Symmetry-Based Design of Protein Nano-Cages

by

Ajitha S. Cristie-David

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Chemistry)
in The University of Michigan
2019

Doctoral Committee:

Professor Neil Marsh, Chair
Assistant Professor Marcos Koutmos
Professor Brandon Routolo
Associate Professor Raymond Trievel

Ajitha S. Cristie-David

scristie@umich.edu

ORCID iD: 0000-0001-7166-6705

Dedication

This thesis is dedicated to

my beloved family,

my thesis advisor Dr. Neil Marsh,

all my teachers,

all my friends,

and everyone who made this journey possible

Acknowledgements

My dissertation would not be possible without the guidance and support of many great people around me. Even though it is not possible to list everyone, I want to thank all those who have challenged, guided and supported throughout this process. First and foremost, I would like to thank Dr. Neil Marsh. I have been tremendously fortunate to have found such a wonderful, knowledgeable and supportive thesis advisor. This dissertation would not have been possible without his enormous guidance, patience and care. It was a great pleasure and honor to have worked under the great supervision of Dr. Marsh.

My next deepest thanks goes to my family; my dearest husband Thushara Karunasekara and loving daughter Nayeli Karunasekara. I am very fortunate to have a very loving, caring and supportive husband who has always by my side encouraging me, lifting me up and strengthening me. Nayeli has and always will be my stress reliever, my forever cheerleader, who can always put a smile on my face no matter what. Having my husband and daughter in my life have been a great source of inspiration and motivation for me during the trials and challenges of graduate school and I am forever grateful.

In addition, I would like to thank my dissertation committee members, Dr. Raymond Trievel, Dr. Brandon Routolo, Dr. Mark Banaszak Holl and Dr. Markos Koutmos for their insight, constructive feedback and guidance. Next, Dr. Aaron

Sciore and Dr. Somaye Badieyan, my former colleagues, for their hard work and great input to elevate this work to its maximum success. Furthermore, I am grateful to my collaborators who have been an enormous help for this work. Here, my special thanks goes to Dr. Min Su for doing Cryo-EM, for Dr. Kai Sun for scanning TEM, Dr. Junjie Chen for AFM, Dr. Philipp Koldewey and Ben Meinen for AUC. Finally I would like to thank everyone in the Marsh Lab. It has been a great pleasure to work with such an amazing group of co-workers.

Table of Contents

List of Figures

List of Tables

Abstract

The self-assembly of protein subunits into large-scale oligomeric structures is a powerful and ubiquitous feature of biology, with viral capsids providing an excellent example. These assemblies perform a diverse set of functional and structural roles in living systems. Additionally, proteins can be modified both genetically and chemically to introduce new properties. Because of these attractive properties, natural and *de novo* designed protein assemblies have already been evaluated for various applications in medicine and materials science. This thesis explores a recently developed, generalizable coiled coil based strategy for *de novo* designing protein cages to utilize for applications in these fields.

The design strategy relies on the combination of 2 rotational symmetry elements, one provided by the natural, building block protein (BBP) and the other provided by the coiled coil, to specify a protein cage of the desired geometry. The oligomerization of the coiled coil brings the copies of the BBP together, leading to the assembly of protein cages. By employing BBPs and coiled coils of different rotational symmetries, cages of various sizes and geometries have been designed, including both tetrahedral and octahedral protein cages.

This thesis extends these studies to more ambitious design targets and explores the generalizability of this approach.

Because the design strategy requires well-specified homo-oligomeric

parallel-coiled coils, the utility of several selected *de novo* designed coiled coils was first evaluated as off-the-shelf components for protein assembly, using green fluorescent protein as a model system. This study revealed context-dependent oligomerization state changes for some of these coiled coils.

Next, the potential of elaborating previously designed protein cages by attaching additional protein domains to free end of the coiled coil was investigated. As a proof-of-concept, an octahedral cage was elaborated by fusing a large monomeric protein to the free end of the coiled coil assembly domain. This design successfully self-assembled into a homogeneous octahedral protein cage of ~ 1.8 MDa, significantly the addition of the extra protein domain dramatically improved the yield and efficiency of protein assembly.

The design strategy was extended to the *de novo* design of an icosahedral protein cage by fusing a pentameric coiled coil to the trimeric BBP previously utilized for octahedral and tetrahedral cage designs. After optimization, a construct with an 8-residue oligo-glycine spacer successfully assembled into a hyperstable 60-subunit protein cage with icosahedral geometry and molecular weight of ~ 2.1 MDa. Surprisingly, these cages captured short DNA strands during purification which were important to maintain the homogeneity of the cages. The cages could be transiently disassembled by treating with Dnase; the re-assembled cages were significantly more heterogeneous. The hyperstability and ability to capture DNA are new emergent properties of this design that arise from assembly and were not evident in previously designed cages.

Finally, the potential of extending this symmetry-based strategy to design protein

cages that assemble in response to environmental stimuli was investigated. This study was conducted by fusing a *de novo* designed metal-dependent coiled coil to the trimeric BBP. The construct successfully assembled into discrete particles in the presence of divalent transition metal ions; adding metal chelators or decreasing pH led to disassembly of these particles into their trimeric form.

Chapter 1: Introduction

1.1 Background

In 1959, Richard Feynman, a world famous physicist mentioned that "there is plenty of room at the bottom" during his speech at the annual American Physical Society Meeting at Caltech, California. The speech went unnoticed for several decades. However, this statement has attracted a great attention within recent years with the swift growth of "nanotechnology" in which scientists explore "the bottom" to design molecular machines for various applications. Particles with sizes ranging from 10-100 nm are considered as nanoparticles. Their small size, large surface/mass ratio, ability to penetrate cellular membranes are among some attractive properties of nanoparticles. Currently various types of nanoparticles, both synthetic and natural, have been developed for a board range of applications in medicine, synthetic biology, as well as materials science. Carbon nanoparticles, gold nanoparticles, polyamidoamine (PAMAM) dendrimers, and quantum dots are few examples for synthetic nanoparticles.[1–12] Natural nanoparticles include nucleic acid based nanoparticles and protein based nanoparticles.[13–27]

The focus of my thesis is on protein-based nanoparticles. Protein-based nanoparticles have potential applications in medicine and industry; they have many attractive properties over other nanoparticles, both natural and synthetic. Proteins can self-assemble into well-organized architectures and these assemblies perform a

diverse set of structural and functional roles in living systems. Proteins are rich in molecular recognition and they interact with all biological materials (DNA, RNA, Lipids and Carbohydrates), small organic molecules and inorganic substances (E.g. metal ions, H_2O, O_2, CO_2). New structural and functional properties can be introduced to proteins with atomic level accuracy, by both chemical and genetic modifications. Furthermore, functional groups can be appended to proteins; both covalently and non-covalently. Proteins can be easily expressed and purified in large scale and with low cost.[26,27] All these properties of proteins make them attractive biomaterials for the development of new functional materials.

1.2 Natural Protein Assemblies

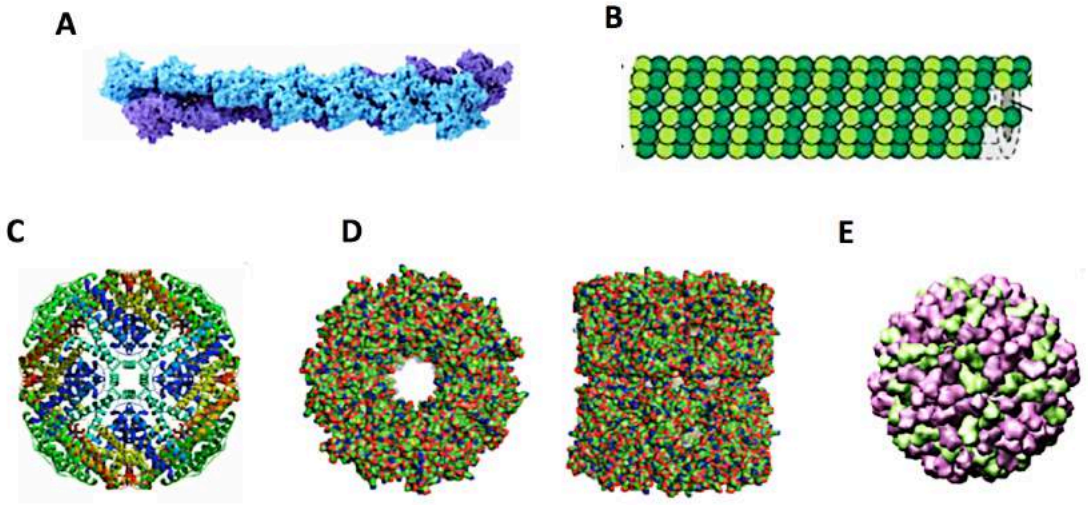

Figure 1. 1 Examples of natural protein assemblies. (A) Actin (*taken from http://ghr.nlm.nih .gov/handbook/illustrations/actin*). (B) Tubulin (*taken from Molecular biology of the Cell, 4th edition*). (C) Ferritin (PDB ID: 3GVY) (D) GroEL (top and side views, PDB ID: 3E76). (E) Cowpea chlorotic mottle virus (PDB ID: 1ZA7).

Proteins are natural nanomachines: The innate ability of proteins to assemble into both homo- and hetero-oligomeric forms increases the stability, co-operativity and functional efficiency of these nanomachines. Assembling into oligomeric structures reduces the solvent-exposed surface area and therefore tends to increase the stability of proteins. Large-scale quaternary structures or higher order assemblies of proteins are mainly grouped into two classes: extended forms are called filamentous proteins and closed-spherical forms are called protein cages (Fig. 1.1).

1.2.1 Filamentous Proteins

Filamentous proteins are mainly involved in structural and mechanical support in living systems. They are often composed of one or two repeating protein subunits and are assembled into fiber-like extended structures. Actin and tubulin are some well-studied examples (Fig. 1.1 A,B).[28,29] These proteins are mainly involved in maintaining the structural integrity of eukaryotic cytoplasm and are the best-studied examples of highly dynamic protein assemblies in nature. Polymerization and de-polymerization of actin and tubulin subunits into filamentous structures are spatiotemporally regulated upon extracellular and intercellular signals/stimuli.

Actin is the most abundant type of protein in most eukaryotic cells.[28] Globular monomeric actin subunits known as G-actin polymerize and form microfilaments or so called F-actin by hydrolyzing ATP. Actin filaments are part of cytoskeleton as well as contractile molecular motors of muscle cells. Actin participates in cell motility and division in all eukaryotes as well as muscle

contraction in animals. Tubulin is found nearly in all domains of life and is also involved in cell division, inter-cellular transport and motility.[29,30] Alpha and beta tubulin subunits, the two most common members in tubulin family, self-assemble to form microtubules, a process that requires GTP.[30] Microtubules are part of the cytoskeleton, form the internal structure of cilia and flagella and are integral to the division of chromosome during mitosis.

1.2.2 Protein Cages

Protein cages comprise, one or a few types of proteins that assemble into well-organized, geometrically constrained structures in highly symmetrical fashion to perform a wide range of functions in all living systems. These assemblies play many important roles in organisms, which include, storage, protection and transport of materials and function as size constrained reaction vessels. GroEL, ferritin, Pyruvate dehydrogenase complex (PDH) and viral capsids are a few examples (Fig. 1.3 C-E).

GroEL is a large double-ringed chaperonin protein found in bacteria.[31,32] GroEL forms a complex with GroES, a lid-shaped protein and this complex plays an important role by facilitating protein folding. Poorly folded proteins are recruited into the hydrophobic interior of GroEL where they undergo ATP-induced re-folding into their native forms.[32] Ferritin is a 24-subunit, octahedral protein cage with a negatively charged interior that acts as an iron storage container in both eukaryotes and prokaryotes.[33–35] The hollow interior of ferritin can hold up to about 45,000 iron atoms.[36] Ferritin reduces the risk of cellular toxicity by preventing the accumulation of free iron in the cytoplasm.

PDH is a multi-enzyme complex found in nearly all the organisms and is composed of multiple copies of three enzymes: pyruvate decarboxylase (E1), dihydrolipoamide acetyltransferase (E2) and dihydrolipoamide dehydrogenase (E3).[37,38] The PDH complex catalyzes the oxidative decarboxylation of pyruvate during cellular respiration and functions as a bridge between the glycolysis and citric acid cycle. This complex is a notable example of how proteins have evolved to perform sequential reactions effectively in living systems. Assembling into cages has helped these multi-enzyme complexes to efficiently channel the unstable intermediates and prevent them from diffusing to the exterior.

Viral capsids are very attractive biological containers and are excellent examples of smart protein assemblies. Viral capsids have evolved to recognize and bind to the host cell membrane; release their genome into the host cell; self-assemble to encapsulate the viral genome and lastly, escape from the host cell and propagate.[39] The most common viral cage geometry is icosahedral, although, prolate and helical capsids are also common. Cooperative binding effects inherent in the assembly of many protein subunits and reduced surface/volume ratios provide a high degree of stability to viral capsids. Some viral capsids form hyperstable protein architectures, which can survive in very harsh environments such as temperatures above the boiling point of water and in extremes of pHs.[40] Additionally, the positively charged, closed interior of viral cages provides a well-protected environment for their genomes. Some viral cages undergo environmentally responsive (for example in response to changes in pH, temperature or ionic strength) structural transitions and gating and which makes them very promising

nanomaterials for applications that require selective release of materials, for example release of drugs during targeted drug delivery.[41–43]

1.3 Functionalizing Protein Cages

Protein cages are versatile biomaterials, which can be repurposed for wide range of applications, and this field has gained a significant attention in nanotechnology in recent years. These nano-scale architectures are assembled from one or few types of subunits into very stable structures. Protein cages possess three distinct regions: an exterior surface, an interior surface and the inter-subunit interface, each of which can be modified both genetically and chemically to introduce new structural and functional properties (Fig. 1.2).[18,22] Multiple different functionalities can be introduced to protein cages at these surfaces. Moreover, protein cages exhibit well-defined molecular symmetries and form highly monodispersed-assemblies, and thus functional groups can be introduced to protein cages in a spatially well define manner.

The intriguing properties of protein cages have made them interesting platforms for developing new materials for use in medicine and industry. Natural protein cages for example ferritin, GroEL and virus like particles (VLPs) have already been developed for various applications in those fields; examples include, gadolinium nanoparticles encapsulated inside ferritin for MRI imaging, both the exterior and narrow interior of tobacco mosaic virus (TMV) have been utilized to template the synthesis of nanowires of Ni and Co and iron-oxide encapsulated ferritin cages have been utilized to fabricate metal oxide semicoductors.[25,44] *De novo* designed cages also have been evaluated for some applications. In one notable

example, a malaria vaccine was developed by fusing malaria coat protein to the exterior of a *de novo* designed polypeptide based protein cage, which conferred long lasting immunity in mice.[45]

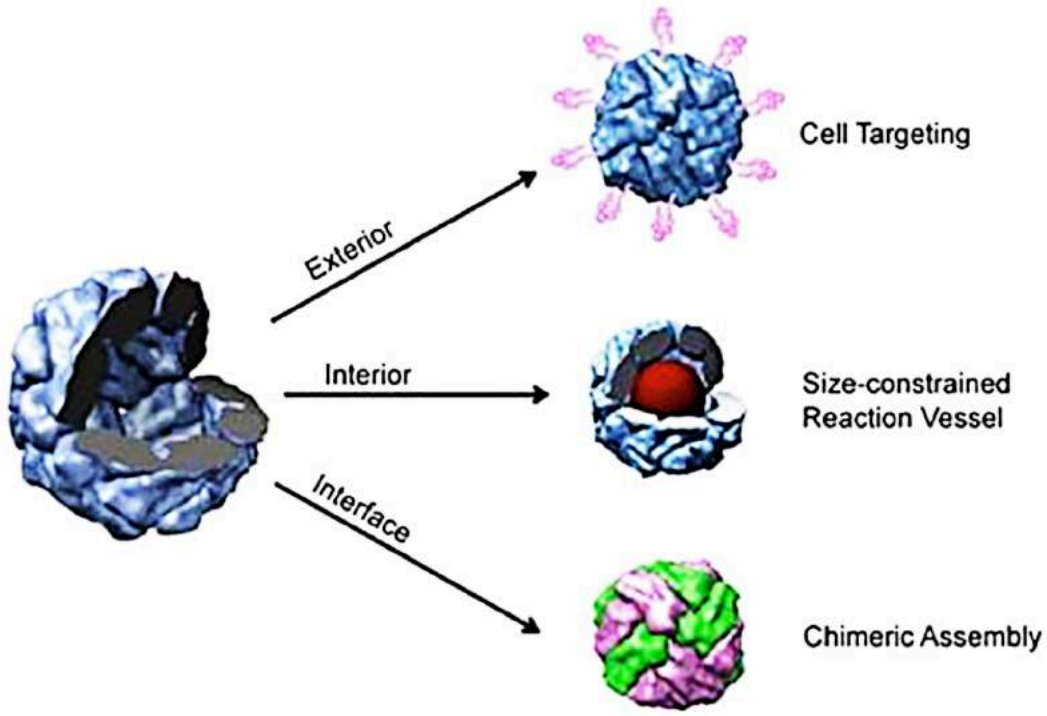

Figure 1. 2 A schematic illustration of protein cage interfaces available for structural and functional modifications (Image adapted from ref. 24).

1.3.1 Functionalizing Interior of Protein Cages

The well-sealed interior of protein cages provides a size-constrained compartment for encapsulation of materials. Natural protein cages utilize their interior for storage, transport of cargos, protection of encapsulated materials and as reaction chambers. A few examples were discussed in section 1.2.2. Introducing charged amino acid residues on the internal surface has allowed encapsulation of

materials via electrostatic interactions. DNA, RNA, proteins, peptides, synthetic polymers, fluorescent molecules, metal ions and other charged substances have been encapsulated inside protein cages using electrostatic interactions.[20,22,23,46,24,47–52] Some noteworthy examples of encapsulating proteins by employing electrostatic interactions comes from the Hilvert group.[53] They engineered negatively charged residues on the inner surface of luminase synthase (AaLS), a 60-subunit protein cage from *Aquifex aeolicus*. In one study, they developed a spectroscopic tool by encapsulating positively charged Green fluorescent protein (GFP) variants inside AaLS using charge complementarity. Additionally, an AaLS-GFP host-guest system was used to encapsulate enzymes by tagging positively charged GFP. Although no significant change in catalytic activity of enzymes was observed, this work demonstrated the potential of internalizing proteins by appending them to charged protein tags. In other work, they encapsulated a peroxidase enzyme and utilized it as a nanoreactor for polymerization of 3,3-diaminobenzidine, which produced more homogeneous polymers.

Additionally proteins, nucleic acids, fluorescent probes, small molecules and polymers have been introduced to the interior of protein cages by covalently linking them to natural or non-natural amino acids through various conjugation chemistries; for example cysteine-maleimide reaction, NHS-ester formation or "click" chemistry.[22,23] In one notable example the Francis group covalently attached prophyrin rings to bacteriophage M22 via cysteine-malemide coupling to develop a photodynamic therapeutic.[19] They selectively delivered these functionalized capsids to Jurkat leukemia cells and the reactive oxygen species generated by the prophyrin

upon light illumination killed these cells with ~ 75% efficiency. In another study by the same group, p-nitrophenylalanine derivatives were internalized inside bacteriophage M22 by covalently linking them to tyrosine a reaction which can be translated to conjugate drugs for biomedical applications.[23] Another group compartmentalized *Candida antarctica* lipase B (CalB) inside CCMV by covalent conjugation, which protected the enzyme from protease degradation.[54]

Proteins and peptides have also been encapsulated inside cages by employing protein-protein interactions as well as genetically fusing them directly to the protein forming the cage. The Comellas-Aragonès group encapsulated GFP into a cage using a heterodimeric coiled coil oligomer and is a notable example for employing protein-protein interactions to encapsulate proteins.[55] The Douglas group encapsulated a number of enzymes inside VLPs by genetically fusing these proteins to a capsid's scaffold protein.[25,56,57] When the hydrogenase enzyme was encapsulated inside a VLP, the catalytic activity was increased by 150-fold compared to the free enzyme.[57]

Hydrogen is a very good, renewable, non-polluting energy source, but bio-synthesizing it, is a challenge. This work is an excellent proof of concept study which can be extended to design highly catalytically active enzyme nano-reactors. The cage interior mimics the crowded environment inside the cell and more likely provides a native-like environment for the enzymes. This presumably increases the catalytic efficiency of enzymes encapsulated in the cage interior. In another study, an enzyme cascade was encapsulated inside a VLP. Even though no significant change in catalytic rate was observed in this study, it demonstrated the potential of

designing confined multi-enzyme systems.[56] Similar strategies can be extended for multi-enzyme reactions involving unstable or toxic intermediates. These studies open up a new field of nanotechnology utilizing protein-based nanoreactors for real world applications, for example efficient synthesis of biofuels such as Hydrogen and constructing new biomimetic catalytic materials.

Nucleic acids have been encapsulated inside protein cages by genetically fusing either RNA or DNA binding peptides to the cage interior. Some research groups have genetically fused nucleic acid binding peptides to the interior of cages and then evolved these cages to selectively capture RNA molecules encoding the cage protein thereby resembling viral capsids.[58,59] The ability to evolve protein cages to selectively capture specific nucleic acid strands suggests a future use in gene therapy. Protein cages encapsulating nucleic acids are already being developed for various biomedical applications including gene delivery, RNA delivery (siRNA and microRNA), vaccines and CRISPR-Cas9 delivery.[60–63]

Functional properties of inorganic nanoparticles highly depend on their shapes and sizes.[12] Synthesizing them in controlled sizes and shapes is a challenge, but the protein cage interior provide a well-defined, size-constrained environment to synthesize nanomaterials with a great precision. Either genetically fusing metal chelating peptides to the interior of the cages or incorporating chemical groups promoting electrostatic interactions have allowed the synthesis of various inorganic nanoparticles; E.g. CoPt, FePt, CdS, ZnSe, SeO_2, FeO, Au, Gd and Ag.[18,21,33,44,64–66] Inorganic nanoparticles templated by protein cages have been developed for wide range of potential applications in medicine as well as industry. Some examples for

medical applications include, nanoparticles utilized as carriers for drug delivery (E.g. Au), as therapeutic agents (E.g. FeO for hyperthermia cancer therapy, Au for photothermal therapy) and for bio-imaging (E.g. Gd^{3+} and FeO as contrasting agents for MRI imaging).[11,67] Industrial applications include semiconductors, transistors, nanowires and catalysts. [18,44,68]

1.3.2 Functionalizing Exterior Surfaces of Protein Cages

Novel functionalities can be incorporated onto the exterior surface of the cages using similar approaches to those described above. Protein cage exteriors have already been functionalized for both medical and industrial applications. Exterior surfaces have been decorated with antigen for vaccine delivery.[18,22,23,69] Several VLP-based vaccines are already available on the market; Gardacil and Cervaric for Human Papilloma Virus (HPV) and Mosquirix for Malaria are some examples.[70] Therapeutic agents such as small molecules and peptides have been covalently attached to the cage exterior.[19,71] Fullerene is a potential theragnostic for cancer and inflammatory disease, however solubility is a concern. In one notable example, the Finn group covalently attached fullerene to the exterior of CPMV via amine reactive cross-linking (NHS) chemistry, which increased its solubility and biocompatibility.[19] Moreover, fullerene conjugated capsids were successfully taken up by HeLa cells with no inhibition by fullerene. In another example, doxorubicin, a chemotherapeutic drug, was conjugated to exterior of CPMV and delivered to HeLa cells, which efficiently killed the cells at a low dosage compared to free doxorubicin.[19]

For targeted drug delivery and similar applications, protein cages are decorated with tissue specific antibodies, peptides, aptamers and tissue specific ligands.[18,22,23,72] The RGD sequence motif from adenovirus, a ligand with natural affinity for upregulated intergrin receptors in endothelial cells of tumor vessels, is a widely used ligand for tumor-targeted drug delivery.[21] Protein cages decorated with RGD have showed enhanced transport of drugs to tumor regions.[21] Fluorescent molecules have been covalently attached to exterior for imaging.[18] Viral capsid exteriors have also been utilized to generate phage display libraries.[73]

The exterior of cages have been functionalized for some industrial applications. Metal-binding peptides have been conjugated to protein cage exteriors to template nanocrystals for nano wire formation. ZnS and CdS nanowire have been synthesized by displaying peptides on M13 bacteriophage.[74] The interior and exterior of TMV virus has been utilized to template Ni and Co nanowires.[20] A conducting network of cowpea chlorotic mottle (CCMV) was designed by decorating its exterior with gold nanoparticles.[20]

More complex, hierarchical architectures: 1D filaments, 2D layers and 3D lattices have been designed by employing protein cages and by modifying their exterior surface to drive the inter-subunit interactions.[25] The exterior surface of GroEL was modified with merocyanine, an organic compound with affinity to Mg^{2+}, which led the protein to assemble into 1D tubes in the presence of Mg^{2+}.[25] In another approach, a thiol group was introduced to the exterior of CCMV capsids and which assembled into 2D layers when incubated with Au.[75] Both 2D and 3D

architectures have been designed by mixing negatively charged CCMV with positively charged substrates.[25]

Tezcan's group designed a metal-directed 3D protein lattice by introducing histidine residues at the 3-fold axes of ferritin.[76] In another example, a catalytically active protein super-lattice was designed by genetically modifying the exterior of a viral capsid to drive the 3D assembly while also encapsulating two enzymes in a cascade reaction, inside the capsid.[77] These particles assembled into 3D arrays, which could be recycled and reused. This is an excellent proof of concept study demonstrating protein-based nanoreactors with advanced properties that can be extended to various industrial applications.

1.3.3 Functionalizing Protein Cage Interfaces

The interface of protein cages can be modified to introduce controlled or environmentally responsive assemblies and gating mechanisms. However, this is a slowly progressing area. Both exterior and interior surfaces of protein cages interact with the solvent and therefore, modifying proteins at those surfaces is less likely to lead to significant alterations of the structural properties of proteins. However, at interfaces, proteins interact with its neighboring protein subunits via protein-protein interactions. Therefore, introducing new properties to protein assemblies at their interfaces often disrupts the assembly and has been a challenging field in protein re-engineering.

Many viral capsids are dynamic structures and some capsids can undergo environmentally responsive changes. Some viral cages undergo environmentally dependent assembly/disassembly and some have gated pores which swell open in

response to environmental changes (E.g. pHs and salt concentrations).[22,78] New interfaces with controlled assemblies have been introduced to a very few natural protein cages. A pH-responsive ferritin protein cage that disassembled at low pHs was designed by genetically introducing a "GALA peptide" sequence.[79] This is a *de novo* designed pH-responsive motif which undergoes acid-induced structural transitions. When introduced into the interface of protein subunits the acid-induced structural transition caused the cage to dissassemble.[79] In another example, a pH-responsive protein cage was designed by introducing histidine residues to the interface of E2 complex from PDH, a 60-subunit protein cage.[80] Decreasing pH to pH 5.0 led to disassembly of the E2 protein cages.

Protein cages with controlled assemblies will be powerful platforms for number of applications. pH-responsive cages will be promising delivery systems for lysosome-targeted drug delivery where low pH results in the disassembly of cages and release of the cargo. Moreover, cages with controlled interfaces or pores can be utilized to design nanoreactors where a control of material flow is important. Furthermore, such an approach will allow development of re-usable nanoreactors.

1.4 *De novo* Design of Protein Cages

The above examples show that protein cages are smart biomaterials, which can be utilized to design advanced nanomaterials for a wide range of applications. Growing interest in protein nanocages also increases the demand for protein cages with new architectures and capabilities that go beyond the ones found in nature. However, *de novo* designing protein cages has proved difficult because of our incomplete understanding of protein folding. Nevertheless, this field has made great

progress within last two decades. Natural proteins that assemble into homo-oligomeric structures form structures that are often highly symmetrical and possess rotational, dihedral and translational symmetries. In 2001, Todd Yeates's group realized that protein cages could be *de novo* designed by employing proteins with rotational symmetries to serve as building block units.

Proteins with rotational symmetries (C_n) often possess regular polygonal shapes and therefore, can serve as promising building blocks to *de novo* design protein cages with various geometries. Various platonic geometries could be generated with regular polygons. In platonic solids, the same number of polygons meet at each vertex with a minimum number of three and at an internal angle of less than 360°. Only equilateral triangles, squares and pentagons can fulfill this requirement and natural proteins with C_3, C_4 and C_5 rotational symmetries respectively often resemble those polygonal shapes.

There are five platonic solids: the tetrahedron, the cube, the octahedron, the dodecahedron and the icosahedron. All these geometries possess three operational (rotational) symmetries: faces and vertices possess C_3+C_3, C_3+C_4 or C_3+C_5 symmetric combinations and the edges have C_2 rotational symmetry (Fig. 1.3). Any platonic solid can be specified by pairwise combination of two of its symmetry elements and the dihedral angle between the principal rotational axes. However, pairing a C_2 symmetry element with any rotational symmetry element can lead to multiple assemblies unless the dihedral angles between subunits are precisely defined in order to satisfy the intended geometry (Fig. 1.3).

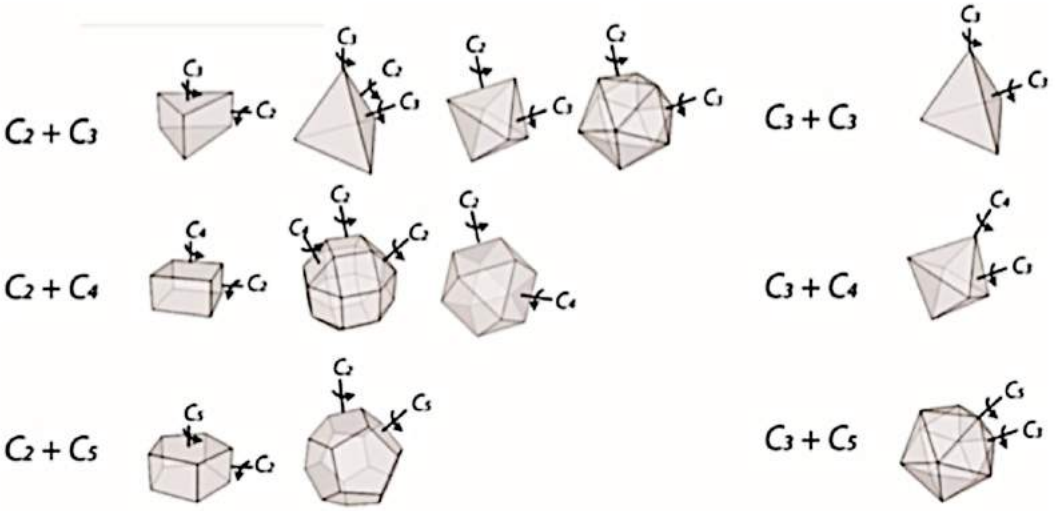

Figure 1. 3 Different cage geometries can be generated by pairwise combination of symmetries. *Left:* Pairing with C_2 symmetry allow multiple geometries. *Right:* Pairing other rotational symmetries allow formation of specific geometries/platonic solid form (Image credits: Dr. Ben Buer).

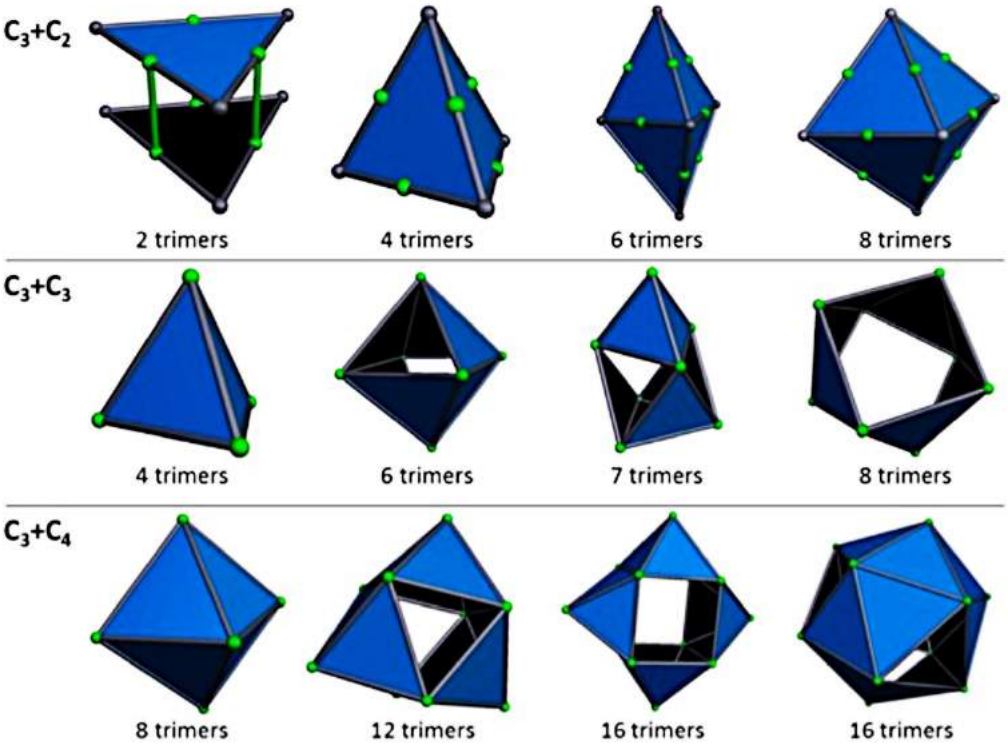

Figure 1. 4 Examples of possible geometries can be formed when a triangular or trimeric building block is paired with C_2, C_3 or C_4 (Image credits: Dr. Aaron Sciore).

We note that pairwise combinations of symmetries can generate many forms of geometries other than platonic solids (Fig 1.4). Nevertheless, the platonic solids are stable geometries with minimized surface/volume ratios and smaller surface area exposed to the exterior compared to other geometric forms. Proteins tend to assemble into energetically more favorable, ordered structures and platonic solids satisfy those requirements. Therefore, we hypothesize that during assembly protein subunits will favor platonic solid geometries over the other forms. Theoretically it is possible to design various cage-like protein architectures using one or more types of proteins with rotational symmetries.

Yeates' lab was the first group to recognize that protein cages could be *de novo* designed by employing the symmetry. In 2001, they designed a tetrahedral protein cage, the first *de novo* designed protein cage, by exploiting the symmetry of natural proteins.[81] Even though the initial design had some heterogeneity, this work opened up a whole new field of nanotechnology. Since then several approaches, both directed and fundamental, have been introduced to *de novo* design protein nanocages by utilizing the symmetry.

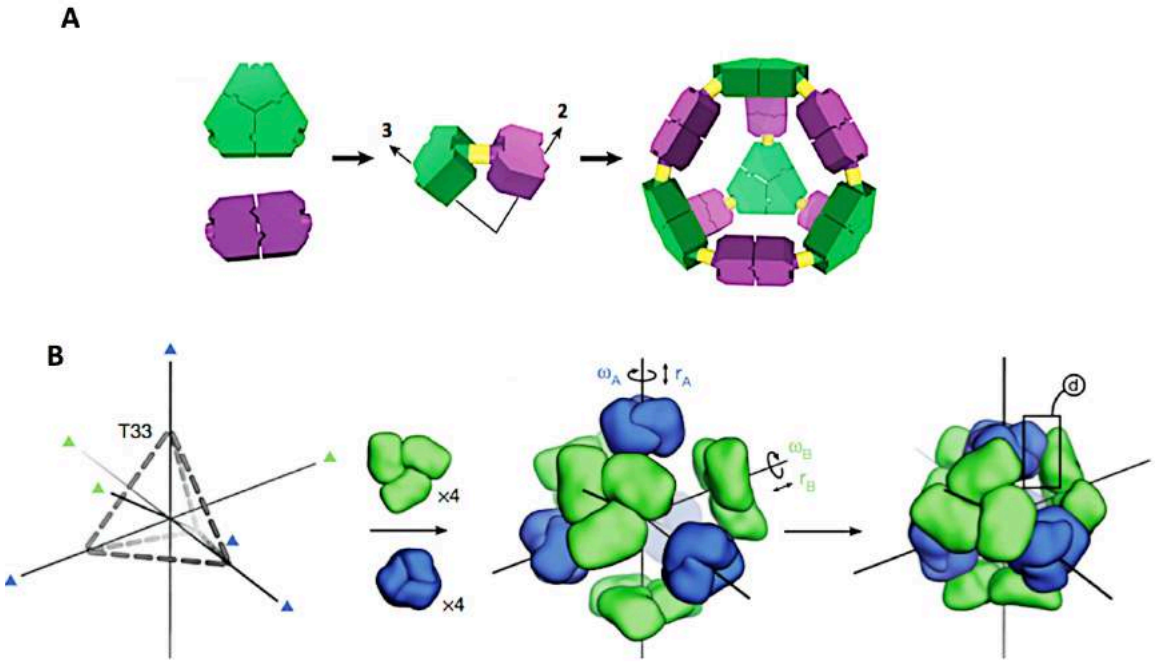

Figure 1. 5 Schematic illustrations of two major protein design approaches. (A) Designing of a tetrahedral protein cage using oligomeric fusion strategy by the Yeates' group (Image adapted from ref. 83). (B) Designing of an octahedral cage using interface design strategy by the Baker group (Image adapted from ref. 86).

Both the oligomeric fusion strategy and the interface design strategy have made significant progress (Fig. 1.5). The oligomeric fusion strategy was introduced by the Yeates' group. In this design strategy, a fusion construct is designed by genetically joining two natural oligomeric proteins with desired rotational symmetries via a short α-helical linker. A rigid α-helical linker is used to constrain the geometry of the cage so that correct dihedral angle is formed. The geometry can be fine tuned by modifying the α-helical pitch by varying the linker length.

The Yeates' group initially designed a tetrahedral cage by fusing a trimeric protein subunit with C_3 symmetry and a dimeric subunit with C_2 symmetry through a helical connector.[81] However, the cages were too heterogeneous to be

characterized by crystallography. In 2012, they revisited the design and removed two amino acid residues from the helical linker. This modification helped them to successfully re-design a homogeneous tetrahedral cage and that's structure could be verified crystallographically.[82] Subsequently, the Yeates group designed cubic and dodecahedral protein cages and extended structures (filaments and lattices) utilizing this approach.[83–85]

The oligomeric fusion strategy opened up a new field of protein design and engineering, which allowed design of both closed-cage like and extended protein architectures utilizing symmetry as a powerful tool. However, in this approach, designing protein cage architectures require the dihedral angle between protein subunits to be precisely controlled to obtain the intended geometries. The rigid α-helical connector is evaluated intensively by varying the residues both *in silico* and experimentally to optimize the dihedral angle. Nevertheless, we note that oligomeric fusion strategy is far more promising for designing extended materials where relaxed linkers can be incorporated and therefore the stress for optimizing the linker is largely reduced. Various research groups have utilized oligomeric fusion approaches but with flexible linkers to design filaments, 2D arrays as well as 3D crystals.[83]

The interface design strategy introduced by David Baker's group has gained impressive success. This design strategy employs sophisticated computational modeling to *de novo* design protein cages and which include symmetric docking of proteins and design of low energy protein-protein interfaces. The approach has been used to design one and two component protein cages with various

geometries.[86–89] Here, one interface is provided by the natural oligomeric protein and a second protein interface is introduced by computational modeling to drive the assembly. [86,87]

First, natural protein subunits with rotational symmetries of interest are symmetrically docked in a geometry of interest using the symmetric modeling framework in Rosetta.[87] The subunits are allowed to freely rotate around and translate along their symmetry axes. Other rigid body movements are restricted in order to prevent the formation of non-specific geometries. A scoring function is used to calculate low energy interfaces with minimum steric clashes. The configurations with repeating symmetrical inter-subunit interfaces and high scoring functions are selected. The models with high scores possess high densities of contacting amino acid residues at their interfaces, which largely reduce the risk for alterations in the properties of protein upon modification of surface amino acid residues.

Next, the amino acid residues on those inter-subunit interfaces are genetically modified to generate novel, geometrically complementary interfaces with the most stable interfaces predicted using Rosetta and interactive design in Foldit. This involves sampling and identifying the configurations of amino acid residues on the inter-subunit interfaces, screening the natural protein interfaces which closely resemble these inter-subunit interfaces and mutating amino acids in inter-subunit interfaces to closely match the properties of those natural interfaces. Natural protein interfaces possess well-packed hydrophobic regions surrounded by polar amino acid residues and similar properties are introduced to the new inter-subunit interfaces in order to drive the protein assembly. One and two component

protein cages with tetrahedral, octahedral, dodecahedral and icosahedral symmetries have been successfully designed using this strategy.[86–89]

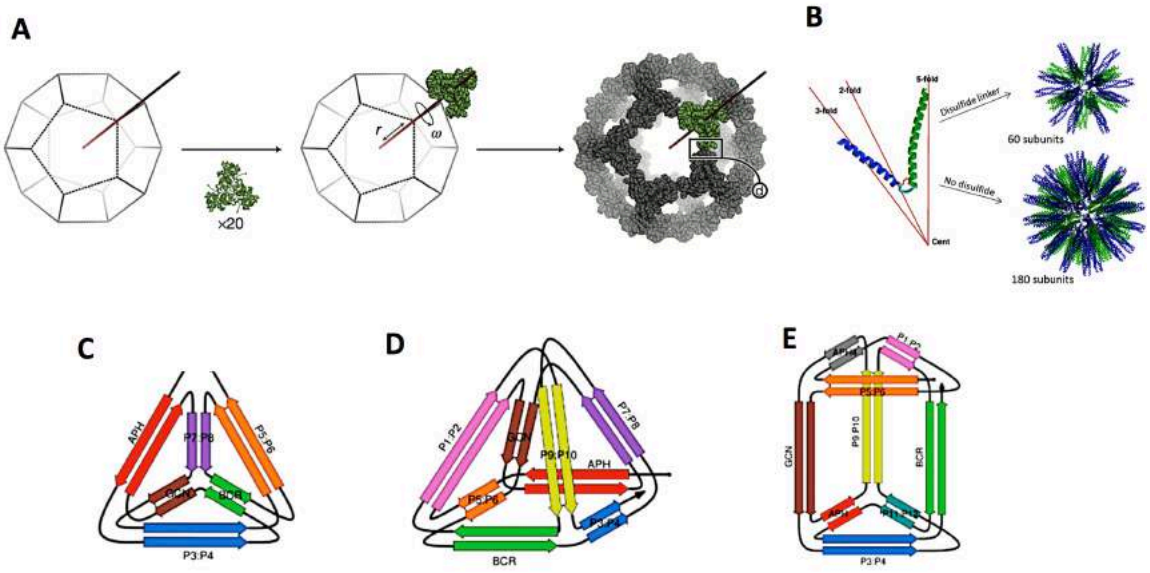

Figure 1. 6 Examples of *de novo* designed protein cages. (A) Designing of a 60-subunit icosahedral protein cage using interface design strategy by the Baker lab (Image adapted from ref. 88). (B) Designing of icosahedral protein cages using coiled coil motifs by the Burkhard group (Image adapted from ref. 90). (C-E) Porous Tetrahedral, pyramidal and prism shape cages designed using single polypeptide chains composed coiled coil motifs, by the Jerala group (Image adapted from ref. 91).

The interface design strategy has made striking progress and can be considered as a very promising approach for *de novo* design of protein cages. However, it is achieved at the high cost of sophisticated protein modeling and intensive protein engineering. In a landmark paper, 271 trimeric proteins were computationally evaluated in order to design a tetrahedral and an octahedral cage.[87] 8 constructs were expressed to screen for tetrahedral assemblies and 33 constructs for octahedral assemblies. However, only one construct from each screen assembled as intended. In 2016, they utilized this strategy and designed an one-component

dodecahedral cage and multiple two-component icosahedra, the highest platonic solid forms. When designing 60-subunit dodecahedral cage, 300 natural trimeric proteins were docked and 17 of them were selected for experimental screening with only one assembled into the intended geometry (Fig. 1.6 A).[88] For two component icosahedra, approximately 350,000 symmetry pairs were docked and about 300 constructs were experimentally screened, only 8 designs assembled into icosahedral assemblies.[89]

Protein cage architectures have also been designed using self-associating motifs such as coiled coils. The Burkhard group designed cage-like assemblies by genetically fusing a trimeric coiled coil with a pentameric coiled coil through a flexible oligo-glycine linker (Fig. 1.6 B).[90] However, the initial assemblies were very heterogeneous. Disulfide bonds were introduced between each side of the oligo-glycine linker to rigidify the dihedral angle in order to increase the homogeneity of assemblies. Subsequently, this construct was slowly refolded under oxidative conditions and which led to formation of a fairly homogeneous 60-subunit icosahedral cage.

The Jerala group designed various cage-like assemblies (E.g. tetrahedral, pyramidal, prism) using single polypeptide chains comprising parallel and antiparallel coiled coils with flexible linkers (Fig. 1.6 C-E).[91,92] Several other groups have explored coiled coil based strategies to design protein cages and extended materials.[93–97] This coiled coil-based design strategy is also referred to as coiled coil protein origami (CCPO) and this strategy provides a simple route to design novel protein architectures. However, coiled coils are quite flexible peptide motifs and

therefore, the resulting cages are often very heterogeneous and overly flexible. Furthermore, only very porous cages can be designed using these motifs. Nevertheless, they are attractive platforms for antigen display and similar applications. One noteworthy example is the functionalization of a coiled coil-based icosahedral cage design by the Burkhard group to develop a malaria vaccine.[45]

1.5 A Generalizable Approach To Design Protein Nano-Cages

To overcome the reliance on computational modeling and intensive protein engineering the Marsh lab has developed a flexible symmetry-based approach. This relaxes the requirement for the symmetry axes to be explicitly oriented at precise angles. This design strategy utilizes small coiled coil domains as modular, off-the-shelf assembly domains that are fused to a homo-oligomeric building block protein through a short, flexible linker sequence. The geometry of the protein cage is thus primarily specified by the rotational symmetries of the coiled coil and building block protein. This flexible approach to protein assembly produces cages that are more conformationally labile than those assembled using designed protein interfaces, but has the advantage that it does not require extensive computational modeling.

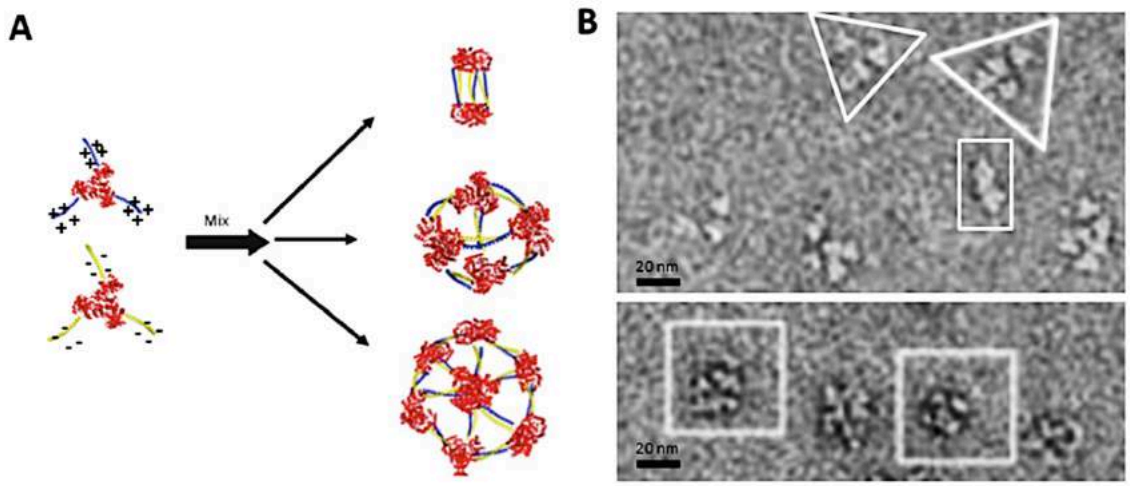

Figure 1. 7 Assembly of Aldolase trimmers into multiple species. (A) The design concept. (B) Negative stain-TEM images of cages assembled into multiple species. (Images adapted from ref. 98)

First step of the design process involves identifying of natural oligomeric "building block" proteins that possess rotational symmetry due to their quaternary structure. The requirements in choosing a building block protein (BBP) were: the protein must possess one of the desired rotational symmetry; the presence of one of its termini in a vertex; it should be easy to express in a cost effective expression system such as *E. coli*. Next, a coiled coil domain is genetically introduced via a flexibly oligo-glycine (Gly) spacer to the terminus located at the vertex of the BBP. The assembly of protein subunits into protein cages is mediated by oligomerization of the coiled coil.

In initial attempts to design protein cages, our group selected KDPG-aldolase from *T. maritima*: a trimer with C_3 rotational symmetry as the BBP.[98] In this design, a hetero-dimeric, anti-parallel coiled coil with C_2 rotational symmetry was utilized to mediate the assembly.[98] This hetero-dimeric coiled coil was composed of a

predominantly positively charged helix and a predominantly negatively charged helix, which formed dimers only when they were mixed together. Therefore, two fusion constructs were designed by genetically fusing each helix to the trimeric protein: fusion constructs with positively charged helix and negatively charged helix were named A-(+) and A-(-) respectively. When the fusion constructs were mixed at 1:1 ratios, they assembled into mixture of complexes (Fig. 1.7). However, this study demonstrated the potential of employing coiled coils as off-the-shelf components to mediate higher order assemblies. As discussed in previous section, pairing a C_2 symmetry element can lead to formation of heterogeneous structures unless the dihedral angel between subunits are precisely defined. Our group identified the drawback of this initial design and attempts were taken to develop this approach to be more robust for *de novo* designing protein cages.

