Design of a hyperstable 60-subunit icosahedral nanocage

Yang Hsia

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2017

Reading Committee:

David Baker, Chair

Wendy E. Thomas

Paul Wiggins

Program Authorized to Offer Degree:

Biochemistry

University of Washington

Abstract

Design of a hyperstable 60-subunit icosahedral nanocage

Yang Hsia

Chair of the Supervisory Committee:
Professor David Baker
Biochemistry

The icosahedron is the largest of the Platonic solids, and icosahedral protein structures are widely used in biological systems for packaging and transport[1,2]. There has been considerable interest in repurposing such structures[3–5] for applications ranging from targeted delivery to multivalent immunogen presentation. The ability to design proteins that self-assemble into precisely specified, highly ordered icosahedral structures would open the door to a new generation of protein 'containers' with properties custom-tailored to specific applications. Here we describe the computational design of a 25 nm icosahedral nanocage that self-assembles from trimeric protein building blocks. The designed protein was produced in *Escherichia coli*, and found by electron microscopy to assemble into a homogenous population of icosahedral particles nearly identical to the design model. The particles are stable in 6.7 molar guanidine

hydrochloride at up to 80 degrees Celsius, and undergo extremely abrupt, but reversible, disassembly at 2–2.25 molar guanidinium thiocyanate. The icosahedron is robust to genetic fusions: one or two copies of superfolder green fluorescent protein (GFP) can be fused to each of the 60 subunits to create highly fluorescent standard candles for use in light microscopy, and a designed protein pentamer can be placed in the center of each of the twelve pentameric voids to gate macromolecule access to the interior of the nanocage. Such robust and customizable nanocages should have considerable utility in targeted drug delivery[6], vaccine design[7], and synthetic biology[8].

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGEMENTS

For the original manuscript titled *Design of a hyperstable 60-subunit protein icosahedron* (DOI: 10.1038/nature18010); J.B.B., N.P.K., and W.S. developed the computational design methodology. Y.H. and J.B.B. performed the design of the icosahedra. Y.H. performed all other unlisted experiments. S.G. and D.S. performed the cryo-EM experiments. K.K.F. performed the fluorescence microscopy experiments. U.N. performed the negative-stain electron microscopy experiments. C.X. provided the pentamer sequence for I3-01(HB). P.S.H. created the computational methodology to model fusions to I3-01. R.R. produced I3-01(HB) proteins. S.Y. produced T33-21 sfGFP fusion proteins. Y.H., N.P.K, and D.B. wrote the manuscript.

This work was supported by the Howard Hughes Medical Institute (D.B. and T.G.), the JRC visitor programme (S.G.), the National Science Foundation (NSF) CHE-1332907 (D.B.), a UW/Hutch CCSG Pilot Award NCI 5 P30 CA015704-41 (D.B. and N.P.K.), Takeda Pharmaceutical Company (N.P.K.), the Bill and Melinda Gates Foundation OPP1120319 (D.B. and N.P.K.), the National Institutes of Health (NIH) P41 GM103533 (T.N.D.), the Defense Advanced Research Projects Agency (D.B. and N.P.K., grant no. W911NF-14-1-0162), and the Air Force Office of Scientific Research (AFOSR) AFOSR FA950-12-10112 (D.B.). Y.H. was supported in part by a NIH Molecular Biology Training Grant (T32GM008268). U.N. was supported in part by a PHS National Research Service Award (T32GM007270) from NIGMS. J.B.B. was supported in part by an NSF Graduate Research Fellowship (DGE-0718124). We thank the Janelia Research Campus Cryo-EM Facility and J. de la Cruz for their assistance with the Titan Krios.

Name Abbriviations:

Yang Hsia[1,2,3], Jacob B. Bale[1,2,4], Shane Gonen[1,2,3,6], Dan Shi[6], William Sheffler[1,2], Kimberly K. Fong[1], Una Nattermann[1,2,3], Chunfu Xu[1,2], Po-Ssu Huang[1,2], Rashmi Ravichandran[1,2], Sue Yi[1,2], Trisha N. Davis[1], Tamir Gonen[5,6], Neil P. King[1,2], and David Baker[1,2,5].

Affiliation:

[1]Department of Biochemistry, University of Washington, Seattle, Washington 98195, USA.

[2]Institute for Protein Design, University of Washington, Seattle, Washington 98195, USA.

[3]Graduate Program in Biological Physics, Structure & Design, University of Washington, Seattle, WA 98195, USA.

[4]Graduate Program in Molecular and Cellular Biology, University of Washington, Seattle, WA 98195, USA.

[5]Howard Hughes Medical Institute, University of Washington, Seattle, Washington 98195, USA.

[6]Janelia Research Campus, Howard Hughes Medical Institute, Ashburn, Virginia, 20147, USA.

Chapter 1. INTRODUCTION

1.1 SYMMETRICAL NANOMACHINES IN NATURE

Nature has evolved proteins to be able to accomplish a wide variety of tasks ranging from enzymes catalyzing reactions to providing structural support like actin filaments. In order to be such a versatile material, proteins must be able to fold into a wide variety of structures. However, as the required task becomes more complicated, so must the protein that needs to perform the task. The obvious solution to this problem is to have increasingly large and complex proteins to face these challenges. Unfortunately, nature does not have an unlimited source of materials, thus limiting absolute protein size because of efficiency. To bypass this, nature utilizes the concept of symmetrical oligomerization of proteins[9], bringing significant advantages to a protein system.

Oligomerization allows for more complex functions that its building blocks alone cannot accomplish, similar to having a single large protein. Membrane channels, for example, oligomerize to create pores that allow molecules to pass through hydrophobic membranes[10]. Many enzymes are oligomeric for stability and cooperativity[11], or in the case of ferritin, oligomerization also separates substrate from product[12]. In structural cases such as clathrin, the subunits work together to encapsulate and transport molecules[13]. Just as advanced function is the default trait of larger proteins, larger complexes inherently are more biochemically stable against denaturation, as they incorporate more internal interactions and have a reduced surface area in general[14].

One of the most important issues that oligomerization solves, as stated previously, is to overcome a limited economy. By having relatively smaller building blocks, the respective gene sequence can also become significantly shorter. In the case for most viruses, this is a trait that is

heavily selected for as the protein capsid needs to encapsulate its own nucleic acid[14]. Similarly, symmetrical assemblies allow for these higher order structures to assemble using minimal amounts of unique interfaces. An icosahedral protein capsid only needs two unique interfaces to assemble; one for the monomers to assemble into cyclic oligomers (C2, C3, or C5) along a cyclic axis, and one for the cyclic oligomers to assemble into the overall icosahedral symmetry. Overall, symmetry is an important concept that can be applied to the design of higher order protein architectures to provide efficiency both computationally through the identical residues, and experimentally through shorter gene sequences.

1.2 Protein Design

Programming protein subunits to self-assemble into well-defined complexes is a promising route to the custom design of macromolecular machines. Protein assemblies have been engineered using metals[15,16], disulfide bonds[17–20], genetic fusions[18,21–23], and ideal helix–helix interactions[17,19,22], but these approaches have generally yielded polydisperse or unanticipated products. Recently, symmetric modelling coupled with computational protein–protein interface design has accurately generated protein assemblies with tetrahedral and octahedral symmetry[24,25], but these relatively small (<16 nm diameter) nanocages have limited use for packaging or delivery applications because they have little internal volume. Here we set out to design larger protein assemblies, specifically aiming at icosahedral architectures.

Rosetta, the computational software suite[26] for protein modelling in recent years have been modified for protein design. Originally, Rosetta was developed for protein fold predictions with two basic features: 1) score functions based on statistically and empirically derived potentials and 2) search algorithms for sampling amino acid identities and side-chain, backbone,

and rigid body conformations. When given a sequence, Rosetta was optimized to predict the lowest energy structure the sequence is likely to adopt. To perform protein design, the pipeline is essentially flipped backwards. Given a protein structure, in this case a target designed structure, Rosetta predicts the lowest energy sequence that best fits that structure. By applying symmetry, Rosetta only needs to calculate the energy of one designed interface and it can duplicate this symmetrical interface to the respective homo-oligomeric building blocks[27].

