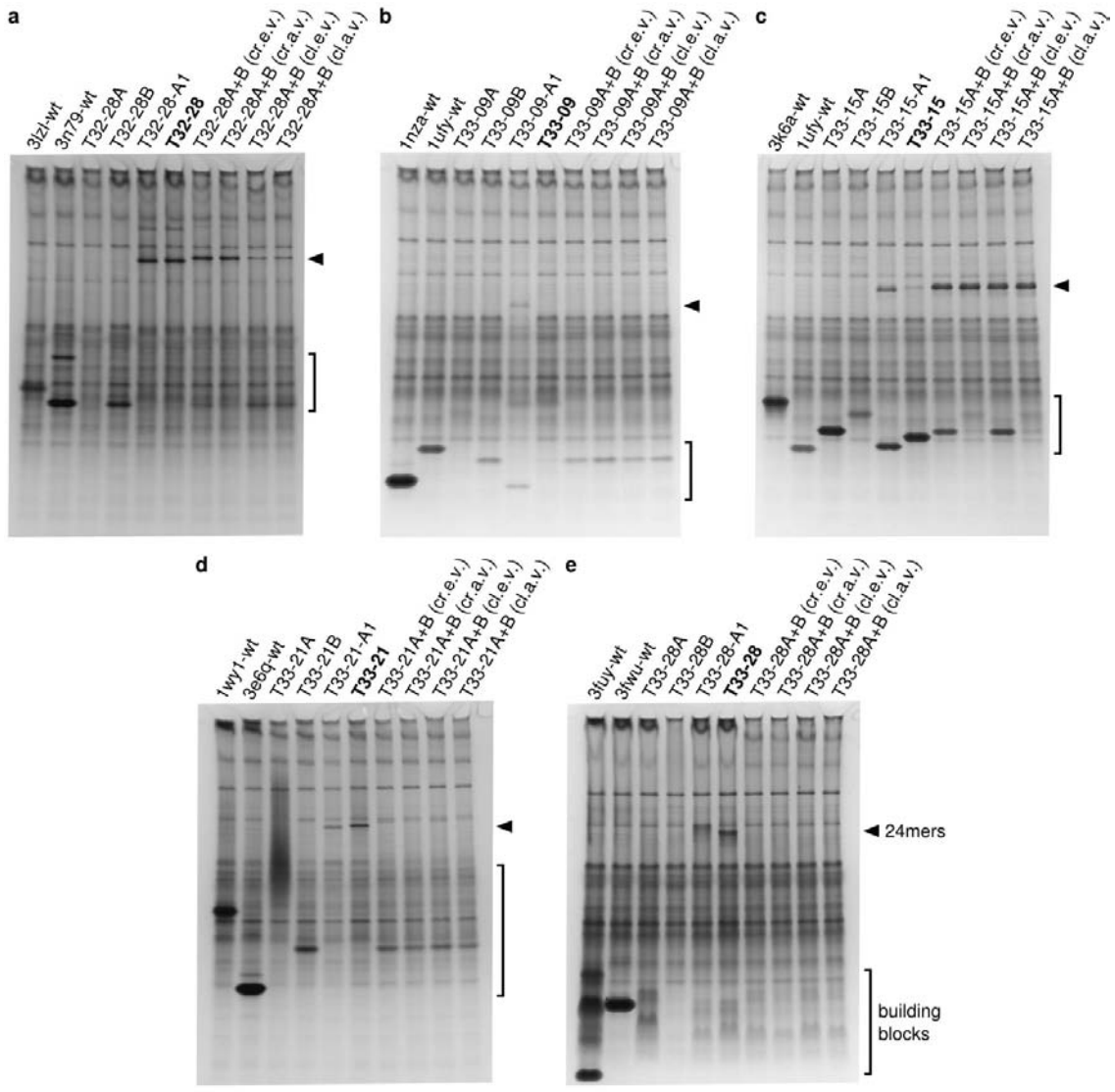


Extended Data Figure 1 | Comparison of one-component and multi-component symmetric fold trees. Three different symmetric fold tree representations of a D32 architecture are shown. In this architecture, two trimeric building blocks (wheat) are aligned along the three-fold rotational axes of D3 point group symmetry and three dimeric building blocks (light blue) are aligned along the two-folds. **a**, The dimer-centric one-component symmetry case. Rigid body degree of freedom (RB DOF, black lines) J_{D3} connecting the master dimer subunit to the master trimer subunit is a child of RB DOFs J_{D1} and J_{D2} controlling the master dimer subunit; in this case the positions of the trimeric subunits depend on the positions of the dimeric subunits. **b**, The trimer-centric one-component symmetry case. RB DOF J_{T3} connecting the master trimer subunit to the master dimer subunit is a child of RB DOFs J_{T1} and J_{T2} controlling the master trimer subunit; in this case the positions of the dimeric subunits depend on the positions of the trimeric subunits. **c**, The multi-component symmetry case. With multi-component symmetric modeling, the RB DOFs controlling the master trimer subunit (J_{T1} and J_{T2}) and the master dimer subunit (J_{D1} and J_{D2}) are independent. In this case the positions of the dimeric subunits do not depend on the positions of the trimeric subunits and vice versa, allowing the internal DOFs for each building block (J_{T2} and J_{D2}) to be maintained while moving the building blocks independently (J_{T1} and J_{D1}). See the Supplementary Methods for additional discussion.



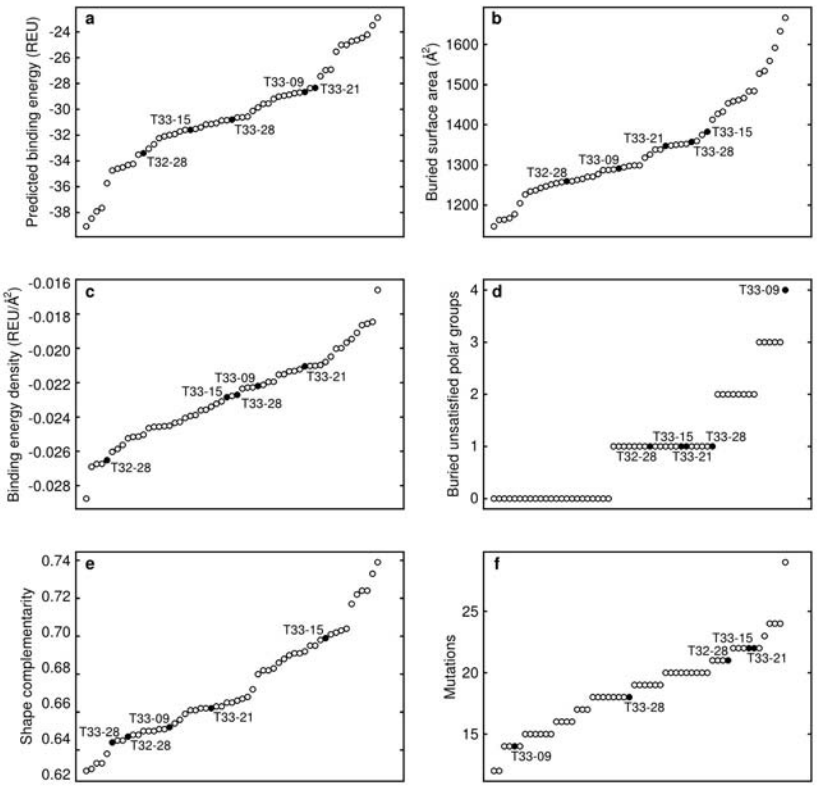
Extended Data Figure 2 | Models of the 57 designs selected for experimental characterization. Smoothed surface representations are shown of each of the 30 T32 and 27 T33 designs. The trimeric component of each T32 design is shown in grey and

the dimeric component in orange. The two different trimeric components of each T33 design are shown in blue and green. The tetrahedral two-fold and three-fold symmetry axes (black lines) are shown passing through the center of each component. Each design is named according to its symmetric architecture (T32 or T33) followed by a unique identification number. The pairs of scaffold proteins from which the designs are also indicated.

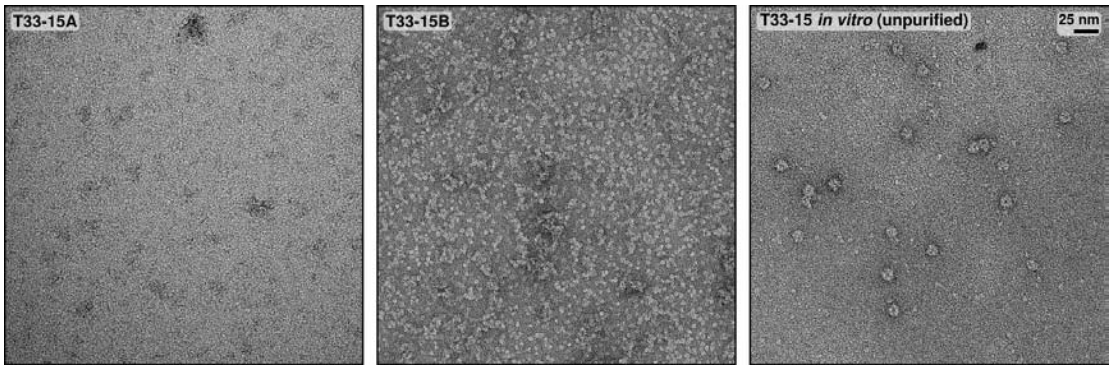


Extended Data Figure 3 | Native PAGE analysis of cleared cell lysates. Each gel contains cleared lysates pertaining to **a**, T32-28, **b**, T33-09, **c**, T33-15, **d**, T33-21, or **e**, T33-28. Lane 1 is from cells expressing the wild-type scaffold for component A and lane 2 the wild-type scaffold for component B. Lanes 3–4 are from cells expressing the individual design components and lanes 5–6 the co-expressed components. Lanes 7–8 are from samples mixed as crude lysates (cr.e.v or cr.a.v), while lanes 9–10 are from samples mixed as cleared lysates (cl.e.v. or cl.a.v.). Lanes 7 and 9 are from lysates mixed with equal volumes (cr.e.v. or cl.e.v.), while lanes 8 and 10 are from lysates mixed with

adjusted volumes (cr.a.v. or cl.a.v.). Lane 5 is from cells expressing the C-terminally A1-tagged constructs; all other lanes are from cells expressing the C-terminally His-tagged constructs. An arrow is positioned next to each gel indicating the migration of 24-subunit assemblies and the gel regions containing unnassembled building blocks are bracketed. Each gel was stained with GelCode Blue (Thermo Scientific). Portions of the gels in **a** and **c** are also shown in Figure 2b.



Extended Data Figure 4 | Structural metrics for the computational design models. Selected metrics related to the designed interfaces are plotted for the 57 designs that were experimentally characterized, including **a**, the predicted binding energy measured in Rosetta Energy Units (REU), **b**, the surface area buried by each instance of the designed interface, **c**, the binding energy density (calculated as the predicted binding energy divided by the buried surface area), **d**, the number of buried unsatisfied polar groups at the designed interface, **e**, the shape complementarity of the designed interface, and **f**, the total number of mutations in each designed pair of proteins. Each circle represents a single design; the five successful materials are plotted as filled circles and labeled. In each plot, the designs are arranged on the x axis in order of increasing value.



Extended Data Figure 5 | Electron micrographs of *in vitro*-assembled T33-15 (unpurified) and T33-15A and T33-15B in isolation. Negative stain electron micrographs of independently purified T33-15 components and unpurified, *in vitro*-assembled T33-15 are shown to scale (scale bar: 25 nm).

Design model	Crystal structure	Global r.m.s.d. (Å)*	Two-chain r.m.s.d. (Å)†	Contents of asymmetric unit	Structure used for superposition‡
T32-28	4NWN	2.586	1.246	One cage (24 subunits)	Asymmetric unit
T33-15	4NWO	1.433	0.876	One chain of each component (2 subunits)	One cage generated from crystallographic 2- and 3-folds
T33-21	4NWP	1.962	0.924	4 chains of each component (8 subunits)	One cage generated from one crystallographic 3-fold
T33-21	4NWQ	1.482	0.765	One chain of each component (2 subunits)	One cage generated from crystallographic 2- and 3-folds
T33-28	4NWR	0.965	0.503	Four complete cages (96 subunits)	One complete cage from the asymmetric unit
T33-28	4NWR	0.965	0.548	Four complete cages (96 subunits)	One complete cage from the asymmetric unit
T33-28	4NWR	1.195	0.567	Four complete cages (96 subunits)	One complete cage from the asymmetric unit
T33-28	4NWR	1.212	0.477	Four complete cages (96 subunits)	One complete cage from the asymmetric unit

Extended Data Table 1 | Root mean square deviations (r.m.s.d.) between crystal structures and design models.

***** Global r.m.s.d. was calculated over all 24 subunits of each design model and corresponding subunits in each crystal structure.

† Two-chain r.m.s.d. was calculated over chains A and B of each design model and corresponding subunits in each crystal structure.

‡ 24 subunits composing one complete cage were derived from each crystal structure as indicated and the chains renamed to match the corresponding names in the design models. In the case of T33-28, four different sets of r.m.s.d. calculations were carried out; one for each of the four cages contained in the asymmetric unit of 4NWR.

References highlighted in Chapter IV

1. Howorka, S. Rationally engineering natural protein assemblies in nanobiotechnology. *Curr Opin Biotechnol* 22, 485-491 (2011).
2. Douglas, T. & Young, M. Viruses: making friends with old foes. *Science* 312, 873-875 (2006).
3. Lai, Y.T., King, N.P. & Yeates, T.O. Principles for designing ordered protein assemblies. *Trends Cell Biol* 22, 653-661 (2012).
4. King, N.P. & Lai, Y.T. Practical approaches to designing novel protein assemblies. *Curr Opin Struct Biol* 23, 632-638 (2013).
5. Sinclair, J.C. Constructing arrays of proteins. *Curr Opin Chem Biol* 17, 946-951 (2013).
6. Salgado, E.N., Radford, R.J. & Tezcan, F.A. Metal-directed protein self-assembly. *Acc Chem Res* 43, 661-672 (2010).
7. King, N.P. et al. Computational design of self-assembling protein nanomaterials with atomic level accuracy. *Science* 336, 1171-1174 (2012).
8. Brodin, J.D. et al. Metal-directed, chemically tunable assembly of one-, two- and three-dimensional crystalline protein arrays. *Nat Chem* 4, 375-382 (2012).
9. Lanci, C.J. et al. Computational design of a protein crystal. *Proc Natl Acad Sci U S A* 109, 7304-7309 (2012).
10. Stranges, P.B., Machius, M., Miley, M.J., Tripathy, A. & Kuhlman, B. Computational design of a symmetric homodimer using beta-strand assembly. *Proc Natl Acad Sci U S A* 108, 20562-20567 (2011).