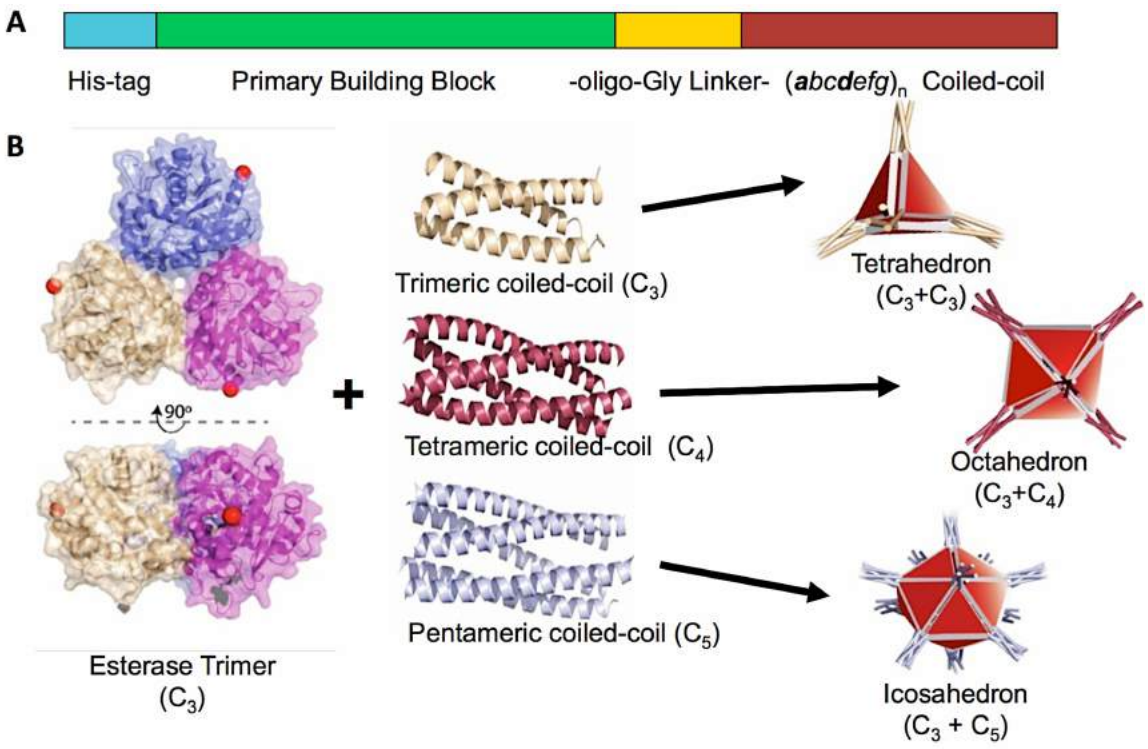

Figure 1. 8 The coiled coil-directed design concept. (A) Genetic fusion construct. (B) Formation of protein cages with different oligomeric states when C-termini of the trimeric esterase (red spheres) genetically fused via a flexible oligo-Gly linker to one of parallel coiled-coils of known oligomerization state. (Image credits: Dr. Ben Buer).

Our group speculated that utilizing homo-oligomeric parallel-coiled coils with higher rotational symmetries (C_3, C_4 or C_5) might alleviate formation of mixed protein assemblies. For this study *Pseudomonas Putida* trimeric esterase (PDB ID: 1ZOI), a C_3 symmetric protein was chosen as the BBP and a coiled coil sequence with the desired oligomeric state was genetically fused to the C-terminus of the BBP through a short oligo-Gly linker. The geometry of the protein cage is specified by the combination of rotational symmetry elements represented by the protein quaternary structure and the oligomerization state of the attached coiled coil (Fig. 1.8). Here, the BBP has C_3 symmetry and pairing it to homo-oligomeric coiled coils

with different symmetries should, theoretically, assemble it into protein cages with different platonic geometries: C_3 BBP + C_3 coiled coil should specify to a 8-subunit tetrahedral cage, C_3 BBP + C_4 coiled coil should specify to a 24-subunit octahedral cage and C_3 BBP + C_5 coiled coil should specify to a 60-subunit icosahedral protein cage.

However, in order to get the intended closed, cage-like geometries it was found to be very important to optimize the linker length and the strength of the coiled coil. Our group employs an oligo-Gly linker to provide flexibility between the two protein domains. If the glycine linker is too short, steric classes will prevent the cages from properly assembling; whereas too long of a linker leads to an overly flexible structure that misassembles. The strength of the coiled coil is also a key-determining factor for getting the subunits to assemble into homogenous cages. In higher order protein assemblies, protein interfaces have very weak interactions. However, once assemble, they make very stable protein cages. Viral capsids are excellent examples: If the subunit interactions are too strong, the partially assembled capsids become trapped in intermediate states that are too stable to proceed to the formation of the final capsid. A very weak interaction is also problematic and can lead failure of the capsid to assemble at all. In the present case, because coiled coils are built on heptad repeats of amino acid residues, the strength of the coiled coils can be controlled by varying the number of heptad repeats.

1.5.1 Designing an Octahedral Protein Cage

Using this design concept, our laboratory first attempted to design an octahedral protein cage by fusing TriEst to a 4-heptad tetrameric coiled coil (Fig.

1.9).[99] This project was led by Dr. Aaron Sciore. As a homo-oligomeric, parallel coiled coil is needed, a *de novo* designed tetrameric coiled coil (PDB ID: 3R4A) with Ile and Leu residues at "a" and "d" positions was selected to specify its geometry. For this first cage design, the inter-terminus distance between the C-terminus of the TriEst and N-terminus of the coiled coil was determined by using a search algorithm implemented in the Rosetta program in collaboration with David Baker's group at the University of Washington, Seattle. TriEst was arranged along the C_3-axes in octahedral geometry and the coiled coil array on the C_4-vertices, the inter-terminus distance was minimized by rotating these subunits and calculating the distance for lowest energy interface. The distance was estimated by modeling to be 9.1 Å, which is the distance of an approximately 3-residue long linker.

Three fusion constructs were designed by starting from a two-residue linker (GT) and by adding one glycine residue at a time, there by measuring the linker length from 6 Å to 12 Å. Constructs were named as Oct-2, Oct-3 and Oct-4 respectively and they were expressed in *E. coli*. From the 3 constructs, Oct-4 with a 4-residue linker, assembled into more homogenous cages and was characterized in detail using negative stain-TEM, cryo-EM, analytical ultracentrifugation (AUC) and native PAGE. About 35,000 particles were collected by cryo-EM, and the 3D construction at a resolution of 17 Å confirmed that the construct assembles into an octahedral protein cage. 2-dimensional sedimentation plots (2DSA) generated by AUC showed that about 75 % of Oct-4 assembles into the intended geometry. A single and symmetric elution profile in size exclusion chromatography (SEC) and native-PAGE gel were further evidence for the homogeneity of Oct-4 assemblies.

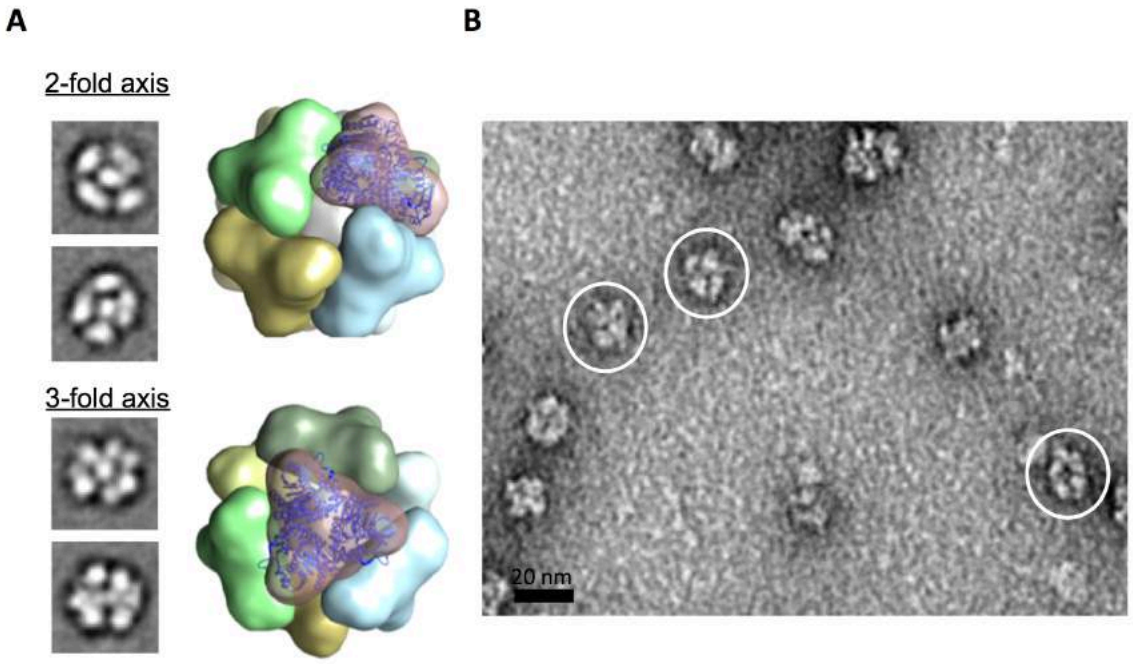

Figure 1. 9 "Oct-4" Octahedral protein cage. (A) Cryo-EM of Oct-4. *Left:* 3D map projections; *Right:* Reconstructed electron density map. (B) Negative stain TEM of Oct-4. (Image adapted from ref. 99)

Additionally, AUC analysis showed that Oct-4 has a frictional ratio (f/f_0) of 1.89 and which suggested that Oct-4 maintains a hollow interior as expected for a protein cage. Because of the flexibility of the cages and low resolution of the cryo-EM 3D-reconstruction, the orientation of the coiled coils (facing inward or outward) could not be determined. Overall, characterization of Oct-4 indicated that it is possible to design an octahedral protein cage by utilizing a simple coiled coil driven design approach. Next, our group wanted to design protein cages in other geometries in order to evaluate the generalizability of our strategy.

1.5.2 Designing a Tetrahedral Protein Cage

Our laboratory next evaluated the potential of designing a tetrahedral

protein cage. This project was led by Dr. Somaye Badieyan. Theoretically, a tetrahedral protein cage could be designed by fusing the TriEst with a trimeric coiled coil. In order to test this, TriEst was fused with a *de novo* designed 4-heptad-homo-trimeric parallel coiled coil (PDB ID: 4DZL). This coiled coil's trimeric state is specified by placing Ile at both "a" and "d" positions. Several constructs were evaluated by varying the glycine linker length and the number of heptad repeats (4 heptads and 5 heptads).[100] First constructs containing a 4-heptad repeat trimeric coiled coil fused to TriEst were evaluated and the linker length varied from 4 to 12 residues, increasing 2 residues at a time. With an 8-residue linker ($GT(G)_6$), about 25 % of the proteins assembled into tetrahedral cages as evidenced by native-PAGE, but other constructs showed a higher heterogeneity and no tetrahedral assemblies were evident. Next, the constructs with 6-, 8- and 10- residues linkers were re-evaluated after adding one more heptad repeat to the coiled coil. The construct with 5-heptads and a 8-residue linker ($GT(G)_6$) assembled into very homogeneous, discrete particles. Characterization of the assemblies was performed using the same set of analytical tools mentioned under the section on octahedral cage design.

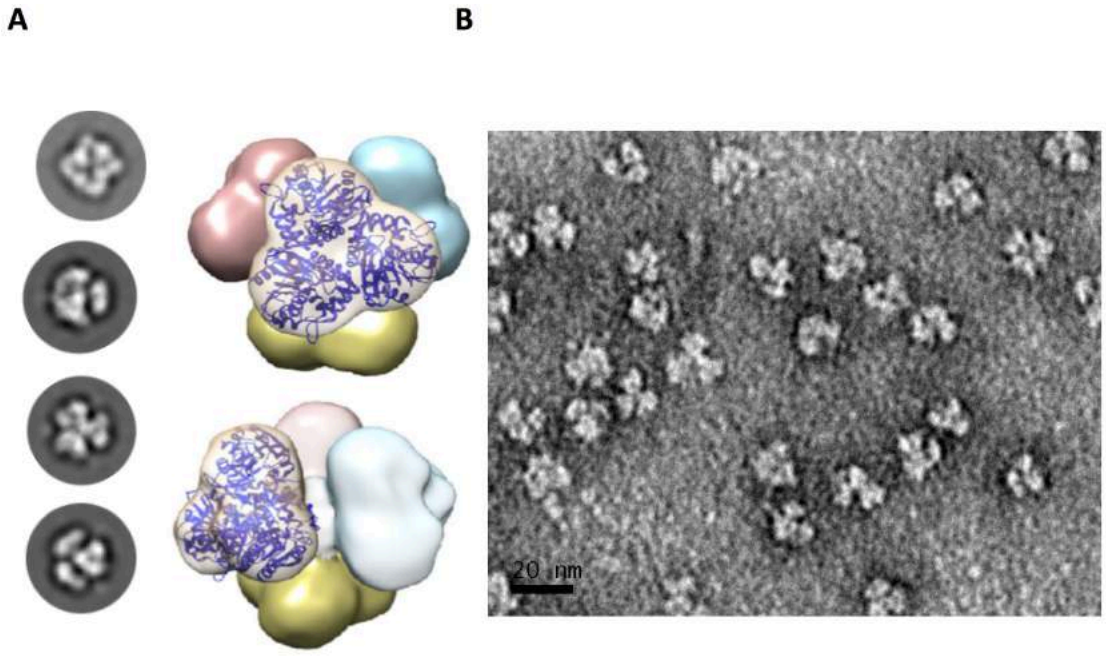

Figure 1. 10 Tetrahedral protein cage design. (A) Cryo-EM of Tet8-5H. *left*: 3D map projections; *Right:* Reconstructed electron density map. (B) Negative stain TEM of Tet-8-H5. (Images adapted from ref. 100).

2DSA plots generated by AUC showed that the construct assembles into a single hydrodynamic species with M_w of 430 kDa and frictional ratio of f/f_0 = 1.5. The calculated M_w obtained by AUC was in a very good agreement with the predicted M_w of 439 kDa. Additionally, the M_w was determined using native mass spectrometry (native MS), which gave a M_w of 434 ± 0.2 kDa and was very close to the predicted M_w. Samples collapsed during the vitrification in cryo-EM attempts, possibly due to the flexibility of the cages. Therefore, the samples were cross-linked with bis(sulfosuccinimidyl)suberate (BS3) cross-linker to prevent them from collapsing during the vitrification step; this cross-linker reacts between amine in the side chains of lysine and has a ~ 11 Å cross-linking span. Cryo-EM images were taken for those cross-linked samples. ~ 35,000 particles were collected and 3D

reconstruction map was generated using class-averaged images. The 3D map at 13 Å resolution revealed that these cages have the tetrahedral symmetry with coiled coils facing outward (Fig. 1.10).

Overall, our laboratory successfully designed both a tetrahedral and an octahedral protein cage from the same BBP using the coiled coil-based design strategy. Only a few constructs needed to be experimentally evaluated to identify successfully assembled proteins. Only three gene constructs were evaluated for the octahedral cage design and nine constructs were evaluated for the tetrahedral cage design. The work described above showed that protein cages with different geometries could be designed by fusing a natural symmetrical protein building block and a coiled coil through an oligo-Gly spacer. Overall, this strategy provides a simple and easily implemented route for *de novo* designing protein cages.

1.5.3 Project Goals

My thesis project was the continuation of this work. First, I evaluated the utility of *de novo* designed coiled coils to use as off-the-shelf components for protein design by fusing a set of *de novo* designed coiled coils with different oligomerization states to GFP. This study demonstrated that coiled coil behavior could be subjected to unanticipated changes when other proteins are attached. The intended oligomerization states of some of these coiled coils could be restored by minor structural modifications. This work is discussed in chapter 2.

In chapter 3, I describe the elaboration of our octahedral protein cage design by fusing a large monomeric protein to the free end of the coiled coil. This work was conducted to increase the yield of octahedral cages and study the potential of

appending large functional groups to the free end of the coiled coil. The protein construct successfully assembled into a ~ 1.8 MDa protein cage with an octahedral core; fusing a large monomeric group did not disrupt the oligomeric state. In this design, maltose binding protein was used as the monomeric protein. This allowed purification of the protein by maltose affinity chromatography which increased the yield of protein by ~ 60-fold and gave ~ 60 mg of protein per 1 L culture.

In chapter 4, I describe work to design an icosahedral protein cage (the highest geometry) using our symmetry-directed design approach. In this work, seven constructs were evaluated by genetically fusing a pentameric coiled coil to the same trimeric building block through a oligo-Gly linker. A construct with 4-heptad coiled coil and a 8-residue linker successfully assembled into a hyperstable icosahedral protein cage under optimized purification protocols. Surprisingly, these assemblies captured short DNA strands, which appeared to be important for the homogeneity of the cage. In the absence of DNA the protein assembled into heterogeneous forms.

In chapter 5, I describe the potential of designing protein cages that assemble in a controlled manner. As there are environment-responsive coiled coils, both natural and *de novo*, fusing them to TriEst should allow design of cages in which the assembly can be controlled. I tested this hypothesis by fusing a metal-dependent trimeric coiled coil to TriEst and was able to successfully control the assembly of the cage.

In chapter 6, I summarize my thesis work and will discuss the future directions of the project.

1.6 References

1. Wang L-Y, Li J-H, Zhou X, Zheng Q-C, Cheng X (2017) Clinical application of carbon nanoparticles in curative resection for colorectal carcinoma. Onco Targets Ther 10:5585–5589.

2. Son KH, Hong JH, Lee JW (2016) Carbon nanotubes as cancer therapeutic carriers and mediators. Int J Nanomedicine 11:5163–5185.

3. Bianco A, Kostarelos K, Partidos CD, Prato M (2005) Biomedical applications of functionalised carbon nanotubes. Chem Commun 5:571-577.

4. Golchin K, Golchin J, Ghaderi S, Alidadiani N, Eslamkhah S, Eslamkhah M, Davaran S, Akbarzadeh A (2018) Gold nanoparticles applications: from artificial enzyme till drug delivery. Artif Cells Nanomed Biotechnol 46:250–254.

5. Chen J, Wiley B, Li Z-Y, Campbell D, Saeki F, Cang H, Au L, Lee J, Li X, Xia Y (2005) Gold Nanocages: Engineering Their Structure for Biomedical Applications. Adv Mater 17:2255–2261.

6. Abbasi E, Aval SF, Akbarzadeh A, Milani M, Nasrabadi HT, Joo SW, Hanifehpour Y, Nejati-Koshki K, Pashaei-Asl R (2014) Dendrimers: synthesis, applications, and properties. Nanoscale Res Lett 9:247.

7. Esfand R, Tomalia DA (2001) Poly(amidoamine) (PAMAM) dendrimers: from biomimicry to drug delivery and biomedical applications. Drug Discovery Today 6:427–436.

8. Alivisatos AP (1996) Semiconductor Clusters, Nanocrystals, and Quantum Dots. Science 271:933–937.

9. Michalet X (2005) Quantum Dots for Live Cells, in Vivo Imaging, and Diagnostics. Science 307:538–544.

10. H. R C, Schiffman JD, Balakrishna RG (2018) Quantum dots as fluorescent probes: Synthesis, surface chemistry, energy transfer mechanisms, and applications. Sens Actuators, B 258:1191–1214.

11. Wang F, Li C, Cheng J, Yuan Z (2016) Recent Advances on Inorganic Nanoparticle-Based Cancer Therapeutic Agents. Int J Environ Res Public Health 13: 1182.

12. Khan I, Saeed K, Khan I (2017) Nanoparticles: Properties, applications and toxicities. Arabian J Chem.

13. Walsh AS, Yin H, Erben CM, Wood MJA, Turberfield AJ (2011) DNA Cage Delivery to Mammalian Cells. ACS Nano 5:5427–5432.

14. Jiang Q, Song C, Nangreave J, Liu X, Lin L, Qiu D, Wang Z-G, Zou G, Liang X, Yan H, et al. (2012) DNA Origami as a Carrier for Circumvention of Drug Resistance. J Am Chem Soc 134:13396–13403.

15. Chandrasekaran AR (2016) DNA origami and biotechnology applications: a perspective: DNA origami and biotechnology applications. J Chem Technol Biot 91:843–846.

16. Guo P (2005) RNA Nanotechnology: Engineering, Assembly and Applications in Detection, Gene Delivery and Therapy. J Nanosci Nanotechnol 5:1964–1982.

17. Dobrovolskaia MA (2016) Self-assembled DNA/RNA nanoparticles as a new generation of therapeutic nucleic acids: immunological compatibility and other translational considerations. DNA and RNA Nanotechnol 3:1-10

18. Flenniken ML, Uchida M, Liepold LO, Kang S, Young MJ, Douglas T A (2009) Library of Protein Cage Architectures as Nanomaterials. Curr Opin Microbiol Immuol 327:71–93.

19. Lee EJ, Lee NK, Kim I-S (2016) Bioengineered protein-based nanocage for drug delivery. Adv Drug Delivery Rev 106:157–171.

20. Uchida M, Klem MT, Allen M, Suci P, Flenniken M, Gillitzer E, Varpness Z, Liepold LO, Young M, Douglas T (2007) Biological Containers: Protein Cages as Multifunctional Nanoplatforms. Adv Mater 19:1025–1042.

21. Kramer RM, Li C, Carter DC, Stone MO, Naik RR (2004) Engineered Protein Cages for Nanomaterial Synthesis. J Am Chem Soc 126:13282–13286.

22. Bhaskar S, Lim S (2017) Engineering protein nanocages as carriers for biomedical applications. NPG Asia Mater 9: e371.

23. Rohovie MJ, Nagasawa M, Swartz JR (2017) Virus-like particles: Next-generation nanoparticles for targeted therapeutic delivery. Bioeng Transl Med 2:43–57.

24. de la Escosura A, Nolte RJM, Cornelissen JJLM (2009) Viruses and protein cages as nanocontainers and nanoreactors. J Mater Chem 19:2274.

25. Aumiller WM, Uchida M, Douglas T (2018) Protein cage assembly across multiple length scales. Chem Soc Rev 47:3433–3469.

26. Diaz D, Care A, Sunna A (2018) Bioengineering Strategies for Protein-Based Nanoparticles. Genes 9:370.

27. Gradišar H, Jerala R (2014) Self-assembled bionanostructures: proteins following the lead of DNA nanostructures. J Nanobiotechnol 12:4.

28. Dominguez R, Holmes KC (2011) Actin Structure and Function. Annu Rev Biophys 40:169–186.

29. Abal M, Andreu JM, Barasoain I (2003) Taxanes: microtubule and centrosome targets, and cell cycle dependent mechanisms of action. Curr Cancer Drug Targets 3:193–203.

30. Vemu A, Atherton J, Spector JO, Moores CA, Roll-Mecak A (2017) Tubulin isoform composition tunes microtubule dynamics. Mol Biol Cell 28:3564–3572.

31. Landry SJ, Gierasch LM (1991) The chaperonin GroEL binds a polypeptide in an .alpha.-helical conformation. Biochemistry 30:7359–7362.

32. Hayer-Hartl M, Bracher A, Hartl FU (2016) The GroEL–GroES Chaperonin Machine: A Nano-Cage for Protein Folding. Trends Biochem Sci 41:62–76.

33. Cohen B, Dafni H, Meir G, Harmelin A, Neeman M (2005) Ferritin as an Endogenous MRI Reporter for Noninvasive Imaging of Gene Expression in C6 Glioma Tumors. Neoplasia 7:109–117.

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

35. Theil EC (1987) Ferritin: Structure, Gene Regulation, and Cellular Function in Animals, Plants, and Microorganisms. Annu Rev Biochem 56:289–315.

36. Linder M (2013) Mobilization of Stored Iron in Mammals: A Review. Nutrients 5:4022–4050.

37. Patel MS, Nemeria NS, Furey W, Jordan F (2014) The Pyruvate Dehydrogenase Complexes: Structure-based Function and Regulation. J Biol Chem 289:16615–16623.

38. Bosma HJ, Kok A, Markwijk BW, Veeger C (1984) The size of the pyruvate dehydrogenase complex of Azotobacter vinelandii. Association phenomena. Eur J Biochem 140:273–280.

39. Roos WH, Ivanovska IL, Evilevitch A, Wuite GJL (2007) Viral capsids: Mechanical characteristics, genome packaging and delivery mechanisms. Cell Mol Life Sci 64:1484–1497.

40. Manchester M, Steinmetz NF (2009) Viruses and nanotechnology. Springer.

41. Ehrlich LS, Liu T, Scarlata S, Chu B, Carter CA (2001) HIV-1 Capsid Protein Forms Spherical (Immature-Like) and Tubular (Mature-Like) Particles in Vitro: Structure Switching by pH-induced Conformational Changes. Biophys J 81:586–594.

42. Taylor DJ, Krishna NK, Canady MA, Schneemann A, Johnson JE (2002) Large-Scale, pH-Dependent, Quaternary Structure Changes in an RNA Virus Capsid Are Reversible in the Absence of Subunit Autoproteolysis. J Virol 76:9972–9980.

43. Tama F, Brooks CL (2002) The Mechanism and Pathway of pH Induced Swelling in Cowpea Chlorotic Mottle Virus. J Mol Biol 318:733–747.

44. Yamada K, Yoshii S, Kumagai S, Miura A, Uraoka Y, Fuyuki T, Yamashita I (2006) Floating Gate Metal–Oxide–Semiconductor Capacitor Employing Array of High-Density Nanodots Produced by Protein Supramolecule. Jpn J Appl Phys 45:8946–8951.

45. Kaba SA, Brando C, Guo Q, Mittelholzer C, Raman S, Tropel D, Aebi U, Burkhard P, Lanar DE (2009) A Nonadjuvanted Polypeptide Nanoparticle Vaccine Confers Long-Lasting Protection against Rodent Malaria. J Immunol 183:7268–7277.

46. Yildiz I, Shukla S, Steinmetz NF (2011) Applications of viral nanoparticles in medicine. Curr Opin Biotech 22:901–908.

47. Edwardson TGW, Mori T, Hilvert D (2018) Rational Engineering of a Designed Protein Cage for siRNA Delivery. J Am Chem Soc 140:10439–10442.

48. Comellas-Aragonès M, de la Escosura A, Dirks A (Ton) J, van der Ham A, Fusté-Cuñé A, Cornelissen JJLM, Nolte RJM (2009) Controlled Integration of Polymers into Viral Capsids. Biomacromolecules 10:3141–3147.

49. Yildiz I, Lee KL, Chen K, Shukla S, Steinmetz NF (2013) Infusion of imaging and therapeutic molecules into the plant virus-based carrier cowpea mosaic virus: Cargo-loading and delivery. J Controlled Release 172:568–578.

50. Dhason MS, Wang JC-Y, Hagan MF, Zlotnick A (2012) Differential assembly of Hepatitis B Virus core protein on single- and double-stranded nucleic acid suggest the dsDNA-filled core is spring-loaded. Virology 430:20–29.

51. Strods A, Ose V, Bogans J, Cielens I, Kalnins G, Radovica I, Kazaks A, Pumpens P, Renhofa R (2015) Preparation by alkaline treatment and detailed characterisation of empty hepatitis B virus core particles for vaccine and gene therapy applications. Sci Rep 5: 11639.

52. Douglas T, Young M (1998) Host–guest encapsulation of materials by assembled virus protein cages. Nature 393:152-155.

53. Timmermans SBPE, van Hest JCM (2018) Self-assembled nanoreactors based on peptides and proteins. Curr Opin Colloid In 35:26–35.

54. Schoonen L, Nolte RJM, van Hest JCM (2016) Highly efficient enzyme encapsulation in a protein nanocage: towards enzyme catalysis in a cellular nanocompartment mimic. Nanoscale 8:14467–14472.

55. Minten IJ, Hendriks LJA, Nolte RJM, Cornelissen JJLM (2009) Controlled Encapsulation of Multiple Proteins in Virus Capsids. J Am Chem Soc 131:17771–17773.

56. Patterson DP, Schwarz B, Waters RS, Gedeon T, Douglas T (2014) Encapsulation of an Enzyme Cascade within the Bacteriophage P22 Virus-Like Particle. ACS Chem Biol 9:359–365.

57. Jordan PC, Patterson DP, Saboda KN, Edwards EJ, Miettinen HM, Basu G, Thielges MC, Douglas T (2016) Self-assembling biomolecular catalysts for hydrogen production. Nat Chem 8:179–185.

58. Butterfield GL, Lajoie MJ, Gustafson HH, Sellers DL, Nattermann U, Ellis D, Bale JB, Ke S, Lenz GH, Yehdego A, et al. (2017) Evolution of a designed protein assembly encapsulating its own RNA genome. Nature 552:415–420.

59. Terasaka N, Azuma Y, Hilvert D (2018) Laboratory evolution of virus-like nucleocapsids from nonviral protein cages. Proc Natl Acad Sci USA 115:5432–5437.

60. Qazi S, Miettinen HM, Wilkinson RA, McCoy K, Douglas T, Wiedenheft B (2016) Programmed Self-Assembly of an Active P22-Cas9 Nanocarrier System. Mol Pharmaceut 13:1191–1196.

61. Storni T, Ruedl C, Schwarz K, Schwendener RA, Renner WA, Bachmann MF (2004) Nonmethylated CG motifs packaged into virus-like particles induce protective cytotoxic T cell responses in the absence of systemic side effects. J. Immunol. 172:1777–1785.

62. Prel A, Caval V, Gayon R, Ravassard P, Duthoit C, Payen E, Maouche-Chretien L, Creneguy A, Nguyen TH, Martin N, et al. (2015) Highly efficient in vitro and in vivo delivery of functional RNAs using new versatile MS2-chimeric retrovirus-like particles. Mol Ther Methods 2:15039.

63. Galaway FA, Stockley PG (2013) MS2 Viruslike Particles: A Robust, Semisynthetic Targeted Drug Delivery Platform. Mol Pharmaceut 10:59–68.

64. Klem MT, Willits D, Solis DJ, Belcher AM, Young M, Douglas T (2005) Bio-inspired Synthesis of Protein-Encapsulated CoPt Nanoparticles. Adv Funct Mater 15:1489–1494.

65. Bhattacharya P, Du D, Lin Y (2014) Bioinspired nanoscale materials for biomedical and energy applications. J R Soc Interface 11:20131067–20131067.

66. Allen M, Willits D, Mosolf J, Young M, Douglas T (2002) Protein Cage Constrained Synthesis of Ferrimagnetic Iron Oxide Nanoparticles. Adv Mater 14:1562–1565.

67. Petryk AA, Giustini AJ, Ryan P, Strawbridge RR, Hoopes PJ (2009) Iron oxide nanoparticle hyperthermia and chemotherapy cancer treatment. Proc SPIE Int Soc opt Eng 7181:71810N.

68. Hikono T, Matsumura T, Miura A, Uraoka Y, Fuyuki T, Takeguchi M, Yoshii S, Yamashita I (2006) Electron confinement in a metal nanodot monolayer embedded in silicon dioxide produced using ferritin protein. Appl Phys Lett 88:023108.

69. Heddle JG, Chakraborti S, Iwasaki K (2017) Natural and artificial protein cages: design, structure and therapeutic applications. Curr Opin Struct Biol 43:148–155.

70. Mohsen MO, Zha L, Cabral-Miranda G, Bachmann MF (2017) Major findings and recent advances in virus–like particle (VLP)-based vaccines. Semin Immunol 34:123–132.

71. Aljabali AAA, Shukla S, Lomonossoff GP, Steinmetz NF, Evans DJ (2013) CPMV-DOX Delivers. Mol Pharmaceut 10:3–10.

72. Lohcharoenkal W, Wang L, Chen YC, Rojanasakul Y (2014) Protein Nanoparticles as Drug Delivery Carriers for Cancer Therapy. Biomed Res Int 2014:1–12.

73. Rami A, Behdani M, Yardehnavi N, Habibi-Anbouhi M, Kazemi-Lomedasht F (2017) An overview on application of phage display technique in immunological studies. Asian Pac J Trop Biomed 7:599–602.

74. Mao C, Flynn CE, Hayhurst A, Sweeney R, Qi J, Georgiou G, Iverson B, Belcher AM (2003) Viral assembly of oriented quantum dot nanowires. Proc Natl Acad Sci USA 100:6946–6951.

75. Klem MT, Willits D, Young M, Douglas T (2003) 2-D Array Formation of Genetically Engineered Viral Cages on Au Surfaces and Imaging by Atomic Force Microscopy. J Am Chem Soc 125:10806–10807.

76. Sontz PA, Bailey JB, Ahn S, Tezcan FA (2015) A Metal Organic Framework with Spherical Protein Nodes: Rational Chemical Design of 3D Protein Crystals. J Am Chem Soc 137:11598–11601.

77. Uchida M, McCoy K, Fukuto M, Yang L, Yoshimura H, Miettinen HM, LaFrance B, Patterson DP, Schwarz B, Karty JA, et al. (2018) Modular Self-Assembly of Protein Cage Lattices for Multistep Catalysis. ACS Nano 12:942–953.

78. Molino NM, Wang S-W (2014) Caged protein nanoparticles for drug delivery. Curr Opin Biotechnol 28:75–82.

79. Choi S-H, Choi K, Chan Kwon I, Ahn HJ (2010) The incorporation of GALA peptide into a protein cage for an acid-inducible molecular switch. Biomaterials 31:5191–5198.

80. Peng T, Lim S (2011) Trimer-Based Design of pH-Responsive Protein Cage Results in Soluble Disassembled Structures. Biomacromolecules 12:3131–3138.

81. Padilla JE, Colovos C, Yeates TO (2001) Nanohedra: Using symmetry to design self assembling protein cages, layers, crystals, and filaments. Proc Natl Acad Sci USA 98:2217–2221.

82. Lai Y-T, Cascio D, Yeates TO (2012) Structure of a 16-nm Cage Designed by Using Protein Oligomers. Science 336:1129–1129.

83. Lai Y-T, King NP, Yeates TO (2012) Principles for designing ordered protein assemblies. Trends Cell Biol 22:653–661.

84. Lai Y-T, Reading E, Hura GL, Tsai K-L, Laganowsky A, Asturias FJ, Tainer JA, Robinson CV, Yeates TO (2014) Structure of a designed protein cage that self-assembles into a highly porous cube. Nat Chem 6:1065–1071.

85. Jorda J, Leibly DJ, Thompson MC, Yeates TO (2016) Structure of a novel 13 nm dodecahedral nanocage assembled from a redesigned bacterial microcompartment shell protein. Chem Commun 52:5041–5044.

86. King NP, Bale JB, Sheffler W, McNamara DE, Gonen S, Gonen T, Yeates TO, Baker D (2014) Accurate design of co-assembling multi-component protein nanomaterials. Nature 510:103–108.

87. King NP, Sheffler W, Sawaya MR, Vollmar BS, Sumida JP, Andre I, Gonen T, Yeates TO, Baker D (2012) Computational Design of Self-Assembling Protein Nanomaterials with Atomic Level Accuracy. Science 336:1171–1174.

88. Hsia Y, Bale JB, Gonen S, Shi D, Sheffler W, Fong KK, Nattermann U, Xu C, Huang P-S, Ravichandran R, et al. (2016) Design of a hyperstable 60-subunit protein icosahedron. Nature 535:136–139.

89. Bale JB, Gonen S, Liu Y, Sheffler W, Ellis D, Thomas C, Cascio D, Yeates TO, Gonen T, King NP, et al. (2016) Accurate design of megadalton-scale two-component icosahedral protein complexes. Science 353:389–394.

90. Raman S, Machaidze G, Lustig A, Aebi U, Burkhard P (2006) Structure-based design of peptides that self-assemble into regular polyhedral nanoparticles. Nanomed-Nanotechnol 2:95–102.

91. Gradišar H, Božič S, Doles T, Vengust D, Hafner-Bratkovič I, Mertelj A, Webb B, Šali A, Klavžar S, Jerala R (2013) Design of a single-chain polypeptide tetrahedron assembled from coiled-coil segments. Nat Chem Biol 9:362–366.

92. Ljubetič A, Lapenta F, Gradišar H, Drobnak I, Aupič J, Strmšek Ž, Lainšček D, Hafner-Bratkovič I, Majerle A, Krivec N, et al. (2017) Design of coiled-coil protein-origami cages that self-assemble in vitro and in vivo. Nat Biotechnol 35:1094.

93. Lapenta F, Aupič J, Strmšek Ž, Jerala R (2018) Coiled coil protein origami: from modular design principles towards biotechnological applications. Chem Soc Rev 47:3530–3542.

94. Lanci CJ, MacDermaid CM, Kang S -g., Acharya R, North B, Yang X, Qiu XJ, DeGrado WF, Saven JG (2012) Computational design of a protein crystal. Proc Natl Acad Sci USA 109:7304–7309.

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

96. Indelicato G, Wahome N, Ringler P, Müller SA, Nieh M-P, Burkhard P, Twarock R (2016) Principles Governing the Self-Assembly of Coiled-Coil Protein Nanoparticles. Biophys J 110:646–660.

97. Kobayashi N, Yanase K, Sato T, Unzai S, Hecht MH, Arai R (2015) Self-Assembling Nano-Architectures Created from a Protein Nano-Building Block Using an Intermolecularly Folded Dimeric *de Novo* Protein. J Am Chem Soc 137:11285–11293.

98. Patterson DP, Su M, Franzmann TM, Sciore A, Skiniotis G, Marsh ENG (2014) Characterization of a highly flexible self-assembling protein system designed to form nanocages: Highly Flexible Self-Assembling Protein System. Protein Sci 23:190–199.

99. Sciore A, Su M, Koldewey P, Eschweiler JD, Diffley KA, Linhares BM, Ruotolo BT, Bardwell JCA, Skiniotis G, Marsh ENG (2016) Flexible, symmetry-directed approach to assembling protein cages. Proc Natl Acad Sci USA 113:8681–8686.

100. Badieyan S, Sciore A, Eschweiler JD, Koldewey P, Cristie-David AS, Ruotolo BT, Bardwell JCA, Su M, Marsh ENG (2017) Symmetry-Directed Self-Assembly of a Tetrahedral Protein Cage Mediated by de Novo-Designed Coiled Coils. ChemBioChem 19: 1888-1892.

Chapter 2: Evaluation of *De Novo* Designed Coiled Coils as off-the-shelf Components for Protein Assembly

*This work is published on "**Cristie-David AS**, Sciore A, Badieyan S, Escheweiler JD, Koldewey P, Bardwell JCA, Ruotolo BT, Marsh ENG (2017) Evaluation of de novo-designed coiled coils as off-the-shelf components for protein assembly. Mol Syst Des Eng 2:140–148"*

2.1 Introduction

α-helical coiled coil domains represent one of the simplest and best understood protein-protein interactions. They occur widely in natural proteins, comprising up to 5% of all protein residues, and were among the first protein structures to be designed *de novo*.[1–3] Simple rules, based on a repeating canonical heptad of amino acid residues, apparently govern the interactions between the α-helices of coiled coil domains. The use of these rules has allowed the *de novo* design of a wide range of small α-helical bundle proteins whose structures comprise between 2 and 7 α-helices (Fig. 2.1).[4–7] The helices may be designed to adopt either parallel or antiparallel topologies and may be identical (homooligomeric) or complementary (heterooligomeric) in sequence.[8–10]

In nature, coiled coil interactions often mediate the assembly of larger protein domains. Well-studied examples include the dimerization of many transcription factors and the assembly of multi-enzyme complexes such as pyruvate dehydrogenase and polyketide synthases.[11–13] Coiled coils are also important in the

interaction of proteins with membranes.[14–17] The ability of coiled coils to mediate the controlled assembly of large-scale protein structures has attracted the interest of synthetic biologists. Coiled coil domains can also be designed *de novo* with well-defined oligomerization states, topologies, and dissociation energies. Hence coiled coils have become very attractive modular protein motifs for *de novo* assembling individual protein subunits into higher order structures in synthetic biology. Both natural and *de novo* designed coiled coil motifs have been widely utilized to *de novo* design protein cages, extended protein structures, hydrogels and various protein based nanoparticles.[18–23]

Our lab identified the potential of using *de novo* designed parallel homo-oligomeric coiled coils to function as simple off-the-shelf connectors to drive homogeneous protein cage assemblies when they are fused to large oligomeric proteins with rotational symmetries. The utility of coiled coil designs as plug-and-play components for such synthetic biology applications depends critically on them robustly maintaining their oligomerization states when fused to larger proteins of interest. However, the properties of the *de novo* designed coiled coils have been almost exclusively investigated in isolation, leaving open the possibility that they may significantly differ in their behavior when genetically fused to much larger, natural proteins.

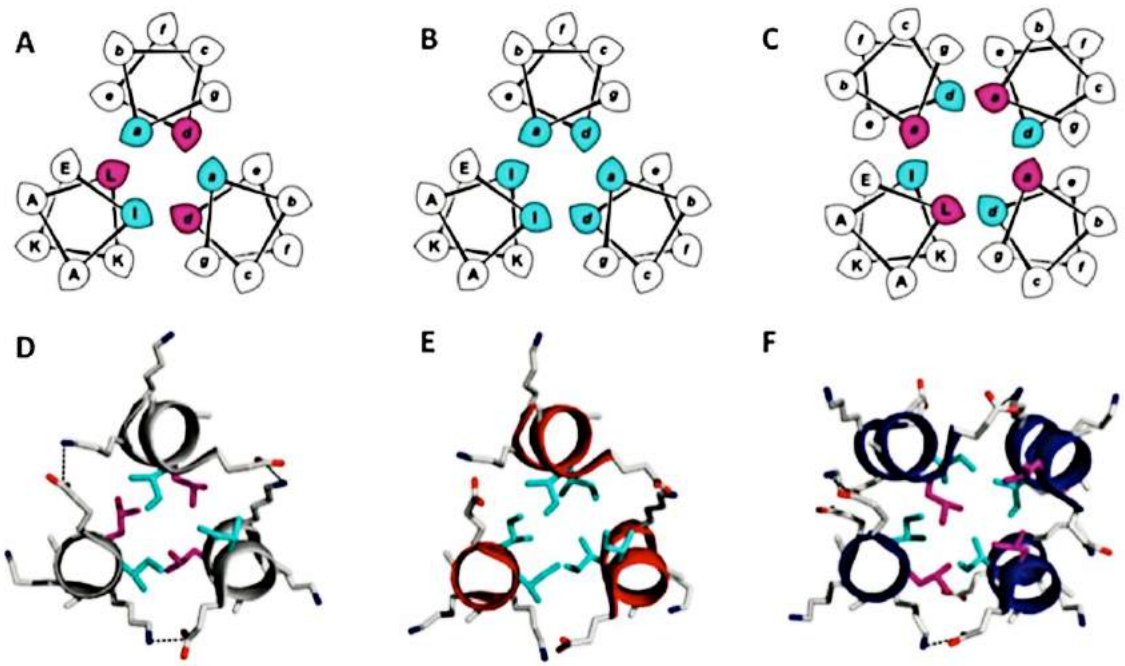

Figure 2. 1 Structures formed by several *de novo* designed coiled coils, which we evaluated by fusing to GFP. (A-C) helical diagrams and (D-F) crystal structures of 4DZN, 4DZL and 3R4A coiled coils respectively. In these structures "a" and "d" residues facing inward make the hydrophobic core region. The mode of arrangement of hydrophobic side chains (mainly Leu and Ile residues) in the core region in "knob into hole" fashion is a main determinant of the oligomeric state of coiled coils. Salt bridges formed between "e" and "g" residues are also important for the stabilization of the coiled coils. (Image adapted from ref. 26)

A very robust set of coiled coils (dimeric, trimeric, tetrameric and pentameric) is required as plug-and-play components for our protein cage design strategy. Therefore, as the first aim of my thesis, I, with my colleagues Dr. Aaron Sciore and Dr. Somaye Badieyan, studied the behavior of coiled coils of interest when attached to a larger monomeric protein domain. The objective was to ensure that these coiled coil designs maintained their intended oligomerization states when used to assemble other proteins. We genetically fused various coiled coils to a model monomeric protein, Green fluorescent protein (GFP) and evaluated the oligomeric state of these fusion constructs using several analytical tools (size exclusion

chromatography, analytical ultracentrifugation, and native mass spectrometry). GFP was chosen as the fusion control because it is a robust, well behaved monomeric protein and its fluorescent properties provide an easy read out for protein expression and folding.

Six coiled coil GFP fusion proteins were initially constructed. Somewhat surprisingly, only two of these initial designs adopted their intended oligomerization states. However, with minor refinements, the intended oligomerization states of two of the four other constructs could be achieved. Parameters found to influence the oligomerization state of the GFP fusions included the number of heptad repeats and the length of the linker sequence separating GFP from the coiled coil. These results demonstrated that even for stable, well-designed coiled coils, the oligomerization state is subject to unanticipated changes when connected to larger protein components. Overall, this study clearly showed that it will be necessary to both carefully validate and optimize the coiled coil interaction to successfully apply coiled coils in the construction of large-scale protein assemblies.

2.2 Materials and Methods

2.2.1 Construction of Genes Encoding Fusion Proteins

Codon-optimized genes encoding the various coiled coil designs and oligo-Gly spacer units were commercially synthesized and introduced into the expression vector pMCSG18 that encodes GFP as a reporter gene, either 3' or 5' to the GFP gene. The complete sequence of each of the fusion proteins is given in appendix A.1.

2.2.2 Protein Expression and Purification

Expression constructs were transformed into *E. coli* BL21(DE3) cells. Cells were grown in 2xYT medium with 100 mg/L ampicillin at 37 °C. At an OD_{600} of 0.8, the temperature was reduced to 18 °C, and at an OD_{600} of 1.0, protein expression was induced by addition of 0.1 mM IPTG. Cells were grown for a further 18 h and harvested by centrifugation.

All purification steps were performed on ice or at 4 °C. Cell pellets were resuspended in 50 mM HEPES buffer, pH 7.5, containing 1 M urea, 300 mM NaCl, 25 mM imidazole, 5% glycerol, SigmaFAST protease inhibitor, and 1 mg/mL lysozyme, then lysed by sonication. The lysate was clarified by centrifugation at 48,000 *g* for 30 min and injected onto a HisTrap Ni-NTA column, washed with several volumes of the same buffer, and eluted with 50 mM HEPES buffer, pH 7.5, containing 300 mM NaCl, 500 mM imidazole, and 5% glycerol. Fractions containing GFP were pooled and dialyzed against 25 mM HEPES buffer, pH 7.5, containing 100 mM NaCl and 2 mM EDTA. The protein was then concentrated using 30-kDa Amicon ultra-centrifugal filter units and further purified by SEC on a Superdex 200 10/300 column equilibrated in the same buffer. Fractions containing proteins of the desired oligomerization state were pooled and further concentrated for analysis.

2.2.3 Size Exclusion Chromatography (SEC)

100 μL of protein sample at a concentration of 1 mg/mL was injected onto a Superdex 200 10/300 column equilibrated at 4 °C in dialysis buffer described above, and proteins were eluted at 0.4 mL/min.

2.2.4 Analytical Ultracentrifugation (AUC)

Sedimentation velocity analysis was performed using a Beckman Proteome Lab XL-I analytical ultracentrifuge (Beckman Coulter, Indianapolis, IN) equipped with an AN60TI rotor. Samples were dialyzed against 25 mM HEPES buffer, pH 7.5, containing 100 mM NaCl and 1 mM EDTA. The hydrodynamic behavior of the various proteins was analyzed at a protein concentration of ~ 0.2 mg/mL (A_{280} = 0.2). Samples were loaded into pre-cooled cells containing standard sector-shaped 2-channel Epon centerpieces with 1.2 cm path-length (Beckman Coulter, Indianapolis, IN) and allowed to equilibrate at 6 °C for 2 h in the non-spinning rotor prior to sedimentation. Proteins were sedimented at 40,000 rpm and the sedimentation of the protein constructs was monitored continuously at a wavelength of 280 nm. Sedimentation velocity data was analyzed with the program SEDFIT (version 15.01b).[24] Sedimentation distribution plots were generated using the continuous c(s) distribution model, with a confidence level for the ME (Maximum Entropy) regularization of 0.7. Molecular weights (M_w) including their standard deviation were calculated by integration of a c(M) continuous distribution. For that, the ME (Maximum Entropy) regularization was set to 0.5. Buffer density and viscosity were calculated using SEDNTERP (http://sednterp.unh.edu/).

2.2.5 Native Mass Spectrometry (Native MS)

Samples prepared for mass spectrometry were purified as described above and concentrated to 40 µL. The minimum concentration of protein required for analysis is 1 µM, with higher concentrations preferred. Samples were then loaded into gold plated needles prepared in house.[25] Nano-electrospray-ion-mobility-TOF

mass spectrometry was performed using a Synapt G2 Traveling-Wave instrument (Waters Corp, Manchester, U.K.). Ions were generated by applying a voltage of 1.5 kV between the needle and the instrument source, with further voltage drops aiding in acceleration and desolvation as ions passed through the skimmer region of the instrument. The quadrupole region was set to RF-only mode for collection of complete mass spectra, and in some cases was tuned to isolate selected peaks for MS/MS analysis. A range of collision energies was tested for enhanced transmission and desolvation of the ions, and in some cases dissociation of the ion into its component subunits. The base values for collision energies were 20-50 V; however, energies up to 150 V were utilized for dissociation experiments. The IMS region of the instrument was operated at 4 mBar of nitrogen, with wave heights and wave velocities of 15 V and 150 m/s, respectively. The instrument time of flight mass analyzer was operated in sensitivity mode, and mass spectra were collected from 1000 to 15000 m/z. Data analysis was performed using the manufacturer-provided Masslynx software.