Chapter 2. COMPUTATIONAL DESIGN

2.1 SCAFFOLD SELECTION

Icosahedral point group symmetry contains two-, three-, and five-fold axes of rotation (Fig. 1a). To generate novel icosahedral protein assemblies, trimeric protein scaffolds of known structure were selected from the Protein Data Bank (PDB). Only structures with resolution better than 2.5 Å, that have well defined secondary structures on potential design surfaces, and that do not have multiple domains were selected.

2.2 SYMMETRICAL DOCKING

For each scaffold, 20 trimeric building blocks were arranged in icosahedral symmetry by aligning the three-fold rotational axis of each trimer with one of the three-fold icosahedral symmetry axes. While preserving symmetry, the building blocks were then docked together by enumeratively sampling their rotations (*w*) about the three-fold symmetry axes and translating (*r*) them into contact along the aligned axes with respect to the center of mass of each trimer. Configurations in which backbone atoms from different building blocks were less than 3.5 Å apart were discarded. Non-clashing design models were ranked based on the number of pairs of b-carbons in adjacent subunits within 12 Å (Fig. 1b, c). Further higher resolution sampling was carried out around the top 208 docked configurations on a 0.5° by 0.2 Å Gaussian grid centered at 0° and 0Å.

2.3 Interface Design

Symmetric Rosetta Design[26,27] calculations were then used to generate low-energy, symmetric hydrophobic interfaces, and the resulting designs were filtered based on shape complementarity[28] (sc), interface surface area (sasa), buried unsatisfied hydrogen bonds (uhb), and binding energy (ddg). Designed substitutions that did not substantially contribute to the interface were reverted to their original identities. A total of 17 designs were selected for experimental characterization.

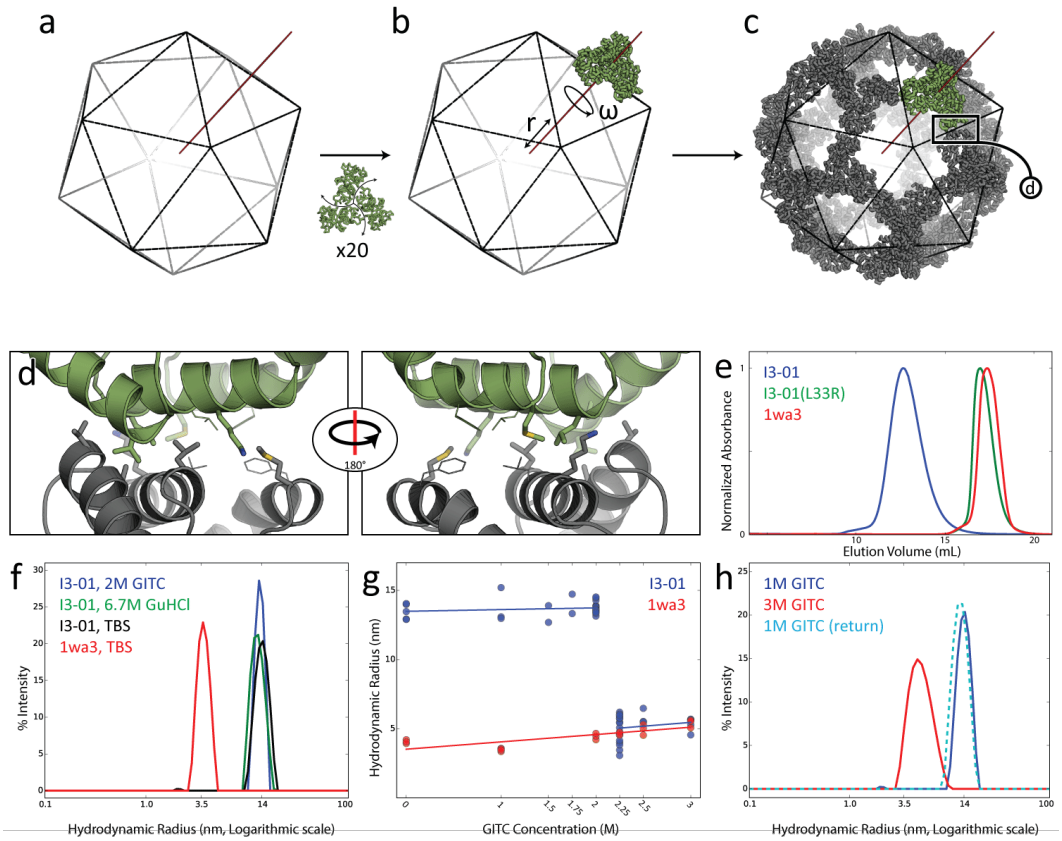

Figure 1: Design methodology and biochemical characterization.

a) Icosahedral symmetry denoted in black while the three-fold axis denoted in red. **b)** The aligned trimeric building block shown in green. A total of 20 copies of the trimer are modelled during docking. **c**) Optimization of *r* and *w* yields closely opposed interfaces between subunits. **d**) Sequence design yields low-energy interfaces. I3-01's interface is composed of five designed residues (thick representations) and two native residues (thin representations). **e**) I3-01 appears larger by SEC than the similarly sized I3-01(L33R) and wild-type trimer (1wa3). **f**) DLS measurement of hydrodynamic radius (note logarithmic scale in **f** and **h**) of 1wa3 (3.5 nm) and I3-01 (14 nm). I3-01 remains assembled in 6.7 M GuHCl and in 2 M GITC. **g**) Extremely sharp disassociation to trimeric building blocks at 2.25 M GITC. Data points represent independent measurements. **h**) I3-01 disassembles into the trimeric building blocks at 3 M GITC and reassembles following dilution to 1 M.

Chapter 3. EXPERIMENTAL CHARACTERIZATION

3.1 CLONING AND PROTEIN EXPRESSION

Codon-optimized genes encoding the wild-type and the designed molecules were generated by recursive polymerase chain reaction (PCR) from sets of synthetic oligonucleotides (Integrated DNA Technologies). After assembly, the genes were cloned into the pET29b+ vector for expression in *E. coli* BL21 DE* cells (Invitrogen). A list of designs mentioned in this document and their respective amino acid sequences can be found in appendix A. After induction with isopropyl β-D-1-thiogalactopyranoside (IPTG) for 4 h at 37 °C, cells were harvested by centrifugation. Cell lysis was accomplished in Tris-buffered saline (TBS; 50 mM Tris, 500 mM NaCl) with lysozyme (0.25 mg ml^{-1}) and sonication (Fisher Scientific) at 20 W for 5 min total 'on' time, using cycles of 10 s on, 10 s off.

For initial screening, all constructs were labelled with the CoA-488 fluorophore (NEB) by the addition of AcpS[29] (NEB) using an A1 peptide tag, allowing the solubility and assembly state of each design to be analysed using SDS-PAGE and native-PAGE (Bio-Rad), following procedures previously described[24]. All subsequent experiments were performed on either $(His)_6$-tagged proteins or proteins that remained untagged.

Most of the designs were found in the insoluble fraction upon cell lysis; of the three soluble designs, two (both based on a KDPG aldolase[30,31]) showed substantial shifts in migration relative to the wild-type scaffold when analysed by native (non-denaturing) polyacrylamide gel electrophoresis (PAGE), suggesting higher-order assembly. We selected the one with fewer mutations, I3-01, for further analysis. Five substitutions (E26K, E33L, K61M, D187V, and

R190A) were made to generate the designed interface between trimers (Fig. 1d; amino acid sequences provided in Supplementary Text).

Larger batches of I3-01 were prepared for further experiments. After lysis and centrifugation at 20,000 *g* for 30 min, the soluble fraction of $(His)_6$-tagged proteins were passed through 2 ml of nickel nitrilotriacetic acid agarose (Ni-NTA) **(**Qiagen), washed with 30 mM imidazole, and eluted with 500 mM imidazole.

For non-$(His)_6$-tagged proteins, cells were lysed as above, and the cleared lysates were treated with serial ammonium sulfate precipitation treatments (20%, 60% w/v). During each step, solid ammonium sulfate was added to the lysate to the desired percentage, and equilibrated at room temperature for 1 h. Ammonium sulfate precipitated protein was then collected by centrifugation at 20,000 *g* for 30 min at 25 °C. After treatment at 60%, the pellet was then solubilized in TBS and heated at 80 °C for 10 min. The soluble fraction was then collected and further purified through size exclusion chromatography (SEC) as described.