11. Sinclair, J.C., Davies, K.M., Venien-Bryan, C. & Noble, M.E. Generation of protein lattices by fusing proteins with matching rotational symmetry. *Nat Nanotechnol* 6, 558-562 (2011).

12. Lai, Y.T., Cascio, D. & Yeates, T.O. Structure of a 16-nm cage designed by using protein oligomers. *Science* 336, 1129 (2012).

13. Der, B.S. et al. Metal-mediated affinity and orientation specificity in a computationally designed protein homodimer. *J Am Chem Soc* 134, 375-385 (2012).

14. Fletcher, J.M. et al. Self-assembling cages from coiled-coil peptide modules. *Science* 340, 595-599 (2013).

15. Boyle, A.L. et al. Squaring the circle in peptide assembly: from fibers to discrete nanostructures by de novo design. *J Am Chem Soc* 134, 15457-15467 (2012).

16. Grigoryan, G. et al. Computational design of virus-like protein assemblies on carbon nanotube surfaces. *Science* 332, 1071-1076 (2011).

17. Seeman, N.C. Nanomaterials based on DNA. *Annu Rev Biochem* 79, 65-87 (2010).

18. Rothemund, P.W. Folding DNA to create nanoscale shapes and patterns. *Nature* 440, 297-302 (2006).

19. Ke, Y., Ong, L.L., Shih, W.M. & Yin, P. Three-dimensional structures self-assembled from DNA bricks. *Science* 338, 1177-1183 (2012).

20. Han, D. et al. DNA gridiron nanostructures based on four-arm junctions. *Science* 339, 1412-1415 (2013).

21. Padilla, J.E., Colovos, C. & Yeates, T.O. Nanohedra: using symmetry to design self assembling protein cages, layers, crystals, and filaments. *Proc Natl Acad Sci U S A* 98, 2217-2221 (2001).

22. Usui, K. et al. Nanoscale elongating control of the self-assembled protein filament with the cysteine-introduced building blocks. *Protein Sci* 18, 960-969 (2009).

23. Goodsell, D.S. & Olson, A.J. Structural symmetry and protein function. *Annu Rev Biophys Biomol Struct* 29, 105-153 (2000).

24. Janin, J., Bahadur, R.P. & Chakrabarti, P. Protein-protein interaction and quaternary structure. *Q Rev Biophys* 41, 133-180 (2008).

25. Huang, P.S., Love, J.J. & Mayo, S.L. A de novo designed protein protein interface. *Protein Sci* 16, 2770-2774 (2007).

26. Jha, R.K. et al. Computational design of a PAK1 binding protein. *J Mol Biol* 400, 257-270 (2010).

27. Karanicolas, J. et al. A de novo protein binding pair by computational design and directed evolution. *Mol Cell* 42, 250-260 (2011).

28. Fleishman, S.J. et al. Computational design of proteins targeting the conserved stem region of influenza hemagglutinin. *Science* 332, 816-821 (2011).

29. Khare, S.D. & Fleishman, S.J. Emerging themes in the computational design of novel enzymes and protein-protein interfaces. *FEBS Lett* 587, 1147-1154 (2013).

30. Kuhlman, B. & Baker, D. Native protein sequences are close to optimal for their structures. *Proc Natl Acad Sci U S A* 97, 10383-10388 (2000).

31. Leaver-Fay, A. et al. ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules. *Methods Enzymol* 487, 545-574 (2011).

32. DiMaio, F., Leaver-Fay, A., Bradley, P., Baker, D. & Andre, I. Modeling symmetric macromolecular structures in Rosetta3. *PLoS One* 6, e20450 (2011).

33. Lawrence, M.C. & Colman, P.M. Shape complementarity at protein/protein interfaces. *J Mol Biol* 234, 946-950 (1993).

34. Arnold, F.H. & Volkov, A.A. Directed evolution of biocatalysts. *Curr Opin Chem Biol* 3, 54-59 (1999).

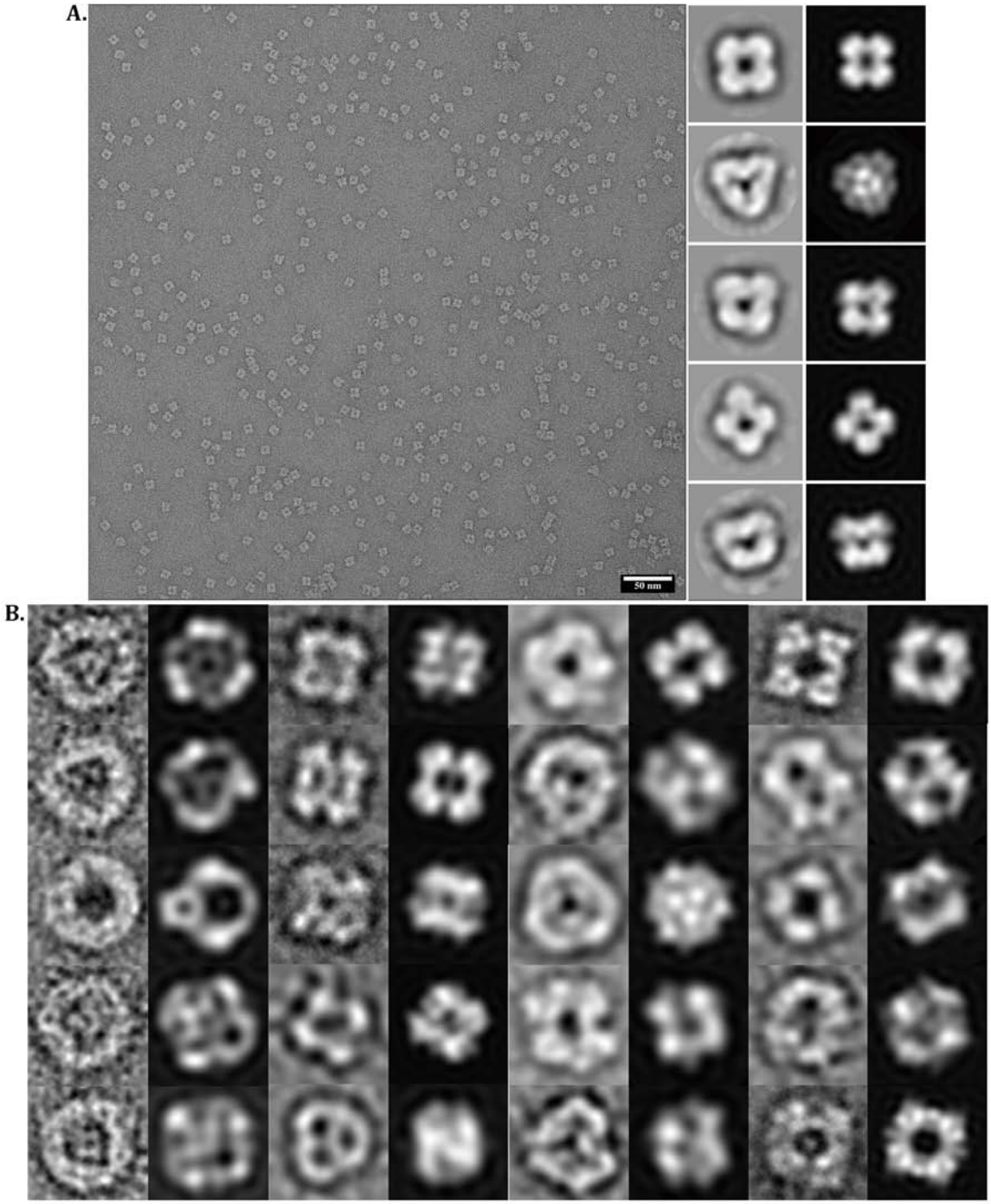
35. Jackel, C., Kast, P. & Hilvert, D. Protein design by directed evolution. *Annu Rev Biophys* 37, 153-173 (2008).

36. Worsdorfer, B., Pianowski, Z. & Hilvert, D. Efficient in vitro encapsulation of protein cargo by an engineered protein container. *J Am Chem Soc* 134, 909-911 (2012).

37. Worsdorfer, B., Woycechowsky, K.J. & Hilvert, D. Directed evolution of a protein container. *Science* 331, 589-592 (2011).

Appendix IV

Additional averages for the tetrahedra described in Chapter IV



Preface to Chapter V

As a follow up to the work described in Chapter IV, the following chapter describes the engineering of both single and two-component Icosahedral cages. The ultimate goal of designing such cages is to create a drug delivery mechanism whereby there are multiple customizable components and many different sized pores to encapsulate and isolate the drug from the surrounding environment. Icosahedra are made of 60 subunits and are much larger than their tetrahedral counterparts. Two-component icosahedra can be made out of different combinations of symmetric subunits and this resulted in many differently sized cages and pores and they were either porous or mostly closed off to the surrounding environment.

The following chapter describes the successful design of both one- and two-component icosahedra. My collaboration on both these projects involved the screening of design conditions for optimization and the structural characterization of the successfully designed cages through both negative-stain and cryo-EM.

The work described in Chapter V is ongoing and has not yet been published with collaborations from both the baker group (Primarily Jacob Bale and Yang Hsai) and the Gonen lab (Dan Shi).

Chapter V

One- and two-component, self-assembling Icosahedral protein cages

Abstract:

Symmetric protein assemblies occur in nature as higher order structures in non-crystallographic form in addition to crystallographic. Tetrahedrons, Octahedrons and Icosahedrons are formed – usually as micro cages to shuttle material around and in and out of the cell, keeping toxins enveloped until the cell can degrade them or as viral capsids. All three of these polyhedra contain 2- and 3- fold axes and additionally, octahedral have 4-fold axes and icosahedra have 5-fold axes. Using cyclic proteins that share an elements of symmetry with the chosen architecture, all three polyhedra can be modeled in Rosetta.

One Component Icosahedral Cages:

The only way to model icosahedral cages using only one protein component using the strategy outlined in Chapter IV is to use cyclic-3, or C3 trimeric proteins. The design I3-01 uses the trimeric protein scaffold deposited with PDB code 1WA3. Each monomer interacts extensively with it's symmetrical neighboring monomer therefore each interface is repeated identically throughout the cage. After purification of the cages by size exclusion chromatography, clear cages could be seen by negative-stain TEM (Fig. 1) and even the symmetrical axes were spotted, particularly the 5-fold.

In order to achieve a clearer picture of the cages, the cages were frozen in liquid ethane and imaged by cryo-EM. At this point, we noticed aggregation of the cages (Fig. 2). After addition of reducing agent, DTT, during purification and before grid

preparation, and alternating the grid preparation we were able to clearly see more separated I3-01 cages and high quality movies were recorded on a FEI Titan Krios TEM with a K2 camera (Fig. 3). These movies allowed for motion correction algorithms to be applied in order to calculate motion corrected images. These images were used to calculate class averages that clearly match the design model in all three symmetrical axes (Fig. 4). The class averages were then used to derive an initial density model for later refinement to high resolution. The model matches the design and the Rosetta output could be clearly fitted into the density (Fig. 5 & 6).

I3-01 fused to superfolder Green Fluorescent Protein (superfolder-GFP) was also created using a short fusion linker at the C-terminus of the protein. In this manner, each of the 60 monomers creating the cage was fused to a superfolder-GFP. This fusion design was screened using negative-stain TEM (Fig. 7) and was found that the superfolder GFP's seem to be encased inside of the cage or close to the edge. Screening showed the same cages seen earlier only this time, the empty spaces were filled with protein (Fig. 8). Further analysis was made using cryo-EM. Using similar grid optimized conditions as with empty I3-01, mono-dispersed superfolder-GFP fused I3-01 cages were clearly observed (Fig. 9). High quality movies were collected and motion corrected to yield micrographs of sufficient quality for class averaging and 3D refinement (Fig. 10). Class averaging yielded similar classes to empty I3-01's (Fig. 11) except a dense cage was observed when compared to the empty spaces from earlier I3-01's. The density seemed to match the model whereby the symmetrical axes were observed, indicating that the superfolder-GFP's were located in or close to the cage surface.

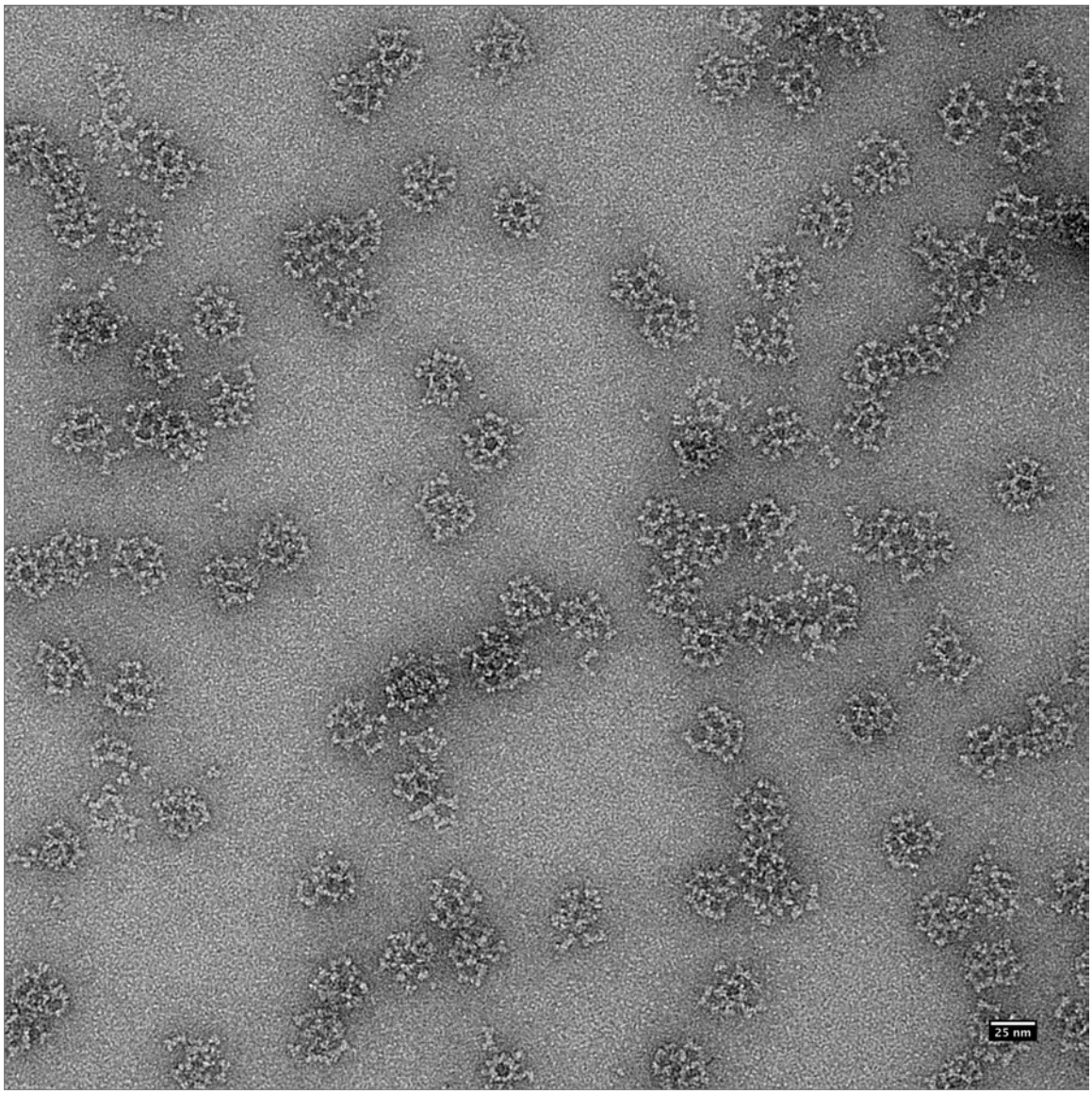


Figure 1. Negative-stain TEM of empty I3-01 cages.