2.3 Results

2.3.1 Selection of Coiled Coils for Evaluation

For our studies, we selected a series of crystallographically characterized four-heptad parallel coiled coils designed by Woolfson and co-workers.[26–28] These coiled coils span the range of oligomeric states from dimer to pentamer, making them potentially useful for the assembly of a wide variety of protein structures. Four of the coiled coils were based on the well-established canonical heptad repeat

sequence $(XAAYKZE)_4$ in which the oligomerization state is primarily determined by the identity of the hydrophobic residues X and Y at the canonical 'a' and 'd' positions. The dimeric coiled coil (Protein Data Bank (PDB) ID 4DZM) employs Ile at the 'a' and Leu at the 'd' positions, with the exception of Asn at the third 'a' position. Two trimeric coiled coils were investigated: one (PDB ID 4DZL) employs Ile at the 'a' and 'd' positions, whereas the other (PDB ID 4DZN) employs Ile at the 'a' and Leu at the 'd' positions (Fig. 2.1 A,B,D,E). The tetrameric coiled coil (PDB ID 3R4A) employs Leu at the 'a' and Ile at the 'd' positions (Fig. 2.1 C,F). The residue 'Z' at the solvent-exposed 'f' position is Gln, Trp or Lys depending upon the coiled coil. The larger size of the pentameric coiled coil (PDB ID 4PN8) results in a solvent-exposed channel at the center of the helical bundle. In this case, the hydrophobic interactions between the helices occur at the 'a', 'd', and 'g' positions, which are occupied by Leu, Ile and Ile respectively. The sequences and oligomerization states of these coiled coils are summarized in Table 2.1.

Table 2. 1 PDB ID, sequences, and oligomerization states of the basis set of coiled coils used in this study. The hydrophobic residues that primarily dictate the oligomerization state are indicated in bold type.

PDB ID	abcdefg	abcdefg	abcdefg	abcdefg ab	Oligomeric state
4DZM	**I**AA**L**KQE	**I**AA**L**KQE	**I**AA**N**KQE	**I**AA**L**KQE	Dimer
4DZN	**I**AA**L**KQE	**I**AA**L**KQE	**I**AA**L**KQE	**I**AA**L**KQE	Trimer
4DZL	**I**AA**I**KQE	**I**AA**I**KQE	**I**AA**I**KQE	**I**AA**I**KQE	Trimer
3R4A	**L**AA**I**KQE	**L**AA**I**KQE	**L**AA**I**KQE	**L**AA**I**KQE	Tetramer
4PN8	K**I**EK**I**	**L**QK**I**EW**I**	**L**QK**I**EQ**I**	**L**QK**I**EQ**I** **L**Q	Pentamer

2.3.2 Fusion of Coiled Coils with GFP

We chose GFP as a model protein domain with which to test the coiled coils ability to mediate protein oligomerization. In addition to GFP being a very well characterized, stable, monomeric protein, its fluorescent property makes it easy to assess whether the addition of the coiled coil domain may have interfered with protein folding. We note that the GFP variant encoded in the PMCSG18 vector we used in these studies has a known tendency to dimerize at high concentrations, but at the lower concentrations used in these studies no dimerization was discernable.

To construct fusion proteins in which the coiled coil sequence was added to either the N- or C-terminus of GFP we used the standard molecular biology techniques described in the experimental section. An oligo-Gly spacer was introduced between the GFP and coiled coil domains to alleviate steric constraints that might cause either domain to misfold (Fig. 2.2). Based on modeling that assumed a hydrodynamic radius for GFP of 2.4 nm, we initially set the length of this spacer to be six-Gly residues, which is sufficient to span a distance of ~ 1.8 nm. This should allow the coiled coil domain to oligomerize without introducing steric clashes between the appended GFP domains. As discussed later, in some designs, the length of the glycine spacer was increased or the coiled coil sequence modified from the initial design. A description of each construct studied is given in Table 2.2. The fusion proteins were over-expressed in *Escherichia coli* BL21 and purified by Ni-NTA affinity chromatography in good yields without difficulty. All constructs exhibited the characteristic fluorescent green color of GFP, indicating that the proteins were correctly folded.

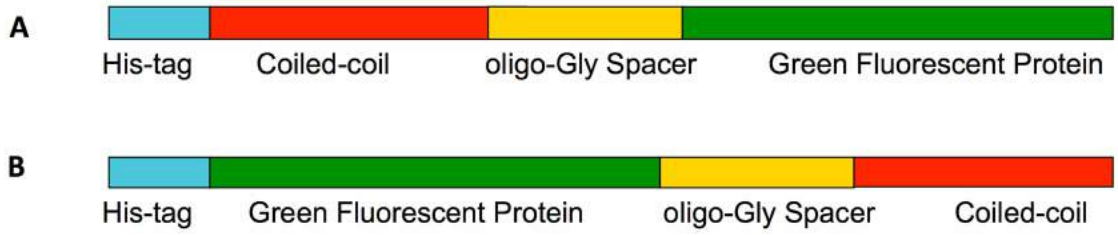

Figure 2. 2 Diagram showing topology of the GFP coiled coil fusion proteins for coiled coils fused to the C-terminus of GFP (A) and to the N-terminus of GFP (B).

Table 2. 2 Description of the GFP coiled coil fusion proteins evaluated in this study**.**

Construct number	Position of coiled coil[a]	Spacer length[b]	Coiled coil sequence[c]	Intended oligomerization state[d]
1	C terminus	6 Gly	IAALKQE IAALKQE IAANKQE IAALKQE	Dimer
2	C terminus	6 Gly	IAALKQE IAALKQE IAALKQE IAALKQE	Trimer
3	C terminus	6 Gly	IAAIKQE IAAIKQE IAAIKQE IAAIKQE	Trimer
4	C terminus	6 Gly	LAAIKQE LAAIKQE LAAIKQE LAAIKQE	Tetramer
5	C terminus	6 Gly	KIEQILQ KIEKILQ KIEWILQ KIEQILQ	Pentamer
6	C terminus	9 Gly	KIEQILQ KIEKILQ KIEWILQ KIEQILQ	Pentamer
7	C terminus	9 Gly	KIEQILQ KIEKILQ KIEQILQ KIEQILQ	Pentamer
8	N terminus	6 Gly	IAAIKQE IAAIKQE IAAIKQE IAAIKQE	Trimer
9	N terminus	6 Gly	IAAIKQE IAAIKQE IAAIKQE IAAIKQE IAAIKQE	Trimer
10	C terminus	6 Gly	LAAIKQE LAAIKQE LAAIKQE LAAIKQE LAAIKQE	Tetramer
11	C terminus	9 Gly	LAAIKQE LAAIKQE LAAIKQE LAAIKQE	Tetramer

[a] Position of the coiled coil relative to GFP.
[b] Number of Gly residues separating the GFP and coiled coil domains.
[c] Constructs **9** and **10** contain coiled coils with 5 repeating heptads.
[d] Based on the crystal structure of the isolated coiled coil.

2.3.3 Evaluation of GFP-Coiled Coil Oligomerization States

To determine whether the oligomerization state of the original coiled coil was maintained in the GFP fusion constructs, we examined the oligomerization states of the various GFP constructs using three complementary techniques: size exclusion chromatography (SEC), sedimentation velocity analytical ultracentrifugation (AUC), and non-denaturing native mass spectrometry (native

MS). Analytical ultracentrifugation allows the molecular mass of macromolecules in solution, as well as their size distribution, to be accurately determined. SEC does not directly measure molecular weights, but is a simple and rapid solution phase method that can report on sample homogeneity and provide an estimate of oligomerization state. In native MS, samples are ionized under very mild conditions so that non-covalent protein-protein interactions are maintained; native MS can thus provide high-resolution measurements of protein masses that allow their oligomerization states to be deduced, with the caveat that these gas phase measurements may be biased by differences in the ease with which various oligomeric species can be ionized.

Characterization of monomeric wild-type GFP (as a control) and fusion proteins **1** (dimer) and **3** (trimer) by these techniques yielded consistent data indicating that each construct was fairly homogeneous and adopted its intended oligomerization state (Fig. 2.3 A,B,D; Table 2.3). However, fusion proteins **2** (expected trimer) and **4** (expected tetramer) (Fig. 2.3 C,E) appeared to be a dimer and a trimer, respectively, by each of the techniques used to assess their oligomerization states, even though the coiled coil domains had been demonstrated by crystallography to be trimeric and tetrameric, respectively. This result may be less surprising for fusion protein **2**, as there are several structures of coiled coil dimers in which the hydrophobic interior is packed with Ile at 'a' and Leu at 'd' positions. However, it was quite unexpected for fusion protein **4**, for which the majority of synthetic coiled coil structures with Leu at 'a' and Ile at 'd' positions are tetrameric.

These results demonstrate that the addition of a large protein domain to the coiled coil has the potential to alter its oligomerization state. This observation is consistent with the previously described lability of some parallel coiled coil designs in which the oligomerization state is sensitive to fairly subtle changes in hydrophobic core packing.[5] Interestingly, in both cases in which the oligomerization state departs from what is expected, the fusion protein adopts a lower oligomerization state. This suggests that unfavorable steric interactions between the GFP domains (despite the introduction of a flexible spacer sequence) may cause reorganization of the coiled coil structure.

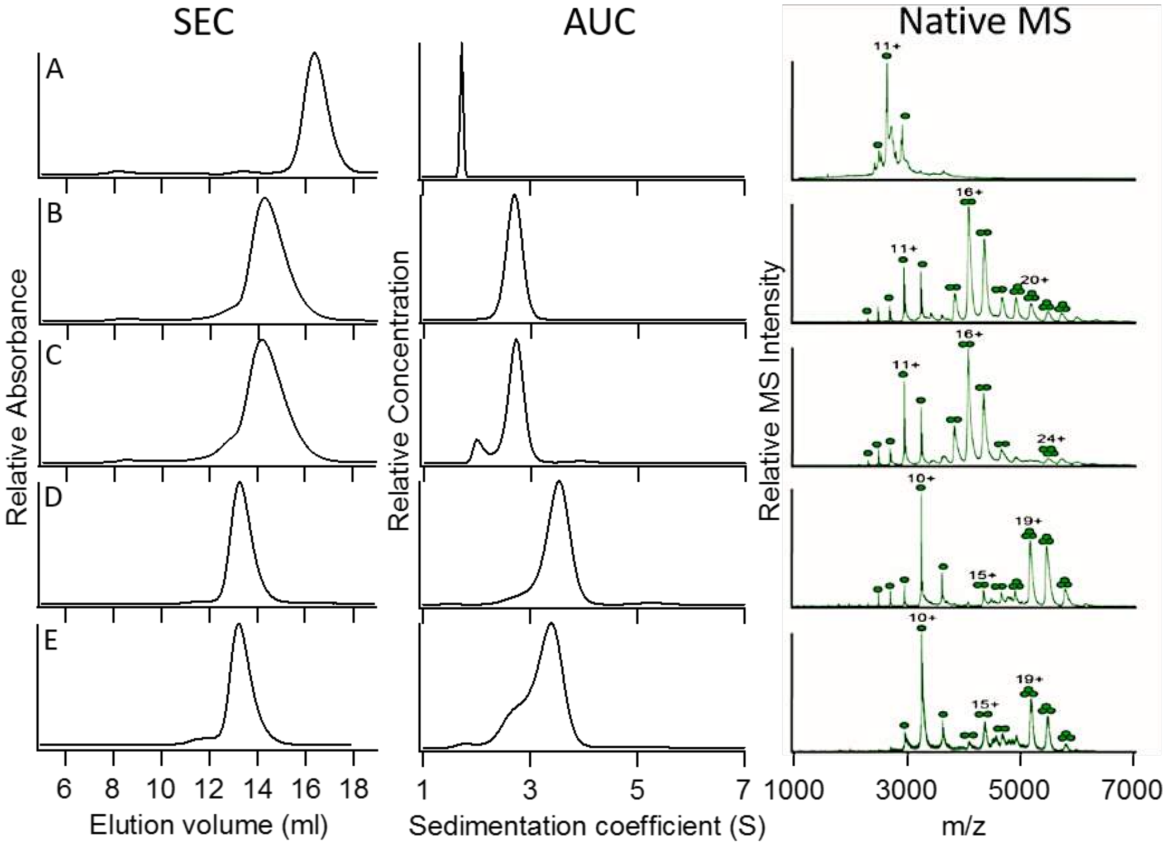

Figure 2. 3 Characterization of the oligomerization state of C-terminal GFP constructs **1**–**4**. (A) Wild-type GFP. (B) Construct **1** (intended dimer). (C) Construct **2** (intended trimer).

Table 2. 3 Summary of data describing the oligomeric properties of the various GFP constructs characterized in this study

GFP fusion construct	Elution volume (mL)	[a]Sedimentation coefficient, (S)	[b]Frictional ratio (f/f_0)	M_r (kDa) Calc. from s and f/f_0	[c]M_r (Da) Native MS	[d]Coiled coil oligomerization state	[e]Observed oligomerization state
WT GFP	16.3	1.73	1.31	30.3 ± 0.4	29053 ± 5	Monomer	Monomer
1	14.3	2.68	1.4	65.0 ± 4.4	64940 ± 80	Dimer	Dimer
2	14.1	2.71	1.37	63.7 ± 0.6	65130 ± 60	Trimer	Dimer
3	13.3	3.53	1.43	101.6 ± 5.9	97743 ± 30	Trimer	Trimer
4	13.2	3.39	1.39	91.4 ± 6.8	98000 ± 120	Tetramer	Trimer
5	8.2	N/D	N/D	N/D	N/D	Pentamer	Aggregate
6	11.9, 10.7, 8.1	4.70	[f]N/D	N/D	166300 ± 80	Pentamer	Pentamer-aggregate
7	12.1	4.65	1.39	144.3 ± 30	153500 ± 160	Pentamer	Pentamer
8	15.5	1.76	1.99	31.7 ± 0.5	32550 ± 6	Trimer	Monomer
9	12.7	3.29	1.52	96.9 ± 14.1	104530 ± 43	Trimer	Trimer
10	13.3	3.44	1.42	97.2 ± 9.0	98400 ± 120	Tetramer	Trimer
11	13.4	3.36	1.41	88.0 ± 0.9	100260 ± 70	Tetramer	Trimer

[a] In samples for which more than one species is present, the sedimentation coefficient of the major species is reported.
[b] Average frictional coefficient measured over all sedimenting species.
[c] The molecular masses determined by native MS include varying numbers of non-specifically bound ions that derive from the buffer. Therefore, the molecular masses are not exact multiples of the protein M_w derived from the sequence.
[d] Oligomerization state determined from the crystal structure.
[e] Consensus from the three experimental methods used to examine the oligomerization state of the fusion proteins.
[f] The presence of aggregated protein in the sample prevented the frictional ratio from being reliably determined from the data.

We next examined the properties of a GFP fusion protein containing a pentameric coiled coil, construct **5**. Although the protein was expressed as soluble, fluorescent protein (indicating that the GFP domain was correctly folded), SEC of **5** yielded a single peak in the void volume (Fig. 2.4 A). This suggested that the protein was forming large aggregates that could not be further characterized by AUC or native MS. In this case, it seemed that a six-Gly spacer might be too short to permit the simultaneous proper folding of the GFP domain and the pentameric coiled coil domain. We therefore increased the length of the spacer to nine-Gly residues,

resulting in construct **6**. This construct showed significantly less tendency to aggregate, but characterization by SEC indicated that it still formed a mixture of oligomeric species (Fig. 2.4 B). Characterization of **6** by AUC allowed the sedimentation coefficient for the major species to be determined as 4.7 S; however, the presence of aggregated proteins prevented the frictional ratio from being reliably determined so that the molecular weight could not be calculated. Characterization of **6** by native MS provided good evidence for the formation of the intended pentameric species, although this technique would not detect high molecular weight aggregates. Further mutation of a Trp residue at a solvent-exposed 'f' position (initially introduced to facilitate spectrophotometric quantification of the synthetic coiled coil) to a more hydrophilic Asn residue (construct **7**) resulted in a more monodisperse protein species that appeared to adopt the intended pentameric oligomerization state as judged by SEC, AUC, and native MS (Fig. 2.4 C). Although the 'f' position of the heptad repeat is solvent-exposed and therefore does not influence inter-helix interactions, this observation indicates that interactions between exterior residues of the coiled coil domain and the appended protein domain (in this case GFP) may need to be considered and optimized for the protein to assemble correctly.

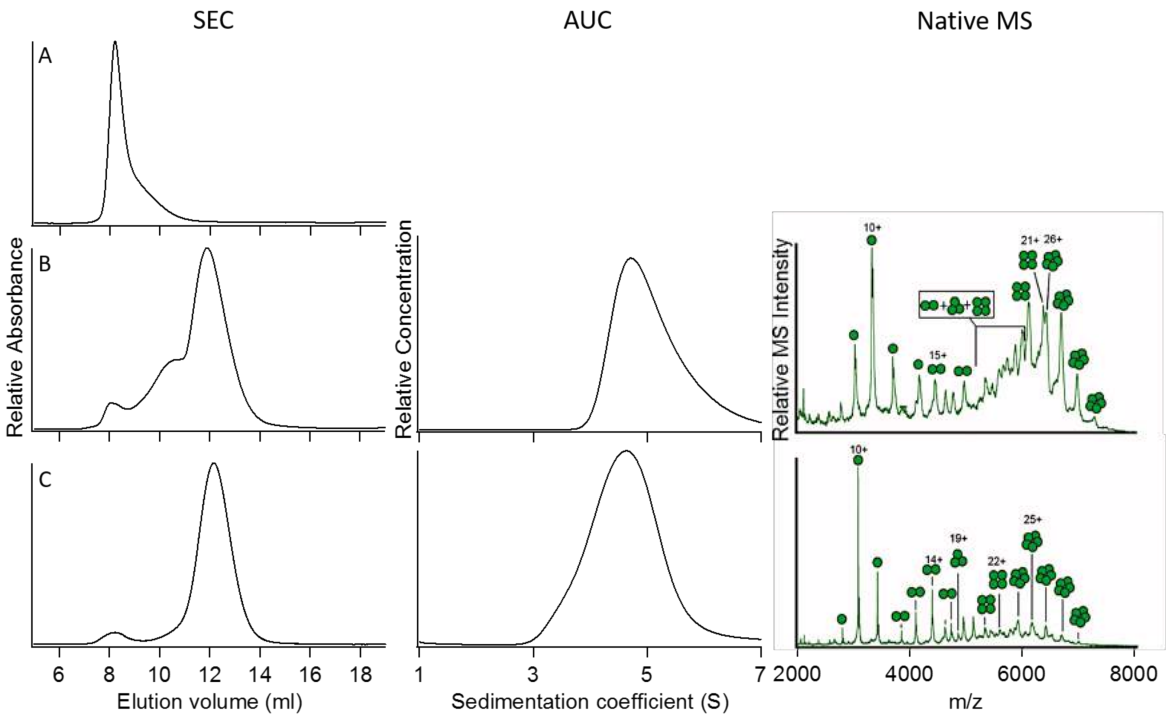

Figure 2. 4 Characterization and optimization of GFP fused with the pentameric coiled coil 4PN8. (A) Construct **5**, for which aggregation prevented its characterization by AUC or native MS. (B) Construct **6**. (C) Construct **7**. For details of the constructs, refer to Table 2.

2.3.4 Transposition of Oligomerization Domain to N-terminus of GFP

Construct **3** successfully mediated the trimerization of GFP when fused to its C-terminus. We therefore extended our studies on this coiled coil domain to investigate whether it would function similarly when placed at the N-terminus of GFP. The coiled coil sequence from 4DZL (containing Ile at 'a' and 'd') was genetically fused to the N-terminus of GFP separated by the same six-Gly spacer sequence to give construct **8**.

Surprisingly, although construct **8** was expressed as a soluble, well-folded protein, it failed to oligomerize at all (Fig. 2.5 A). Its elution volume (determined by SEC), sedimentation coefficient, and native mass spectrum all indicated that the

protein was exclusively monomeric. In this case, it seemed unlikely that steric crowding between the GFP domains would explain the protein's failure to oligomerize. However, it is possible that placing the coiled coil domain at the N-terminus may have resulted in unintended interactions between the coiled coil and the N-terminal 6-His sequence used to purify the protein that interfered with the coiled coil's ability to oligomerize. In particular, electrostatic interactions between the (partially) positively charged 6-His sequence and negatively charged glutamate residues in the coiled coil domain could potentially disrupt the coiled coil structure.

Therefore, we investigated the effect of strengthening the coiled coil interaction by increasing the number of heptad repeats in the coiled coil domain from 4 to 5, resulting in construct **9**. The addition of the extra heptad appeared to restore the original trimeric oligomerization state of the coiled coil domain. This suggests that increasing the strength of the coiled coil interaction was sufficient to overcome the unintended interactions between the coiled coil and either the His-tag or GFP that initially prevented the fusion protein from trimerizing.

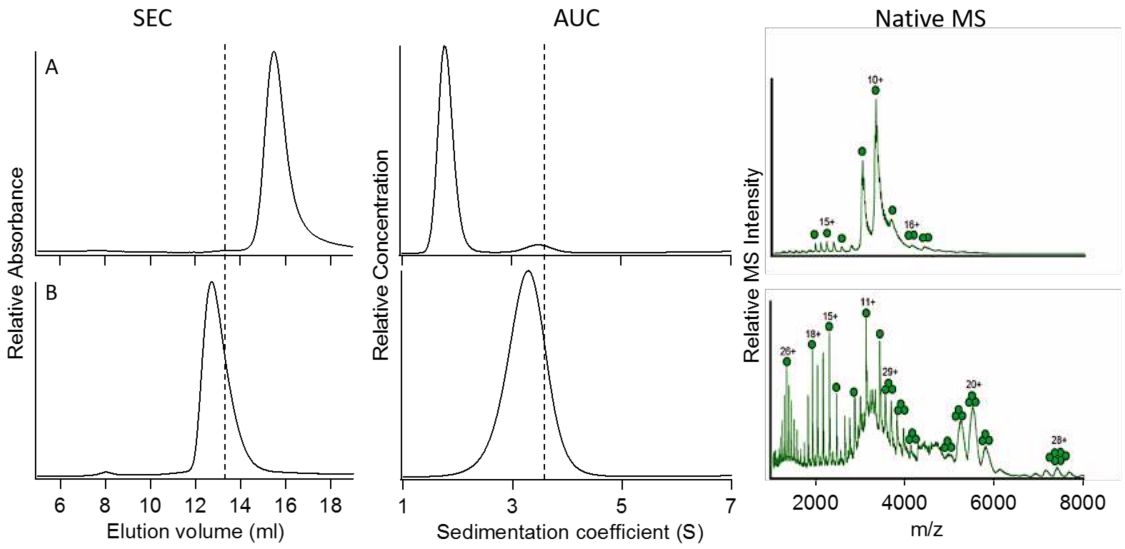

Figure 2. 5 Characterization of fusion proteins in which the trimeric coiled coil 4DZL was linked to the N-terminus of GFP. (A) Construct **8.** (B) Construct **9**. The vertical dotted lines denote the elution volume and sedimentation coefficient measured for the C-terminal fusion protein, **3**. For details of the constructs, refer to Table 2.

2.3.5 Attempts to Construct a Tetrameric GFP Fusion Protein

Based on what we learned from optimizing the trimeric and pentameric coiled coil designs, we attempted to construct a tetrameric fusion protein of GFP by either increasing the length of the spacer sequence between GFP and the coiled coil domain or increasing the strength of the coiled coil interaction. In construct **10,** we increased the number of Gly residues in the spacer from 6 to 9, and in construct **11,** we increased the number of heptad repeat units from 4 to 5. Both these constructs expressed as soluble, folded protein and could be purified to homogeneity. However, analysis of their oligomerization states by SEC, AUC, and native MS each indicated that **10** and **11** remained trimeric (Fig. 2.6). Why these GFP fusion proteins appear to adopt a trimeric rather than a tetrameric oligomerization state is not very clear.

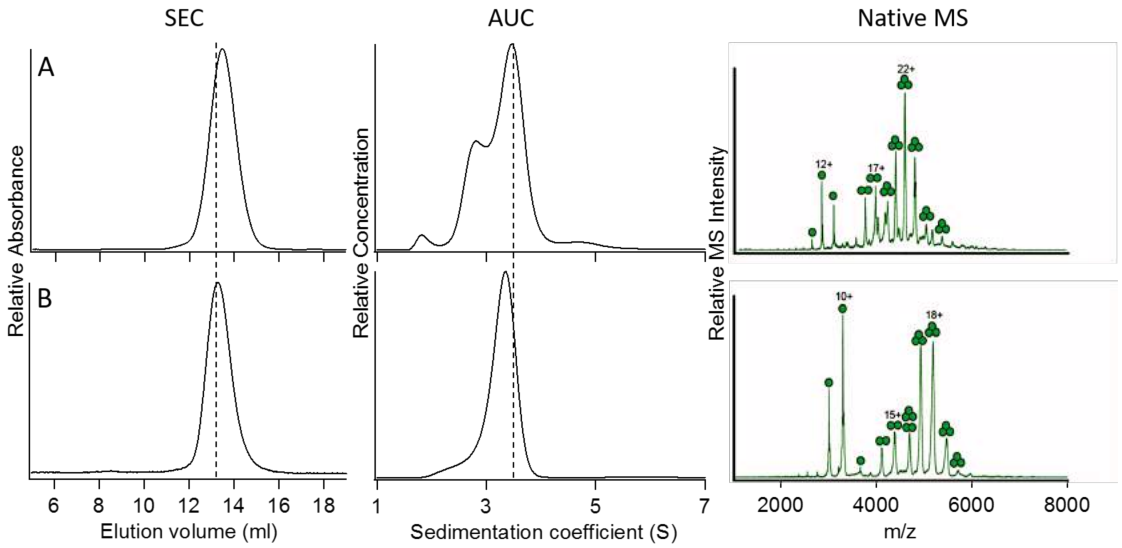

Figure 2. 6 Characterization of modified versions of construct **4**. (A) Construct **10**, heptad strength increased to 5 repeats. (B) Construct **11**, spacer length increased to 9 Gly residues. The vertical dotted lines denote the elution volume and sedimentation coefficient measured for the C-terminal fusion protein, **4**. For details of the constructs, refer to Table 2.

2.4 Discussion

Overall, of the five crystallographically characterized coiled coils we examined, only two constructs, **1** and **3**, retained their oligomerization state when fused to GFP, which served here as a model protein domain. Although they formed discrete, monodisperse assemblies, the oligomerization states of both constructs **2** and **4**, which formed dimers and trimers respectively, were lower than expected based on the trimeric and tetrameric crystal structures of their respective synthetic coiled coil domains. Interestingly, we recently employed the same tetrameric coiled coil motif with Leu at 'a' and Ile at 'd' positions to successfully assemble an octahedral protein cage;[29] in that case the coiled coil functioned as a tetramer, as intended. In contrast, the GFP fusion protein with the pentameric coiled coil, as initially constructed, formed high molecular weight aggregates, suggesting that the protein

was misfolded. Lastly, the trimeric coiled coil, which was well-behaved when fused to the C-terminus of GFP, failed to mediate oligomerization of GFP when transferred to the N-terminus, possibly as a result of interactions between the coiled coil domain and the N-terminal His-tag.

These results illustrated both the limitations and, in some cases, advantages of coiled coil domains as plug-and-play adaptors for the assembly of proteins. As pointed out early on by Betz and DeGrado, among others, the lability of parallel coiled coils derives from the fact that their oligomerization states are primarily specified by hydrophobic packing interactions contributed by the 'a' and 'd' side chains.[5,30–32] The stabilizing interactions at the interfaces of the α-helices arise from complementary electrostatic interactions between residues and 'b' and 'e' and 'c' and 'g' positions and are essentially the same for all oligomerization states between dimer and tetramer (Fig. 2.1). In contrast, the interactions at the interfaces of the α-helices in anti-parallel coiled coils are different for each oligomerization state, rendering this topology inherently more robust. It seems that in several of the cases we studied, the addition of the relatively large GFP domain was sufficient to perturb the delicate balance of hydrophobic core interactions specifying the oligomerization state. Set against this, in the case of construct **9** the modularity of the coiled coil design could be exploited to achieve the desired oligomerization state simply by adding a further heptad repeat to the coiled coil.

Our primary focus in this study was to evaluate the potential of the *de novo* coiled coils to act as simple components to assemble larger, more complex protein subunits and *de novo* protein cages in our case. However, it should be noted that the

choice of protein might also influence how the coiled coils behave; i.e. the protein may not be an entirely neutral component in the assembly process. In this case, for example, the latent tendency of GFP to dimerize at high concentrations could influence the assembly process when individual GFP subunits are brought into close proximity by the coiled coil domain (although we observed no direct evidence for this occurring). It is possible that if had we chosen a different protein, we may have found this set of coiled coils to have been more, or less, successful as oligomerization domains.

After evaluating, some of these *de novo* coiled coils were utilized for our *de novo* protein cage design strategy. As mentioned above, we were able to successfully design an octahedral protein cage by fusing the tetrameric coiled coil to TriEst building block protein (discussed in chapter 1) even though this coiled coil motif failed to assemble into a tetramer when appended to GFP. This octahedral cage was elaborated by fusing a large monomeric protein of ~ 40 kDa to the free end of the coiled coil and in that case it was still able to assemble into discrete particles with M_w of ~ 1.8 MDa and maintained the octahedral core (will be discussed in chapter 3).[33] A tetrahedral cage was designed by adding 1 more heptad repeat to 4DZL trimer and fusing it to the TriEst (discussed in chapter 1).[34] Finally, an icosahedral protein cage with a M_w in the mega-dalton range (~ 2.1 MDa) was designed by fusing the pentameric coiled coil to TriEst protein building block (will be discussed in chapter 4).

Coiled coils are among the most well understood protein-protein interactions and are attractive protein motifs often utilized as off-the-shelf components for *de*

novo designing protein architectures. Both natural and *de novo* coiled coils have been successfully utilized to design metalloenzyme mimics, pH-responsive hydrogels, two-dimensional protein lattices, and three-dimensional protein nanoparticles.[18,19,35–38,38–40] These nanoparticles have subsequently been adapted for the polyvalent display of viral epitopes to create a novel vaccine.[41] Nevertheless, evaluating the coiled coils by fusing to a monomeric protein revealed that coiled coils could subject to unanticipated behaviors when appended to large proteins. Furthermore, monomeric fusion system provided a simple and an easy route to evaluate the coiled coils' behaviors *in vivo* and to carry out modifications to restore intended oligomeric states.

2.5 Conclusion

Overall, this study demonstrated that *de novo* designed coiled coil domains could be subjected to unexpected changes in oligomerization state when coupled to larger protein domains. Thus, of the six fusion proteins we initially constructed (constructs **1**–**5** and **8**), only two behaved as expected (constructs **1** and **3**), which formed dimeric and trimeric assemblies. The relatively low success rate in mediating the desired self-assembly of a simple, monomeric protein comprising only a single domain should serve as a cautionary note for those considering coiled coils for use in synthetic biology applications. It appears that context effects are likely to play a greater role in modulating the structures of coiled coils than has previously been appreciated. Encouragingly, by varying the strength of the coiled coil and the length of the spacer between the two protein domains (parameters that might intuitively be expected to affect the ability of the coiled coil domain to

oligomerize), we were able to achieve the intended oligomerization states of two of the fusion proteins (constructs **7** and **9)**, representing pentameric and trimeric assemblies. These observations suggest that by screening a fairly sparse matrix of spacer lengths and coiled coil strengths, it should be possible to identify conditions for successful protein assembly in many cases.

2.6 Acknowledgements

I acknowledge Dr. Aaron sciore for designing and evaluating wild type GFP and GFP constructs 1 and 2, Dr. Somaye Badieyan for evaluating GFP construct 3. AUC experiments were performed by Dr. Philipp Koldewey in Dr. James Bardwell's lab in the Department of Molecular, Cellular and Developmental Biology at the University of Michigan, Ann Arbor. Native MS experiments were performed by Dr. Joseph Eschweiler in collaboration with Dr. Brandon Routolo's lab in the Department of Chemistry at the University of Michigan, Ann Arbor.

This research was supported by grants from the Army Research Office ARO W911NF-11-1-0251 and W911NF-16-1-0147 to E.N.G.M.

2.7 References

1. Wolf E, Kim PS, Berger B (1997) MultiCoil: A program for predicting two-and three-stranded coiled coils. Protein Sci 6:1179–1189.

2. Apostolovic B, Danial M, Klok H-A (2010) Coiled coils: attractive protein folding motifs for the fabrication of self-assembled, responsive and bioactive materials. Chem Soc Rev 39:3541-3575.

3. Burkhard P, Stetefeld J, Strelkov SV (2001) Coiled coils: a highly versatile protein folding motif. Trends Cell Biol. 11:82–88.

4. Liu J, Yong W, Deng Y, Kallenbach NR, Lu M (2004) Atomic structure of a tryptophan-zipper pentamer. Proc Natl Acad Sci USA 101:16156–16161.

5. Harbury PB, Zhang T, Kim PS, Alber T (1993) A switch between two-, three-, and four-stranded coiled coils in GCN4 leucine zipper mutants. Science 262:1401–1407.

6. Zaccai NR, Chi B, Thomson AR, Boyle AL, Bartlett GJ, Bruning M, Linden N, Sessions RB, Booth PJ, Brady RL, et al. (2011) A de novo peptide hexamer with a mutable channel. Nat Chem Biol 7:935–941.

7. Buer BC, Meagher JL, Stuckey JA, Marsh ENG (2012) Comparison of the structures and stabilities of coiled-coil proteins containing hexafluoroleucine and *t* -butylalanine provides insight into the stabilizing effects of highly fluorinated amino acid side-chains. Protein Sci 21:1705–1715.

8. Oakley MG, Kim PS (1998) A buried polar interaction can direct the relative orientation of helices in a coiled coil. Biochemistry 37:12603–12610.

9. Kashiwada A, Hiroaki H, Kohda D, Nango M, Tanaka T (2000) Design of a Heterotrimeric α-Helical Bundle by Hydrophobic Core Engineering. J Am Chem Soc 122:212–215.

10. Marsh ENG, DeGrado WF (2002) Noncovalent self-assembly of a heterotetrameric diiron protein. Proc Natl Acad Sci USA 99:5150–5154.

11. Miller M (2009) The importance of being flexible: the case of basic region leucine zipper transcriptional regulators. Curr Protein Pept Sci 10:244–269.

12. Jung H-I, Cooper A, Perham RN (2003) Interactions of the peripheral subunit-binding domain of the dihydrolipoyl acetyltransferase component in the assembly of the pyruvate dehydrogenase multienzyme complex of Bacillus stearothermophilus. Eur J Biochem 270:4488–4496.

13. Buchholz TJ, Geders TW, Bartley FE, Reynolds KA, Smith JL, Sherman DH (2009) Structural Basis for Binding Specificity between Subclasses of Modular Polyketide Synthase Docking Domains. ACS Chem Biol 4:41–52.

14. Last NB, Schlamadinger DE, Miranker AD (2013) A common landscape for membrane-active peptides: A Common Landscape for Membrane-Active Peptides. Protein Sci 22:870–882.

15. Marsh ENG, Suzuki Y (2014) Using ^{19}F NMR to Probe Biological Interactions of Proteins and Peptides. ACS Chem Biol 9:1242–1250.

16. Polyansky AA, Chugunov AO, Vassilevski AA, Grishin EV, Efremov RG (2012) Recent advances in computational modeling of α-helical membrane-active peptides. Curr Protein Pept Sci 13:644–657.

17. Buer BC, Chugh J, Al-Hashimi HM, Marsh ENG (2010) Using Fluorine Nuclear Magnetic Resonance To Probe the Interaction of Membrane-Active Peptides with the Lipid Bilayer. Biochemistry 49:5760–5765.

18. McAllister KA, Zou H, Cochran FV, Bender GM, Senes A, Fry HC, Nanda V, Keenan PA, Lear JD, Saven JG, et al. (2008) Using α-Helical Coiled-Coils to Design Nanostructured Metalloporphyrin Arrays. J Am Chem Soc 130:11921–11927.

19. Dånmark S, Aronsson C, Aili D (2016) Tailoring Supramolecular Peptide–Poly(ethylene glycol) Hydrogels by Coiled Coil Self-Assembly and Self-Sorting. Biomacromolecules 17:2260–2267.

20. Ryadnov MG (2007) Peptide α-helices for synthetic nanostructures. Biochem Soc Trans 35:487–491.

21. Gradišar H, Božič S, Doles T, Vengust D, Hafner-Bratkovič I, Mertelj A, Webb B, Šali A, Klavžar S, Jerala R (2013) Design of a single-chain polypeptide tetrahedron assembled from coiled-coil segments. Nat Chem Biol 9:362–366.

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

23. Raman S, Machaidze G, Lustig A, Aebi U, Burkhard P (2006) Structure-based design of peptides that self-assemble into regular polyhedral nanoparticles. Nanomed-Nanotechnol 2:95–102.

24. Schuck P (2000) Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. Biophys J 78:1606–1619.

25. Ruotolo BT, Benesch JLP, Sandercock AM, Hyung S-J, Robinson CV (2008) Ion mobility–mass spectrometry analysis of large protein complexes. Nat Protoc 3:1139–1152.

26. Fletcher JM, Boyle AL, Bruning M, Bartlett GJ, Vincent TL, Zaccai NR, Armstrong CT, Bromley EHC, Booth PJ, Brady RL, et al. (2012) A Basis Set of *de Novo* Coiled-Coil Peptide Oligomers for Rational Protein Design and Synthetic Biology. ACS Synth Biol 1:240–250.

27. Thomas F, Boyle AL, Burton AJ, Woolfson DN (2013) A Set of *de Novo* Designed Parallel Heterodimeric Coiled Coils with Quantified Dissociation Constants in the Micromolar to Sub-nanomolar Regime. J Am Chem Soc 135:5161–5166.

28. Thomson AR, Wood CW, Burton AJ, Bartlett GJ, Sessions RB, Brady RL, Woolfson DN (2014) Computational design of water-soluble α-helical barrels. Science 346:485–488.

29. Sciore A, Su M, Koldewey P, Eschweiler JD, Diffley KA, Linhares BM, Ruotolo BT, Bardwell JCA, Skiniotis G, Marsh ENG (2016) Flexible, symmetry-directed approach to assembling protein cages. Proc Natl Acad Sci USA 113:8681–8686.

30. Betz SF, DeGrado WF (1996) Controlling Topology and Native-like Behavior of *de Novo* -Designed Peptides: Design and Characterization of Antiparallel Four-Stranded Coiled Coils. Biochemistry 35:6955–6962.

31. Gottler LM, de la Salud Bea R, Shelburne CE, Ramamoorthy A, Marsh ENG (2008) Using Fluorous Amino Acids To Probe the Effects of Changing Hydrophobicity on the Physical and Biological Properties of the β-Hairpin Antimicrobial Peptide Protegrin-1. Biochemistry 47:9243–9250.

32. Buer BC, de la Salud-Bea R, Al Hashimi HM, Marsh ENG (2009) Engineering Protein Stability and Specificity Using Fluorous Amino Acids: The Importance of Packing Effects. Biochemistry 48:10810–10817.

33. Cristie-David AS, Koldewey P, Meinen BA, Bardwell JCA, Marsh ENG (2018) Elaborating a coiled coil-assembled octahedral protein cage with additional protein domains: A multi-domain coiled coil-assembled protein cage. Protein Sci 27:1883-1890.

34. Badieyan S, Sciore A, Eschweiler JD, Koldewey P, Cristie-David AS, Ruotolo BT, Bardwell JCA, Su M, Marsh ENG (2017) Symmetry-Directed Self-Assembly of a Tetrahedral Protein Cage Mediated by de Novo-Designed Coiled Coils. ChemBioChem 18:1888-1892.

35. Zastrow ML, Pecoraro VL (2013) Designing functional metalloproteins: from structural to catalytic metal sites. Coord Chem Rev 257:2565–2588.

36. Haines-Butterick L, Rajagopal K, Branco M, Salick D, Rughani R, Pilarz M, Lamm MS, Pochan DJ, Schneider JP (2007) Controlling hydrogelation kinetics by peptide design for three-dimensional encapsulation and injectable delivery of cells. Proc Natl Acad Sci USA 104:7791–7796.

37. Rajagopal K, Lamm MS, Haines-Butterick LA, Pochan DJ, Schneider JP (2009) Tuning the pH Responsiveness of β-Hairpin Peptide Folding, Self-Assembly, and Hydrogel Material Formation. Biomacromolecules 10:2619–2625.

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

39. Raman S, Machaidze G, Lustig A, Aebi U, Burkhard P (2006) Structure-based design of peptides that self-assemble into regular polyhedral nanoparticles. Nanomed-Nanotechnol 2:95–102.

40. Magnotti EL, Hughes SA, Dillard RS, Wang S, Hough L, Karumbamkandathil A, Lian T, Wall JS, Zuo X, Wright ER, et al. (2016) Self-Assembly of an α-Helical Peptide into a Crystalline Two-Dimensional Nanoporous Framework. J Am Chem Soc 138:16274–16282.

41. Kaba SA, Brando C, Guo Q, Mittelholzer C, Raman S, Tropel D, Aebi U, Burkhard P, Lanar DE (2009) A Nonadjuvanted Polypeptide Nanoparticle Vaccine Confers Long-Lasting Protection against Rodent Malaria. J Immunol 183:7268–7277.

Chapter 3: Elaborating Coiled Coil-driven Octahedral Cage Design by Fusion with Additional Protein Domains

*This work is published on "**Cristie-David AS,** Koldewey P, Meinin B and Marsh ENG. (2018) Elaborating coiled coil-driven octahedral cage design by fusion with a large monomeric protein. Protein Sci* 27:1883-1890*"*

3.1 Introduction

An interesting feature of our coiled coil base design strategy is the potential to elaborate protein cages in a modular fashion by the addition of further protein domains to the free end of the coiled coil domain (Fig. 3.1). We aimed to test this idea by fusing a large monomeric protein free end of the coiled coil used to assemble an octahedral cage.[1] Furthermore, although the tetrahedral cage could be purified in good yields (~ 20 mg/L of culture), the original octahedral cage could only be purified in low yields (~ 1 mg/L of culture). This was because the octahedral cages bound poorly to the Ni-NTA affinity column used to purify them, and a significant amount of the protein formed inclusion bodies. The primary goal of *de novo* designing cages is to utilize them for various applications in medicine and materials science. However, the proteins must be easily purified and produced in good yields in order to make them useful for such applications. Therefore, we needed to establish a better purification approach for our octahedral cage design.

Furthermore, although cryo-EM reconstruction of the tetrahedral cage

resolved the coiled coil domains, showing them as protruding from the surface of the cage, we were unable to resolve these domains in a similar reconstruction of the octahedral cage. This suggested that there was heterogeneity in the orientation of the coiled coil domains in the octahedral design. We reasoned that we could take advantage of our design strategy to elaborate the octahedral cage by attaching an additional large protein domain to the C-terminal ends of the coiled coil domains. This would force the large protein to point outwards, potentially reducing the formation of inclusion bodies and improving the homogeneity of the resulting elaborated octahedral cage.

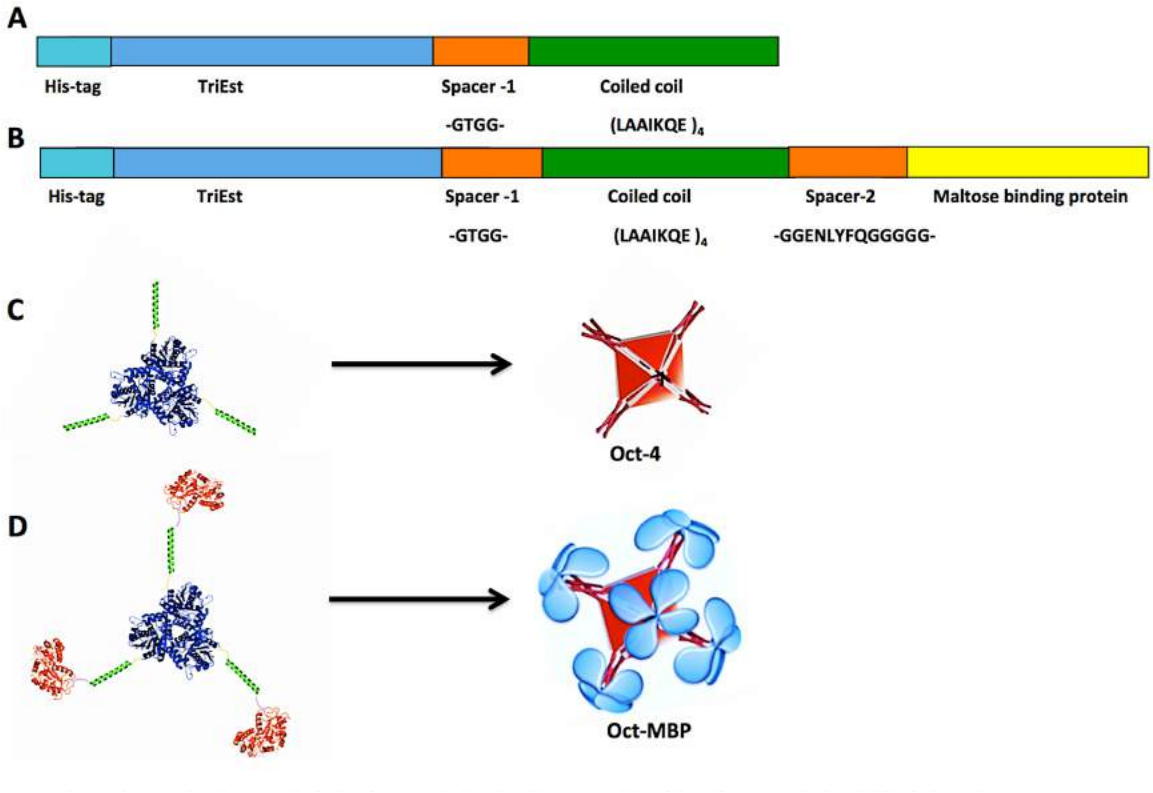

Figure 3. 1 Outline of the strategy used to construct octahedral cages. (A) and (B) display linear maps of the protein sequences of Oct-4 and Oct-MBP, respectively. (C) and (D) show cartoons illustrating the coiled coil-mediated oligomerization of Oct-4 and Oct-MBP, respectively to form octahedral protein cages.

Maltose binding protein (MBP), $M_w \sim 40$ kDa, is a monomeric protein widely used to help solubilize and affinity purify proteins. Additionally, MBP is larger than the TriEst monomers (~ 30 kDa). Therefore, MBP was selected as a solubilizing domain to elaborate the octahedral protein cage. MBP was genetically appended to the C-terminal end of the coiled coil of the Oct-4 octahedral cage construct. In this case fusing MBP protein to Oct-4 produced a structure with a designed molecular weight of ~ 1.8 MDa. The resulting complex protein assembly, Oct-MBP, (Fig. 3.1) retained its octahedral structure and was more homogenous than the original Oct-4 protein cage. Importantly, addition of the MBP domain dramatically improved the efficiency of assembly, resulting in ~ 60-fold greater yield of purified protein compared to the original cage design. This study shows the potential of using small, coiled coil motifs as off-the-shelf components to design MDa-sized protein cages to which additional structural or functional elements can be added in a modular manner.

3.2 Materials and Methods

3.2.1 Construction of Genes Encoding Fusion Proteins

The MBP gene was PCR amplified from pMAL-c5X using commercially synthesized primers. The 5'end of the forward primer was designed to overlap with the sequence encoding the C-terminus of the coiled coil and 13-residue linker in the middle. The reverse primer incorporated two stop codons and overlapped the T7 terminator region of pET-28B at its 5' end. The codon-optimized gene encoding the tetrameric coiled coil and the oligo-Gly spacer units was commercially synthesized.

Both the tetrameric coiled coil and the MBP construct were introduced into the expression vector pET-28b by Gibson assembly using the NEB Gibson assembly protocol. The complete sequence of the Oct-MBP design is shown in Appendix A.2.

3.2.2 Protein Expression and Purification

Proteins were expressed in *E. coli* BL21 by standard methods, as described in previous chapter (see 2.2.2), and cell pellets were stored at -80 ˚C. All purification steps were performed at 4 ˚C and all the buffers used were at pH 7.5. Cell pellets (~ 8 g) were thawed on ice for 20–30 min, then resuspended in buffer containing 50 mM HEPES, 100 mM ammonium acetate, 5% glycerol, 1 FAST protease inhibitor tablet (Sigma), and 50 mg of hen egg lysozyme (Sigma). This suspension was kept on ice and gently agitated for 20 min. Cells were lysed by sonication. The lysate was clarified by centrifugation at 40,000 *g* for 30 min and filtered through a 0.2 μM filter. During Ni-purification attempts, the supernatant was loaded onto a HisTrap Ni-NTA column, washed with 60 - 80 mL of the same buffer, and eluted with 50 mM HEPES buffer, pH 7.5, containing 100 mM ammonium acetate, 750 mM imidazole, and 5% glycerol. Ni-affinity chromatography gave a very low yield of protein and therefore Oct-MBP was purified using its MBP tag. The supernatant was loaded on to a 5 mL MBP-trap column (GE Healthcare) at a flow rate of 0.5 mL/min. The column was subsequently washed with 60 - 80 mL of the same buffer at a flow rate of 1 mL/min and then eluted with buffer containing 10 mM maltose, 50 mM HEPES, 100 mM ammonium acetate and 5% glycerol. Fractions containing the desired protein (approximately 15 mL) were combined and treated with benzonase (1 μL). The resulting solution was dialyzed against buffer containing 20 mM HEPES, 100 mM

ammonium acetate, 2 mM EDTA and 10% glycerol for two days.

3.2.3 Size Exclusion Chromatography (SEC)

SEC was used as an additional purification step and also an analytical step. SEC was performed on a Superose 6 10/300 column equilibrated with buffer containing 20 mM HEPES, 100 mM ammonium acetate, and 2 mM EDTA at 4 °C. Injections comprised 400 μL of sample at 1 mg/mL. Fractions with desired oligomeric state were pooled and kept in the column elution buffer for further characterization using the techniques described below. If needed, protein was concentrated using 100-kDa Amicon ultra-centrifugal filter units.

3.2.4 Analytical Ultracentrifugation (AUC)

Sedimentation velocity analysis was performed using a Beckman Proteome Lab XL-I analytical ultracentrifuge (Beckman Coulter, Indianapolis, IN) equipped with an AN60TI rotor. For AUC, samples were prepared by buffer exchanging into PBS, 100 mM NaCl, pH 7.4 by SEC on a Superose 6 10/300 FPLC column equilibrated with the above buffer. Samples were loaded into pre-cooled cells containing standard sector-shaped 2-channel Epon centerpieces with 1.2 cm path-length (Beckman Coulter, Indianapolis, IN) and allowed to equilibrate at 6 °C for 2 h in the non-spinning rotor prior to sedimentation. Proteins were sedimented at 22,500 rpm at 6 °C. Absorbance data were collected at a wavelength of 220 nm.

