Matthew C. Jenkins

Date

Protein Engineering of Encapsulin Nanocontainers for Biocatalysis Applications

By

Matthew C. Jenkins
Doctor of Philosophy

Chemistry

Stefan Lutz, Ph.D.
Advisor

Vincent Conticello, Ph.D.
Committee Member

Khalid Salaita, Ph.D.
Committee Member

Accepted:

Lisa A. Todesco, Ph.D.
Dean of the James T. Laney School of Graduate Studies

Date

Protein Engineering of Encapsulin Nanocontainers for Biocatalysis Applications

By

Matthew C. Jenkins

B.A., Emory University, 2010

Advisor: Stefan Lutz, Ph.D.

An abstract of
A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Chemistry

2020

Abstract

Protein Engineering of Encapsulin Nanocontainers for Biocatalysis Applications

By Matthew C. Jenkins

Nature has been shown to repeatedly employ proteinaceous containers spanning a broad range of length scales (i.e. approximately 10 to >2000 nm) to suit myriad biological functions including the propagation and infectious behavior of viruses, creating distinct microenvironments designed to facilitate specific metabolic processes, and the generation of metabolite storehouses for maintaining intracellular homeostasis. Such macromolecular cage assemblies are evolutionary marvels, forming highly symmetrical and monodisperse architectures in a hierarchical fashion from either singular or small subsets of structurally-related protein building blocks. In recognition of their vast diversity in terms of sizes, morphologies, physiochemical attributes, and dynamic functional behaviors, synthetic biologists have increasingly sought to repurpose naturally occurring protein containers for applications in a breadth of biotechnologically-relevant fields. Along these lines, this dissertation specifically focuses on the rational engineering of a recently discovered class of proteinaceous nanocontainers, referred to as encapsulins, in order to generate catalytically functional multienzyme nanoreactors.

The first chapter provides general context for this dissertation by presenting a broad overview of select protein-based container structures found in nature, followed by several brief reviews of therapeutic, catalytic, and biomaterials applications for which these protein-based containers have been employed in recent decades. The second chapter describes efforts to rationally engineer the exterior surface of encapsulin nanocontainers derived from the

hyperthermophilic bacterium *Thermotoga maritima* to present a series of solvent-exposed peptide interaction domains. *T. maritima* encapsulins (TmE) presenting external SpyCatcher covalent interaction domains were shown to capture up to 60 copies of an *Escherichia coli* dihydrofolate reductase (DHFR) variant enzyme both *in vitro* and *in vivo*. Surface-tethered DHFR enzymes maintained catalytic functionality with minimal deviations from their untethered Michaelis-Menten profiles. The third chapter expands upon the DHFR-decorated nanocontainers generated in chapter 2 to construct a bi-enzymatic nanoreactor metabolon in which the reduction of dihydrofolate by DHFR is used to fuel the demethylation of an aryl substrate by LigM, a tetrahydrofolate-dependent aryl-*O*-demethylase enzyme isolated from *Sphingomonas paucimobilis* SYK-6, which was encapsulated within the TmE lumen. The resulting bi-enzymatic nanoreactors were shown to be functional, though mutations previously used to enlarge the 5-fold symmetry pores natively distributed throughout the TmE shell were needed to facilitate efficient exchange of pathway metabolites between the interior and exterior spaces.

The fourth chapter shifts focus to describe attempts to generate a novel *in vitro* cargo loading mechanism for TmE under benign solvent conditions by abrogating the native *in vivo* container assembly process using engineered steric obstructions. Recombinant fusion of a bulky protein domain to the lumen-oriented N-terminus of TmE was shown to prevent full container assembly, and subsequent proteolytic liberation of the TmE coat proteins resulted in rapid initiation of container assembly. However, cargo loading attempts in tandem with protease treatment proved unsuccessful, highlighting the need for refinement of the steric-based assembly design.

The final chapter provides a general summary of the works presented in the preceding chapter. Additionally, several commentaries concerning the successes and failures of the nanocontainer engineering strategies employed in this dissertation are presented, along with general opinions pertaining to possible future directions within the field of nanocontainer engineering.

Protein Engineering of Encapsulin Nanocontainers for Biocatalysis Applications

By

Matthew C. Jenkins

B.A., Emory University, 2010

Advisor: Stefan Lutz, Ph.D.

A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Chemistry

2020

Acknowledgements

Having spent a decade working in Stefan's lab as both a technician and then as a graduate student, I have had the great fortune of working with a number of individuals over the years who have helped to shape both my early career as a scientist and me as an individual. I would first like to thank my advisor, Dr. Stefan Lutz, for his mentorship, his constructive criticisms, and his unwavering dedication to my education. I would also like to thank my committee members, Dr. Khalid Salaita and Dr. Vincent Conticello, for their academic input and their continued support. Additionally, I would like to thank Dr. Emily Weinert for her invaluable knowledge, insights, and encouragement.

I would next like to thank several lab members whose influences have truly made my graduate experience a joy. First, I would like to thank my initial mentor Dr. Joel Panay for his unending patience and for his unforgettable teaching style. I could not have asked for a better introduction into laboratory science. Next, I would like to thank Dr. Ashley Daugherty, Dr. Parvin Muthu, Dr. Leann Teadt, Dr. Charles Modlin, and Dr. Spencer Hughes, whose camaraderie I will always cherish. I would also like to thank Dr. Roberto Orru, Dr. Elsie Williams, and Dr. Huanyu Zhao, whose knowledge and assistance were indispensable. I also want to thank the current lab members, Dr. Samantha Iamurri, Evy Kimbrough (and Poppy), David White, Ayda Gonzalez de la Nuez, Tamra Blue, and Kendra Ireland, who have made coming to lab each day both fun and memorable. It has been a pleasure working with each and every one of you.

Next, I would like to thank my parents for their immeasurable love, their belief in me, and their dedication to helping me succeed no matter what. I would also like to thank my brother for his friendship, his support, and the escapism he has provided through our weekly D&D games. Finally, I want to especially thank my wife Michelle and our son Anson. To Michelle: you are the love of my life and I never could not have done this without you by my side. You have been my

rock, celebrating the highs with me and helping me to endure the lows. I thank you from the bottom of my heart. To my son Anson: in these last days of my graduate career, your smiles and the glimmer in your eyes as you begin to explore the world around you have made each day a little brighter for me as your father. I love you son.

Table of Contents

List of Figures

List of Tables

List of Commonly Used Abbreviations

5-CH_3-THF	N^5-methyltetrahydrfolate
AaLS	*A. aeolicus* Lumazine Synthase
Aftn	Apoferritin
Bfn	Bacterioferritin
BMC	Bacterial Microcompartment
CA	Carbonic Anhydrase
CAC_{app}	Apparent Critical Assembly Concentration
CCMV	Cowpea Chlorotic Mottle Virus
Clp	Encapsulin C-terminal Cargo-loading Peptide
CPA	Cleavage-prompted Assembly
CPMV	Cowpea Mosaic Virus
DHF	Dihydrofolate
DHFR	*E. coli* Dihydrofolate Reductase
DHFR*	*E. coli* DHFR with the M42W Point Mutation
DHFR*-CST	*E. coli* DHFR* Variant with a C-terminal SpyTag Ligand
DHFR*-NST	*E. coli* DHFR* Variant with an N-terminal SpyTag Ligand
DHFR-NST(DA)	*E. coli* DHFR with a Non-functional N-terminal SpyTag Ligand

DLS	Dynamic Light Scattering
Dps	DNA-binding Protein from Starved Cells
EP	Encapsulation Peptide
FLP	Ferritin-like Protein
Ftn	Ferritin
GdnHCl	Guanidine Hydrochloride
HBV	Hepatitis B Virus
IMAC	Immobilized Metal Affinity Chromatography
LbL	Layer-by-Layer Assembly of Nanocontainers
LigM-SC	*S. paucimobilis* SYK-6 LigM with a C-terminal SpyCatcher Domain
LS	Lumazine Synthase
MBP	*E. coli* Maltose Binding Protein
MBP-SC	*E. coli* MBP with a C-terminal SpyCatcher Domain
MetE	*C. roseus* Methionine Synthase
MRI	Magnetic Resonance Imaging
MTX-F	Fluorescein Methotrexate
NP	Nanoparticle
p-ABG	*p*-(aminobenzoyl)-L-glutamate

PBD	Peptide-binding Domain
PCA	Protocatechuate
Pdu	1,2-propanediol Utilizing Bacterial Microcompartment
PDZ	PSD95/DlgA/Zo-1 Peptide-binding Domain
PET	Positron Emission Tomography
PL	Peptide Ligand
RS	Riboflavin Synthase
RuBisCO	Ribulose-1,5-bisphosphate Carboxylase/Oxygenase
SEC	Size Exclusion Chromatography
sfGFP	Superfolder GFP
sfGFP-ST	Superfolder GFP with an N-terminal SpyTag Ligand
SH3	Src Homology 3 Peptide-binding Domain
siRNA	Silencing RNA
TEM	Transmission Electron Microscopy
TEV	Tobacco Etch Virus
THF	Tetrahydrofolate
TmE	*T. maritima* Encapsulin
TmE-SC	TmE with a C-terminal SpyCatcher Domain

TmE-SC:DHFR*-CST	TmE-SC with Bioconjugated DHFR*-CST
TmE-SC:DHFR*-NST	TmE-SC with Bioconjugated DHFR*-NST
TmE-SC:DHFR-NST	TmE-SC with Bioconjugated DHFR-NST
TmE-SC●LigM	TmE with Encapsulated LigM Cargo
TmEΔ9Gly2	TmE Nanocontainers Containing Pore-enlarging Δ9Gly2 Mutation
TMV	Tobacco Mosaic Virus
VLP	Virus-like Particle

Chapter 1:

Introduction

1.1 – Overview of Proteinaceous Micro-and Nanocontainers Used in Biotechnology

1.1.1 – Bacterial Microcompartments

Representing perhaps the largest and most sophisticated proteinaceous cage structures found in nature, bacterial microcompartments (BMCs) encompass a class of diverse prokaryotic organelles responsible for co-localizing core sets of functionally-related enzymes within the luminal spaces of the protein cages in order to sequester specific metabolic pathways from the larger cytosolic environment (1-3). BMCs are ubiquitous in nature with recent bioinformatics analyses having positively identified BMC operons in organisms from 23 of the currently established bacterial phyla, indicating that approximately 20 to 25% of all species within the bacterial kingdom possess these organelle macromolecules based on current estimates (4, 5). Identified BMC operons within bacterial genomes typically consist of two primary gene cassettes with one cassette encoding the genes for several related structural proteins required to construct the BMC outer shell and another cassette encoding the core enzymes responsible for the intended metabolic function of the BMC (1). Additional genes encoded in the operons can include regulatory proteins, substrate-specific importer proteins designed to increase cellular uptake of BMCs' target substrates, and cytoskeletal-associated proteins responsible for uniformly distributing assembled BMCs throughout the bacterial cytosol (1, 6). In total, most BMC operons consist of a total of roughly 10 to 20 genes tightly clustered in the host genome (3, 7-10). Several studies have subsequently suggested that BMC operons behave as superloci, resulting in the wide distribution of BMCs throughout the bacterial kingdom via facile horizontal transfer of BMC gene clusters among bacterial species (11). This assertion is strengthened by the fact that roughly 40%

of BMC-containing bacteria sharing common natural habitats (e.g. anaerobic soil, marine waters, etc.) have also been shown to contain multiple BMC operons within their genomes (11).

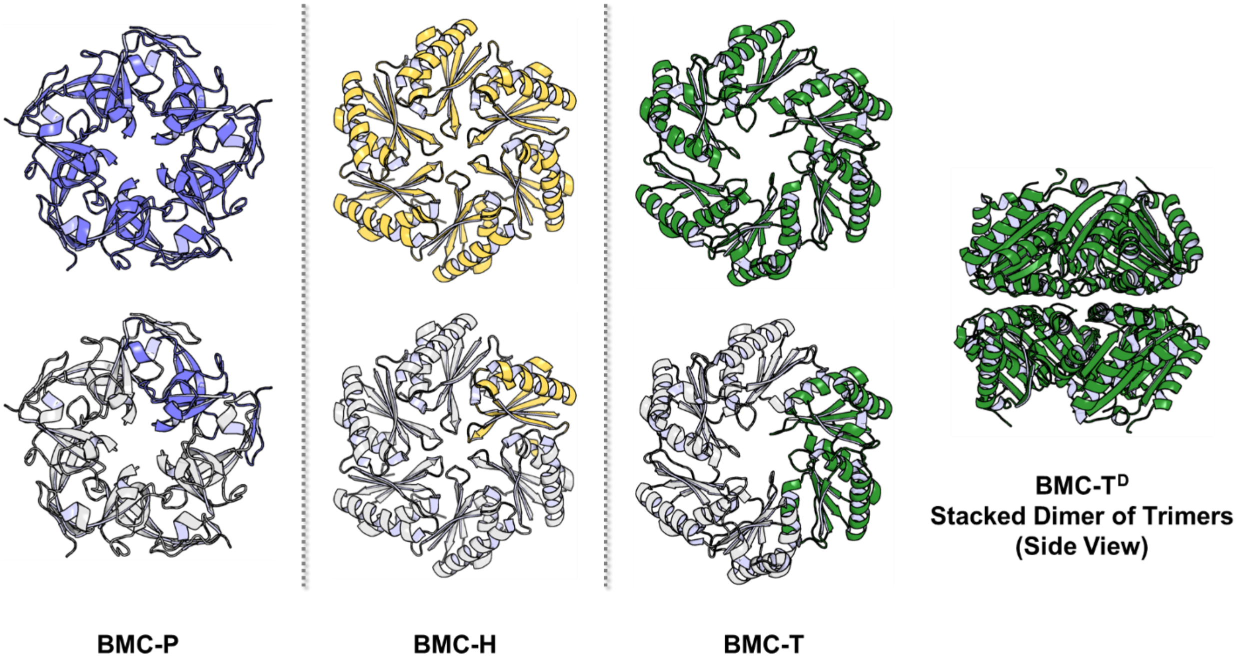


Figure 1.1 – Atomic structures of BMC-P, BMC-H, and BMC-T coat proteins used in the construction of BMC outer shells. All coat protein oligomers are displayed in the top down view, with the exception of the BMC-T^D dimer of trimers on the far right. The BMC-P (PDB ID: 2QW7) and BMC-H (PDB ID: 3BN4) structures above originate from the β-carboxysome of *Synechocystis* sp. PCC 6803 while the BMC-T (PDB ID: 4HT5) and the BMC-T^D (PDB ID: 4HT7) originate from the β-carboxysome of *Synechococcus elongatus* sp. PCC 6301.

Structurally, BMCs are heteropolymers comprised of roughly 10,000 to 20,000 copies of several highly conserved shell protein subunits that self-assemble *in vivo* to form polyhedral cages approximately 100 to 500 nm in external diameter (2, 5, 12) (**Figure 1.1**). Specifically, three classes of shell proteins, denoted BMC-H (Pfam00936), BMC-P (Pfam03319), and BMC-T (tandem fusion of Pfam00936 proteins), are used for BMC construction (13-16). BMC-H proteins are the most abundant shell components, forming planar hexameric tiles that subsequently polymerize in an edge-wise fashion to create the BMCs' polyhedral facets, while BMC-P proteins form puckered pentamers designed to cap the resulting vertices where the facets meet in the assembled BMC structure. BMC-T proteins appear sporadically in the BMC facets, forming trimeric “pseudo-hexamers” that are structurally and functionally similar to the hexameric tiles produced by BMC-H subunits. Both BMC-H hexamers and BMC-T pseudo-hexamers possess pores approximately 4 to 7 Å in diameter located at their respective central axes, which serve important roles in establishing the selective-permeability intrinsic to BMC shells (15, 17-20). Interestingly, both the BMC-H and BMC-T proteins have been shown to also form vertically stacked dodecameric tiles that are believed to be further involved in the dynamic regulation of BMC shell permeability. In the former case, BMC-H hexamers that are not incorporated into the BMC superstructure have been shown to reversibly bind in a slightly offset manner on top of BMC-embedded hexamers to form the dodecameric stacks, which occludes the central pore of the embedded BMC-H hexamer and allows the BMCs to rapidly restrict metabolite flux across the

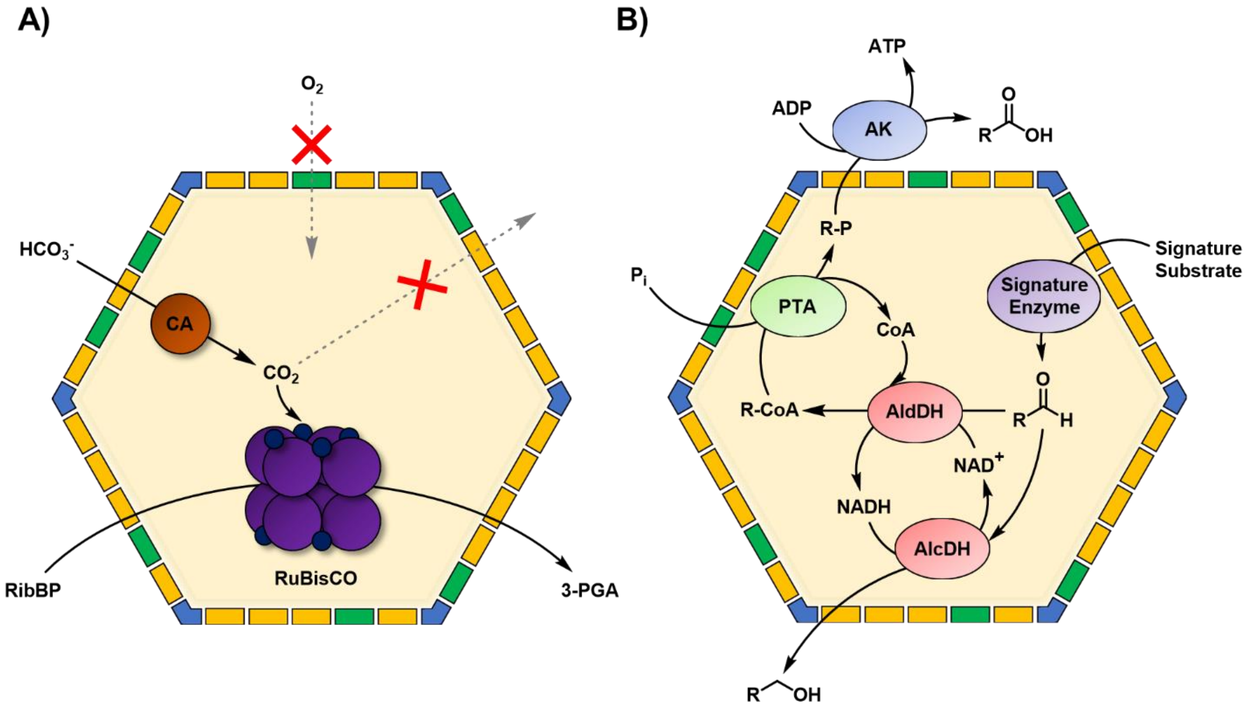


Figure 1.2 – Cartoon representations of the native metabolic functions of A) carboxysomes, and B) metabolosomes. The shell protein colors in the cartoons correspond to the colors used for the individual shell components depicted in **Figure 1.1**. Figure adapted from Ref. (1). Abbreviations: CA – carbonic anhydrase; RuBisCO – ribulose 1,5-bisphosphate carboxylase/oxygenase; HCO_3^- – carbonic acid; CO_2 – carbon dioxide; O_2 – molecular oxygen; RibBP – ribulose-1,5-bisphosphate; 3-PGA – 3-phosphoglycerate; AlcDH – alcohol dehydrogenase; AldDH – aldehyde dehydrogenase; NAD(H) – nicotinamide adenine dinucleotide; PTA – phosphotransacylase; CoA – coenzyme A; AK – acetyl kinase; P_i – inorganic phosphate; ATP – adenosine triphosphate; ADP – adenosine disphosphate

BMC shell (21, 22). In contrast, vertical stacking of BMC-T pseudo-hexamers creates a pocket of void volume between the two trimeric pseudo-hexamers, thus creating gated, airlock-style pores in the BMC shell that can be dynamically opened and closed independently on the cytosolic and luminal faces in response to the binding of specific ligands (14, 23, 24).

In terms of their metabolic functions, BMCs can largely be divided into two primary classes consisting of anabolic carboxysomes used to perform carbon dioxide (CO_2) fixation in autotrophs, and catabolic metabolosomes employed by heterotrophs for the utilization of specific niche carbon sources for growth and respiration (1, 25). Carboxysomes are natively found in all photosynthetically-capable cyanobacteria and are used as a CO_2-concentrating mechanism to fuel the Calvin cycle enzymes carbonic anhydrase (CA) and ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO) encapsulated within the BMC lumen (2, 26, 27) (**Figure 1.2A**). To achieve CO_2 concentration, cell membrane transport proteins designed to facilitate carbonic acid (HCO_3^-) influx encoded in the carboxysome operon first act to increase the cytosolic concentrations of HCO_3^- within the host organism. Cationic pores located at the central axes of BMC-H hexameric tiles in the carboxysome shell then draw anionic HCO_3^- and ribulose-1,5-bispohosphate substrates into the BMC lumen where encapsulated CA dehydrates the HCO_3^- to CO_2 (20). The nascent CO_2 then feeds directly into the co-encapsulated RuBisCO alongside ribulose-1,5-bisphosphate to form two equivalents of the three carbon sugar 3-phosphoglycerate via RuBisCO’s carboxylase activity. One equivalent of the anionic 3-phosphoglycerate product subsequently diffuses out of the carboxysome lumen through the same cationic pores facilitating HCO_3^- entry in order to feed into cellular metabolic pathways for energy production while the other equivalent is used to recharge the RuBisCO for further rounds of catalysis. The cationic pores of the carboxysome shell are critically important to the functional success of these organelle structures as they not only promote the influx of anionic HCO_3^- and ribulose-1,5-bisphosphate substrates for CA and RuBisCO, respectively, but they also act as permeability barriers that prevent the transport of non-polar gasses, such as CO_2 and molecular oxygen (O_2), through the BMC shell (2, 27). Thus, volatile CO_2 generated in the luminal space is retained in order to feed the notoriously inefficient RuBisCO enzymes while O_2, the opposing substrate known to fuel RuBisCO's non-productive oxygenase activity, is largely excluded from entering the BMC.

While anabolic carboxysomes utilize their innate permeability restrictions to concentrate a volatile substrate in a specified locus, catabolic metabolosomes largely use their BMC shells to prevent the efflux of cytotoxic metabolites generated within the BMC lumen (28). Specifically, metabolosomes use a common mechanism involving the enzymatic evolution of a reactive aldehyde intermediate from a target substrate that is unique to each metabolosome (**Figure 1.2B**). Given that each metabolosome converts only a single target substrate, the initial aldehyde-forming enzyme is often referred to as a BMC's "signature enzyme" (25). The best studied signature enzymes to date are B_{12}-dependent enzyme complexes that generate aldehydes through cofactor-mediated radical mechanisms, though an increasing number of B_{12}-independent glycyl radical enzymes have been identified as BMC-associated signature enzymes in recent years (29, 30). Following conversion of the target substrate by the signature enzyme, a functionally conserved set of core enzymes are then used to process the resulting aldehyde intermediate into order to ultimately yield product alcohols and phosphate esters that can be utilized for cellular energy production (1, 5, 25).

To date, characterized metabolosomes have predominantly been identified in the genomes of anaerobic soil and enteric bacterial species, and have been accordingly shown to use alternative carbon sources, such as 1,2-propanediol (8, 31), ethanolamine (7), 1-amino-2-propanol (32), choline (33, 34), and the rare sugars L-fucose and L-rhamnose (10, 35) as target substrates for BMC catabolism. Using the 1,2-propanediol utilizing BMC (Pdu) as an example to illustrate target substrate flux though the BMC-associated enzyme cascade, the initial aldehyde intermediate is generated by encapsulated PduCDE, a B_{12}-dependent diol dehydratase, to form propionaldehyde from 1,2-propanediol (**Figure 1.2B**). A portion of the propionaldehyde pool is immediately converted into propionyl-CoA using an equivalent each of NADH and CoA by the aldehyde dehydrogenase PduP. The remainder of the propionaldehyde is used to generate 1-propanol by the alcohol dehydrogenase PduQ, which simultaneously regenerates the NADH cofactor consumed by PduP. The 1-propanol evolved by PduP passes out of the BMC lumen as one of the two final metabolic products while the propionyl-CoA is converted to propionyl phosphate by the phosphotransacylase PduL, which liberates the equivalent of CoA for subsequent rounds of catalysis. The propionyl phosphate then also leaves the BMC lumen where is can enter the host cell's central metabolism via the methylcitrate metabolic pathway for energy production (36). Emphasizing the importance of sequestering the toxic aldehyde intermediate, previous research efforts have shown that interfering with Pdu shell assembly can lead to the accumulation of propionaldehyde in cell culture media to final concentrations between 15 to 20 mM, which in turn causes extensive damage to host DNA and can eventually lead to total culture arrest (28, 31).

Generally speaking, the native encapsulation of β-carboxysome and metabolosome cargo enzymes tends to occur in an "inside-out" fashion in which the individual cargo enzymes coalesce into condensed core aggregate prior to encapsulation (37-40). Recruitment of BMC shell proteins around the pre-formed enzyme core is subsequently accomplished via specific protein-protein interactions that have been shown to occur between the interior surfaces of BMC shell proteins and ~20 amino acid α-helical encapsulation peptides (EPs) predominantly localized to the N-termini of several metabolon enzymes, or on the C-terminus of the CcmN protein in β-carboxysomes (41, 42). Interestingly, not all metabolon enzymes possess EPs; however, non-EP presenting enzymes have been shown to be packaged within BMC containers nonetheless due to "piggybacking" within the condensed enzyme core alongside EP-containing enzymes (43, 44). In contrast to β-carboxysomes and metabolons, the more evolutionarily distinct α-carboxysomes instead actively package cargo enzymes during the process of shell assembly. Specifically, a highly conserved intrinsically disordered scaffolding protein, CsoS2, has been shown to serve the dual purpose of both binding to RuBisCO and CA enzymes while simultaneously recruiting the formation of α-carboxysome shell proteins around the nascent cargo (45). Following recruitment of the BMC shells, short N-terminal EPs present on the large subunit protein of RuBisCO bind to the interior faces of CsoS1C shell proteins with submicromolar affinities to lock the tethered enzymes in place within the BMC lumen (46).