Purified I3-01 without any fusions was stained with Uranyl Formate and imaged by TEM. Icosahedral cages could clearly be viewed, including the 5-fold symmetric axis.

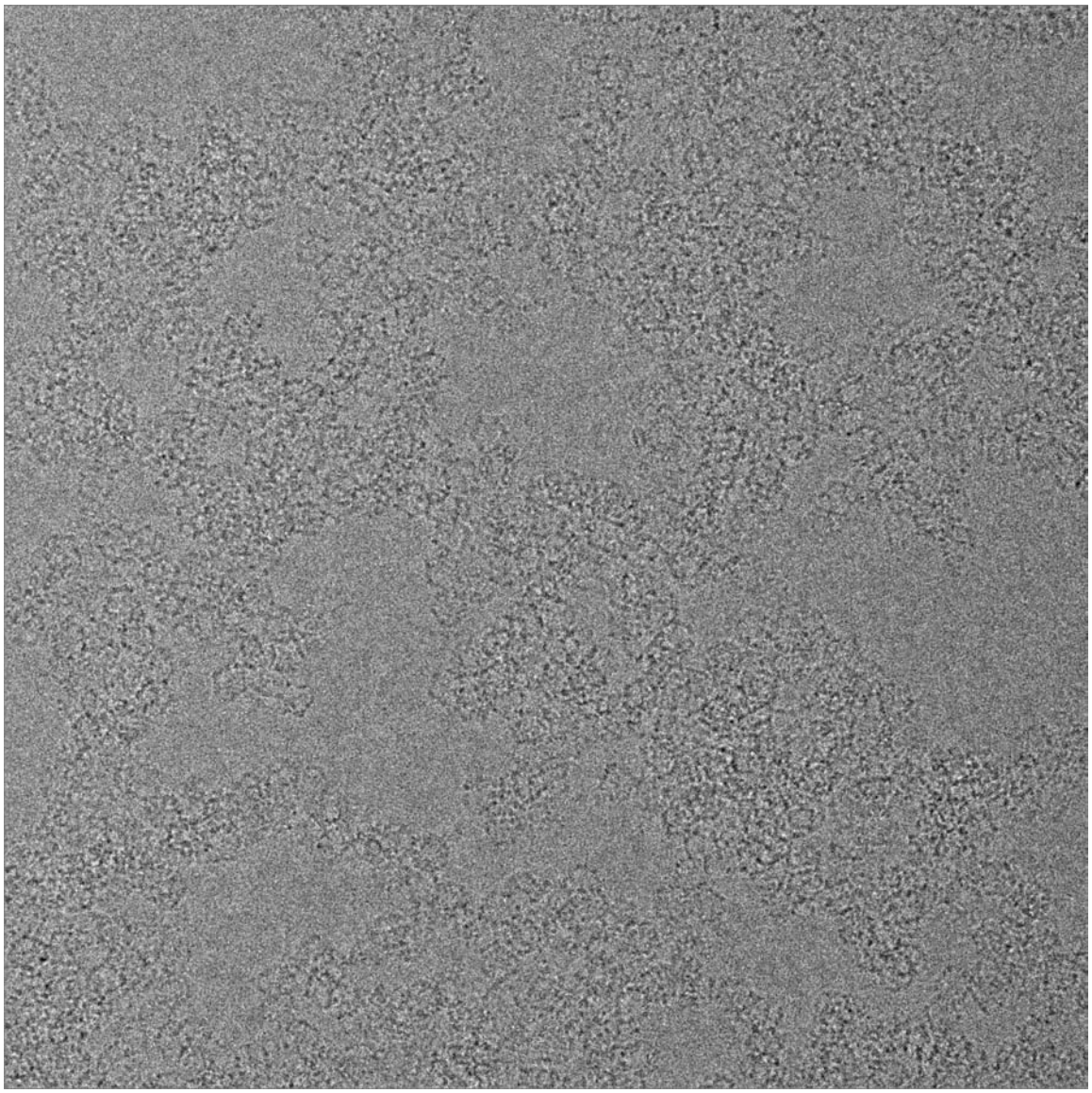


Figure 2. Initial Cryo-EM image of I3-01 empty cages.

Purified I3-01 without any fusions was frozen in liquid ethane and imaged by TEM.

Cages could clearly be discerned but were aggregated.

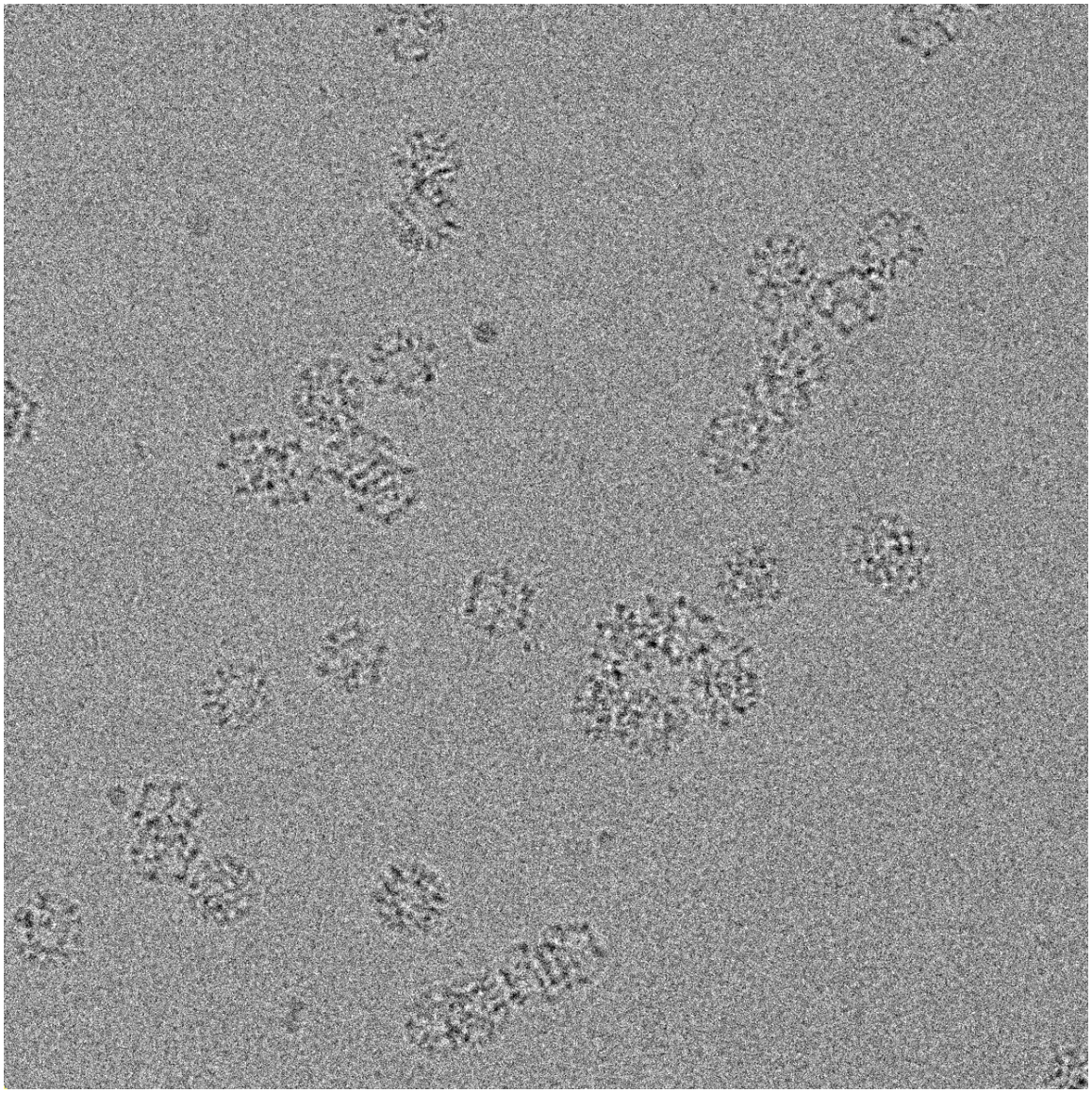


Figure 3. Empty I3-01 cages imaged by cryo-EM after sample and grid optimization. I3-01 without any fusions and after the addition of DTT during sample and grid prep was viewed using optimized grid conditions yielding more monodispersed cages.

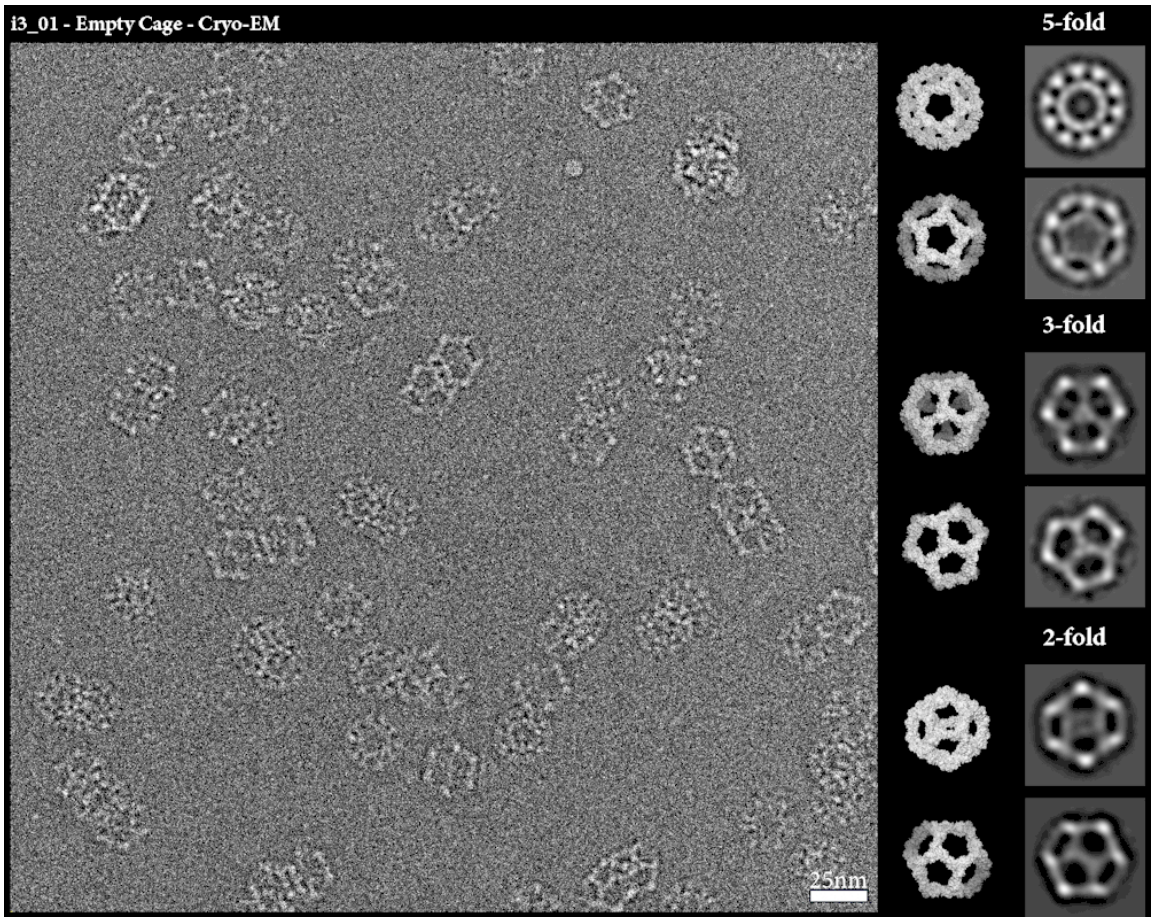


Figure 4. Cryo-EM averages of empty I3-01.

Motion corrected movies of empty I3-01 cages were used to create class averages. All three axes (2-, 3- and 5- folds) can be clearly discerned and match the design.

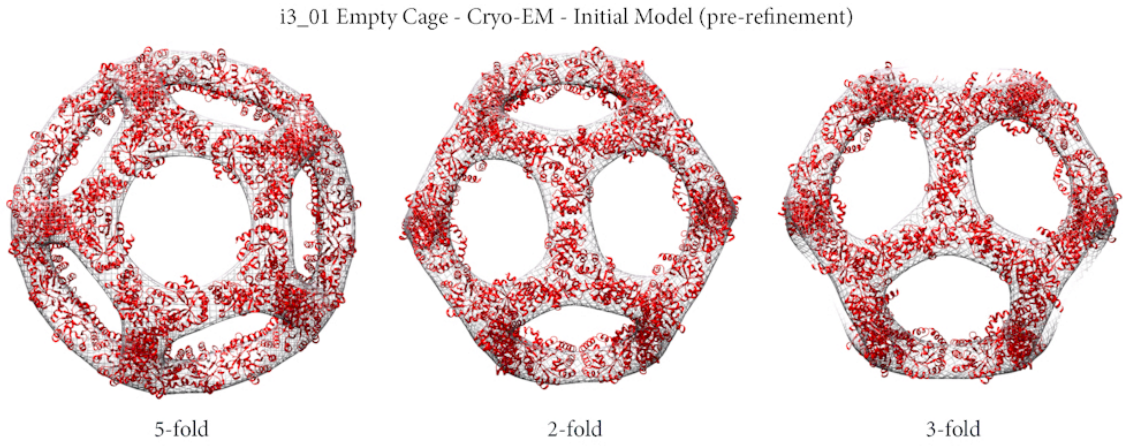


Figure 5. Initial density model of empty I3-01 cage.