3.2.5 Negative-stain TEM Imaging

Protein samples were adjusted to a concentration 0.04 mg/mL and fixed on Formvar/Carbon 400 Mesh, Cu grids using conventional negative staining

procedures. Imaging was performed using a JEOL 1500 electron microscope equipped with tungsten filament, XR401 high sensitivity sCMOS camera and operated at 90 kV.

3.2.6 Activity Assays

Catalytic activity for TriEst, Oct-MBP and cross-linked and TEV-cleaved Oct-MBP was assessed by incubating 1 μM of protein with 1 mM *2,4*-di-nitrophenyl acetate (*2,4*-DNPA) substrate at 45 °C in 20 mM HEPES, 100 mM ammonium acetate, and 2 mM EDTA with 5% v/v methanol (Reaction volume: 150 μl). Changes in absorbance were measured at 405 nm.

3.2.7 Dynamic Light Scattering (DLS)

DLS was performed using a DynaPro NanoStar ZS instrument using a standard 90° scattering geometry. The samples were centrifuged to remove any suspended particles before being analyzed in 10 μL cuvettes at 25 °C. The refractive index and absorption coefficient for the particles were set at 1.45 and 0.001, respectively. Runs were performed in triplicate and each run is an average of 15 scans. At the concentration ranges used non-ideality effects were assumed to be negligible.

3.2.8 Protein Crosslinking and TEV Cleavage

The Oct-MBP protein cages were cross-linked using bis(sulfosuccinimidyl)-suberate (BS3). BS3 concentrations were varied between 1 mM and 2 mM whereas the protein concentrations were varied between 5 and 10 μM. The reaction was performed in dialysis buffer at pH 7.5 and the sample was kept at 4 °C for 3-4 days

to ensure that the cross-linker is fully destabilized. ~ 0.5 μM TEV protease was added and incubated over night at 4 °C. The cleaved material was purified by SEC on a Superose 6 10/300 column and the efficiency of cleavage was assessed by SDS-PAGE.

3.3 Results

3.3.1 Construction of the Elaborated Octahedral Protein Cage, Oct-MBP

MBP was genetically fused to the C-terminus of Oct-4 through a 13-residue spacer sequence (GGENLYFQGGGGG) that incorporates a TEV protease cleavage site, which potentially allows the MBP domain to be removed from the cage, and a series of glycine residues designed to impart flexibility. This design was designated Oct-MBP. The choice of this relatively long spacer sequence was informed by our previous studies, which have shown that coiled coil oligomerization states can be affected, leading to aggregation, by the proteins attached to them if the spacer is too short.[2]

Oct-MBP was over-expressed in *Escherichia coli*. Attempts to purify Oct-MBP by Ni-affinity chromatography using its N-terminal His-tag gave a low yield of protein (Fig 3.2 A) ~ 2 mg/ mL) suggesting that the N-terminus is buried within the protein cage. Next we purified Oct-MBP by MBP utilizing a MBP-trap column (Fig 3.2 B). Size exclusion chromatography followed by MBP-affinity chromatography gave more homogeneous protein as judged by SDS-PAGE (Fig. 3.3 B). The yield of purified protein (~ 60 mg/L culture) represented a dramatic improvement over the yields obtained with the original Oct-4 design. The esterase activity of Oct-MBP, measured

as turn-over number with *2,4*-di-nitrophenyl acetate as substrate, was 12.4 ± 0.7 min^{-1}, which is very similar to the activity of the unassembled trimeric esterase (14.2 ± 2.3 min^{-1}). This indicates that the assembly process did not result in any gross structural perturbation of the protein building blocks.

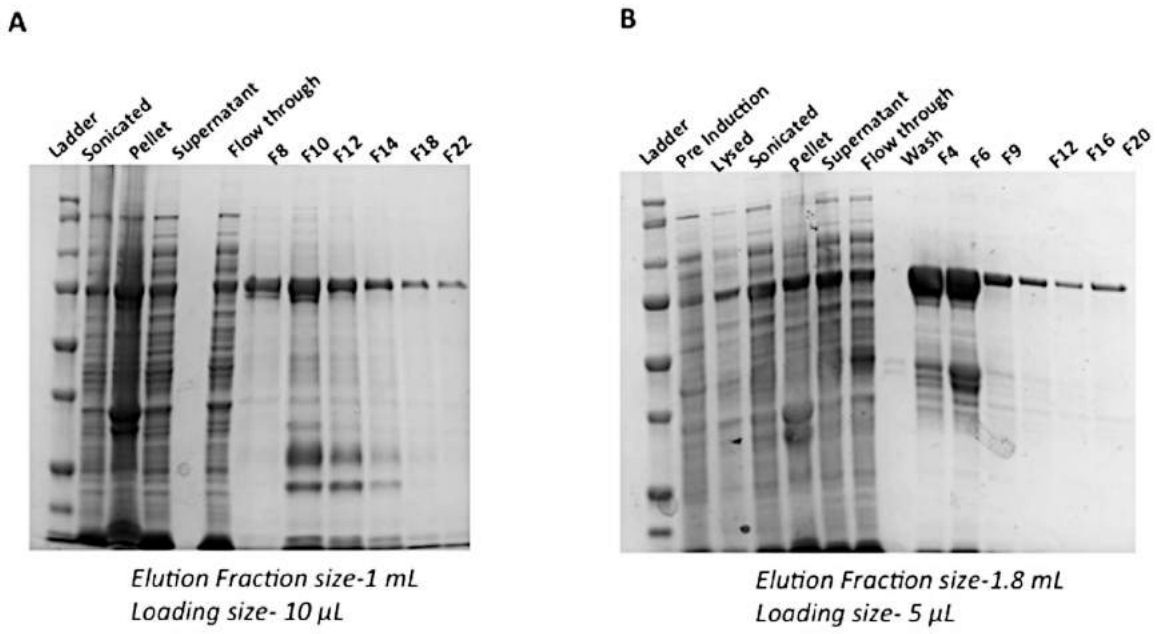

Figure 3. 2 Comparing SDS PAGE gels Oct-MBP purified by Ni-affinity chromatography vs MBP affinity chromatography. (A) Purification of Oct-MBP by Ni-affinity chromatography. F8-F12 are the elution fractions and 10 µL of fractions were loaded on to the gel. (B) Purification of Oct-MBP by MBP-trap column. F4-F20 are the elution fractions and 5 µL of fractions were loaded onto the gel.

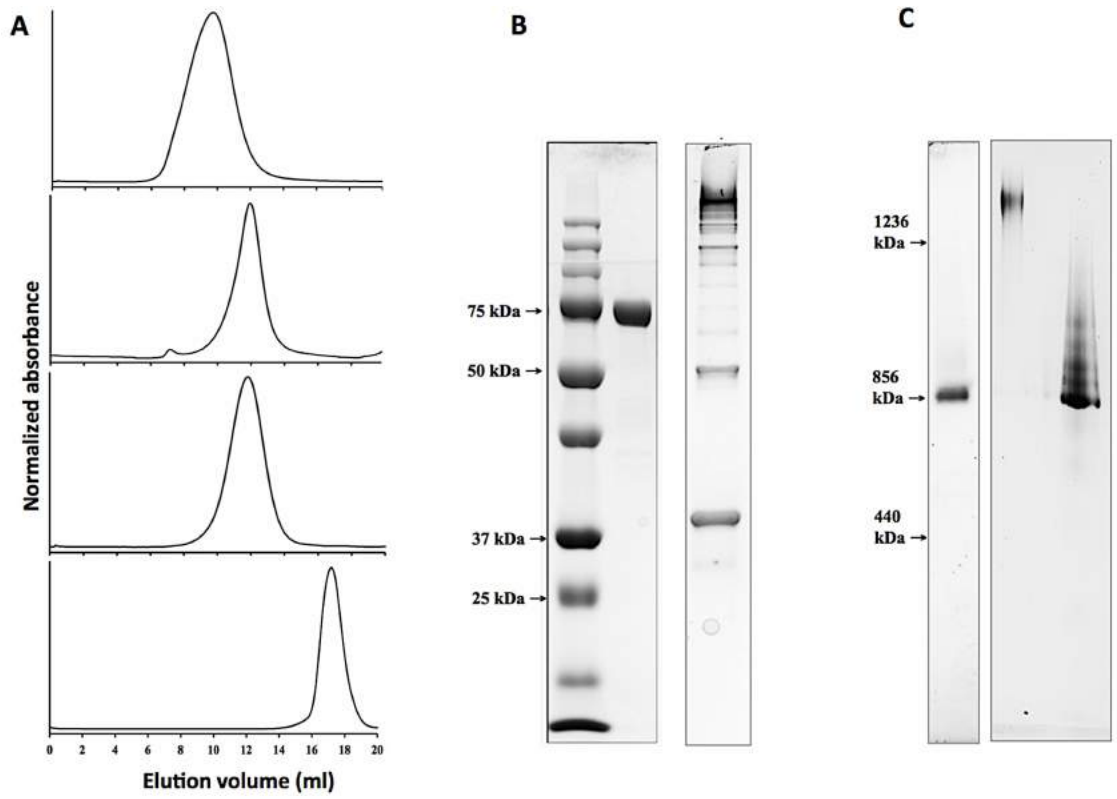

Figure 3. 3 Initial characterization of Oct-MBP. (A) Size exclusion chromatographs (Superose 6 10/300 column) of, *Top to bottom*: Oct-MBP, Oct-4, TEV protease-cleaved Oct-MBP and Tri-EST. (B) SDS-PAGE analysis of Oct-MBP. *Lane 1*: marker proteins; *lane 2*: purified Oct-MBP; *lane 3*: cross-linked and TEV-cleaved Oct-MBP. For details see the text. (C) Native PAGE analysis of Oct-MBP and Oct-4. *Lane 1* Oct-4; *lane 2*: Oct-MBP *lane 3*: Oct-MBP after cross-linking and TEV protease cleavage.

3.3.2 Characterization of the Elaborated Octahedral Cage

To examine whether the construct assembled into the intended octahedral cage geometry, we characterized the complex using size exclusion chromatography (SEC), native PAGE, sedimentation velocity analytical ultracentrifugation (SV-AUC), negative-stain transmission electron microscopy (TEM) and dynamic light scattering (DLS). These techniques provide complementary information on the size and shape of protein complexes.

3.3.3 Size Exclusion Chromatography

We initially used SEC to examine whether Oct-MBP assembled into a high molecular weight complex. When subjected to FPLC chromatography on a Superose 6 10/ 300 SEC column, (molecular cutoff ~ 5 x 10^6 Da; void volume ~ 8 mL), purified Oct-MBP eluted a single, symmetrical peak, with an approximate elution volume of 10 mL (Fig. 3.3 A). For comparison, Oct-4 (~ 854 kDa), the unelaborated octahedral protein cage, eluted at ~12 mL, Tet8-5H (~ 439 kDa), a tetrahedral protein cage, eluted at ~ 14 mL and the unassembled trimeric protein building block (TriEst, ~ 97-kDa) eluted at 18 mL.[1,3] These indicated that, as expected, Oct-MBP assembles into a complex that is significantly larger than Oct-4, from which it is derived.

3.3.4 Native PAGE

The approximate size and homogeneity of Oct-MBP was further analyzed by native PAGE on a 3-12% gradient gel. Oct-MBP migrated as a smeared band that ran significantly more slowly than Oct-4 (Fig. 3.3 C). The observed smearing may result from interactions of the large protein complex with the gel matrix, or possibly the presence incorrectly assembled material. The electrophoretic behavior of Oct-MBP is also consistent with Oct-MBP forming fairly homogenous cages that are significantly larger than Oct-4.

3.3.5 Dynamic Light Scattering

The size of the particles formed by Oct-MPB was further investigated by dynamic light scattering. These revealed the particles formed by Oct-MBP to be

homogeneous and, intensity distribution analysis gave a hydrodynamic diameter of 28.1 ± 0.8 nm (Fig. 3.4). This value is in good agreement with diameter of ~ 24 nm for Oct-MBP, which was simply estimated by modeling the structure of MBP on to the vertices of the Oct-4 protein cage.

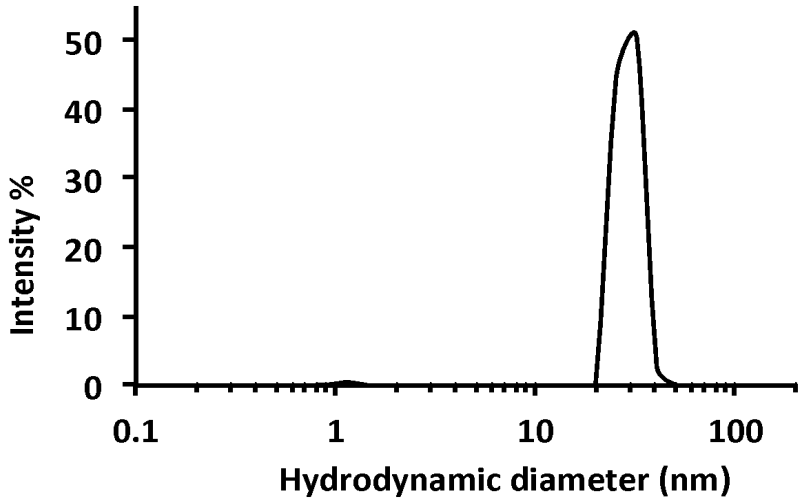

Figure 3. 4 Characterization of Oct-MBP using dynamic light scattering.

3.3.6 Sedimentation-Velocity Analytical Ultracentrifugation (SV-AUC)

Sedimentation traces for Oct-MBP were recorded at 22,500 rpm and 6 °C on samples with initial protein concentrations ranging between 0.014 and 0.07 mg/mL. The sedimentation velocity experiments were first analyzed by the enhanced van-Holde Weischet method to assess the homogeneity of the sample (Fig. 3.5 A).[4,5] At higher Oct-MBP concentrations the observed sedimentation coefficient, *s*, decreased, which indicates non-ideal behavior. This non-ideal behavior likely arises from non-specific interactions between protein cages at higher concentrations, a phenomenon that is generally more pronounced for larger macromolecules. At lower

concentrations the samples were characterized by a major component with a sedimentation coefficient (*s*) of ~ 23 S and another less abundant component comprising faster sedimenting species.

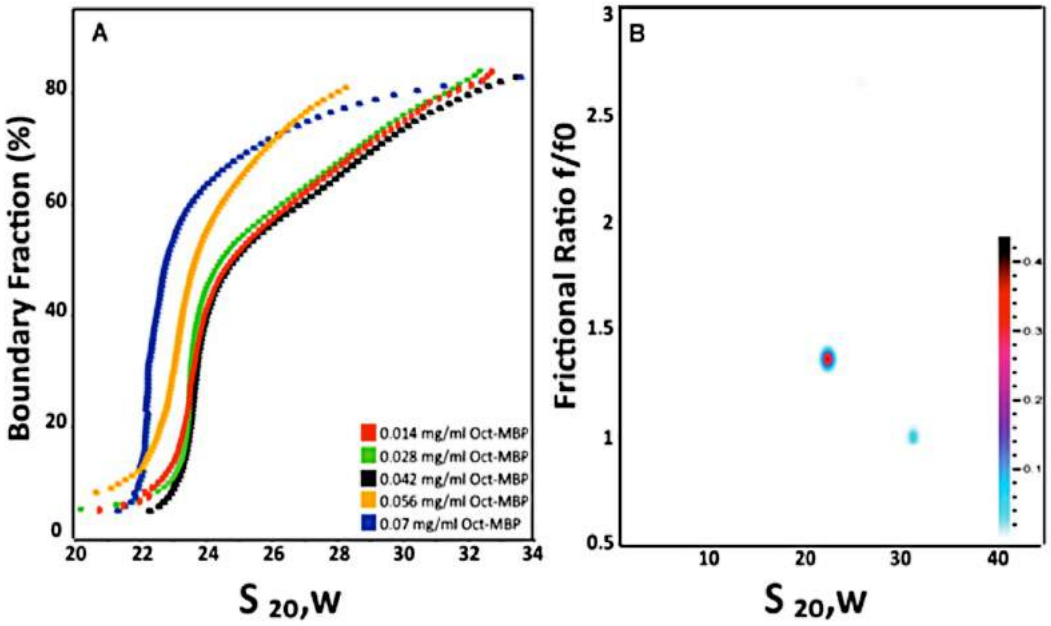

Figure 3. 5 Characterization of Oct-MBP using sedimentation velocity-analytical ultracentrifugation: (A) enhanced van Holde-Weischet analysis of Oct-MBP sedimenting at a range of initial concentrations, as indicated on plot. (B) Representative plot obtained by 2-dimensional sedimentation analysis of Oct-MBP initial protein concentration 0.028 mg/mL.

We further analyzed the data from the lower concentration samples by 2-dimensional sedimentation spectrum analysis (2DSA) followed by a genetic algorithm analysis, which was validated by a Monte Carlo analysis (GA-MC). This is a model-independent analysis approach to fit sedimentation boundaries that allows the shape and molecular mass distribution of macromolecular mixtures to be determined.[6] We have used this methodology to characterize our smaller protein cage designs and determine their molecular weights.[1,3,7] In this case, 2DSA analysis

of Oct-MBP traces recorded at initial concentrations ranging between 0.014 and 0.042 mg/mL consistently resolved two sedimenting species (Fig. 3.5 B). The principle component (comprising ~ 75 % of the sample) had s = 23 ± 1 S, consistent with the van-Holde Weischet analysis, and the minor component had s = 30 ± 1 S. The frictional ratio (f/f_0) of the major 23 S species was 1.7 ± 0.04 for the major component, although this varied somewhat and was dependent upon the sample preparation and concentration. The minor 30 S species has a surprisingly small frictional ratio, f/f_0 = 1.005 ± 0.03, that might indicate that this is a more compact cage conformation or aggregated form of the protein. Overall the AUC analyses indicate that Oct-MBP is predominantly a single species that appears significantly larger than Oct-4 for which s = 17.5 S. The range of frictional ratios associated with major Oct-MBP species falls within the range reported for other porous protein cages such as ferritin, f/f_0 = 1.3, the E2 complex of pyruvate dehydrogenase, f/f_0 = 2.5 and Oct-4, f/f_0 = 1.9, suggesting that it too maintains an open, porous structure.[1,8,9]

3.3.7 Electron Microscopy

The assemblies formed by Oct-MBP were imaged by negative-stain TEM (Fig. 3.6 A,B). The protein cages appeared as discrete particles with a fairly uniform size distribution. Analysis of ~ 300 particle images using the program Image J gave average diameter of 23.4 ± 2.9 nm (Fig. 3.7) and which closely matches with the modeled diameter of ~ 24 nm. Comparison of the images with those obtained from Oct-4 showed that they appeared less well defined and in particular, unlike the parent Oct-4 design, no symmetry was apparent in the particle images. This is most

likely a result of the appended monomeric MBP domains masking the core octahedral cage. In most images, poorly defined peripheral structures were discernable surrounding the core particle that may represent individual MBP domains.

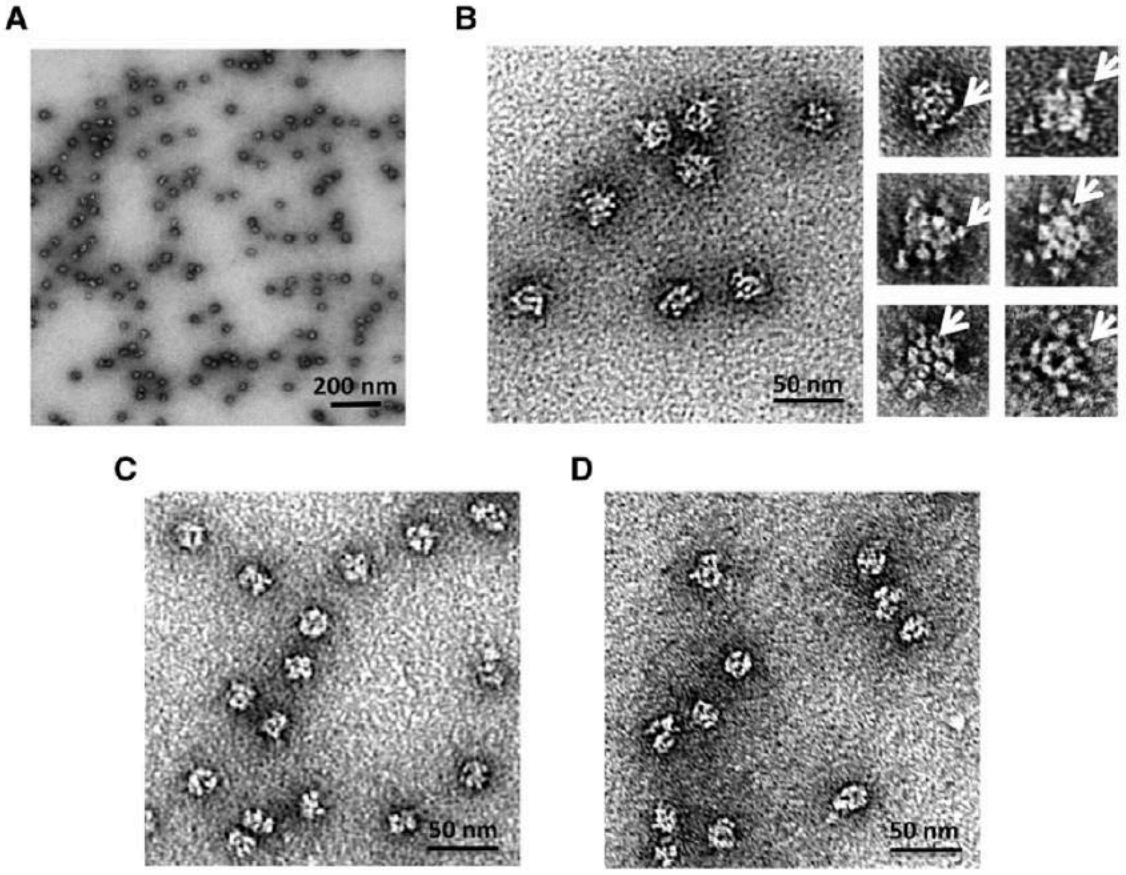

Figure 3. 6 Negative-stain transmission electron microscopy images of Oct-MBP. (A) Wide field view of particles formed by Oct-MBP (B) *left*: Oct-MBP particles at higher magnification exhibit a less compact structure that Oct-4 (compare with panel C) *right:* individual particles at high magnification; arrows indicate peripheral structures that likely represent MBP domains. (C) images of Oct-4 particles lack peripheral structures (compare with panel B) (D) images of Oct-MBP after removal of MBP domains by treatment with TEV protease. These particles more closely resemble the compact structures formed by Oct-4 shown in panel C.

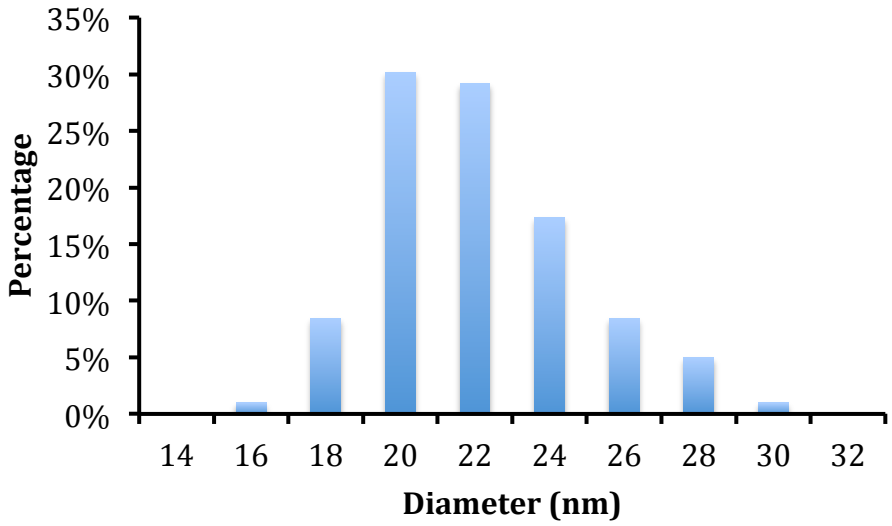

Figure 3. 7 Size distribution analysis of protein cages formed by Oct-MBP obtained from measurement of ~ 300 particles imaged by negative stain TEM using the software package Image J.

3.3.8 Cleavage of the MBP Domain

To further establish the multidomain nature of Oct-MBP we used tobacco etch virus protease (TEV) to cleave the maltose-binding domain from the octahedral core. Initial attempts at proteolysis resulted in dissociation of Oct-MBP to its component trimers (Fig. 3.8). The reason for this is unclear. Therefore, we first covalently cross-linked the Oct-MBP subunits using the lysine-reactive cross linker, bis(sulfosuccinimidyl)suberate (BS3). The protein was then incubated with TEV protease overnight and the material re-purified by SEC. The cross-linking preserved the structure of the cages and the TEV-cleaved material chromatographed with an elution volume similar to Oct-4 (Fig. 3.3 A). Analysis by native PAGE showed that most of the TEV-cleaved material migrated similarly to Oct-4, although some higher molecular weight bands are evident (Fig. 3.3 C). These may represent cages from

which the MBP domain has been incompletely cleaved or cages in which the MBP domain was covalently cross-linked to the Oct core. Examination of the cleaved cages by TEM showed that the particles have a similar appearance and size to Oct-4, consistent with the removal of the outer MBP domain (Fig. 3.6 C,D). The cross-linked and TEV-cleaved cages retained their catalytic activity (14.8 ± 1.8 min^{-1}), indicating that the structure of the esterase building block was not significantly altered by these manipulations.

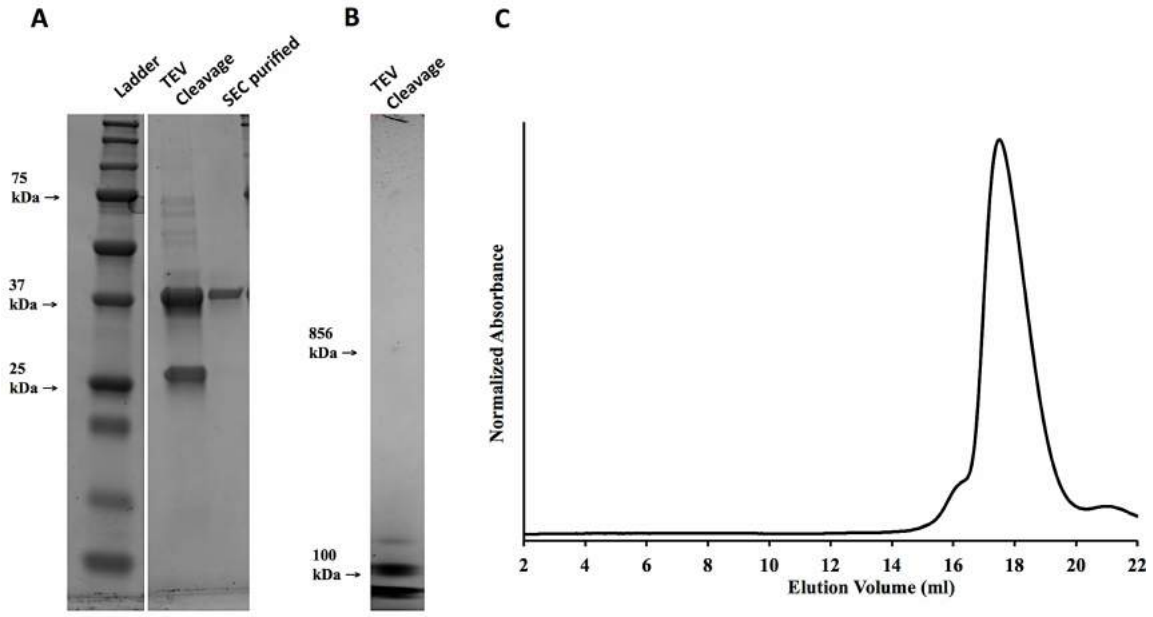

Figure 3. 8 Initial attempts to cleave MBP tag from Oct-MPB protein cages. (A) SDS-PAGE for TEV cleaved Oct-MBP. As shown in the gel, SEC followed by TEV-cleavage completely removed the MBP domain. (B) 3-12 % Native PAGE for TEV cleaved Oct-MBP. The cleaved proteins migrated rapidly in the gel which indicated that MBP cleaved protein does not assemble into an octahedral protein cage with Mw ~ 856 kDa (C) Re-chromatography of MBP removed protein. It eluted at elution volume 18 mL, the elution volume of TriEst. Native PAGE analysis and SEC data indicated that removal of MBP from Oct-MBP leads to complete dissociation of cages into its trimeric form.

3.4 Discussion

Growing interest in the use of protein nanoparticles for applications in medicine and materials science has stimulated efforts to design new protein cages from a range of protein building blocks.[10–17] Our efforts have focused on the use of *de novo* designed coiled coils as "off-the-shelf" components to assemble proteins into cages.[1,3,7] In this study, we demonstrated the potential of our design strategy for the modular expansion of a protein cage by fusing a relatively large monomeric protein, MBP, to the free end of the coiled coil domain used to mediate cage assembly. The resulting construct self-assembled into a fairly homogeneous protein cage in which MBP was displayed on the exterior. The addition of the MBP domain both improved protein folding and allowed the cage to be easily purified by affinity chromatography. Furthermore, it was possible to remove the MBP domain by TEV protease cleavage and recover the core octahedral protein cage.

A valuable consequence of appending MBP to the octahedral cage was to increase the purified yield of protein by ~ 60-fold. This is important for real-world applications where proteins need to be produced in high yields. The purified yield of our initial Oct-4 design was only about 1 mg/L culture, both because a significant fraction of the protein was produced as inclusion bodies, presumably because it was misfolded, and because the His-tag-based affinity chromatography used during its purification was inefficient. We suspect that this inefficiency arose because the N-terminus of the trimeric esterase, which contained the His-tag, was not exposed to the exterior of the cage, so it could not bind to the Ni-NTA affinity column material. Thus, elaborating the original octahedral cage with MBP domains has provided

three benefits over the original design. First it allowed an alternative method of affinity purification that resulted in increased efficiency. Second, it forced the coiled coil domains to point outwards from the octahedral cage (because there is insufficient space within the cage to accommodate the MBP domain) thereby improving homogeneity. Third, it improved folding and assembly, thereby largely eliminating the formation of inclusion bodies and improving the yield of protein.

3.5 Conclusion

In this work, we successfully elaborated our octahedral protein cage design by fusing a larger monomeric protein to the free end of the coiled coils. This project supports the idea that the coiled coil-mediated assembly of protein superstructures based primarily on symmetry considerations is a generalizable and robust strategy. By appending a large protein domain at each end of the small, coiled coil domain (~ 3 kDa) we have constructed discrete cage-like assemblies that extend to the mega-Dalton size range. The domain fusion strategy that we demonstrate here potentially allows the cages to be elaborated towards various applications such as the polyvalent display of antigens for vaccine development or the construction of artificial multi-enzyme complexes. More generally, this design strategy provides a conceptually simple route to design complex, higher order protein assemblies that is largely independent of the structure of the proteins and does not require extensive computational modeling and protein re-design.

3.5 Acknowledgements

AUC was performed by Dr. Philipp Koldeway and Ben Meinin in Dr. James Bardwell's lab in the Department of Molecular, Cellular and Developmental Biology at the University of Michigan, Ann Arbor.

This research was supported in part by a grant from the Army Research Office ARO W911NF-16-1-0147 to E.N.G.M. Electron microscopy was performed at the Microscope Imaging Laboratory (MIL) core facility at University of Michigan. Dynamic light scattering was performed at Biophysics Research Facility at University of Michigan.

3.6 References

1. Sciore A, Su M, Koldewey P, Eschweiler JD, Diffley KA, Linhares BM, Ruotolo BT, Bardwell JCA, Skiniotis G, Marsh ENG (2016) Flexible, symmetry-directed approach to assembling protein cages. Proc Natl Acad Sci USA 113:8681–8686.

2. Cristie-David AS, Sciore A, Badieyan S, Escheweiler JD, Koldewey P, Bardwell JCA, Ruotolo BT, Marsh ENG (2017) Evaluation of de novo-designed coiled coils as off-the-shelf components for protein assembly. Mol Syst Des Eng 2:140–148.

3. Badieyan S, Sciore A, Eschweiler JD, Koldewey P, Cristie-David AS, Ruotolo BT, Bardwell JCA, Su M, Marsh ENG (2017) Symmetry-Directed Self-Assembly of a Tetrahedral Protein Cage Mediated by de Novo-Designed Coiled Coils. ChemBioChem 18:1888-1892.

4. Demeler B, Brookes E, Nagel-Steger L (2009) Chapter 4 Analysis of Heterogeneity in Molecular Weight and Shape by Analytical Ultracentrifugation Using Parallel Distributed Computing. In: Methods in Enzymology. Elsevier 454:87–113.

5. Demeler B, van Holde KE (2004) Sedimentation velocity analysis of highly heterogeneous systems. Anal Biochem 335:279–288.

6. Brookes E, Cao W, Demeler B (2010) A two-dimensional spectrum analysis for sedimentation velocity experiments of mixtures with heterogeneity in molecular weight and shape. Eur Biophys J 39:405–414.

7. Patterson DP, Desai AM, Holl MMB, Marsh ENG (2011) Evaluation of a symmetry-based strategy for assembling protein complexes. RSC Adv 1:1004.

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

9. Bosma HJ, Kok A, Markwijk BW, Veeger C (1984) The size of the pyruvate dehydrogenase complex of Azotobacter vinelandii. Association phenomena. Eur J Biochem 140:273–280.

10. Cohen B, Dafni H, Meir G, Harmelin A, Neeman M (2005) Ferritin as an Endogenous MRI Reporter for Noninvasive Imaging of Gene Expression in C6 Glioma Tumors. Neoplasia 7:109–117.

11. Klem MT, Willits D, Solis DJ, Belcher AM, Young M, Douglas T (2005) Bio-inspired Synthesis of Protein-Encapsulated CoPt Nanoparticles. Adv Funct Mater 15:1489–1494.

12. Kaba SA, Brando C, Guo Q, Mittelholzer C, Raman S, Tropel D, Aebi U, Burkhard P, Lanar DE (2009) A Nonadjuvanted Polypeptide Nanoparticle Vaccine Confers Long-Lasting Protection against Rodent Malaria. J Immunol 183:7268–7277.

13. Yildiz I, Shukla S, Steinmetz NF (2011) Applications of viral nanoparticles in medicine. Curr Opin Biotechnol 22:901–908.

14. Yamada K, Yoshii S, Kumagai S, Miura A, Uraoka Y, Fuyuki T, Yamashita I (2006) Floating Gate Metal–Oxide–Semiconductor Capacitor Employing Array of High-Density Nanodots Produced by Protein Supramolecule. Jpn J Appl Phys 45:8946–8951.

15. Lawrence AD, Frank S, Newnham S, Lee MJ, Brown IR, Xue W-F, Rowe ML, Mulvihill DP, Prentice MB, Howard MJ, et al. (2014) Solution Structure of a Bacterial Microcompartment Targeting Peptide and Its Application in the Construction of an Ethanol Bioreactor. ACS Synth Biol 3:454–465.

16. Bhattacharya P, Du D, Lin Y (2014) Bioinspired nanoscale materials for biomedical and energy applications. J R Soc Interface 11:20131067–20131067.

17. Raeeszadeh-Sarmazdeh M, Hartzell E, Price JV, Chen W (2016) Protein nanoparticles as multifunctional biocatalysts and health assessment sensors. Curr Opin Chem Eng13:109–118.

Chapter 4: Designing a Coiled Coil-Mediated Assembly of an Icosahedral Protein Cage with Extremely High Thermal and Chemical Stability

4.1 Introduction

Protein-based compartments are intimately involved in cellular functions, where they serve as storage vessels, assist in protein folding and protein degradation, and in some cases contain enzymatic reactions.[1–5] Viruses provide an excellent example of the importance of protein cages in the packaging, transport and delivery of biological cargoes.[6,7] We discussed in chapter 1 the development of natural protein cages towards a variety of applications: these include targeted drug delivery, polyvalent display of antigens, *in vivo* imaging, templating of nano-particles, and encapsulation of enzymes in protein nano-reactors.

The protein cages most commonly used for such applications are virus like particles (VLPs) derived from icosahedral viruses.[6] VLPs comprise the assembled viral coat protein(s) but lack the viral genome and so are both empty and non-infectious. Examples include cages derived from bacteriophages P22 and Qβ, and from cowpea chlorotic mottle and cowpea mosaic viruses.[6] Viral protein cages have evolved to assemble into very robust cages; however, this places constraints on the extent to which these cages can be re-purposed. The Baker group has explored the potential of *de novo* designing cages with icosahedral geometry ressembling viral

capsids, and within last two years, several icosahedral cages have been reported using the interface design strategy.[8,9] Yeates group also attempted to design a dodecahedral protein cage by utilizing their fusion strategy, but only a small percentage of the protein assembled into the intended form.[10] These design strategies were discussed in detail under chapter 1.

The coiled coil driven cage design strategy developed in our lab, in which protein assembly is mediated by short, parallel, coiled coil domains, provides a more modular approach to design protein cages. Our approach greatly simplifies the design process because the angle between the symmetry axes need not be specified, eliminating the requirement for intensive computational modeling and protein engineering. By using combinations of symmetry axes that are unique to the geometry of the desired protein cage, it is possible in principle to assemble well-defined protein cages using symmetry considerations alone. In chapter 1, I described the construction of octahedral and tetrahedral cages by combining a C_3-symmetric building block protein with either a C_4-symmetric or C_3-symmetric coiled coil assembly domain, respectively. [11,12]

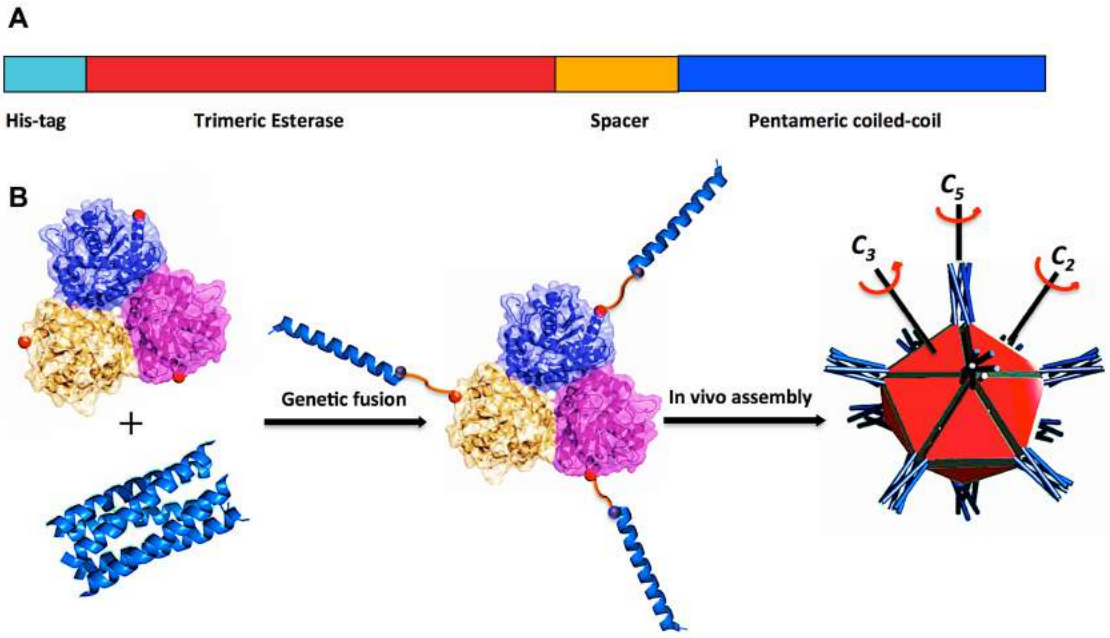

Figure 4. 1 Design scheme for a self-assembling icosahedral protein cage. (A) The Ico8 fusion protein comprises a trimeric esterase, an oligo-Gly linker sequence and a pentameric coiled coil. (B) Structures of the trimeric esterase and pentameric coiled coil with a schematic illustration of the assembly of the fusion protein into an icosahedral cage.

In this chapter, I describe the extension of this approach to design a ~ 2.1 MDa icosahedral cage of a size similar to a small viral capsid. As the largest Platonic solid, icosahedral protein cages can encapsulate large volumes and as such are often used in Nature for packaging and transport.[13] Large, robust and customizable protein cages should be useful in synthetic biology, vaccine design and targeted drug delivery.[7,14–16] However, icosahedral cages remain challenging to construct, as they require the coordinated assembly of 60 protein subunits (in this case 20 copies of a trimeric protein) to achieve the desired icosahedral geometry.

For the combination of C_3 and C_5-symmetric proteins necessary to specify an icosahedral cage, we selected, respectively, same trimeric esterase (TriEst) which we

utilize as building block protein (BBP) for our other designs and a well-characterized *de novo* pentameric parallel coiled coil domain. These were linked together through a flexible oligo-Gly sequence. Characterization of the resulting protein assemblies revealed that they form hollow cages with features consistent with the intended icosahedral geometry. Notably, assembly of the trimeric esterase into a cage resulted in a remarkable increase in both its thermal stability and resistance to chemical denaturation by guanidium hydrochloride (GnHCl) and extremes of pH.

4.2 Materials and Methods

4.2.1 Construction of Genes Encoding Fusion Proteins

The codon-optimized gene encoding the pentameric coiled coil and the oligo-Gly spacer units were commercially synthesized and were introduced into the expression vector pET-28b using Gibson assembly protocol. The sequences of the proteins are given in Appendix A.3.

4.2.2 Protein Expression and Purification

Expression constructs were transformed into *E. coli* BL21(DE3) cells and protein expression induced by IPTG using standard methods as described in chapter 2 (section 2.2.2). For initial characterization, proteins were purified by standard methods using an N-terminal 6-His tag to facilitate affinity chromatography on a Ni-NTA column as described in section 2.2.2. For further characterization of Ico8, the protein was purified as follows: ~ 5 g of cells was resuspended on ice in 45 mL of 50 mM HEPES buffer, pH 7.5 containing 300 mM ammonium acetate, 50 mM imidazole,

1.5 M urea and 5% (v/v) glycerol; a protease inhibitor tablet and 50 mg of lysozyme were added and the suspension gently shaken for 20 min. The cells were lysed by sonication or passage through a microfluidizer and the lysate was clarified by centrifugation for 30 min at 40,000 g. The supernatant was passed through a 5 mL Ni-NTA column at 0.25 mL/min and the flow-through heated at 70 °C for ~ 75 min. Precipitated proteins were removed by centrifugation at 40,000 g for 30 min and the supernatant dialyzed at 4 °C against 20 mM, pH 7.5, HEPES, 2 mM EDTA, 100 mM ammonium acetate for one week. Dialyzed samples were further purified by 2 rounds of SEC using a Superose 6 10/300 column equilibrated in the same buffer (flow rate 0.3 mL/min). Fractions containing Ico8 were pooled, concentrated by ultrafiltration and re-chromatographed on the same column. Samples were then concentrated and stored at room temperature.

4.2.3 Enzyme Activity

The esterase activity of Ico8 and TriEst were assayed using *2,4*-di-nitrophenol acetate as the substrate and measuring the increase in absorbance at 405 nm, as described in 3 chapter (section 3.2.6).

4.2.4 Nucleic Acid Analysis

Nucleic acids were purified from protein samples by phenol:chloroform extraction using standard methods. The presence of DNA, rather than RNA, was confirmed by digestion with either DNase I (RNase free) or RNase, followed by agarose gel electrophoresis. 500 ng of phenol:chloroform extracted nucleic acids were incubated with either 1 unit of DNase I or RNase A for 30 min at 25 °C in the

manufacturers' buffer. The amount of DNA was quantified by analysis of agarose gels and comparison to DNA standards using the program Image J.

To study the effect of DNAse I on the stability of Ico8, 0.5 mg/mL protein samples were incubated overnight at room temperature with 5 units of DNAse I and purified by SEC on a superose 6 10/300. Fractions collected from elution volume ~ 18 mL was used for further characterizations.

4.2.5 Electron Microscopy

Proteins (0.03 – 0.1 mg/mL) were fixed on Formvar/Carbon 400 Mesh, Cu grids using conventional procedures and staining with uranyl formate. For negative stain TEM, imaging was performed at room temperature with JEOL 1500 electron microscope equipped with tungsten filament, XR401 high sensitivity CMOS camera and operated at 90 keV. Samples for scanning transmission electron microscopy (STEM) were prepared similarly. Images were acquired using a JEOL JEM-2100F transmission electron microscope (TEM) with a CEOS probe corrector. Microscopy was performed at 200 keV in STEM mode with the lens setting corrected by the Cs-corrector to produce a sub-Angstrom beam size. Both high-angle annular dark-field and bright-field images were acquired simultaneously. To better display the internal microstructures of the virus, fast Fourier transform image filtering was performed for bright field images.

4.2.6 Dynamic Light Scattering

DLS measurements were made using a DynaPro Nanosizer ZS instrument. Measurements were made at 25 °C with the refractive index and absorption

coefficient for the particles were set at 1.45 and 0.001 respectively. Diameter distributions by intensity were recorded. Measurements were made in triplicate with each measurement comprising an average of 15 scans. Temperature-dependent DLS was performed using the same instrument. 10 μL samples with protein concentration 0.2 mg/mL were heated at a rate of 5 °C/min from 40 to 100 °C. The refractive index and absorption coefficient for the particles were set at 1.45 and 0.001 respectively, and the diameter distribution by intensity was recorded.

4.2.7 Zeta Potential Measurements

Zeta Potential measurements were performed using Malvern Instruments Nanosizer ZS. Samples were buffer-exchanged into 1XPBS buffer containing 20 mM ammonium acetate, pH 7.4. The protein concentration was adjusted to ~ 0.2 mg/mL and measurements were performed in triplicate and averaged readings reported.

4.2.8 Circular Dichroism (CD)

CD measurements were performed using Jasco J815 CD spectrometer. Protein concentrations were between 0.2 – 0.5 mg/mL, (1 mm path length cuvette). Samples were heated from 40 °C to 98 °C at a rate of 1 °C /min.

4.2.9 Differential Scanning Calorimetry (DSC)

DSC scans were performed using a TA instruments nanoDSC, with dialysis buffer was used as reference. ~ 0.6 mL of samples of protein, 0.3 - 0.4 mg/mL, were heated at 1 °C/min from 50 °C to 120 °C at 3 atm pressure and the changes in heat taken up by the samples recorded.

4.2.10 Atomic Force Microscopy (AFM)

Samples for AFM analysis were first chemically cross-linked using the lysine-specific reagent bis(sulfosuccinimidyl)suberate (Thermo Fisher) as described in chapter 3 (section 3.2.8) to achieve ~ 50% cross-linking as determined by SDS-PAGE. The protein solution (20 μL ~ 0.02 mg/mL) was spin-coated at 3500 rpm onto a freshly cleaved mica surface and further rinsed with 40 μL milli-Q water while spinning to remove salt from the surface. AFM imaging was carried out under AC mode using PicoPlus 5500 (Agilent Electronic Measurement, K-tek CNT tip) AFM system. A total of 20 images containing ~ 1100 cages were analyzed by the automated particle detection function in SPIP (version 6.2.6, Image Metrology, 0.8 nm z-range cutoff threshold). Since the protein cage flattened on the mica surface, we computed the undistorted icosohedral diameter from the measured particle volume using the icosahedral geometric relationship: $V = \frac{5(3+\sqrt{5})}{12} * (\frac{2d}{\sqrt{10+2\sqrt{5}}})^3$, where V is the volume and d is the diagonal diameter. This procedure was carried out so that the extent of cage assembly, as assessed by AFM, could be compared to the information obtained from TEM measurements.

The IR-PiFM measurements were performed using a Molecular Vista Inc, Vista-IR microscope. A QCL based light source illuminated the AFM probe (300 kHz NCHR-Au, NanoSensors AG) providing a spectral range from 770-1885 cm^{-1}. The protein solution (20 uL, 0.02 mg/mL) was drop-casted onto freshly cleaved mica and air dried before imaging. The spectral acquisition time was 1.0 seconds with image acquisition time of 4.2 min at 256 x 256 pixels.

4.2.11 Cryo-electron Microscopy (Cryo-EM)

Samples were concentrated to ~ 0.3 mg/mL and loaded onto glow-discharged QUANTIFOIL R2/2 200 mesh grids and flash-frozen using Vitrobot (FEI Mark IV). The samples were visualized at liquid nitrogen temperature on a Talos Arctica electron microscope (FEI) operating at 200 keV. Cryo-EM images were recorded at a nominal magnification of ×34014 using a K2 Summit direct electron detector (Gatan Inc.) in counted mode.

A total of 560 micrographs of cages assembled in the presence of DNA and 240 micrographs for cages assembled in the absence of DNA were taken using automation software SerialEM, Dose-fractionated image Images were subjected to motion correction using MotionCor2, and binned by 2 resulting in a sampling of 2.94 Å per pixel for particle picking and processing.[17,18] For the sample with DNA, 18,707 particles were selected automatically, then subjected to reference-free 2D classifications using RELION.[19] Well-defined class average images were selected for further 3D reconstruction using cisTEM.[20] For the sample without DNA, 9,380 particles were selected manually and subject to reference-free 2D classifications using RELION.

4.2.12 DPH binding

Ico8 and TriEst were buffer exchanged into 100 mM HEPES, 100 mM NaCl, pH 7.0 and DMSO added to 5 % final concentration. Protein samples, 5 μM-10 μM final concentration were incubated with 5 μM DPH for 20-30 min with gently agitating at 25 °C. Fluorescence was visualized under UV light.