Proteins were then passed through SEC using a Superose 6 10/300 GL SEC column (GE Healthcare), yielding a single peak with an apparent molecular weight much larger than that of the wild-type trimeric protein and consistent with the expected elution volume for the 60-subunit assembly (Fig. 1e). A mutant bearing a leucine-to-arginine substitution predicted to disrupt the designed interface eliminated the high-molecular-weight species and returned the elution volume to that of the wild-type scaffold (Fig. 1e).

3.2 Dynamic Light Scattering

To probe assembly kinetics, we performed dynamic light scattering (DLS) on I3-01. Purified protein was measured using a DynaPro NanoStar (Wyatt) DLS setup. 0.5 mg ml^{-1} of I3-

01 and 1wa3 were measured at 25 °C. Measurements of I3-01 showed a monodisperse population of particles with a hydrodynamic radius of 14 nm, consistent with the design model (Fig. 1f).

The temperature was then ramped up to 90 °C, then ramped back down to 25 °C for temperature scans at 2 °C min^{-1}. In addition, chemical denaturation was attempted with guanidine hydrochloride (GuHCl) and guanidinium thiocyanate (GITC). Measurements were taken in the presence of TBS (25 mM Tris, 500 mM NaCl), buffered GuHCl (25 mM Tris, 500 mM NaCl, 1–6.7 M GuHCl), or buffered GITC (25 mM Tris, 500 mM NaCl, 1–4 M GITC). Different concentrations of GITC equilibrated samples were achieved by combining stocks of 0 M and 4 M equilibrated solutions in different ratios while GuHCl equilibrated samples were equilibrated individually. Re-annealing experiments were performed by diluting I3-01 equilibrated in 3 M GITC down to 1 M GITC final concentration (0.166 mg ml^{-1} protein). Each sample was allowed to equilibrate in their respective buffer for at least 24 h before measurement. Data analysis was performed using DYNAMICS v7 (Wyatt), reporting regularization fits (with D10/D50/D90) except for temperature ramp experiments, where cumulant fits were used. The ~1 nm radius particle consistent with GITC buffer alone was disregarded for analysis, and monodispersity was assumed when less than 15% peak polydispersity was observed[32,33].
No disassembly to the trimeric building block was observed at 80 °C or, remarkably, in 6.7 M (GuHCl) (Fig. 2). This hyperstability is a property of both the trimeric scaffold from which I3-01 was derived and of the designed interface; both are completely resistant to GuHCl denaturation. An exceptionally sharp disassociation into the constituent trimers was observed between 2 M and 2.25 M GITC: at 2 M the dominant species is the nanocage, while at 2.25 M only the trimeric building block is observed (Fig. 1g). Notably, for cargo packaging applications, the

disassociation is fully reversible; the hydrodynamic radius of particles formed by diluting disassembled protein in 3 M GITC down to 1 M GITC is identical to those originally produced in *E. coli* (Fig. 1h).

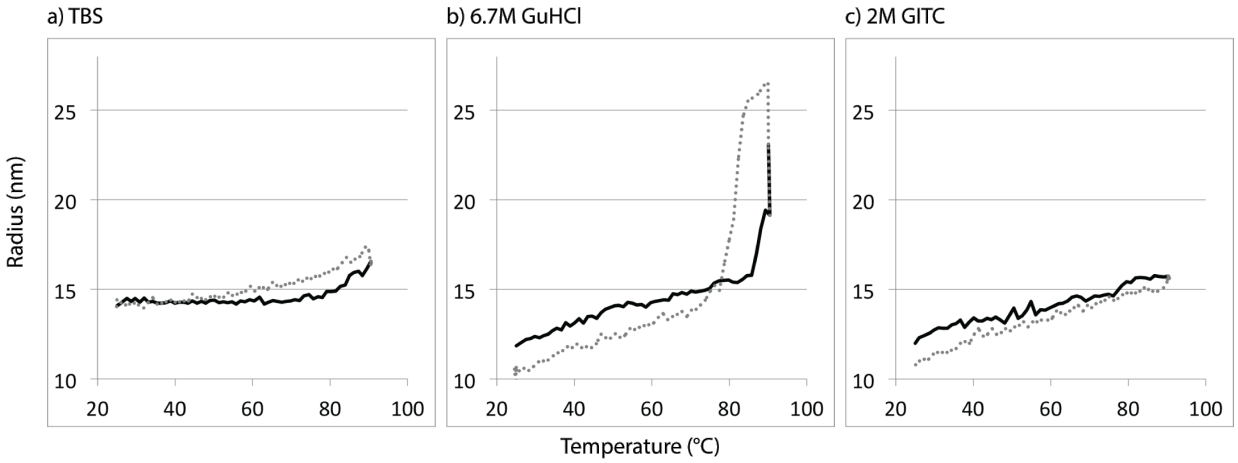

Figure 2: I3-01 tolerance to temperature.

DLS measurements as I3-01 is subjected to heating to 90 °C (solid line), then cooling to 25 °C (dotted line) in TBS (**a**), 6.7 M GuHCl (**b**), and 2 M GITC (**c**). Under all three conditions, any indications of aggregation or increase in size due to temperature appear to be completely reversible.

3.3 Native Enzymatic Activity

As the native scaffold of I3-01 is a KDPG aldolase[30,31], we investigated if I3-01 retained its native enzymatic activity. An additional mutant, I3-01(K129A), was designed where the catalytic lysine in the active site was mutated to an alanine. The reactions were carried out in 25 mM HEPES, 20 mM NaCl buffer at pH 7.0 with the presence of NADH (0.1 mM), L-lactate dehydrogenase (LDH, 0.11 U μl^{-1}), and 2-keto-3-deoxy-6-phosphogluconate (KDPG, 1 mM) at

25 °C, based on previously described methods[30]. Native 1wa3, I3-01, or I3-01(K129A) was added at 0.02 μM final concentration to each well and immediately monitored for 339 nm ultraviolet absorbance over time. It is evident that I3-01 still retains the native enzymatic activity, which is not surprising as the active site was not mutated in the design process (Fig. 3). This shows as a proof of concept that nanocages have the potential to be designed as nano-reactors for synthetic biology.

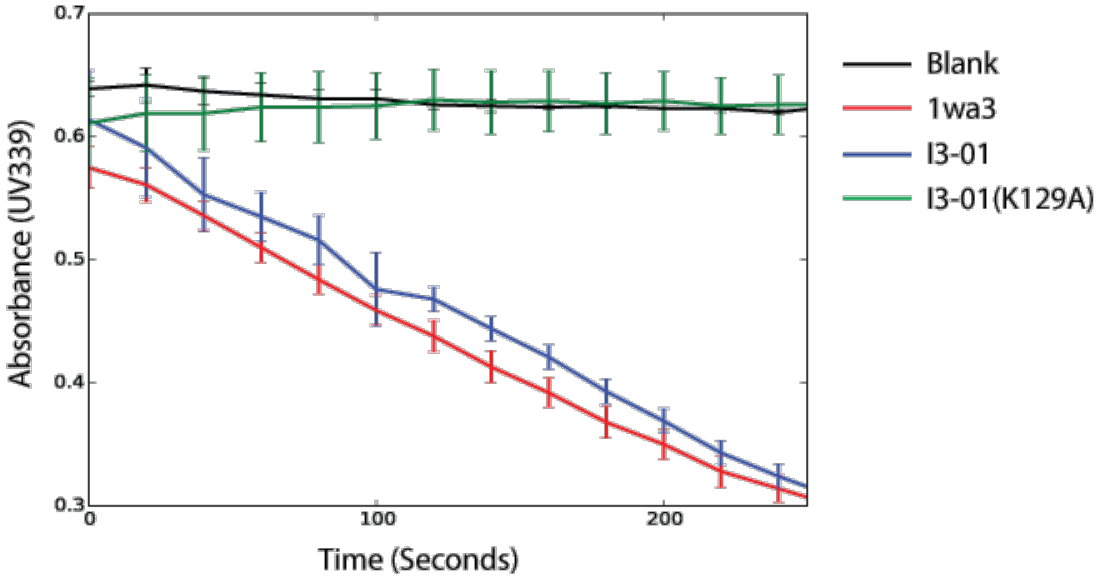

Figure 3: I3-01 retains native enzyme activity.