Using the calculated class averages as templates, an initial density model was calculated in EMAN2 using the picked particles. Fitting the Rosetta output model reveals a close match between the design and model and all three axes were clearly present.

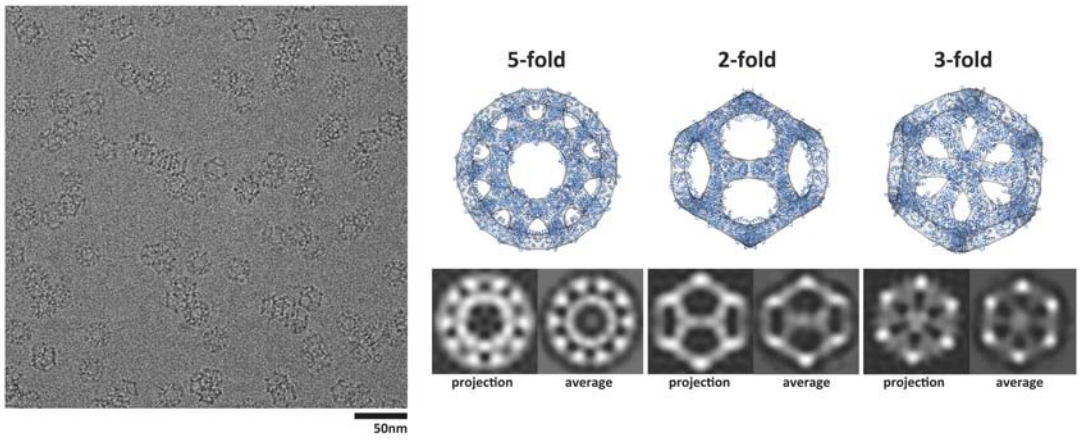


Figure 6. Montage showing Cryo-EM results from the empty I3-01 cage. Cryo-EM micrograph is shown highlighting the monodispersed cages from an optimized cryo grid. Back projections calculated from the I3-01 Rosetta model are shown next to averages calculated from cryo-EM micrographs.

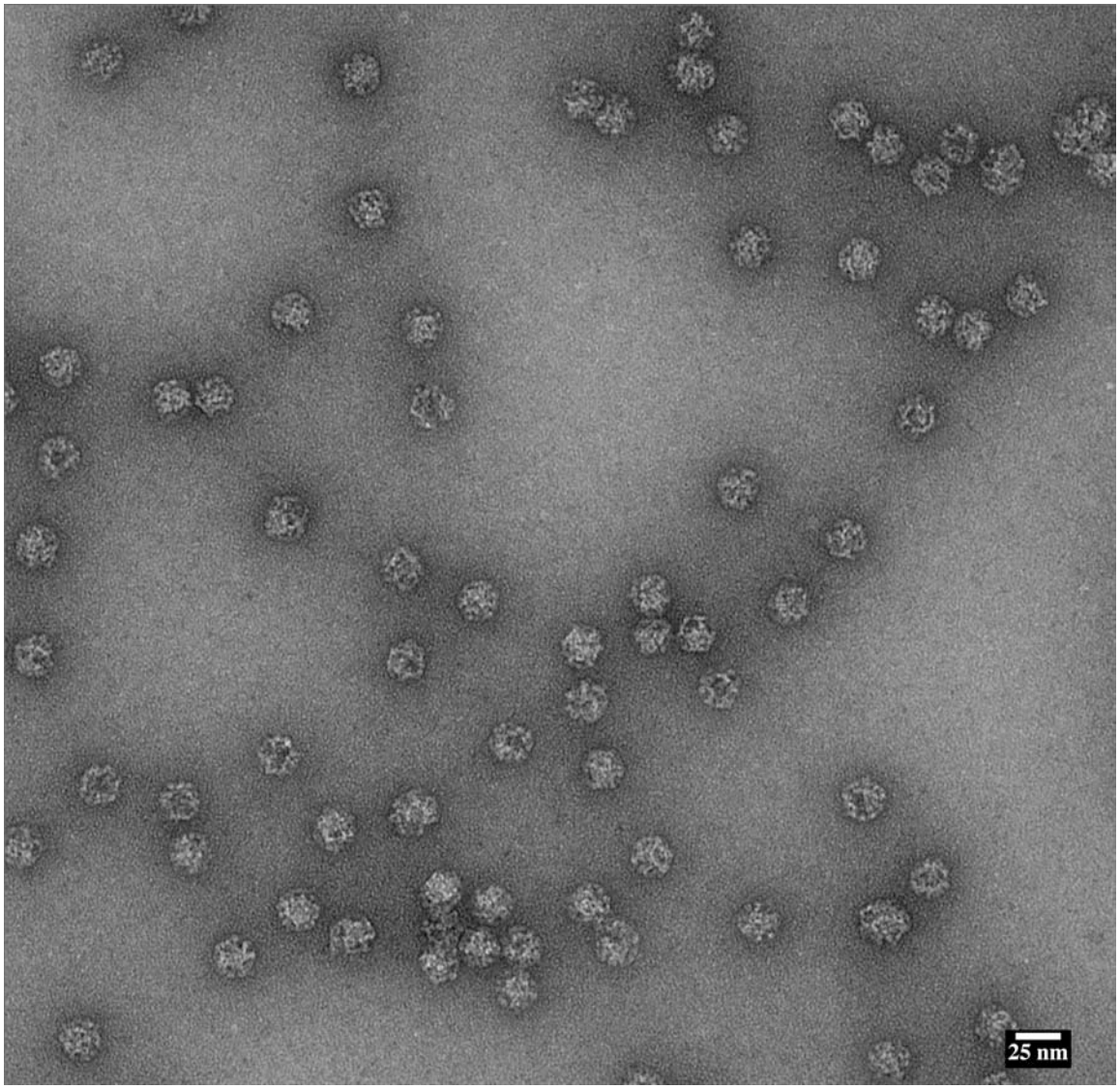


Figure 7. Negative-stain TEM of I3-01 cages fused to superfolder GFP.

Purified I3-01 without any fusions was stained with Uranyl Formate and imaged by TEM. Icosahedral cages could clearly be viewed and the cage density was clearly higher when compared to the empty I3-01 cages. Symmetrical features could also be discerned, suggesting the GFP was sticking to the inside of the cages.

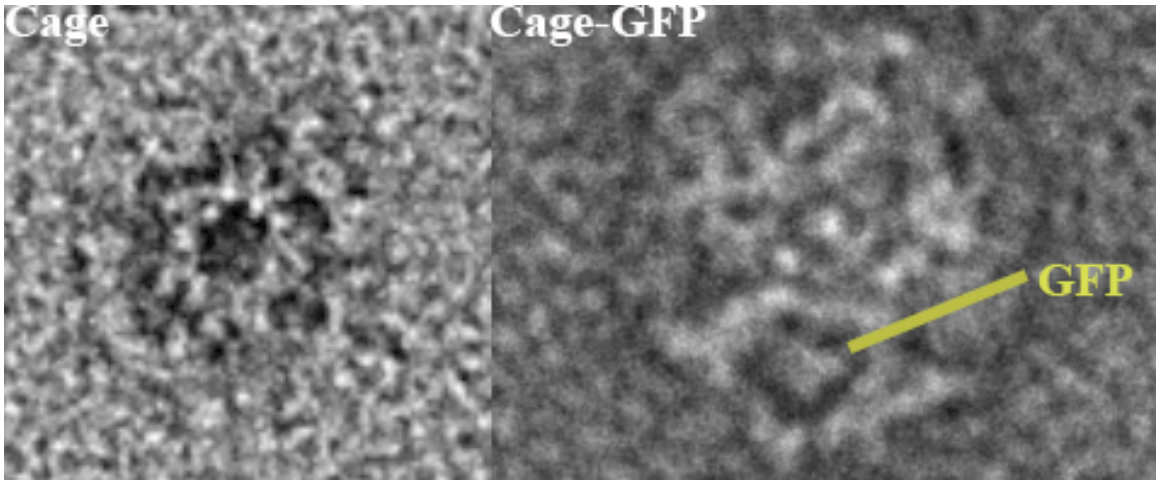


Figure 8. Initial interpretation of the GFP density compared to empty I3-01. Negatively-stained samples of both empty and superfolder-GFP fused I3-01's are compared. 5-fold symmetry could clearly be discerned in both and the added density is hypothesized to be that of GFP

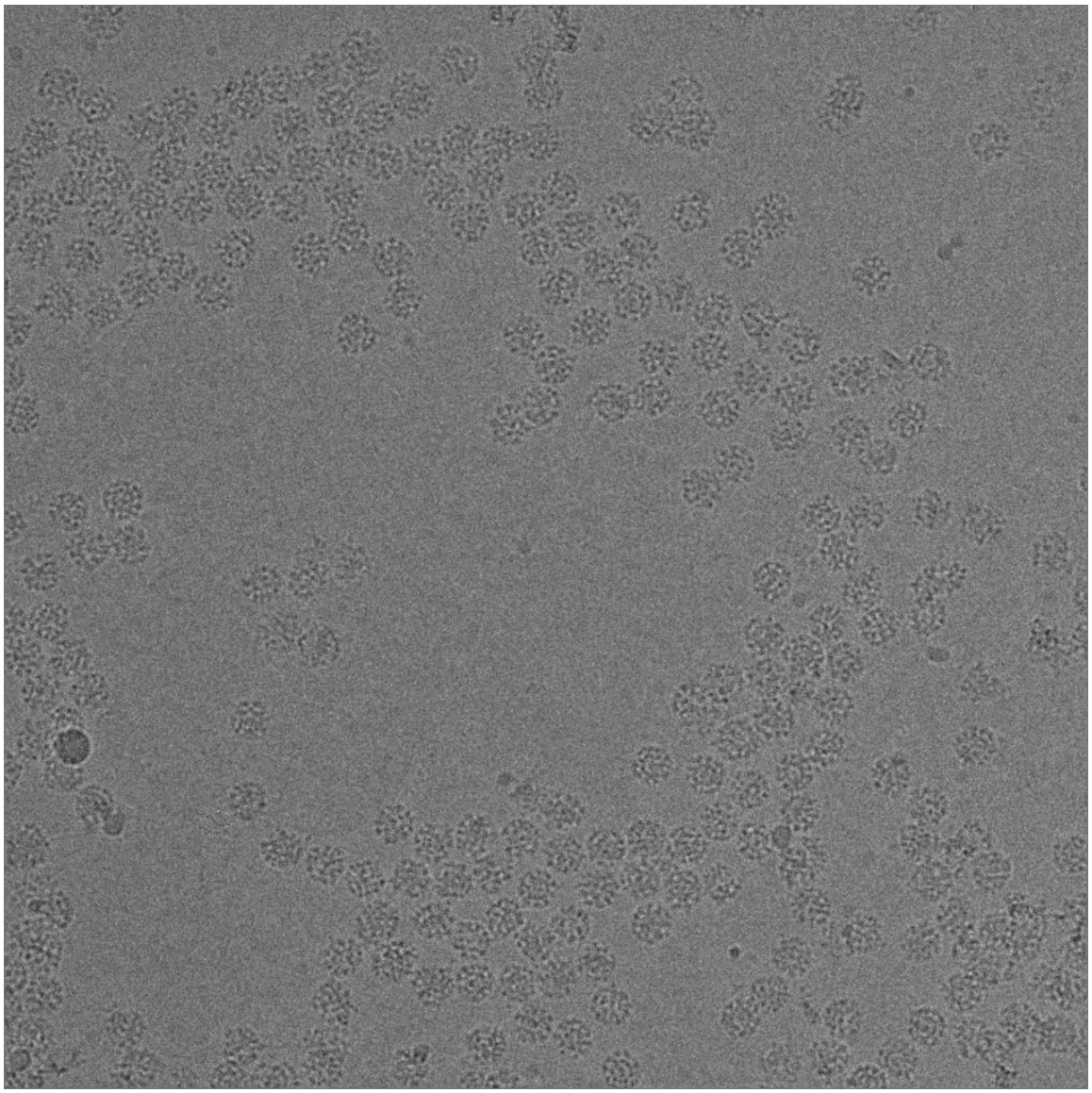


Figure 9. Initial screening of superfolder-GFP fused I3-01.

Initial images recorded by cryo-EM from optimized frozen grids of superfolder-GFP fused I3-01 showing monodispersed particles and the symmetric axes could clearly be discerned.