1.1.2 – Microbial Encapsulins

Encapsulins are a recently discovered class of proteinaceous cage structures intrinsic to both bacterial and archaeal species which have several initial similarities to bacterial microcompartments, such as the abilities to self-assemble into polyhedral shells and to encapsulate enzyme cargoes *in vivo* to fulfill specific metabolic functions (47, 48). However, while the diameters of individual BMCs typically fall within the range of 100 to 500 nm, encapsulins are significantly smaller cage assemblies with currently characterized containers ranging from approximately 24 to 42 nm in external diameter (**Figure 1.3**). Accordingly, encapsulins are often referred to as "nanocontainers" or "nanocompartments" to reflect their smaller size. Likewise, whereas BMCs are heteropolymers assembled from several structurally homologous shell proteins, encapsulins are entirely homopolymeric macromolecules consisting of a single protomer shell protein that self-assembles into the final nanocontainer construct (49). The encapsulin protomers (Pfam04454) are typically between 31 to 33 kDa in size and share neither sequence nor structural homology with BMC shell proteins (47, 48). Collectively, these discrepancies indicate that while encapsulins and BMCs may share functionally similar roles in sequestering certain metabolic operations within microbial cellular environments, the two classes of prokaryotic protein cages are evolutionarily distinct (47-49).

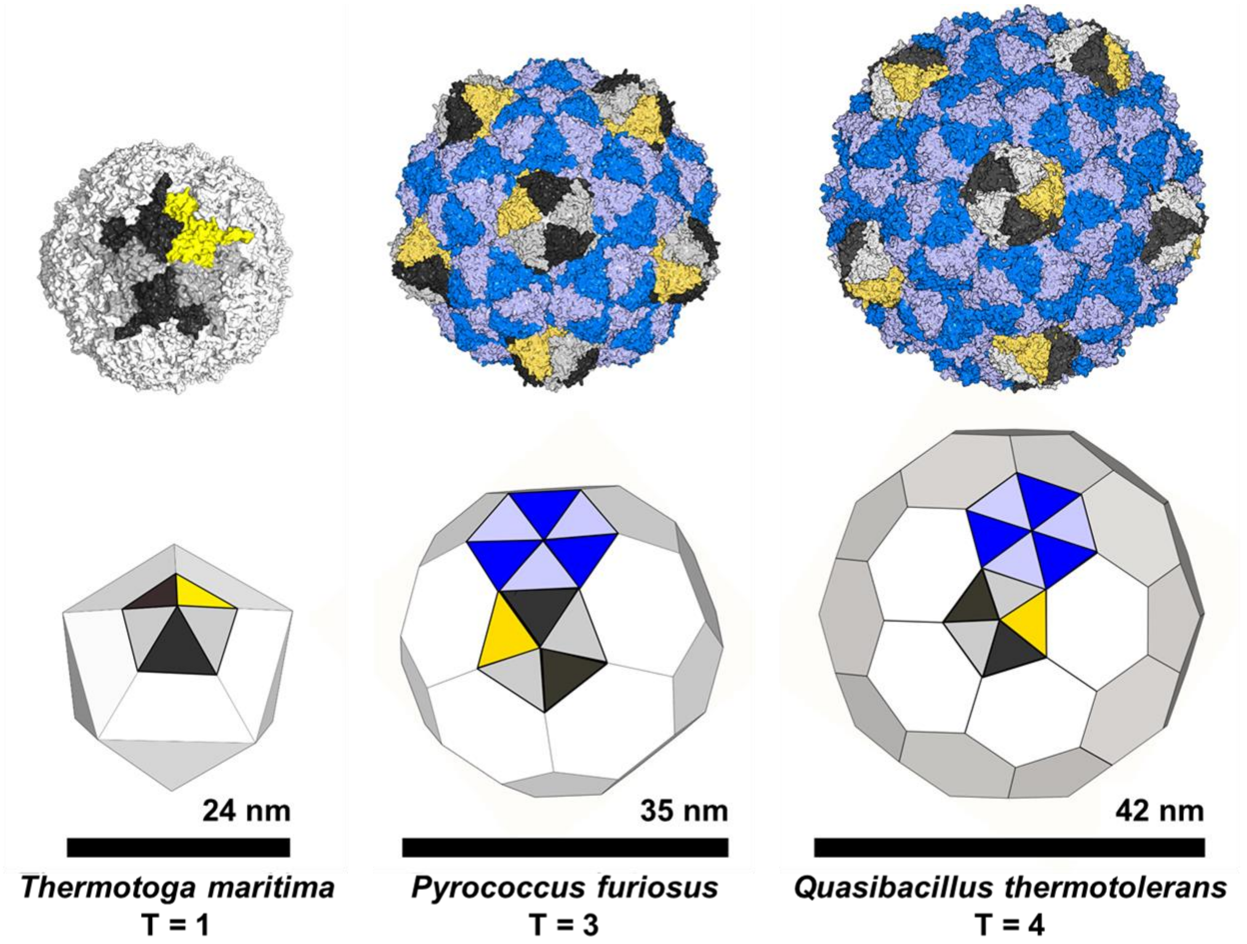
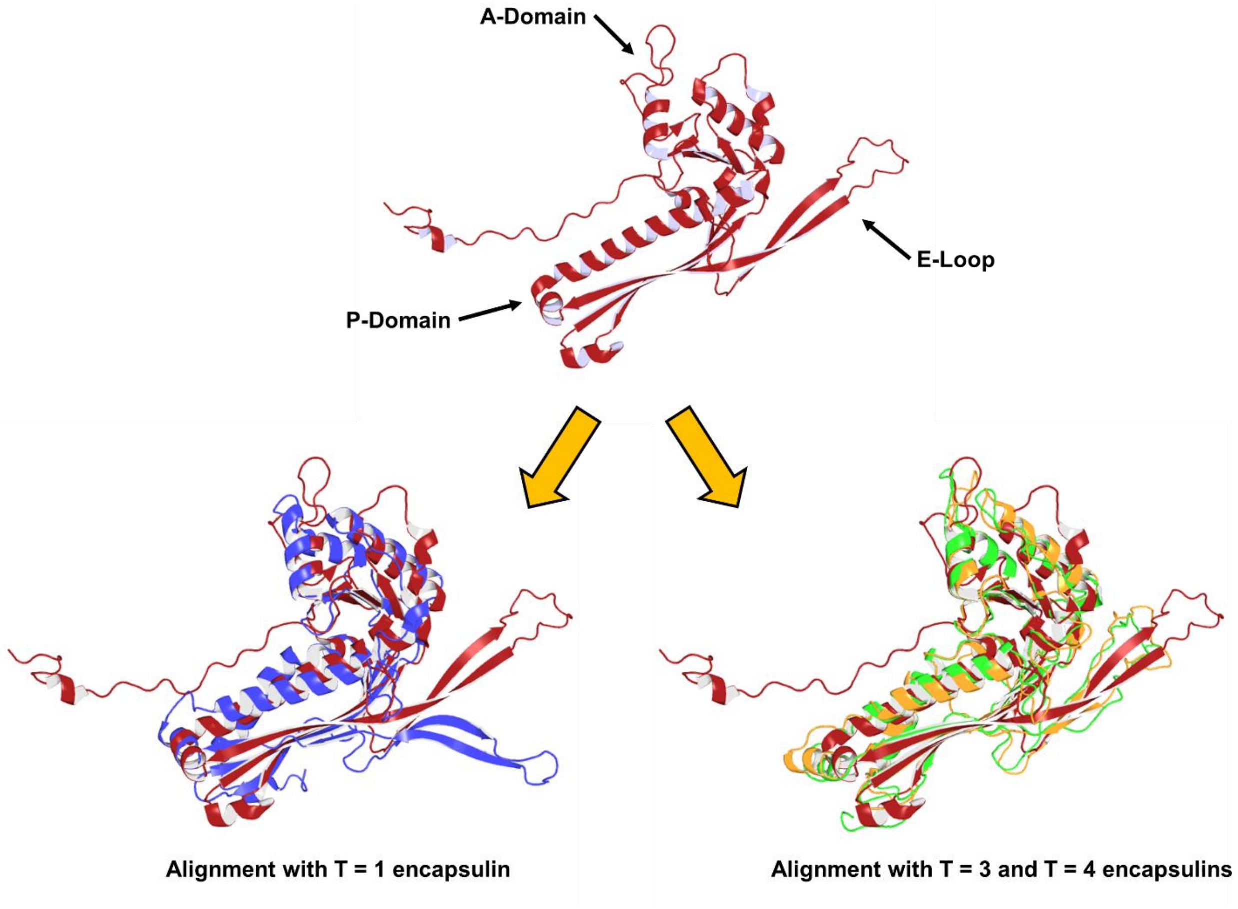


Figure 1.3 – Representative atomic structures of several encapsulins resolved in recent years. Cartoon representations of the containers' overall geometries and quasiequivalent monomers arrangements are depicted below. PDB IDs for presented encapsulins: *T. maritima* encapsulin – 3DKT; *P. furiosus* encapsulin – 2E0Z; Q. thermotolerans encapsulin – 6NJ8.

Genome mining efforts within the last few years have revealed that encapsulin containers are widely distributed among prokaryotes with conservative estimates indicating that between 1 to 3.5% of all microbial organisms likely contain encapsulin nano-organelles at a minimum (48). Given that encapsulin operons are broadly dispersed across microbial species from different phyla and physical environments, it is likely that encapsulin genes, like those of BMCs, are transmitted among prokaryotic species via horizontal transfer rather than vertical inheritance (47-49). Encapsulin operons are usually quite simple, often consisting of a single gene or a two gene cassette in which the gene for the encapsulin protomer and its affiliated cargo protein are either directly linked via genetic fusion or are immediately juxtaposed within the host genome (47, 49-51). In some organisms, however, a primary cargo protein is positioned adjacent to the protomer gene while other functionally related cargo proteins are spread across different loci within the host genomic DNA (52). For cargo proteins that are not fused to encapsulin protomer genes *in vivo*, encapsulation of the translated cargo protein is accomplished in a process referred to as "selective self-sorting" using programmed protein-protein interactions between aliphatic "cargo-loading peptides" (Clp) genetically encoded on one of the cargo protein's native termini (usually the C-terminus) and a corresponding hydrophobic binding cleft located on the luminal face of each encapsulin protomer (49). Of the cargo proteins characterized to date, all have been shown to exhibit some type of antioxidant activity, suggesting that the nascent encapsulin family may play a functional role in combating oxidative or nitrosative stresses within cells (48, 52, 53). The current repertoire of known cargo proteins includes iron storage proteins such as ferritin-like proteins (FLP) and iron-mineralizing encapsulin-associated Firmicute (IMEF) proteins, dye-decolorizing peroxidases, dihydroneopterin aldolases, di-iron hemerythrins, and even multidomain nitrate reductase-hydroxlamine oxidases used for anaerobic ammonium oxidation (48).



Alignment with T = 1 encapsulin

Alignment with T = 3 and T = 4 encapsulins

Figure 1.4 – Structural assessment of the gp5 major coat protein of HK97 (top; PDB ID: 1OHG). The characteristic A-Domain, P-Domain, and extended loop (i.e. E-Loop) substructures are labeled. Alignments of the gp5 coat protein with the T = 1 encapsulin from *T. maritima* (bottom left; PDB ID: 3DKT), and the larger T = 3 and T = 4 encapsulins from *M. xanthus* and *Q. thermotolerans* (bottom right; PDB IDs: 4PT2 and 6NJ8, respectively) are depicted below. The observed deviation in the E-loop orientation of the *T. maritima* encapsulin has been suggested as a possible explanation for its inability to form the quasiequivalent structures of higher T-number encapsulins (47).

Atomic-level structures for the encapsulins from *Thermotoga maritima* (49), *Myxococcus xanthus* (53), *Pyrococcus furiosus* (50), and *Quasibacillus thermotolerans* (54, 55) have been resolved within the last thirteen years. Analyses of structural data for these distinct containers have yielded valuable information pertaining to shell assembly and critical protomer-protomer interactions, cargo loading and distribution, and native physiological functions. Encapsulin protomers adopt a homologous fold to the gp5 major coat protein of the HK97 bacteriophage (56-58) and self-assemble into icosahedral container structures presenting 20 equivalent triangular faces. The number of subunits used to construct the assembled nanocontainers is, however, variable, and consequently dictates the overall size of the final encapsulin structure (characterized encapsulins span the mass range of 1.8 to 9.6 MDa upon assembly) (49, 54). Nanocompartments composed of 60 coat protomers, such as the *T. maritima* encapsulin, are approximately 24 nm in external diameter upon assembly and are constructed from 12 pentameric vertices in accordance with Casper and Klug's T = 1 triangulation symmetry model (59, 49). Larger containers composed of 180 protomers, such as the encapsulins from *M. xanthus* and *P. furiosus*, are approximately 30 to 32 nm in external diameter and are collectively constructed from 12 pentameric vertices and 20 hexameric faces in accordance with a T = 3 symmetry model (50, 53). Finally, the recently discovered container from *Q. thermotolerans* utilizes 240 subunits to form a container roughly 42 nm in diameter, which adopts a T = 4 triangulation arrangement consisting of 12 pentameric vertices and 30 hexameric faces (54, 55). The ability of encapsulin protomers in the T =3 and T = 4 containers to adopt both pentameric and hexameric substructures indicates that these protomers can adopt quasi-equivalent folds in a manner similar to many viral capsids found in nature (59, 60) (**Figure 1.3**). Accordingly, the morphological similarities at both the protomer and macromolecular levels between encapsulins and known virus capsids has led to the speculation that encapsulins likely share a common evolutionary ancestor with many HK97-like viruses (49, 61).

Each nanocompartment, regardless of triangulation number, contains several pores (3-7 Å in diameter) located at three major symmetry axes of the shell structure (49). For most encapsulins, pores located at the two-fold symmetry axis are heavily lined with negatively-charged residues while another set of pores located at the five-fold symmetry axis of the container vertices are more neutral in character. Perhaps the most intriguing of the three pores are those located at the three-fold symmetry axis whose amino acid composition is not conserved and seems to vary across different orthologs (47). The encapsulin from *Q. thermotolerans* deviates from this trend somewhat as all of the native symmetry pores are heavily lined with negatively charged residues, a phenomenon which is likely directly associated with this particular encapsulin's function as an iron mineralizing container (estimates indicate that the *Q. thermotolerans* encapsulin can sequester up to 23,000 iron atoms within the cage interior) (54). While there is still some ambiguity as to which of the multiple symmetry pores on a given encapsulin are critical for biological function, it is thought that collectively these openings are responsible for regulating the flux of small molecules and ions from the cytosol to encapsulin lumen. Simultaneously, these same pores are also thought to serve as a permeability barrier precluding the entry of larger compounds, thus facilitating the creation of organelle-like nanoenvironments *in vivo* (49).

1.1.3 – Virus-like Particles

Viruses constitute the most abundant biological entities on Earth with recent estimates indicating that viral species outnumber both the known bacterial and archaeal species combined by roughly 15-fold (62). As a result of their prevalence and their ability to persist in multifarious biomes across the globe, viruses encompass the largest pool of genetic diversity found in nature, the vast majority of which remains unexplored (62-65). The evolutionary and ecological success of viruses is due, in part, to their sophisticated capsid shells, which simultaneously package and protect the viral genome, facilitate infectivity while mediating host immunogenic responses, and dynamically disassemble and reassemble during the viral replication cycle (66). Given these impressive qualities, synthetic biologists have actively sought to recapitulate similar multifaceted properties in both synthetic and biologically-derived scaffolds for diverse applications in nanomedicine, the development of novel nanobiomaterials, biologically-templated nanoelectronics, and many other fields (65, 67, 68).

Viral capsids are typically constructed form one or several protein subunits that self-assemble into hierarchical cage structures within a host cell during viral replication. Research efforts in recent decades, however, have shown that capsid proteins can often be recombinantly expressed independently from the viral genome to produce monodisperse, hollow cage structures with identical or highly similar morphologies to the parent virus (69, 70). These hollow constructs, referred to as “virus-like particles” (VLPs), are incapable of self-replication in the absence of the full viral genomic material, and are thus entirely non-pathogenic. Additionally, much of the functional plasticity observed for the progenitor viral capsids is preserved in the corresponding VLPs as these properties and pleomorphic behaviors often originate from tertiary and quaternary protein-protein interactions intrinsic to the capsid’s three-dimensional structure (71). Coupled with the fact that viral capsids have been shown to tolerate a variety of genetic and chemical modifications, VLPs have become highly attractive scaffolds for synthetic biology and protein engineering endeavors (70, 72, 73).

As of 2013, approximately 110 unique VLPs had been generated and reported in literature sources (74). Of those reported containers, the majority of VLP nanocages currently used for biotechnology applications originate from plant-pathogenic viruses, animal-pathogenic viruses, or bacteriophages, many of which can be produced in high yield and purity via heterologous expression in common laboratory hosts such as *E. coli* or yeast strains (70, 74) (**Figure 1.5**). As such, these VLPs have been extensively characterized both biophysically and structurally, permitting researchers atomic-level detail when engineering and derivatizing VLP scaffolds (66). Morphologically, commonly utilized VLPs can broadly be divided into two structural classes:

roughly spherical capsids adopting icosahedral symmetries and rod-shaped filamentous assemblies (66, 69). Spherical VLPs typically span the range of 20 to 500 nm in external diameter and include the plant pathogens cowpea chlorotic mottle virus (CCMV), cowpea mosaic virus (CPMV), and Brome mosaic virus (BMV), the animal pathogens Simian vacuolating virus 40 (SV40) and hepatitis B virus (HBV), and the bacteriophages P22, MS2, and Qβ as the most commonly utilized structures (66, 70). Filamentous VLPs, by contrast, are typically between 5 to 20 nm in diameter and can extend to over 2 μm along their longitudinal axes. The two most prevalent filamentous VLPs are the plant-pathogenic tobacco mosaic virus (TMV) and the bacteriophage M13 (66, 70).

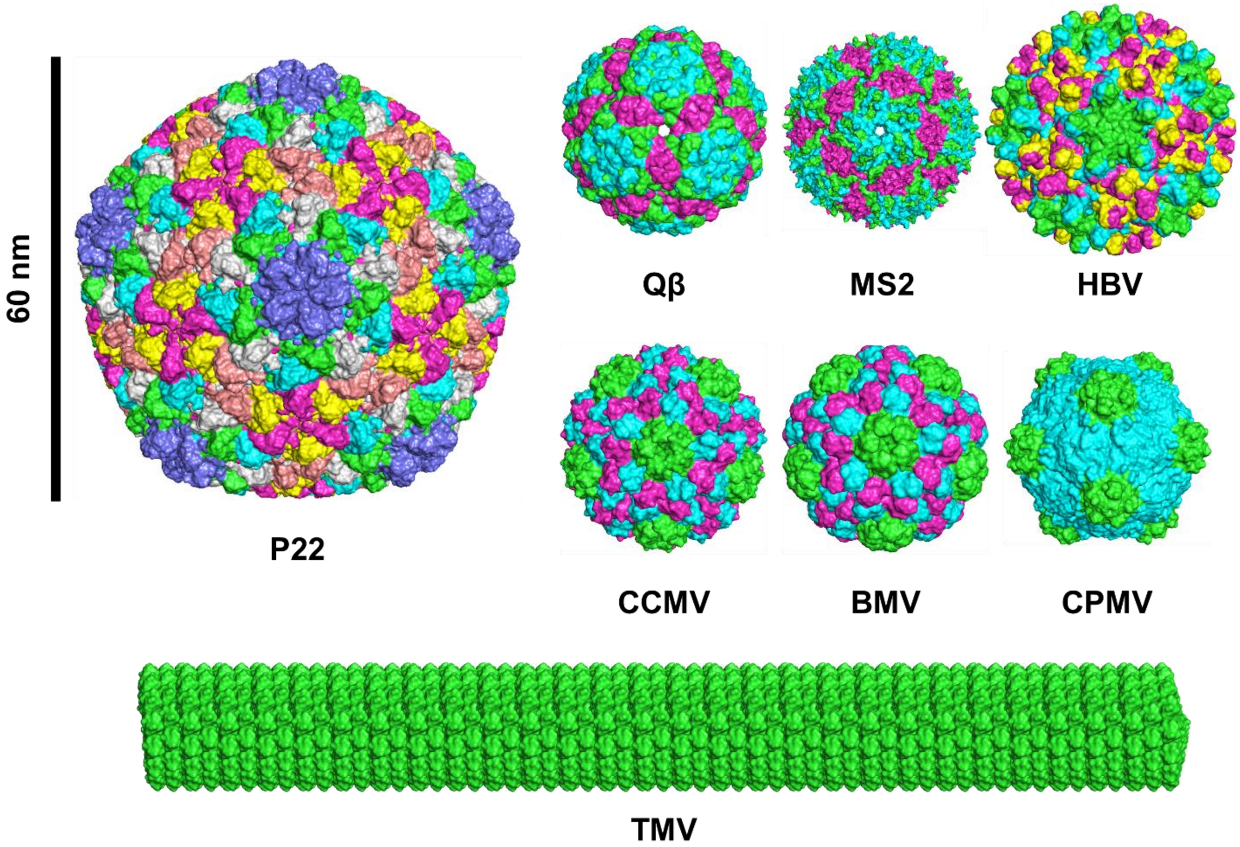


Figure 1.5 – Atomic structures for several of the most commonly employed VLPs in biotechnology. Structures are approximately scaled relative to one another. PDB IDs: P22 – 5UU5; Qβ – 1QBE; MS2 – 2MS2; HBV – 1QGT; CCMV – 1CWP; BMV – 1JS9; CPMV – 1NY7; TMV – 4UDV.

While the current repertoire of VLPs is impressive, it is by no means expansive given that less than 1% of the viral biosphere is estimated to have been explored to date. In light of this fact, novel VLPs possessing unique physiochemical and structural properties will undoubtedly be revealed in the years and decades to come as bioinformatic technologies continue to advance. Indeed, several recent reviews have described new subsets of archaeal viruses abstracted from extremophilic organisms presenting novel capsid architectures previously unobserved in viral communities (75-77).

1.1.4 – Lumazine Synthase

Though animal species must obligately acquire riboflavin (i.e. vitamin B_2) through regular dietary intake, plants and certain microbial species possess the ability to biosynthesize riboflavin *in vivo* as part of their standard metabolism (78, 79). In such organisms, lumazine synthase (LS) and riboflavin synthase (RS) represent the final two enzymes from the riboflavin biosynthetic pathway (80) (**Figure 1.6A**). LS specifically is responsible for the generation of 6,7-dimethyl-8-ribityllumazine (**3**) from 5-amino-6-(D-ribitylamino)uracil (**1**) and L-3,4-dihydroxy-2-butanone 4-phosphate (**2**). Subsequently, two equivalents of 6,7-dimethyl-8-ribityllumazine are used by RS to generate one equivalent of riboflavin (**4**) and an equivalent of **1**, with the latter being recycled for subsequent rounds of catalysis by LS (80, 81). Given that riboflavin biosynthesis is not an endogenous process in the animal kingdom, both LS and RS have been extensively studied in recent decades as potential targets for the development of selective antimicrobial therapeutics designed to treat bacterial and fungal infections (80).

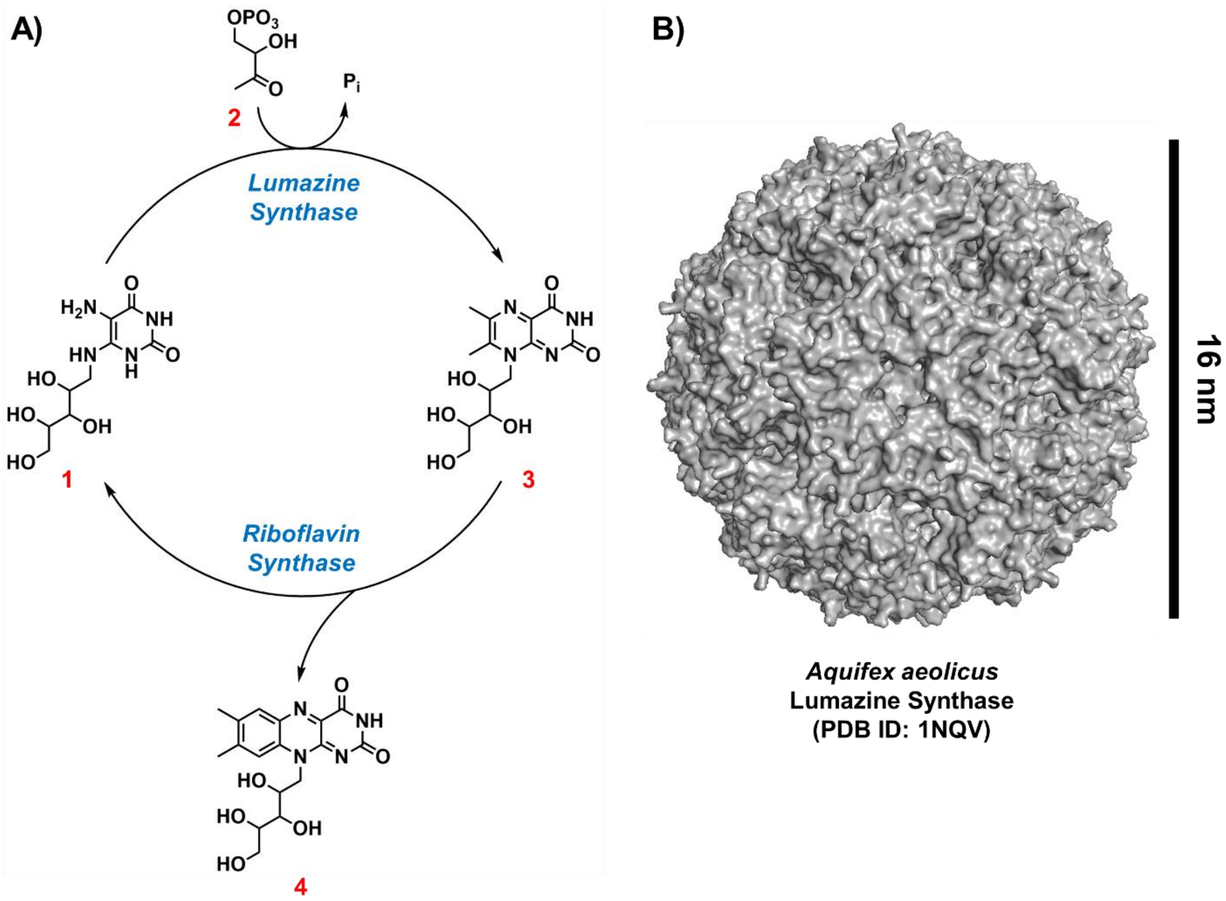


Figure 1.6 – A) Atomic structure of *A. aeolicus* lumazine synthase. B) Final two steps of the riboflavin biosynthetic pathway.