4.3 Results

To design a self-assembling icosahedral protein cage, we utilized the same coiled coil driven design approach previously used to assemble tetrahedral and octahedral cages. For the C_3-symmetric building block we used the same trimeric esterase (TriEst) (PDB ID: 1ZOI) as in our previous studies to facilitate comparisons with our earlier designs and assess the generality of the design strategy.[21] To provide the C_5 component necessary for icosahedral geometry we selected a *de novo* designed, parallel, pentameric coiled coil (PDB ID: 4PN8); this component was modified by mutating W13Q to alleviate problems with aggregation, and was described in chapter 2.[22,23]

4.2.1 Screening and Optimization of Assembly

Synthetic genes were constructed to encode proteins in which the C-terminus of TriEst was linked to the N-terminus of the coiled coil through a flexible sequence containing increasing numbers of Gly residues (Fig. 4.1 and Table 4.1). The number of Gly residues were increased two at a time, resulting in six constructs that contained between 2 and 12 Gly, herein referred to as Ico2 through Ico12, with each additional di-Gly unit increasing by ~ 5 Å the distance between the C_3 and C_5 domains. These constructs were expressed and purified by standard procedures as described in the Methods section.

Construct Name	Spacer sequence	Pentameric coiled coil sequence
Ico2	GT	KIEQILQKIEKILQKIEQILQKIEQILQ
Ico4	$GT(G)_2$	KIEQILQKIEKILQKIEQILQKIEQILQ
Ico6	$GT(G)_4$	KIEQILQKIEKILQKIEQILQKIEQILQ
Ico8	$GT(G)_6$	KIEQILQKIEKILQKIEQILQKIEQILQ
Ico10	$GT(G)_8$	KIEQILQKIEKILQKIEQILQKIEQILQ
Ico12	$GT(G)_{10}$	KIEQILQKIEKILQKIEQILQKIEQILQ
Ico8-5h	$GT(G)_6$	KIEQILQKIEKILQKIEQILQKIEQILQKIEQILQ

Table 4. 1 Spacer and coiled coil sequences of fusion proteins generated in this study.

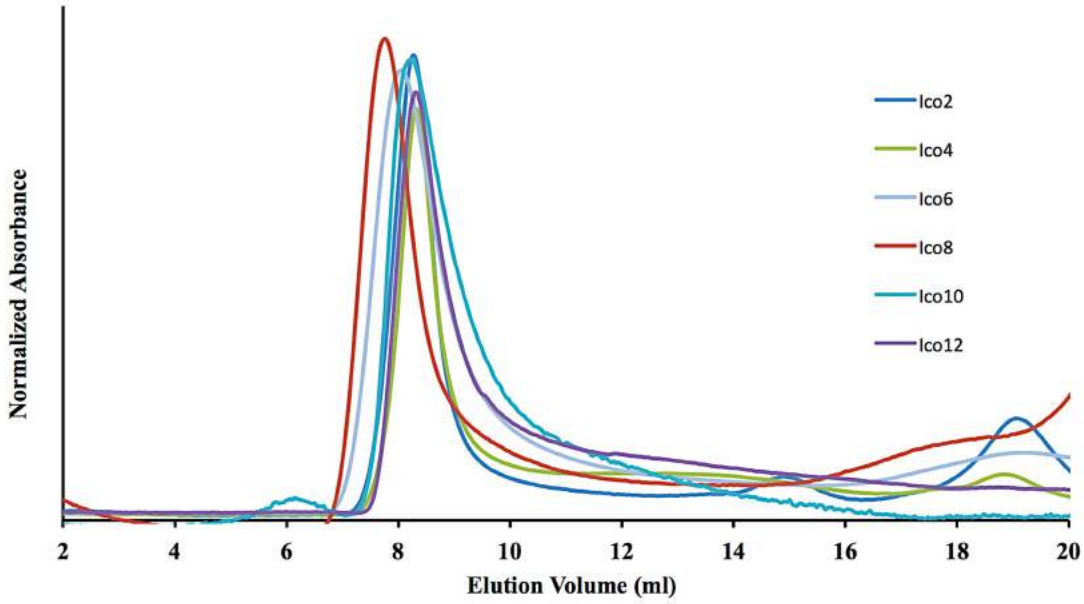

Figure 4. 2 Size exclusion chromatography of Ico constructs.

All the constructs eluted in the void volume (~ 8 mL) on a superose 6 10/300 (Fig. 4.2). Superose 6 10/300 column's M_w cut off is 5 MDa and calculated M_w for the 60-subunit icosahedral cage is ~ 2.1 MDa. These results indicated that the constructs are either aggregrating or assembling into very large molecular weight species. The Ico2 and Ico4 proteins were very unstable and rapidly precipitated, presumably because the linker length was too short for the two protein domains to fold properly. The other constructs were sufficiently stable to be characterized by negative stain TEM. Ico6, Ico10 and Ico12 each formed a heterogeneous mixture of assembled proteins, aggregated material and unassembled trimers (Fig. 4.3 A-C). As evidenced by negative stain TEM, Ico8 assembled into sperical particles with a wide diameter distribution, and for the protein to completely assemble into sperical cages, it took approximately 2-3 months (Fig. 4.3 D,E). Although Ico8 contained an N-terminal His-tag, the protein bound only weakly to Ni-NTA resin used in the initial purification and the yield of protein was less than 1 mg/L culture. From all 6 constructs evaluated, the construct with 8-residue linker was identified as the most promising construct. After optimizing the linker length, we evaluated another protein construct by adding 1 more heptad to the 8-residue linker (Ico8-5h) (Table 4.1). If the heterogeneity of cages result from the weak interaction strength of the 4-heptad pentameric coiled coil, we hypothesized that adding another heptad to its C-terminus will alleviate this problem. However, Ico8-5h construct also formed a mixture of assemblies similar to Ico6, 10 and 12 constructs as evidenced by both SEC and negative stain TEM (Fig. 4.4).

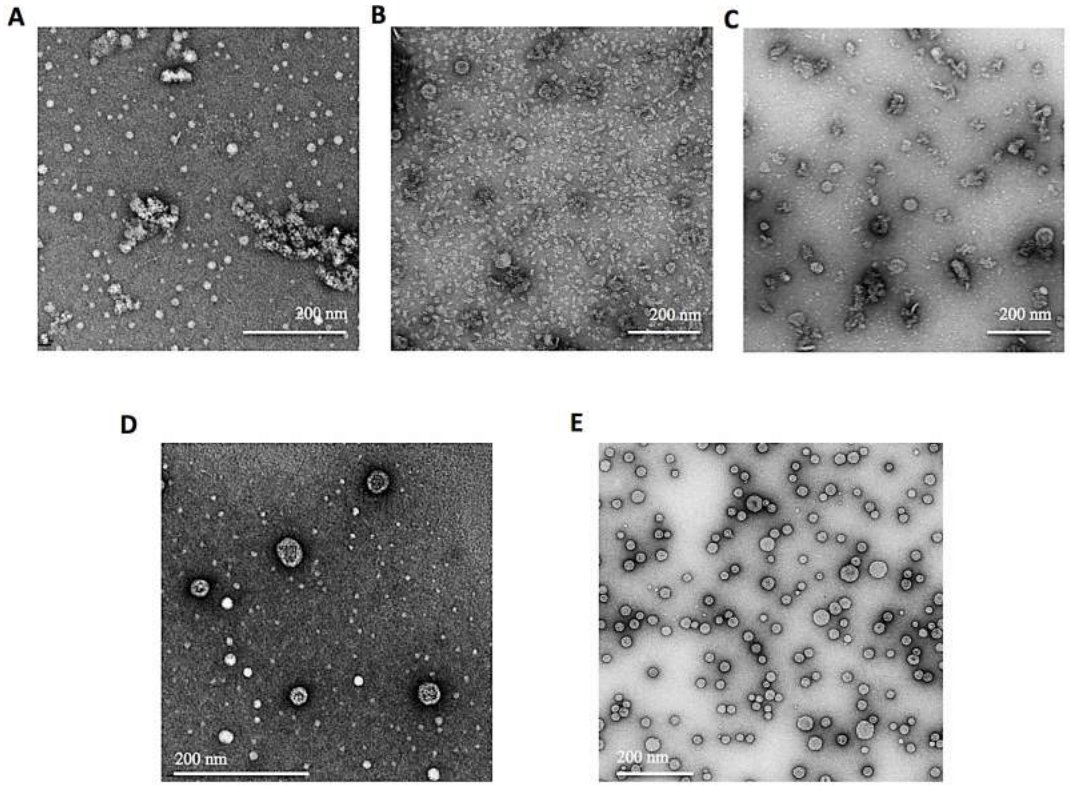

Figure 4. 3 Initial evaluation of protein cage formation by Ico6, Ico8, Ico10 and Ico12 fusion proteins visualized by negative stain TEM. (A) Ico6; (B) Ico10; (C) Ico12. (D-E) Ico8 purified by Ni-affinity column chromatography visualized by negative stain TEM. (D) Few days after purification (E) Same sample after 10 weeks.

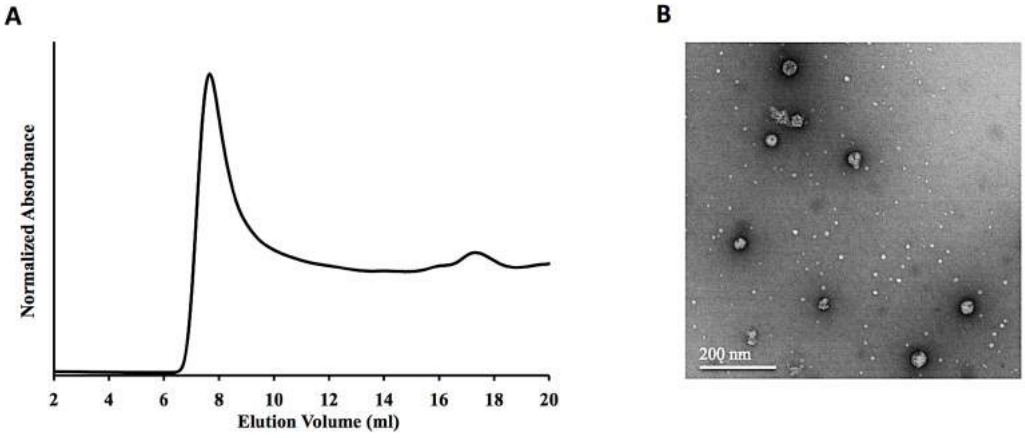

Figure 4. 4 Characterization of of Ico8-5h construct. (A) Size and homogeneity assessed by size exclusion chromatography. (B) Negative stain TEM of particles formed by Ico8-5h.

Ico8 cages were very thermostable (discussed later) and we therefore took advantage of its heat stability to purify the protein. This was achieved by heating the cell lysate at 70 °C for ~ 75 min to precipitate most *E. coli* proteins, which were removed by centrifugation. After dialysis, the cages were purified further by two rounds of SEC and the protein eluted at 10 mL with a symmetrical elution profile on the superose 6 10/300 column. This purification procedure increased the yield of protein to ~ 5 mg/L of culture and produced electrophoretically pure protein (Fig. 4.5 A) that assembled into cages of a more uniform size than those purified by Ni-NTA affinity chromatography. Surprisingly, this purification procedure also accelerated Ico8 assembly from 2-3 months to ~ 2 weeks. This material was used for further characterization.

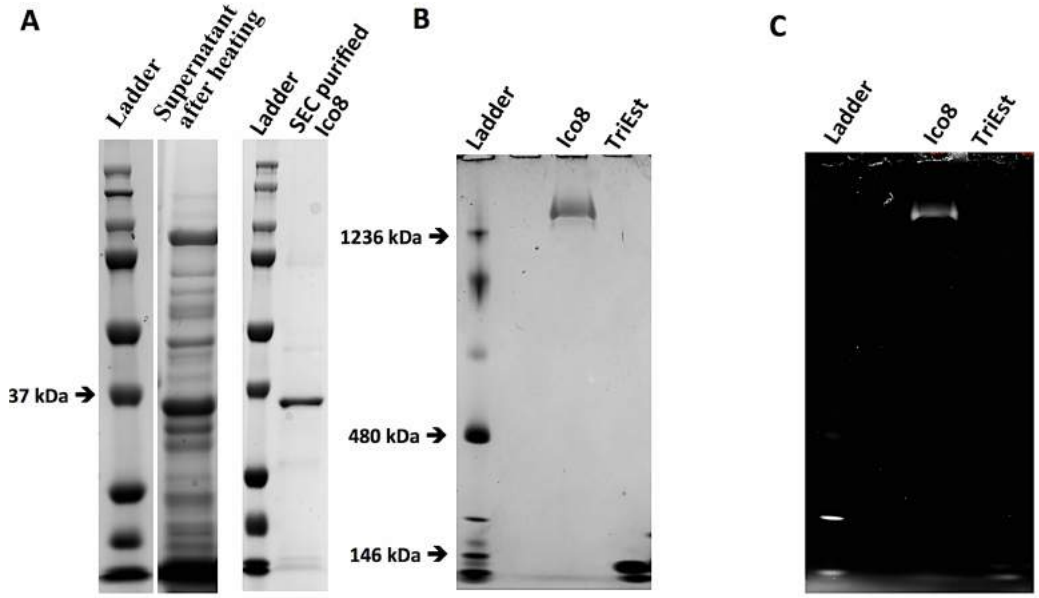

Figure 4. 5 Analysis of Ico8 by polyacrylamide gel electrophoresis (A) SDS-PAGE of heat purified Ico8; (B, C) Co-migration of Ico8 cages and nucleic acid: (B) 3-12% Native PAGE of purified Ico8 and TriEst stained for protein with Instant Blue. (C) The same gel stained with Gel red to detect nucleic acids. Ico8 protein band is stained with Gel red, which confirmed that DNA strands are associated with Ico8.

SEC and dynamic light scattering (DLS) analysis indicated that Ico8 formed quite homogenous assemblies of the correct size (Fig. 4.6 A,B). Analysis of negative stain TEM images of Ico8 showed that Ico8 assembled into particles ranging in diameter from ~ 24 – 30 nm (Fig. 4.6 C). However Ico8 appeared more spherical and icosahedral features were not evident in these images. The sizes of the cages were of approximately the size expected for the assembly of TriEst into an icosahedral cage, which was estimated to be ~ 27 nm in diameter based on the structure of the TriEst trimer when arrayed in icosahedral geometry. These findings were consistent with the flexible nature of the design, as the oligo-Gly spacer sequence has the potential to span a length of ~ 25 Å. However, it is also possible that some particles may have misassembled and contain either fewer or greater than the intended 60 protein subunits. Higher resolution images were obtained using scanning transmission electron microscopy (STEM) (Fig. 4.6 C). These images revealed hexagonal morphologies in many of the particles, which is consistent with the projection onto a 2D surface of an icosahedron lying on one face.

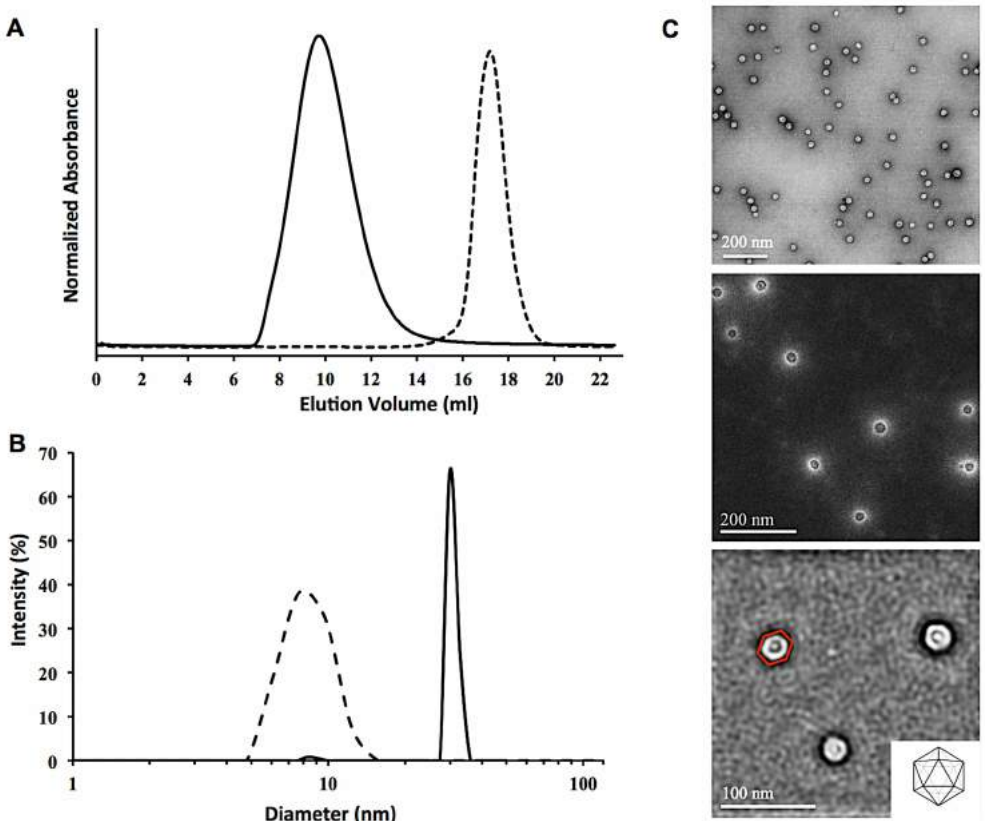

Figure 4. 6 Characterization of the protein cages formed by Ico8. (A) Size and homogeneity of Ico8 assessed by size exclusion chromatography: solid line, Ico8; dashed line TriEst; (B) Number-averaged diameter distribution of Ico8 cages measured by dynamic light scattering: solid line, Ico8; dashed line TriEst (C) *Top:* Conventional negative stain TEM of particles formed by Ico8; *middle*: high-angle annular dark-field scanning EM of Ico8 particles; *bottom*: bright field Fourier-filtered scanning EM images reveal hexagonal features associated with cages.

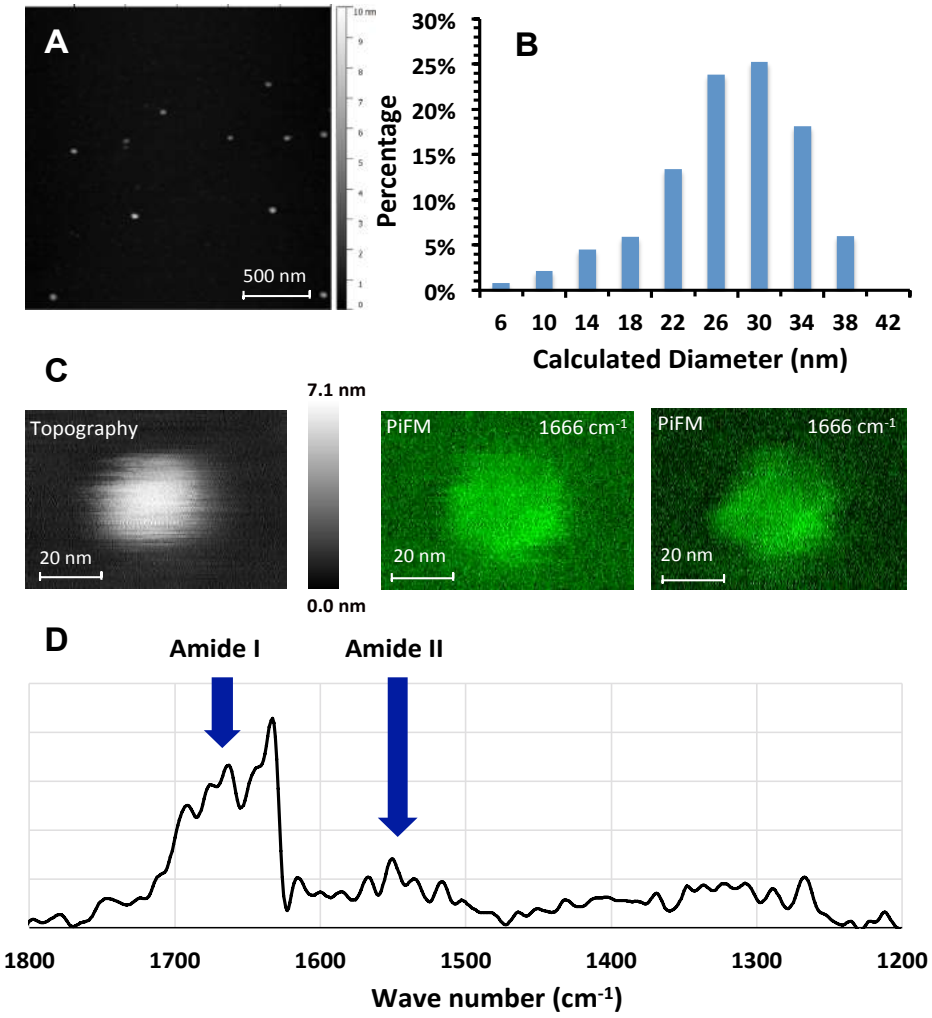

Figure 4. 7 Atomic force microscopy and IR-photo-induced force microscopy of Ico8 cages. (A) Representative field of view of Ico8 particles imaged by AFM. (B) Distribution of particle diameters calculated from AFM images. (C) *left:* Topography of a representative Ico8 particle imaged by conventional AFM; middle and *right:* images of Ico8 particles obtained by scanning PiFM at 1666 cm^{-1} (D) IR spectrum of an individual Ico8 particle.

4.2.2 Atomic Force Microscopy of Ico8

To obtain further information on the size of the cages formed by Ico8, we imaged the cages using atomic force microscopy (AFM, Fig. 4.7 A), which has the advantage of directly measuring the height and volume of particles. Particles were spin-coated onto a mica substrate as described in the Methods section and imaged

using a carbon nanotube (CNT) tip in AC mode in the air. ~ 1100 particles were imaged and subjected to automated particle analysis and classification (Fig. 4.7 B and Fig. 4.8). In air, the cages flattened onto the freshly cleaved mica surface resulting in an average particle height of 3.6 ± 1.1 nm. To compare the theoretically expected volume and the experimental volume obtained from cryo-electron microscopy (Cryo-EM), the measured particle volumes were used to calculate the diameter of an ideal icosahedral geometry. This yielded an average diameter of 24.9 ± 7.0 nm for the Ico8 cages, with the distribution of particle sizes being in good agreement with that measured by both cryo-EM and negative stain TEM. AFM characterization indicated a lower volume tail that may result from collaped protein cages during spin-coating.

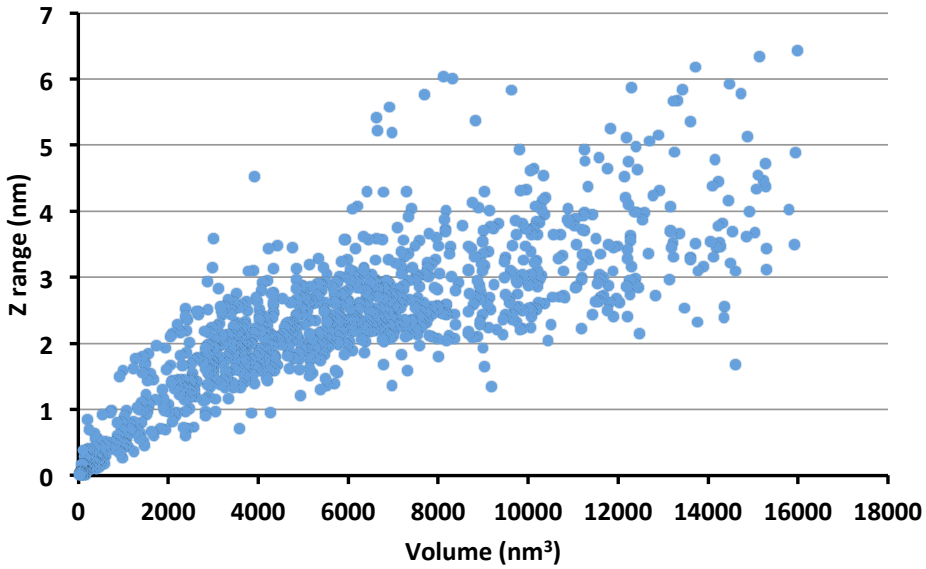

Figure 4. 8 Analysis of Ico8 particle size by atomic force microscopy. Plots of particle thickness (Z range) of Ico8 particles spin-coated on mica surfaces as of particle volume.

Further images and infrared (IR) spectra for individual Ico8 particles were measured using IR-photo-induced force microscopy IR-PiFM (Fig. 4.7 C,D). This technique records the vibrational spectrum of the material located in close proximity of an AFM tip operated in dynamic mode.[24,25] IR spectra recorded from individual particles showed characteristic amide-I and amide-II bands, confirming that the particles consist of proteins (Fig. 4.7 D). Images obtained using the amide I band at 1666 cm^{-1} were consistent with the formation of icosahedral cages and of similar dimensions to those obtained by AFM and TEM.

4.2.3 Cryo-electron Microscopy of Ico8

Further structural information on the cages formed by Ico8, was obtained by cryo-EM single particle analysis. A total of 18,707 particles were excised then subjected to reference-free classification and averaging using the program RELION.[47] 2D class-averaged cryo-EM images (Fig. 4.9) indicated that Ico8 forms hollow cages that range in diameter from 20 to 30 nm, with an average diameter of ~ 25 nm (Fig. 4.10 A,B). Unexpectedly, two concentric shells of density were observed in all the class averages, with most classes exhibiting short, inward-pointing features that appear to connect the inner and outer shells at multiple sites. The thickness and gap between the two shells appeared independent of the sizes of these cages (Fig. 4.11). Interestingly, these inward-pointing features were approximately the size of the pentameric coiled coil domain (Fig. 4.10 B).

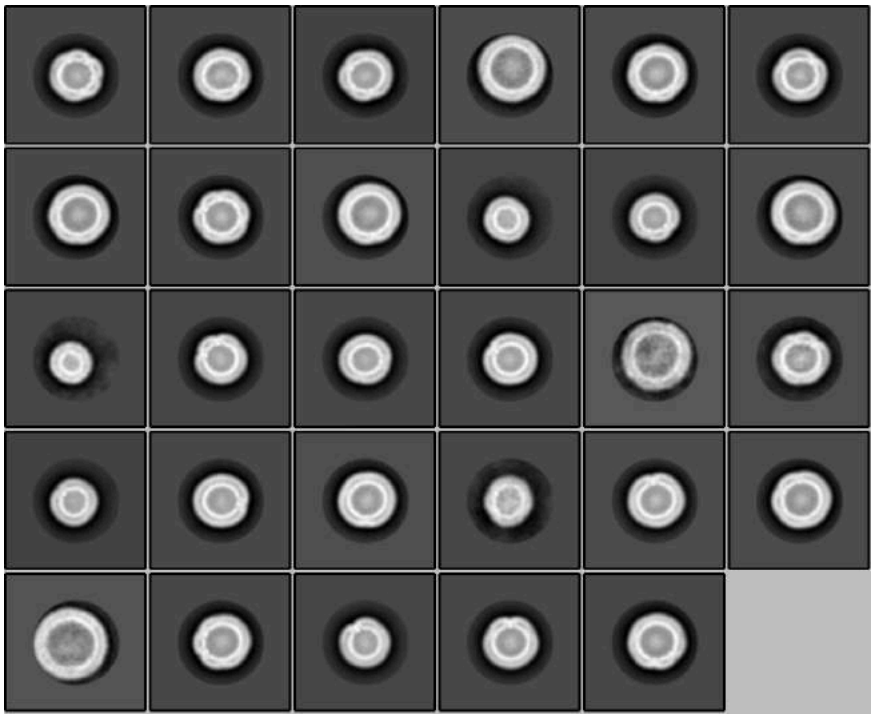

Figure 4. 9 Class averaged images of Ico8 particles determined by automated classification of Ico8 particles imaged by cryo-EM. All the class averages exhibit two concentric shells of density. Many class averages show connections between the inner and outer shells.

To obtain more detailed structural information on Ico8, we selected 10,363 particle projections, representing the average size of the protein cage, to calculate a 3D cryo-EM reconstruction of Ico8 using the program cisTEM.[20] Icosahedral symmetry was enforced during the refinement and the final reconstruction indicated a resolution of 12 Å. The symmetrically-reconstructed electron density map clearly reveals the arrangement of TriEst subunits that surround a region of inward-pointing density attributable to the 5-fold symmetric coiled coil domain (Fig. 4.10 C-E). This is consistent with the inward-pointing density seen in the 2D-class averages. The orientation of the coiled coil, inward or outward, was not specified in the design. So, although the cryo-EM reconstruction indicates an inward orientation,

it is possible that there is a minor population of coiled coils that point outwards but are not resolved by cryo-EM. The inner shell of electron density is also clearly evident in the reconstruction. The origin of this additional electron density is unclear.

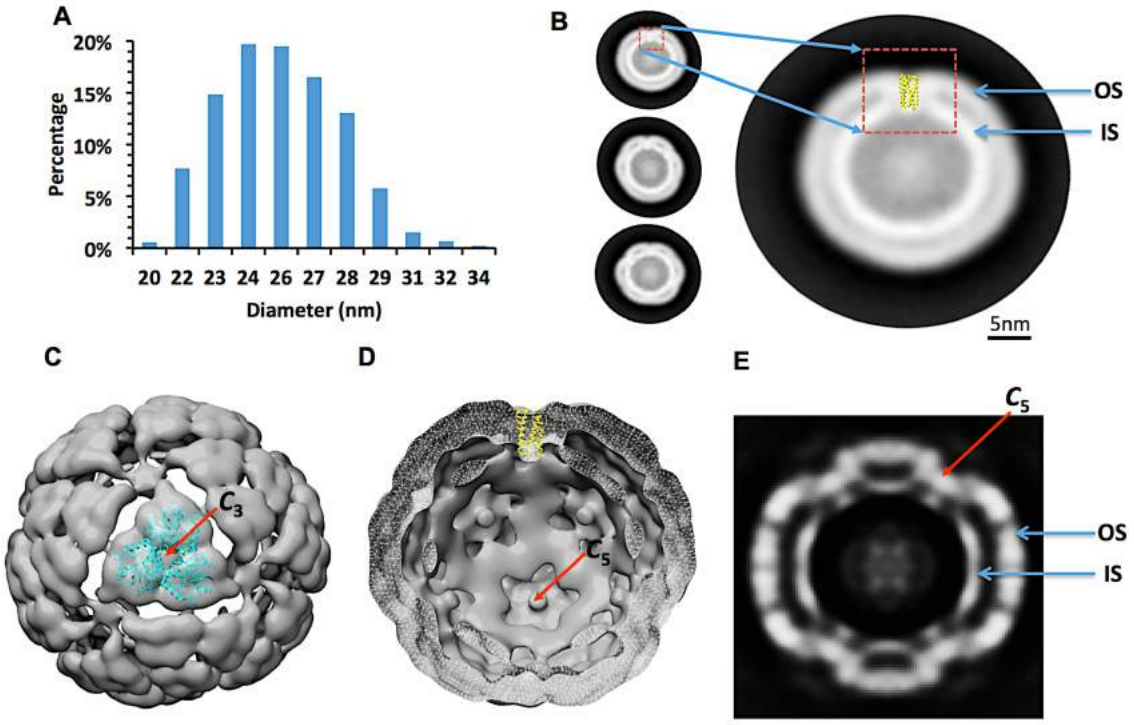

Figure 4. 10 Cryo-EM analysis of Ico8 cages. (A) Size distribution of Ico8 cage measured from 2D class averages. (B) Representative reference-free 2D class-averaged images of Ico8 cages illustrating the concentric shells of density exhibited by the cages and regularly spaced constrictions between the inner and outer walls that likely represent the coiled coil domain. The expanded view of one class-averaged image shows the coiled coil domain (yellow) superimposed on the density attributed the coiled coil. (Arrows indicate: IS, inner shell of density; OS, outer shell of density). (C, D) Cross-sections through the 3D electron density map of Ico8, with the crystal structures of TriEst (C) and the pentameric coiled coil (D) docked. E) Cross-sectional view of the density map perpendicular to 2-fold axis. The red arrow indicates the 5-fold symmetry element specified by the coiled coil.

It is evident that the electron density poorly accommodates the structure of TriEst, as the density is thinner and more spread out than expected (Fig. 4.10 C). This is attributable to the flexible nature of the protein cage attributable to the flexible glycine spacer connecting the two domains, which leads to smearing out of the averaged density. However, although of low resolution, the structure could be reproducibly reconstructed from the data. Importantly, the icosahedral features of the cage were maintained when the symmetry enforced in the reconstruction was relaxed from icosahedral to C_5 symmetry. This observation provides confidence that the icosahedral geometry is correct.

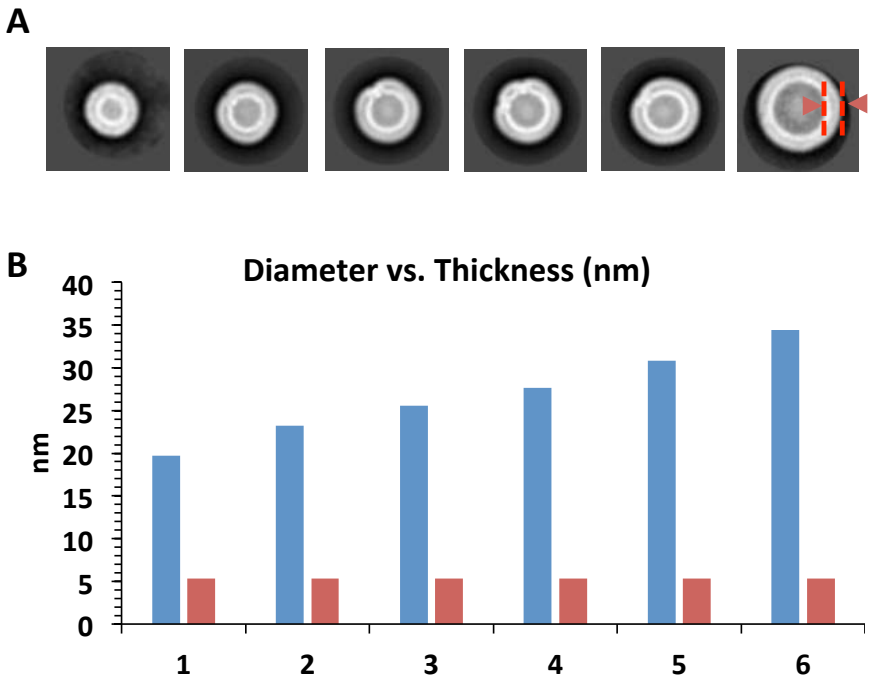

Figure 4. 11 Analysis of cage wall thickness as a function of cage diameter. (A) Representative class-averaged cryo-EM images of Ico8 cages of increasing diameter. Measurement of wall thickness is indicated on the image on the right. (B) Plots of diameter (blue bars) and wall thickness (red bars) for the class-averaged images shown in A.

Construct	Turnover (min^{-1})
TriEst	15.2 ± 0.72
TriEst (boiled)	N.D.
Ico8	21.5 ± 0.6
Ico8 (boiled)	24.2 ± 0.42

Table 4. 2 Enzymatic activity of TriEst and Ico8 determined with *2,4*-di-nitrophenyl acetate as substrate (N.D. – none detected).

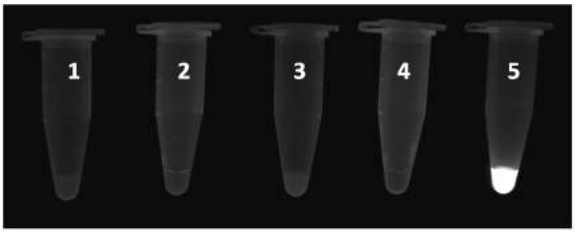

Figure 4. 12 DPH binding to Ico8. *Tube 1:* DPH in buffer; *Tube 2:* TriEst; *Tube 3:* TriEst and DPH; *Tube 4:* Ico8; *Tube 5* Ico8 and DPH.

4.2.4 Structural Integrity of Ico8

As a further check on the structural integrity of the cage, we measured the catalytic activity of the esterase building block and the ability of the coiled coil domain to bind hydrophobic fluorophores. The cages formed by Ico8 retained full catalytic activity, determined with *2,4*-di-nitrophenyl acetate as substrate (Table 4.2), indicating that the structure of the esterase is not compromised by the assembly process. The 5 helices of the pentameric coiled coil form a well-defined hydrophobic pore that has been shown to selectively bind hydrophobic fluorophores

such as diphenylhexatriene (DPH).[26] We exploited this property to evaluate the integrity of the coiled coil component of Ico8. Consistent with the coiled coil domain assembling correctly, Ico8 cages fluoresced when DPH was added to the buffer (Fig. 4.12), whereas TriEst exhibited no fluorescence in the presence of DPH.

4.2.5 Ico8 Binds DNA

Although neither TriEst nor the coiled coil bind nucleic acids, preparations of Ico8 always absorbed strongly at 260 nm (A_{260}:A_{280} ratios were typically between 1.6 – 1.8), which indicated that nucleic acids were associated with the cages. Native PAGE analysis of Ico8 demonstrated that the nucleic acids and Ico8 protein cage co-migrate, indicating that the nucleic acids are associated with the protein cage rather than adventitiously co-purifying (Fig. 4.5 B,C).

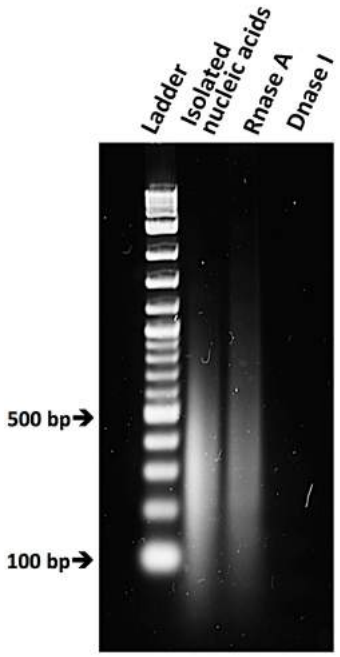

Figure 4. 13 Electrophoretic analysis of nucleic acid content of Ico8 cages. *Left to right* DNA ladder; phenol-chloroform extracted nucleic acids from purified Ico8 cages; nucleic acids after digestion with RNAse A; nucleic acids after digestion with DNAse I.

To further characterize the nucleic acids, the protein cages were extracted with phenol:chloroform to remove the protein and the nucleic acids analyzed by electrophoresis, either before or after digestion with RNase A or DNase I. The recovered nucleic acids fragments were approximately 300 – 500 bp in size and could be digested by DNase I treatment (but not RNase A treatment) indicating that they comprise DNA (Fig. 4.13). Quantification of the nucleic acid and protein content of the cages, determined by agarose gel electrophoresis and SDS-PAGE respectively, indicated that the nucleic acid component comprises ~ 30 % by weight of the purified Ico8 cages.

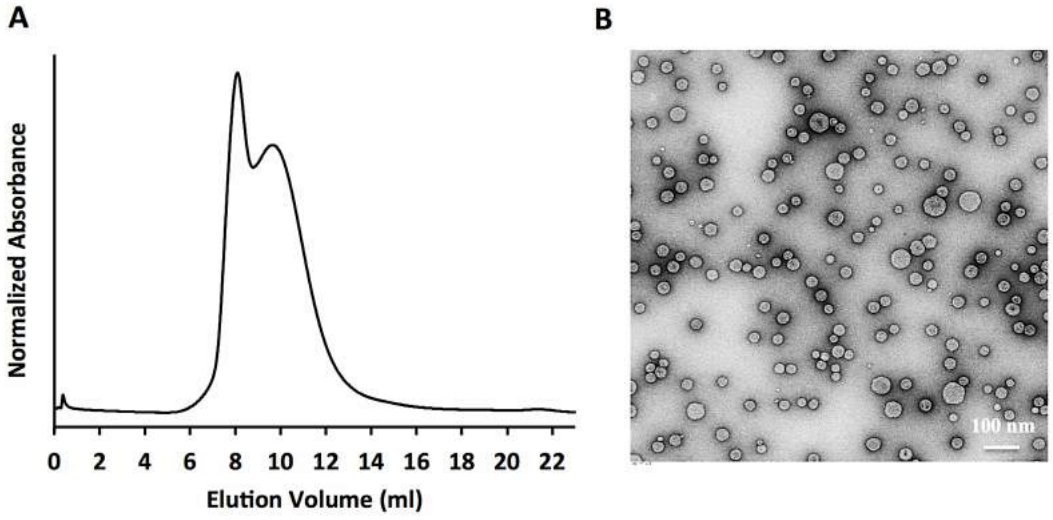

Figure 4. 14 Characterization of Ico8 cages purified from cells lysed using a microfluidizer that lack associated nucleic acids. (A) Size and homogeneity of Ico8 assessed by size exclusion chromatography. (B) Conventional negative stain TEM of Ico8 assemblies show them to be more heterogenous than those containing DNA fragments.

The short strands of DNA associated with Ico8 most likely derive from the sonication step used to lyse the cells during purification, because when cells were

lysed using a microfluidizer, which does not shear DNA so effectively, no nucleic acids were associated with the Ico8 protein purified from them. However, the Ico8 cages purified in this way were noticeably more heterogeneous in size as judged by SEC and negative stain TEM (Fig. 4.14).

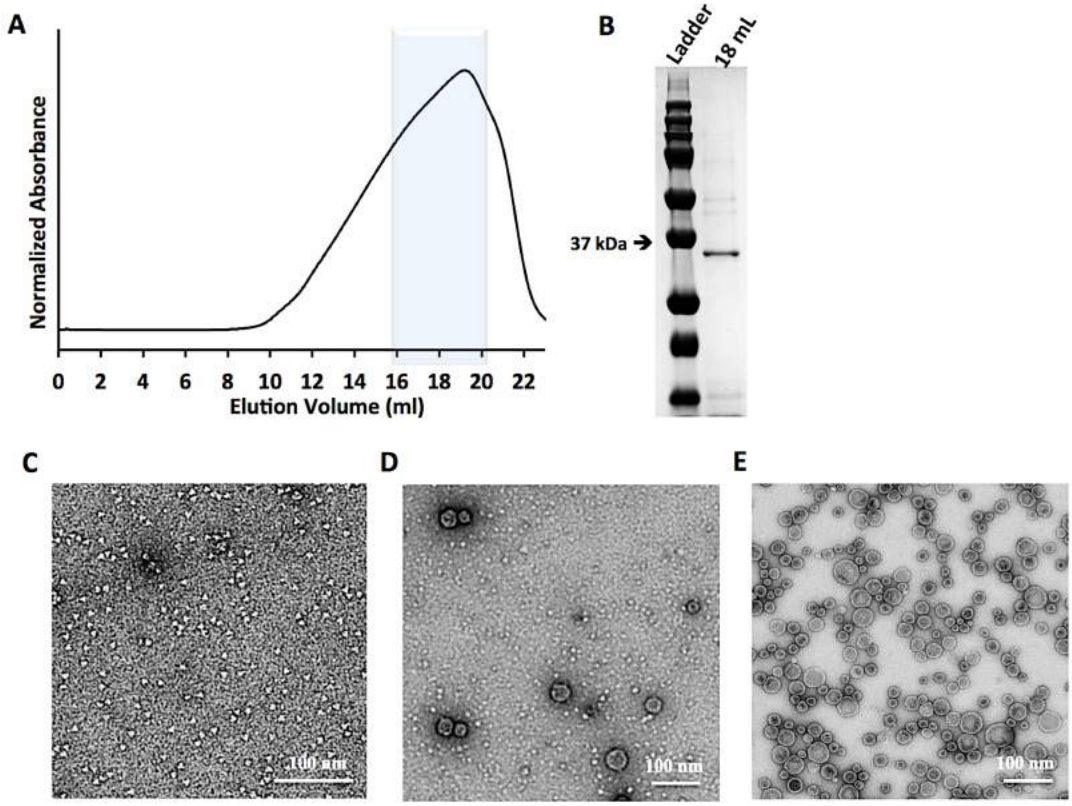

Figure 4. 15 Disassembly and reassembly of Ico8 in response to DNAse I digestion. (A) Size exclusion chromatography of Ico8 after treatment with DNAse I. The shaded fractions were pooled and analyzed further by SDS-PAGE and EM. (B) SDS-PAGE of DNAse I-treated Ico8 after SEC purification. (C-E) Negative stain TEM images of Ico8: completely disassembled trimers images immediately after purification (C); Partially reassembled cages imaged after 3 days (D); fully reassembled Ico8 cages imaged after 14 days (E).

To better understand the effect on nucleic acids on cage morphology, we attempted to remove the DNA from purified Ico8 cages by DNase I digestion. Overnight incubation at room temperature with DNase I removed the nucleic acids,

suggesting that the cages are sufficiently porous to allow access of the nuclease to the DNA. Interestingly, a *de novo* designed protein cage was recently described that encapsulated RNA and was, similarly, porous to RNase.[27] Surprisingly, digestion of the DNA also resulted in transient disassembly of the Ico8 cages. Thus, immediately after DNase I treatment, Ico8 appeared to be predominantly trimeric, as judged both by negative stain TEM and SEC (Fig. 4.15 A,C). However, after storage at room temperature for ~ 2 weeks almost all of the protein had reassembled into cages (Fig. 4.15 E). These were more heterogeneous in size than the original material and their morphology resembled the cages formed by Ico8 purified from cells lysed using the microfluidizer.

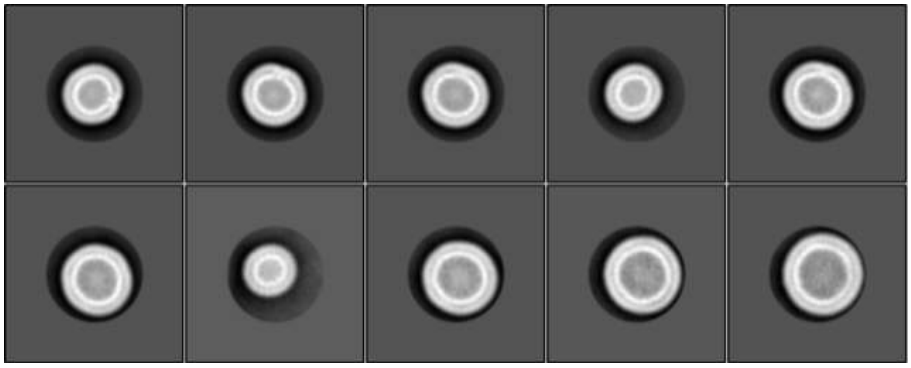

Figure 4. 16 Class averaged images of Ico8 particles after DNAse I digestion and re-assembly determined by automated classification of Ico8 particles imaged by cryo-EM. All the class averages exhibit two concentric shells of density, similar to Ico8 particles imaged before removal of DNA (see Fig. 4.9).

The reassembled cages were further analyzed by cryo-EM. 9380 particles were excised then subjected to reference-free classification as described above. The 2D class-averaged images (Fig. 4.16) show that the DNA-free protein forms similar hollow cages, although they appear slightly less regular in shape. These observations

suggest that DNA may exert an unanticipated templating effect on Ico8 leading to the assembly of more homogeneous cages.

As a further point of comparison, we also measured the zeta potential (ZP) of Ico8 cages to examine the surface charge of the particles. If DNA was bound to the exterior of the cages, then this should be reflected in a more negative zeta potential. ZP measurements were conducted in PBS buffer containing 20 mM ammonium acetate at pH 7.4 for Ico8 preparations that contained DNA and those for which DNA had been removed as described above and the cages allowed to reassemble. The ZP for both Ico8 samples were within error the same ζ = -21.8 ± 1.7 mV and ζ = -23.3 ± 2.6 mV for the DNA-containing and DNase I treated samples respectively. These measurements suggest that DNA is less likely to be associated with the outer surface of Ico8 cages.

4.2.6 Stability of Ico8

Many natural protein cages, e.g. viral capsids, exhibit high stability towards unfolding. We therefore examined how assembly of the TriEst building block into an icosahedral cage affected its stability and catalytic activity. Circular dichroism (CD) was used to follow the thermally-induced unfolding of both Ico8 and TriEst by monitoring the decrease in ellipticity at 222 nm (Fig. 4.17). Whereas TriEst irreversibly denatured between 70 °C and 80 °C, Ico8 remained folded at 98 °C (the highest temperature examined). The thermal stability of Ico8 was further investigated using differential scanning calorimetry (DSC), which permitted thermal unfolding of the protein to be studied at temperatures up to 120 °C at a pressure of 3

atm (Fig. 4.18). Again, TriEst underwent an irreversible exothermic transition between 70 °C and 80 °C, indicative of protein unfolding; however, Ico8 exhibited no thermal transitions up to 120 °C, implying that it remained folded. Consistent with this, no discernable change in morphology was apparent when the protein cages were imaged by negative stain TEM after cooling (Fig. 4.19). Significantly, the Ico8 cages remained fully catalytically active after heating and cooling (Table 4.2) and retained their ability to bind DPH (Fig. 4.20). In contrast, as expected, heating TriEst above its unfolding temperature precipitated the protein and completely inactivated it.

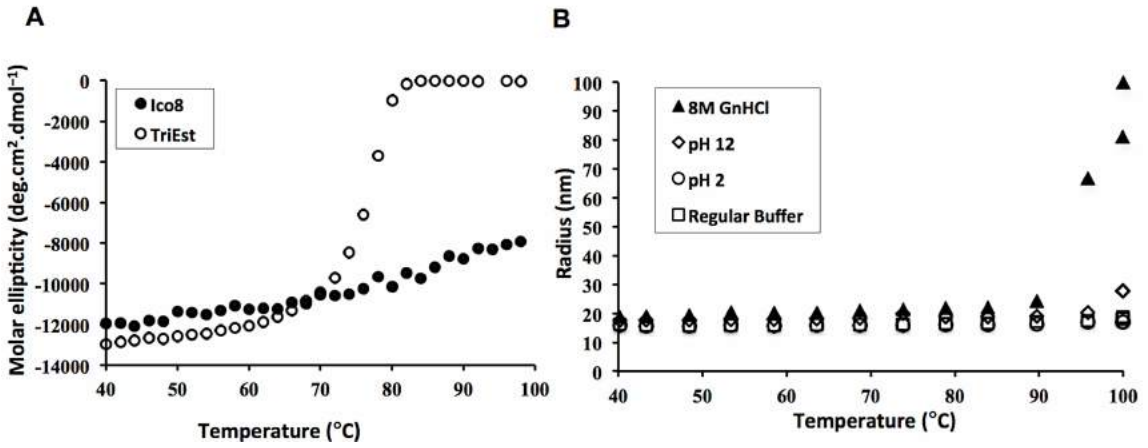

Figure 4. 17 Thermal and chemical stability of Ico8 protein cage. (A) The thermally-induced unfolding of TriEst and Ico8 monitored by changes in molar ellipticity at 222 nm. (B) Thermally-induced unfolding of Ico8 monitored by light scattering as a function of temperature at both high and low pH and in 8 M GnHCl.