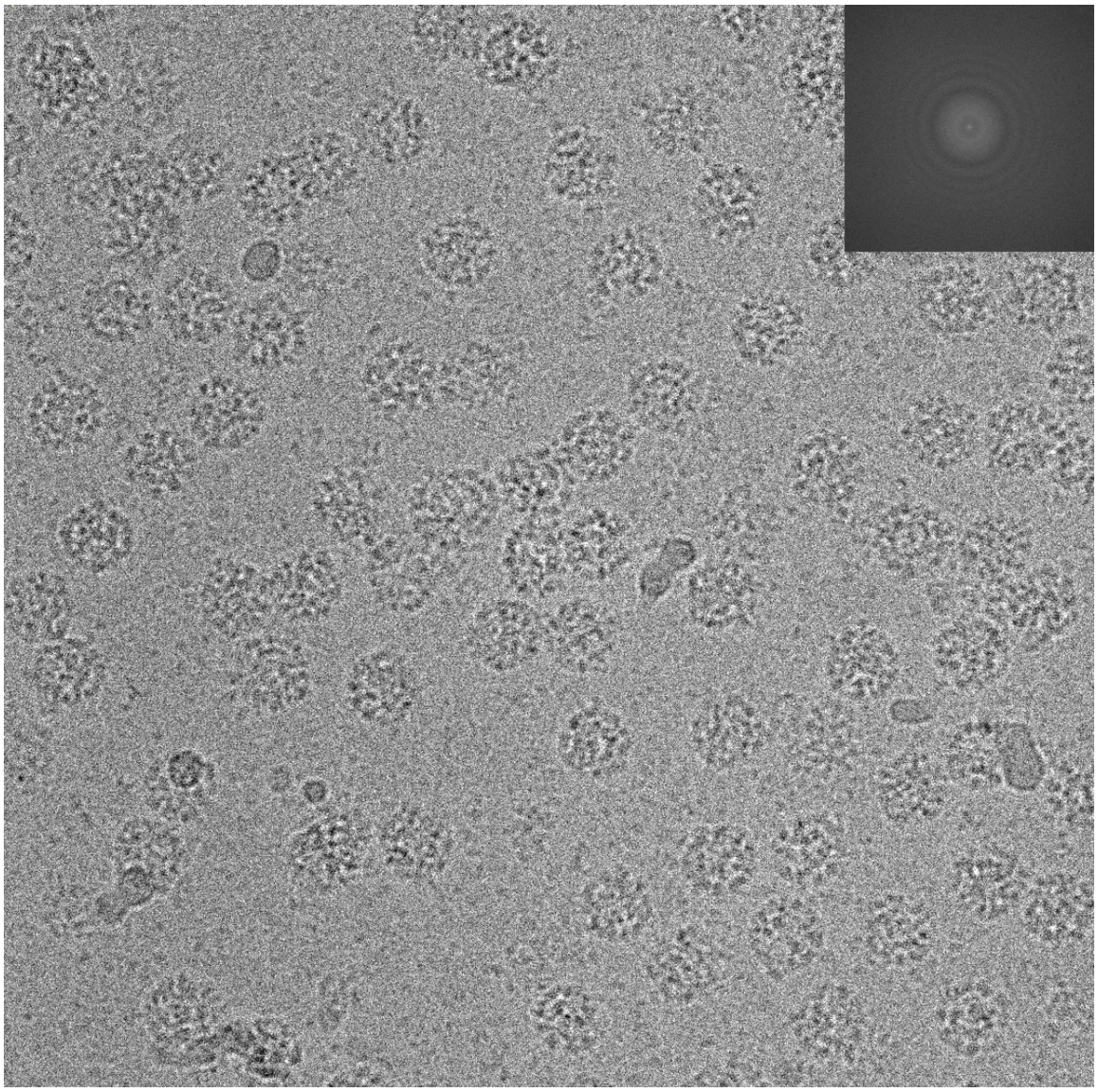


Figure 10. Motion corrected movie of superfolder-GFP fused I3-01.

Movies of superfolder-GFP fused I3-01's were collected and motion corrected showing high contrast I3-01 cages for class averaging.

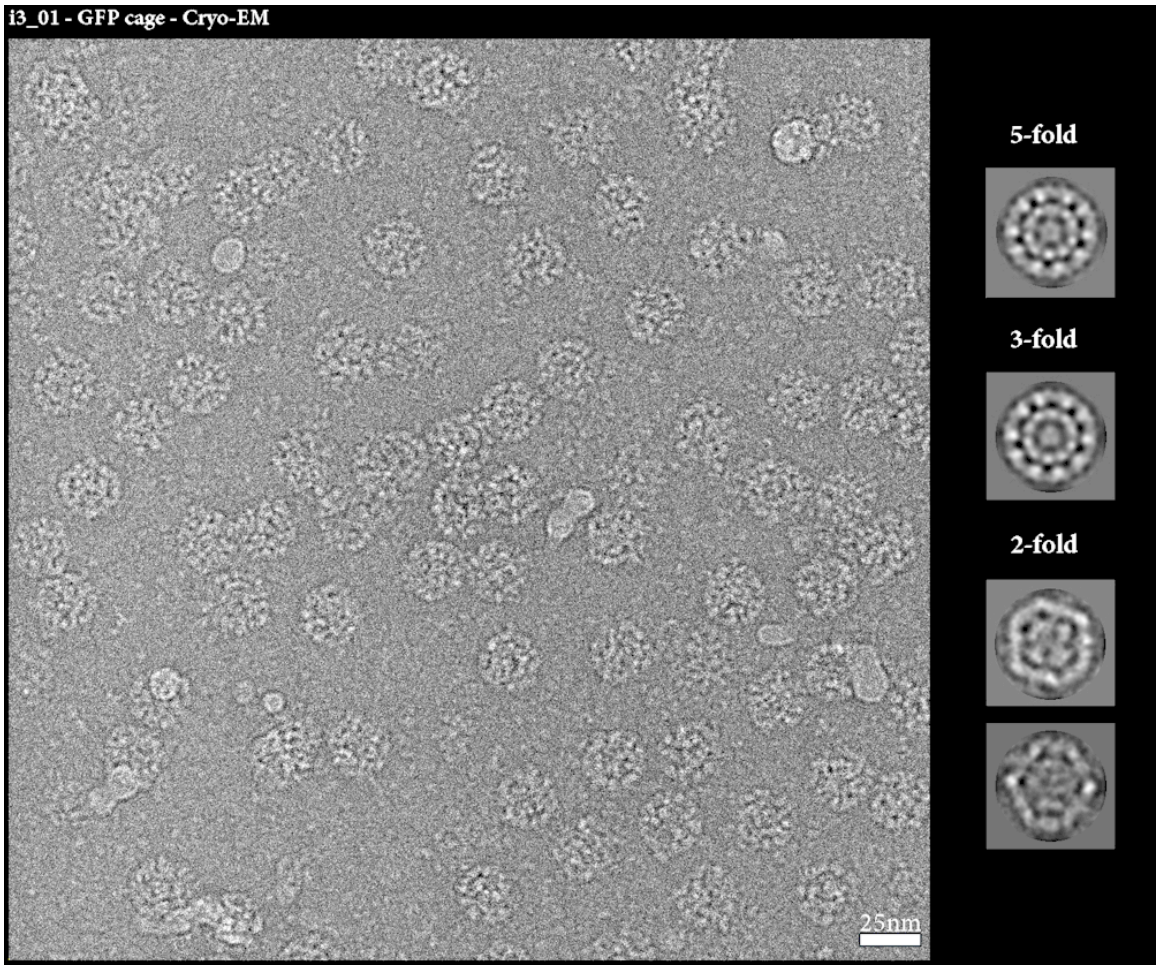


Figure 11. Class averages of superfolder-GFP fused I3-01

Motion corrected movies were used to create class averages of superfolder-GFP fused I3-01. The symmetric axes could be discerned and contain extra density compared to empty I3-01.

Two-component, Icosahedral protein cages:

The methodology applied to designing two-component tetrahedral cages highlighted in chapter IV can also be applied to designing nanomaterial with icosahedral symmetry. Icosahedral structures contain three distinct symmetrical axes, five-, three- and two-fold axes with a total of 60 subunits. Designing icosahedra from two different cyclic proteins while retaining symmetry can be done using several different protein combinations but in order to only have one designable interface, only certain building blocks can be used. The possible combinations are:

1. Trimers (C3) and dimers (C2) (I32)
2. Pentamers (C5) and dimers (C2) (I52) and;
3. Pentamers (C5) and trimers (C3) (I53)

In order to maximize the diversity of the outputted cages, design calculations on all three of the scaffolds combinations were performed. After rounds of biochemistry and screening by negative-stain TEM and optimization, a total of 12 designs ultimately showed promise for assembly called:

I32-06, I32-10, I32-19, I32-28

I52-03, I52-32, I52-33

I53-34, I53-40, I53-47, I53-50 and I53-51

I32-10, I32-19, I52-32 and I53-51 had various issues, including breaking of assembly and aggregation or in the case of I32-10, the wrong assembly size. These 4 designs were not continued with further class averaging.

A total of 8 class averages were made using optimized negative-stain grids and are highlighted in the following pages:

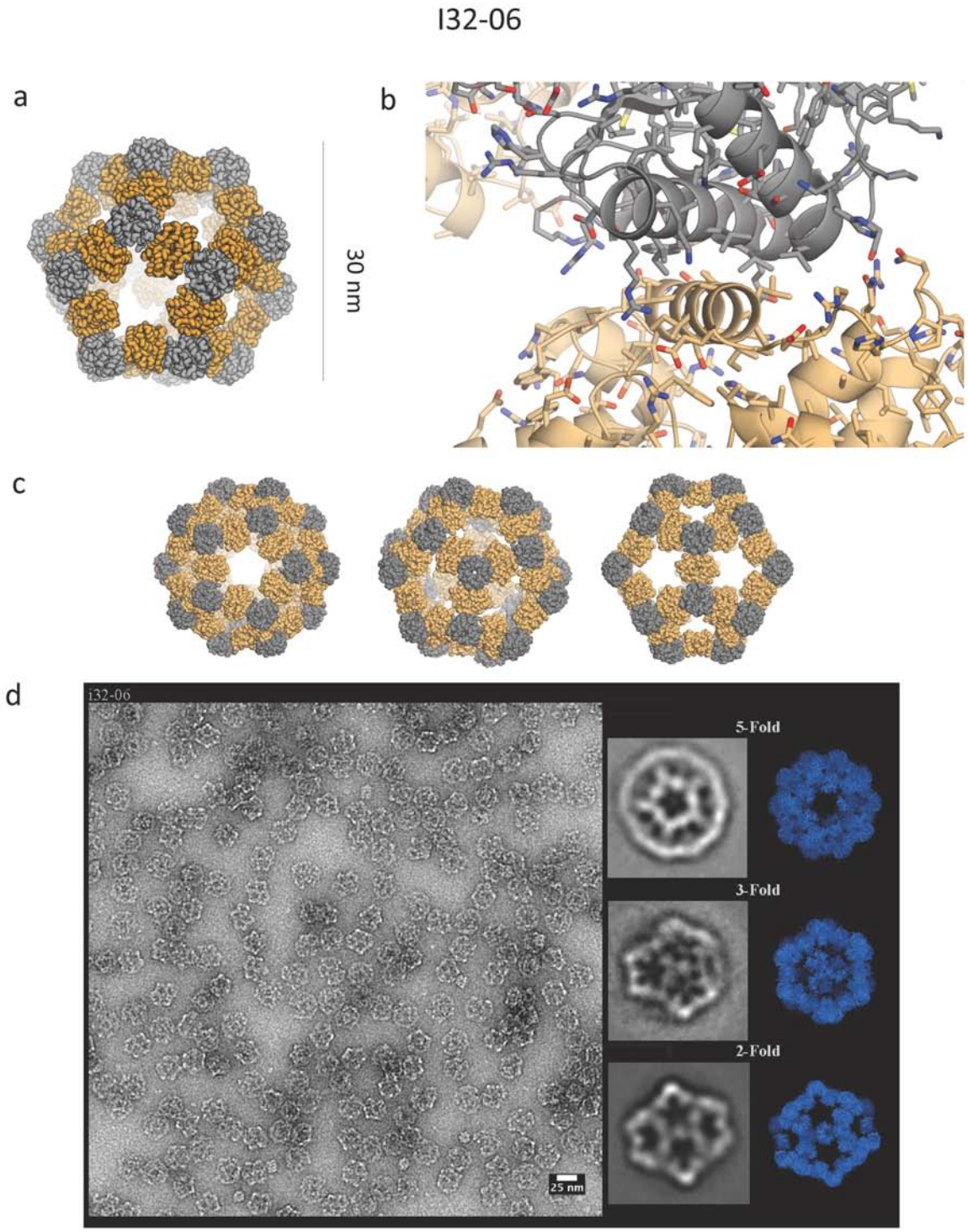


Figure 12. Icosahedral design I32-06 characterization. (a) Ribbon diagram version of the design model. (b) Interface between each trimer and dimer. (c) 5- 3- and 2- fold ribbon representations of the design. (d) TEM characterization of the design using negative-stain. Averages for each symmetric operator are shown alongside cartoon models of the same view from the design (blue)

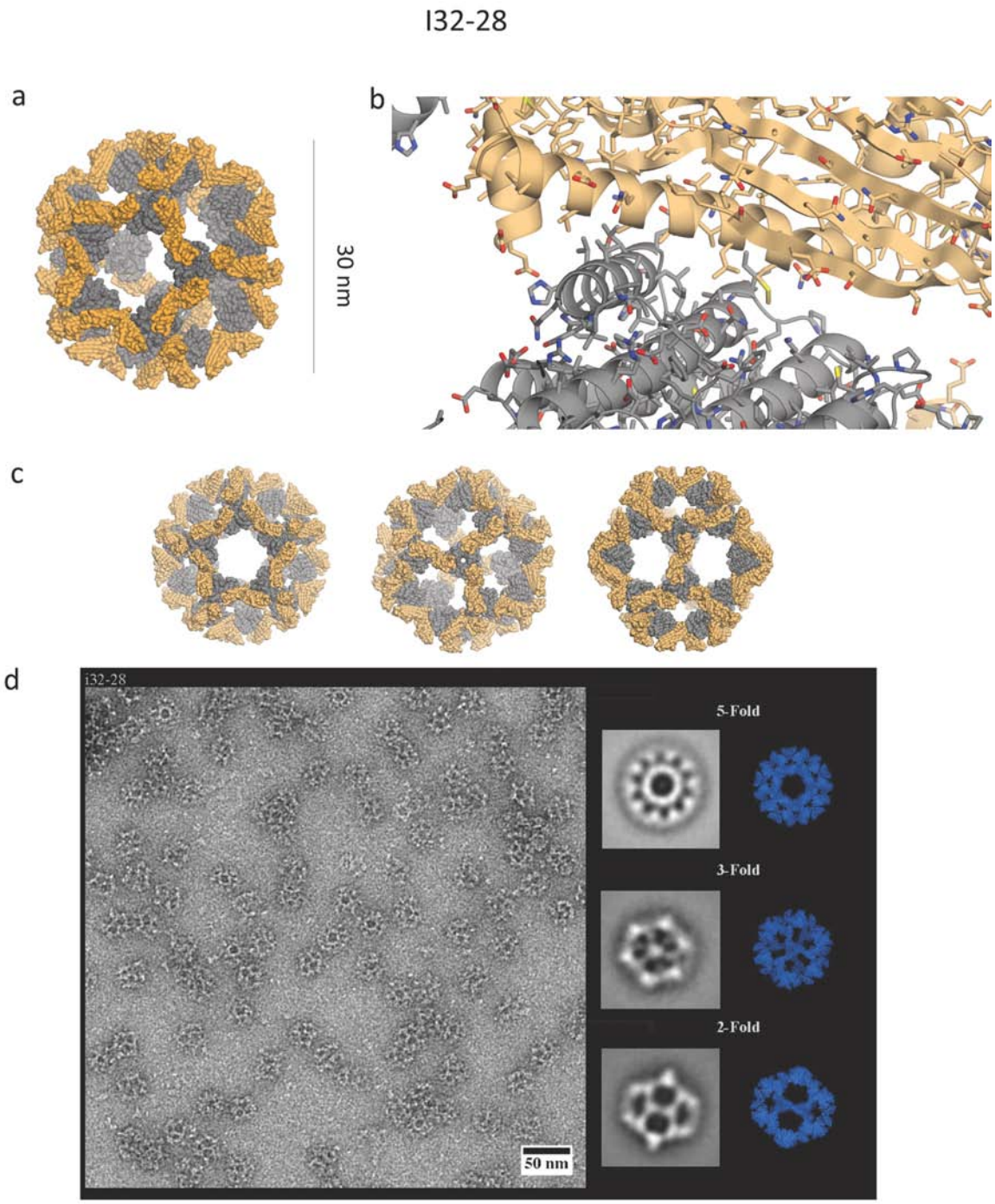


Figure 13. Icosahedral design I32-28 characterization. (a) Ribbon diagram version of the design model. (b) Interface between each trimer and dimer. (c) 5- 3- and 2- fold ribbon representations of the design. (d) TEM characterization of the design using negative-stain. Averages for each symmetric operator are shown alongside cartoon models of the same view from the design (blue)