Coupled KDPG aldolase assay showing native-like enzymatic activity in I3-01. The K129A knockout shows no enzyme activity, similar to buffer alone. UV339, absorbance at 339 nm; error bars are standard deviation.

3.4 Electron Microscopy

We investigated the structure of I3-01 using cryo-electron microscopy (cryo-EM). 5 ml of purified untagged I3-01 and I3-01(ctGFP), diluted to ~0.1 mg ml^{-1} using TBS buffer (25 mM Tris pH 8.0, 150 mM NaCl) with an additional 2 mM dithiothreitol were applied to glow discharged 1.2/1.3 Quantifoil grids, blotted, and plunged into liquid ethane using a Vitrobot (FEI). Screening and grid optimization was performed on a 200 kV TF20 transmission electron microscope (FEI) with a bottom-mount TVIPS F416 CMOS 4k camera. 4-6 s movies were recorded on a 300 kV Titan Krios (FEI) using a Gatan K2 direct detector at either 29,000× or 37,000× magnification at the specimen level at ~10 electrons per pixel per second.

Movies were motion-corrected using previously described methods[34]. Coordinates for 6,461 (I3-01) and 13,297 (I3-01(ctGFP)) unique particles were obtained for averaging using EMAN2[35]. Extracted frames of these particles were used to calculate class averages by refinement in IMAGIC[36] using multiple rounds of multivariate statistical analysis and multi-reference alignment. An initial density model was calculated based on the calculated averages using EMAN2[35] and the fitting of the model and correlation were calculated using UCSF Chimera[37]. Low-resolution (17–30 Å) volumes from the I3-01 design model were calculated using SPIDER[38] and inspected in UCSF Chimera[37]. Back-projection images were computed in SPIDER[38] on the low-resolution volumes and visualized using WEB[38]. The contrast of all micrographs was enhanced in Fiji[39].

The individual particles in large fields of view are homogenous in size and shape (Fig. 4a), and in class averages, the three projections along the symmetry axes and the overall icosahedral architecture are clearly discernible (Fig. 4b, c). A three-dimensional model

calculated from the cryo-EM data matches the I3-01 design model very well with a correlation coefficient of 0.92 at 20 Å and 1.5 s (Fig. 4d, e), clearly indicating that I3-01 forms the designed structure: an icosahedron with a diameter of 25 nm and an interior volume of approximately 3,000 nm^3, values that are within the range of those observed in small viral capsids[40].

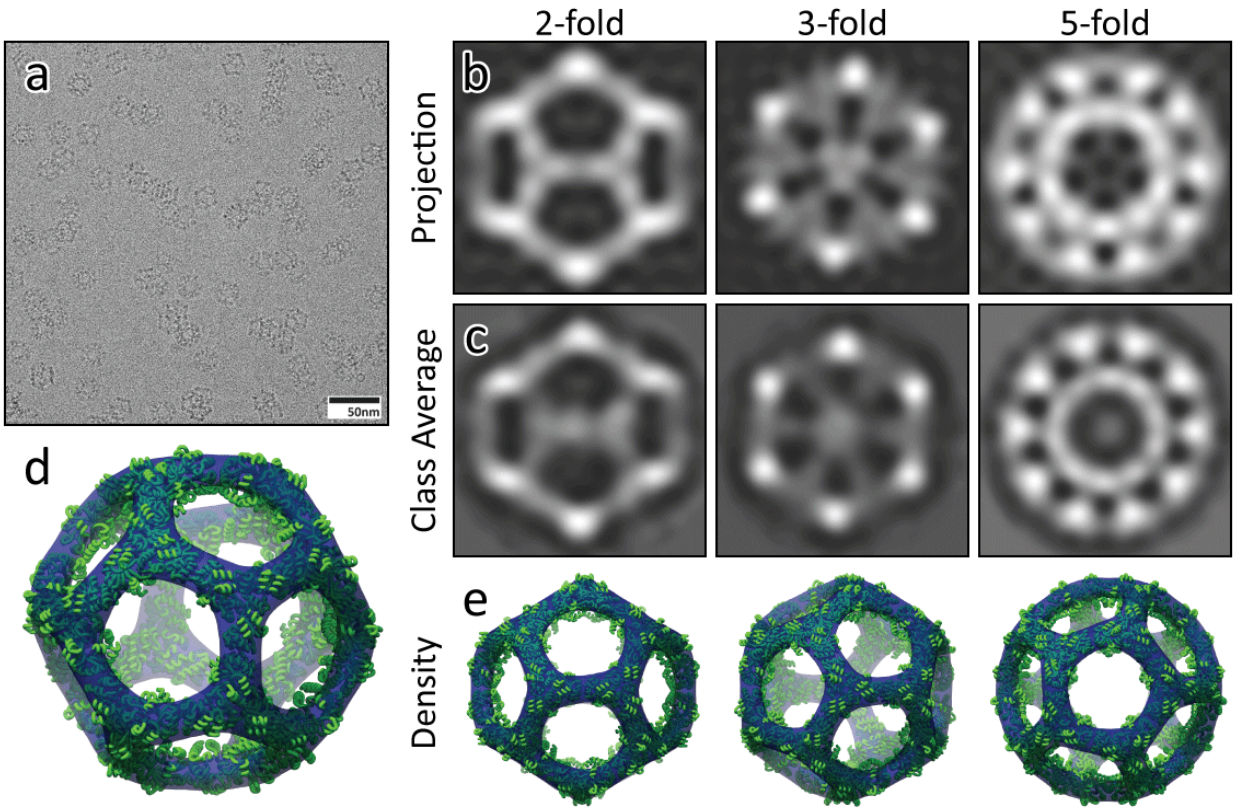

Figure 4: Cryo-EM.

a) Field-of-view cryo-EM micrograph showing homogeneous icosahedral particles in various orientations. **b**) Back-projections of I3-01 from the design model. **c**) Cryo-EM class averages closely match the design projections along all three symmetry axes. **d**, **e**) The calculated initial, unrefined density (blue, 3.22 *s*) closely matches the design model (green).

Chapter 4. GENETIC FUSIONS

4.1 GFP FUSIONS

To probe the robustness of I3-01 to genetic fusion, we fused superfolder GFP (sfGFP)[41] to one or both termini of the monomeric subunit and produced the resulting proteins in *E. coli*. SEC analysis showed that the fusion proteins had hydrodynamic radii consistent with cage formation (Fig. 5a). Analysis of I3-01 with a carboxy (C)-terminal sfGFP fusion (called I3-01(ctGFP)) by cryo-EM revealed icosahedral particles with overall shapes very similar to those of the original design. Class averages of 13,297 particles revealed considerable internal density compared to the original I3-01 averages, consistent with computational models of the fusion complex (Fig. 6a). Rosetta Remodel[41] was used to model I3-01(ctGFP). I3-01 was held static while the linker was sampled via fragment insertion, placing the sfGFP molecules at the end of the linker. The overall model was sampled symmetrically with icosahedral symmetry. The I3-01 sfGFP fusions are robust to denaturation of the amino (N)- or C-terminal fused sfGFP in GuHCl; the particles remain assembled as GFP signal is lost[42] (Fig. 5b).

It is presently challenging to infer subunit copy number in GFP-tagged assemblies from their fluorescence intensity. What is needed are ‘standard candles’ with known fluorescent protein copy numbers that can be used to correlate fluorescence intensity to copy number. To complement the icosahedra with 60 and 120 copies of sfGFP described above, we fused sfGFP to one or both components of a previously described two-component tetrahedron[25] (T33-21) to generate assemblies with 12 or 24 copies of sfGFP (Fig. 5a). Different constructs used for fluorescence microscopy were generated by genetically fusing sfGFP to the termini of

nanocages. For T33-21, the sfGFP was fused to either the C terminus of the first component (12 sfGFP molecules), or the C terminus of both components (24 sfGFP molecules). For I3-01, the sfGFP was fused to either terminus of I3-01 (60 sfGFP molecules), or both termini of I3-01 **(**120 sfGFP molecules). For mTurquoise2 and SYFP2 versions, sfGFP was replaced with the sequence of the respective fluorescent protein bearing additional surface mutations identical to sfGFP[41]. GFP nanocages were mounted on agarose pads for microscopy as previously described[43]. Images of the GFP nanocages were obtained using a DeltaVision system (Applied Precision) with an IX70 inverted microscope (Olympus), a U Plan Apo 100× objective (1.35 NA), and a CoolSnap HQ digital camera (Photometrics). GFP images were taken with a 0.4 s exposure, in a single focal plane, and binned 1×1. The fluorescence intensities of GFP puncta were identified and quantified using custom Matlab programs as previously described[44]. Fluorescent intensity histograms of individual sfGFP-fused cages were fitted with Gaussian distributions, shown with mean total arbitrary fluorescence unit (AFU) intensity ± one standard deviation.