The cages formed by Ico8 also proved extremely stable towards chemical denaturation. We compared the stability of Ico8 and TriEst at pH 2.0 and pH 12.0. At low pH the TriEst building block was unfolded and/or aggregated. However, both CD spectroscopy and DLS measurements indicated that the cages formed by Ico8

remain intact at extremely low as well as extremely high pHs. Similarly, Ico8 remained folded even in the presence of 8 M GnHCl, whereas TriEst was irreversibly unfolded by ~ 1.5 M GnHCl (Fig. 4.21). We extended these measurements by using temperature-dependent DLS to examine the thermal stability of Ico8 at pH 2.0, at pH 12.0, and in 8 M GnHCl (Fig. 4.17 B). The number-averaged diameter distribution of particles recorded by DLS provides a sensitive measure of protein aggregation, which is indicative of partial unfolding. At pH 2.0, Ico8 showed no evidence for unfolding/aggregation at temperatures up to 100 °C. At pH 12.0, or in 8 M GnHCl, an increase in particle diameter, indicative of aggregation, was only evident when the protein samples were heated to temperatures above 90 °C. Again, no discernable change in morphology was apparent when the protein cages were imaged by negative stain TEM after treatment with denaturants (Fig. 4.19). The only exception was samples heated in 8 M GnHCl, for which some aggregation was evident by EM (Fig. 4.19 G). Interestingly, when the cages heated in GnHCl were re-imaged after storing at room temperature for 2 weeks, the aggregated material was no longer evident although the cages appeared more heterogeneous in size (Fig. 4.19 H). The thermal unfolding of Ico8 cages from which nucleic acids had been removed by DNase I treatment was also examined by CD and DSC. Although more heterogeneous in size, these cages did not show any evidence for unfolding by either technique when heated to the maximum experimentally accessible temperature: 98 °C and 120 °C for CD and DSC respectively. (Fig. 4.22)

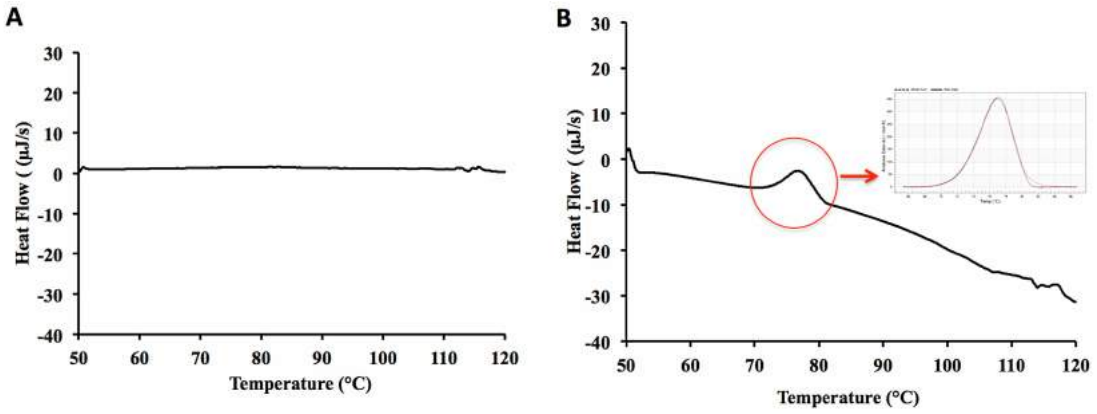

Figure 4. 18 Thermal unfolding of Ico8 and TriEst measured by scanning differential calorimetry. (A) Baseline subtracted plot of heat flow as a function of temperature for Ico8. (B) Baseline subtracted plot of heat flow as a function of temperature for TriEst; the unfolding transition is circled and the expanded trace shown inset.

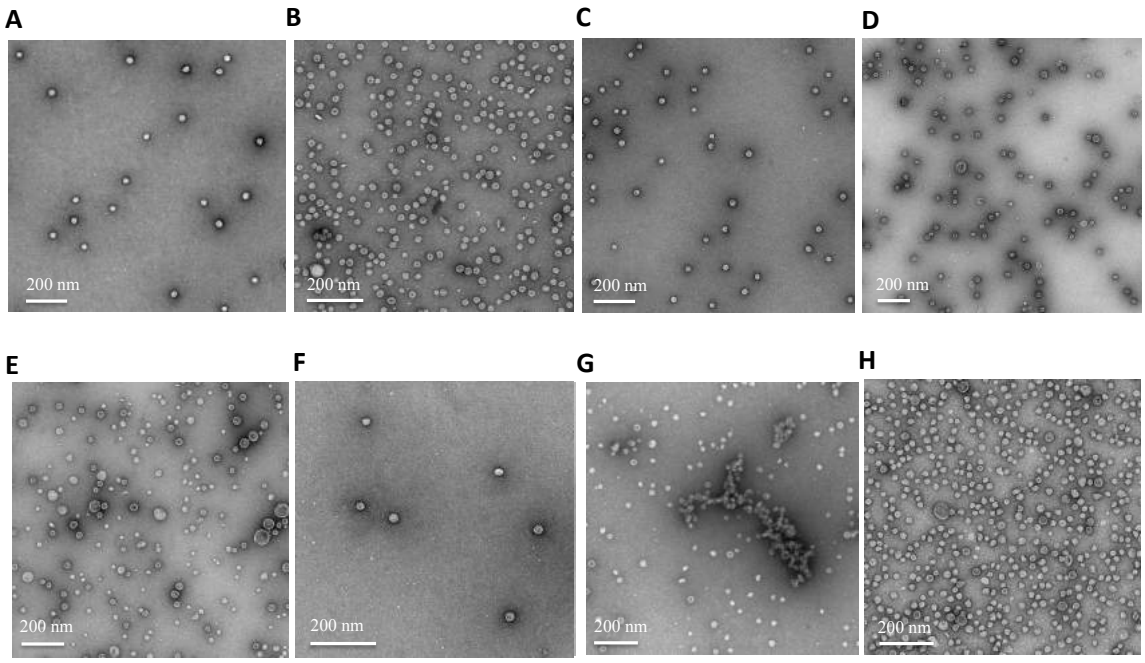

Figure 4. 19 Negative stain TEM images of Ico8 cages after heating and/or exposure to denaturants or extremes of pH. (A) pH 2, 25 °C; (B) pH 2, 95 °C; (C) pH 11, 25 °C; (D) pH 12, 25 °C; (E) pH 12, 95 °C; (F) 120 °C, 3 atm; G) 8 M GnHCl, 100 °C; and imaged after two days, (H) Same sample as G imaged after storing at 25 °C for 2 weeks.

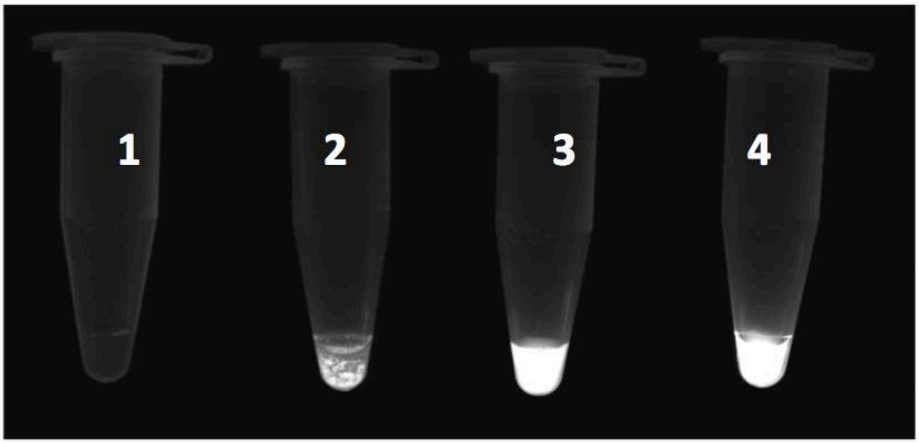

Figure 4. 20 Stability of Ico8 at higher temperatures assessed by DPH binding. *Tube 1* TriEST at 25 °C; *Tube 2* TriEst after boiling; *Tube 3* Ico8 at 25 °C; *Tube 4* Ico8 after boiling.

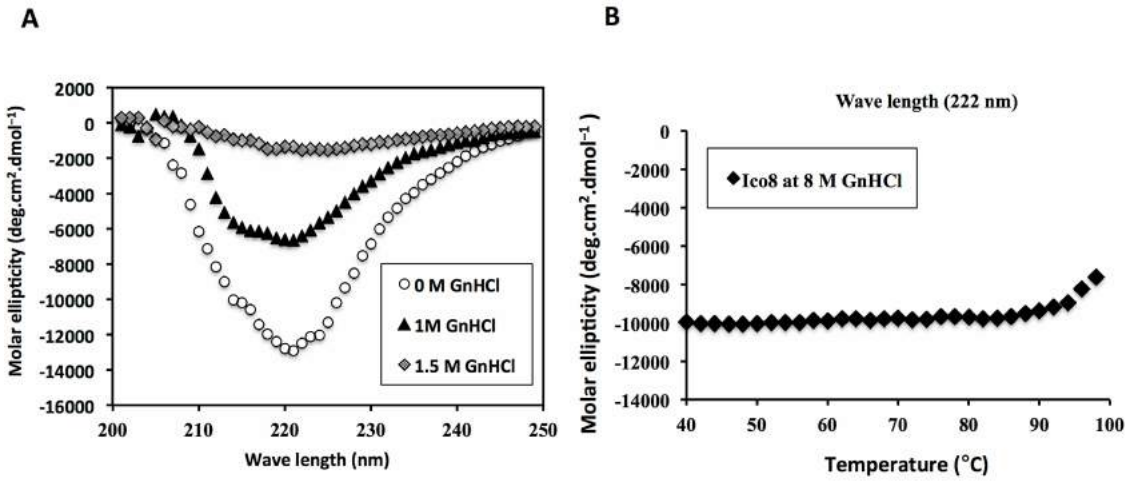

Figure 4. 21 Denaturation of TriEst and Ico8 by GnHCl. (A) Denaturation of TriEst at 0 M, 1 M and 1.5 M GnHCl followed by CD spectroscopy. (B) The thermally-induced unfolding of Ico8 at 8M GnHCl monitored by changes in molar ellipticity at 222 nm by CD spectroscopy.

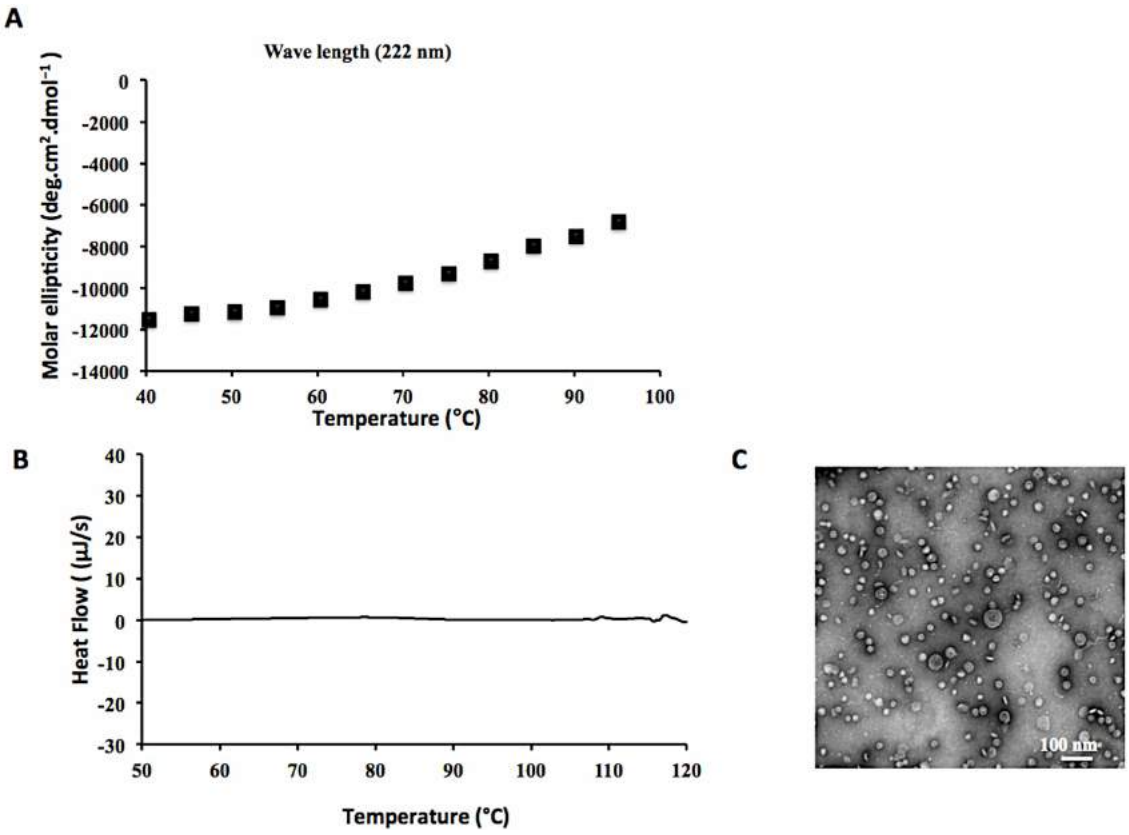

Figure 4. 22 Thermal stability of Ico8 cages re-assembled in the absence of DNA. (A) The thermally-induced unfolding of reassembled Ico8 monitored by changes in molar ellipticity at 222 nm by CD spectroscopy; (B) measured by scanning differential calorimetry; (C) Negative stain TEM of reassembled Ico8 after heating to 120 °C at 3 atm during DSC.

4.3 Discussion

Protein-based nanomaterials, formed by the ordered assembly of protein building blocks, have the potential to impact wide range applications in material science, biotechnology and medicine: many examples were discussed in chapter 1. For this potential to be fully realized, simple and generalizable approaches to protein assembly are needed. In this study we have demonstrated the assembly of a nano-scale protein container with a size comparable to a small viral capsid using a very simple, symmetry-based approach.

Remarkably, the combination of a C_3-symmetric building block protein and a C_5-symmetric coiled coil, with no other constraints, appears to be sufficient to assemble 20 copies of a trimeric enzyme into a 60-subunit icosahedral protein cage. We have now shown that it is possible to assemble the same trimeric esterase into tetrahedral, octahedral or icosahedral protein cages simply by choosing the appropriate coiled coil design. To achieve the efficient assembly of each of these cages it was necessary to optimize the linker length between the esterase and coiled coil domain, but this required screening only a small number of constructs (7 in this case) to identify a successful design. Thus, our design approach appears to be both an efficient and general method to assemble proteins that requires minimal computational modeling to implement. In comparison, designing new protein-protein interfaces to assemble proteins produces cages that are both more rigid and structurally better defined.[8,9] However, this precision requires extensive computational design methods that screen out many proteins that do not meet designability criteria and requires that many design variants be examined to identify soluble, correctly assembled cages.

Importantly, the assembly of Ico8 did not compromise the enzyme activity, which was actually slightly higher for Ico8 than the esterase from which Ico8 is assembled (Table 4.1). The relative ease with which tetrahedral, octahedral and icosahedral protein cages are formed suggests that natural protein cages may perhaps have evolved through a similar route involving gene fusion of C_3-symmetric proteins with C_3-, C_4- or C_5-symmetric proteins. If the cages conferred an advantage

on the organisms in which they arose, then further rounds of natural selection may have led to the development of more rigid cages with extensive protein interfaces.

An interesting and valuable feature of the cages formed by Ico8 is their extremely high thermal and chemical stability. Whereas TriEst is irreversibly unfolded at ~ 75 °C, and in the presence of 1.5 M GnHCl, the Ico8 cages remain intact at 120 °C and in 8 M GnHCl. This places Ico8 among some of the most thermostable proteins so far described, either natural or designed. The enzymatic activity of the cages is not affected by exposure to these harsh conditions, which is clearly a valuable feature for many biotechnological applications. An icosahedral protein cage was recently reported by Baker and co-workers that was thermostable to 90 °C, which was engineered by designing new subunit interfaces.[8] However, in this case the protein building block was obtained from a hyperthermophile and so was intrinsically thermostable. As shown in Fig. 4.17 A, TriEst is not itself very thermostable, posing a question as to the origin of the remarkable stability exhibited by Ico8. This may partly derive from the stability of the coiled coil domain, which was previously shown to withstand thermal unfolding up to 90 °C.[22]

Cooperative effects that arise from the increased number of inter-subunit interactions in the assembled cage also may contribute to Ico8's stability. Similar cooperative interactions between capsid proteins are believed to be responsible the high stability of many viral capsids, and studies on the assembly pathways of icosahedral viruses have highlighted the importance of cooperative effects in assembling viral capsids.[28–34] The strength of the protein-protein interactions is crucial to the successful assembly of the capsid: too weak and the virus will not

assemble; too strong and many kinetically-trapped intermediate structures accumulate. One advantage of using coiled coils to direct protein assembly is the ability to easily optimize the interaction strength by changing either the number of heptad repeats or the number of electrostatic interactions between adjacent α-helices.[35] The stability of viral capsids has also been suggested to be coupled with their high rigidity; however, our results suggest that, on the contrary, protein rigidity is not a prerequisite for assembling extremely stable protein cages.[28–30]

An unanticipated property of the cages formed by Ico8 is that they appear to bind nucleic acids. This was unexpected because neither TriEst, nor the coiled coil domain, bind nucleic acids. Furthermore, neither the octahedral nor tetrahedral cages that we previously designed bound nucleic acids.[11,12] It is possible that the small DNA fragments associated with the cage might be encapsulated within the cages, as the DNA co-migrates with Ico8 when subjected to native PAGE (Fig. 4.5). However, it was not possible to identify the location of the DNA in the cryo-EM images. The inner shell of electron density identified in the cryo-EM images is present in cages independently of whether DNA is present, indicating that this material is not DNA. We consider it more likely the additional material comprises proteins that are encapsulated, possibly by non-specific electrostatic interactions. Tentative support for this possibility is provided by the observation that Ico8 cage preparations always contained a background of contaminating proteins which are evident in SDS-PAGE gels (Fig. 4.5 A), but which are not resolved as separate species by native PAGE (Fig. 4.5 B).

The transient disassembly of the cages when the DNA was removed by nuclease digestion remains to be fully understood. However, we note that for many viruses, capsid assembly occurs concomitant with packaging their genomes. Furthermore, it has been found that the size of the genome packaged can alter viral morphology.[36,37] We suspect that in this case the nucleic acid fragments may help to template the assembly of Ico8, because although Ico8 preparations lacking nucleic acids assembled into cage-like structures, these particles were significantly more heterogenous, as judged by EM, than cages containing nucleic acids (Fig. 4.14). We suggest that electrostatic effects arising from the highly negatively charged DNA polymer, rather than specific DNA-protein interactions, underlie the templating effect of DNA. Further experiments are needed to examine this intriguing aspect of Ico8 assembly.

4.4 Conclusion

In conclusion, by selecting the appropriate coiled coil and optimizing the spacer length between the TriEst building block and the coiled coil domains, we have now been able to construct well-defined cages with tetrahedral, octahedral and icosahedral geometries. Taken together, these results demonstrate the modularity of this simple approach that relies on small off-the-shelf coiled coil sequences of the appropriate rotational symmetries to direct assembly. The icosahedral cage that we describe here represents the most challenging geometric structure to assemble as it is necessary to bring together 60 protein subunits. However, it is potentially the most useful protein cage because of the large internal volume enclosed. The extreme stability of the cage is an emergent property that is not inherent to the

enzyme used in its construction. This allows enzyme activity to be preserved under harsh conditions, which is a valuable property for any industrial process relying on enzymatic catalysis.

4.6 Acknowledgements

I acknowledge Junji Chen for performing AFM, Derek B. Nowak for and Sung I. Park PiFM, Kai Sun for STEM, Min Su for CryoEM particle analysis and 3D reconstructions and Amy L. Bondy for collecting CryoEM images.

This work was supported in part by grants from the National Institutes of Health (GM 093088) and the Army Research Office (W911NF-11-1-0251) to E.N.G.M. Transmission electron microscopy was performed at the Microscope Imaging Laboratory (MIL) core facility at University of Michigan. Scanning electron microscopy was performed at the Michigan Center for Materials Characterization at University of Michigan. Dynamic Light Scattering was performed at Biophysics Research Facility at University of Michigan.

4.7 References

1. Liu X, Theil EC (2005) Ferritins: Dynamic Management of Biological Iron and Oxygen Chemistry. Acc Chem Res 38:167–175.

2. Walsby AE (1994) Gas vesicles. Microbiol Rev 58:94–144.

3. Saibil H (2013) Chaperone machines for protein folding, unfolding and disaggregation. Nat Rev Mol Cell Biol 14:630–642.

4. Coux O, Tanaka K, Goldberg AL (1996) Structure and Functions of the 20S and 26S Proteasomes. Annu Rev Biochem 65:801–847.

5. Kerfeld CA, Aussignargues C, Zarzycki J, Cai F, Sutter M (2018) Bacterial microcompartments. Nat Rev Microbiol 16:277-290.

6. Aumiller WM, Uchida M, Douglas T (2018) Protein cage assembly across multiple length scales. Chem Soc Rev 47:3433–3469.

7. Ma Y, Nolte RJM, Cornelissen JJLM (2012) Virus-based nanocarriers for drug delivery. Adv Drug Deliv Rev 64:811–825.

8. Bale JB, Gonen S, Liu Y, Sheffler W, Ellis D, Thomas C, Cascio D, Yeates TO, Gonen T, King NP, et al. (2016) Accurate design of megadalton-scale two-component icosahedral protein complexes. Science 353:389–394.

9. Hsia Y, Bale JB, Gonen S, Shi D, Sheffler W, Fong KK, Nattermann U, Xu C, Huang P-S, Ravichandran R, et al. (2016) Design of a hyperstable 60-subunit protein icosahedron. Nature 535:136–139.

10. Jorda J, Leibly DJ, Thompson MC, Yeates TO (2016) Structure of a novel 13 nm dodecahedral nanocage assembled from a redesigned bacterial microcompartment shell protein. Chem Commun 52:5041–5044.

11. Sciore A, Su M, Koldewey P, Eschweiler JD, Diffley KA, Linhares BM, Ruotolo BT, Bardwell JCA, Skiniotis G, Marsh ENG (2016) Flexible, symmetry-directed approach to assembling protein cages. Proc Natl Acad Sci USA 113:8681–8686.

12. Badieyan S, Sciore A, Eschweiler JD, Koldewey P, Cristie-David AS, Ruotolo BT, Bardwell JCA, Su M, Marsh ENG (2017) Symmetry-Directed Self-Assembly of a Tetrahedral Protein Cage Mediated by de Novo-Designed Coiled Coils. ChemBioChem 18:1888-1892.

13. Zandi R, Reguera D, Bruinsma RF, Gelbart WM, Rudnick J (2004) From The Cover: Origin of icosahedral symmetry in viruses. Proc Natl Acad Sci USA 101:15556–15560.

14. Howorka S (2011) Rationally engineering natural protein assemblies in nanobiotechnology. Curr Opin Biotechnol 22:485–491.

15. Bauler P, Huber G, Leyh T, McCammon JA (2010) Channeling by Proximity: The Catalytic Advantages of Active Site Colocalization Using Brownian Dynamics. J Phys Chem Lett 1:1332–1335.

16. Smith ML, Lindbo JA, Dillard-Telm S, Brosio PM, Lasnik AB, McCormick AA, Nguyen LV, Palmer KE (2006) Modified Tobacco mosaic virus particles as scaffolds for display of protein antigens for vaccine applications. Virology 348:475–488.

17. Mastronarde DN (2005) Automated electron microscope tomography using robust prediction of specimen movements. J Struct Biol 152:36–51.

18. Zheng SQ, Palovcak E, Armache J-P, Verba KA, Cheng Y, Agard DA (2017) MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. Nat Methods 14:331–332.

19. Scheres SHW (2012) RELION: Implementation of a Bayesian approach to cryo-EM structure determination. J Struct Biol 180:519–530.

20. Grant T, Rohou A, Grigorieff N (2018) cisTEM, user-friendly software for single-particle image processing. eLife 7:e35383.

21. Elmi F, Lee H-T, Huang J-Y, Hsieh Y-C, Wang Y-L, Chen Y-J, Shaw S-Y, Chen C-J (2005) Stereoselective esterase from Pseudomonas putida IFO12996 reveals alpha/beta hydrolase folds for D-beta-acetylthioisobutyric acid synthesis. J Bacteriol 187:8470–8476.

22. Thomson AR, Wood CW, Burton AJ, Bartlett GJ, Sessions RB, Brady RL, Woolfson DN (2014) Computational design of water-soluble α-helical barrels. Science 346:485–488.

23. Cristie-David AS, Koldewey P, Meinen BA, Bardwell JCA, Marsh ENG (2018) Elaborating a coiled coil-assembled octahedral protein cage with additional protein domains: A multi-domain coiled coil-assembled protein cage. Protein Sci 27:1883-1890.

24. Nowak D, Morrison W, Wickramasinghe HK, Jahng J, Potma E, Wan L, Ruiz R, Albrecht TR, Schmidt K, Frommer J, et al. (2016) Nanoscale chemical imaging by photoinduced force microscopy. Science Adv 2:e1501571.

25. Rajapaksa I, Uenal K, Wickramasinghe HK (2010) Image force microscopy of molecular resonance: A microscope principle. Appl Phys Lett 97:073121.

26. Thomas F, Dawson WM, Lang EJM, Burton AJ, Bartlett GJ, Rhys GG, Mulholland AJ, Woolfson DN (2018) *De Novo* -Designed α-Helical Barrels as Receptors for Small Molecules. ACS Synth Biol 7:1808–1816.

27. Butterfield GL, Lajoie MJ, Gustafson HH, Sellers DL, Nattermann U, Ellis D, Bale JB, Ke S, Lenz GH, Yehdego A, et al. (2017) Evolution of a designed protein assembly encapsulating its own RNA genome. Nature 552:415–420.

28. Perlmutter JD, Hagan MF (2015) Mechanisms of virus assembly. Annu Rev Phys Chem 66:217–239.

29. Roos WH, Bruinsma R, Wuite GJL (2010) Physical virology. Nat Phys 6:733.

30. Roos WH, Gertsman I, May ER, Brooks CL, Johnson JE, Wuite GJL (2012) Mechanics of bacteriophage maturation. Proc Natl Acad Sci USA 109:2342–2347.

31. Moisant P, Neeman H, Zlotnick A (2010) Exploring the Paths of (Virus) Assembly. Biophys J 99:1350–1357.

32. Zlotnick A, Aldrich R, Johnson JM, Ceres P, Young MJ (2000) Mechanism of Capsid Assembly for an Icosahedral Plant Virus. Virology 277:450–456.

33. Garmann RF, Comas-Garcia M, Knobler CM, Gelbart WM (2016) Physical Principles in the Self-Assembly of a Simple Spherical Virus. Acc Chem Res 49:48–55.

34. Garmann RF, Comas-Garcia M, Gopal A, Knobler CM, Gelbart WM (2014) The Assembly Pathway of an Icosahedral Single-Stranded RNA Virus Depends on the Strength of Inter-Subunit Attractions. J Mol Biol 426:1050–1060.

35. Fletcher JM, Boyle AL, Bruning M, Bartlett GJ, Vincent TL, Zaccai NR, Armstrong CT, Bromley EHC, Booth PJ, Brady RL, et al. (2012) A Basis Set of *de Novo* Coiled-Coil Peptide Oligomers for Rational Protein Design and Synthetic Biology. ACS Synth Biol 1:240–250.

36. Zlotnick A, Porterfield JZ, Wang JC-Y (2013) To Build a Virus on a Nucleic Acid Substrate. Biophys J 104:1595–1604.

37. Kler S, Wang JC-Y, Dhason M, Oppenheim A, Zlotnick A (2013) Scaffold Properties Are a Key Determinant of the Size and Shape of Self-Assembled Virus-Derived Particles. ACS Chem Biol 8:2753–2761.

Chapter 5: Designing a Metal-dependent Protein Cage

5.1 Introduction

Our research group showed that protein cages can be designed *de novo* by fusing a natural protein with a rotational symmetry (C_n) to a homo-oligomeric parallel coiled coil through a flexible oligo-glycine linker. By pairwise combination of these two elements various geometries can be generated. Using this approach, our group successfully designed a tetrahedral, an octahedral and an icosahedral protein cages by fusing a timeric natural protein to a trimeric coiled coil, tetrameric or a pentameric coiled coil respectively, which demonstrated the generalizability of the design approach.[1,2] Next, we wanted to extend our coiled coil based strategy to design protein cages that assemble in response to environmental stimuli.

Multimeric proteins such as actin, vertebrate nuclear pore complex and some viral capsids undergo environmentally responsive assemblies and disassemblies.[3–5] Inspired by Nature, various strategies have been developed to control the assembly of extended protein-based nanomaterials (filaments, 2D lattices and 3D crystals); for example by incorporating di-sulfide linkages, small molecules and metal ions to mediate assembly.[6–11] The interfaces of natural protein cages have been engineered to undergo environmentally responsive assembly-disassembly.[12–14] For example a Cu(II) responsive ferritin cage has been designed by introducing His residues at the

inter-subunit interface of ferritin.[13] A pH responsive ferritin cage has been engineered by introducing GALA peptide, a pH responsive amphipathic helix, to protein-protein interface.[14]

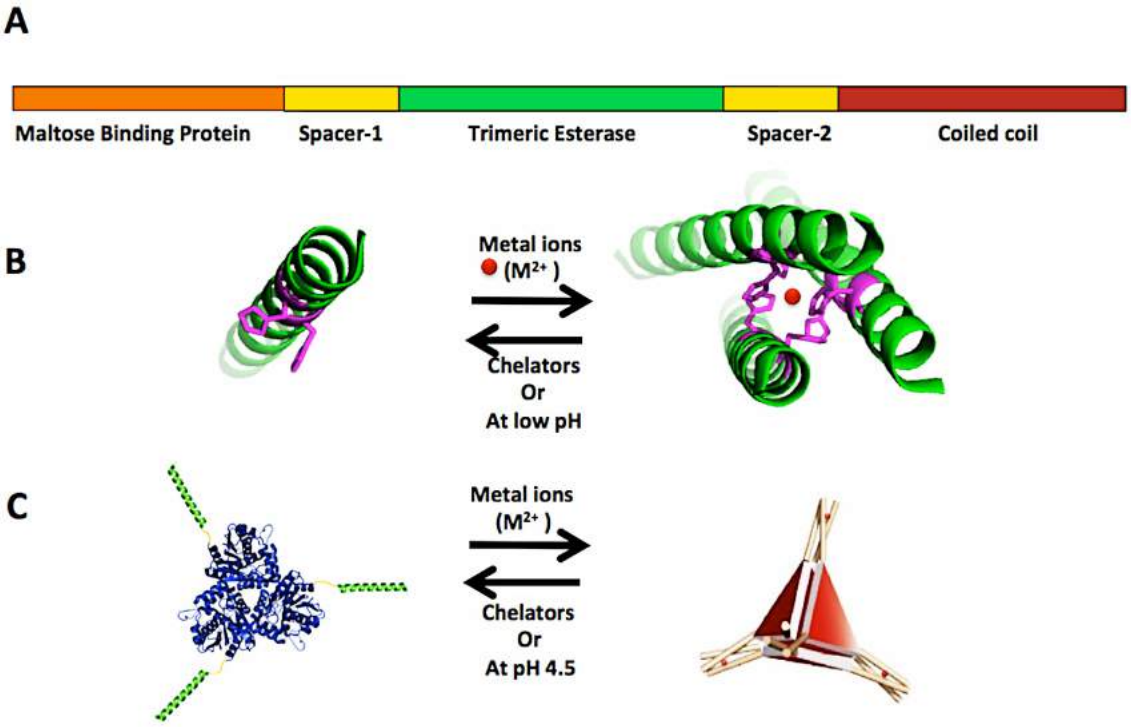

Figure 5. 1 The metal-dependent protein cage design strategy. (A) Genetic fusion. (B) Metal-dependent coiled coil motif. (C) A cartoon illustrating the controlled assembly of the protein design.

Re-engineering interfaces is challenging as modifications in these regions are likely disrupt inter-subunit interactions. Indeed re-engineering interfaces to achieve environmentally responsive interfaces in *de novo* designed protein cages has not been reported. However, various environmentally responsive coiled coil designs have been reported, that undergo assembly only under certain conditions (E.g. pH,

presence of metal ions).[15–17] Therefore, our coiled coil based design strategy can potentially be utilized as a promising route for controlling the assembly of *de novo* protein cages by overcoming these limitations.

Here, we have evaluated the design of a metal-dependent tetrahedral protein cage, which was constructed by fusing a trimeric esterase from *Pseudomonas putida* (TriEst, PDB ID: 1ZOI) to a trimeric metal-dependent coiled coil (Fig. 5.1). We utilized a metal-dependent coiled coil design (IZ-3adH) that was reported to assembled into a homo-trimeric parallel coiled coil in the presence of divalent transition metal ions and disassembled in the absence of metal ions or at low pH.[17] In the presence of divalent metal ions, this construct assembled into discrete, high M_w materials consistent with a tetrahedral protein cage. Adding metal chelators (E.g. EDTA) or decreasing pH led to disassembly of the cages into their trimeric form. This coiled coil-based design strategy therefore provides a simple and generalizable route for *de novo* designing protein cages with controlled assemblies; a challenging problem in protein nanotechnology.

5.2 Materials and Methods

5.2.1 Construction of Genes Encoding Fusion Proteins

A metal-dependent coiled coil with a 8-residue spacer at its 5' end was commercially synthesized. This ds-DNA sequence was genetically fused to 3' end of TriEst in expression vector pET-28b plasmid, using Gibson assembly protocol. This fusion construct was named T8H4-1. To design T8H4-2 and T8H4-3 constructs, ds-DNA with a TEV protease cleavage site and ds-DNA with a TEV protease cleavage

site and oligo-Gly spacers respectively were commercially synthesized. These ds-DNA sequences were fused genetically to 5' end of TriEst in expression vector pET-28b plasmid, using Gibson assembly protocol. The complete sequences of T8H4-1,-2 and -3 are given in Appendix A.4.

To design MBP-T8H4, His-tag at N-terminus of TriEst was replaced by MBP as follows. The MBP gene was PCR-amplified from pMAL-c5X using commercially synthesized primers. The 5' end of the forward primer was designed to overlap with the sequence encoding the T7 promoter region of pET-28b at its 3' end. The reverse primer was designed to overlap with the spacer (spacer 1 sequence: GGGGGGENLYFQGGGGH) region between MBP and TriEst. The codon-optimized gene encoding the 17-residue long spacer 1 with a TEV protease cleavage site was commercially synthesized. Both the MBP construct and spacer 1 were introduced into 5' end of the above TriEst-metal-dependent coiled coil fusion construct using Gibson assembly protocol. The complete sequence of the MBP-T8H4 design is given in Appendix A.4.

5.2.2 Protein Expression and Purification

Proteins were expressed as described in chapter 2 (see 2.2.2). T8H4-1 and T8H4-2 were purified by Ni-column chromatography as described in chapter 2 (see 2.2.2). MBP-T8H4 was purified by maltose-affinity chromatography as described in chapter 3 (see 3.2.2). Fractions containing the desired protein (approximately 10-15 mL) were pooled, combined and treated with 2 μL of benzonase. T8H4-1 was dialyzed against buffer containing 20 mM HEPES, 100 mM NaCl, 2 mM EDTA and 10% glycerol approximately for 1 day at 4 °C. MBP-T8H4 was dialyzed against

buffer containing 20 mM HEPES, 100 mM ammonium acetate, 2 mM EDTA and 10% glycerol approximately for 2 days at 4 °C.

Size exclusion chromatography (SEC) was performed for both T8H4-1 and MBP-T8H4 as an additional purification and buffer exchange step as well as an analytical step. SEC was performed on a superose 6 10/300 column equilibrated with Tris HCl SEC buffer (10 mM Tris HCl, 100 mM NaCl, pH 8) at 4 °C. 400 µL samples with ~ 2-3 mg/mL concentration were injected at 0.3 mL/min flow rate.

Fractions at elution volume 17 mL for T8H4-1 injections and fractions at elution volume 14.5 mL for MBP-T8H4 injections were pooled and stored at 4 °C in the column elution buffer. If needed, protein was concentrated using 100-kDa Amicon ultra-centrifugal filter units.

5.2.3 Sample Preparation and SEC Procedure for Studying the Controlled Assembly

200 µL samples of 20 µM T8H4-1 was equilibrated with 40 µM and 100 µM $NiCl_2$ and incubated over night (o/n) and re-chromatographed on a superose 6 10/300 column equilibrated with same TrisHCl SEC buffer containing metal ions at same concentrations. 200 µL samples of 20 µM MBP-T8H4 was equilibrated with various concentrations of $NiCl_2$, $CoCl_2$, $CuCl_2$ and $ZnCl_2$ to evaluate the metal-dependent assembly of protein cages (concentrations are given in the Results section). These samples were incubated at 4 °C for 2 h, o/n or o/n while gently rocking. Samples were injected to a superose 6 10/300 column equilibrated with TrisHCl SEC buffer at same metal ion concentrations. The fractions from elution

volume 10.5 mL were re-chromatographed under the same buffer conditions.

In order to further characterize assembled MBP-T8H4, 400 µl of 50-200 µM protein was incubated with 200 µM $NiCl_2$ o/n at 4 °C with gently agitating and purified by SEC using a superose 6 10/300 column equilibrated with Tris HCl containing 200 µM $NiCl_2$. The fractions at elution volume 10.5 mL were pooled and used for further characterizations by the techniques described below.

5.2.4 Thrombin Cleavage of T8H4-1

0.2 mg of protein was incubated o/n with 3 units of thrombin from Bovine plasma (Sigma-Aldrich) to remove the His-tag and cleavage was assayed by SDS-PAGE analysis.

5.2.5 Cross-linking with BS3

Assembled MBP-T8H4 were cross-linked using the lysine-specific reagent (bis(sulfosuccinimidyl)suberate) (Thermo Fisher) as described in chapter 3 (see 3.2.8).

5.2.6 Negative stain TEM imaging

Negative stain TEM was performed as described in chapter 3 (see 3.2.5).

5.2.7 Activity assays

Catalytic activity for TriEst and, assembled and unassembled MBP-T8H4 were assessed using *2,4*-di-nitrophenyl acetate (*2,4*-DNPA) as the substrate and at 45 °C as described in chapter 3 (see 3.2.6). The assays were performed in Tris HCl SEC buffer. Absorbance was measured at 405 nm.

5.2.8 Dynamic Light Scattering

DLS was performed using DynaPro NanoStar ZS instrument as described in chapter 3 (see 3.2.7). Protein concentrations were kept at 20 μM.

5.3 Results

5.3.1 Metal-dependent Protein Cage Design

TriEst (PDB ID: 1ZOI) was selected as the primary building block protein (BBP), to facilitate comparison with our previous studies. In order to design the metal-dependent protein cage, we genetically fused IZ-3adH, a *de novo* designed metal-binding 4-heptad trimeric coiled coil (Sequence: IEKKIEA IEKKIEA **H**EKK**H**EA IEKKIEA) to the C-terminus of TriEst through a flexible 8-residue Gly linker (GTGGGGGG). IZ-3adH incorporates His residues at *a* and *d* positions in the 3rd heptad repeat of the trimeric coiled coil to provide a 6-cordinate His-X3-His site.[17] Oligomerization of the construct into a parallel trimeric coiled coil in the presence of metal ions has been evaluated by sedimentation equilibrium analytical ultracentrifugation (SE-AUC), circular dichroism (CD), fluorescence quenching and NMR.[17] These studies confirmed that the construct assembles into a trimeric coiled coil in the presence of divalent transition metals such as Ni(II), Co(II), Cu(II) and Zn(II) and returns to an unstructured monomeric form in the absence of metals or at acidic pH.

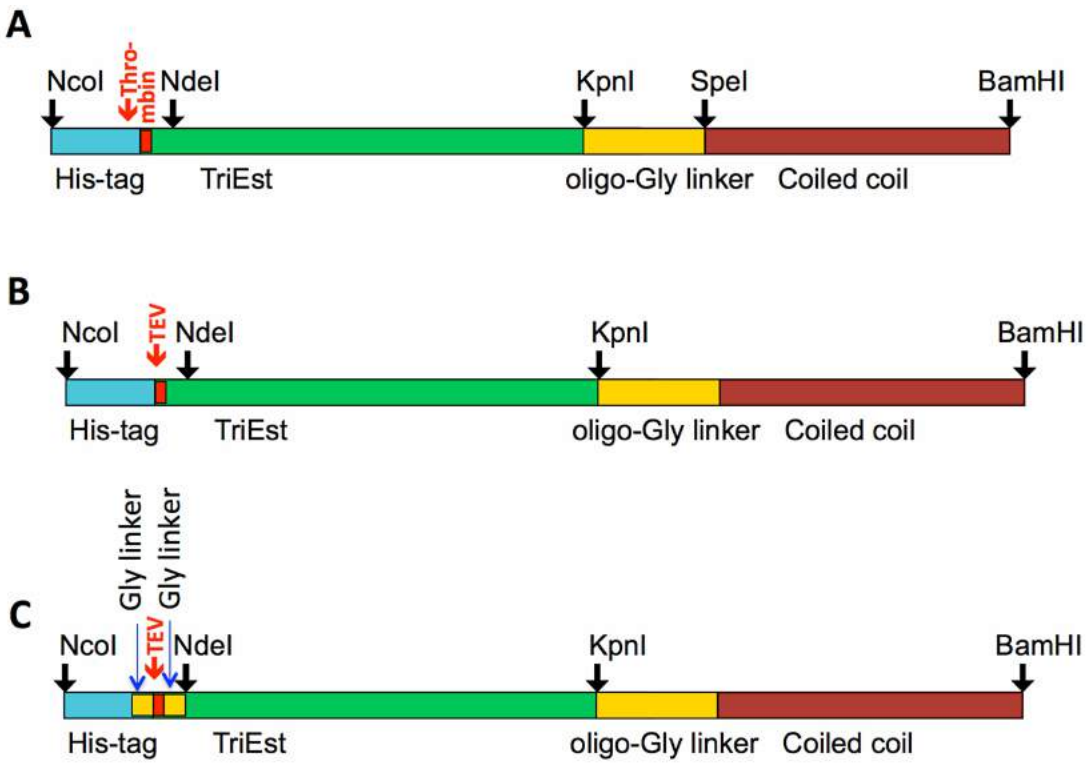

Figure 5. 2 T8H4 fusion constructs. (A) T8H4-1; (B) T8H4-2; (C) T8H4-3.

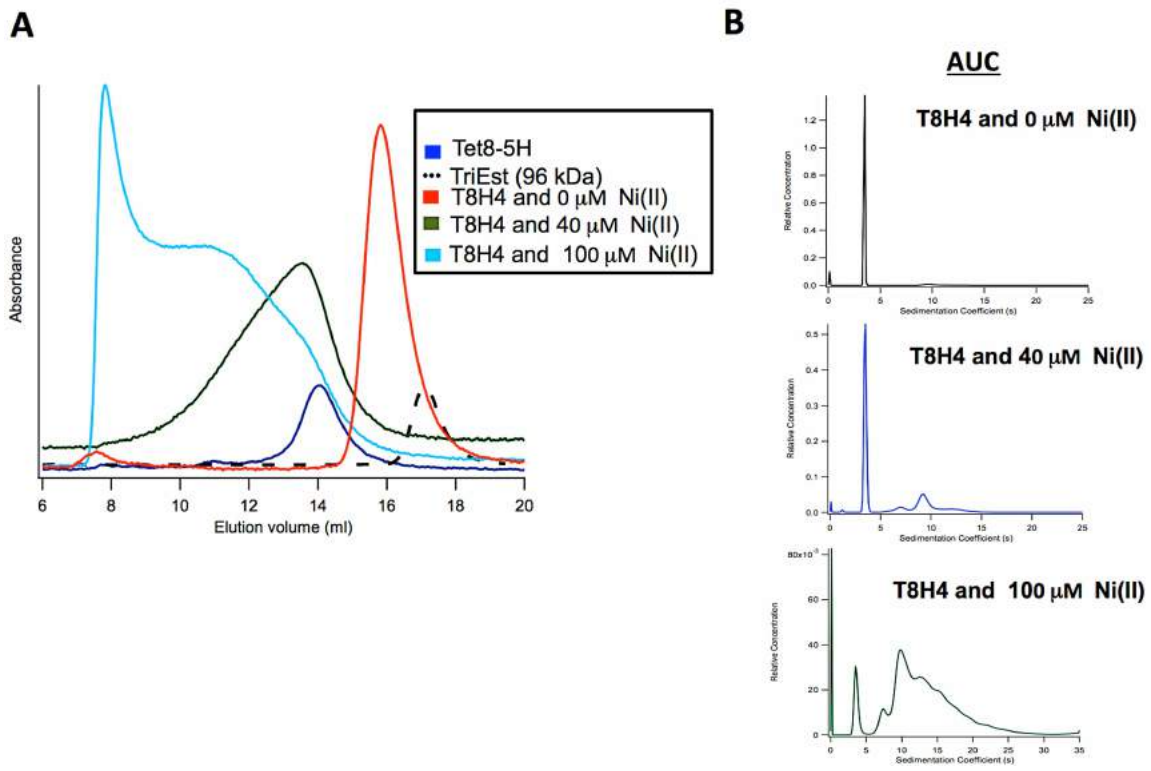

Figure 5. 3 Characterization of T8H4-1. (A) Size exclusion chromatography of T8H4-1. Tet8-5H tetrahedral cage and TriEst were used as standards and which eluted at 14 mL and 17 mL respectively; T8H4-1 trimer eluted at 16 mL; In the presence of Ni(II) T8H4-1 gave broad, asymmetrical elution profiles indicating a heterogeneity. (B) Sedimentation profile analysis of T8H4-1.

Our original tetrahedral design was optimized with a 8-residue linker, therefore the same linker was incorporated between the TriEst and the metal-dependent coiled coil. Three fusion constructs (T8H4-1/-2/-3) were initially evaluated and the sequences are given in appendix A.4 (Fig. 5.2). The first construct was named T8H4-1. We presumed that, in the presence of above metal ions, trimerization of coiled coil domain in this fusion construct will bring together four copies of the TriEst and assemble into a tetrahedral protein cage. In the absence of metal ions, the construct stayed in its trimeric form as evidenced by SEC and AUC

data: the protein eluted at elution volume 16 mL on a superose 6 10/300 column and gave a sedimentation coefficient of ~ 4, which is consistent with the expected molecular weight (Fig. 5.3). 20 μM of protein was equilibrated with 40 μM and 100 μM Ni(II) and studied by both SEC and AUC. In the presence of Ni(II), T8H4-1 gave broad and asymmetric size exclusion and sedimentation profiles indicating the heterogeneity of assembled species (Fig. 5.3). A His-tag is placed at N-terminus of TriEst as a purification tag. We questioned whether the heterogeneity of protein assemblies resulted from interactions of Ni(II) with these His-residues. We thus attempted to remove the His-tag by using the thrombin cleavage site located between His-tag and N-terminus of TriEst. However, thrombin cleavage was unsuccessful as evident by SDS PAGE (Fig. 5.4).

Therefore, T8H4-2 fusion construct was designed by replacing the thrombin cleavage site with a TEV cleavage site in order to remove the His-tag followed by purification of the protein by Ni-affinity chromatography. However, this fusion construct was not expressed: no post-induction protein band was observed by SDS-PAGE (Fig. 5.5 A,B). Western Blot analysis of pre- and post-induction samples of T8H4-2 stained for His-tag further confirmed that T8H4-2 construct does not express (Fig.5.5 C). Original TriEst also failed to express when its thrombin cleavage site was replaced with TEV. Therefore, a third fusion construct was designed by adding oligo-Gly spacers in each side of TEV cleavage site (Sequence: GGGGGGGENLYFQGGG) in an attempt to express the protein, the construct was named T8H4-3 (Fig. 5.2 C). However, this fusion construct also did not express and original TriEst with the same genetic modification behaved similarly. The reason

why these fusion constructs were not expressed when the thrombin site was replaced with TEV protease cleavage site is unclear.

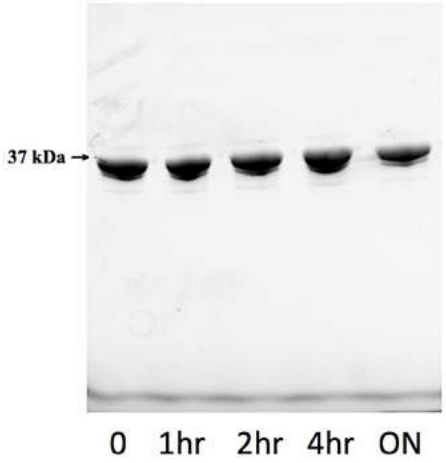

Figure 5. 4 SDS-PAGE analysis of T8H4-1 incubated with thrombin. No cleavage was evident.