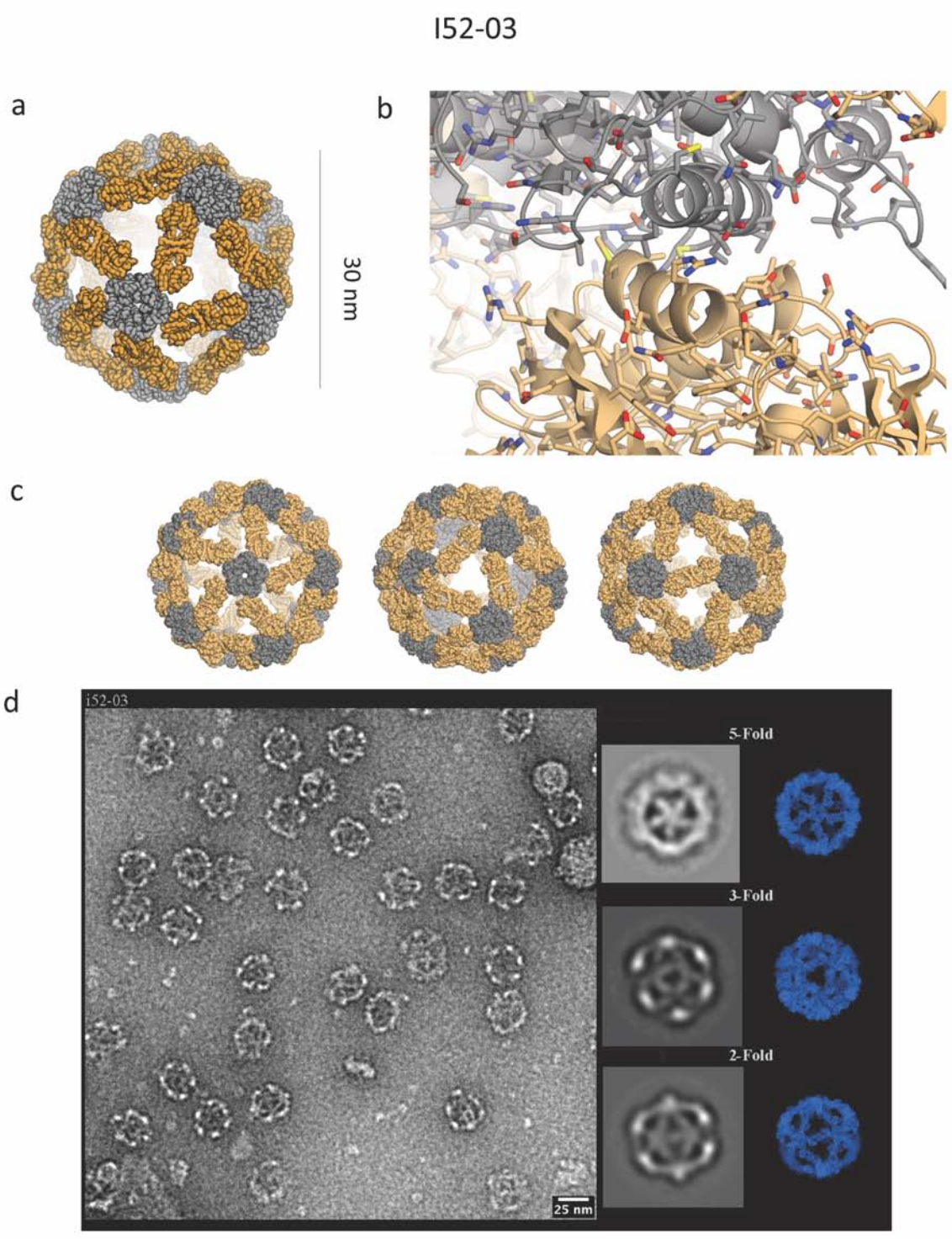


Figure 14. Icosahedral design I52-03 characterization. (a) Ribbon diagram version of the design model. (b) Interface between each pentamer and dimer. (c) 5- 3- and 2- fold ribbon representations of the design. (d) TEM characterization of the design using negative-stain. Averages for each symmetric operator are shown alongside cartoon models of the same view from the design (blue)

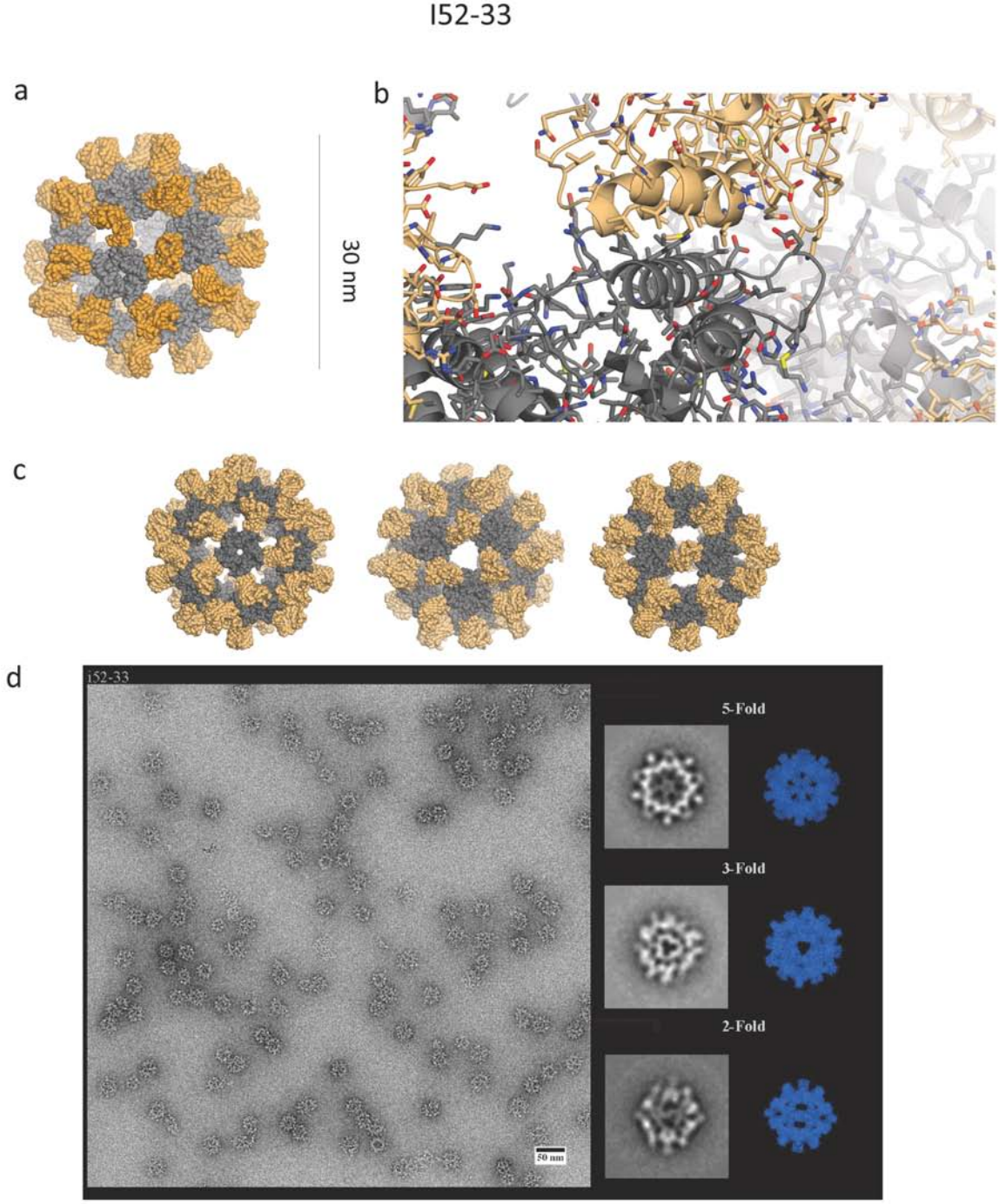


Figure 15. Icosahedral design I52-33 characterization. (a) Ribbon diagram version of the design model. (b) Interface between each pentamer and dimer. (c) 5- 3- and 2- fold ribbon representations of the design. (d) TEM characterization of the design using negative-stain. Averages for each symmetric operator are shown alongside cartoon models of the same view from the design (blue)

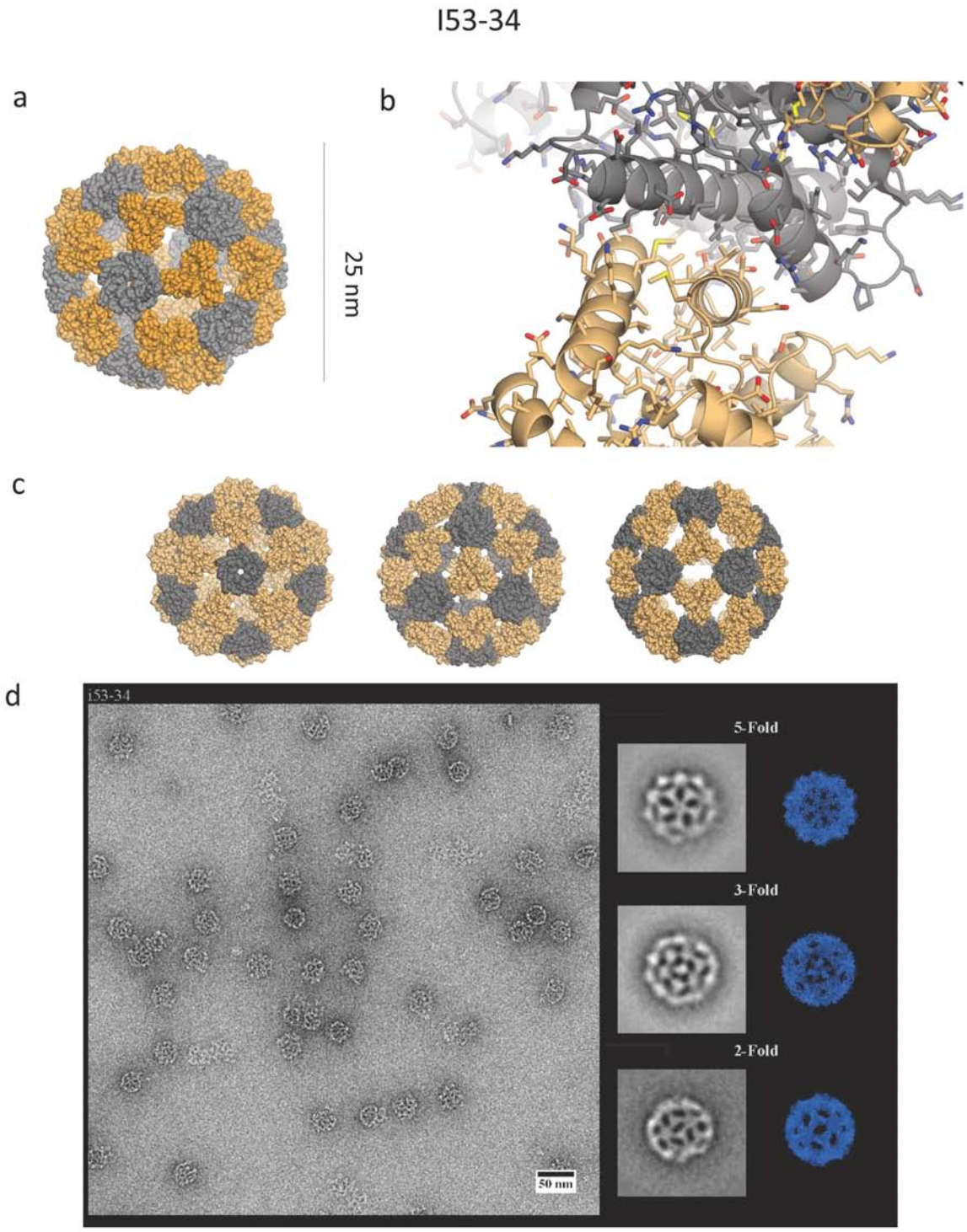


Figure 16. Icosahedral design I53-34 characterization. (a) Ribbon diagram version of the design model. (b) Interface between each pentamer and trimer. (c) 5- 3- and 2- fold ribbon representations of the design. (d) TEM characterization of the design using negative-stain. Averages for each symmetric operator are shown alongside cartoon models of the same view from the design (blue)