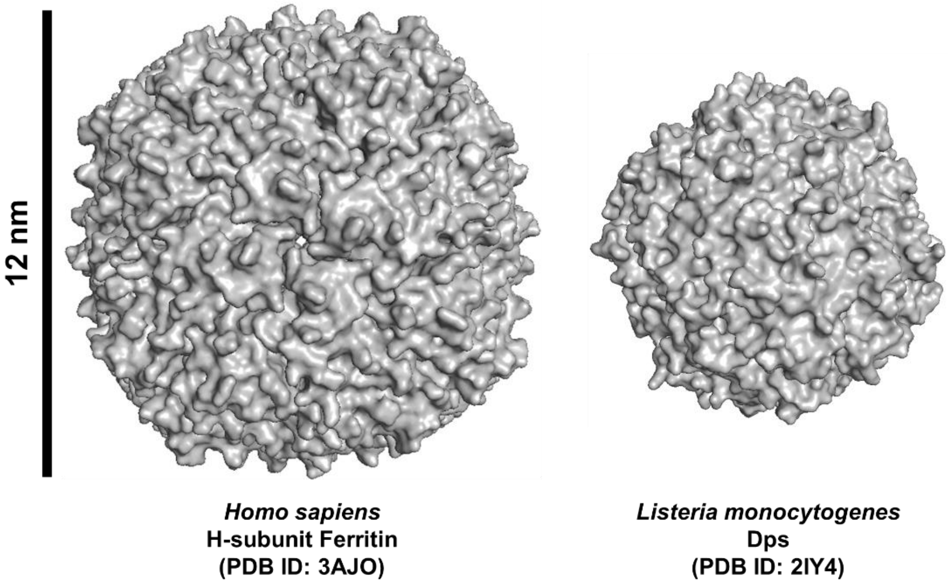
The riboflavin biosynthetic pathway is particularly noteworthy in relation to naturally occurring proteinaceous containers as several crystallographic studies have revealed that LS enzymes natively self-assemble into porous, 60 subunit cage structures as their final quaternary structure in some organisms (82-87) (**Figure 1.6B**). Such cages adopt a T = 1 icosahedral 532 symmetry constructed from 12 pentameric substructures with a final assembled mass of approximately 1 MDa, and usually have external and internal diameters of approximately 16 and 12 nm, respectively (81, 85). Pores in the LS shell are found at the 5-fold symmetry axes located at the center of each pentamer, permitting the exchange of substrates and products between the cage lumen and the cytosol. Combined sedimentation and crystallographic experiments dating back to the 1970s have also revealed that most organisms producing LS cages package a trimer of RS enzymes within the cages to form a bifunctional enzyme complex referred to as the "heavy riboflavin synthase" (82, 88), though *E. coli* has been shown to be an exception as it natively produces a LS cage structure devoid of RS cargo (84). Previous kinetic assessments of LS cages containing internalized RS trimers has shown that cascade catalysis is possible for the two enzymes when initial substrates are present in low concentrations, perhaps representing an evolutionary method designed to ensure a consistent riboflavin supply for the host during both abundant and scarce nutrient conditions (78, 84).

RS trimers are packaged into LS shells *in vivo* using short, C-terminal peptide tags, with the peptide tag of each RS monomer arranged near the 3-fold symmetry axis of the trimeric quaternary structure (89, 90). Given that each LS cage possesses 20 identical 3-fold symmetry axes, it has been speculated that the efficiency of RS packaging is facilitated by symmetry matching of cargo-loading peptides and luminal cargo peptide binding sites (90). This symmetry matching phenomenon has been suggested before for encapsulin nanocontainers possessing ferritin-like protein and dye-decolorizing peroxidase cargoes (49), though whereas the cargo peptide binding site is known for several encapsulins, a defined cargo peptide binding cleft has yet to be identified for LS cages (81, 90). Interestingly, LS shells from several species have been shown to possess pleomorphic behavior in which they can swell into expanded cage structures with external diameters of approximately 30 nm at slightly alkaline pH and under certain buffer conditions (82, 90, 91). While in the expanded state, encapsulated RS cargo can diffuse out of the LS cages, indicating that the expanded shells likely possess increased porosity compared to the compacted LS shells (78, 82). While LS enzymatic activity is maintained for expanded cages, RS activity is obviously lost, suggesting that the ability to reversibly transition between the compacted and expanded quaternary states may represent a form of dynamic regulation for the overall LS/RS complex (91).

Nearly two decades ago, an icosahedral LS cage from the hyperthermophilic organism *Aquifex aeolicus* was identified and subsequently crystallized (85) (**Figure 1.6B**). This particular *A. aeolicus* LS (AaLS) possesses several key qualities that has led to its increasing utilization in synthetic biology pursuits in the years succeeding its discovery. First, the hyperthermophilic nature of the host organism is recapitulated in the AaLS cages, which have been shown to withstand heating up to 120°C due in part to increased numbers of inter-subunit ion-pairs relative to other LS cages (85). Additionally, AaLS has been shown to undergo the same pleomorphic expansions observed previously for LS cages from *B. subtilis* (91). Finally, the 12 amino acid sequence of the C-terminal cargo loading peptide from *A. aeolicus* RS has been identified and has been shown to permit the artificial loading of AaLS capsids both *in vivo* and *in vitro* via fusion onto the C-termini of non-native cargo proteins (89). Collectively, these traits have cemented LS, and especially AaLS, as versatile scaffolds for diverse biotechnology applications.

1.1.5 – Ferritins

Ferritins (Ftn) represent one of the most prevalent (and accordingly, one of the best studied) classes of protein nanocages found in nature due to their universal importance in maintaining intracellular iron homeostasis (92). Intrinsic to all kingdoms of life, the Ftn family of proteins consists of spherical hollow cage structures between 9 to 12 nm in external diameter that are responsible for scavenging Fe(II) ions *in vivo* (**Figure 1.7**), mineralizing them as poorly soluble ferrihydrites within their luminal spaces in order to prevent the Fe(II)-mediated generation of cytotoxic hydroxide radicals from intracellular hydrogen peroxide via the Fenton reaction (i.e. $Fe^{2+} + H_2O_2 \rightarrow Fe^{3+} + OH^- + OH\bullet$) (93-95). All Ftn cages are constructed from either one or several related monomeric proteins, with each monomer having a total mass of approximately 20 kDa and consisting of a structurally conserved 4-helix bundle connected by a series of flexible loops (94, 95). A short fifth helix, known as the E-helix, resides near the C-terminus of each monomer and extends roughly perpendicularly to the central axis of the 4-helical bundle (94, 96). Generally, the various Ftn cages utilize distinct ferroxidase catalytic sites located on the luminal faces of their monomeric subunits to promote the oxidation of Fe(II), using either O_2 or H_2O_2 as the corresponding oxidant (97). However, while all Ftn cages have similar biological functions, three distinct subfamilies of Ftn nanocontainers with unique structural and biophysical properties have emerged in recent decades.



Homo sapiens
H-subunit Ferritin
(PDB ID: 3AJO)

Listeria monocytogenes
Dps
(PDB ID: 2IY4)

Figure 1.7 – Atomic structures of ferritin and Dps nanocages. The two cage structures are approximately scaled relative to one another.

The first subclass of ferritins are the “classical” ferritins found in eukaryotes and some bacterial species (92, 93). These nanocages consist of 24 ferritin monomers that self-assemble into hollow cages adopting an octahedral 432 symmetry structure and possessing external and internal diameters of 12 and 8 nm, respectively. Once assembled, the cages have a total molecular weight of approximately 0.45 MDa, and possess eight defined pores lined with anionic residues at their 3-fold symmetry axes designed to facilitate the influx of cationic Fe(II) (94, 98). While some of the classical ferritins are homopolymeric entities, mammalian ferritins are heteropolymeric, consisting of heavy (H) and light (L) subunits (93). The two subunits are highly similar, sharing 55% sequence identity with one another, though they have been found to serve distinct, yet complimentary roles *in vivo* (99). Specifically, the H-subunits are 182 amino acids in length and contain luminally-oriented ferroxidase sites that bind two iron atoms simultaneously in order to catalyze the oxidation of Fe(II) using molecular O_2 as the oxidant, resulting in the formation of H_2O_2 (93). The L-subunit, which is slightly smaller at a total of 174 amino acids, lacks this ferroxidase site, and instead possesses a series of anionic glutamate residues designed to generate a microenvironment within the ferritin lumen that promotes the mineralization of nascent Fe(III) originating at the H-subunit ferroxidase sites as hydrous ferric oxides (93). By siphoning away the Fe(III) formed by the H-subunits in this manner, the L-subunits effectively accelerate the processing rate of Fe(II) by the H-subunit ferroxidase centers. Once a sufficiently large ferrihydrite core has been formed within the ferritin lumen, additional ferrihydrite mineralization begins to occur at the surface of the iron core, thus limiting the net amount of H_2O_2 generated by the ferroxidase sites within the ferritin. In total, the interior volume of the ferritin cage allows for the accumulation of up to 4000 to 4500 iron atoms as poorly soluble ferrihydrite, though the ferritin shell itself ensures that the mineralized iron core remains soluble within the host organism (93-96). Interestingly, the ratio of H:L subunits incorporated into ferritin shells has been shown to be tissue specific, and represents a means through which Fe(II) remediation can be optimized for specific cell types (99).

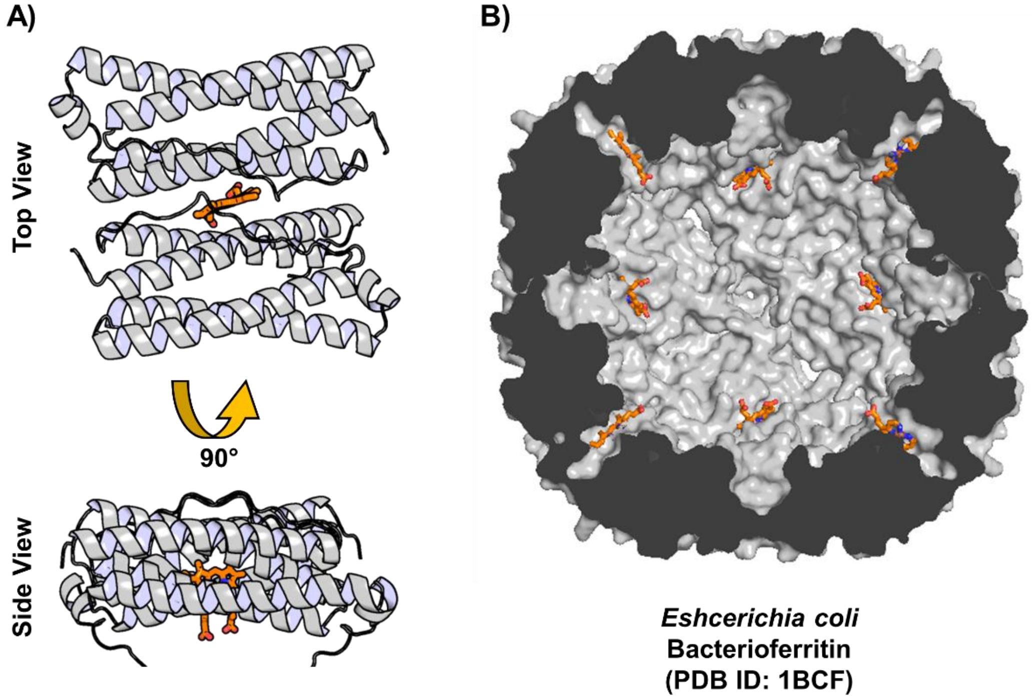


Figure 1.8 – A) Atomic structure of Bfr dimers depicting the heme cofactor (orange sticks) resting in the binding cleft generated at the 2-fold symmetry axis between the two monomers. The side-view below shows the propionate moieties of the heme cofactor extending downward past the luminal face of the dimer. B) Cross-sectional view of the *E. coli* Bfr showing the symmetry of the embedded heme cofactors.

Largely similar to the classical ferritin subclass, bacterioferritins (Bfr) represent the second ferritin subclass and are obligately restricted to the archaeal and bacterial kingdoms (94, 96). Bfr cages generally adopt the same 24-mer superstructure, the same 432 icosahedral symmetry, and the same general dimensions as classical ferritins; however, the Bfr subclass is unique in that many of its members possess 12 b-type heme cofactors bound in distinct clefts located at the containers' 2-fold symmetry axes between two fold-adjacent monomers (**Figure 1.8A**) (96, 100). The heme groups are positioned deep (approximately 13 Å) in the Bfr shell such that their propionate moieties extend into the Bfr lumen (**Figure 1.8B**) (101), and are coaxially ligated into their individual binding clefts by a pair of symmetry-related methionine residues with one methionine originating from each of the juxtaposed monomers (100, 102, 103). This particular methionine residue is not completely conserved among Bfr family members, though, and thus not all Bfr cages are capable of incorporating heme cofactors (96). Subsequent crystallographic and biochemical investigations have revealed that Bfr-associated ferredoxins containing NAD(P)H binding sites and coordinated [2Fe-2S] clusters can bind above the Bfr-bound heme groups in order to transfer electrons directly to the heme (101). The heme groups, in turn, then transfer the ferredoxin-delivered electrons into the Bfr interior, rapidly reducing mineralized Fe(III) on the surface of the iron core back to Fe(II) for iron mobilization and release from the Bfr lumen (104).

The final subclass of ferritins encompasses the DNA-binding proteins from starved cells (Dps), also referred to as “mini-ferritins.” As their name implies, Dps containers are natively induced in nutrient starved stationary phase bacterial cells and function to protect cellular DNA from oxidative damage resulting from Fenton-based radicals (105, 106, 107). Structurally, Dps containers are spherical, dodecameric assemblies adopting a tetrahedral symmetry and possessing total exterior and interior diameters of 9 and 4.5 nm, respectively (**Figure 1.7**). As a result of their smaller dimensions relative to Ftn and Bfr, Dps cages are capable of only mineralizing up to approximately 500 iron atoms per nanocontainer (108). At the monomeric level, structural analyses indicate that Dps proteins likely share an evolutionary lineage with Bfr monomers (95).

Though Dps monomers adopt the same 4-helical bundle fold typical of Ftn and Bfr monomers, several key differences distinguish the Dps proteins from those of their fellow subclasses. First, whereas both Ftn and Bfr monomers possess intrasubunit ferroxidase centers located near the center of their respective 4-helical bundles, Dps cages utilize an intersubunit ferroxidase site that is assembled at the luminal 2-fold symmetry axis from residues belonging to adjacent shell monomers (95, 109). Once assembled into complete cages, Dps containers have similar 3-fold symmetry pores as those found in Ftn and Bfr, though the Dps pores are slightly smaller and possess relatively less net anionic charge density (106). Perhaps the largest structural difference intrinsic to Dps monomers, though, is the presence of an extended segment of polycationic residues on the N-terminus that extends outward into the surrounding solvent (108, 110). These polycationic tags have subsequently been shown to facilitate the binding of expressed Dps cages to host cell DNA in a sequence-independent manner. This physical tethering of Dps containers to host DNA affords the genomic material specific protection from reactive oxygen species both by locally sequestering Fe(II) ions and also by effectively capping the polyanionic charges of the DNA backbone that would otherwise attract Fe(II) via Coulombic attraction (108, 110).

1.1.6 – Synthetic Protein Cage Assemblies

One of the consistent themes of living systems is that biological molecules assembled from simple building blocks often present physical, chemical, and functional properties greater than those of their constituent components. This phenomenon is evident as well in the plethora of protein-based cage assemblies found in nature as such micro and nanoscale containers typically display attributes that are absent in their corresponding protomers, including biophysical robustness, dynamic assembly and disassembly behaviors, specific permeability restrictions, and the capacity to generate distinct microenvironments. Over the last 20 years, researchers have attempted to emulate this nonpareil of evolution through the rational design of protein cages that self-assemble in a bottom-up fashion from either small peptides (111, 112) or from proteins of natural origin which do not natively form cage-like structures *in vivo* (113-119). Macromolecular cages generated in this manner have the potential to present novel morphologies and/or tailored higher-order functions for varied biotechnological applications (66, 120). Several key findings have emerged from these early efforts, the first of which is that the assembly of proteins into macromolecular complexes in nature is largely dictated by the principles of symmetry (120, 121). Specifically, symmetrical designs are believed to be evolutionarily favorable as they inherently minimize the total number of unique interfacial contacts requiring optimization between subunits within an assembly (66, 120, 121). The second finding was that a minimum of two distinct subunit rotations (e.g. 2-fold, 3-fold, etc., referred to as "symmetry groups," and abbreviated C_2, C_3, etc.) are necessary for assembly of complete macromolecular cage structures; propagation of a single symmetric interface, by contrast, tends to yield more simplistic oligomeric structures such as cyclic rings or head-to-tail filaments (121). The primary design challenge researchers must address is thus to design specific protein-protein interactions between protomeric subunits that will yield the

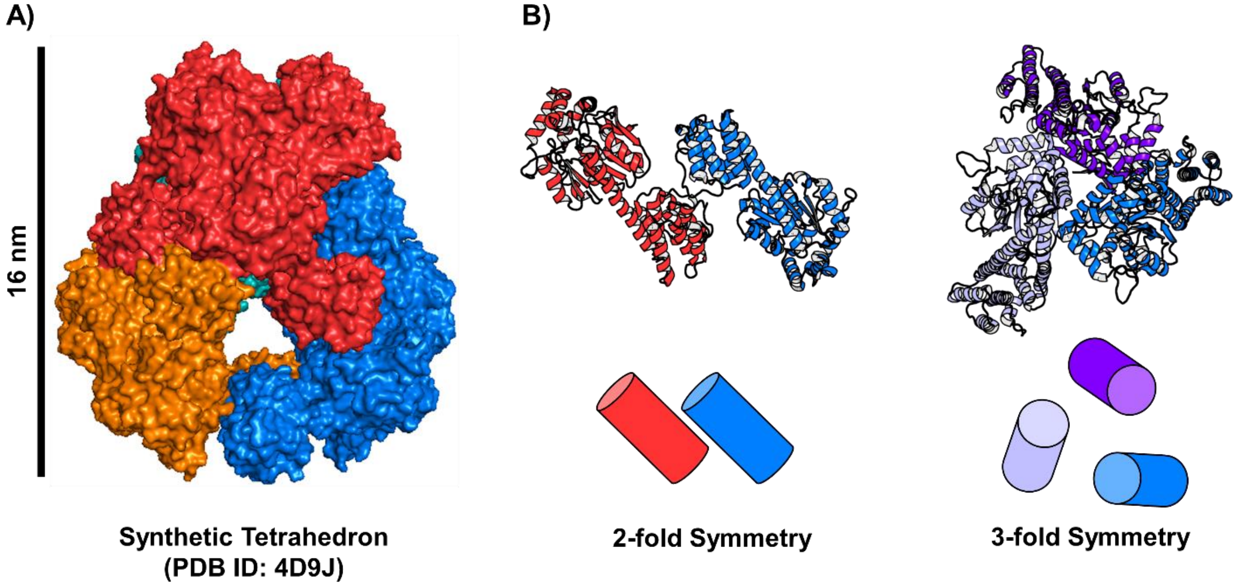


Figure 1.9 – A) Synthetically designed 12-mer tetrahedral cage (114). B) Engineered symmetry interfaces used to generate the synthetic cage. The 3-fold symmetry subunits were taken from the lower-right corner of the full cage structure in A) and the colors were changed slightly for clarity.

necessary symmetry elements at the proper geometries for self-assembly into cage structures (113, 115, 116).

Fortuitously, the design of non-native self-assembling structures can be simplified by employing naturally oligomeric proteins as the composite protomeric subunits, utilizing the natural symmetry interfaces intrinsic to the protomers to aid in cage construction rather than designing the requisite symmetry groups entirely *de novo* (113, 122). Accordingly, two separate methodologies have been employed to accomplish cage assembly using oligomeric protomers. The first method, pioneered by Padilla *et al.* in 2001, involves the genetic fusion of two or more protein subunits that natively display different oligomeric states to fulfill the symmetry requirements for a given architecture (113). As an initial proof-of-principle, the trimeric bromoperoxidase A2 from *Streptomyces aureofaciens* (PDB ID: 1BRO) was fused to a dimeric portion of the Influenza virus matrix protein M1 (PDB ID: 1AA7) using a rigid α-helical linker, which subsequently generated a tetrahedral 12-mer cage upon recombinant expression of the fusion protein in *E. coli* according to TEM and sedimentation analyses (113). This design principle was later refined and used to generated both nanoscale tetrahedral (114, 123) and cubic (115) assemblies with sufficient ordering to obtain atomic-resolution crystal structures. In a similar method, engineered α-helical coils known to self-assemble into C_3, C_4, and C_5-symmetrical coiled-coils were fused onto the C-terminus of a trimeric esterase from *Pseudomonas putida* IFO12996 (PDB ID: 1ZOI) in separate experiments (118, 124, 125). Using flexible peptide linker sequences between the C-terminus of the esterase and the start of the coil sequence, the resulting container assemblies were shown to adopt tetrahedral, octahedral, and icosahedral morphologies using the C_3, C_4, and C_5 coiled-coils, respectively; however, none of these assemblies have been verified at atomic scale to date as the coiled-coils have been shown to yield a mixture of homo-oligomers during recombinant expression, presumably due to the flexibility afforded to the α-helical coils by the designed peptide linkers (118, 124, 125). This problem of homo-oligomerization has been observed in several of previous studies as well, and will require further optimization in future experiments to elucidate the geometric considerations necessary to generate homogeneous assemblies (115).

The second method for generating non-native cage structures involves the use of computational algorithms, such as the symmetry algorithms built into the Rosetta software suite (126, 127), to engineer *de novo* intersubunit interactions in order to generate novel protomer interfaces capable of prompting self-assembly (122). This process typically follows two key steps in which the oligomeric protein subunits are symmetrically docked into a desired target architecture, and then multiple low energy protein-protein interactions are designed at the resulting interfaces between symmetrical units (122). Using this methodology, King *et al.* were successful in constructing both 12-mer and 24-mer cages adopting tetrahedral and octahedral symmetries, respectively, from the same trimeric variant $\text{PduT}_{\text{C38S}}$ coat protein originally native to a propanediol-utilization BMC (PDB ID: 3N79) (122). Crystal structures obtained for both cage architectures revealed that the novel protein-protein sidechain interactions at the engineered interfaces closely matched the computationally designed orientations, though the computational designs were not entirely successful as additional point mutations were subsequently introduced into the tetrameric cage to eliminate a slow oligomeric exchange between the trimeric subunits and the assembled 12-mer cages (122). This same methodology was subsequently used to generate a robust 60-mer wireframe dodecahedron from a trimeric class I KDPG aldolase native to *Thermotoga maritima* (PDB ID: 1WA3); the resulting cage structure was found to be stable up to 80°C, could withstand guanidine hydrochloride treatment up to a concentration of 6.7 M, and was found to be capable of reversible disassembly in the presence of 2.25 M guandinium thiocyanate (128).

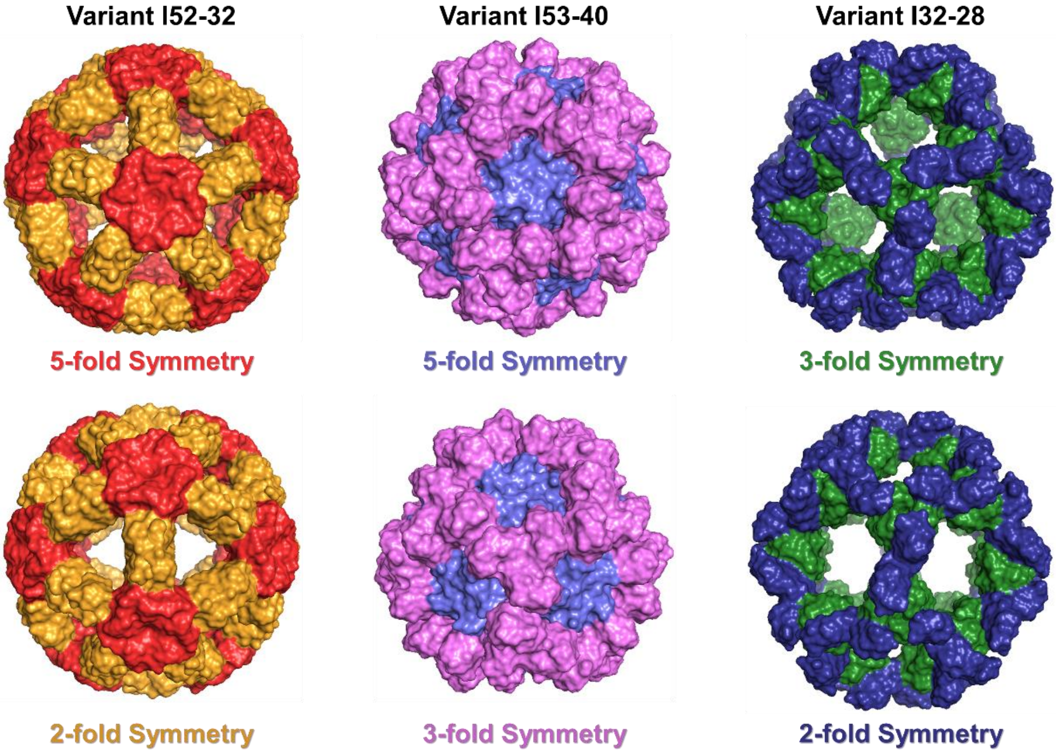


Figure 1.10 – Computationally-designed icosahedral nanocontainers using either $C_5 + C_2$ symmetry interfaces (left, PDB ID: 5IM4), $C_5 + C_3$ interfaces (middle, PDB ID: 5IM5), or $C_3 + C_2$ interfaces (right, PDB ID: 5IM6). All structures originate from (117). Structures not to represented to scale.

Expanding the computational methodology even further, two-component cage structures have also been designed using novel interfaces generated between two different oligomeric protein subunits, which has attractive biotechnological implications as multi-component assemblies have the potential to form a larger repertoire of geometric morphologies that cannot be accessed with a single protomeric subunit (117, 129). In one such study, atomic resolution structures were obtained for two sets of 24-mer architectures adopting tetrahedral point symmetries, which were generated using either four copies of two distinct trimeric protein oligomers ($C_3 + C_3$ subunit symmetries), or four copies of a trimeric subunit in conjunction with six copies of a dimer subunit protein (C_3 +

C_2 subunit symmetries) (129). Interestingly, the tetrahedral cages could be formed both through *in vivo* coexpression of the two oligomeric protomers, or by *in vitro* mixing of separately purified protomers, representing the ability to impart user-defined temporal control over cage assembly (129). Following the initial success with the 24-mer assemblies, larger 120-mer icosahedral cages with molecular weights of 1.8 to 2.8 MDa were constructed in a follow-up study from two-component designs employing paired sets of either 12 pentameric and 20 trimeric protomers (C_5 + C_3), 12 pentameric and 30 dimeric protomers (C_5 + C_2), or 20 trimeric and 30 dimeric subunits (C_3 + C_2) (117). These larger cages (24 to 40 nm in external diameter) were capable of *in vitro* self-assembly like their 24-mer tetrahedral predecessors, and were also capable of internalizing cargo proteins during *in vitro* assembly on the basis of charge complementarity (117).