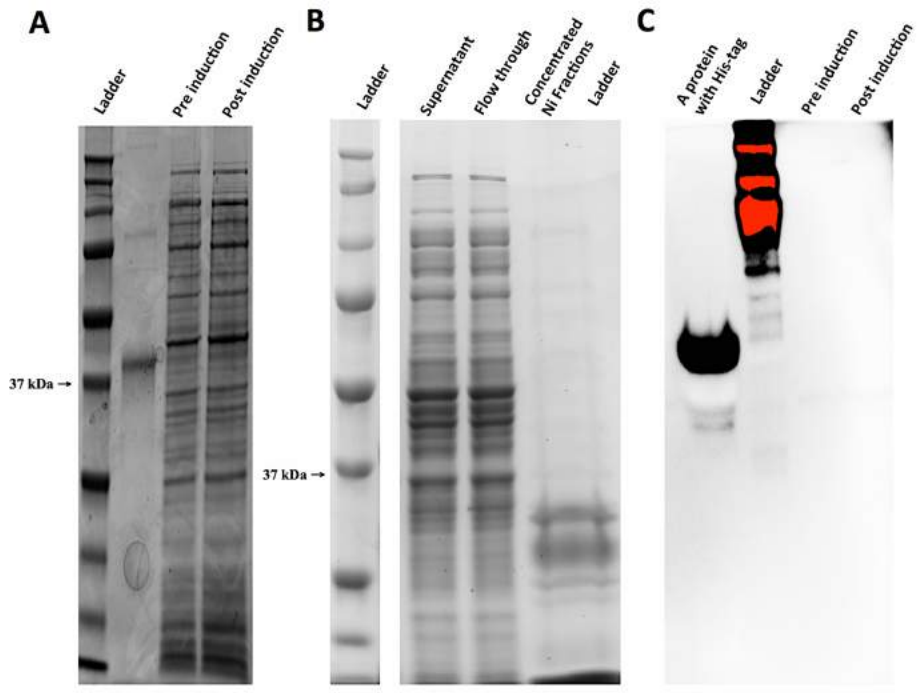

Figure 5. 5 Analysis of T8H4-2. (A) SDS PAGE analysis of pre- and post-induction samples of T8H4-2 (B) SDS PAGE analysis of supernatant, flow through and concentrated elution fractions from Ni-column chromatography of T8H4-2. (C) Western Blot analysis of pre- and post-induction samples of T8H4-2, visualized by staining with anti-His antibody.

Next a fourth fusion construct was designed by replacing the His-tag with maltose binding protein (MBP) domain to facilitate purification (Fig. 5.1 A). A 17-residue spacer (Spacer 1: GGGGGGGGGENLYFQGG) was genetically introduced between the MBP domain and TriEst to provide enough freedom for these two protein domains to fold while minimizing possible steric clashes (the N-terminus of TriEst is located on one of the faces of TriEst and is in closer proximity to its rotational axis). A TEV cleavage site was placed in this spacer region to facilitate the removal of MBP tag if needed. This fusion construct was named as MBP-T8H4.

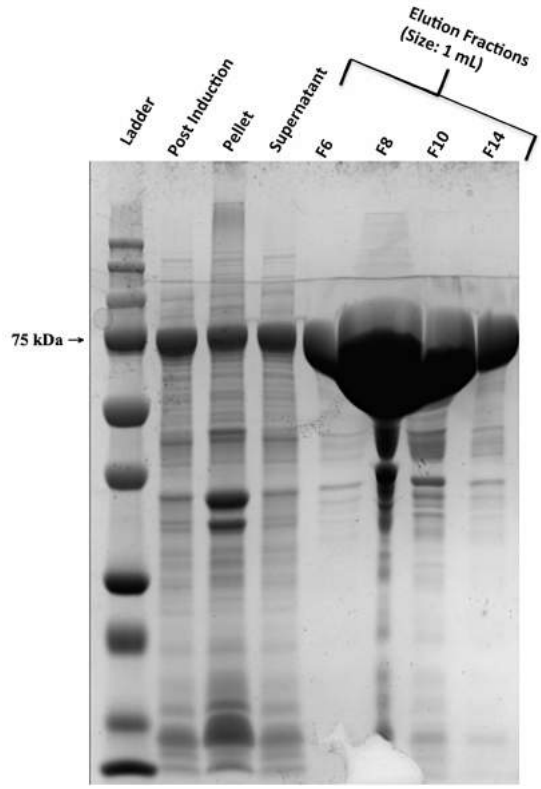

Figure 5. 6 SDS-PAGE analysis of MBP-T8H4 samples from post-induction and maltose-affinity chromatography.

MBP-T8H4 was overexpressed in *E.coli* and, purified by MBP affinity column chromatography (Fig. 5.6). The protein was purified to homogeneity as verified by SDS-PAGE and with a yield of ~ 400 mg/L culture. The purified protein was dialyzed over night at 4 °C in buffer containing 2 mM EDTA to remove any bound metal ions. The protein was further purified by size exclusion chromatography (SEC) using a superose 6 10/300 column and the protein was eluted at 14.5 mL in the column as a single peak, which is consistent with the expected M_w = 227.4 kDa for the trimeric form of the protein (5.7 A).

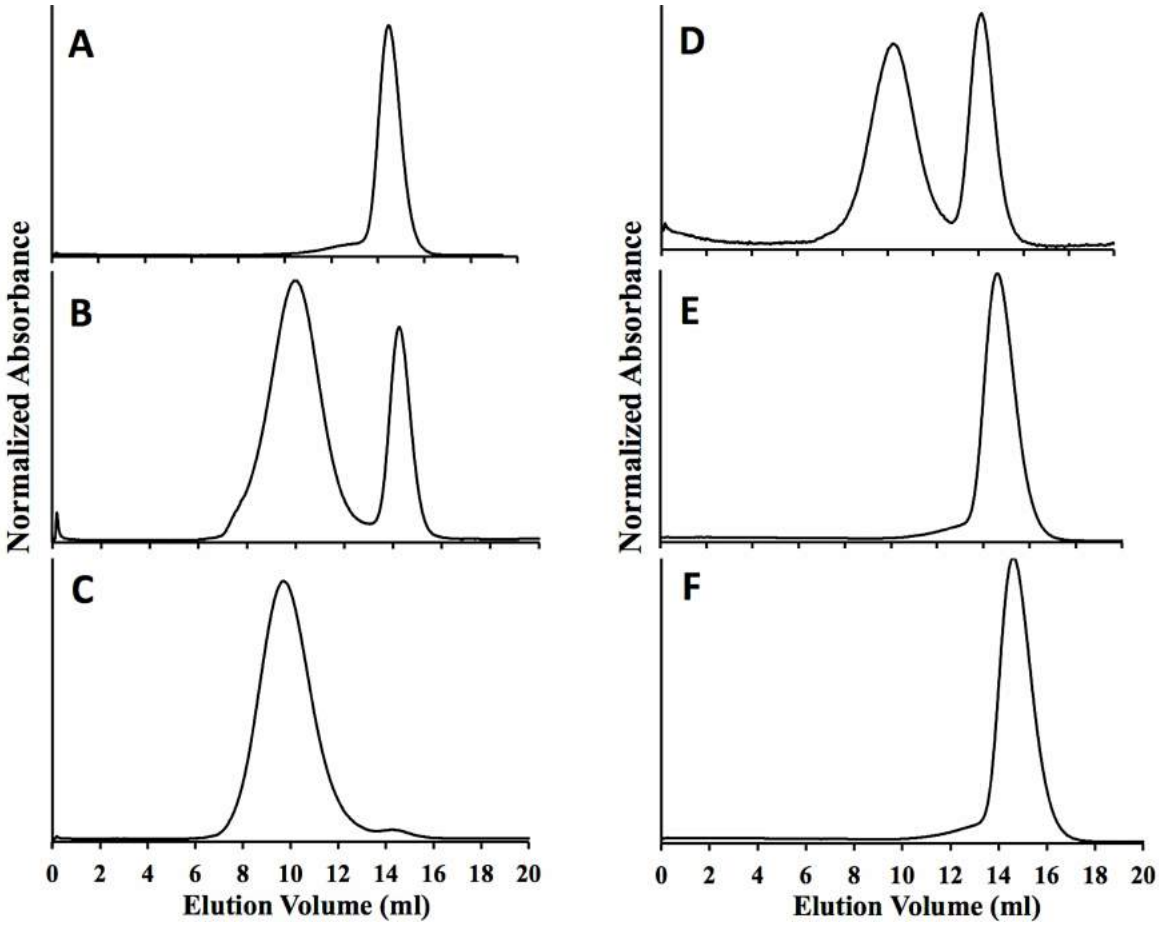

Figure 5. 7 Size Exclusion Chromatography of MBP-T8H4. (A) SEC of MBP-T8H4 purified by MBP-trap column. (B) 20 uM MBP-T8H4 equilibrated with 80 uM Ni(II). (C) Re-chromatography of MBP-T8H4 assemblies in the presence of Ni(II) in the buffer. (D) Re-chromatography of MBP-T8H4 assemblies in the absence of Ni(II) in the buffer. (E) Re-chromatography of MBP-T8H4 assemblies in the presence of a metal chelator (EDTA) in the buffer. (F) Re-chromatography of MBP-T8H4 assemblies at pH 4.5 in TrisHCl buffer containing Ni(II).

5.3.2 Equilibrating MBP-T8H4 with Metal Ions

The *de novo* designed metal-binding coiled coil (IZ-3adH) was reported to have low micromolar affinities for divalent metal ions with relative affinities in the order Ni(II) > Cu(II)> Zn(II) ≈ Co(II).[17] We therefore evaluated the potential of MBP-T8H4 construct to assemble into protein cages in the presence of the above metal

ions. Because Ni(II) was reported to have the highest binding affinity for IZ-3adH (K_d = 5.0 ± 0.3 μM) we first studied the binding of Ni(II) by MBP-T8H4. 200 μL of 20 μM MBP-T8H4 was equilibrated with concentrations of $NiCl_2$ varying from 5 μM to 200 μM $NiCl_2$ at 4 °C with gently agitating o/n. The samples were chromatographed on a superose 6 10/300 column equilibrated with the Tris HCl SEC buffer (at pH 8) containing the same $NiCl_2$ concentration. Two peaks were observed for all the samples, at 10.5 mL and 14.5 mL indicating that some of the protein was assembling into higher molecular weight species in the presence of Ni(II) (Fig. 5.7 B). Maximum percentage of assemblies were observed at Ni(II) concentrations between 20 μM-200 μM (Fig. 5.8). Analysis of SEC peak areas indicated that ~ 65% - 75% of the proteins assembled into higher order species at these Ni(II) concentrations (Fig. 5.8). MBP-T8H4 was studied by equilibrating at higher Ni(II) concentrations (1 mM and 2 mM), however a smaller fraction (~ 30- 40%) of proteins assembled at these concentrations and a 100% assembly could not be achieved at studied Ni(II) concentrations. Equilibrating MBP-T8H4 with Ni(II) at room temperature or for a longer time period (up to 2 weeks) did not make any significant difference on the fraction of assembled vs. unassembled MBP-T8H4.

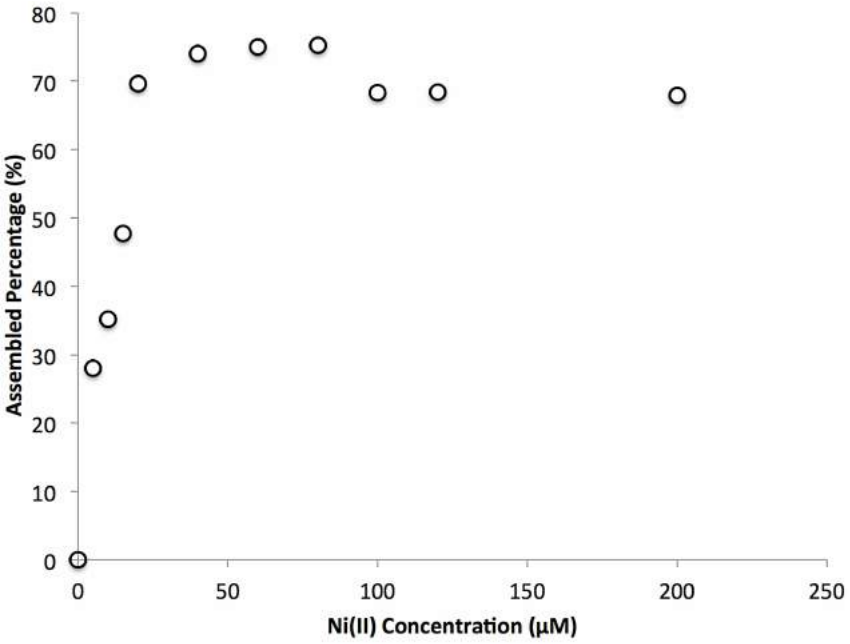

Figure 5. 8 Assembled percentage of MBP-T8H4 in the presence of Ni(II) studied by analyzing area under the peak of SEC elution profiles.

Next, the assembly of MBP-T8H4 in the presence of Co(II), Cu(II) and Zn(II) was evaluated. 20 μM of MBP-T8H4 samples were equilibrated at 20 μM, 80 μM and 200 μM concentrations of each metal ion and evaluated by SEC. At 20 μM, 80 μM and 200 μM Co(II), at 20 μM and 80 μM Cu(II) and at 20 μM Zn(II) concentrations, MBP-T8H4 gave size exclusion profiles similar to the elution profiles in the presence of Ni(II) at those concentrations (Fig. 5.9). However, at 200 μM Cu(II), no peak was observed at 10.5 mL; rather, a small peak was observed at 12 mL, in addition to the trimer peak at 14.5 mL. 4-coordinate complexes are also common for Cu(II), so the peak at elution volume 12 mL might result from MBP-T8H4 forming dimers at higher Cu(II) concentrations. In support of this possibility, Suzuki, et al. observed dimer formation with IZ-3adH peptides at higher Cu(II) concentrations whereas at

low Cu(II) IZ-3adH formed trimers.[17] Moreover, only a single peak at elution volume 14.5 mL was observed at 80 μM and 200 μM Zn(II) concentrations studied. This experiment was repeated with two independent Zn(II) stocks and similar results were observed.

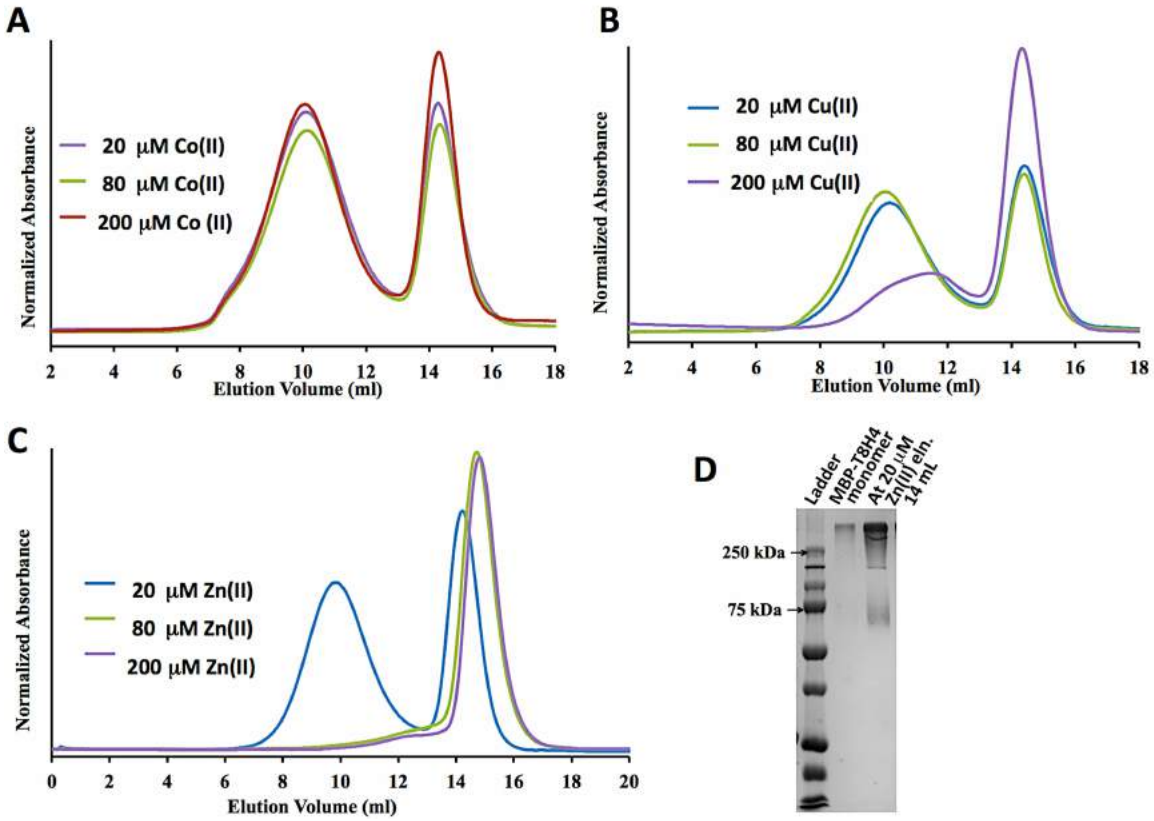

Figure 5. 9 Studying the assembly of MBP-T8H4 in the presence of Co(II),Cu(II) and Zn(II). (A-C) Size exclusion chromatography of MBP-T8H4 at, A) Co(II); B); Cu(II); C) Zn(II). (D) SDS PAGE gel comparing cross-linked protein from elution volume 14 mL of 20 μM Zn(II) chromatograph with cross-linked MBP-T8H4 trimer, the same protein band heights for both samples confirmed that the shift of trimer peak of MBP-T8H4 at 20 μM Zn(II) is not from dimerizing of the protein.

5.3.3 Studying the Controlled Assembly of MBP-T8H4

50-200 μM of MBP-T8H4 were equilibrated with 200 μM $NiCl_2$ and assembled fractions were purified by SEC. When re-chromatographed, the

assembled protein eluted at the same elution volume in a symmetrical peak, which suggested the assembled species was fairly homogeneous (Fig. 5.7 C). However, when the assembled species were re-chromatographed in buffer containing no metal ions, the protein eluted as a mixture of assembled and unassembled forms (Fig. 5.7 D). This is consistent with the loss of Ni(II) ions from the complex during chromatography and suggested that assembly/disassembly may be a dynamic process.

Next, we studied the potential of disassembling the assembled species by changing the buffer conditions. When the MBP-T8H4 assemblies were chromatographed on the superose 6 10/300 column equilibrated with the Tris HCl buffer containing 2 mM EDTA, a single protein peak at elution volume ~ 14.5 mL was observed indicating a complete dissociation of the cages into their trimeric form (Fig. 5.7 E). Additionally, similar results were observed for assembled protein re-chromatographed with Tris HCl buffer adjusted to pH 4.5 (Fig. 5.7 F). A small percentage (~ 30-40%) of the proteins disassembled at pH 5 and pH 5.5. Next, the potential for the re-assembly of the cages was studied by either removing EDTA followed by equilibrating with Ni(II) or re-adjusting the pH of pH 4.5 to pH 8.0. When these samples were re-chromatographed on a superose 6 10/300 size exclusion column, a large fraction of protein eluted at 10.5 mL indicating that the protein had reassembled in the presence of metal ions and under near neutral pH conditions (pH 8). However, the chromatograph also indicated a small fraction of aggregated protein after low pH treatment, which might have resulted from denaturation of some of TriEst at low pH (Fig. 5.10).[18]

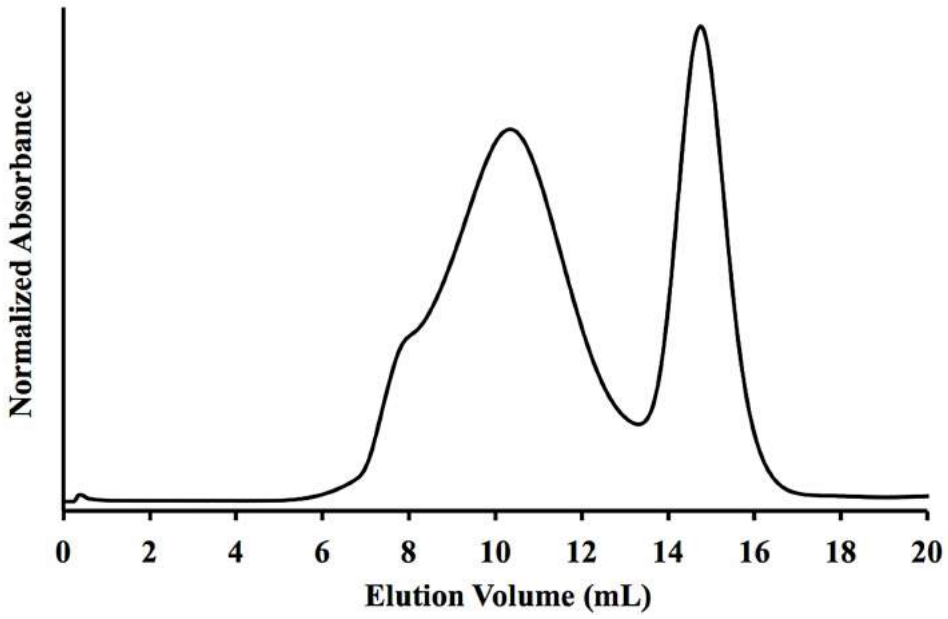

Figure 5. 10 Size exclusion chromatography of MBP-T8H4 after re-adjusting pH from 4.5 to pH 8.0.

5.3.4 Characterization of Metal-dependent Assembly of MBP-T8H4

5.3.4.1 Dynamic Light Scattering (DLS)

DLS is widely used to measure the hydrodynamic radius of particles and to assess the homogeneity of samples. The trimeric and assembled forms of MBP-T8H4 were studied by DLS at 20 μM protein concentrations and, at 100 μM Ni(II) for assembled cages. In the presence of Ni(II), MBP-T8H4 assembled into a homogeneous species with a larger hydrodynamic radius (23.8 ± 0.9 nm). The unassembled form gave a much smaller hydrodynamic radius (12.6 ± 2.0 nm) (Fig. 5.11).

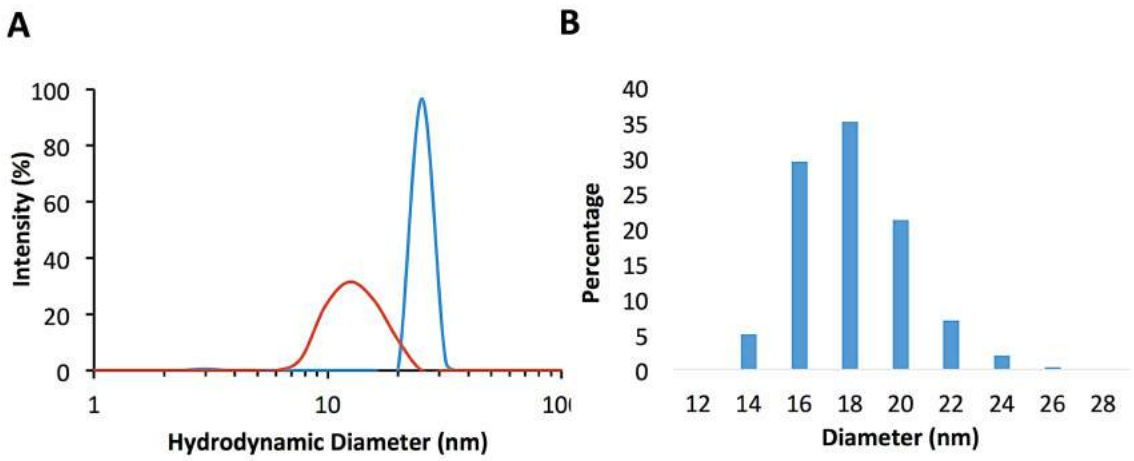

Figure 5. 11 Diameter distribution of MBP-T8H4 (A). Intensity distribution of assembled *(blue)* and unassembled *(Red)* MBP-T8H4 by DLS. (B) Size distribution analysis of assembled MBP-T8H4 obtained from measurement of ~ 300 particles imaged by negative stain TEM using the software package Image J.

5.3.4.2 Native-PAGE

The assemblies were further studied by native-PAGE to assess the homogeneity of the assembled proteins. The assembled protein was cross-linked with bis(sulfosuccinimidyl)suberate (BS3) to prevent dissociation into the trimeric form during electrophoresis because Ni(II) ions would be stripped from the protein during electrophoresis. On a 3-12% native-PAGE gel, cross-linked assemblies migrated as a much larger but single species and the trimeric MBP-T8H4 migrated fast as a small species and with a single gel band (Fig. 5.12 B).

5.3.4.3 Electron Microscopy

The samples were visualized by negative stain TEM. The trimeric form of the protein (fractions from elution volume 14.5 mL) appeared as small, unassembled species by EM (Fig. 5.13). Metal-assembled proteins collapsed on the grids during sample preparation possibly due to interactions between the coiled coils motifs and

Cu on the surface of the TEM grid (which was not seen for our previous cage designs). This was overcome by covalently cross-linking the assembled protein with BS3 reagent (Fig. 5.12). The cross-linked material appeared as discrete particles and showed a homogeneous size distribution (Fig. 5.13). However, neither the symmetry nor the shape of the cages could be discerned in TEM images, possibly because the MBP domain fused to the surface of TriEst masked the core structure. The diameter distribution of TEM images analyzed using image j software gave an average diameter of ~ 18 nm for the particles (Fig. 5.11).

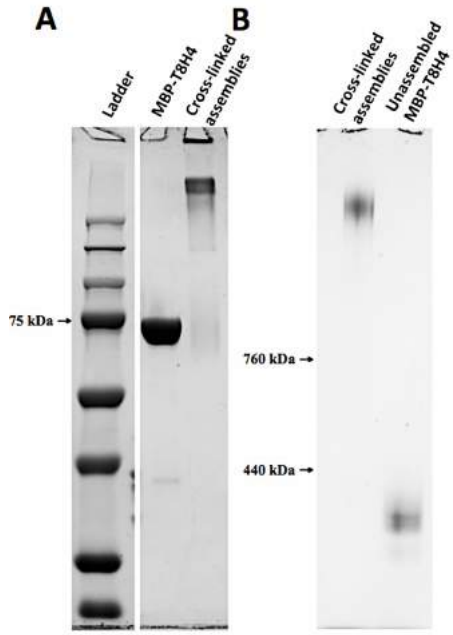

Figure 5. 12 Characterization of MBP-T8H4 by gel analysis. (A) SDS-PAGE analysis of MBP-T8H4, both uncross-linked and cross-linked samples. (B) Native PAGE analysis of trimeric and assembled MBP-T8H4.

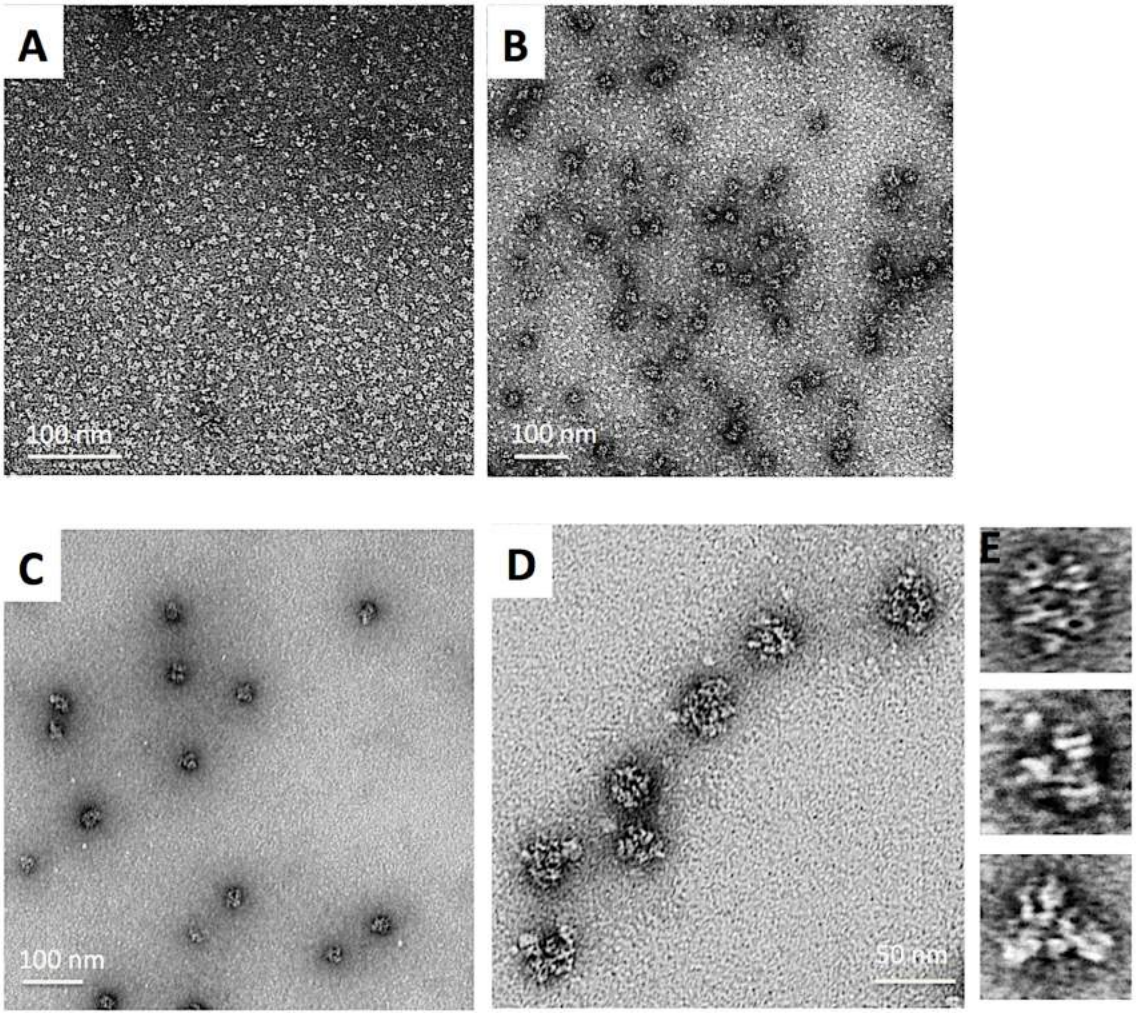

Figure 5. 13 Negative stain TEM images of MBP-T8H4. (A) MBP-T8H4 trimer (unassembled form) purified by SEC followed by MBP trap column. (B) 20 μM MBP-T8H4 equilibrated with 80 μM Ni(II) (C-E) SEC purified and cross-linked MBP-T8H4 assemblies, (C) A wide field view of the assembled cages, (D) A view of cages at higher magnification, (E) Some selected particles showing the three fold symmetry of the cages.

5.3.4.4 Catalytic Activity

The esterase activities of the trimeric MBP-T8H4 construct and the high molecular weight species assembled from MBP-T8H4 in the presence of Ni(II) were compared with the unmodified esterase, TriEst, from which MBP-T8H4 is derived. With *2,4*-di-nitrophenyl acetate (*2,4*-DNPA) as a substrate the turnover number for unmodified TriEst was 14.2 ± 2.3 min^{-1} whereas the turnover number for trimeric MBP-T8H4 was

22.3 ± 1.9 min^{-1} and turnover number for the assembled MBP-T8H4 complex was 15.5 ± 1.7 min^{-1}. These results confirmed that neither the assembly process nor fusion of the coiled coil with MBP domain significantly alters the functional properties of TriEst.

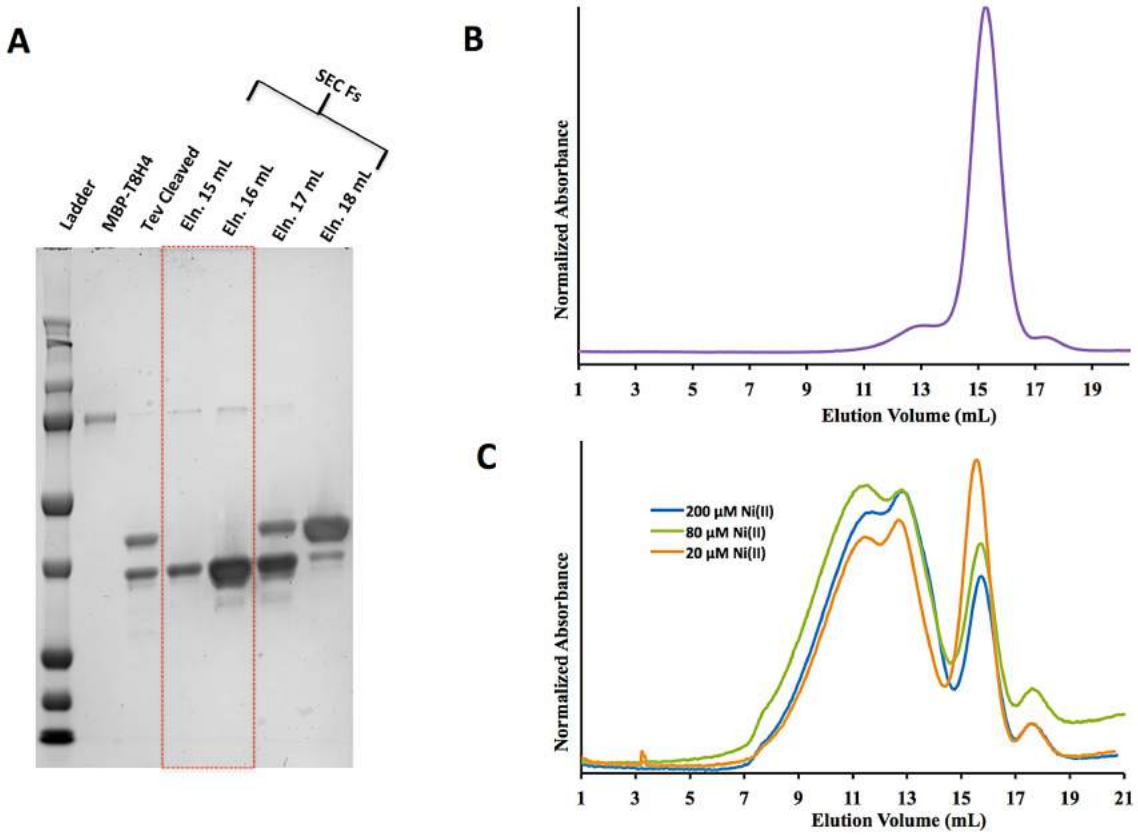

Figure 5. 14 Analysis of MBP cleaved MBP-T8H4. (A) SDS PAGE for MBP cleaved protein samples. (B) Re-chromatography of protein sample from elution volume 16 mL. (C) Size exclusion chromatography of MBP cleaved protein equilibrated at 20 µM, 80 µM and 200µM Ni(II) concentrations.

5.3.5 Removing MBP Tag from MBP-T8H4

The MBP tag was removed by TEV cleavage. MBP cleavage was very efficient for Oct-MBP where o/n incubation of the protein with TEV protease cleaved MBP with ~ 100% efficiency (Chapter 3). However, a longer incubation with TEV

protease (agitating at room temperature for ~ 2 weeks) was needed to cleave MBP tag from MBP-T8H4. As both MBP-T8H4 and TEV cleaved MBP-T8H4 elution profiles overlapped on the superose 6 10/300 column, almost 100% of TEV cleavage efficiency was necessary to purify the MBP cleaved protein to homogeneity. The MBP-cleaved construct was eluted at elution volume of ~ 16 mL and homogeneity of the collected samples were assayed by SDS-PAGE (Fig. 5.14 A,B).

20 μM of the SEC purified protein was equilibrated with 20 μM, 80 μM or 200 μM Ni(II) o/n at 4 ˚C while gently agitating. Three elution profiles, at 11 mL, 13 mL and 16 mL, were observed for all the samples (Fig. 5.14 C). Elution profiles at 11 mL and 13 mL indicated assembled species and the elution profile at 16 mL corresponded to unassembled trimeric form. When re-chromatographed, fractions collected from elution profiles 11 mL and 13 mL gave symmetric elution profiles at elution volumes 11 mL and 13 mL respectively (Fig. 5.15 A). The samples from elution volume 11 mL and 13 mL were cross-linked with BS3 and analyzed by native-PAGE gel electrophoresis and negative stain TEM. The proteins migrated as smeary bands in native-PAGE gel where the species from elution volume 11 mL migrated slowly compared to the species from elution volume 13 mL (Fig. 5.15 B). Negative stain TEM images of the samples collected from both elution volumes indicated that they assembled into highly heterogeneous species (Fig. 5.15 C). Therefore, no further studies were performed for MBP-cleaved protein.

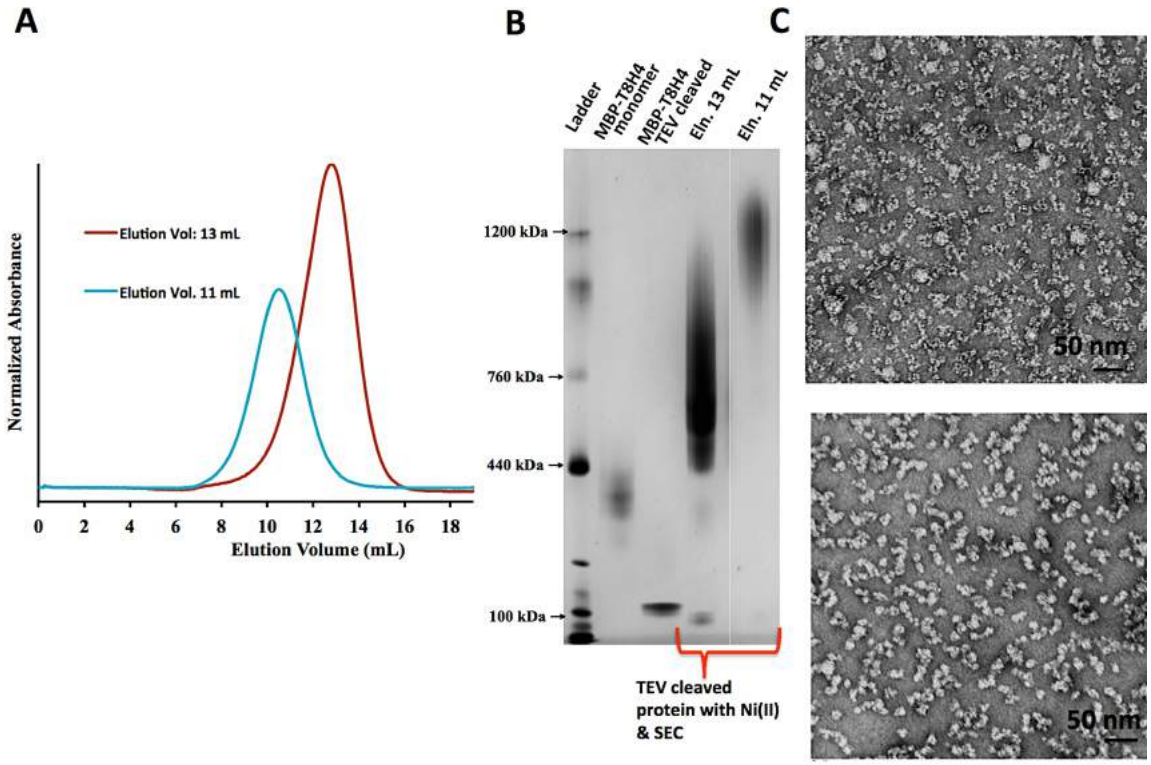

Figure 5. 15 Characterization of assemblies of MBP cleaved protein in the presence of Ni(II). (A) Re-chromatography of protein fractions from elution volume 11 mL (Cyan) and elution volume 13 mL (Red) fractions. (B) Native PAGE analysis of cross-linked samples from elution volume 11 mL and 13 mL. (C) Negative Stain TEM of cross-linked samples from elution volume 11 mL (*Top*) and 13 mL (*Bottom*).

5.4 Discussion

In this work we described designing a metal-dependent protein cage by exploiting the coiled-coil based design strategy developed in our lab. A *de novo* designed metal-binding, parallel, trimeric coiled coil (IZ-3adH) was genetically fused to the C-terminus of TriEst through a oligo-Gly linker: several constructs were designed. T8H4-1 assembled into heterogeneous species in the presence of Ni(II) as evidenced by both SEC and AUC. In TriEst, a His-tag is incorporated at N-terminus for purification purposes, which also interacts with metal ions. In attempts to remove the His-tag (to minimize possible interferences by Ni(II) binding to the His-

tag), two fusion constructs were designed: T8H4-2 and T8H4-3. However, these constructs did not express and the reason is unclear. MBP-T8H4 fusion protein designed by replacing the His-tag with MBP was identified as the most promising construct. This protein could be purified in large scale (~ 400 mg/L culture) by maltose-affinity chromatography.

The assembly of MBP-T8H4 into protein cages in the presence of several divalent transition metal ions was evaluated. The MBP-T8H4 construct could assemble into higher order species in most cases, but despite varying the conditions assembly was never complete. Under the studied metal ion concentrations, the highest percentage observed for the assembled form was 65-75%. Only about 30-40% of the protein assembled at higher Ni(II) concentrations (1 mM and 2 mM) and the protein formed assemblies only at very low concentrations (20 μM) of Zn(II). The reason for these observations is not clear. One possibility is that steric shielding by protein subunits could affect the formation of metal coordination complexes. Furthermore, higher concentration of metal ions could promote the formation of coordination complexes within the same fusion protein; for example with two His-residues in a single helix and its neighboring amine groups, or with the two His-residues in the helix and solvent or buffer molecules (E.g. H_2O, Cl^-, OH^-). However, we do not have clear evidence or experimental data to support these hypotheses.

Nevertheless, when the fractions from the peak with elution volume 10.5 mL were re-chromatographed in the presence of metal ions, they eluted at the same elution volume showing the stability of metal-assembled particles. Removing the metal from the buffer or lowering pH led to disassembly of the cages. Interestingly

the protein could be reassembled by incubating with metal ions or increasing pH. This confirmed that the cage assembly is dynamic and could be controlled by both metal ions and pH.

Overall, in the presence of Ni(II), Co(II), Cu(II) or Zn(II), MBP-T8H4 construct successfully assembled into discrete particles as evident by DLS, TEM and native-PAGE gel analysis . The hydrodynamic radius by DLS and the diameter analysis by negative stain TEM were in a good agreement with the modeled diameter (~ 18-20 nm) of the tetrahedral cage. The symmetry of assembled structures was not apparent in most of the particles imaged by negative stain TEM as MBP domain probably masks the core cage structure. 3-fold symmetry of a tetrahedral cage was apparent in some images (Fig. 5.13 E). Furthermore, MBP-T8H4 needed to be incubated o/n with gentle agitating in order for the protein to assemble. MBP-T8H4 stayed in trimeric form when equilibrated with metal ions for a short time or o/n and therefore the agitating step was necessary for the protein to assemble (Fig 5.16). Agitation step probably position the subunits very close to each other and in an orientation that facilitates the assembly.

Some research groups have designed metal-dependent extended materials (1-, 2- and 3-dimensional) by introducing metal coordinating groups to the exterior surfaces or interfaces of natural proteins.[6,11,19–21] Furthermore, natural protein cages have been engineered to undergo environment-responsive assemblies.[12–14] Additionally, metal-dependent protein motifs (E.g. coiled coils and beta sheets) have been *de novo* designed.[17,22–26] Nevertheless, *de novo* designing protein cages with controlled interfaces in 3-dimensions remains a challenge. Here, we were able to

overcome this limitation by utilizing the inherently modular protein cage design approach, developed in our lab.

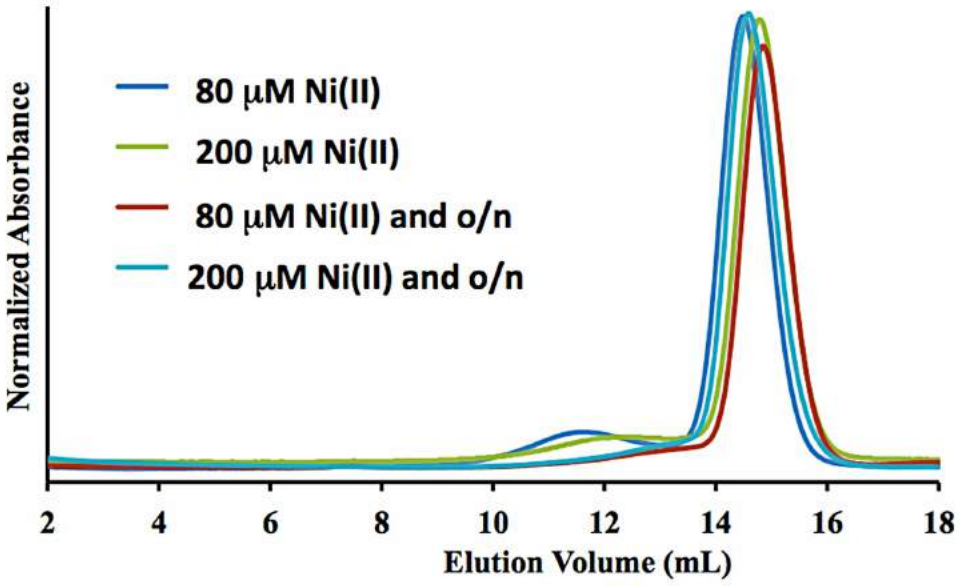

Figure 5. 16 Size exclusion chromatography of MBP-T8H4 equilibrated with Ni(II) for ~ 2 h or o/n without agitating the samples.

Additionally, in this design, we fused MBP (a large monomeric protein) to N-terminus of TriEst located on one of its two triangular surfaces (the surface opposite to its active site) to facilitate purification. In chapter 3, I described fusing a large monomeric protein to the free end of the coiled coil of our octahedral protein cage design. In that case, the design successfully assembled into an octahedral protein cage with M_w ~ 1.8 MDa. This work shows the potential of TriEst BBP to assemble into cages when a large functional group is fused to its N-terminus. As TriEst could be decorated with large groups, proteins in our case, at either of its termini, these protein cages show potential to be developed for various biomedical applications

(E.g. polyvalent antigen display, fusing tissue-specific tags for targeted drug delivery or decorating with nucleic acid strands for gene therapy).

One drawback of our all the previous cage designs was our inability to control their surface orientation. In this design, by fusing a large group onto one surface of TriEst, we were able to force its active site to face inward and there by control the surface orientation of the BBP. Activity assays showed that adding a bulky group (MBP in our case) or assembly process did not alter the enzymatic properties of TriEst.

5.5 Conclusion

De novo designing protein cages with controlled interfaces has been a challenge for other protein cage *de novo* design approaches. Here, we successfully designed a metal-dependent protein cage by utilizing our coiled coil-driven symmetry-based design strategy. Our approach requires no structural modifications to the BBP and therefore preserves its structural and functional properties. Coiled coil motifs govern the assembly of the BBP to cage-like geometries. Utilizing off-the-shelf environmental responsive coiled coils, we extended the study to *de novo* design of protein cages to create metal-dependent assemblies. These cages have potential applications as metal ion sensors, as nanoreactors, and for *in vivo* bio medical applications such as for controlled release of drugs.

5.6 Acknowledgements

I would like to acknowledge Dr. Aaron Sciore for designing and evaluating T8H4-1 fusion construct and Phil Koldewey for performing analytical ultracentrifugation for T8H4-1 construct.

Transmission electron microscopy was performed at the Microscope Imaging Laboratory (MIL) core facility at University of Michigan. Scanning electron microscopy was performed at the Michigan Center for Materials Characterization at University of Michigan. Dynamic light scattering was performed at Biophysics Research Facility at University of Michigan.

This work was supported in part by grants from the National Institutes of Health (GM 093088) and the Army Research Office (W911NF-11-1-0251).

5.7 References

1. Badieyan S, Sciore A, Eschweiler JD, Koldewey P, Cristie-David AS, Ruotolo BT, Bardwell JCA, Su M, Marsh ENG (2017) Symmetry-Directed Self-Assembly of a Tetrahedral Protein Cage Mediated by de Novo-Designed Coiled Coils. ChemBioChem 18:1888-1892.

2. Sciore A, Su M, Koldewey P, Eschweiler JD, Diffley KA, Linhares BM, Ruotolo BT, Bardwell JCA, Skiniotis G, Marsh ENG (2016) Flexible, symmetry-directed approach to assembling protein cages. Proc Natl Acad Sci USA 113:8681–8686.

3. Dominguez R, Holmes KC (2011) Actin Structure and Function. Annu Rev Biophys 40:169–186.

4. Rabut G, Doye V, Ellenberg J (2004) Mapping the dynamic organization of the nuclear pore complex inside single living cells. Nat Cell Biol 6:1114–1121.

5. Shoemaker GK, van Duijn E, Crawford SE, Uetrecht C, Baclayon M, Roos WH, Wuite GJL, Estes MK, Prasad BVV, Heck AJR (2010) Norwalk Virus Assembly and Stability Monitored by Mass Spectrometry. Mol Cell Proteomics 9:1742–1751.

6. Liljeström V, Mikkilä J, Kostiainen MA (2014) Self-assembly and modular functionalization of three-dimensional crystals from oppositely charged proteins. Nat Commun 5:4445.

7. Ringler P (2003) Self-Assembly of Proteins into Designed Networks. Science 302:106–109.

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

9. Zhang W, Luo Q, Miao L, Hou C, Bai Y, Dong Z, Xu J, Liu J (2012) Self-assembly of glutathione S-transferase into nanowires. Nanoscale 4:5847.

10. Ballister ER, Lai AH, Zuckermann RN, Cheng Y, Mougous JD (2008) In vitro self-assembly of tailorable nanotubes from a simple protein building block. Proc Natl Acad Sci USA 105:3733–3738.

11. Bailey JB, Zhang L, Chiong JA, Ahn S, Tezcan FA (2017) Synthetic Modularity of Protein–Metal–Organic Frameworks. J Am Chem Soc 139:8160–8166.

12. Peng T, Lim S (2011) Trimer-Based Design of pH-Responsive Protein Cage Results in Soluble Disassembled Structures. Biomacromolecules 12:3131–3138.

13. Huard DJE, Kane KM, Tezcan FA (2013) Re-engineering protein interfaces yields copper-inducible ferritin cage assembly. Nat Chem Biol 9:169–176.