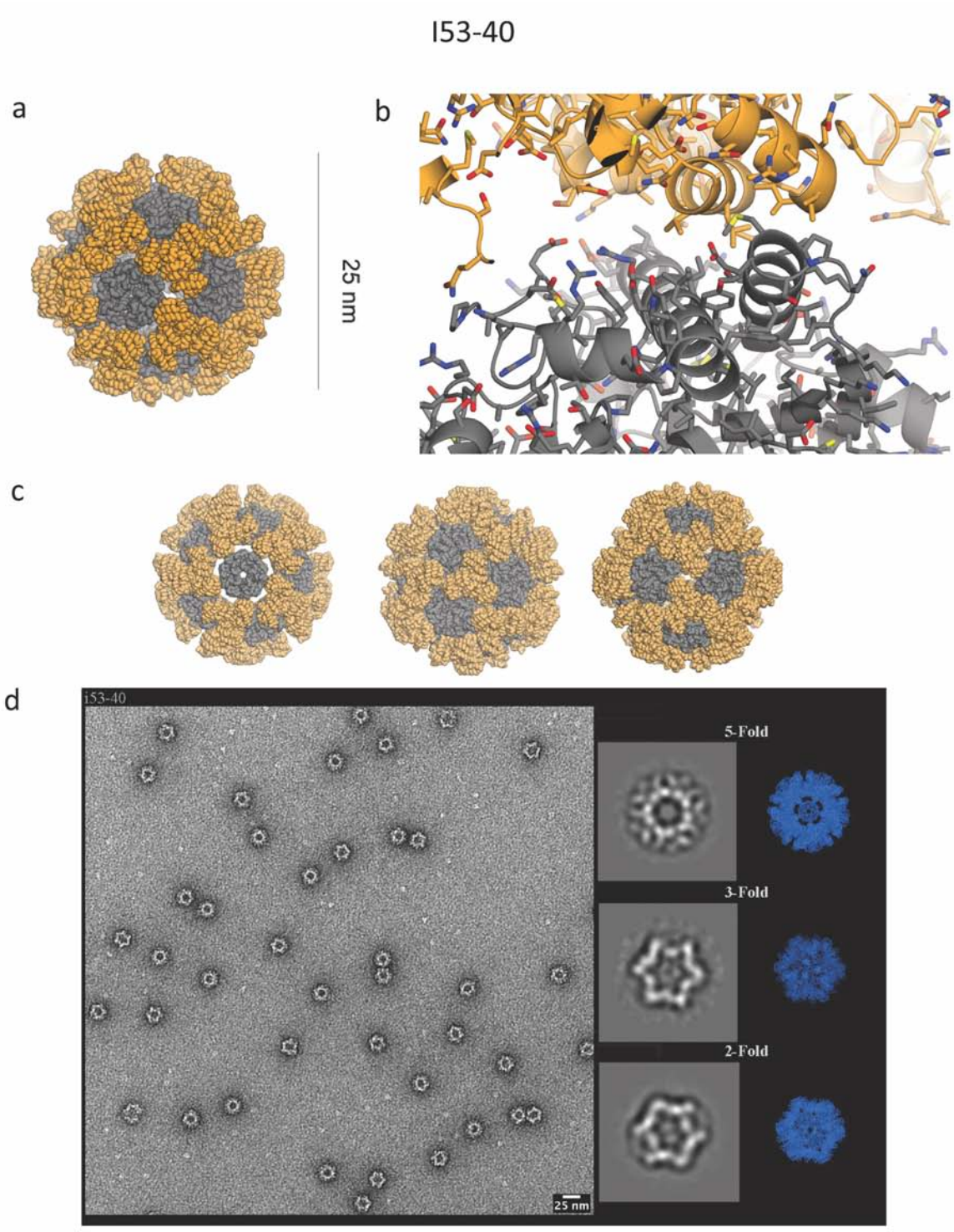


Figure 17. Icosahedral design I53-40 characterization. (a) Ribbon diagram version of the design model. (b) Interface between each pentamer and trimer. (c) 5- 3- and 2- fold ribbon representations of the design. (d) TEM characterization of the design using negative-stain. Averages for each symmetric operator are shown alongside cartoon models of the same view from the design (blue)

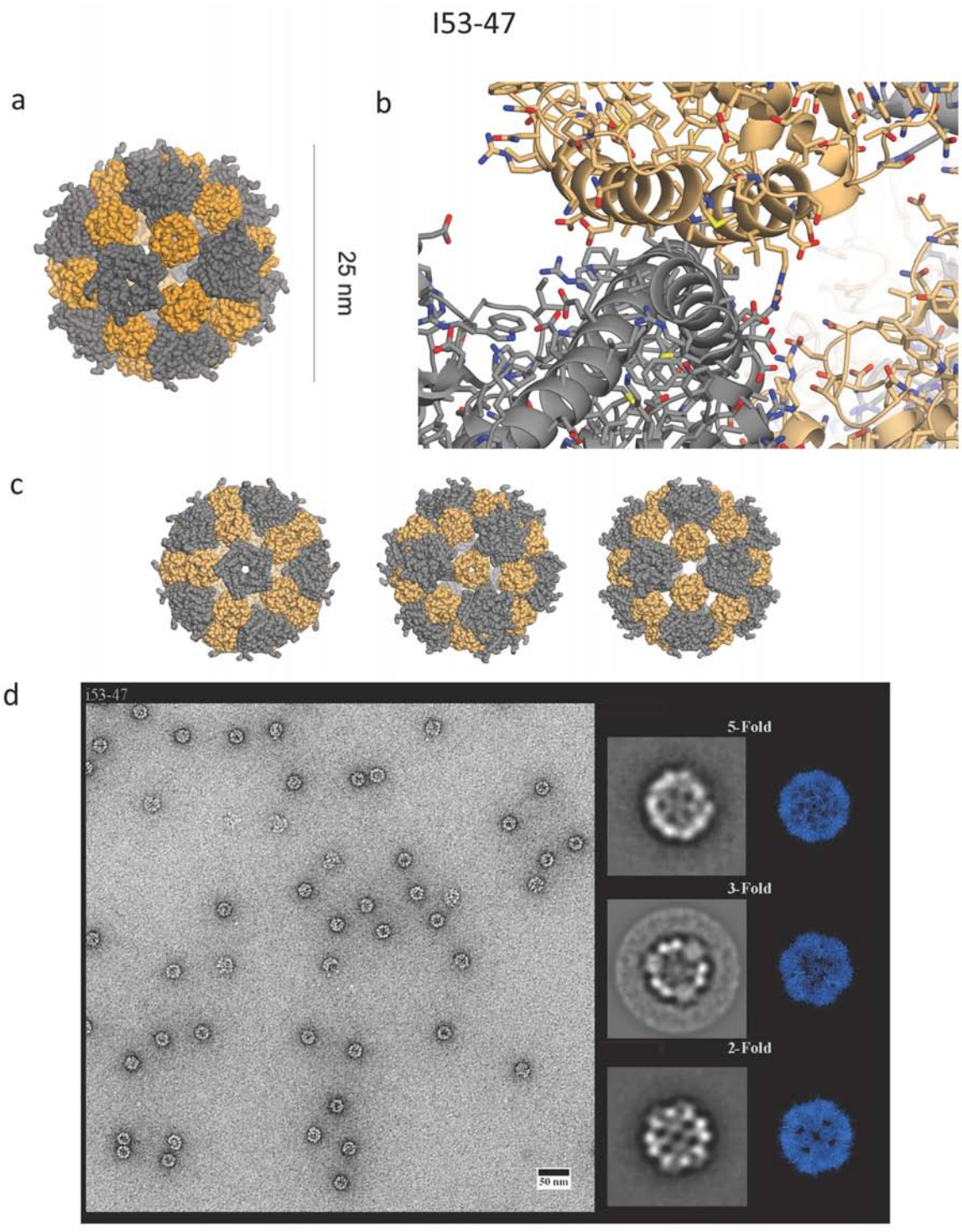


Figure 18. Icosahedral design I53-47 characterization. (a) Ribbon diagram version of the design model. (b) Interface between each pentamer and trimer. (c) 5- 3- and 2- fold ribbon representations of the design. (d) TEM characterization of the design using negative-stain. Averages for each symmetric operator are shown alongside cartoon models of the same view from the design (blue)

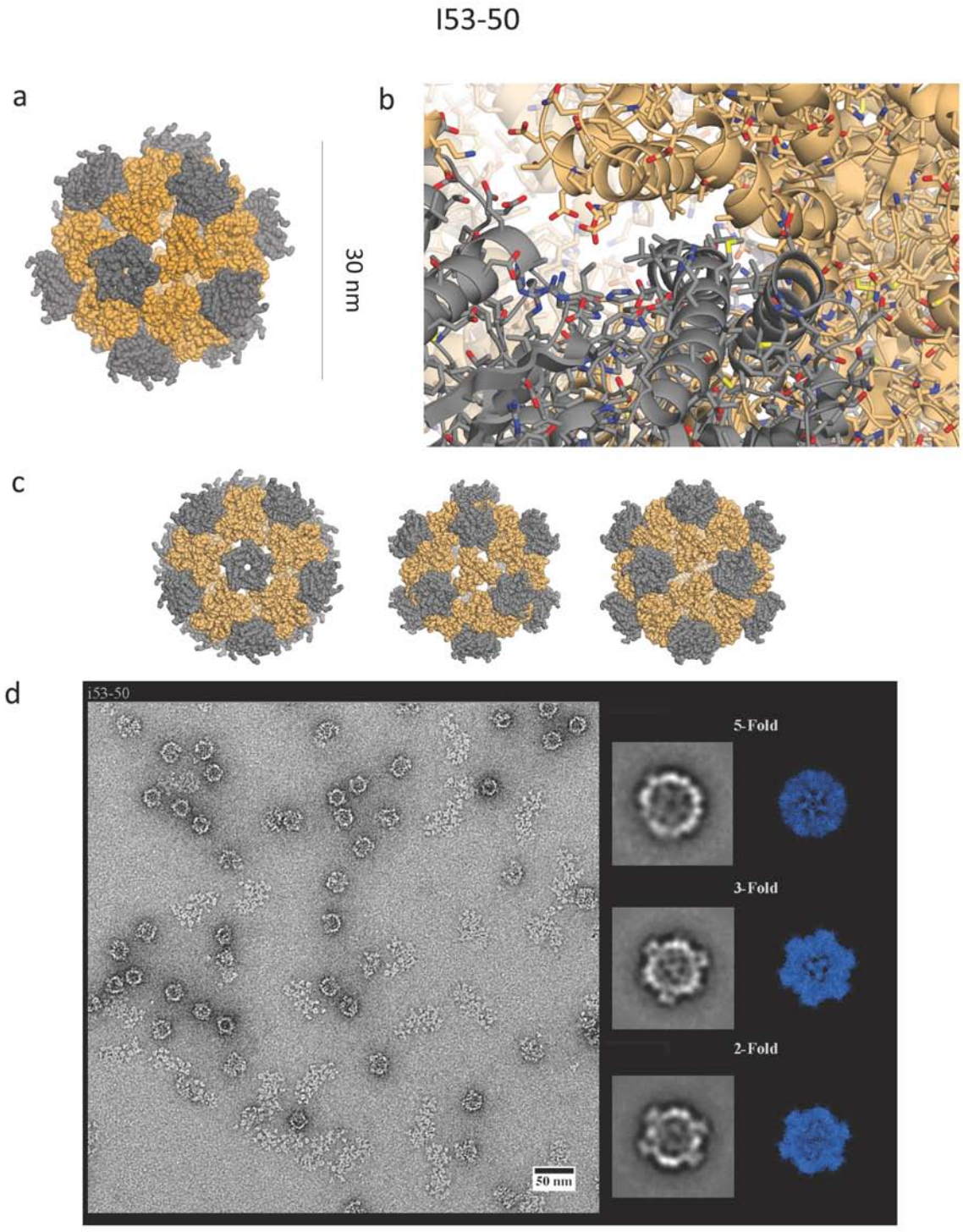


Figure 19. Icosahedral design I53-50 characterization. (a) Ribbon diagram version of the design model. (b) Interface between each pentamer and trimer. (c) 5- 3- and 2- fold ribbon representations of the design. (d) TEM characterization of the design using negative-stain. Averages for each symmetric operator are shown alongside cartoon models of the same view from the design (blue)

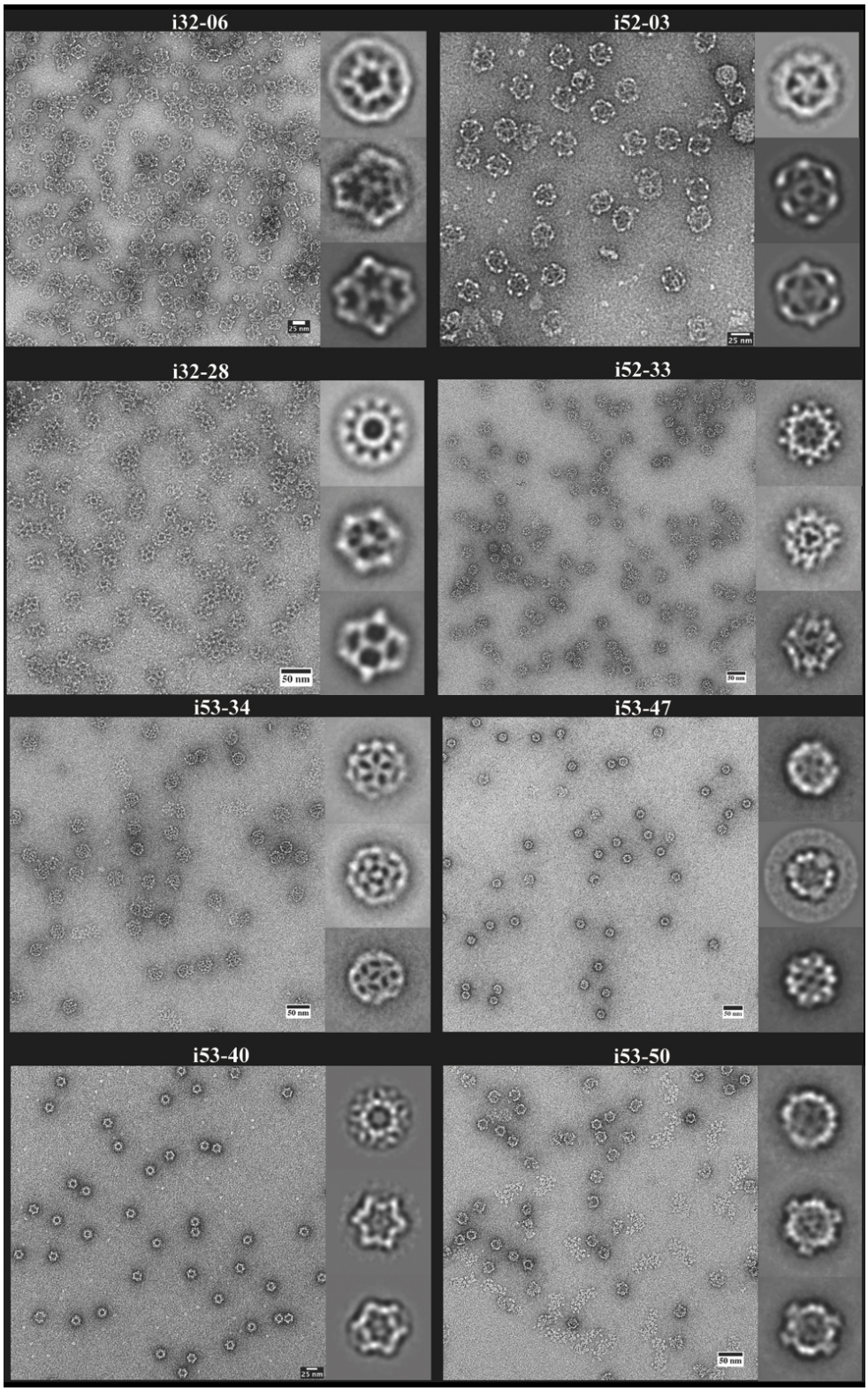


Figure 20. Overview of the 8 designs where averages were obtained.