While the computational design of novel cage architectures represents an incredibly powerful tool for *de novo* nanocontainer design, the methodology is not without its limitations. One particularly persistent problem is that the majority of the both single and multi-component designer cages are insoluble upon recombinant expression in host strains, which necessitates the computational design and screening of large subsets of candidates (117, 128-130). This insolubility factor, which is currently difficult to predict *a priori*, is likely a result of the novel interface designs, which often involve the replacement of hydrophilic residues in favor of hydrophobic patches intended to drive self-assembly (66, 119, 129). Additionally, the algorithmic analyses needed to generate the large subsets of candidate designs are both extensive, time consuming, and require significant computing power (118). However, continuous advances in both computational technologies and algorithmic designs will likely reduce these burdens in the years to come.

1.2 – Therapeutic Applications of Protein Containers

1.2.1 – Polyvalent Vaccine Scaffolds

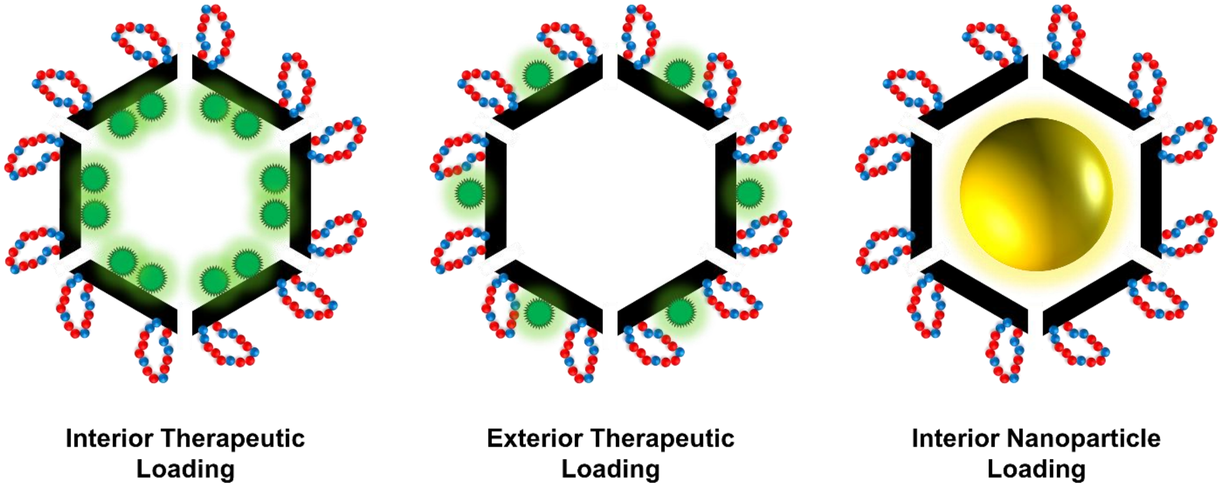
Vaccines are undoubtedly one of mankind's greatest innovations given their widespread and beneficial impacts on human health, wellbeing, and longevity. Historically, the first systematic implementation of prophylactic vaccinations dates back to Edward Jenner's development of a live attenuated smallpox vaccine in the late 18th century (131, 132). Over the last 200 years following this epidemiological breakthrough, vaccination strategies have continually advanced with modern licensed vaccines now including an array of live attenuated whole organisms, inactivated (i.e. deceased) organisms, or isolated protein and polysaccharide epitopes abstracted from pathogenic species (132). Recent estimates from the World Health Organization indicate that current vaccination strategies save up to 2.5 million lives annually across the globe (133), and routine vaccination regimens have either eliminated or can effectively prevent infections from poliovirus, smallpox, measles, mumps, diphtheria, rubella, and hepatitis A (72, 134). In spite of their many successes, however, there still remain a vast number of microbial and viral pathogens for which no effective vaccines have been developed; consequently, vaccine development remains a critical area of study among the scientific, clinical and public health fields.

VLPs, and to a lesser extent other proteinaceous cage structures such as ferritins and encapsulins, have been increasingly incorporated into novel vaccine designs over the last 40 years following the commercial licensing of a VLP-like vaccine derived from the HBV surface antigen (HBsAg) protein in 1981 (135). Several VLP-derived vaccines have subsequently been licensed for prophylactic use in humans or animals, including both Gardasil® and Cervarix® for the prevention of cervical cancers resulting from human papilloma virus infection, Hecolin® for prevention of hepatitis E virus infection, and both Porcilis® PCV and CircoFLEX® for the prevention of porcine circoviruses in commercial swine populations (72, 136). The success of these vaccines is a result of several key structural and practical features of VLPs that make them ideal candidates for vaccine development. As described previously in Section 1.1.3 (above), VLPs are hollow cage structures assembled from recombinantly expressed viral capsid proteins. Given that VLPs are assembled devoid of the native viral genome, they are non-infectious and thus safer than traditional attenuated live vaccines, which can sometimes recover virulence upon introduction into a host (137). Additionally, VLPs mimic the morphological traits of viral capsids, including the high density display and regular ordering of surface exposed antigens. This dense display of surface antigens has been shown to elicit stronger host immunogenic responses relative to monovalent antigen vaccines due to more extensive cross-linking of B-cell surface receptors resulting from polyvalent B-cell/VLP interactions (131, 138). Likewise, the overall size of VLPs, typically within the range of 20 to 200 nm, has been shown to favor their uptake by antigen presenting cells and to permit them direct access to the lymphatic system for effective immune cell targeting (131, 139). Finally, the recombinant expression of VLPs is often facile and scalable in common industrial expression hosts, including *E. coli*, yeast, insect, plant, and mammalian cells, or even in cell-free expression systems (132, 140, 141).

In addition to their capacity to serve directly as antigenic vaccine agents, VLPs and other proteinaceous cages have become attractive scaffolds for the generation of polyvalent chimeric nanocages displaying non-native antigens due to their robustness and their tolerance to both chemical and genetic modifications (70). This concept was proven viable in the late 1980s when a 19 amino acid protein epitope derived from a portion of the VP1 protein originating from the foot and mouth disease virus (FMDV) was recombinantly fused into the VLP-forming hepatitis B core antigen (HBcAg) (142). Chimeric HBcAg-VP1 epitope VLP assemblies were subsequently shown to elicit immunogenic responses nearly as strong as the FMDV itself while the same VP1 epitope provoked a significantly weaker response when administered alone or fused to the N-terminus of a monomeric β-galactosidase carrier protein (142). This particular route of vaccine candidate development holds great promise given the ever increasing sophistication of bioinformatic and molecular biology technologies that have continued to rapidly advance over the last half century.

1.2.2 – Nanoscale Delivery Vehicles and Imaging Agents

The ability to individually target specific tissues and cell types *in vivo* is of great importance in modern medicine for effectively treating various disease states while simultaneously minimizing deleterious off-target side effects. For example, the ability to specifically deliver cytotoxic antineoplastic drugs to tumor cells represents a persistent challenge in numerous chemotherapeutic treatment strategies given that many such pharmaceuticals are non-specific and have well-known propensities to also target rapidly proliferating non-cancerous tissues, such as blood cells, hair follicles, and the gastrointestinal epithelium (143-145). Similar to the development of vaccine scaffolds, many protein-based cages natively possess structural and biophysical features making them well-suited for the development of specific nanoscale vehicles capable of facilitating the delivery of small molecule drugs, genetic materials, antibodies, and therapeutic biocatalysts both *in vitro* and *in vivo* (146-149). Particularly, naturally-derived protein cages are biocompatible, they are within the appropriate size range for efficient cellular uptake via endocytosis, they have the potential to support high loading capacities of therapeutic drugs and biologics due to their large surface areas and internal volumes, and they can be specifically modified to present a variety of targeting ligands, including peptides, aptamers, and carbohydrates, on their exterior surfaces for tissue-specific delivery of their cargoes (134, 144, 148, 150). Several studies have also shown that conjugation of hydrophobic drug molecules to proteinaceous nanocages can increase their solubility, thereby increasing both their bioavailability and efficacy *in vivo* (134, 149, 151).



Interior Therapeutic Loading

Exterior Therapeutic Loading

Interior Nanoparticle Loading

Figure 1.11 – Cartoon representations of several commonly applied nanovehicle designs. In all designs above, the porous protein cages are depicted as black hexagons while the red and blue surface loops represent cell-specific targeting ligands decorated on the cages' exterior surfaces. In the left-hand and middle designs, therapeutic cargoes are represented by the glowing green spheres. In the right-hand design, a model nanoparticle is represented by the glowing golden sphere.

Generally, drug molecules are either chemically functionalized onto protein cage exterior surfaces or are confined within the cages' lumens for therapeutic delivery applications (70, 144). Internalization of drug molecules presents the added benefits of protecting the drug from the surrounding environment while also preventing harmful cytotoxic effects due to non-specific drug exposure during circulation of the nanocarrier vehicle (144). Luminal loading of protein nanocontainers is typically accomplished using either pores natively distributed throughout the nanocontainers' shells, or by reversibly disassembling and reassembling the containers in the presence of the drug molecule (146, 149, 150). Additionally, therapeutic cargoes can be either passively encapsulated, or they can be directed into a container's lumen using either some form of internal targeting ligand or via chemical conjugation (70, 149). For example, given that the native function of viral capsids is to encapsulate viral genetic material, multiple groups have successfully loaded recombinant VLP nanocontainers with silencing RNAs (siRNAs) or antisense oligonucleotides by allowing the nucleic materials to passively bind to the viral coat proteins' luminal faces (152-155). In a particularly eloquent example, Ashley *et al.* utilized both passive and directed encapsulation strategies to package and deliver several therapeutic cargoes to Hep3B human hepatocellular carcinoma cells *in vitro* using MS2 VLPs presenting surface-conjugated SP94 and H5WYG peptide targeting ligands (152). Specifically, the MS2 nanovehicles were charged with either siRNAs (i.e. passive packaging) or several RNA oligonucleotide-labeled therapeutic cargoes and imaging agents (i.e. directed packaging), including cadmium selenide (CdSe) and zinc sulfide (ZnS) quantum dots, the ricin toxin A-chain, and the anticancer drugs doxorubicin, cisplatin, and 5-fluorouracil (152). Further examples in which protein nanocontainers were utilized for the targeted delivery of therapeutics are listed in **Table 1.1**, below.

As alluded to in the example above, proteinaceous nanocontainers have also gained significant attention in recent decades as modular scaffolds for tissue-specific cell imaging applications. Similar to the delivery of therapeutic drugs and biologics, nanocontainers employed for cellular imaging are first either decorated or filled with a variety of molecular contrast agents, including fluorophores, quantum dots, and metals. The loaded nanovehicles can then be directed towards certain cell types *in vivo* using externally-presented targeting ligands for use in fluorescence-based imaging, positron emission tomography (PET), and magnetic resonance imaging (MRI) technologies (68, 156, 157). The use of protein nanocontainers as delivery

Nanocontainer	Therapeutic Cargo	Targeting ligand	Reference
Ferritin/Apoferritin	Cisplatin	Antibody against melanoma antigen CSPG4	Falvo *et al.* (2013)
	Trastuzumab	Fc-binding peptide	Kang *et al.* (2012)
	Doxorubicin ZW800	RGD4C peptide	Zhen *et al.* (2013)
CCMV	Doxorubicin	Folic acid	Barwal *et al.* (2016)
	IR-780 iodide dye	F3 peptide	Wu *et al.* (2017)
CPMV	Proflavine	--	Yildiz *et al.* (2013)
HBV	siRNA	RGD peptides	Choi *et al.* (2013)
M13	Doxorubicin	SPARC binding peptide	Ghosh *et al.* (2012)
MS2	Cisplatin Doxorubicin 5-fluorouracil Ricin-toxin A chain siRNA	SP94 peptide	Ashley *et al.* (2011)
	Antisense oligonucleotide	Transferrin	Wu *et al.* (2005)
	Porphyrin	Jurkat leukemia T-cell aptamers	Stephanopoulos *et al.* (2010)
	MicroRNA 146a	HIV1 Tat$_{47\text{-}57}$ peptide	Pan *et al.* (2012)
	AuNP	EGFR affibody peptide	Kwon *et al.* (2014)
P22	Aldoxorubicin	EGFR affibody peptide HER2 affibody peptide	Kim *et al.* (2019)
	Bortezomib	SP94 peptide	Min *et al.* (2014)
Qβ		Human epidermal growth factor	Pokorski *et al.* (2011)
	Porphyrin	B-cell targeting glycan ligand	Rhee *et al.* (2012)
TMV	Doxorubicin	RGD peptides	Tian *et al.* (2016)
T. maritima encapsulin	Doxorubicin	SP94 peptide	Moon *et al.* (2014)
Aquifex aeolicus lumazine synthase	Doxorubicin Bortezomib	RGD4C peptide SP94 peptide	Min *et al.* (2014)
	Doxorubicin	Fc-binding peptide	Kim *et al.* (2016)

Table 1.1 - Recently employed protein nanocontainers as delivery vehicles for therapeutic agents. Abbreviations: CCMV – cowpea chlorotic mottle virus; CPMV – cowpea mosaic virus; HBV – hepatitis B virus; TMV – tobacco mosaic virus.

platforms for contrast agents is particularly appealing due to several key factors. First, the nanoscale sizes, uniform morphologies, and large payload capacities of nanocontainers allows them to carry high concentrations of imaging agents within small spatial loci, which can greatly improve contrast resolution for more accurate imaging (156-159). Additionally, the biocompatibility of many protein cage species allows both the cages and their affiliated cargoes to be cleared from the body quickly, whereas heavy metal quantum dots and paramagnetic nanoparticles (NPs) have been shown to persist in certain tissues where toxic leaching of heavy metals can occur over time (156, 157, 160). Finally, the ability of protein containers to stabilize metal NPs of varying sizes has also been shown to enhance MRI imaging by significantly increasing the particles' transverse and longitudinal relaxivities (161, 162). **Table 1.2** below details several recent studies describing the use of various nanocontainer platforms as targeted delivery vehicles for contrast imaging agents.

Nanocontainer	**Imaging Agent**	**Targeting ligand**	**Reference**
Ferritin/Apoferritin	Fluorophore ^{64}Cu	RGD4C peptide	Lin *et al.* (2011)
	Gadolinium	C3d peptide	Crich *et al.* (2006)
	Co(II)-doped iron oxide NP	Melanoma targeting peptide	Fantechi *et al.* (2014)
	Fe_3O_4 NP	RGD peptide	Kitagawa *et al.* (2017)
	Fe_3O_4 NP	EGFR affibody peptide	Li *et al.* (2012)
CPMV	Fluorophore	Pan-brombesin peptide analog	Steinmetz *et al.* (2011)
	Fluorophore	VEGFR-1 peptide ligand	Brunel *et al.* (2010)
HK97	Fluorophore	Transferrin	Huang *et al.* (2010)
MS2	QDs	SP94 peptide	Ashley *et al.* (2011)
	Fluorophore	Jurkat leukemia T-cell aptamers	Tong *et al.* (2009)
P22	Fluorophore	EGFR affibody peptide HER2 affibody peptide	Kim *et al.* (2019)
PVX	Fluorophore	GE11 peptide	Chariou *et al.* (2015)
TMV	Gadolinium Fluorophore	VCAM-1 targeting peptide	Bruckman *et al.* (2015)
T. maritima encapsulin	Fluorophore	SP94 peptide	Moon *et al.* (2014)
	Fluorophore	Fc-binding peptide	Moon *et al.* (2014)

Table 1.2 - Recently employed protein nanocontainers as delivery vehicles for contrast imaging agents. Abbreviations: CPMV – cowpea mosaic virus; PVX – potato virus x; TMV – tobacco mosaic virus.

1.3 – Protein Containers as Nanoscale Catalytic Scaffolds

1.3.1 – Enzymatic Catalysis

Over the last two decades, significant research efforts have been devoted to repurposing natural protein cage architectures of varying sizes and morphologies to create micro- and nanoscale storehouses for functional enzyme biocatalysts (i.e. nanoreactors). In particular, synthetic biologists have sought to generate novel catalytic scaffolds and biomaterials for *in vitro* and *in vivo* applications by exploiting the inherent biocompatibility, plasticity, and monodispersities of naturally-derived protein containers (163, 164). In many reported cases, encapsulation affords several distinct benefits to non-native enzyme cargoes, such as increased thermostability, resistance to proteolytic degradation, the prevention of catalyst loss due to aggregation, and even tolerance to elevated concentrations of organic solvents (165-169). Additionally, several research groups have shown that the concentrations of non-native cargoes can be user-modulated, allowing researchers to tune both the molecular crowding of guest enzymes and, by extension, the catalytic properties of the resulting nanoreactor (170-172). For example, Azuma *et al.* recently showed that the encapsulated concentration of a computationally engineered retroaldolase variant could be linearly controlled up to a maximum of 45 copies of enzyme within an engineered lumazine synthase nanocontainer (170). The authors subsequently found that at low enzyme packing densities (~8 enzymes per container), the k_{cat} of the retroaldolase catalysts was unaffected while the K_M for a fluorogenic test substrate was increased 4.5-fold. However, increasing the concentration of retroaldolase enzymes further resulted in concurrent decreases in the measured k_{cat} value.

Similar to the non-native therapeutic and imaging agents discussed in Section 1.2 above, biocatalyst cargoes can also be packaged into nanocontainers both *in vivo* and *in vitro* using an assortment of recombinant and post-translational means. As a general overview, enzymes can be non-specifically encapsulated by stochastic capture during *in vitro* assembly of capsid shells (173, 174), through natural or engineered non-covalent interactions between the capsid interior surface and the enzyme cargo, and through covalent means of tethering the enzyme within the container lumen (175). Natural non-covalent encapsulation relies on repurposing lumen-orienting peptide or nucleic acid affinity ligands derived from either capsid scaffolding proteins or from the native biological cargo for a given container, such as the EPs of metabolosome BMCs, the Clps of microbial encapsulins, or the scaffolding protein of the P22 VLP (41, 42, 47, 176). Alternatively, engineered interactions are typically accomplished using paired coiled-coil interactions, Coulombic attraction between the biocatalyst and the container interior, or particularly in the case of VLPs, RNA aptamers designed to advantageously imitate the native genomic cargo of a viral capsid (81, 177, 178). By contrast, covalent attachment of cargoes within nanocontainers is typically accomplished by recombinantly fusing the cargo enzyme to a luminally-oriented terminus of the nanocontainer protomer, by Sortase A-mediated peptide ligation, or through the use of isopeptide bonds formed by SpyCatcher/SpyTag pairs (179-181).

In addition to the encapsulation of single enzymes, the use of nanocontainers as templates for the development of artificial multienzyme pathways performing sequentially-acting catalytic transformations (i.e. metabolons) has also been explored in recent years. In a seminal work in the field, Patterson *et al.* developed a three-enzyme cascade within P22 VLPs by recombinantly fusing a β-glycosidase, a galactokinase, and a glucokinase together with flexible peptide linkers distributed between them (182). The three tethered enzymes were successfully directed into the

P22 lumen via a truncated scaffolding protein motif that natively binds on the capsid interior wall (183). Interestingly, while the three-enzyme metabolon was shown to be functional within the P22 lumen, competitive inhibition of the initial β-glycosidase enzyme was needed in order to artificially repress its efficiency to increase substrate flux through the entire metabolon. These results indicate that the relative catalytic efficiencies and stoichiometries of metabolon enzymes must be considered in multistep pathway design, similar to what has been observed previously when synthetic metabolons were generated using flexible peptide scaffolds (184, 185).

In another example, Giessen and Silver employed the SpyCatcher/SpyTag bacterial superglue system (181) to covalently load a pyridoxal phosphate-dependent tryptophanase, TnaA, and a flavin mononucleotide-dependent monooxygenase, FMO, simultaneously into the lumen of MS2 VLPs *in vivo* for the two-step production of indigo dye from two equivalents of L-tryptophan (179). Measured indigo output from several nanoreactors containing different ratios of the two enzyme catalysts showed that optimal metabolon production occurred when approximately 4 copies of TnaA and 2 copies of FMO were loaded into the MS2 VLPs. A subsequent *in vitro* comparison of the optimized nanoreactors with the same stoichiometric ratio of non-encapsulated enzymes showed that overall indigo production was nearly equivalent between the two samples. However, the catalytic activities of the non-encapsulated enzymes dropped to < 30% and < 10% at three and seven days post-purification, respectively, while the MS2-encapsulated enzymes maintained their starting activity levels over the same seven day timespan, indicating that encapsulation greatly extended the functional lifetimes of the individual biocatalysts. Additionally, the indigo-producing nanoreactors were shown to function *in vivo* as well, with *E. coli* host cells expressing the MS2-based metabolons exhibiting a 60% increase in indigo production relative to a parallel *E. coli* strain containing non-encapsulated TnA and FMO enzymes. Collectively, these works indicate that nanocontainer-based metabolons can be utilized as artificial organelle structures *in vivo*, which could be a beneficial addition to preexisting metabolic engineering platforms (186).

1.3.2 – Non-enzymatic Catalysis

In addition to the encapsulation of enzyme biocatalysts, the symmetrical structures and limited internal volumes of protein cages have been increasingly exploited as scaffolds for the size-constrained synthesis of highly uniform metal NPs. This particular application of protein nanocontainers has gained significant interest over the last decade following the discovery that many such metal, metal oxide, and quantum-dot NPs are capable of performing catalytic transformations akin to those of enzymes and traditional chemocatalysts (187). Specifically, Gao *et al.* discovered in 2007 that Fe_3O_4 ferromagnetic NPs, similar to those naturally mineralized within the cores of Ftn nanocages, possess intrinsic peroxidase-like catalytic activities highly

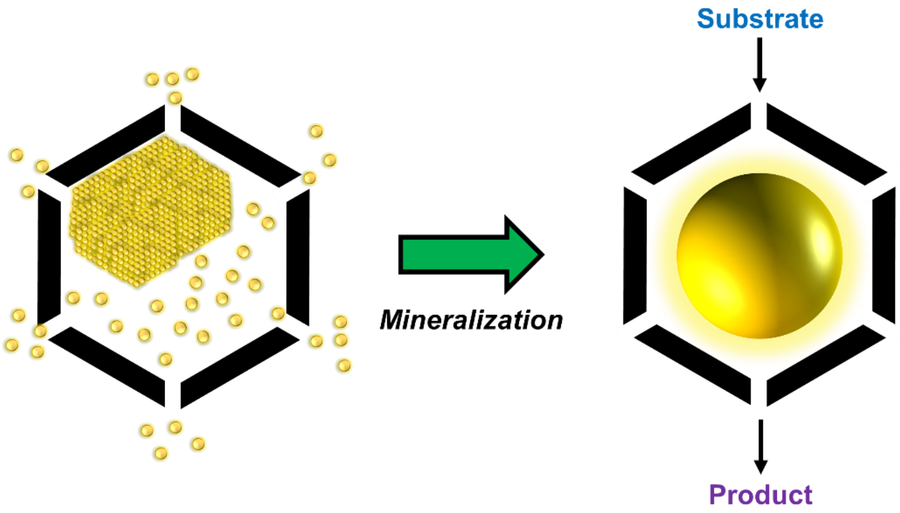


Figure 1.12 – General approach to mineralizing NPs within nanocontainer shells for subsequent nanozyme-based catalysis with the encapsulated NP.

reminiscent of heme-containing peroxidase enzymes (188). In the years following this initial discovery, a variety of nanomaterials possessing inherent enzyme-like catalytic activities, which have since been classified as “nanozymes” (189), have been developed, such as Prussian blue and Mn_3O_4 NPs collectively possessing peroxidase-, catalase-, glutathione peroxidase-, and superoxide dismutase-like activities (187, 190).

Studies with catalytic nanozymes have shown that NP size directly affects the associated catalytic behavior as smaller NPs possess higher surface areas, and thus present a larger number of available reactive sites (188). Accordingly, the ability to develop monodisperse nanozyme NPs within the constrained interior volumes of protein nanocages is highly appealing. Ftn and apoferritin (i.e. ferritins lacking the native ferrihydrite core particles; Aftn) nanocages have been extensively exploited for the development of ~4-6 nm nanozyme NPs in recent years, either through natural Ftn-catalyzed mineralization of various metals, through the accumulation of metal ions within the nanocage interior (mediated by the net anionic charge of the Ftn lumen) followed by treatment with reducing agents such as sodium borohydride ($NaBH_4$), or through the disassembly and reconstruction of the Ftn shell (191, 192). As an example, Ueno *et al.* generated palladium NPs within Aftn cages by reducing Pd(II) metal ions with $NaBH_4$ that were capable of performing the catalytic hydrogenation of a series of olefin compounds in aqueous media (193). Given the relatively small diameter and highly anionic character of Aftn pores, the authors observed that smaller substrates with more cationic character were granted preferential access to the encapsulated Pd-NPs over larger or more anionic substrates. In a similar work, Liu *et al.* successfully encapsulated 4.5 nm nanoceria (CeO_2) NPs within Aftn shells by reversibly disassembling the Aftn cages *in vitro*. Subsequently, encapsulation of the nanoceria NPs within Aftn shells allowed the NPs to be efficiently endocytosed into HepG2 cells *in vivo* via ferritin specific receptor-mediated endocytosis where they were able to catalytically scavenge reactive oxygen species more effectively than superoxide dismutase (194). In addition to Ftn cages, size-constrained NP synthesis has also been accomplished within the larger void spaces of VLPs and encapsulins using either direct capsid-promoted mineralization or specific mineralization initiated by genetically introduced metal-binding affinity tags (195-197).