14. Choi S-H, Choi K, Chan Kwon I, Ahn HJ (2010) The incorporation of GALA peptide into a protein cage for an acid-inducible molecular switch. Biomaterials 31:5191–5198.

15. Salom D, Hill BR, Lear JD, DeGrado WF (2000) pH-dependent tetramerization and amantadine binding of the transmembrane helix of M2 from the influenza A virus. Biochemistry 39:14160–14170.

16. Zimenkov Y, Dublin SN, Ni R, Tu RS, Breedveld V, Apkarian RP, Conticello VP (2006) Rational Design of a Reversible pH-Responsive Switch for Peptide Self-Assembly. J Am Chem Soc 128:6770–6771.

17. Suzuki K, Hiroaki H, Kohda D, Nakamura H, Tanaka T (1998) Metal Ion Induced Self-Assembly of a Designed Peptide into a Triple-Stranded α-Helical Bundle: A Novel Metal Binding Site in the Hydrophobic Core. J Am Chem Soc 120:13008–13015.

18. Shaw S-Y, Chen Y-J, Ou J-J, Ho L (2006) Enzymatic resolution of methyl dl-β-acetylthioisobutyrate and dl-β-acetylthioisobutyramide using a stereoselective esterase from Pseudomonas putida IFO12996. J Mol Catal B-Enzym 38:163–170.

19. Sontz PA, Bailey JB, Ahn S, Tezcan FA (2015) A Metal Organic Framework with Spherical Protein Nodes: Rational Chemical Design of 3D Protein Crystals. J Am Chem Soc 137:11598–11601.

20. Schurke P, Freeman JC, Dabrowski MJ, Atkins WM (1999) Metal-dependent Self-assembly of Protein Tubes from *Escherichia coli* Glutamine Synthetase: Cu^{2+} EPR Studies of the Ligation and Stoichiometry of Intermolecular Metal-binding Sites. J Biol Chem 274:27963–27968.

21. Bailey JB, Subramanian RH, Churchfield LA, Tezcan FA (2016) Metal-Directed Design of Supramolecular Protein Assemblies. Methods Enzymol 580: 223–250.

22. Tanaka T, Mizuno T, Fukui S, Hiroaki H, Oku J, Kanaori K, Tajima K, Shirakawa M (2004) Two-Metal Ion, Ni(II) and Cu(II), Binding α-Helical Coiled Coil Peptide. J Am Chem Soc 126:14023–14028.

23. Tebo AG, Hemmingsen L, Pecoraro VL (2015) Variable primary coordination environments of Cd(II) binding to three helix bundles provide a pathway for rapid metal exchange. Metallomics 7:1555–1561.

24. Plegaria JS, Pecoraro VL (2016) De Novo Design of Metalloproteins and Metalloenzymes in a Three-Helix Bundle. Methods Mol Biol 1414:187–196.

25. Schneider JP, Kelly JW (1995) Synthesis and Efficacy of Square Planar Copper Complexes Designed to Nucleate .beta.-Sheet Structure. J Am Chem Soc 117:2533–2546.

26. Scotter AJ, Guo M, Tomczak MM, Daley ME, Campbell RL, Oko RJ, Bateman DA, Chakrabartty A, Sykes BD, Davies PL (2007) Metal ion-dependent, reversible, protein filament formation by a designed beta-roll polypeptide. BMC Struct Biol 7:63.

Chapter 6: Conclusion

Our lab has recently developed a fundamental design strategy for assembling protein cages based on small coiled coil sequences. In this approach, a natural oligomeric protein possessing C_x rotational symmetry is genetically fused to a homo-oligomeric parallel coiled coil domain with C_y symmetry and the geometry of the cage is specified by the combination of those two symmetry elements. The natural protein constitutes the primary building block (BBP) of the protein cage while the coiled coil drives the assembly of these BBP units into cage-like geometries. A flexible oligo-Gly linker is incorporated between the BBP and the coiled coil to provide the necessary degree of freedom and flexibility between these two symmetry elements for assembly to occur. The fusion construct is optimized by varying the strength of the coiled coil and the length of the oligo-Gly linker to obtain the intended geometry. The strength of the coiled coil can be controlled by varying the number of heptad repeats.

This approach provides a simple route for designing protein cages *de novo* which does not depend explicitly on structural details of the BBP. Therefore, our design strategy avoids the need for intensive computational modeling and protein engineering that other protein cage design approaches require. Our group successfully designed a tetrahedral protein cage and an octahehedral protein cage by utilizing this approach and fusing a trimeric BBP to a trimeric coiled coil and a

tetrameric coiled coil respectively, which were discussed in chapter 1.[1,2] My thesis is a continuation of this work, which included evaluating and optimizing coiled coil motifs to mediate protein assemblies and designing protein cages with new geometries and new properties.

6.1 Evaluating *De novo* Designed Coiled Coils

The first focus of my thesis was evaluating *de novo* designed coiled coils as off-the-shelf components to mediate these protein cage assemblies. These *de novo* coiled coils were originally chemically synthesized and evaluated crystallographically and their behavior when fused to other proteins had not been investigated.[3,4] Therefore, we examined how reliable these coiled coils are as off-the-shelf component for *de novo* designing complex protein structures. We decided to evaluate the utility of these coiled coils by fusing to a monomeric protein that does not interfere with the oligomeric state of the coiled coils. Green fluorescent protein (GFP) was chosen as it is a widely used and well behaved monomeric protein whose fluorescence provides an easy read out of the protein's expression and folding.

We evaluated several *de novo* designed coiled coils of potential utility for protein assembly.[5] Interestingly, these results clearly showed that the oligomeric state of these coiled coils could be influenced by the protein fused to it, even in the presence of a fairly long linker between the monomer and GFP (9 residues in this case, spanning approximately ~ 30 Å). This work was the first study to report the context-dependent behavior of coiled coils. Overall, this study suggested the importance of evaluating the *de novo* designed coiled coils by fusing them to a

monomeric protein before utilizing them to assemble oligomeric proteins into more complex architectures.

6.2 Elaborating an Octahedral Cage with a Monomeric Protein

In this work, I explored the potential of elaborating protein cages with additional protein domains by fusing maltose binding protein (MBP) to the free end of the coiled coil in the octahedral protein cage (Oct-4). With the addition of MBP, a protein domain of ~ 40 kDa, the construct successfully assembled into an octahedral protein cage with M_w of ~ 1.8 MDa.[6] Importantly, fusing a larger protein to the coiled coil domain did not perturb the geometry of the cage. The ability to elaborate the cages by appending additional domains to the to free end of the coiled coil has potential uses for polyvalent antigen display and designing artificial multi-enzyme complexes.

Additionally, fusing MBP to the Oct-4 cage allowed the cage to be purified efficiently using maltose-affinity chromatography. Initially Oct-4 octahedral cages were purified using Ni-affinity chromatography which gave a very low yield of protein (~ 1 mg/L culture). The inability to purify the protein assemblies on large scale is a drawback for using these cages for applications which need large quantities of proteins (E.g. vaccine development, drug delivery, nanoreactors for synthesizing inorganic nanoparticles). Fusing MBP to the octahedral cage increased expression level significantly and purification by MBP-trap column further increased the yield of protein by 60-fold. In our design strategy, coiled coils can face inward or outward from the cage wall. However, fusing large group to the free end of the coiled coil prevents them from facing inward because of steric clashes.

Therefore, fusing MBP provided a way to control the orientation of coiled coils by forcing them to point outward.

6.3 Designing a hyperstable, 60-subunit Icosahededral Protein Cage

The third aim of my thesis was exploring the potential of designing an icosahedral protein cage by fusing a pentameric coiled coil (PDB ID: 4PN8) to the same TriEst BBP (Trimeric Esterase from *Pseudomonos putida*; PDB ID: 1ZOI) that our group employed for designing tetrahedral and octahedral protein cages. Although the strategy is simple, the fact that 20 copies of the BBP need to be brought together by a small, ~ 3 kDa coiled coil domain in order to assemble an icosahedral protein cage represents a significant challenge. Several constructs were screened that contained increasing oligo-Gly linker lengths from 2 residues to 12 residues. The construct with the 8-residue linker (Ico8) assembled into discrete-cages, however the assemblies were highly heterogeneous and the assembly took about 2-3 months to occur.

Next I studied a better purification protocol and a method to increase the homogeneity of the proteins and to accelerate the time required for assembly. Interestingly Ico8 cages showed a high thermostability and this property of the cages was used to purify Ico8. Heating the cell lysate at 70 °C for 1.15 h allowed precipitation of most of non-specific proteins. Dialyzing the supernatant followed by size exclusion chromatography gave pure protein. This purification procedure was far more effective in purifying Ico8. Additionally, purifying by heating increased the homogeneity of the assemblies and accelerated the assembly process to less than two weeks.

Interestingly Ico8 cages showed thermal hyperstability which was not seen in either octahedral or tetrahedral designs. Ico8 stayed folded at temperatures up to 120 °C, at extremes of pHs between pH 2 and 12 and in high concentrations of denaturants (8 M urea or 8 M guanidinium chloride). This stability likely results from the higher order assembly and is often evident in viral capsids where a small surface area is exposed to the external environment as a large number of protein subunits assemble into pseudo-spherical geometries with a small surface-to-volume ratio, e.g. icosahedral.

Furthermore, Ico8 captured short DNA strands that were adventitiously formed by shearing of DNA during the sonication of cells. Treating these cages with benzonase led to disassembly of the cages into the trimeric protein building blocks. Interestingly, they reassembled into heterogeneous cages after storage at room temperature for about two weeks. This indicated that these short DNA strands play a role in assembling Ico8 into homogeneous cages. Ico8 preparations purified using a microfludizer to break the cells, where little shearing of DNA occurs, also had significant size variations. However, the cages assembled in the absence of DNA were still very stable, similarly to cages assembled in the presence of DNA. The morphological changes of these cages in the presence or absence of DNA is an emergent property of Ico8 and further studies are needed to understand the underlying mechanism of these changes.

Cryo-EM 3D reconstruction of Ico8 containing DNA showed that it has the C_3+C_5 symmetry of a 60-subunit (T1) icosahedral cage. These cages still varied in diameter between ~ 20-30 nm. This variation may result from the flexibility of the

8-residue Gly linker. Furthermore, a Cryo-EM 3D reconstruction map could not be generated for Ico8 cages assembled in the absence of DNA because of their significant size variations (~ 20-50 nm). Interestingly 2D class averages of both Ico8 either containing DNA or not containing DNA showed an inner wall. SDS-PAGE analysis of SEC purified Ico8 cages showed a small amount of non-specific proteins co-purified with Io8 but these were not evident in the Native-PAGE gel analysis of Ico8. This indicated that the inner layer might be formed by these non-specific proteins and possibly by other unidentified bio-polymers. Some viral capsids (E.g. P22 viral capsid) assemble around scaffold proteins and Ico8 may follow a similar route for assembly. This is another emergent property of Ico8 resulting from higher order structure. Understanding the mechanism underlying the assembly process of Ico8 may inform the evolution of natural proteins into complex nano-machines with diverse structural and functional properties (E.g. viral capsids, PDH, ferritin, GroEL). Overall, the successful design of an icosahedral cage clearly demonstrates the generalizability of coiled-coil driven protein cage design approach to design complex protein architectures.

6.4 Design of a Metal-dependent Protein Cage

Lastly, I explored the potential of utilizing this approach to design protein cages with controlled assembly. As there are environment-responsive coiled coils available, this design approach would provide a simple approach to *de novo* design cages whose assembly could be controlled. This was studied by fusing a *de novo* designed metal dependent coiled coil, containing six-His ligands, to the TriEst BBP. Additionally the His-Tag of TriEst was replaced by MBP to avoid interference that

can result from interaction of the His-Tag with metal ions. The construct was named MBP-T8H4. The construct successfully assembled into discrete particles, presumably of a tetrahedral geometry (although this remains to be confirmed) with M_w of ~ 900 kDa in the presence of Ni(II), Co(II), Cu(II) and Zn(II). The cages could be completely disassembled by adding chelators or by lowering the pH to pH 4.5. Removing EDTA followed by incubating with metal ions or increasing the pH to near neutral led MBP-T8H4 to reassemble into cages. This study provided a proof-of-concept for utilizing the coiled coil mediated protein cage design approach to design protein cages with controlled interfaces.

6.5 Future Directions

Both natural and *de novo*-designed protein cages are becoming very attractive nano-particles for various applications in medicine and materials science. The shape, size, geometry and the internal volume are some key considerations in choosing protein cages for applications in these fields. Our approach provides a systematic method to design cages that match these particular needs. So far, we have designed protein cages only by utilizing one natural BBP. As the demand for protein cages is rapidly growing, one very interesting future direction of the project would be to extend this work to design various protein cage architectures by pairing other natural proteins with different rotational symmetries with the appropriate coiled coils.

Attractive 3D structures, which we have yet to explore, are cubic and dodecahedral protein cage geometries. By pairing a tetrameric BBP with a trimeric coiled coil, a cubic protein cage with C_4+C_3 symmetry could be designed. Although

there are many tetrameric proteins, most of them exist in pseudo-D_2 symmetries and are thus not well suited for our design approach. Nevertheless, Ruv-A (PDB ID: 1CUK) is a well-characterized DNA-binding protein with C_4 symmetry and which could be utilized to design a protein cage with cubic geometry.[7] As Ruv-A binds to DNA, the structural integrity of assembled protein could be easily verified by a DNA gel shift assay. Dodecahedral protein cages with C_5+C_3 symmetry could be designed by fusing a BBP with C_5 symmetry to a trimeric coiled coil. Here, the well-characterized lumazine synthase (PDB ID: 2I0F) from *Saccharomyces cerevisiae* could be utilized as the BBP.[8] Lumazine synthase is an enzyme in the riboflavin biosynthesis pathway and therefore, structural properties of the BBP of assembled dodecahedral protein cage can be easily assayed by an existing chromogenic assay.[9]

Additionally, protein cages that assemble under controlled conditions are very useful systems for applications that require selective encapsulation and release of materials: for example, targeted drug delivery and designing reusable nanoreactors. However, *de novo* designing protein cages that assemble in a controlled manner is a challenging problem because of our limited knowledge in protein-protein interactions. The ability to use small coiled coils as off-the-shelf connectors to mediate protein assembly and the availability of environmentally responsive coiled coils make our design approach promising for designing protein cages whose assembly can be controlled. As a proof-of-concept study, I successfully designed a metal-dependent protein cage.

pH-responsive protein cages could be very useful for lysosome-targeted drug-delivery. Here, one could explore the potential of designing a pH-responsive

tetrahedral or octahedral protein cage. The required trimeric and tetrameric pH-responsive homo-oligomeric parallel coiled coils, could be designed by incorporating His residues at either “a” or “d” positions of trimeric and tetrameric coiled coils, which provide a simple route to pH-responsive coiled coils. Once the coiled coils are optimized, one could fuse them to TriEst through the appropriate linkers to design a pH-responsive tetrahedral and a pH-responsive octahedral protein cages respectively.

A further extension of our approach would be to design two-component protein cages, which assemble only upon mixing the two different fusion constructs. Such cage designs would be very useful for applications that require controlled encapsulation of materials. To test this, one could genetically fuse TriEst to a hetero-tetrameric, parallel coiled coil with complementary coiled coil domains and design two fusion constructs with complementary coiled coils. The octahedral protein cage should form only when the two fusion constructs are mixed. The hetero-tetrameric parallel coiled coil designed by Fairman and coworkers by modifying the GCN4-pLI parallel coiled-coil tetramer would be well suited for such a study.[10] Our Oct-4 octahedral protein cage required a 4-residue Gly-linker and therefore the same linker could be utilized for this cage design as well.

Functionalizing protein cages would be another very interesting area to explore as there are many potential applications for such biomaterials. Chapter 1 discussed efforts to functionalize the exterior, interior and interfaces of protein cages in order to utilize them for medical as well as industrial applications. Additionally, chapter 1 also discussed in depth about genetic and chemical

approaches to functionalizing protein cages including methods for attaching functional groups both covalently and transiently for various applications. Our cage designs would also appear suitable for all of these various functionalization strategies.

In chapter 3, I described the potential of elaborating our protein cages by fusing large functional groups to free end of the coiled coil, demonstrating it was possible to fuse large protein tags to an octahedral protein cage. Furthermore, in chapter 5, MBP was fused to the N-terminus of TriEst for purification purposes without interfering with assembly of the construct into discrete particles. Both studies demonstrated the potential of the BBP protein to make homogeneous, discrete assemblies in the presence of large groups at either its termini. These properties could make them useful for vaccine design by attaching antigens or for targeted drug delivery by attaching tissue specific ligands to their exteriors. As discussed in chapter 4, TMA-DPH, a fluorogenic molecule, fluoresced upon binding to the pentameric coiled coil of Ico8 cages. The Woolfson group demonstrated the potential of using this interaction to develop a biosensor to detect certain lipid molecules.[11] These are some examples that show the potential of utilizing these cages for actual applications.

In future studies, the hollow interior of these cages could be utilized to encapsulate cargos, synthesize inorganic nanoparticles and as bioreactors. For *in vivo* biomedical applications, drugs and imaging agents could be covalently attached to protein cage interior. Aberrant vascular architecture in tumor tissues causes accumulation of nanoparticles in these areas and this enhanced permeability and

retention (EPR) effect is utilized to treat cancer by nano-medicines.[12] Our protein cages could in principle be loaded with anti-cancer drugs to develop effective cancer therapeutics. For targeted delivery, the protein cage exterior could be decorated with tissue specific ligands; for example, epitope tags and aptamers. PEGylation or decorating our protein cages with humanized proteins might be necessary to make them less immunogenic nano-carriers. A pH-responsive cage design could be developed for lysosome-targeted delivery where the low pH inside the lysosome causes disassembly of the protein cage and release of drugs.

The functional properties of inorganic nanoparticles depend on their size, shape and density, but controlling these properties and synthesizing homogeneous nanoparticles are challenging problems. Another future direction would be repurposing our protein cages for synthesizing inorganic nanoparticles. The well-defined interior of a protein cage provides a size-constrained environment to synthesis nanoparticles with well-defined geometries, volumes and densities. Examples of nanoparticles synthesized inside protein cages and their applications were discussed in chapter 1 (see section 1.3.1). Introducing negatively charged amino acid residues and decorating the cage interior with metal chelating peptides are some popular approaches utilized to mineralize inorganic nanoparticles within the protein cage interior. Our metal dependent protein cage design could be utilized to develop a re-usable nanoreactor where mineralized products can be released by disassembling the protein cages and reassembled cages can be re-used.

The hyper-stability and the large internal volume of the Ico8 protein cage makes it a promising system for designing enzyme-nanoreactors. The extremely

stable nature of this protein cage should provide a well-protected environment for the enzymes. In chapter one we discussed how a significant increase in the catalytic activity of an enzyme could be obtained when it was encapsulated inside a protein cage (see section 1.3.1). The large interior of Ico8 could be utilized for similar applications and to develop highly catalytically active enzyme nanoreactors. Furthermore, compartmentalization of enzymes in cascade reactions, which involve unstable or toxic intermediates, provides an efficient and a safer platform for conducting sequential chemical reactions.

The discussion above outlines a few potential ways to functionalize the protein cages synthesized as part of this thesis and develop them for real world applications. Additionally, various examples of functionalized protein cages were discussed in chapter 1 (see section 1.3) and most of these strategies could be extended to our protein cage designs. With the growing interest in protein-based nanoparticles for both medical and industrial applications, this is a very exciting field of science that our research group would like to explore further in the future.

6.6 Conclusion

In conclusion, in our first attempts to design protein cages, we evaluated a strategy by fusing a trimeric building block protein (Adolase) to a heterodimeric anti-parallel coiled coil through a flexible oligo-Gly linker, however this construct assembled into multiple species.[13] As discussed in chapter 1, pairwise combination of C_3+C_2 symmetry elements can lead to a mixture of geometries unless the dihedral angle is controlled precisely. Next, we explored an approach, which is still modular but far more promising in designing protein cages to homogeneity. As there are

homo-oligomeric parallel coiled coils with higher oligomeric states (Trimer, Tetramer and Pentamer), we hypothesized that utilizing these coiled coils will provide a better route for *de novo* design of coiled coil driven protein cages while mitigating the potential for multiple geometries. Utilizing this approach our group successfully designed a tetrahedral, octahedral and an icosahedral protein cage. Furthermore, we elaborated the octahedral cage with a protein domain at its free coiled coil end. Moreover, using this modular approach, we were able to successfully design a metal responsive protein cage. Overall, all these examples shows the generalizability of the coiled coil based design strategy introduced by the Marsh group.

6.7 References

1. Badieyan S, Sciore A, Eschweiler JD, Koldewey P, Cristie-David AS, Ruotolo BT, Bardwell JCA, Su M, Marsh ENG (2017) Symmetry-Directed Self-Assembly of a Tetrahedral Protein Cage Mediated by de Novo-Designed Coiled Coils. ChemBioChem 18:1888-1892.

2. Sciore A, Su M, Koldewey P, Eschweiler JD, Diffley KA, Linhares BM, Ruotolo BT, Bardwell JCA, Skiniotis G, Marsh ENG (2016) Flexible, symmetry-directed approach to assembling protein cages. Proc Natl Acad Sci USA 113:8681–8686.

3. Fletcher JM, Boyle AL, Bruning M, Bartlett GJ, Vincent TL, Zaccai NR, Armstrong CT, Bromley EHC, Booth PJ, Brady RL, et al. (2012) A Basis Set of *de Novo* Coiled-Coil Peptide Oligomers for Rational Protein Design and Synthetic Biology. ACS Synthetic Biology 1:240–250.

4. Thomson AR, Wood CW, Burton AJ, Bartlett GJ, Sessions RB, Brady RL, Woolfson DN (2014) Computational design of water-soluble α-helical barrels. Science 346:485–488.

5. Cristie-David AS, Sciore A, Badieyan S, Escheweiler JD, Koldewey P, Bardwell JCA, Ruotolo BT, Marsh ENG (2017) Evaluation of de novo-designed coiled coils as off-the-shelf components for protein assembly. Mol Syst Des Eng 2:140–148.

6. Cristie-David AS, Koldewey P, Meinen BA, Bardwell JCA, Marsh ENG (2018) Elaborating a coiled coil-assembled octahedral protein cage with additional protein domains: A multi-domain coiled coil-assembled protein cage. Protein Sci 27:1883-1890.

7. Rafferty JB, Sedelnikova SE, Hargreaves D, Artymiuk PJ, Baker PJ, Sharples GJ, Mahdi AA, Lloyd RG, Rice DW (1996) Crystal structure of DNA recombination protein RuvA and a model for its binding to the Holliday junction. Science 274:415–421.

8. Klinke S, Zylberman V, Bonomi HR, Haase I, Guimarães BG, Braden BC, Bacher A, Fischer M, Goldbaum FA (2007) Structural and Kinetic Properties of Lumazine Synthase Isoenzymes in the Order Rhizobiales. J Mol Biol 373:664–680.

9. Kaiser J, Illarionov B, Rohdich F, Eisenreich W, Saller S, den Brulle JV, Cushman M, Bacher A, Fischer M (2007) A high-throughput screening platform for inhibitors of the riboflavin biosynthesis pathway. Anal Biochem 365:52–61.

10. Root BC, Pellegrino LD, Crawford ED, Kokona B, Fairman R (2009) Design of a heterotetrameric coiled coil. Protein Sci 18:329–336.

11. Thomas F, Dawson WM, Lang EJM, Burton AJ, Bartlett GJ, Rhys GG, Mulholland AJ, Woolfson DN (2018) *De Novo* -Designed α-Helical Barrels as Receptors for Small Molecules. ACS Synth Biol 7:1808–1816.

12. Greish K (2010) Enhanced Permeability and Retention (EPR) Effect for Anticancer Nanomedicine Drug Targeting. Methods Mol Biol 624:25-37.

13. Patterson DP, Su M, Franzmann TM, Sciore A, Skiniotis G, Marsh ENG (2014) Characterization of a highly flexible self-assembling protein system designed to form nanocages: Highly Flexible Self-Assembling Protein System. Protein Science 23:190–199.

Appendix A: Protein Sequences

A.1 Chapter 2 constructs

Vector plasmid: pMCSG18

Expression cell line: E. Coli BL21 DE3

Wild Type GFP

M_w: 30.02 kDa

```
         10          20          30          40          50          60
MHHHHHHSSG VDLGTENLYF QSNIGSGLLA SKGEELFTGV VPILVELDGD VNGHKFSVSG

         70          80          90         100         110         120
EGEGDATYGK LTLKFICTTG KLPVPWPTLV TTLCYGVQCF SRYPDHMKRH DFFKSAMPEG

        130         140         150         160         170         180
YVQERTIFFK DDGNYKTRAE VKFEGDTLVN RIELKGIDFK EDGNILGHKL EYNYNSHNVY

        190         200         210         220         230         240
IMADKQKNGI KVNFKTRHNI EDGSVQLADH YQQNTPIGDG PVLLPDNHYL STQSALSKDP

        250         260
NEKRDHMVLL EFVTAAGITH GMDELYN
```

GFP Construct 1 (GFP-IL17N)

M_w: 32.42 kDa

```
         10          20          30          40          50          60
MHHHHHHSSG VDLGTENLYF QSNIGSGLLA SKGEELFTGV VPILVELDGD VNGHKFSVSG

         70          80          90         100         110         120
EGEGDATYGK LTLKFICTTG KLPVPWPTLV TTLCYGVQCF SRYPDHMKRH DFFKSAMPEG

        130         140         150         160         170         180
YVQERTIFFK DDGNYKTRAE VKFEGDTLVN RIELKGIDFK EDGNILGHKL EYNYNSHNVY

        190         200         210         220         230         240
IMADKQKNGI KVNFKTRHNI EDGSVQLADH YQQNTPIGDG PVLLPDNHYL STQSALSKDP
```

```
       250         260         270         280         290
NEKRDHMVLL EFVTAAGITG GGGGGEIAAL KQEIAALKQE IAANKQEIAA LKQG
```

GFP Construct 2 (GFP-IL)

M_w: 32.42 kDa

```
        10          20          30          40          50          60
MHHHHHHSSG VDLGTENLYF QSNIGSGLLA SKGEELFTGV VPILVELDGD VNGHKFSVSG

        70          80          90         100         110         120
EGEGDATYGK LTLKFICTTG KLPVPWPTLV TTLCYGVQCF SRYPDHMKRH DFFKSAMPEG

       130         140         150         160         170         180
YVQERTIFFK DDGNYKTRAE VKFEGDTLVN RIELKGIDFK EDGNILGHKL EYNYNSHNVY

       190         200         210         220         230         240
IMADKQKNGI KVNFKTRHNI EDGSVQLADH YQQNTPIGDG PVLLPDNHYL STQSALSKDP

       250         260         270         280         290
NEKRDHMVLL EFVTAAGITG GGGGGEIAAL KQEIAALKQE IAALKQEIAA LKQG
```

GFP construct 3 (GFP-II)

M_w:32.42 kDa

```
        10          20          30          40          50          60
MHHHHHHSSG VDLGTENLYF QSNIGSGLLA SKGEELFTGV VPILVELDGD VNGHKFSVSG

        70          80          90         100         110         120
EGEGDATYGK LTLKFICTTG KLPVPWPTLV TTLCYGVQCF SRYPDHMKRH DFFKSAMPEG

       130         140         150         160         170         180
YVQERTIFFK DDGNYKTRAE VKFEGDTLVN RIELKGIDFK EDGNILGHKL EYNYNSHNVY

       190         200         210         220         230         240
IMADKQKNGI KVNFKTRHNI EDGSVQLADH YQQNTPIGDG PVLLPDNHYL STQSALSKDP

       250         260         270         280         290
NEKRDHMVLL EFVTAAGITG GGGGGEIAAI KQEIAAIKQE IAAIKQEIAA IKQG
```

GFP construct 4 (GFP-LI)

M_w: 32.42 kDa

```
        10          20          30          40          50          60
MHHHHHHSSG VDLGTENLYF QSNIGSGLLA SKGEELFTGV VPILVELDGD VNGHKFSVSG

        70          80          90         100         110         120
EGEGDATYGK LTLKFICTTG KLPVPWPTLV TTLCYGVQCF SRYPDHMKRH DFFKSAMPEG
```

```
       130         140         150         160         170         180
YVQERTIFFK DDGNYKTRAE VKFEGDTLVN RIELKGIDFK EDGNILGHKL EYNYNSHNVY

       190         200         210         220         230         240
IMADKQKNGI KVNFKTRHNI EDGSVQLADH YQQNTPIGDG PVLLPDNHYL STQSALSKDP

       250         260         270         280         290
NEKRDHMVLL EFVTAAGITG GGGGGELAAI KQELAAIKQE LAAIKQELAA IKQG
```

GFP construct 5 (GFP-pent)

M_w: 32.88 kDa

```
        10          20          30          40          50          60
MHHHHHHSSG VDLGTENLYF QSNIGSGLLA SKGEELFTGV VPILVELDGD VNGHKFSVSG

        70          80          90         100         110         120
EGEGDATYGK LTLKFICTTG KLPVPWPTLV TTLCYGVQCF SRYPDHMKRH DFFKSAMPEG

       130         140         150         160         170         180
YVQERTIFFK DDGNYKTRAE VKFEGDTLVN RIELKGIDFK EDGNILGHKL EYNYNSHNVY

       190         200         210         220         230         240
IMADKQKNGI KVNFKTRHNI EDGSVQLADH YQQNTPIGDG PVLLPDNHYL STQSALSKDP

       250         260         270         280         290
NEKRDHMVLL EFVTAAGITG GGGGGKIEQI LQKIEKILQK IEWILQKIEQ ILQG
```

GFP construct 6 (GFP-pent-9)

M_w: 33.05 kDa

```
        10          20          30          40          50          60
MHHHHHHSSG VDLGTENLYF QSNIGSGLLA SKGEELFTGV VPILVELDGD VNGHKFSVSG

        70          80          90         100         110         120
EGEGDATYGK LTLKFICTTG KLPVPWPTLV TTLCYGVQCF SRYPDHMKRH DFFKSAMPEG

       130         140         150         160         170         180
YVQERTIFFK DDGNYKTRAE VKFEGDTLVN RIELKGIDFK EDGNILGHKL EYNYNSHNVY

       190         200         210         220         230         240
IMADKQKNGI KVNFKTRHNI EDGSVQLADH YQQNTPIGDG PVLLPDNHYL STQSALSKDP

       250         260         270         280         290
NEKRDHMVLL EFVTAAGITG GGGGGGGGKI EQILQKIEKI LQKIEWILQK IEQILQG
```

GFP construct 7 (GFP-pent9-W)

M_w: 32.97 kDa

```
        10          20          30          40          50          60
MHHHHHHSSG VDLGTENLYF QSNIGSGLLA SKGEELFTGV VPILVELDGD VNGHKFSVSG

        70          80          90         100         110         120
EGEGDATYGK LTLKFICTTG KLPVPWPTLV TTLCYGVQCF SRYPDHMKRH DFFKSAMPEG

       130         140         150         160         170         180
YVQERTIFFK DDGNYKTRAE VKFEGDTLVN RIELKGIDFK EDGNILGHKL EYNYNSHNVY

       190         200         210         220         230         240
IMADKQKNGI KVNFKTRHNI EDGSVQLADH YQQNTPIGDG PVLLPDNHYL STQSALSKDP

       250         260         270         280         290
NEKRDHMVLL EFVTAAGITG GGGGGGGGKI EQILQKIEKI LQKIENILQK IEQILQG
```

GFP construct 8 (GFP-II-Nterm)

M_w: 33.35 kDa

```
        10          20          30          40          50          60
MHHHHHHSSG VDHIAAIKQE IAAIKQEIAA IKQEIAAIKQ EGGGGGGTEN LYFQSNIGSG

        70          80          90         100         110         120
LLASKGEELF TGVVPILVEL DGDVNGHKFS VSGEGEGDAT YGKLTLKFIC TTGKLPVPWP

       130         140         150         160         170         180
TLVTTLCYGV QCFSRYPDHM KRHDFFKSAM PEGYVQERTI FFKDDGNYKT RAEVKFEGDT

       190         200         210         220         230         240
LVNRIELKGI DFKEDGNILG HKLEYNYNSH NVYIMADKQK NGIKVNFKTR HNIEDGSVQL

       250         260         270         280         290         300
ADHYQQNTPI GDGPVLLPDN HYLSTQSALS KDPNEKRDHM VLLEFVTAAG ITHGMDELYN
```

GFP construct 9 (GFP-II5hep-Nterm)

M_w: 34.10 kDa

```
        10          20          30          40          50          60
MHHHHHHSSG VDHIAAIKQE IAAIKQEIAA IKQEIAAIKQ EIAAIKQEGG GGGGTENLYF

        70          80          90         100         110         120
QSNIGSGLLA SKGEELFTGV VPILVELDGD VNGHKFSVSG EGEGDATYGK LTLKFICTTG

       130         140         150         160         170         180
KLPVPWPTLV TTLCYGVQCF SRYPDHMKRH DFFKSAMPEG YVQERTIFFK DDGNYKTRAE

       190         200         210         220         230         240
VKFEGDTLVN RIELKGIDFK EDGNILGHKL EYNYNSHNVY IMADKQKNGI KVNFKTRHNI
```

```
       250         260         270         280         290         300
EDGSVQLADH YQQNTPIGDG PVLLPDNHYL STQSALSKDP NEKRDHMVLL EFVTAAGITH

GMDELYN
```

GFP construct 10 (GFP-LI-5hep)

M_w: 33.17 kDa

```
        10          20          30          40          50          60
MHHHHHHSSG VDLGTENLYF QSNIGSGLLA SKGEELFTGV VPILVELDGD VNGHKFSVSG

         70          80          90         100         110         120
EGEGDATYGK LTLKFICTTG KLPVPWPTLV TTLCYGVQCF SRYPDHMKRH DFFKSAMPEG

       130         140         150         160         170         180
YVQERTIFFK DDGNYKTRAE VKFEGDTLVN RIELKGIDFK EDGNILGHKL EYNYNSHNVY

       190         200         210         220         230         240
IMADKQKNGI KVNFKTRHNI EDGSVQLADH YQQNTPIGDG PVLLPDNHYL STQSALSKDP

       250         260         270         280         290         300
NEKRDHMVLL EFVTAAGITG GGGGGELAAI KQELAAIKQE LAAIKQELAA IKQELAAIKQ
```

GFP construct 11 (GFP-LI-9Gly)

M_w: 32.59 kDa

```
        10          20          30          40          50          60
MHHHHHHSSG VDLGTENLYF QSNIGSGLLA SKGEELFTGV VPILVELDGD VNGHKFSVSG

         70          80          90         100         110         120
EGEGDATYGK LTLKFICTTG KLPVPWPTLV TTLCYGVQCF SRYPDHMKRH DFFKSAMPEG

       130         140         150         160         170         180
YVQERTIFFK DDGNYKTRAE VKFEGDTLVN RIELKGIDFK EDGNILGHKL EYNYNSHNVY

       190         200         210         220         230         240
IMADKQKNGI KVNFKTRHNI EDGSVQLADH YQQNTPIGDG PVLLPDNHYL STQSALSKDP

       250         260         270         280         290
NEKRDHMVLL EFVTAAGITG GGGGGGGGEL AAIKQELAAI KQELAAIKQE LAAIKQ
```

A.2 Chapter 3 Constructs

Vector plasmid: pET 28b

Expression cell line: E. Coli BL21 DE3

Oct-MBP

M_w: 76.9 kDa

```
        10         20         30         40         50         60
MGSSHHHHHH SSGLVPRGSH MSYVTTKDGV QIFYKDWGPR DAPVIHFHHG WPLSADDWDA

        70         80         90        100        110        120
QLLFFLAHGY RVVAHDRRGH GRSSQVWDGH DMDHYADDVA AVVAHLGIQG AVHVGHSTGG

       130        140        150        160        170        180
GEVVRYMARH PEDKVAKAVL IAAVPPLMVQ TPGNPGGLPK SVFDGFQAQV ASNRAQFYRD

       190        200        210        220        230        240
VPAGPFYGYN RPGVEASEGI IGNWWRQGMI GSAKAHYDGI VAFSQTDFTE DLKGIQQPVL

       250        260        270        280        290        300
VMHGDDDQIV PYENSGVLSA KLLPNGALKT YKGYPHGMPT THADVINADL LAFIRSGTGG

       310        320        330        340        350        360
LAAIKQELAA IKQELAAIKQ ELAAIKQDGG GENLYFQGGG GGKIEEGKLV IWINGDKGYN

       370        380        390        400        410        420
GLAEVGKKFE KDTGIKVTVE HPDKLEEKFP QVAATGDGPD IIFWAHDRFG GYAQSGLLAE

       430        440        450        460        470        480
ITPDKAFQDK LYPFTWDAVR YNGKLIAYPI AVEALSLIYN KDLLPNPPKT WEEIPALDKE

       490        500        510        520        530        540
LKAKGKSALM FNLQEPYFTW PLIAADGGYA FKYENGKYDI KDVGVDNAGA KAGLTFLVDL

       550        560        570        580        590        600
IKNKHMNADT DYSIAEAAFN KGETAMTING PWAWSNIDTS KVNYGVTVLP TFKGQPSKPF

       610        620        630        640        650        660
VGVLSAGINA ASPNKELAKE FLENYLLTDE GLEAVNKDKP LGAVALKSYE EELVKDPRIA

       670        680        690        700
ATMENAQKGE IMPNIPQMSA FWYAVRTAVI NAASGRQTVD EALKDAQ
```

TriEst

M_w: 32.3 kDa

```
        10          20          30          40          50          60
MGSSHHHHHH SSGLVPRGSH MSYVTTKDGV QIFYKDWGPR DAPVIHFHHG WPLSADDWDA

         70          80          90         100         110         120
QLLFFLAHGY RVVAHDRRGH GRSSQVWDGH DMDHYADDVA AVVAHLGIQG AVHVGHSTGG

        130         140         150         160         170         180
GEVVRYMARH PEDKVAKAVL IAAVPPLMVQ TPGNPGGLPK SVFDGFQAQV ASNRAQFYRD

        190         200         210         220         230         240
VPAGPFYGYN RPGVEASEGI IGNWWRQGMI GSAKAHYDGI VAFSQTDFTE DLKGIQQPVL

        250         260         270         280         290
VMHGDDDQIV PYENSGVLSA KLLPNGALKT YKGYPHGMPT THADVINADL LAFIRS
```

A.3 Chapter 4 Constructs

Vector plasmid: pET 28b

Expression cell line: E. Coli BL21 DE3

Ico2

M_w: 35.9 kDa

```
        10          20          30          40          50          60
MGSSHHHHHH SSGLVPRGSH MSYVTTKDGV QIFYKDWGPR DAPVIHFHHG WPLSADDWDA

         70          80          90         100         110         120
QLLFFLAHGY RVVAHDRRGH GRSSQVWDGH DMDHYADDVA AVVAHLGIQG AVHVGHSTGG

        130         140         150         160         170         180
GEVVRYMARH PEDKVAKAVL IAAVPPLMVQ TPGNPGGLPK SVFDGFQAQV ASNRAQFYRD

        190         200         210         220         230         240
VPAGPFYGYN RPGVEASEGI IGNWWRQGMI GSAKAHYDGI VAFSQTDFTE DLKGIQQPVL

        250         260         270         280         290         300
VMHGDDDQIV PYENSGVLSA KLLPNGALKT YKGYPHGMPT THADVINADL LAFIRSGTKI

        310         320
EQILQKIEKI LQKIEQILQK IEQILQG
```

Ico4

M_w: 36 kDa

```
        10          20          30          40          50          60
MGSSHHHHHH SSGLVPRGSH MSYVTTKDGV QIFYKDWGPR DAPVIHFHHG WPLSADDWDA

        70          80          90         100         110         120
QLLFFLAHGY RVVAHDRRGH GRSSQVWDGH DMDHYADDVA AVVAHLGIQG AVHVGHSTGG

       130         140         150         160         170         180
GEVVRYMARH PEDKVAKAVL IAAVPPLMVQ TPGNPGGLPK SVFDGFQAQV ASNRAQFYRD

       190         200         210         220         230         240
VPAGPFYGYN RPGVEASEGI IGNWWRQGMI GSAKAHYDGI VAFSQTDFTE DLKGIQQPVL

       250         260         270         280         290         300
VMHGDDDQIV PYENSGVLSA KLLPNGALKT YKGYPHGMPT THADVINADL LAFIRSGTGG

       310         320
KIEQILQKIE KILQKIEQIL QKIEQILQG
```

Ico6

M_w: 36.2 kDa

```
        10          20          30          40          50          60
MGSSHHHHHH SSGLVPRGSH MSYVTTKDGV QIFYKDWGPR DAPVIHFHHG WPLSADDWDA

        70          80          90         100         110         120
QLLFFLAHGY RVVAHDRRGH GRSSQVWDGH DMDHYADDVA AVVAHLGIQG AVHVGHSTGG

       130         140         150         160         170         180
GEVVRYMARH PEDKVAKAVL IAAVPPLMVQ TPGNPGGLPK SVFDGFQAQV ASNRAQFYRD

       190         200         210         220         230         240
VPAGPFYGYN RPGVEASEGI IGNWWRQGMI GSAKAHYDGI VAFSQTDFTE DLKGIQQPVL

       250         260         270         280         290         300
VMHGDDDQIV PYENSGVLSA KLLPNGALKT YKGYPHGMPT THADVINADL LAFIRSGTGG

       310         320         330
GGKIEQILQK IEKILQKIEQ ILQKIEQILQ G
```

Ico8

M_w: 36.2 kDa

```
        10          20          30          40          50          60
MGSSHHHHHH SSGLVPRGSH MSYVTTKDGV QIFYKDWGPR DAPVIHFHHG WPLSADDWDA

        70          80          90         100         110         120
QLLFFLAHGY RVVAHDRRGH GRSSQVWDGH DMDHYADDVA AVVAHLGIQG AVHVGHSTGG
```

```
       130         140         150         160         170         180
GEVVRYMARH PEDKVAKAVL IAAVPPLMVQ TPGNPGGLPK SVFDGFQAQV ASNRAQFYRD

       190         200         210         220         230         240
VPAGPFYGYN RPGVEASEGI IGNWWRQGMI GSAKAHYDGI VAFSQTDFTE DLKGIQQPVL

       250         260         270         280         290         300
VMHGDDDQIV PYENSGVLSA KLLPNGALKT YKGYPHGMPT THADVINADL LAFIRSGTGG

       310         320         330
GGGGKIEQIL QKIEKILQKI EQILQKIEQI LQG
```

Ico10

M_w: 36.3 kDa

```
        10          20          30          40          50          60
MGSSHHHHHH SSGLVPRGSH MSYVTTKDGV QIFYKDWGPR DAPVIHFHHG WPLSADDWDA

        70          80          90         100         110         120
QLLFFLAHGY RVVAHDRRGH GRSSQVWDGH DMDHYADDVA AVVAHLGIQG AVHVGHSTGG

       130         140         150         160         170         180
GEVVRYMARH PEDKVAKAVL IAAVPPLMVQ TPGNPGGLPK SVFDGFQAQV ASNRAQFYRD

       190         200         210         220         230         240
VPAGPFYGYN RPGVEASEGI IGNWWRQGMI GSAKAHYDGI VAFSQTDFTE DLKGIQQPVL

       250         260         270         280         290         300
VMHGDDDQIV PYENSGVLSA KLLPNGALKT YKGYPHGMPT THADVINADL LAFIRSGTGG

       310         320         330
GGGGKIEQIL QKIEKILQKI EQILQKIEQI LQG
```

Ico12

M_w: 36.4 kDa

```
        10          20          30          40          50          60
MGSSHHHHHH SSGLVPRGSH MSYVTTKDGV QIFYKDWGPR DAPVIHFHHG WPLSADDWDA

        70          80          90         100         110         120
QLLFFLAHGY RVVAHDRRGH GRSSQVWDGH DMDHYADDVA AVVAHLGIQG AVHVGHSTGG

       130         140         150         160         170         180
GEVVRYMARH PEDKVAKAVL IAAVPPLMVQ TPGNPGGLPK SVFDGFQAQV ASNRAQFYRD

       190         200         210         220         230         240
VPAGPFYGYN RPGVEASEGI IGNWWRQGMI GSAKAHYDGI VAFSQTDFTE DLKGIQQPVL

       250         260         270         280         290         300
VMHGDDDQIV PYENSGVLSA KLLPNGALKT YKGYPHGMPT THADVINADL LAFIRSGTGG
```

```
       310        320        330
GGGGGGKIEQ ILQKIEKILQ KIEQILQKIE QILQG
```

A.4 Chapter 5 Constructs

Vector plasmid: pET 28b

Expression cell line: E. Coli BL21 DE3

T8H4-1

M_w: 36.0 kDa

```
        10         20         30         40         50         60
MGSSHHHHHH SSGLVPRGSH MSYVTTKDGV QIFYKDWGPR DAPVIHFHHG WPLSADDWDA

        70         80         90        100        110        120
QLLFFLAHGY RVVAHDRRGH GRSSQVWDGH DMDHYADDVA AVVAHLGIQG AVHVGHSTGG

       130        140        150        160        170        180
GEVVRYMARH PEDKVAKAVL IAAVPPLMVQ TPGNPGGLPK SVFDGFQAQV ASNRAQFYRD

       190        200        210        220        230        240
VPAGPFYGYN RPGVEASEGI IGNWWRQGMI GSAKAHYDGI VAFSQTDFTE DLKGIQQPVL

       250        260        270        280        290        300
VMHGDDDQIV PYENSGVLSA KLLPNGALKT YKGYPHGMPT THADVINADL LAFIRSGTGG

       310        320        330
GGGGIEKKIE AIEKKIEAHE KKHEAIEKKI E
```

T8H4-2

M_w: 36.8 kDa

```
        10         20         30         40         50         60
MHHHHHHSSG VDLGTENLYF QSNHMSYVTT KDGVQIFYKD WGPRDAPVIH FHHGWPLSAD

        70         80         90        100        110        120
DWDAQLLFFL AHGYRVVAHD RRGHGRSSQV WDGHDMDHYA DDVAAVVAHL GIQGAVHVGH

       130        140        150        160        170        180
STGGGEVVRY MARHPEDKVA KAVLIAAVPP LMVQTPGNPG GLPKSVFDGF QAQVASNRAQ

       190        200        210        220        230        240
FYRDVPAGPF YGYNRPGVEA SEGIIGNWWR QGMIGSAKAH YDGIVAFSQT DFTEDLKGIQ

       250        260        270        280        290        300
QPVLVMHGDD DQIVPYENSG VLSAKLLPNG ALKTYKGYPH GMPTTHADVI NADLLAFIRS
```

```
       310         320         330
GTGGGGGGIE KKIEAIEKKI EAHEKKHEAI EKKIEA
```

T8H4-3

M_w: 36.8 kDa

```
        10          20          30          40          50          60
MGSSHHHHHH SSGGGGGGGE NLYFQGGGHM SYVTTKDGVQ IFYKDWGPRD APVIHFHHGW

         70          80          90         100         110         120
PLSADDWDAQ LLFFLAHGYR VVAHDRRGHG RSSQVWDGHD MDHYADDVAA VVAHLGIQGA

       130         140         150         160         170         180
VHVGHSTGGG EVVRYMARHP EDKVAKAVLI AAVPPLMVQT PGNPGGLPKS VFDGFQAQVA

       190         200         210         220         230         240
SNRAQFYRDV PAGPFYGYNR PGVEASEGII GNWWRQGMIG SAKAHYDGIV AFSQTDFTED

       250         260         270         280         290         300
LKGIQQPVLV MHGDDDQIVP YENSGVLSAK LLPNGALKTY KGYPHGMPTT HADVINADLL

       310         320         330         340
AFIRSGTGGG GGGIEKKIEA IEKKIEAHEK KHEAIEKKIE A
```

MBP-T8H4

M_w: 76.8 kDa

```
        10          20          30          40          50          60
MKIEEGKLVI WINGDKGYNG LAEVGKKFEK DTGIKVTVEH PDKLEEKFPQ VAATGDGPDI

         70          80          90         100         110         120
IFWAHDRFGG YAQSGLLAEI TPDKAFQDKL YPFTWDAVRY NGKLIAYPIA VEALSLIYNK

       130         140         150         160         170         180
DLLPNPPKTW EEIPALDKEL KAKGKSALMF NLQEPYFTWP LIAADGGYAF KYENGKYDIK

       190         200         210         220         230         240
DVGVDNAGAK AGLTFLVDLI KNKHMNADTD YSIAEAAFNK GETAMTINGP WAWSNIDTSK

       250         260         270         280         290         300
VNYGVTVLPT FKGQPSKPFV GVLSAGINAA SPNKELAKEF LENYLLTDEG LEAVNKDKPL

       310         320         330         340         350         360
GAVALKSYEE ELVKDPRIAA TMENAQKGEI MPNIPQMSAF WYAVRTAVIN AASGRQTVDE

       370         380         390         400         410         420
ALKDAQTGGG GGGGGGENLY FQGGHMSYVT TKDGVQIFYK DWGPRDAPVI HFHHGWPLSA

       430         440         450         460         470         480
```

```
DDWDAQLLFF LAHGYRVVAH DRRGHGRSSQ VWDGHDMDHY ADDVAAVVAH LGIQGAVHVG

       490        500        510        520        530        540
HSTGGGEVVR YMARHPEDKV AKAVLIAAVP PLMVQTPGNP GGLPKSVFDG FQAQVASNRA

       550        560        570        580        590        600
QFYRDVPAGP FYGYNRPGVE ASEGIIGNWW RQGMIGSAKA HYDGIVAFSQ TDFTEDLKGI

       610        620        630        640        650        660
QQPVLVMHGD DDQIVPYENS GVLSAKLLPN GALKTYKGYP HGMPTTHADV INADLLAFIR

       670        680        690
SGTGGGGGGI EKKIEAIEKK IEAHEKKHEA IEKKIEA
```