Intensity histograms obtained for each of the sfGFP-nanocage constructs using widefield fluorescence microscopy were well fitted with Gaussians (Fig. 6b, c), and the mean fluorescence intensity for each cage was found to be linearly proportional ($r^2 = 0.9925$) to sfGFP copy number (Fig. 6d). The fluorescent properties of the particles were readily manipulated by substituting sfGFP with mTurqoise2 and sYFP2 (Fig. 7). In addition to serving as genetically encoded, water-soluble fluorescent standard candles, the fluorescent protein cage fusions could be useful for correlative light and electron microscopy[45] since the icosahedral shape is quite distinctive.

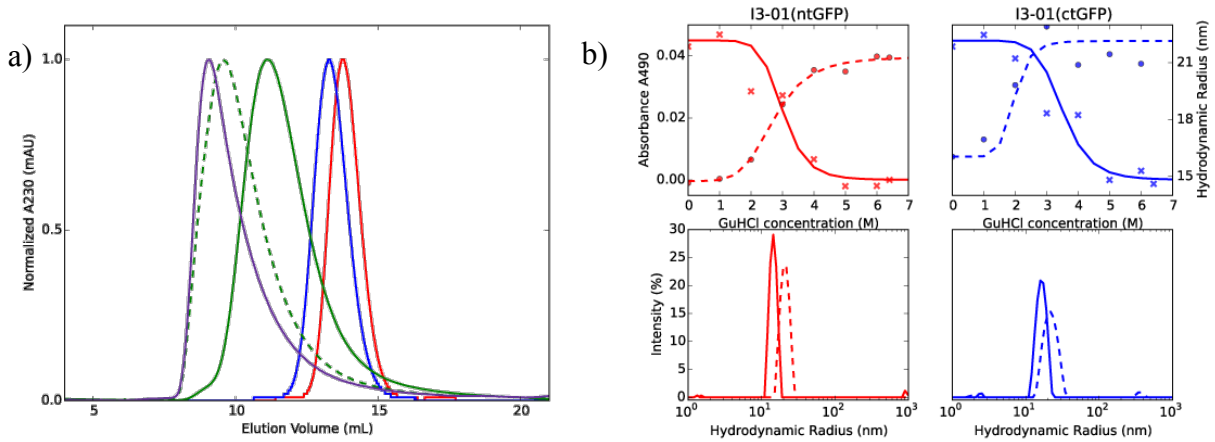

Figure 5: Nanocages fused with sfGFP.

a) Size exclusion chromatography traces, for T33-21 (12mer in red and 24mer in blue) and I3-01 (60mer in green and 120mer in purple) sfGFP fusions, display increased particle sizes with increasing copies of GFP, but retain monodispersed populations. The N-terminal fusion of sfGFP (dashed line) is expected to extend mostly outward from the icosahedron, thus greatly increasing the hydrodynamic radius while the C-terminal fusion is predicted to occupy the internal void space. A230, ultraviolet absorbance at 230 nm; mAU, milli-absorbance units. **b)** Tolerance of I3-01–sfGFP fusions to GuHCl. N-terminal (red) and C-terminal (blue) sfGFP fusions were equilibrated to 0–6.4 M GuHCl. Ultraviolet absorbance at 490 nm (A490) monitors the unfolding of sfGFP (top, solid line and crosses). DLS experiments (top, dotted line and dots) reveal as sfGFP unfolds, the hydrodynamic radius increases slightly, and then stabilizes. The bottom panels show that in 1 M GuHCl (solid line) and in 6 M GuHCl (dotted line), the icosahedral assemblies remain relatively monodisperse.

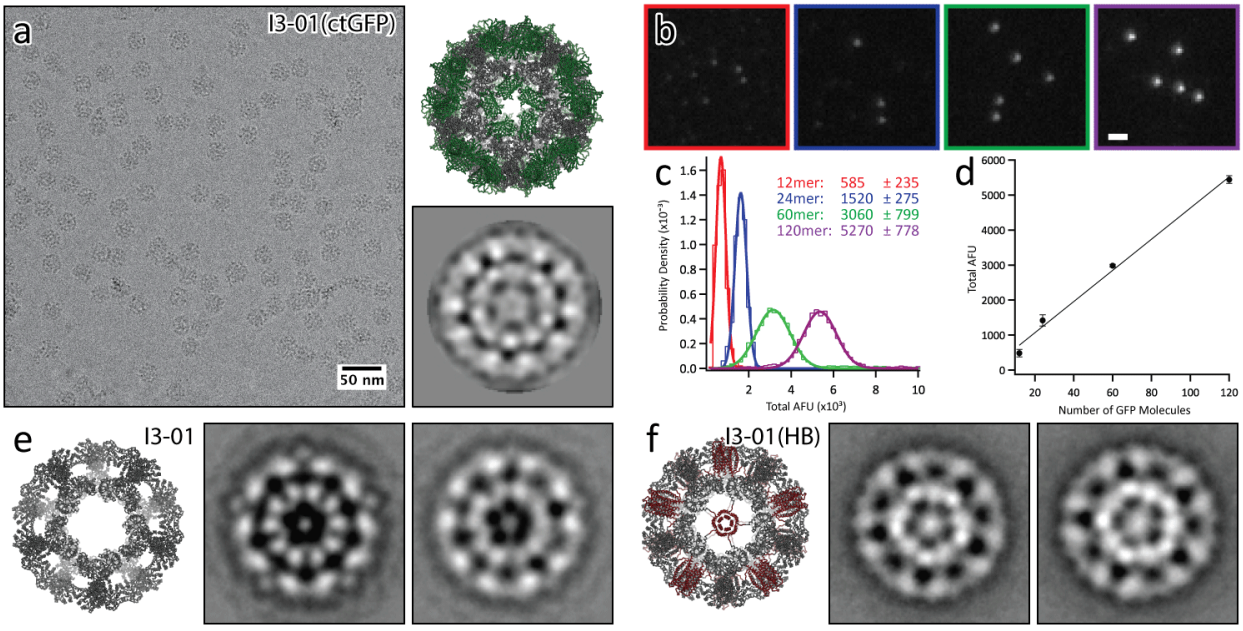

Figure 6: Tuning nanocage structure and function with genetic fusions.

a) The left panel shows a cryo-EM micrograph of I3-01(ctGFP); the top right panel shows the computational model with sfGFP in green; the bottom right panel shows the class average along the fivefold axis. **b**) Fluorescence microscopy fields of view. **c**) Fluorescence intensity histograms. AFU, arbitrary fluorescence units; standard deviation of Gaussian fit reported (n=3). **d**) Correlation between the mean fluorescence intensity and sfGFP copy number for nanoparticles with different numbers of fused sfGFP molecules. Error bars are s.e.m. **e**, **f**) Computational model and class averages along the five-fold axis of negatively stained I3-01 (**e**) and I3-01(HB) (**f**); the helical bundle is shown in red. Weak density in the centre of the pentameric faces in I3-01 may reflect randomly packaged material. There is clear density in the centre of the pentameric faces in the I3-01(HB) class averages consistent with the model.