Within the last decade several protein nanocages have also been used as scaffolds for the specific templating of organometallic chemocatalysts. Ftn and Aftn cages have been the predominantly employed nanocages for these applications due to their particular affinities for the direct coordination of various metals (198-200). Abe *et al.* showed that the Aftn cages generated using the L-subunits from horse liver (L-subunits lack the catalytic ferroxidase center, but retain the metal accumulation center; see Section 1.1.5 above) were capable of coordinating a 2,5-norbornadiene-rhodium(I) ligand on their interior surfaces following incubation of Aftn cages with 2,5-norbornadiene-rhodium(I) chloride dimer complexes in a 9% acetonitrile aqueous solution (198). Subsequently, the rhodium-functionalized Aftn cages were used to catalyze the size-constrained synthesis of phenylacetylene polymers within the containers' luminal spaces. Several years later, Maity *et al.* employed a similar strategy with the same horse liver L-subunit Aftn cages to coordinate both palladium and iridium-containing chemocatalysts within lumen of Aftn in order to perform the two-step generation of biphenyl ethanol from Suzuki-Miyaura coupling and hydrogenation reactions in an aqueous medium using 4-iodoacetophenone as the initial substrate (199). While the overall reaction yield was low (~10% total conversion), a control reaction performed under the same aqueous conditions with the corresponding $[IrCp^{*}Cl_2]_2$ and $[Pd(allyl)Cl]_2$ complexes in the absence of the Aftn cage showed that only the palladium-mediated coupling reaction occurred (~85% conversion), indicating that coordination of the iridium catalyst with the Aftn cage provides stabilization for the chemocatalyst in aqueous solvent.

1.3.3 – Dual Chemo-biocatalytic Cascades

Given that numerous protein containers have been derivatized with both biocatalysts and chemocatalysts, several groups have recently used protein cages as scaffolds to template combined chemo-biocatalytic reaction complexes. Patterson *et al.* recently employed the P22 VLP as a scaffold capable of housing both a covalently-tethered rhodium-based chemocatalyst and a non-covalently scaffolded NADH-dependent alcohol dehydrogenase enzyme (AdhD) within the container lumen (201). The authors showed that the encapsulated AdhD biocatalysts were able to successfully reduce acetoin to 2,3-butanediol, generating NAD^+ in the process. Subsequently, the covalently-encapsulated rhodium complex was used to regenerate the NADH cofactor to fuel additional rounds of AdhD-mediated acetoin conversion by coupling the reduction of NAD^+ to the oxidation of sodium formate with carbon dioxide liberated as the oxidized byproduct. While the combined chemo-biocatalyst cascade was functional, the authors did note that the activity of the encapsulated rhodium-complex was approximately 48-fold slower as compared to the non-encapsulated chemocatalyst.

In a more recent work, Bari *et al.* successfully utilized the highly uniform exterior surface topology of Pdu BMCs as nucleation sizes for the templated formation of catalytically active gold NPs (AuNPs) (202). Purified Pdu metabolosomes exposed to auric chloride were shown to subsequently facilitate the growth of ~2-3 nm AuNPs evenly distributed across the Pdu exterior surfaces. Interestingly, not only were the surface-bound AuNPs able to catalyze the reduction of

4-nitrophenol to 4-aminophenol on the container exterior, but the Pdu compartments also retained their native metabolic function as they were observed to readily convert 1,2-propanediol into propionaldehyde. While the biocatalytic and chemocatalytic reactivities of these augmented BMCs were not directly coupled to one another, they were nonetheless compatible and showed that the BMC shell could serve as an efficient barrier between the two catalytic species. This type of physical separation of biocatalysts and chemocatalysts using the shells of protein cages may be advantageous for future designs as it has been well established in a number of one-pot dynamic kinetic resolution experiments involving organometallic catalysts and lipases that the two catalysts must often be separated to prevent destruction of the organometallic complexes or poisoning of the lipases (203, 204).

1.4 – Novel Biomaterials from Nanocontainer-derived Supramolecular Assembly

1.4.1 – Applications and Advantages of Supramolecular Assembly

As the breadth of research efforts into the catalytic functionalization of VLPs, encapsulins, and other nanoscale containers continues to expand, there has naturally emerged an increasing interest in translating these advances with individual nanoreactors into larger scale, multidimensional assemblies. Accordingly, such biological compartments inherently possess many appealing qualities for the construction of higher-order supramolecular structures, such as their exquisite uniformity with respect to their sizes, symmetries, and anisotropic topological features, as well as the ease with which exterior surface modifications can be introduced either genetically or post-translationally (69, 70). As such, the self-assembly of protein nanocontainers into ordered structures on nano- to micrometer length scales presents a number of unique opportunities for the precise templating of biocatalysts, chemocatalysts, and/or catalytic metal nanoparticles into porous, geometrically-defined biomaterials (205-208). Recent works have exemplified some of the unique benefits of such supramolecular array structures, including the generation of artificial multistep catalytic pathways (208), and the dense packing of catalysts within a small spatial locus, allowing for higher overall catalyst loads within a given reaction volume as compared to reactions containing catalysts dispersed homogeneously in solution (207, 209). Additionally, arranging proteinaceous nanoreactors in defined lattice arrays can permit the facile observation of container structural features, including morphological changes resulting from capsid engineering or cargo packaging, using techniques such as cryo-EM (210). Finally, the recovery and reuse of bulk superstructures is typically feasible using gentle separation methods, such as centrifugation or filtration (208), while some superstructure lattices can even be reversibly disassembled and reassembled as needed via select changes to the solution phase conditions (206, 211, 212).

1.4.2 – Two-dimensional Assemblies

Functionally, the simplest methods for the formation of multidimensional biomolecular assemblies typically involve the deposition and subsequent drying of nanocontainers on solid surfaces (213), non-covalent adsorption of containers via electrostatic interactions with a charged surface (209, 214, 215), or the direct covalent attachment of containers to functionalized supports (216, 217). Alternatively, dense two-dimensional (2D) arrays of viral capsids can be electrostatically preassembled at the liquid-air interfaces of Langmuir-Blodgett monolayers and subsequently transferred to solid supports for downstream applications (210, 214, 218, 219). Finally, both top-down and bottom-up nanolithography technologies have been successfully implemented to create specified 2D arrangements of nanocontainers on surfaces for various applications, such as the development of nanoelectronics, biosensors, and catalytically active chips (209, 220-224). One particularly interesting example was recently presented by Koch et al. in which the authors used engineered TMV capsids adsorbed to solid supports as scaffolds capable of binding high quantities of functionalized glucose oxidase and horseradish peroxidase enzymes for the development of a two-step catalytic biosensor (209). Subsequent analysis of the sensor performance revealed that virus-immobilized enzymes exhibited a 45-fold catalytic rate enhancement as compared to equal solution volumes containing the two enzymes non-specifically adhered to the same solid support without the virus scaffold. Additionally, virus-immobilization increased the functional lifetime of the two sensor catalysts such that the biosensors were able to maintain >50% catalytic activity for up to three weeks. In light of these results, the authors conclude that such dense packing and stabilization of target proteins, in conjunction with stable surface adsorption of the viral scaffolds to solid supports, could be used for the development of both flow-cell based sensor and catalytic systems.

1.4.3 – Three-dimensional Assemblies

In addition to the formation of 2D nanocontainer arrays on surfaces, three-dimensional (3D) bulk assemblies can similarly be generated, typically in either a bottom-up layer-by-layer (LbL) fashion originating from a solid support, or as free-floating bulk crystals formed in solution (215, 225). Collective research efforts over the last few decades have elucidated many of the underlying physical principles that dictate the formation these types of superstructures. In almost all cases, synthesis of ordered 3D arrays is principally facilitated by the presence of some form of multivalent ion (i.e. Ca^{2+}, Mg^{2+}, etc.) (211, 226), chemical moiety (i.e. polyionic dendrimers, Ni-NTA conjugates, etc**.)** (227-229), or biological bridging element (i.e. protein binding domains, ssDNA hybridization, etc.) (206, 225, 230-232) capable of non-covalently binding to two or more individual nanocontainers simultaneously. Additionally, extensive research regarding the interplay of solution pH and ionic strength for electrostatic-dependent assemblies has been reported (211, 233).

The construction of LbL assemblies typically mimics the initial formation of 2D lattices as protein nanocontainers are adsorbed onto solid supports via electrostatic or covalent interactions. Higher-order assembly of additional nanocontainer strata is subsequently achieved in an iterative fashion by adhering multivalent bridging mediators to the base monolayer of container particles, followed by the addition of another layer of nanocontainers, and so on (231). In a recent work, Moon *et al.* used the covalent isopeptide bond-forming SpyCatcher/SpyTag system (181) to functionalize the exterior surfaces of AaLS containers with FKBP12 and FRB protein domains, which form heterodimers in the presence of rapamycin (234). Using the FKBP12/FRB-decorated containers for LbL assembly in the presence of rapamycin led to facile formation of discrete strata. Additionally, by only partially occupying the available SpyTag ligands presented on the AaLS shell, the authors were able to use the remaining free SpyTag ligands to simultaneously functionalize the AaLS containers with β-lactamase-SpyCatcher fusion enzymes. In so doing, LbL assembly promoted the formation of catalytically active superstructures exhibiting successively higher substrate turnover rates in a given reaction volume as the number of nanocontainer strata, and thus the overall amount of lattice-confined β-lactamase, were increased.

In another clever example of superstructure formation, Liljestrom *et al.* advantageously used the difference in the isoelectric points of CCMV capsids (pI ~ 3.8) and avidin (pI ~10.5), along with the anisotropic distribution of the surface charges on the two entities, to electrostatically assemble ordered 3D lattice structures (206). Specifically, the negative charges on the exterior surface of CCMV capsids are largely centered around the 60 pores distributed across the nanocontainer surface while the tetrameric avidin has positively charged patches on each monomer that adopt a tetrahedral arrangement in the oligomeric structure. Given the relatively small size of the avidin tetramers relative to the CCMV capsids (~7 versus ~28 nm, respectively), the two proteins efficiently crystallized into densely packed body-centered cubic lattices on the micrometer length scale. While the crystal formation itself is impressive, the beauty of the electrostatic assembly used to form the superstructures is that the biotin binding sites of the avidin tetramers were left unaltered. As such, the authors subsequently showed that biotinylated fluorophores, gold nanoparticles, and even horseradish peroxidase enzymes could be incorporated into the crystal lattices both pre- and post-formation. Kinetic analyses of the crystals containing horseradish peroxidase showed that both pre- and post-functionalized samples exhibited >2-fold higher specific activity as compared to free enzyme in solution.

1.5 – Aims and Scope of the Dissertation

The general theme of this dissertation focuses on the rational engineering of *T. maritima* encapsulin nanocontainers in order to generate biocatalytic scaffolds capable of performing multi-step catalytic transformations in a facile and effective manner.

Chapter 2 focuses on the genetic incorporation of both non-covalent and covalent peptide-binding domains into the *T. maritima* encapsulin monomer at strategic locations such that the peptide domains are displayed on the external surface of the assembled nanocontainers following *in vivo* translation. These domains were subsequently used in attempts to generate ordered supramolecular assemblies of nanocontainers and to decorate the exterior surfaces of encapsulin nanocontainers with functional enzyme catalysts. In the latter case, detailed Michaelis-Menten assessments were used to elucidate the consequence of enzyme immobilization in regard to catalytic performance.

Chapter 3 expands on the surface functionalization work detailed in Chapter 2 to describe a dual-enzyme nanoreactor generated through the covalent attachment of one enzyme biocatalyst on the encapsulin surface and the simultaneous encapsulation of a second enzyme biocatalyst within the nanocontainer's lumen. Substrate exchange between the two enzymes was investigated using encapsulin variants containing wildtype-sized and genetically enlarged pores in their assembled shells.

Lastly, Chapter 4 details the engineering of encapsulin nanocontainers in order to prevent their natural self-assembly using recombinantly imposed steric obstruction. Subsequently, these variants were used in attempts to perform *in vitro* loading of encapsulin nanocontainers with a non-native fluorogenic cargo protein under benign solvent conditions via proteolytic alleviation of the engineered steric burdens.

1.6 – References

1. Kerfeld, C. A.; Aussignargues, C.; Zarzycki, J.; Cai, F.; Sutter, M., Bacterial microcompartments. *Nature Reviews Microbiology* **2018**, *16* (5), 277-290.
2. Kerfeld, C. A.; Sawaya, M. R.; Tanaka, S.; Nguyen, C. V.; Phillips, M.; Beeby, M.; Yeates, T. O., Protein Structures Forming the Shell of Primitive Bacterial Organelles. *Science* **2005**, *309* (5736), 936-938.
3. Yeates, T. O.; Crowley, C. S.; Tanaka, S., Bacterial microcompartment organelles: protein shell structure and evolution. *Annual Review of Biophysics* **2010**, *39*, 185-205.
4. Axen, S. D.; Erbilgin, O.; Kerfeld, C. A., A Taxonomy of Bacterial Microcompartment Loci Constructed by a Novel Scoring Method. *PLOS Computational Biology* **2014**, *10* (10), e1003898.
5. Bobik, T. A.; Lehman, B. P.; Yeates, T. O., Bacterial microcompartments: widespread prokaryotic organelles for isolation and optimization of metabolic pathways. *Molecular Microbiology* **2015**, *98* (2), 193-207.
6. Savage, D. F.; Afonso, B.; Chen, A. H.; Silver, P. A., Spatially Ordered Dynamics of the Bacterial Carbon Fixation Machinery. *Science* **2010**, *327* (5970), 1258.
7. Kofoid, E.; Rappleye, C.; Stojiljkovic, I.; Roth, J., The 17-Gene Ethanolamine (eut) Operon of Salmonella typhimurium Encodes Five Homologues of Carboxysome Shell Proteins. *Journal of Bacteriology* **1999**, *181* (17), 5317.
8. Bobik, T. A.; Havemann, G. D.; Busch, R. J.; Williams, D. S.; Aldrich, H. C., The propanediol utilization (pdu) operon of Salmonella enterica serovar Typhimurium LT2 includes genes necessary for formation of polyhedral organelles involved in coenzyme

B(12)-dependent 1, 2-propanediol degradation. *Journal of Bacteriology* **1999**, *181* (19), 5967-75.

9. Roberts, E. W.; Cai, F.; Kerfeld, C. A.; Cannon, G. C.; Heinhorst, S., Isolation and Characterization of the Prochlorococcus Carboxysome Reveal the Presence of the Novel Shell Protein CsoS1D. *Journal of Bacteriology* **2012**, *194* (4), 787.

10. Petit, E.; LaTouf, W. G.; Coppi, M. V.; Warnick, T. A.; Currie, D.; Romashko, I.; Deshpande, S.; Haas, K.; Alvelo-Maurosa, J. G.; Wardman, C.; Schnell, D. J.; Leschine, S. B.; Blanchard, J. L., Involvement of a Bacterial Microcompartment in the Metabolism of Fucose and Rhamnose by Clostridium phytofermentans. *PloS One* **2013**, *8* (1), e54337.

11. Abdul-Rahman, F., The Distribution of Polyhedral Bacterial Microcompartments Suggests Frequent Horizontal Transfer and Operon Reassembly. *Journal of Phylogenetics & Evolutionary Biology* **2013**, *1* (4), 1-7

12. Cheng, S.; Liu, Y.; Crowley, C. S.; Yeates, T. O.; Bobik, T. A., Bacterial microcompartments: their properties and paradoxes. *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology* **2008**, *30* (11-12), 1084-1095.

13. Tanaka, S.; Kerfeld, C. A.; Sawaya, M. R.; Cai, F.; Heinhorst, S.; Cannon, G. C.; Yeates, T. O., Atomic-level models of the bacterial carboxysome shell. *Science* **2008**, *319* (5866), 1083-1086.

14. Klein, M. G.; Zwart, P.; Bagby, S. C.; Cai, F.; Chisholm, S. W.; Heinhorst, S.; Cannon, G. C.; Kerfeld, C. A., Identification and Structural Analysis of a Novel Carboxysome Shell Protein with Implications for Metabolite Transport. *Journal of Molecular Biology* **2009**, *392* (2), 319-333.

15. Sagermann, M.; Ohtaki, A.; Nikolakakis, K., Crystal structure of the EutL shell protein of the ethanolamine ammonia lyase microcompartment. *Proceedings of the National Academy of Sciences* **2009**, *106* (22), 8883.

16. Sutter, M.; Wilson, S. C.; Deutsch, S.; Kerfeld, C. A., Two new high-resolution crystal structures of carboxysome pentamer proteins reveal high structural conservation of CcmL orthologs among distantly related cyanobacterial species. *Photosynthesis Research* **2013**, *118* (1), 9-16.

17. Tanaka, S.; Sawaya, M. R.; Phillips, M.; Yeates, T. O., Insights from multiple structures of the shell proteins from the beta-carboxysome. *Protein Science: A Publication of the Protein Society* **2009**, *18* (1), 108-120.

18. Crowley, C. S.; Cascio, D.; Sawaya, M. R.; Kopstein, J. S.; Bobik, T. A.; Yeates, T. O., Structural Insight into the Mechanisms of Transport across the Salmonella enterica Pdu Microcompartment Shell. *Journal of Biological Chemistry* **2010**, *285* (48), 37838-37846.

19. Yeates, T. O.; Thompson, M. C.; Bobik, T. A., The protein shells of bacterial microcompartment organelles. *Current Opinion in Structural Biology* **2011**, *21* (2), 223-31.

20. Mahinthichaichan, P.; Morris, D. M.; Wang, Y.; Jensen, G. J.; Tajkhorshid, E., Selective Permeability of Carboxysome Shell Pores to Anionic Molecules. *The Journal of Physical Chemistry B* **2018**, *122* (39), 9110-9118.

21. Samborska, B.; Kimber, Matthew S., A Dodecameric CcmK2 Structure Suggests β-Carboxysomal Shell Facets Have a Double-Layered Organization. *Structure* **2012**, *20* (8), 1353-1362.

22. Sommer, M.; Sutter, M.; Gupta, S.; Kirst, H.; Turmo, A.; Lechno-Yossef, S.; Burton, R. L.; Saechao, C.; Sloan, N. B.; Cheng, X.; Chan, L.-J. G.; Petzold, C. J.; Fuentes-Cabrera, M.; Ralston, C. Y.; Kerfeld, C. A., Heterohexamers Formed by CcmK3 and CcmK4 Increase the Complexity of Beta Carboxysome Shells. *Plant Physiology* **2019**, *179* (1), 156.

23. Takenoya, M.; Nikolakakis, K.; Sagermann, M., Crystallographic Insights into the Pore Structures and Mechanisms of the EutL and EutM Shell Proteins of the Ethanolamine-Utilizing Microcompartment of Escherichia coli. *Journal of Bacteriology* **2010**, *192* (22), 6056.

24. Cai, F.; Sutter, M.; Cameron, J. C.; Stanley, D. N.; Kinney, J. N.; Kerfeld, C. A., The Structure of CcmP, a Tandem Bacterial Microcompartment Domain Protein from the β-Carboxysome, Forms a Subcompartment Within a Microcompartment. *Journal of Biological Chemistry* **2013**, *288* (22), 16055-16063.

25. Kirst, H.; Kerfeld, C. A., Bacterial microcompartments: catalysis-enhancing metabolic modules for next generation metabolic and biomedical engineering. *BMC Biology* **2019**, *17* (1), 79.

26. Shively, J. M.; Ball, F.; Brown, D. H.; Saunders, R. E., Functional Organelles in Prokaryotes: Polyhedral Inclusions (Carboxysomes) of Thiobacillus neapolitanus. *Science* **1973**, *182* (4112), 584-586.

27. Rae, B. D.; Long, B. M.; Badger, M. R.; Price, G. D., Functions, Compositions, and Evolution of the Two Types of Carboxysomes: Polyhedral Microcompartments That Facilitate CO2 Fixation in Cyanobacteria and Some Proteobacteria. *Microbiology and Molecular Biology Reviews* **2013**, *77* (3), 357.

28. Sampson, E. M.; Bobik, T. A., Microcompartments for B12-Dependent 1,2-Propanediol Degradation Provide Protection from DNA and Cellular Damage by a Reactive Metabolic Intermediate. *Journal of Bacteriology* **2008**, *190* (8), 2966.

29. Zarzycki, J.; Erbilgin, O.; Kerfeld, C. A., Bioinformatic Characterization of Glycyl Radical Enzyme-Associated Bacterial Microcompartments. *Applied and Environmental Microbiology* **2015**, *81* (24), 8315.

30. Ferlez, B.; Sutter, M.; Kerfeld, C. A., Glycyl Radical Enzyme-Associated Microcompartments: Redox-Replete Bacterial Organelles. *mBio* **2019**, *10* (1), e02327.

31. Havemann, G. D.; Sampson, E. M.; Bobik, T. A., PduA Is a Shell Protein of Polyhedral Organelles Involved in Coenzyme B12-Dependent Degradation of 1,2-Propanediol in Salmonella enterica Serovar Typhimurium LT2. *Journal of Bacteriology* **2002**, *184* (5), 1253.

32. Mallette, E.; Kimber, M. S., Structural and kinetic characterization of (S)-1-amino-2-propanol kinase from the aminoacetone utilization microcompartment of Mycobacterium smegmatis. *Journal of Biological Chemistry* **2018**, *293* (51), 19909-19918.

33. Craciun, S.; Balskus, E. P., Microbial conversion of choline to trimethylamine requires a glycyl radical enzyme. *Proceedings of the National Academy of Sciences* **2012**, *109* (52), 21307.

34. Herring, T. I.; Harris, T. N.; Chowdhury, C.; Mohanty, S. K.; Bobik, T. A., A Bacterial Microcompartment Is Used for Choline Fermentation by Escherichia coli 536. *Journal of Bacteriology* **2018**, *200* (10), e00764.

35. Erbilgin, O.; McDonald, K. L.; Kerfeld, C. A., Characterization of a Planctomycetal Organelle: a Novel Bacterial Microcompartment for the Aerobic Degradation of Plant Saccharides. *Applied and Environmental Microbiology* **2014**, *80* (7), 2193.

36. Horswill, A. R.; Escalante-Semerena, J. C., Salmonella typhimuriumLT2 Catabolizes Propionate via the 2-Methylcitric Acid Cycle. *Journal of Bacteriology* **1999**, *181* (18), 5615.

37. Parsons, J. B.; Lawrence, A. D.; McLean, K. J.; Munro, A. W.; Rigby, S. E. J.; Warren, M. J., Characterisation of PduS, the pdu Metabolosome Corrin Reductase, and Evidence of Substructural Organisation within the Bacterial Microcompartment. *PloS One* **2010**, *5* (11), e14009.

38. Cameron, Jeffrey C.; Wilson, Steven C.; Bernstein, Susan L.; Kerfeld, Cheryl A., Biogenesis of a Bacterial Organelle: The Carboxysome Assembly Pathway. *Cell* **2013**, *155* (5), 1131-1140.

39. Lee, M. J.; Brown, I. R.; Juodeikis, R.; Frank, S.; Warren, M. J., Employing bacterial microcompartment technology to engineer a shell-free enzyme-aggregate for enhanced 1,2-propanediol production in Escherichia coli. *Metabolic Engineering* **2016**, *36*, 48-56.

40. Wang, H.; Yan, X.; Aigner, H.; Bracher, A.; Nguyen, N. D.; Hee, W. Y.; Long, B. M.; Price, G. D.; Hartl, F. U.; Hayer-Hartl, M., Rubisco condensate formation by CcmM in β-carboxysome biogenesis. *Nature* **2019**, *566* (7742), 131-135.

41. Fan, C.; Cheng, S.; Liu, Y.; Escobar, C. M.; Crowley, C. S.; Jefferson, R. E.; Yeates, T. O.; Bobik, T. A., Short N-terminal sequences package proteins into bacterial microcompartments. *Proceedings of the National Academy of Sciences of the United States of America* **2010**, *107* (16), 7509-7514.

42. Lawrence, A. D.; Frank, S.; Newnham, S.; Lee, M. J.; Brown, I. R.; Xue, W.-F.; Rowe, M. L.; Mulvihill, D. P.; Prentice, M. B.; Howard, M. J.; Warren, M. J., Solution Structure of a Bacterial Microcompartment Targeting Peptide and Its Application in the Construction of an Ethanol Bioreactor. *ACS Synthetic Biology* **2014**, *3* (7), 454-465.

43. Plegaria, J. S.; Kerfeld, C. A., Engineering nanoreactors using bacterial microcompartment architectures. *Current Opinion in Biotechnology* **2018**, *51*, 1-7.

44. Planamente, S.; Frank, S., Bio-engineering of bacterial microcompartments: a mini review. *Biochemical Society Transactions* **2019**, *47* (3), 765-777.

45. Cai, F.; Sutter, M.; Bernstein, S. L.; Kinney, J. N.; Kerfeld, C. A., Engineering Bacterial Microcompartment Shells: Chimeric Shell Proteins and Chimeric Carboxysome Shells. *ACS Synthetic Biology* **2015**, *4* (4), 444-453.

46. Liu, Y.; He, X.; Lim, W.; Mueller, J.; Lawrie, J.; Kramer, L.; Guo, J.; Niu, W., Deciphering molecular details in the assembly of alpha-type carboxysome. *Scientific Reports* **2018**, *8* (1), 15062.

47. Nichols, R. J.; Cassidy-Amstutz, C.; Chaijarasphong, T.; Savage, D. F., Encapsulins: molecular biology of the shell. *Critical Reviews in Biochemistry and Molecular Biology* **2017**, *52* (5), 1-12.

48. Giessen, T. W.; Silver, P. A., Widespread distribution of encapsulin nanocompartments reveals functional diversity. *Nature Microbiology* **2017**, *2*, 17029.

49. Sutter, M.; Boehringer, D.; Gutmann, S.; Guenther, S.; Prangishvili, D.; Loessner, M. J.; Stetter, K. O.; Weber-Ban, E.; Ban, N., Structural basis of enzyme encapsulation into a bacterial nanocompartment. *Nature Structural and Molecular Biology* **2008**, *15* (9), 939-947.

50. Akita, F.; Chong, K. T.; Tanaka, H.; Yamashita, E.; Miyazaki, N.; Nakaishi, Y.; Suzuki, M.; Namba, K.; Ono, Y.; Tsukihara, T.; Nakagawa, A., The Crystal Structure of a Virus-like Particle from the Hyperthermophilic Archaeon Pyrococcus furiosus Provides Insight into the Evolution of Viruses. *Journal of Molecular Biology* **2007**, *368* (5), 1469-1483.

51. Tracey, J. C.; Coronado, M.; Giessen, T. W.; Lau, M. C. Y.; Silver, P. A.; Ward, B. B., The Discovery of Twenty-Eight New Encapsulin Sequences, Including Three in Anammox Bacteria. *Scientific Reports* **2019**, *9* (1), 20122.

52. Contreras, H.; Joens, M. S.; McMath, L. M.; Le, V. P.; Tullius, M. V.; Kimmey, J. M.; Bionghi, N.; Horwitz, M. A.; Fitzpatrick, J. A. J.; Goulding, C. W., Characterization of a Mycobacterium tuberculosis Nanocompartment and Its Potential Cargo Proteins. *Journal of Biological Chemistry* **2014**, *289* (26), 18279-18289.