Appendix V

Protocols & Recipes

Recipes & Buffers:

Lysis Buffer (without protease inhibitor):

150mM NaCl

25mM Tris pH 8.0

1mM DTT

Denaturing Buffer

6M Guanidinim Hydrochloride

150mM NaCl

25mM Tris pH 8.0

1mM DTT

20% PEG 8000/ 2.5M NaCl

Add the chemicals into a beaker with the required amount of MQ and add a magnetic stir bar – Autoclave

S.O.C. Media

1. Add the following for a 100mL SOC solution:

2g Tryptone

0.5g Yeast Extract

0.05g or 1mL of 1M NaCl

0.25mL 1M KCl

1mL 1M $MgCl_2$

1mL 1M $MgSO_4$

ddH_2O to make the solution reach 100mL

2. Autoclave

3. Once cool, add 2mL sterile 1M glucose

4. Aliquot

Wash and Elution Buffers for Ni/Co Resin

Wash – Add between 20mM to 40mM final Imidazole to Lysis buffer. Ramp up washes and check with Gels the correct amounts to use next time.

Elution – Add 500mM final Imidazole to Lysis Buffer

Glycerol stocks

1. Use 80% sterile glycerol 1:1 with overnight cell culture and freeze in the -80^{o}C

BCA Assay to check Protein concentration:

1. Make standards using 1mg/mL BSA – 0, 0.1, 0.2….0.6, 0.8, 1.0 mg/mL

2. Take 50ul of each into glass falcon brand tube

3. Make a 50x and 100x dilution of your protein into 50ul and 100ul respectively (take 50ul of each) (I.e. 1ul sample +49ul buffer and 1ul sample +99ul buffer)

4. BCA Kit: A 10mL (Large bottle)

B 200ul (Small blue bottle)

Mix and put 1mL into each tube, vortex and place in 37c for 25min

5. Check abs @562nm and compare to standard

Uranyl Formate Recipe:

1. Add 5mL boiled ddH2O to 37.5mg Uranyl Formate and stir for 5min under foil

2. Add 6uL 5M NaOH and continue stirring for 5min

3. Filter through 0.22um filter into foil wrapped tube

Transformation Protocol:

Keep competent cells on ice

1. On ice, thaw competent cells and place 15mL culture tubes to cool.

2. Aliquot 30ul of competent cells (BL21 DE3* for protein expression or 100ul of XL10-Gold or Top10 competent cells for cloning) (in-house stock) to each culture tube.

3. Add 1ul (more for cloning and if concentration is an issue) of pure DNA to each (add MQ Water to the last as a negative control and keep on ice for 5 min.

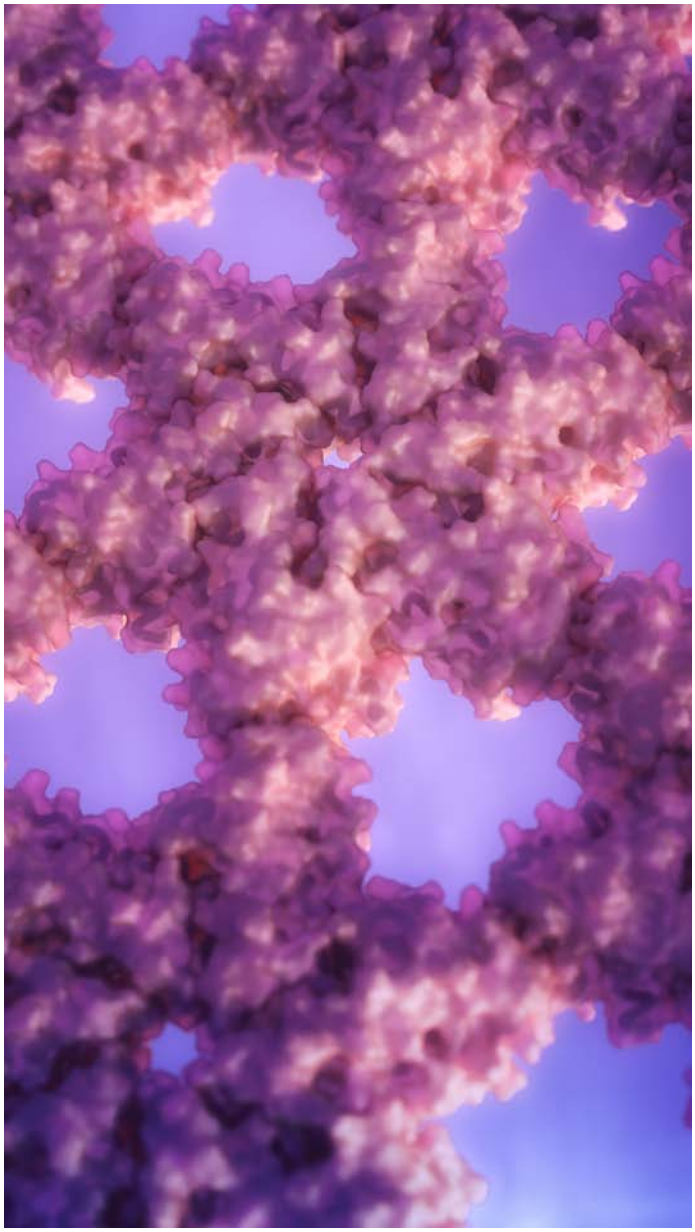
4. Plunge into 42^{o}C water bath for 45sec and back on ice for 2 min

5. Add 250ul SOC media to each tube and put in shaker on 225rpm for 1hr

6. Plate ~200ul for cloning or ~50ul for protein expression (dependent on the clone, this will likely need tinkering) on appropriate antibiotic LB-Agar plates with beads or spreader and put in 37^{o}C Incubator for 12-16 hours.

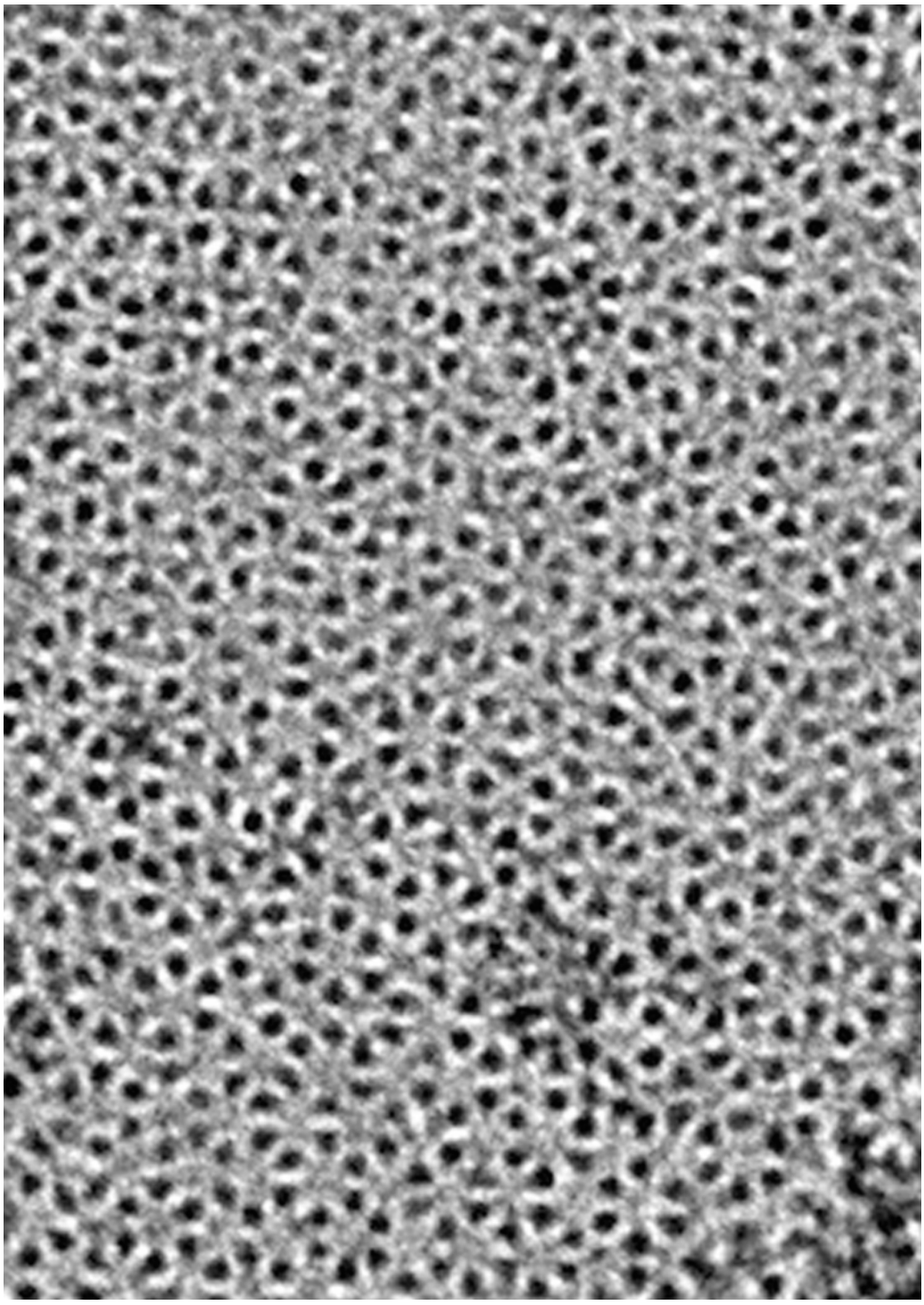
7. Wrap with parafilm around the outside and keep in the cold room for no more than a few days (even less for Ampicillin plates).

Covers and Posters



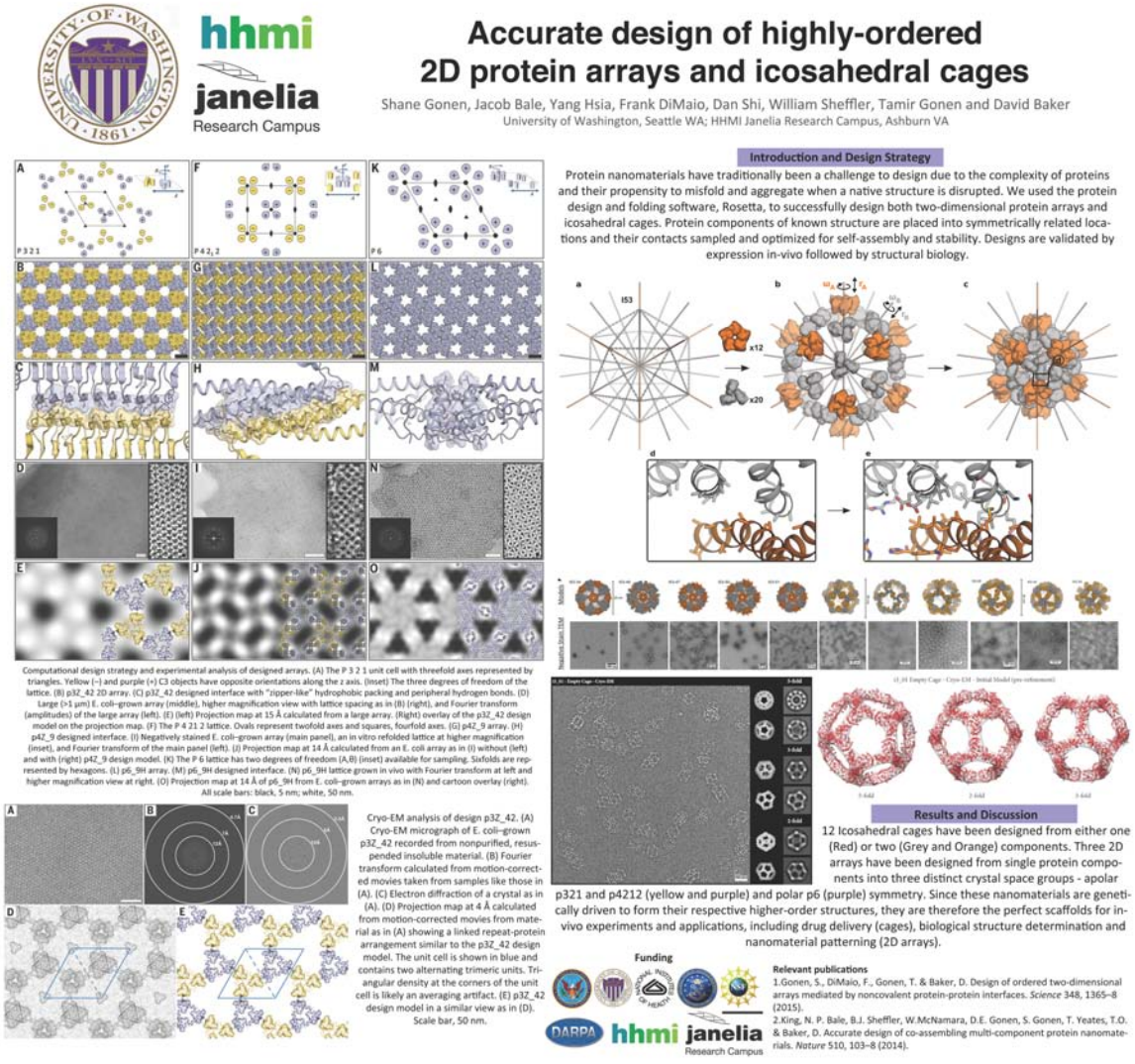
Cover suggestion – p6_9H design

Rendering created by Vikram Mulligan from the p6_9H rosetta model for a cover suggestion to Science. It was featured as a research highlight.



Cover suggestion – p6_9H_KDKCKXX construct

Negatively-stained p6_9H_KDKCKXX construct for a cover suggestion to Science



Poster highlighting 2D arrays and Icosahedral cages – Shown at a Janelia Research Campus Graduate Student Symposium

Vita

Shane Gonen was born in the 20th century. He grew up partly on a small island in the south pacific and began his graduate studies in 2011 in Seattle, WA.