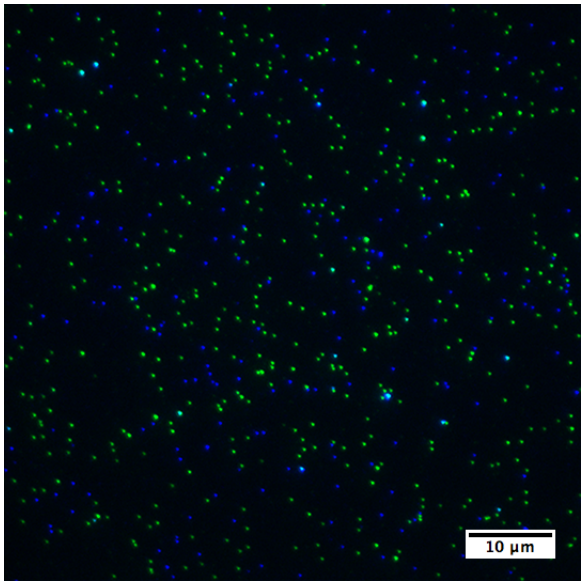

Figure 7: I3-01 C-terminal fusions with other fluorescent proteins.

Fluorescent proteins mTurquoise2 (blue) or sYFP2 (green) were fused to the C-terminus of I3-01. The field of view using widefield fluorescence microscopy shows distinct signals of each type when the two types are mixed together.

4.2 Helical Bundle Fusions

In I3-01, the trimeric building blocks are aligned with the three-fold axes while the designed interface is along the icosahedral two-folds. To explore the possibility of symmetry-matched fusions to designed nanocages, we modelled a designed pentameric helical bundle[46] into the center of the large 9-nm pore at the fivefold axis with a C-terminal linker; this fusion was named I3-01(HB).

Rosetta Remodel[47] was used to generate linkers for I3-01(HB). I3-01 was held static while linkers of different lengths (7–12 residues) were sampled via fragment insertion. The

resulting placement of the helical bundle at the end of the linker was filtered with pentameric assembly constraints to determine linker lengths that could satisfy formation of the pentameric helical bundle. The shorter linkers that allowed unstrained helical assembly were selected for experimental testing.

For negative-stain electron microscopy, 6 ml of purified I3-01 and I3-01(HB) at 0.05–0.1 mg ml^{-1} were applied to glow discharged, carbon-coated 400-mesh copper grids (Ted Pella), washed with Milli-Q water, and stained with 0.75% uranyl formate. Grids were visualized for assembly validation and optimized for data collection. Screening and sample optimization was performed on a 100 kV Morgagni M268 transmission electron microscope (FEI) equipped with an Orius charge-coupled device (CCD) camera (Gatan). Data collection was performed on a 120 kV Tecnai G2 Spirit transmission electron microscope (FEI). All final images were recorded using an Ultrascan 4000 4k × 4k CCD camera (Gatan) at 52,000× magnification at the specimen level. Coordinates for 6,576 I3-01 and 4,131 I3-01(HB) unique particles were obtained for averaging using EMAN2[35]. Boxed particles were used to obtain two-dimensional class averages by refinement in EMAN2. Additional image analysis was performed using ImageJ[48].

Negative-stain electron microscopy showed monodisperse particles of the expected size and symmetry; the incorporation of the pentamer does not interfere with icosahedral assembly. Particle averages showed a structure similar to that of the original icosahedron, with additional density at the centre of each fivefold axis, consistent with computational models of the fusion protein (Fig. 6e, f). The capability of incorporating symmetry-matched substructures into designed nanocages offers considerable flexibility and modularity; for example, pentamers filling otherwise open pentameric faces could control the release of cargo contained within the nanocage.

Chapter 5. CONCLUDING REMARKS

5.1 ICOSAHEDRA VS DODECAHEDRA

Shortly after publication of the manuscript, a reader suggested that the nanocage we have designed was in fact a "dodecahedron" instead of an "icosahedron" as we called it. However, we do not agree that it is an error; it is dependent on the viewer's interpretation of the assembly, as icosahedrons and dodecahedrons are duals and they share the same underlying icosahedral symmetry. To be truly technical, referring to a given protein assembly using the name of a regular polyhedron will always be an approximation, as Platonic solids are defined as convex polyhedra containing congruent, regular, polygonal faces. In reality, I3-01 is neither an icosahedron nor a dodecahedron. As I3-01 was constructed from trimers, it was classified as I3, "icosahedral symmetry from trimers". If the trimeric building blocks are conceptualized as the faces of an icosahedron, which we believe makes intuitive sense given that the icosahedron is constructed from triangular faces, then referring to this class of structures as icosahedra conveys useful information about the hierarchy of interactions in the assemblies and the way in which they are constructed. It just so happens that I3-01 also resembles the wireframe of a dodecahedron, which could also be correct if the trimers are placed at the vertices of the icosahedral symmetry. If a series of docked configurations are analyzed side by side, it is clear that some configurations by eye resemble more of an icosahedron while others resemble dodecahedrons (Fig. 8). However, it is important to remember that they all share the same underlying icosahedral symmetry.

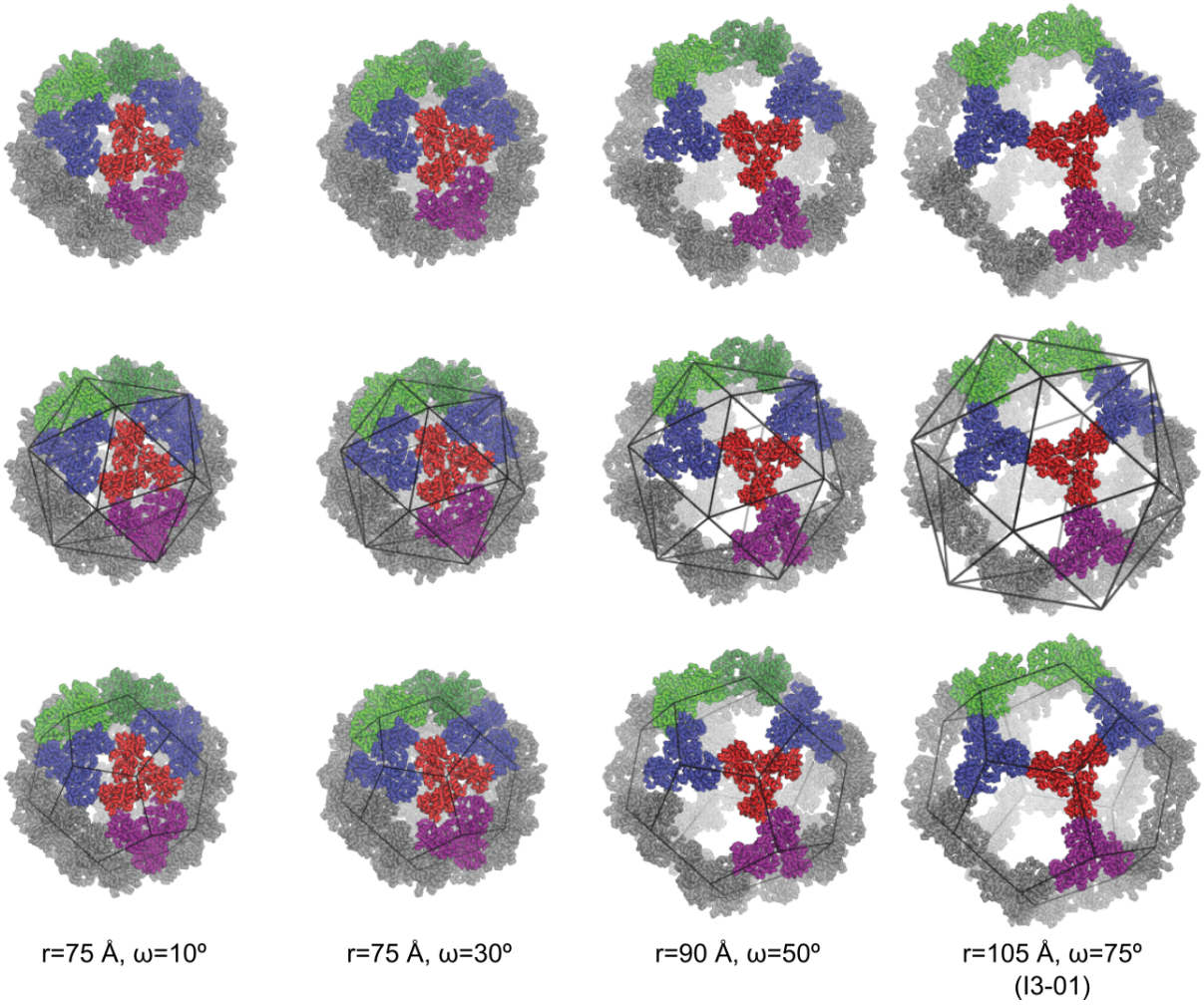

Figure 8: Fitting icosahedra and dodecahedra shapes into docked configurations.