53. McHugh, C. A.; Fontana, J.; Nemecek, D.; Cheng, N. Q.; Aksyuk, A. A.; Heymann, J. B.; Winkler, D. C.; Lam, A. S.; Wall, J. S.; Steven, A. C.; Hoiczyk, E., A virus capsid-like nanocompartment that stores iron and protects bacteria from oxidative stress. *The EMBO Journal* **2014**, *33* (17), 1896-1911.

54. Giessen, T. W.; Orlando, B. J.; Verdegaal, A. A.; Chambers, M. G.; Gardener, J.; Bell, D. C.; Birrane, G.; Liao, M.; Silver, P. A., Large protein organelles form a new iron sequestration system with high storage capacity. *eLIFE* **2019**, *8*, e46070.

55. Sigmund, F.; Pettinger, S.; Kube, M.; Schneider, F.; Schifferer, M.; Schneider, S.; Efremova, M. V.; Pujol-Martí, J.; Aichler, M.; Walch, A.; Misgeld, T.; Dietz, H.; Westmeyer, G. G., Iron-Sequestering Nanocompartments as Multiplexed Electron Microscopy Gene Reporters. *ACS Nano* **2019**, *13* (7), 8114-8123.

56. Wikoff, W. R.; Liljas, L.; Duda, R. L.; Tsuruta, H.; Hendrix, R. W.; Johnson, J. E., Topologically linked protein rings in the bacteriophage HK97 capsid. *Science* **2000**, *289* (5487), 2129-33.

57. Hendrix, R. W.; Johnson, J. E., Bacteriophage HK97 Capsid Assembly and Maturation. In Viral Molecular Machines, Rossmann, M. G.; Rao, V. B., Eds. Springer US: Boston, MA, 2012; pp 351-363.

58. Suhanovsky, M. M.; Teschke, C. M., Nature′s favorite building block: Deciphering folding and capsid assembly of proteins with the HK97-fold. *Virology* **2015**, *479-480*, 487-497.

59. Caspar, D. L.; Klug, A., Physical Principles in the Construction of Regular Viruses. *Cold Spring Harbor Symposia on Quantitative Biology* **1962**, *27*, 1-24.

60. Prasad, B. V. V.; Schmid, M. F., Principles of Virus Structural Organization. In Viral Molecular Machines, Rossmann, M. G.; Rao, V. B., Eds. Springer US: Boston, MA, 2012; pp 17-47.

61. Heinemann, J.; Maaty, W. S.; Gauss, G. H.; Akkaladevi, N.; Brumfield, S. K.; Rayaprolu, V.; Young, M. J.; Lawrence, C. M.; Bothner, B., Fossil record of an archaeal HK97-like provirus. *Virology* **2011**, *417* (2), 362-368.

62. Suttle, C. A., Marine viruses — major players in the global ecosystem. *Nature Reviews Microbiology* **2007**, *5* (10), 801-812.

63. Mokili, J. L.; Rohwer, F.; Dutilh, B. E., Metagenomics and future perspectives in virus discovery. *Current Opinion in Virology* **2012**, *2* (1), 63-77.

64. Paez-Espino, D.; Eloe-Fadrosh, E. A.; Pavlopoulos, G. A.; Thomas, A. D.; Huntemann, M.; Mikhailova, N.; Rubin, E.; Ivanova, N. N.; Kyrpides, N. C., Uncovering Earth's Virome. *Nature* **2016**, *536* (7617), 425-430.

65. Pasin, F.; Menzel, W.; Daròs, J.-A., Harnessed viruses in the age of metagenomics and synthetic biology: an update on infectious clone assembly and biotechnologies of plant viruses. *Plant Biotechnology Journal* **2019**, *17* (6), 1010-1026.

66. Aumiller, W. M.; Uchida, M.; Douglas, T., Protein cage assembly across multiple length scales. *Chemical Society Reviews* **2018**, *47* (10), 3433-3469.

67. Petrescu, D. S.; Blum, A. S., Viral-based nanomaterials for plasmonic and photonic materials and devices. *WIREs Nanomedicine and Nanobiotechnology* **2018**, *10* (4), e1508.

68. Eiben, S.; Koch, C.; Altintoprak, K.; Southan, A.; Tovar, G.; Laschat, S.; Weiss, I. M.; Wege, C., Plant virus-based materials for biomedical applications: Trends and prospects. *Advanced Drug Delivery Reviews* **2019**, *145*, 96-118.

69. Douglas, T.; Young, M., Viruses: making friends with old foes. *Science* **2006**, *312* (5775), 873-875.

70. Schoonen, L.; van Hest, J. C. M., Functionalization of protein-based nanocages for drug delivery applications. *Nanoscale* **2014**, *6* (13), 7124-7141.

71. Janin, J.; Bahadur, R. P.; Chakrabarti, P., Protein–protein interaction and quaternary structure. *Quarterly Reviews of Biophysics* **2008**, *41* (2), 133-180.

72. Charlton Hume, H. K.; Vidigal, J.; Carrondo, M. J. T.; Middelberg, A. P. J.; Roldão, A.; Lua, L. H. L., Synthetic biology for bioengineering virus-like particle vaccines. *Biotechnology and Bioengineering*. **2019**, *116* (4), 919-935.

73. Selivanovitch, E.; Douglas, T., Virus capsid assembly across different length scales inspire the development of virus-based biomaterials. *Current Opinion in Virology* **2019**, *36*, 38-46.

74. Zeltins, A., Construction and characterization of virus-like particles: a review. *Molecular Biotechnology* **2013**, *53* (1), 92-107.

75. Prangishvili, D., The Wonderful World of Archaeal Viruses. *Annual Review of Microbiology* **2013**, *67* (1), 565-585.

76. Prangishvili, D.; Bamford, D. H.; Forterre, P.; Iranzo, J.; Koonin, E. V.; Krupovic, M., The enigmatic archaeal virosphere. *Nature Reviews Microbiology* **2017**, *15* (12), 724-739.

77. Krupovic, M.; Cvirkaite-Krupovic, V.; Iranzo, J.; Prangishvili, D.; Koonin, E. V., Viruses of archaea: Structural, functional, environmental and evolutionary genomics. *Virus Research* **2018**, *244*, 181-193.

78. Kis, K.; Volk, R.; Bacher, A., Biosynthesis of riboflavin. Studies on the reaction mechanism of 6,7-dimethyl-8-ribityllumazine synthase. *Biochemistry* **1995**, *34* (9), 2883-92.

79. Bacher, A.; Eberhardt, S.; Fischer, M.; Kis, K.; Richter, G., Biosynthesis of Vitamin B2 (Riboflavin). *Annual Review of Nutrition* **2000**, *20* (1), 153-167.

80. Ladenstein, R.; Fischer, M.; Bacher, A., The lumazine synthase/riboflavin synthase complex: shapes and functions of a highly variable enzyme system. *The FEBS Journal* **2013**, *280* (11), 2537-2563.

81. Azuma, Y.; Edwardson, T. G. W.; Hilvert, D., Tailoring lumazine synthase assemblies for bionanotechnology. *Chemical Society Reviews* **2018**, *47* (10), 3543-3557.

82. Bacher, A.; Ludwig, H. C.; Schnepple, H.; Ben-Shaul, Y., Heavy riboflavin synthase from Bacillus subtilis: Quaternary structure and reaggregation. *Journal of Molecular Biology* **1986**, *187* (1), 75-86.

83. Ritsert, K.; Huber, R.; Turk, D.; Ladenstein, R.; Schmidt-Bäse, K.; Bacher, A., Studies on the Lumazine Synthase/Riboflavin Synthase Complex ofBacillus subtilis: Crystal Structure Analysis of Reconstituted, Icosahedral β-subunit Capsids with Bound Substrate Analogue Inhibitor at 2.4 Å Resolution. *Journal of Molecular Biology* **1995**, *253* (1), 151-167.

84. Mörtl, S.; Fischer, M.; Richter, G.; Tack, J.; Weinkauf, S.; Bacher, A., Biosynthesis of Riboflavin: LUMAZINE SYNTHASE OF ESCHERICHIA COLI. *Journal of Biological Chemistry* **1996**, *271* (52), 33201-33207.

85. Zhang, X.; Meining, W.; Fischer, M.; Bacher, A.; Ladenstein, R., X-ray structure analysis and crystallographic refinement of lumazine synthase from the hyperthermophile Aquifex aeolicus at 1.6 Å resolution: determinants of thermostability revealed from structural comparisons. *Journal of Molecular Biology* **2001**, *306* (5), 1099-1114.

86. Morgunova, E.; Illarionov, B.; Saller, S.; Popov, A.; Sambaiah, T.; Bacher, A.; Cushman, M.; Fischer, M.; Ladenstein, R., Structural study and thermodynamic characterization of inhibitor binding to lumazine synthase from Bacillus anthracis. *Acta Crystallography D Biological Crystallography* **2010**, *66* (Pt 9), 1001-1011.

87. Kumar, P.; Singh, M.; Karthikeyan, S., Crystal structure analysis of icosahedral lumazine synthase from Salmonella typhimurium, an antibacterial drug target. *Acta Crystallography D Biological Crystallography* **2011**, *67* (Pt 2), 131-9.

88. Bacher, A.; Baur, R.; Eggers, U.; Harders, H. D.; Otto, M. K.; Schnepple, H., Riboflavin synthases of Bacillus subtilis. Purification and properties. *Journal of Biological Chemistry* **1980**, *255* (2), 632-637.

89. Azuma, Y.; Zschoche, R.; Hilvert, D., The C-terminal peptide of Aquifex aeolicus riboflavin synthase directs encapsulation of native and foreign guests by a cage-forming lumazine synthase. *Journal of Biological Chemistry* **2017**, *292* (25), 10321-10327.

90. Han, X.; Woycechowsky, K. J., Encapsulation and Controlled Release of Protein Guests by the Bacillus subtilis Lumazine Synthase Capsid. *Biochemistry* **2017**, *56* (47), 6211-6220.

91. Zhang, X.; Konarev, P. V.; Petoukhov, M. V.; Svergun, D. I.; Xing, L.; Cheng, R. H.; Haase, I.; Fischer, M.; Bacher, A.; Ladenstein, R.; Meining, W., Multiple Assembly States of Lumazine Synthase: A Model Relating Catalytic Function and Molecular Assembly. *Journal of Molecular Biology* **2006**, *362* (4), 753-770.

92. He, D.; Marles-Wright, J., Ferritin family proteins and their use in bionanotechnology. *New Biotechnology* **2015**, *32* (6), 651-657.

93. Chasteen, N. D.; Harrison, P. M., Mineralization in Ferritin: An Efficient Means of Iron Storage. *Journal of Structural Biology* **1999**, *126* (3), 182-194.

94. Arosio, P.; Levi, S., Cytosolic and mitochondrial ferritins in the regulation of cellular iron homeostasis and oxidative damage. *Biochimica et Biophysica Acta (BBA) - General Subjects* **2010**, *1800* (8), 783-792.

95. Andrews, S. C., The Ferritin-like superfamily: Evolution of the biological iron storeman from a rubrerythrin-like ancestor. Biochimica et biophysica acta 2010, 1800 (8), 691-705.

96. Carrondo, M. A., Ferritins, iron uptake and storage from the bacterioferritin viewpoint. *The EMBO Journal* **2003**, *22* (9), 1959-1968.

97. Le Brun, N. E.; Andrews, S. C.; Guest, J. R.; Harrison, P. M.; Moore, G. R.; Thomson, A. J., Identification of the ferroxidase centre of Escherichia coli bacterioferritin. *The Biochemical Journal* **1995**, *312* (Pt 2) (Pt 2), 385-392.

98. Treffry, A.; Bauminger, E. R.; Hechel, D.; Hodson, N. W.; Nowik, I.; Yewdall, S. J.; Harrison, P. M., Defining the roles of the threefold channels in iron uptake, iron oxidation and iron-core formation in ferritin: a study aided by site-directed mutagenesis. *Biochemical Journal* **1993**, *296* (3), 721-728.

99. Arosio, P.; Adelman, T. G.; Drysdale, J. W., On ferritin heterogeneity. Further evidence for heteropolymers. *The Journal of Biological Chemistry* **1978**, *253* (12), 4451-4458.

100. Frolow, F.; Kalb, A. J.; Yariv, J., Structure of a unique twofold symmetric haem-binding site. *Nature Structural Biology* **1994**, *1* (7), 453-460.

101. Yao, H.; Wang, Y.; Lovell, S.; Kumar, R.; Ruvinsky, A. M.; Battaile, K. P.; Vakser, I. A.; Rivera, M., The Structure of the BfrB–Bfd Complex Reveals Protein–Protein Interactions Enabling Iron Release from Bacterioferritin. *Journal of the American Chemical Society* **2012**, *134* (32), 13470-13481.

102. Macedo, S.; Romão, C. V.; Mitchell, E.; Matias, P. M.; Liu, M. Y.; Xavier, A. V.; LeGall, J.; Teixeira, M.; Lindley, P.; Carrondo, M. A., The nature of the di-iron site in the bacterioferritin from Desulfovibrio desulfuricans. *Nature Structural and Molecular Biology* **2003**, *10* (4), 285-290.

103. Cobessi, D.; Huang, L. S.; Ban, M.; Pon, N. G.; Daldal, F.; Berry, E. A., The 2.6 Å resolution structure of Rhodobacter capsulatus bacterioferritin with metal-free dinuclear site and heme iron in a crystallographic `special position'. *Acta Crystallographica Section D* **2002**, *58* (1), 29-38.

104. Weeratunga, S. K.; Gee, C. E.; Lovell, S.; Zeng, Y.; Woodin, C. L.; Rivera, M., Binding of Pseudomonas aeruginosa Apobacterioferritin-Associated Ferredoxin to Bacterioferritin B Promotes Heme Mediation of Electron Delivery and Mobilization of Core Mineral Iron. *Biochemistry* **2009**, *48* (31), 7420-7431.

105. Almirón, M.; Link, A. J.; Furlong, D.; Kolter, R., A novel DNA-binding protein with regulatory and protective roles in starved Escherichia coli. *Genes & Development* **1992**, *6* (12b), 2646-2654.

106. Grant, R. A.; Filman, D. J.; Finkel, S. E.; Kolter, R.; Hogle, J. M., The crystal structure of Dps, a ferritin homolog that binds and protects DNA. *Nature Structural Biology* **1998**, *5* (4), 294-303.

107. Haikarainen, T.; Papageorgiou, A. C., Dps-like proteins: structural and functional insights into a versatile protein family. *Cellular and Molecular Life Sciences* **2010**, *67* (3), 341-351.

108. Chiancone, E.; Ceci, P., The multifaceted capacity of Dps proteins to combat bacterial stress conditions: Detoxification of iron and hydrogen peroxide and DNA binding. *Biochimica et Biophysica Acta (BBA) - General Subjects* **2010**, *1800* (8), 798-805.

109. Ilari, A.; Stefanini, S.; Chiancone, E.; Tsernoglou, D., The dodecameric ferritin from Listeria innocua contains a novel intersubunit iron-binding site. *Nature Structural Biology* **2000**, *7* (1), 38-43.

110. Stillman, T. J.; Upadhyay, M.; Norte, V. A.; Sedelnikova, S. E.; Carradus, M.; Tzokov, S.; Bullough, P. A.; Shearman, C. A.; Gasson, M. J.; Williams, C. H.; Artymiuk, P. J.; Green, J., The crystal structures of Lactococcus lactis MG1363 Dps proteins reveal the presence of an N-terminal helix that is required for DNA binding. *Molecular Microbiology* **2005**, *57* (4), 1101-1112.

111. Fletcher, J. M.; Harniman, R. L.; Barnes, F. R. H.; Boyle, A. L.; Collins, A.; Mantell, J.; Sharp, T. H.; Antognozzi, M.; Booth, P. J.; Linden, N.; Miles, M. J.; Sessions, R. B.; Verkade, P.; Woolfson, D. N., Self-Assembling Cages from Coiled-Coil Peptide Modules. *Science* **2013**, *340* (6132), 595.

112. Kobayashi, N.; Yanase, K.; Sato, T.; Unzai, S.; Hecht, M. H.; Arai, R., Self-Assembling Nano-Architectures Created from a Protein Nano-Building Block Using an Intermolecularly Folded Dimeric de Novo Protein. *Journal of the American Chemical Society* **2015**, *137* (35), 11285-11293.

113. Padilla, J. E.; Colovos, C.; Yeates, T. O., Nanohedra: Using symmetry to design self assembling protein cages, layers, crystals, and filaments. *Proceedings of the National Academy of Sciences* **2001**, *98* (5), 2217.

114. Lai, Y.-T.; Cascio, D.; Yeates, T. O., Structure of a 16-nm Cage Designed by Using Protein Oligomers. *Science* **2012**, *336* (6085), 1129.

115. Lai, Y.-T.; Reading, E.; Hura, G. L.; Tsai, K.-L.; Laganowsky, A.; Asturias, F. J.; Tainer, J. A.; Robinson, C. V.; Yeates, T. O., Structure of a designed protein cage that self-assembles into a highly porous cube. *Nature Chemistry* **2014**, *6* (12), 1065-1071.

116. Huang, P.-S.; Boyken, S. E.; Baker, D., The coming of age of de novo protein design. *Nature* **2016**, *537* (7620), 320-327.

117. Bale, J. B.; Gonen, S.; Liu, Y.; Sheffler, W.; Ellis, D.; Thomas, C.; Cascio, D.; Yeates, T. O.; Gonen, T.; King, N. P.; Baker, D., Accurate design of megadalton-scale two-component icosahedral protein complexes. *Science* **2016**, *353* (6297), 389.

118. Badieyan, S.; Sciore, A.; Eschweiler, J. D.; Koldewey, P.; Cristie-David, A. S.; Ruotolo, B. T.; Bardwell, J. C. A.; Su, M.; Marsh, E. N. G., Symmetry-Directed Self-Assembly of

a Tetrahedral Protein Cage Mediated by de Novo-Designed Coiled Coils. *ChemBioChem* **2017**, *18* (19), 1888-1892.

119. Cannon, K. A.; Nguyen, V. N.; Morgan, C.; Yeates, T. O., Design and Characterization of an Icosahedral Protein Cage Formed by a Double-Fusion Protein Containing Three Distinct Symmetry Elements. *ACS Synthetic Biology* **2020**, *9* (3), 517-524.

120. King, N. P.; Lai, Y.-T., Practical approaches to designing novel protein assemblies. *Current Opinion in Structural Biology* **2013**, *23* (4), 632-638.

121. Yeates, T. O.; Liu, Y.; Laniado, J., The design of symmetric protein nanomaterials comes of age in theory and practice. *Current Opinion in Structural Biology* **2016**, *39*, 134-143.

122. King, N. P.; Sheffler, W.; Sawaya, M. R.; Vollmar, B. S.; Sumida, J. P.; André, I.; Gonen, T.; Yeates, T. O.; Baker, D., Computational Design of Self-Assembling Protein Nanomaterials with Atomic Level Accuracy. *Science* **2012**, *336* (6085), 1171.

123. Lai, Y.-T.; Tsai, K.-L.; Sawaya, M. R.; Asturias, F. J.; Yeates, T. O., Structure and Flexibility of Nanoscale Protein Cages Designed by Symmetric Self-Assembly. *Journal of the American Chemical Society* **2013**, *135* (20), 7738-7743.

124. Sciore, A.; Su, M.; Koldewey, P.; Eschweiler, J. D.; Diffley, K. A.; Linhares, B. M.; Ruotolo, B. T.; Bardwell, J. C. A.; Skiniotis, G.; Marsh, E. N. G., Flexible, symmetry-directed approach to assembling protein cages. *Proceedings of the National Academy of Sciences* **2016**, *113* (31), 8681.

125. Cristie-David, A. S.; Chen, J.; Nowak, D. B.; Bondy, A. L.; Sun, K.; Park, S. I.; Banaszak Holl, M. M.; Su, M.; Marsh, E. N. G., Coiled-Coil-Mediated Assembly of an Icosahedral Protein Cage with Extremely High Thermal and Chemical Stability. *Journal of the American Chemical Society* **2019**, *141* (23), 9207-9216.

126. Leaver-Fay, A.; Tyka, M.; Lewis, S. M.; Lange, O. F.; Thompson, J.; Jacak, R.; Kaufman, K. W.; Renfrew, P. D.; Smith, C. A.; Sheffler, W.; Davis, I. W.; Cooper, S.; Treuille, A.; Mandell, D. J.; Richter, F.; Ban, Y.-E. A.; Fleishman, S. J.; Corn, J. E.; Kim, D. E.; Lyskov, S.; Berrondo, M.; Mentzer, S.; Popović, Z.; Havranek, J. J.; Karanicolas, J.; Das, R.; Meiler, J.; Kortemme, T.; Gray, J. J.; Kuhlman, B.; Baker, D.; Bradley, P., Chapter nineteen - Rosetta3: An Object-Oriented Software Suite for the Simulation and Design of Macromolecules. In Methods in Enzymology, Johnson, M. L.; Brand, L., Eds. Academic Press: 2011; Vol. 487, pp 545-574.

127. DiMaio, F.; Leaver-Fay, A.; Bradley, P.; Baker, D.; André, I., Modeling Symmetric Macromolecular Structures in Rosetta3. *PloS One* **2011**, *6* (6), e20450.

128. Hsia, Y.; Bale, J. B.; Gonen, S.; Shi, D.; Sheffler, W.; Fong, K. K.; Nattermann, U.; Xu, C.; Huang, P.-S.; Ravichandran, R.; Yi, S.; Davis, T. N.; Gonen, T.; King, N. P.; Baker, D., Design of a hyperstable 60-subunit protein icosahedron. *Nature* **2016**, *535* (7610), 136-139.

129. King, N. P.; Bale, J. B.; Sheffler, W.; McNamara, D. E.; Gonen, S.; Gonen, T.; Yeates, T. O.; Baker, D., Accurate design of co-assembling multi-component protein nanomaterials. *Nature* **2014**, *510* (7503), 103-108.

130. Yeates, T. O., Geometric Principles for Designing Highly Symmetric Self-Assembling Protein Nanomaterials. *Annual Review of Biophysics* **2017**, *46* (1), 23-42.

131. Bachmann, M. F.; Jennings, G. T., Vaccine delivery: a matter of size, geometry, kinetics and molecular patterns. *Nature Reviews Immunology* **2010**, *10* (11), 787-796.

132. Plotkin, S., History of vaccination. *Proceedings of the National Academy of Sciences* **2014**, *111* (34), 12283.

133. Rappuoli, R.; Pizza, M.; Del Giudice, G.; De Gregorio, E., Vaccines, new opportunities for a new society. *Proceedings of the National Academy of Sciences* **2014**, *111* (34), 12288.

134. Yan, D.; Wei, Y.-Q.; Guo, H.-C.; Sun, S.-Q., The application of virus-like particles as vaccines and biological vehicles. *Applied Microbiology and Biotechnology* **2015**, *99* (24), 10415-10432.

135. Hilleman, M. R., Vaccines in historic evolution and perspective: a narrative of vaccine discoveries. *Vaccine* **2000**, *18* (15), 1436-1447.

136. López-Vidal, J.; Gómez-Sebastián, S.; Bárcena, J.; Nuñez, M. d. C.; Martínez-Alonso, D.; Dudognon, B.; Guijarro, E.; Escribano, J. M., Improved Production Efficiency of Virus-Like Particles by the Baculovirus Expression Vector System. *PloS One* **2015**, *10* (10), e0140039.

137. Zepp, F., Principles of vaccine design—Lessons from nature. *Vaccine* **2010**, *28*, C14-C24.

138. López-Sagaseta, J.; Malito, E.; Rappuoli, R.; Bottomley, M. J., Self-assembling protein nanoparticles in the design of vaccines. *Computational and Structural Biotechnology Journal* **2016**, *14*, 58-68.

139. Keller, S. A.; Bauer, M.; Manolova, V.; Muntwiler, S.; Saudan, P.; Bachmann, M. F., Cutting edge: limited specialization of dendritic cell subsets for MHC class II-associated presentation of viral particles. *Journal of Immunology* **2010**, *184* (1), 26-9.

140. Bundy, B. C.; Franciszkowicz, M. J.; Swartz, J. R., Escherichia coli-based cell-free synthesis of virus-like particles. *Biotechnology and Bioengineering* **2008**, *100* (1), 28-37.

141. Wahome, N.; Cooper, A.; Thapa, P.; Choudhari, S.; Gao, F. P.; Volkin, D. B.; Middaugh, C. R., Production of Well-Characterized Virus-like Particles in an Escherichia coli-Based Expression Platform for Preclinical Vaccine Assessments. In Vaccine Design: Methods

and Protocols, Volume 2: Vaccines for Veterinary Diseases, Thomas, S., Ed. Springer New York: New York, NY, 2016; pp 437-457.

142. Clarke, B. E.; Newton, S. E.; Carroll, A. R.; Francis, M. J.; Appleyard, G.; Syred, A. D.; Highfield, P. E.; Rowlands, D. J.; Brown, F., Improved immunogenicity of a peptide epitope after fusion to hepatitis B core protein. *Nature* **1987**, *330* (6146), 381-384.

143. Beck, P. L.; Wong, J. F.; Li, Y.; Swaminathan, S.; Xavier, R. J.; Devaney, K. L.; Podolsky, D. K., Chemotherapy- and radiotherapy-induced intestinal damage is regulated by intestinal trefoil factor. *Gastroenterology* **2004**, *126* (3), 796-808.

144. Mateu, M. G., Assembly, Engineering and Applications of Virus-Based Protein Nanoparticles. *Advances in Experimental Medicine and Biology* **2016**, *940*, 83-120.

145. Singh, Y.; Palombo, M.; Sinko, P. J., Recent trends in targeted anticancer prodrug and conjugate design. *Current Medicinal Chemistry* **2008**, *15* (18), 1802-1826.

146. Belletti, D.; Pederzoli, F.; Forni, F.; Vandelli, M. A.; Tosi, G.; Ruozi, B., Protein cage nanostructure as drug delivery system: magnifying glass on apoferritin. *Expert Opinion on Drug Delivery* **2017**, *14* (7), 825-840.

147. Jeevanandam, J.; Pal, K.; Danquah, M. K., Virus-like nanoparticles as a novel delivery tool in gene therapy. *Biochimie* **2019**, *157*, 38-47.

148. Karimi, M.; Mirshekari, H.; Moosavi Basri, S. M.; Bahrami, S.; Moghoofei, M.; Hamblin, M. R., Bacteriophages and phage-inspired nanocarriers for targeted delivery of therapeutic cargos. *Advanced Drug Delivery Reviews* **2016**, *106*, 45-62.

149. Lee, E. J.; Lee, N. K.; Kim, I.-S., Bioengineered protein-based nanocage for drug delivery. *Advanced Drug Delivery Reviews* **2016**, *106*, 157-171.

150. Molino, N. M.; Wang, S.-W., Caged protein nanoparticles for drug delivery. *Current Opinion in Biotechnology* **2014**, *28*, 75-82.

151. De Jong, W. H.; Borm, P. J., Drug delivery and nanoparticles: applications and hazards. *International Journal Nanomedicine* **2008**, *3* (2), 133-49.

152. Ashley, C. E.; Carnes, E. C.; Phillips, G. K.; Durfee, P. N.; Buley, M. D.; Lino, C. A.; Padilla, D. P.; Phillips, B.; Carter, M. B.; Willman, C. L.; Brinker, C. J.; Caldeira, J. d. C.; Chackerian, B.; Wharton, W.; Peabody, D. S., Cell-Specific Delivery of Diverse Cargos by Bacteriophage MS2 Virus-like Particles. *ACS Nano* **2011**, *5* (7), 5729-5745.

153. Fang, P.-Y.; Gómez Ramos, Lizzette M.; Holguin, S. Y.; Hsiao, C.; Bowman, J. C.; Yang, H.-W.; Williams, L. D., Functional RNAs: combined assembly and packaging in VLPs. *Nucleic Acids Research* **2016**, *45* (6), 3519-3527.

154. Nuñez-Rivera, A.; Fournier, P. G. J.; Arellano, D. L.; Rodriguez-Hernandez, A. G.; Vazquez-Duhalt, R.; Cadena-Nava, R. D., Brome mosaic virus-like particles as siRNA nanocarriers for biomedical purposes. *Beilstein Journal of Nanotechnology* **2020**, *11*, 372-382.

155. Wu, M.; Sherwin, T.; Brown, W. L.; Stockley, P. G., Delivery of antisense oligonucleotides to leukemia cells by RNA bacteriophage capsids. *Nanomedicine: Nanotechnology, Biology and Medicine* **2005**, *1* (1), 67-76.

156. Shukla, S.; Steinmetz, N. F., Virus-based nanomaterials as positron emission tomography and magnetic resonance contrast agents: from technology development to translational medicine. *WIREs Nanomedicine and Nanobiotechnology* **2015**, *7* (5), 708-721.

157. Truffi, M.; Fiandra, L.; Sorrentino, L.; Monieri, M.; Corsi, F.; Mazzucchelli, S., Ferritin nanocages: A biological platform for drug delivery, imaging and theranostics in cancer. *Pharmacological Research* **2016**, *107*, 57-65.

158. Guerrero, Y.; Singh, S. P.; Mai, T.; Murali, R. K.; Tanikella, L.; Zahedi, A.; Kundra, V.; Anvari, B., Optical Characteristics and Tumor Imaging Capabilities of Near Infrared Dyes in Free and Nano-Encapsulated Formulations Comprised of Viral Capsids. *ACS Applied Materials & Interfaces* **2017**, *9* (23), 19601-19611.

159. Schwarz, B.; Uchida, M.; Douglas, T., Biomedical and Catalytic Opportunities of Virus-Like Particles in Nanotechnology. *Advances in Virus Research* **2017**, *97*, 1-60.

160. Semmler-Behnke, M.; Kreyling, W. G.; Lipka, J.; Fertsch, S.; Wenk, A.; Takenaka, S.; Schmid, G.; Brandau, W., Biodistribution of 1.4- and 18-nm gold particles in rats. *Small* **2008**, *4* (12), 2108-2111.

161. Sánchez, P.; Valero, E.; Gálvez, N.; Domínguez-Vera, J. M.; Marinone, M.; Poletti, G.; Corti, M.; Lascialfari, A., MRI relaxation properties of water-soluble apoferritin-encapsulated gadolinium oxide-hydroxide nanoparticles. *Dalton Transactions* **2009**, (5), 800-804.

162. Song, Y.; Kang, Y. J.; Jung, H.; Kim, H.; Kang, S.; Cho, H., Lumazine Synthase Protein Nanoparticle-Gd(III)-DOTA Conjugate as a T1 contrast agent for high-field MRI. *Scientific Reports* **2015**, *5* (1), 15656.

163. Diaz, D.; Care, A.; Sunna, A., Bioengineering Strategies for Protein-Based Nanoparticles. *Genes* **2018**, *9* (7), 370.

164. Ren, H.; Zhu, S.; Zheng, G., Nanoreactor Design Based on Self-Assembling Protein Nanocages. *International Journal of Molecular Sciences* **2019**, *20* (3).

165. Cassidy-Amstutz, C.; Oltrogge, L.; Going, C. C.; Lee, A.; Teng, P.; Quintanilla, D.; East-Seletsky, A.; Williams, E. R.; Savage, D. F., Identification of a Minimal Peptide Tag for in Vivo and in Vitro Loading of Encapsulin. *Biochemistry* **2016**, *55* (24), 3461-3468.

166. Johnson, H. R.; Hooker, J. M.; Francis, M. B.; Clark, D. S., Solubilization and stabilization of bacteriophage MS2 in organic solvents. *Biotechnology and Bioengineering* **2007**, *97* (2), 224-34.

167. Jordan, P. C.; Patterson, D. P.; Saboda, K. N.; Edwards, E. J.; Miettinen, H. M.; Basu, G.; Thielges, M. C.; Douglas, T., Self-assembling biomolecular catalysts for hydrogen production. *Nature Chemistry* **2015**, *8*, 179.

168. Williams, E. M.; Jung, S. M.; Coffman, J. L.; Lutz, S., Pore Engineering for Enhanced Mass Transport in Encapsulin Nanocompartments. *ACS Synthetic Biology* **2018**, *7* (11), 2514-2517.

169. Patterson, D. P.; LaFrance, B.; Douglas, T., Rescuing recombinant proteins by sequestration into the P22 VLP. *Chemical Communications* **2013**, *49* (88), 10412-10414.

170. Azuma, Y.; Zschoche, R.; Tinzl, M.; Hilvert, D., Quantitative Packaging of Active Enzymes into a Protein Cage. *Angewandte Chemie International Edition* **2016**, *55* (4), 1531-1534.

171. Rurup, W. F.; Verbij, F.; Koay, M. S. T.; Blum, C.; Subramaniam, V.; Cornelissen, J. J. L. M., Predicting the Loading of Virus-Like Particles with Fluorescent Proteins. *Biomacromolecules* **2014**, *15* (2), 558-563.

172. Sharma, J.; Uchida, M.; Miettinen, H. M.; Douglas, T., Modular interior loading and exterior decoration of a virus-like particle. *Nanoscale* **2017**, *9* (29), 10420-10430.

173. Comellas-Aragones, M.; Engelkamp, H.; Claessen, V. I.; Sommerdijk, N. A.; Rowan, A. E.; Christianen, P. C.; Maan, J. C.; Verduin, B. J.; Cornelissen, J. J.; Nolte, R. J., A virus-based single-enzyme nanoreactor. *Nature Nanotechnology* **2007**, *2* (10), 635-9.

174. Sánchez-Sánchez, L.; Cadena-Nava, R. D.; Palomares, L. A.; Ruiz-Garcia, J.; Koay, M. S. T.; Cornelissen, J. J. M. T.; Vazquez-Duhalt, R., Chemotherapy pro-drug activation by biocatalytic virus-like nanoparticles containing cytochrome P450. *Enzyme and Microbial Technology* **2014**, *60*, 24-31.

175. Wilkerson, J. W.; Yang, S.-O.; Funk, P. J.; Stanley, S. K.; Bundy, B. C., Nanoreactors: Strategies to encapsulate enzyme biocatalysts in virus-like particles. *New Biotechnology* **2018**, *44*, 59-63.

176. Patterson, D. P.; Prevelige, P. E.; Douglas, T., Nanoreactors by Programmed Enzyme Encapsulation Inside the Capsid of the Bacteriophage P22. *ACS Nano* **2012**, *6* (6), 5000-5009.

177. Fiedler, J. D.; Fishman, M. R.; Brown, S. D.; Lau, J.; Finn, M. G., Multifunctional Enzyme Packaging and Catalysis in the Qβ Protein Nanoparticle. *Biomacromolecules* **2018**, *19* (10), 3945-3957.

178. Minten, I. J.; Claessen, V. I.; Blank, K.; Rowan, A. E.; Nolte, R. J. M.; Cornelissen, J. J. L. M., Catalytic capsids: the art of confinement. Chemical Science 2011, 2 (2), 358-362.

179. Giessen, T. W.; Silver, P., A catalytic nanoreactor based on in vivo encapsulation of multiple enzymes in an engineered protein nanocompartment. *ChemBioChem* **2016**, *17* (20), 1931-1935.

180. Schoonen, L.; Nolte, R. J. M.; van Hest, J. C. M., Highly efficient enzyme encapsulation in a protein nanocage: towards enzyme catalysis in a cellular nanocompartment mimic. *Nanoscale* **2016**, *8* (30), 14467-14472.

181. Zakeri, B.; Fierer, J. O.; Celik, E.; Chittock, E. C.; Schwarz-Linek, U.; Moy, V. T.; Howarth, M., Peptide tag forming a rapid covalent bond to a protein, through engineering a bacterial adhesin. *Proceedings of the National Acadademy of Science USA* **2012**, *109* (12), E690-7.

182. Patterson, D. P.; Schwarz, B.; Waters, R. S.; Gedeon, T.; Douglas, T., Encapsulation of an Enzyme Cascade within the Bacteriophage P22 Virus-Like Particle. *ACS Chemical Biology* **2014**, *9* (2), 359-365.

183. O'Neil, A.; Reichhardt, C.; Johnson, B.; Prevelige, P. E.; Douglas, T., Genetically programmed in vivo packaging of protein cargo and its controlled release from bacteriophage P22. *Angewandte Chemie International Edition* **2011**, *50* (32), 7425-8.

184. Dueber, J. E.; Wu, G. C.; Malmirchegini, G. R.; Moon, T. S.; Petzold, C. J.; Ullal, A. V.; Prather, K. L. J.; Keasling, J. D., Synthetic protein scaffolds provide modular control over metabolic flux. *Nature Biotechnology* **2009**, *27* (8), 753-U107.

185. Moon, T. S.; Dueber, J. E.; Shiue, E.; Prather, K. L., Use of modular, synthetic scaffolds for improved production of glucaric acid in engineered E. coli. *Metabolic Engineering* **2010**, *12* (3), 298-305.

186. Jullesson, D.; David, F.; Pfleger, B.; Nielsen, J., Impact of synthetic biology and metabolic engineering on industrial production of fine chemicals. *Biotechnology Advances* **2015**, *33* (7), 1395-1402.

187. Liang, M.; Yan, X., Nanozymes: From New Concepts, Mechanisms, and Standards to Applications. *Accounts of Chemical Research* **2019**, *52* (8), 2190-2200.

188. Gao, L.; Zhuang, J.; Nie, L.; Zhang, J.; Zhang, Y.; Gu, N.; Wang, T.; Feng, J.; Yang, D.; Perrett, S.; Yan, X., Intrinsic peroxidase-like activity of ferromagnetic nanoparticles. *Nature Nanotechnology* **2007**, *2* (9), 577-83.

189. Wei, H.; Wang, E., Nanomaterials with enzyme-like characteristics (nanozymes): next-generation artificial enzymes. *Chemical Society Reviews* **2013**, *42* (14), 6060-6093.

190. Zhang, W.; Zhang, Y.; Chen, Y.; Li, S.; Gu, N.; Hu, S.; Sun, Y.; Chen, X.; Li, Q., Prussian Blue Modified Ferritin as Peroxidase Mimetics and Its Applications in Biological Detection. *Journal of Nanoscience and Nanotechnology* **2013**, *13* (1), 60-67.

191. Jiang, B.; Fang, L.; Wu, K.; Yan, X.; Fan, K., Ferritins as natural and artificial nanozymes for theranostics. *Theranostics* **2020**, *10* (2), 687-706.

192. Maity, B.; Fujita, K.; Ueno, T., Use of the confined spaces of apo-ferritin and virus capsids as nanoreactors for catalytic reactions. *Current Opinion in Chemical Biology* **2015**, *25*, 88-97.

193. Ueno, T.; Suzuki, M.; Goto, T.; Matsumoto, T.; Nagayama, K.; Watanabe, Y., Size-selective olefin hydrogenation by a Pd nanocluster provided in an apo-ferritin cage. *Angewandte Chemie International Edition* **2004**, *43* (19), 2527-30.

194. Liu, X.; Wei, W.; Yuan, Q.; Zhang, X.; Li, N.; Du, Y.; Ma, G.; Yan, C.; Ma, D., Apoferritin–CeO2 nano-truffle that has excellent artificial redox enzyme activity. *Chemical Communications* **2012**, *48* (26), 3155-3157.

195. Bedwell, G. J.; Zhou, Z.; Uchida, M.; Douglas, T.; Gupta, A.; Prevelige, P. E., Selective Biotemplated Synthesis of TiO2 Inside a Protein Cage. *Biomacromolecules* **2015**, *16* (1), 214-218.

196. Douglas, T.; Young, M., Host–guest encapsulation of materials by assembled virus protein cages. *Nature* **1998**, *393* (6681), 152-155.

197. Giessen, T. W.; Silver, P. A., Converting a Natural Protein Compartment into a Nanofactory for the Size-Constrained Synthesis of Antimicrobial Silver Nanoparticles. *ACS Synthetic Biology* **2016**, *5* (12), 1497-1504.

198. Abe, S.; Hirata, K.; Ueno, T.; Morino, K.; Shimizu, N.; Yamamoto, M.; Takata, M.; Yashima, E.; Watanabe, Y., Polymerization of Phenylacetylene by Rhodium Complexes within a Discrete Space of apo-Ferritin. *Journal of the American Chemical Society* **2009**, *131* (20), 6958-6960.

199. Maity, B.; Fukumori, K.; Abe, S.; Ueno, T., Immobilization of two organometallic complexes into a single cage to construct protein-based microcompartments. *Chemical Communications* **2016**, *52* (31), 5463-5466.

200. Wang, Z.; Takezawa, Y.; Aoyagi, H.; Abe, S.; Hikage, T.; Watanabe, Y.; Kitagawa, S.; Ueno, T., Definite coordination arrangement of organometallic palladium complexes accumulated on the designed interior surface of apo-ferritin. *Chemical Communications* **2011**, *47* (1), 170-172.

201. Patterson, D.; Edwards, E.; Douglas, T., Hybrid Nanoreactors: Coupling Enzymes and Small-Molecule Catalysts within Virus-Like Particles. *Israel Journal of Chemistry* **2015**, *55* (1), 96-101.

202. Bari, N. K.; Kumar, G.; Bhatt, A.; Hazra, J. P.; Garg, A.; Ali, M. E.; Sinha, S., Nanoparticle Fabrication on Bacterial Microcompartment Surface for the Development of Hybrid Enzyme-Inorganic Catalyst. *ACS Catalysis* **2018**, *8* (9), 7742-7748.

203. Gelman, F.; Blum, J.; Avnir, D., One-Pot Sequences of Reactions with Sol-Gel Entrapped Opposing Reagents: An Enzyme and Metal-Complex Catalysts. *Journal of the American Chemical Society* **2002**, *124* (48), 14460-14463.

204. Hoyos, P.; Buthe, A.; Ansorge-Schumacher, M. B.; Sinisterra, J. V.; Alcántara, A. R., Highly efficient one pot dynamic kinetic resolution of benzoins with entrapped Pseudomonas stutzeri lipase. *Journal of Molecular Catalysis B: Enzymatic* **2008**, *52-53*, 133-139.

205. Courchesne, N. M.; Klug, M. T.; Chen, P. Y.; Kooi, S. E.; Yun, D. S.; Hong, N.; Fang, N. X.; Belcher, A. M.; Hammond, P. T., Assembly of a bacteriophage-based template for the organization of materials into nanoporous networks. *Advanced Materials* **2014**, *26* (21), 3398-404.

206. Liljeström, V.; Mikkilä, J.; Kostiainen, M. A., Self-assembly and modular functionalization of three-dimensional crystals from oppositely charged proteins. *Nature Communications* **2014**, *5*, 4445.

207. McCoy, K.; Uchida, M.; Lee, B.; Douglas, T., Templated Assembly of a Functional Ordered Protein Macromolecular Framework from P22 Virus-like Particles. *ACS Nano* **2018**, *12* (4), 3541-3550.

208. Uchida, M.; McCoy, K.; Fukuto, M.; Yang, L.; Yoshimura, H.; Miettinen, H. M.; LaFrance, B.; Patterson, D. P.; Schwarz, B.; Karty, J. A.; Prevelige, P. E., Jr.; Lee, B.;

Douglas, T., Modular Self-Assembly of Protein Cage Lattices for Multistep Catalysis. *ACS Nano* **2018**, *12* (2), 942-953.

209. Koch, C.; Wabbel, K.; Eber, F. J.; Krolla-Sidenstein, P.; Azucena, C.; Gliemann, H.; Eiben, S.; Geiger, F.; Wege, C., Modified TMV Particles as Beneficial Scaffolds to Present Sensor Enzymes. *Frontiers in Plant Science* **2015**, *6* (1137).

210. Yoshimura, H.; Edwards, E.; Uchida, M.; McCoy, K.; Roychoudhury, R.; Schwarz, B.; Patterson, D.; Douglas, T., Two-Dimensional Crystallization of P22 Virus-Like Particles. *The Journal of Physical Chemistry B* **2016**, *120* (26), 5938-5944.

211. Asor, R.; Ben-Nun-Shaul, O.; Oppenheim, A.; Raviv, U., Crystallization, Reentrant Melting, and Resolubilization of Virus Nanoparticles. *ACS Nano* **2017**, *11* (10), 9814-9824.

212. Aumiller, W. M.; Uchida, M.; Biner, D. W.; Miettinen, H. M.; Lee, B.; Douglas, T., Stimuli Responsive Hierarchical Assembly of P22 Virus-like Particles. *Chemistry of Materials* **2018**, *30* (7), 2262-2273.

213. Alloyeau, D.; Stéphanidis, B.; Zhao, X.; Larquet, E.; Boisset, N.; Ricolleau, C., Biotemplated Synthesis of Metallic Nanoclusters Organized in Tunable Two-Dimensional Superlattices. *The Journal of Physical Chemistry C* **2011**, *115* (43), 20926-20930.

214. Johnson, C. A.; Yuan, Y.; Lenhoff, A. M., Adsorbed Layers of Ferritin at Solid and Fluid Interfaces Studied by Atomic Force Microscopy. *Journal of Colloid and Interface Science* **2000**, *223* (2), 261-272.

215. Suci, P. A.; Klem, M. T.; Douglas, T.; Young, M., Influence of electrostatic interactions on the surface adsorption of a viral protein cage. *Langmuir* **2005**, *21* (19), 8686-93.

216. Klem, M. T.; Willits, D.; Young, M.; Douglas, T., 2-D array formation of genetically engineered viral cages on au surfaces and imaging by atomic force microscopy. *Journal of the American Chemical Society* **2003**, *125* (36), 10806-7.

217. Royston, E.; Ghosh, A.; Kofinas, P.; Harris, M. T.; Culver, J. N., Self-assembly of virus-structured high surface area nanomaterials and their application as battery electrodes. *Langmuir* **2008**, *24* (3), 906-12.

218. Fukuto, M.; Nguyen, Q. L.; Vasilyev, O.; Mank, N.; Washington-Hughes, C. L.; Kuzmenko, I.; Checco, A.; Mao, Y.; Wang, Q.; Yang, L., Crystallization, structural diversity and anisotropy effects in 2D arrays of icosahedral viruses. *Soft Matter* **2013**, *9* (40), 9633-9642.

219. Kewalramani, S.; Wang, S.; Lin, Y.; Nguyen, H. G.; Wang, Q.; Fukuto, M.; Yang, L., Systematic approach to electrostatically induced 2D crystallization of nanoparticles at liquid interfaces. *Soft Matter* **2011**, *7* (3), 939-945.

220. Blum, A. S.; Soto, C. M.; Sapsford, K. E.; Wilson, C. D.; Moore, M. H.; Ratna, B. R., Molecular electronics based nanosensors on a viral scaffold. *Biosensors and Bioelectronics* **2011**, *26* (6), 2852-2857.

221. Cheung, C. L.; Camarero, J. A.; Woods, B. W.; Lin, T.; Johnson, J. E.; De Yoreo, J. J., Fabrication of Assembled Virus Nanostructures on Templates of Chemoselective Linkers Formed by Scanning Probe Nanolithography. *Journal of the American Chemical Society* **2003**, *125* (23), 6848-6849.

222. Shin, Y.-H.; Yun, S.-H.; Pyo, S.-H.; Lim, Y.-S.; Yoon, H.-J.; Kim, K.-H.; Moon, S.-K.; Lee, S. W.; Park, Y. G.; Chang, S.-I.; Kim, K.-M.; Lim, J.-H., Polymer-Coated Tips

for Patterning of Viruses by Dip-Pen Nanolithography. *Angewante Chemie International Edition* **2010**, *49* (50), 9689-9692.

223. Smith, J. C.; Lee, K.-B.; Wang, Q.; Finn, M. G.; Johnson, J. E.; Mrksich, M.; Mirkin, C. A., Nanopatterning the Chemospecific Immobilization of Cowpea Mosaic Virus Capsid. *Nano Letters* **2003**, *3* (7), 883-886.

224. Yoo, P. J.; Nam, K. T.; Belcher, A. M.; Hammond, P. T., Solvent-Assisted Patterning of Polyelectrolyte Multilayers and Selective Deposition of Virus Assemblies. *Nano Letters* **2008**, *8* (4), 1081-1089.

225. Strable, E.; Johnson, J. E.; Finn, M. G., Natural Nanochemical Building Blocks: Icosahedral Virus Particles Organized by Attached Oligonucleotides. *Nano Letters* **2004**, *4* (8), 1385-1389.

226. Künzle, M.; Eckert, T.; Beck, T., Metal-Assisted Assembly of Protein Containers Loaded with Inorganic Nanoparticles. *Inorganic Chemistry* **2018**, *57* (21), 13431-13436.

227. Kostiainen, M. A.; Kasyutich, O.; Cornelissen, J. J. L. M.; Nolte, R. J. M., Self-assembly and optically triggered disassembly of hierarchical dendron–virus complexes. *Nature Chemistry* **2010**, *2*, 394.

228. Mikkilä, J.; Rosilo, H.; Nummelin, S.; Seitsonen, J.; Ruokolainen, J.; Kostiainen, M. A., Janus-Dendrimer-Mediated Formation of Crystalline Virus Assemblies. *ACS Macro Letters* **2013**, *2* (8), 720-724.

229. Moon, H.; Kim, W. G.; Lim, S.; Kang, Y. J.; Shin, H.-H.; Ko, H.; Hong, S. Y.; Kang, S., Fabrication of uniform layer-by-layer assemblies with complementary protein cage nanobuilding blocks via simple His-tag/metal recognition. *Journal of Materials Chemistry B* **2013**, *1* (35), 4504-4510.

230. Kang, Y. J.; Uchida, M.; Shin, H.-H.; Douglas, T.; Kang, S., Biomimetic FePt nanoparticle synthesis within Pyrococcus furiosus ferritins and their layer-by-layer formation. *Soft Matter* **2011**, *7* (23), 11078-11081.

231. Suci, P. A.; Klem, M. T.; Arce, F. T.; Douglas, T.; Young, M., Assembly of Multilayer Films Incorporating a Viral Protein Cage Architecture. *Langmuir* **2006**, *22* (21), 8891-8896.

232. Uchida, M.; LaFrance, B.; Broomell, C. C.; Prevelige, P. E., Jr.; Douglas, T., Higher order assembly of virus-like particles (VLPs) mediated by multi-valent protein linkers. *Small* **2015**, *11* (13), 1562-70.

233. Kostiainen, M. A.; Hiekkataipale, P.; Laiho, A.; Lemieux, V.; Seitsonen, J.; Ruokolainen, J.; Ceci, P., Electrostatic assembly of binary nanoparticle superlattices using protein cages. *Nature Nanotechnology* **2012**, *8*, 52.

234. Choi, H.; Choi, B.; Kim, G. J.; Kim, H.-u.; Kim, H.; Jung, H. S.; Kang, S., Fabrication of Nanoreaction Clusters with Dual-Functionalized Protein Cage Nanobuilding Blocks. *Small* **2018**, *14* (35), 1801488.

235. Lin, X.; Xie, J.; Niu, G.; Zhang, F.; Gao, H.; Yang, M.; Quan, Q.; Aronova, M. A.; Zhang, G.; Lee, S.; Leapman, R.; Chen, X., Chimeric Ferritin Nanocages for Multiple Function Loading and Multimodal Imaging. *Nano Letters* **2011**, *11* (2), 814-819.

236. Geninatti Crich, S.; Bussolati, B.; Tei, L.; Grange, C.; Esposito, G.; Lanzardo, S.; Camussi, G.; Aime, S., Magnetic Resonance Visualization of Tumor Angiogenesis by Targeting Neural Cell Adhesion Molecules with the Highly Sensitive Gadolinium-Loaded Apoferritin Probe. *Cancer Research* **2006**, *66* (18), 9196.

237. Fantechi, E.; Innocenti, C.; Zanardelli, M.; Fittipaldi, M.; Falvo, E.; Carbo, M.; Shullani, V.; Di Cesare Mannelli, L.; Ghelardini, C.; Ferretti, A. M.; Ponti, A.; Sangregorio, C.; Ceci, P., A Smart Platform for Hyperthermia Application in Cancer Treatment: Cobalt-Doped Ferrite Nanoparticles Mineralized in Human Ferritin Cages. *ACS Nano* **2014**, *8* (5), 4705-4719.

238. Li, X.; Qiu, L.; Zhu, P.; Tao, X.; Imanaka, T.; Zhao, J.; Huang, Y.; Tu, Y.; Cao, X., Epidermal Growth Factor–Ferritin H-Chain Protein Nanoparticles for Tumor Active Targeting. *Small* **2012**, *8* (16), 2505-2514.

239. Steinmetz, N. F.; Ablack, A. L.; Hickey, J. L.; Ablack, J.; Manocha, B.; Mymryk, J. S.; Luyt, L. G.; Lewis, J. D., Intravital imaging of human prostate cancer using viral nanoparticles targeted to gastrin-releasing Peptide receptors. *Small* **2011**, *7* (12), 1664-72.