Top) Docked configurations of 1wa3, the scaffold that was used to design I3-01. *r* and *w* denote the displacement and angle used to generate each of the docks. **Middle)** A canonical icosahedra fit into each of the docks, aligned to their respective symmetrical axes. **Bottom)** A canonical dodecahedra fit into each of the docks, aligned to their respective symmetrical axes.

5.2 CONCLUSIONS

The designed I3-01 icosahedron is exceptionally stable, robust to genetic fusion, and has a considerably larger internal volume than previously designed nanocages with well-defined and prespecified structures[20,23,25]. Since publication of the manuscript, additional work has been done to show that it is possible to design larger, 2-component icosahedral nanocages[49]. It is evident that the ability to accurately design even larger icosahedral protein structures opens the door to new approaches to many application fields such as vaccine generation and targeted drug delivery.

BIBLIOGRAPHY

1. Zandi, R., Reguera, D., Bruinsma, R. F., Gelbart, W. M. & Rudnick, J. From The Cover: Origin of icosahedral symmetry in viruses. *Proc. Natl. Acad. Sci.* **101,** 15556–15560 (2004).
2. Ritsert, K. *et al.* Studies on the lumazine synthase/riboflavin synthase complex of Bacillus subtilis: crystal structure analysis of reconstituted, icosahedral beta-subunit capsids with bound substrate analogue inhibitor at 2.4 A resolution. *J. Mol. Biol.* **253,** 151–167 (1995).
3. Howorka, S. Rationally engineering natural protein assemblies in nanobiotechnology. *Curr. Opin. Biotechnol.* **22,** 485–491 (2011).
4. Roldão, A., Mellado, M. C. M., Castilho, L. R., Carrondo, M. J. T. & Alves, P. M. Virus-like particles in vaccine development. *Expert Rev. Vaccines* **9,** 1149–1176 (2010).
5. Effio, C. L. & Hubbuch, J. Next generation vaccines and vectors: Designing downstream processes for recombinant protein-based virus-like particles. *Biotechnol. J.* **10,** 715–727 (2015).
6. Ma, Y., Nolte, R. J. M. & Cornelissen, J. J. L. M. Virus-based nanocarriers for drug delivery. *Adv. Drug Deliv. Rev.* **64,** 811–825 (2012).
7. Smith, M. L. *et al.* Modified Tobacco mosaic virus particles as scaffolds for display of protein antigens for vaccine applications. *Virology* **348,** 475–488 (2006).
8. Bauler, P., Huber, G., Leyh, T. & McCammon, J. A. Channeling by Proximity: The Catalytic Advantages of Active Site Colocalization Using Brownian Dynamics. *J. Phys. Chem. Lett.* **1,** 1332–1335 (2010).
9. Levy, E. D., Erba, E. B., Robinson, C. V. & Teichmann, S. A. Assembly reflects evolution of protein complexes. *Nature* **453,** 1262–1265 (2008).

10. Clarke, O. B. & Gulbis, J. M. Oligomerization at the Membrane. in *Protein Dimerization and Oligomerization in Biology* (ed. Matthews, J. M.) **747,** 122–136 (Springer New York, 2012).

11. Koshland, D. E. & Hamadani, K. Proteomics and Models for Enzyme Cooperativity. *J. Biol. Chem.* **277,** 46841–46844 (2002).

12. Jutz, G., van Rijn, P., Santos Miranda, B. & Böker, A. Ferritin: A Versatile Building Block for Bionanotechnology. *Chem. Rev.* **115,** 1653–1701 (2015).

13. Royle, S. J. The cellular functions of clathrin. *Cell. Mol. Life Sci.* **63,** 1823–1832 (2006).

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

15. 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).

16. 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).

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

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

19. Raman, S., Machaidze, G., Lustig, A., Aebi, U. & Burkhard, P. Structure-based design of peptides that self-assemble into regular polyhedral nanoparticles. *Nanomedicine Nanotechnol. Biol. Med.* **2,** 95–102 (2006).

20. Raman, S. *et al.* Design of Peptide Nanoparticles Using Simple Protein Oligomerization Domains. *Open Nanomedicine J.* **2,** 15–26 (2009).

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

22. 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).

23. Lai, Y.-T. *et al.* Structure of a designed protein cage that self-assembles into a highly porous cube. *Nat. Chem.* **6,** 1065–1071 (2014).

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

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

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

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

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

29. Zhou, Z. *et al.* Genetically encoded short peptide tags for orthogonal protein labeling by Sfp and AcpS phosphopantetheinyl transferases. *ACS Chem. Biol.* **2,** 337–346 (2007).

30. Griffiths, J. S. *et al.* Cloning, isolation and characterization of the Thermotoga maritima KDPG aldolase. *Bioorg. Med. Chem.* **10,** 545–550 (2002).

31. Fullerton, S. W. B. *et al.* Mechanism of the Class I KDPG aldolase. *Bioorg. Med. Chem.* **14,** 3002–3010 (2006).

32. Baalousha, M. & Lead, J. R. Nanoparticle dispersity in toxicology. *Nat. Nanotechnol.* **8,** 308–309 (2013).

33. Zulauf, M. & D'Arcy, A. Light scattering of proteins as a criterion for crystallization. *J. Cryst. Growth* **122,** 102–106 (1992).

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

35. Tang, G. *et al.* EMAN2: an extensible image processing suite for electron microscopy. *J. Struct. Biol.* **157,** 38–46 (2007).

36. van Heel, M., Harauz, G., Orlova, E. V., Schmidt, R. & Schatz, M. A New Generation of the IMAGIC Image Processing System. *J. Struct. Biol.* **116,** 17–24 (1996).

37. Pettersen, E. F. *et al.* UCSF Chimera--a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25,** 1605–1612 (2004).

38. Frank, J. *et al.* SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields. *J. Struct. Biol.* **116,** 190–199 (1996).

39. Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9,** 676–682 (2012).

40. Perlmutter, J. D. & Hagan, M. F. Mechanisms of Virus Assembly. *Annu. Rev. Phys. Chem.* **66,** 217–239 (2015).

41. Pédelacq, J.-D., Cabantous, S., Tran, T., Terwilliger, T. C. & Waldo, G. S. Engineering and characterization of a superfolder green fluorescent protein. *Nat. Biotechnol.* **24,** 79–88 (2006).

42. Andrews, B. T., Schoenfish, A. R., Roy, M., Waldo, G. & Jennings, P. A. The Rough Energy Landscape of Superfolder GFP Is Linked to the Chromophore. *J. Mol. Biol.* **373,** 476–490 (2007).

43. Muller, E. G. D. *et al.* The organization of the core proteins of the yeast spindle pole body. *Mol. Biol. Cell* **16,** 3341–3352 (2005).

44. Shimogawa, M. M., Wargacki, M. M., Muller, E. G. & Davis, T. N. Laterally attached kinetochores recruit the checkpoint protein Bub1, but satisfy the spindle checkpoint. *Cell Cycle* **9,** 3619–3628 (2010).

45. Cortese, K., Diaspro, A. & Tacchetti, C. Advanced Correlative Light/Electron Microscopy: Current Methods and New Developments Using Tokuyasu Cryosections. *J. Histochem. Cytochem.* **57,** 1103–1112 (2009).

46. Huang, P.-S. *et al.* High thermodynamic stability of parametrically designed helical bundles. *Science* **346,** 481–485 (2014).

47. Huang, P.-S. *et al.* RosettaRemodel: A Generalized Framework for Flexible Backbone Protein Design. *PLoS ONE* **6,** e24109 (2011).

48. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **9,** 671–675 (2012).

49. Bale, J. B. *et al.* Accurate design of megadalton-scale two-component icosahedral protein complexes. *Science* **353,** 389–394 (2016).

APPENDIX A

Protein Sequences

Tags and linkers underlined, mutations **bolded**, and fusions in respective colors.

>>1wa3

MHHHHHHGGMKMEELFKKHKIVAVLRANSVEEAKEKALAVFEGGVHLIEITFTVPDADTVIKELSFLKEKGAIIGAGTVTSVEQCR
KAVESGAEFIVSPHLDEEISQFCKEKGVFYMPGVMTPTELVKAMKLGHTILKLFPGEVVGPQFVKAMKGPFPNVKFVPTGGVNLDN
VCEWFKAGVLAVGVGSALVKGTPDEVREKAKAFVEKIRGCTE

>>I3-01

MHHHHHHGGSGGSGGSGGSMKMEELFKKHKIVAVLRANSVEEAK**K**KALAVF**L**GGVHLIEITFTVPDADTVIKELSFLKE**M**GAIIGA
GTVTSVEQCRKAVESGAEFIVSPHLDEEISQFCKEKGVFYMPGVMTPTELVKAMKLGHTILKLFPGEVVGPQFVKAMKGPFPNVKF
VPTGGVNLDNVCEWFKAGVLAVGVGSALVKGTP**V**EV**A**EKAKAFVEKIRGCTE

>>I3-01(L33R)

MHHHHHHGGMKMEELFKKHKIVAVLRANSVEEAKKKALAVF**R**GGVHLIEITFTVPDADTVIKELSFLKEMGAIIGAGTVTSVEQCR
KAVESGAEFIVSPHLDEEISQFCKEKGVFYMPGVMTPTELVKAMKLGHTILKLFPGEVVGPQFVKAMKGPFPNVKFVPTGGVNLDN
VCEWFKAGVLAVGVGSALVKGTPVEVAEKAKAFVEKIRGCTE

>>I3-01(K129A)

MHHHHHHGGSGGSGGSGGSMKMEELFKKHKIVAVLRANSVEEAKKKALAVFLGGVHLIEITFTVPDADTVIKELSFLKEMGAIIGA
GTVTSVEQCRKAVESGAEFIVSPHLDEEISQFCKEKGVFYMPGVMTPTELVKAMKLGHTIL**A**LFPGEVVGPQFVKAMKGPFPNVKF
VPTGGVNLDNVCEWFKAGVLAVGVGSALVKGTPVEVAEKAKAFVEKIRGCTE

>>T33-21(12-mer)

A:MRITTKVGDKGSTRLFGGEEVWKDSPIIEANGTLDELTSFIGEAKHYVDEEMKGILEEIQNDIYKIMGEIGSKGKIEGISEERI
AWLLKLILRYMEMVNLKSFVLPGGTLESAKLDVCRTIARRALRKVLTVTREFGIGAEAAAYLLALSDLLFLLARVIEIEKNKLKEV
RSGSGSGSSKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLTLKFICTTGKLPVPWPTLVTTLTYGVQCFARYPDHMK
QHDFFKSAMPEGYVQERTISFKDDGTYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNFNSHNVYITADKQKNGIKANF
KIRHNVEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSVLSKDPNEKRDHMVLLEFVTAAGITHGMDELY
B:MPHLVIEATANLRLETSPGELLEQANKALFASGQFGEADIKSRFVTLEAYRQGTAAVERAYLHACLSILDGRDIATRTLLGASL
CAVLAEAVAGGGEEGVQVSVEVREMERLSYAKRVVARQRLEHHHHHH

>>T33-21(24-mer)

A:MRITTKVGDKGSTRLFGGEEVWKDSPIIEANGTLDELTSFIGEAKHYVDEEMKGILEEIQNDIYKIMGEIGSKGKIEGISEERI
AWLLKLILRYMEMVNLKSFVLPGGTLESAKLDVCRTIARRALRKVLTVTREFGIGAEAAAYLLALSDLLFLLARVIEIEKNKLKEV
RSGSGSGSSKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLTLKFICTTGKLPVPWPTLVTTLTYGVQCFARYPDHMK
QHDFFKSAMPEGYVQERTISFKDDGTYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNFNSHNVYITADKQKNGIKANF
KIRHNVEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSVLSKDPNEKRDHMVLLEFVTAAGITHGMDELY
B:MPHLVIEATANLRLETSPGELLEQANKALFASGQFGEADIKSRFVTLEAYRQGTAAVERAYLHACLSILDGRDIATRTLLGASL
CAVLAEAVAGGGEEGVQVSVEVREMERLSYAKRVVARQRSKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLTLKFIC
TTGKLPVPWPTLVTTLTYGVQCFARYPDHMKQHDFFKSAMPEGYVQERTISFKDDGTYKTRAEVKFEGDTLVNRIELKGIDFKEDG
NILGHKLEYNFNSHNVYITADKQKNGIKANFKIRHNVEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSVLSKDPNEKRDHMVL
LEFVTAAGITHGMDELYLEHHHHHH

>>I3-01(ctGFP, 60-mer)

MHHHHHHGGSGGSGGSGGSMKMEELFKKHKIVAVLRANSVEEAKKKALAVFLGGVHLIEITFTVPDADTVIKELSFLKEMGAIIGA
GTVTSVEQCRKAVESGAEFIVSPHLDEEISQFCKEKGVFYMPGVMTPTELVKAMKLGHTILKLFPGEVVGPQFVKAMKGPFPNVKF
VPTGGVNLDNVCEWFKAGVLAVGVGSALVKGTPVEVAEKAKAFVEKIRGCTEGSGSGSGSGSSKGEELFTGVVPILVELDGDVNGH
KFSVRGEGEGDATNGKLTLKFICTTGKLPVPWPTLVTTLTYGVQCFARYPDHMKQHDFFKSAMPEGYVQERTISFKDDGTYKTRAE
VKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNFNSHNVYITADKQKNGIKANFKIRHNVEDGSVQLADHYQQNTPIGDGPVLLPD
NHYLSTQSVLSKDPNEKRDHMVLLEFVTAAGITHGMDELY

>>I3-01(nt/ctGFP, 120-mer)

MHHHHHHGSSKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLTLKFICTTGKLPVPWPTLVTTLTYGVQCFARYPDHM
KQHDFFKSAMPEGYVQERTISFKDDGTYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNFNSHNVYITADKQKNGIKAN
FKIRHNVEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSVLSKDPNEKRDHMVLLEFVTAAGITHGMDELYGGSGGSGGSGGSM
KMEELFKKHKIVAVLRANSVEEAKKKALAVFLGGVHLIEITFTVPDADTVIKELSFLKEMGAIIGAGTVTSVEQCRKAVESGAEFI
VSPHLDEEISQFCKEKGVFYMPGVMTPTELVKAMKLGHTILKLFPGEVVGPQFVKAMKGPFPNVKFVPTGGVNLDNVCEWFKAGVL
AVGVGSALVKGTPVEVAEKAKAFVEKIRGCTEGSGSGSGSGSSKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLTLK
FICTTGKLPVPWPTLVTTLTYGVQCFARYPDHMKQHDFFKSAMPEGYVQERTISFKDDGTYKTRAEVKFEGDTLVNRIELKGIDFK
EDGNILGHKLEYNFNSHNVYITADKQKNGIKANFKIRHNVEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSVLSKDPNEKRDH
MVLLEFVTAAGITHGMDELY

>>I3-01(HB)

MHHHHHHGGSGGSGGSGGSMKMEELFKKHKIVAVLRANSVEEAKKKALAVFLGGVHLIEITFTVPDADTVIKELSFLKEMGAIIGA

GTVTSVEQCRKAVESGAEFIVSPHLDEEISQFCKEKGVFYMPGVMTPTELVKAMKLGHTILKLFPGEVVGPQFVKAMKGPFPNVKF

VPTGGVNLDNVCEWFKAGVLAVGVGSALVKGTPVEVAEKAKAFVEKIRGCTEGGSGGSGGTKEEYDIEELRKKLKEQNKEMEKLKE

ELKKMEKLPKSPIAKLLILQMKLLLLQIENLQMQITMLEILSKR

VITA

Yang grew up in Taipei, Taiwan, and obtained both a BS and MSc in Biological Sciences from the University of the Pacific in Stockton, CA. He then moved to Seattle, WA to attend the University of Washington for his PhD.