240. Brunel, F. M.; Lewis, J. D.; Destito, G.; Steinmetz, N. F.; Manchester, M.; Stuhlmann, H.; Dawson, P. E., Hydrazone ligation strategy to assemble multifunctional viral nanoparticles for cell imaging and tumor targeting. *Nano Letters* **2010**, *10* (3), 1093-7.

241. Huang, R. K.; Steinmetz, N. F.; Fu, C.-Y.; Manchester, M.; Johnson, J. E., Transferrin-mediated targeting of bacteriophage HK97 nanoparticles into tumor cells. *Nanomedicine* **2010**, *6* (1), 55-68.

242. Tong, G. J.; Hsiao, S. C.; Carrico, Z. M.; Francis, M. B., Viral Capsid DNA Aptamer Conjugates as Multivalent Cell-Targeting Vehicles. *Journal of the American Chemical Society* **2009**, *131* (31), 11174-11178.

243. Kim, H.; Choi, H.; Bae, Y.; Kang, S., Development of target-tunable P22 VLP-based delivery nanoplatforms using bacterial superglue. *Biotechnology and Bioengineering* **2019**, *116* (11), 2843-2851.

244. Chariou, P. L.; Lee, K. L.; Wen, A. M.; Gulati, N. M.; Stewart, P. L.; Steinmetz, N. F., Detection and Imaging of Aggressive Cancer Cells Using an Epidermal Growth Factor Receptor (EGFR)-Targeted Filamentous Plant Virus-Based Nanoparticle. *Bioconjugate Chemistry* **2015**, *26* (2), 262-269.

245. Bruckman, M. A.; VanMeter, A.; Steinmetz, N. F., Nanomanufacturing of Tobacco Mosaic Virus-Based Spherical Biomaterials Using a Continuous Flow Method. *ACS Biomaterials Science & Engineering* **2015**, *1* (1), 13-18.

246. Moon, H.; Lee, J.; Min, J.; Kang, S., Developing Genetically Engineered Encapsulin Protein Cage Nanoparticles as a Targeted Delivery Nanoplatform. *Biomacromolecules* **2014**, *15* (10), 3794-3801.

247. Moon, H.; Lee, J.; Kim, H.; Heo, S.; Min, J.; Kang, S., Genetically engineering encapsulin protein cage nanoparticle as a SCC-7 cell targeting optical nanoprobe. *Biomaterials Research* **2014**, *18* (1), 21.

248. Falvo, E.; Tremante, E.; Fraioli, R.; Leonetti, C.; Zamparelli, C.; Boffi, A.; Morea, V.; Ceci, P.; Giacomini, P., Antibody–drug conjugates: targeting melanoma with cisplatin encapsulated in protein-cage nanoparticles based on human ferritin. *Nanoscale* **2013**, *5* (24), 12278-12285.

249. Kang, H. J.; Kang, Y. J.; Lee, Y.-M.; Shin, H.-H.; Chung, S. J.; Kang, S., Developing an antibody-binding protein cage as a molecular recognition drug modular nanoplatform. *Biomaterials* **2012**, *33* (21), 5423-5430.

250. Zhen, Z.; Tang, W.; Chen, H.; Lin, X.; Todd, T.; Wang, G.; Cowger, T.; Chen, X.; Xie, J., RGD-Modified Apoferritin Nanoparticles for Efficient Drug Delivery to Tumors. *ACS Nano* **2013**, *7* (6), 4830-4837.

251. Barwal, I.; Kumar, R.; Kateriya, S.; Dinda, A. K.; Yadav, S. C., Targeted delivery system for cancer cells consist of multiple ligands conjugated genetically modified CCMV capsid on doxorubicin GNPs complex. *Scientific Reports* **2016**, *6* (1), 37096.

252. Wu, Y.; Li, J.; Yang, H.; Seoung, J.; Lim, H.-D.; Kim, G.-J.; Shin, H.-J., Targeted cowpea chlorotic mottle virus-based nanoparticles with tumor-homing peptide F3 for photothermal therapy. *Biotechnology and Bioprocess Engineering* **2017**, *22* (6), 700-708.

253. Yildiz, I.; Lee, K. L.; Chen, K.; Shukla, S.; Steinmetz, N. F., Infusion of imaging and therapeutic molecules into the plant virus-based carrier cowpea mosaic virus: Cargo-loading and delivery. *Journal of Controlled Release* **2013**, *172* (2), 568-578.

254. Choi, K.-m.; Kim, K.; Kwon, I. C.; Kim, I.-S.; Ahn, H. J., Systemic Delivery of siRNA by Chimeric Capsid Protein: Tumor Targeting and RNAi Activity in Vivo. *Molecular Pharmaceutics* **2013**, *10* (1), 18-25.

255. Ghosh, D.; Kohli, A. G.; Moser, F.; Endy, D.; Belcher, A. M., Refactored M13 Bacteriophage as a Platform for Tumor Cell Imaging and Drug Delivery. *ACS Synthetic Biology* **2012**, *1* (12), 576-582.

256. Stephanopoulos, N.; Tong, G. J.; Hsiao, S. C.; Francis, M. B., Dual-Surface Modified Virus Capsids for Targeted Delivery of Photodynamic Agents to Cancer Cells. *ACS Nano* **2010**, *4* (10), 6014-6020.

257. Pan, Y.; Zhang, Y.; Jia, T.; Zhang, K.; Li, J.; Wang, L., Development of a microRNA delivery system based on bacteriophage MS2 virus-like particles. *The FEBS Journal* **2012**, *279* (7), 1198-1208.

258. Kwon, K. C.; Ryu, J. H.; Lee, J.-H.; Lee, E. J.; Kwon, I. C.; Kim, K.; Lee, J., Proteinticle/Gold Core/Shell Nanoparticles for Targeted Cancer Therapy without Nanotoxicity. *Advanced Materials* **2014**, *26* (37), 6436-6441.

259. Min, J.; Moon, H.; Yang, H. J.; Shin, H.-H.; Hong, S. Y.; Kang, S., Development of P22 viral capsid nanocomposites as anti-cancer drug, bortezomib (BTZ), delivery nanoplatforms. *Macromolecular Bioscience* **2014**, *14* (4), 557-564.

260. Pokorski, J. K.; Hovlid, M. L.; Finn, M. G., Cell Targeting with Hybrid Qβ Virus-Like Particles Displaying Epidermal Growth Factor. *ChemBioChem* **2011**, *12* (16), 2441-2447.

261. Rhee, J.-K.; Baksh, M.; Nycholat, C.; Paulson, J. C.; Kitagishi, H.; Finn, M. G., Glycan-Targeted Virus-like Nanoparticles for Photodynamic Therapy. *Biomacromolecules* **2012**, *13* (8), 2333-2338.

262. Tian, Y.; Gao, S.; Wu, M.; Liu, X.; Qiao, J.; Zhou, Q.; Jiang, S.; Niu, Z., Tobacco Mosaic Virus-Based 1D Nanorod-Drug Carrier via the Integrin-Mediated Endocytosis Pathway. *ACS Applied Materials & Interfaces* **2016**, *8* (17), 10800-10807.

263. Min, J.; Kim, S.; Lee, J.; Kang, S., Lumazine synthase protein cage nanoparticles as modular delivery platforms for targeted drug delivery. *RSC Advances* **2014**, *4* (89), 48596-48600.

264. Kim, H.; Kang, Y. J.; Min, J.; Choi, H.; Kang, S., Development of an antibody-binding modular nanoplatform for antibody-guided targeted cell imaging and delivery. *RSC Advances* **2016**, *6* (23), 19208-19213.

Chapter 2:

Functionalization of Encapsulin Surfaces via Genetically Embedded Peptide-binding Domains

2.1 – Introduction

Protein-based micro- and nano-scale containers are specialized three-dimensional architectures found throughout nature that are inherently designed to perform a variety of biologically essential tasks, including the packaging, protection, and transportation of genetic material, the generation of chemically distinct microenvironments, the co-localization of functionally-related enzymes, and the dynamic capture and release of specific metabolites according to cellular needs (1-6). Structural characterizations of proteinaceous containers from all kingdoms of life have revealed highly uniform macromolecular assemblies adopting a range of sizes and morphologies, the most common of which include spherical containers displaying polyhedral symmetries and rod-like containers forming extended tubular filaments (7-9). Assembled containers are typically generated from the polymerization of one or several protomeric subunit(s), and, in accordance with their diverse native functions, generally display unique biophysical properties. As a result of their natural polyvalency, monodispersity, structural plasticity, and biocompatibility, protein containers have attracted increasing attention from researchers over the last half century seeking to employ these macromolecular assemblies as functionalizable nano-scale vehicles for applications in medicine, industrial catalytic processes, and the development of next-generation biomaterials (9-11).

Among the natural proteinaceous cages, encapsulins are a recently established class of nanocontainer structures that are intrinsic to a variety of bacterial and archaeal species which function as simple organelles *in vivo* (12). Morphologically, encapsulins are polyhedral shells constructed in a homopolymeric fashion from a single conserved protomeric coat protein. Atomic resolution structures obtained for several encapsulins reveal cage assemblies on the order of 20 to

42 nm in external diameter adopting T = 1, T = 3, or T = 4 icosahedral symmetries constructed from 60, 180 or 240 subunits, respectively (12-15). Currently, encapsulin containers are believed to participate in the intracellular mitigation of oxidative stressors, and have been found to selectively encapsulate enzyme cargoes such as dye-decolorizing peroxidases, ferritin-like iron mineralizing proteins, haemerythrins, and two-domain nitrite reductase-hydroxylamine oxidases (6, 12, 16). Transport of metabolites to and from the encapsulin lumen is accomplished via a series of largely conserved pore structures located at the 2-fold, 3-fold, and 5-fold symmetry axes (12). Due to their small sizes and uniform construction, encapsulins have become attractive candidates for the development of tailored, biorthogonal protein scaffolds in recent years (16).

The exterior decoration of proteinaceous nanocontainers represents a powerful tool for the introduction of novel properties into these robust biological scaffolds. This phenomenon has been observed to occur naturally with certain viral capsids possessing associated surface proteins, many of which are involved in either facilitating viral tropism and infectivity or in stabilizing the assembled capsid superstructure (17-20). Mimicking this natural enhancement of container structures, previous works have shown a number of nanocontainers to be amenable to both chemical and genetic derivatizations as means for tailored surface engineering (10, 21, 22). While both methodologies are viable, genetic derivatization is particularly appealing as modifications made to promoters are subsequently presented in a uniform, symmetrical fashion on the surfaces of the resulting assembled cages.

In the last two decades, a variety of small peptide binding domains (PBDs) isolated from natural protein sources have been used to scaffold disparate biocatalysts, structural proteins, or other biological molecules on the basis of programmed protein-protein interactions that occur between the PBDs and corresponding peptide ligand (PL) sequences (23-27). Each PBD recognizes a specific ligand sequence, allowing researchers to template biorthogonal associations to suit both *in vitro* and *in vivo* applications. Several prominent PBDs include the Src homology 3 (SH3) and the PSD95/DlgA/Zo-1 (PDZ) domains isolated from mouse metazoan signaling pathways, which non-covalently bind the peptide ligand sequences PPPALPPKRRR ($K_d = 0.1$ μM) and GVKESLV ($K_d = 8$ μM), respectively (23, 28, 29) (**Figure 2.1**). In more recent years, the SpyCatcher/SpyTag bacterial superglue system was developed by Zakeri *et al.* as a means of covalently tethering proteins together by exploiting the native ability of the CnaB2 domain from the *Streptococcus pyogenes* fibronectin binding protein FbaB to form intramolecular isopeptide bonds (30) (**Figure 2.1**). Collectively, these PBDs have been applied to the development of artificial metabolic pathways (23, 24), the enhancement of protein stability through self-cyclization (31), and the production of functionalizable biofilm matrices (32).

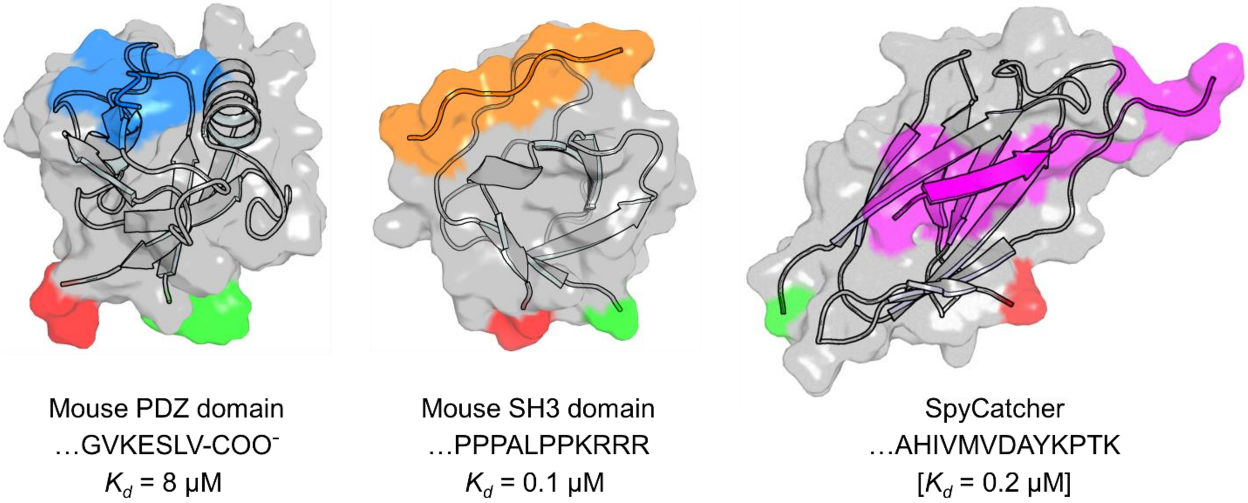


Figure 2.1 – Atomic structures for the PDZ (PDB: 2PDZ), SH3 (PDB: 1CKB), and SpyCatcher (PDB: 4MLI) protein binding domains. Surface densities corresponding to the N- and C-termini for each PBD are colored in green and red, respectively. The amino acid sequences of the individual peptide ligands specific to the PDZ, SH3, and SpyCatcher domains are displayed under each structure along with their respective K_d values. The K_d value for SpyCatcher is bracketed as it is an assumed theoretical value due to the covalent nature of the SpyCatcher/SpyTag interaction (30). The peptide ligands are depicted in blue, orange, and magenta in the atomic structures for the PDZ, SH3, and SpyCatcher domains, respectively.

In this study, we report the presentation of both non-covalent and covalent PBDs on the exterior surface of encapsulins abstracted from the hyperthermophilic bacterium *Thermotoga maritima* (TmE). Surface presentation of PBDs was accomplished by either genetically incorporating PBD genes within solvent-exposed flexible loops of the TmE protomer subunit or by appending PBDs directly to the TmE's exterior-oriented C-terminus. TmE fusion proteins containing embedded PDZ or SH3 domains were subsequently used in attempts to generate ordered supramolecular assemblies in a heterogeneous fashion with partnered TmEs displaying recombinantly fused PDZ or SH3 ligand sequences on their C-termini. Additionally, TmEs containing C-terminally fused SpyCatcher domains were used as polyvalent scaffolds for the high-density localization of *Escherichia coli* dihydrofolate reductase (DHFR) variants possessing corresponding SpyTag sequences. The results presented herein show that TmEs are tolerant to genetic insertions at a number of locations within the protomer sequence, and thus the generation of TmE-PBD chimeras represents a general strategy for introducing new functionalities into these robust nanocontainers.

2.2 – Results and Discussion

2.2.1 – External Display of Non-covalent Peptide Binding Domains

Our first research aim was to attempt to generate higher-order supramolecular structures using paired TmE variants displaying non-covalent PBDs and their corresponding PLs on their exterior surfaces. Accordingly, we first set about designing the two populations of nanocontainer variants with one population consisting of TmEs presenting surface exposed PDZ or SH3 peptide-binding domains and the other population consisting of TmEs displaying the corresponding PDZ or SH3 ligand sequences. For TmEs presenting PBDs, several segments of the TmE monomer containing flexible, solvent-exposed loops were chosen for the genetic insertion of the PBD sequences. Loop insertion was reasoned to be a feasible design strategy due to the small sizes of the PDZ (10.5 kDa) and SH3 (6.9 kDa) domains, the close spatial proximity of the native N- and C-termini in both PBDs, and due to the PL binding clefts being on the opposing faces of the PBDs with respect to their termini (**Figure 2.1**). PBDs were subsequently inserted between TmE residues 59 and 60 (TmE_{59}-PDZ and TmE_{59}-SH3), residues 124 and 125 (TmE_{124}-PDZ and TmE_{124}-SH3), or between residues 135 and 136 (TmE_{135}-PDZ and TmE_{135}-SH3) (**Figure 2.2A**). Once assembled, PBDs inserted at these positions should adopt 3-fold, 2-fold, or 5-fold symmetry arrangements, respectively (**Figure 2.2B**). An additional glycine residue was encoded at both the 5' and 3' ends of the PBDs to generate slight flexibility into the backbone of the fusion protein where the ends of the PBDs meet the TmE monomer. For TmE variants presenting the corresponding PLs, amino acid sequences for the PDZ and SH3 ligands were fused to the TmE C-terminus (denoted TmE-PDZlig or TmE-SH3lig, respectively), which has been shown previously to be both solvent exposed and amenable to small peptide affinity-tag fusions (12, 33-35). In the case of TmE-PDZlig, C-terminal fusion of the PDZ ligand sequence was an obligate requirement as the terminal carboxylate moiety is essential for proper ligand binding within the PDZ domain (29).

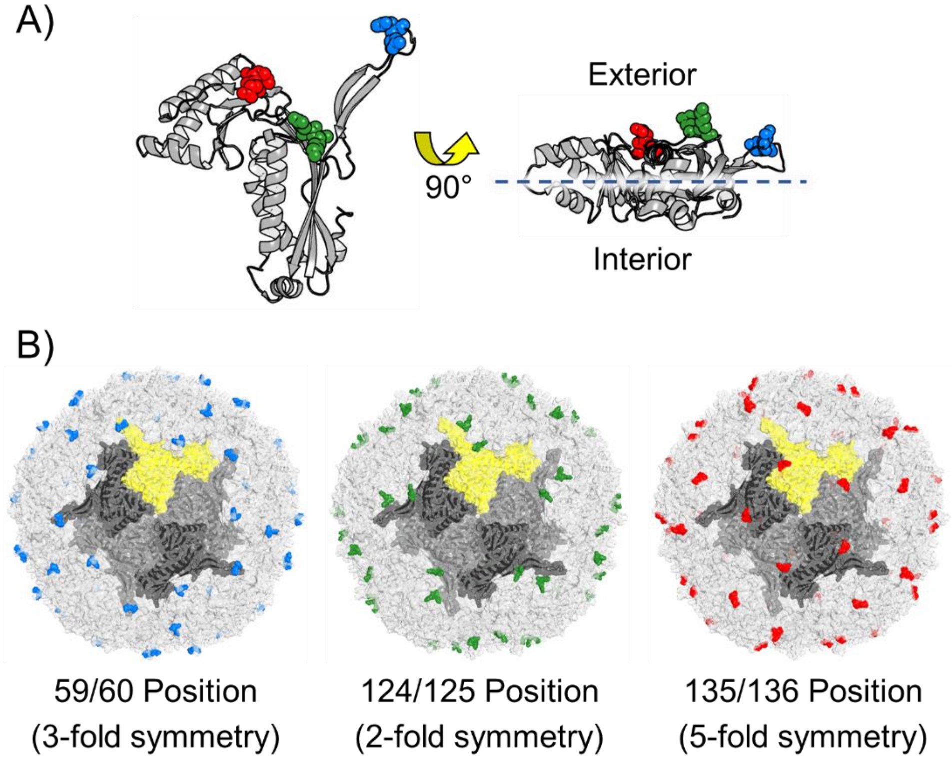


Figure 2.2 – A) TmE protomeric subunit (PDB: 3DKT) with residues 59-60, 124-125, and 135-136 depicted as blue, green, and red spheres, respectively. B) Symmetrical distributions of PBD insertion positions across assembled nanocontainer surfaces.

TmE-PBD variants were purified to homogeneity using C-terminal hexahistidine affinity tags in conjunction with immobilized metal affinity chromatography (IMAC). Purified samples of both wildtype and variant TmEs were visually assessed via transmission electron microscopy (TEM) and were found to appear morphologically identical under the staining conditions and acceleration voltage used for imaging with external diameters matching previously published values of approximately 24-25 nm (**Figure 2.3**) (12). Given the overall size of the TmE nanocontainers in relation to the small sizes of the two PBDs (PDZ domain = 10.5 kDa; SH3 domain = 6.9 kDa), it is perhaps unsurprising that the PBDs do not appear visible in the TEM images collected. In conjunction with TEM analyses, dynamic light scattering (DLS) was used to assess the solution hydrodynamic diameters of the TmE variants (**Table 2.1**). Consistent with the

TEM data, wildtype TmE containers were shown to possess hydrodynamic diameters of approximately 24 nm. Interestingly, the TmE_{59}-PDZ, TmE_{124}-PDZ and TmE_{135}-PDZ variants also possessed hydrodynamic diameters of roughly 24 nm while the TmE_{59}-SH3, TmE_{124}-SH3 and TmE_{135}-SH3 variants exhibited diameters of approximately 30 to 31 nm, which seems counterintuitive given the smaller size of the SH3 domains. However, while the increases in hydrodynamic diameter are seemingly consistent for the SH3-containing variants, the increases are also both relatively minor and within error of the values collected for wildtype TmE and the PDZ-containing variants. TEM images were not collected for TmE-PDZlig or TmE-SH3lig variants as it was assumed that the flexible PLs would not be observable under the imaging conditions used. However, DLS measurements show that both TmE-PL variants possess hydrodynamic diameters identical to wildtype TmE (**Table 2.1**).

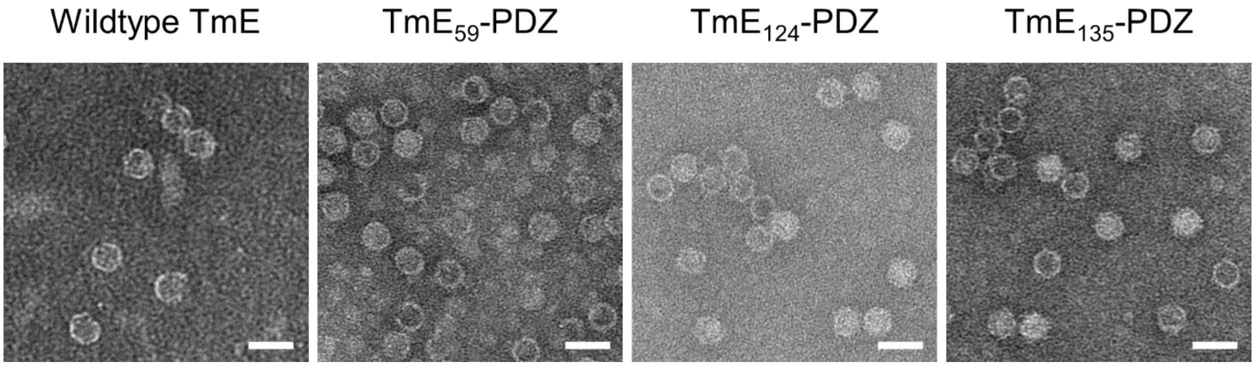
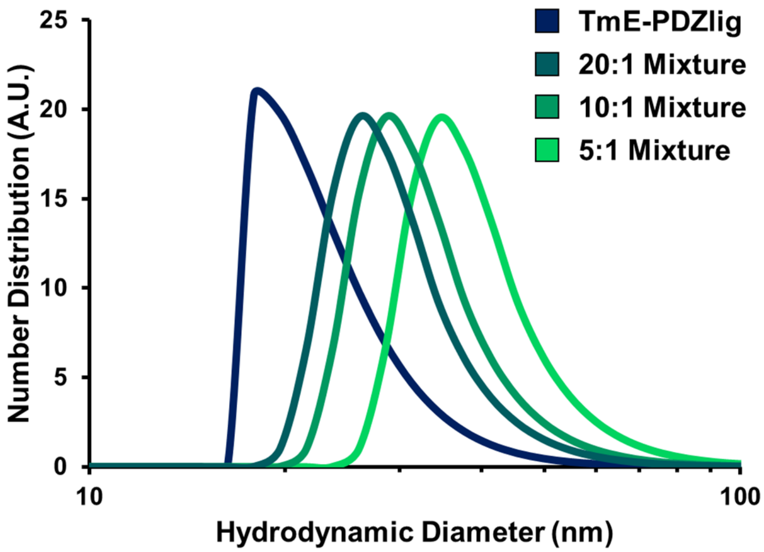


Figure 2.3 – Transmission electron microscopy images of TmE-PDZ encapsulin fusion proteins. TmE-SH3 fusion proteins are not shown as they appeared visually indistinguishable from either wildtype TmEs or TmE-PDZ variants under the imaging conditions used. Scale bars = 40 nm.

Following purification and initial biophysical characterization of the nanocontainer variants, we next attempted to form higher-order heterogeneous assemblies by creating *in vitro* mixtures of TmE-PBD and TmE-PL variants. As an initial test, we chose to create mixtures of TmE_{135}-PDZ and TmE-PDZlig as a representative test case given that the 135/136 insertion position appears to have the most uniform distribution across the surface of the TmE containers. For the preliminary test, a static mass of 0.2 mg of purified TmE-PDZlig was mixed with varying masses of TmE_{135}-PDZ (final mass values between 0.0125 and 0.0500 mg), representing mole ratios of TmE-PDZlig to TmE_{135}-PDZ between 20:1 and 5:1 . Following an overnight incubation at 4°C to promote the association of the PBDs and PLs, the mixtures were assessed via DLS (**Figure 2.4**), showing consistent increases in the average particle size correlated with the decreasing mass ratio of TmE-PDZlig to TmE_{135}-PDZ. Interestingly, these assemblies appeared to be somewhat stable as the same increases in particle size were still present upon re-assessment following three days' storage at 4°C, though the average particle sized decreased slightly for each of the heterogeneous mixtures.

Variant	Hydrodynamic Diameter (nm)
Wildtype TmE	24.3 ± 6.7
TmE_{59}-PDZ	24.6 ± 5.1
TmE_{124}-PDZ	25.2 ± 6.4
TmE_{135}-PDZ	23.7 ± 6.7
TmE-PDZlig	23.5 ± 6.6
TmE_{59}-SH3	30.7 ± 7.9
TmE_{124}-SH3	30.3 ± 7.4
TmE_{135}-SH3	30.7 ± 7.3
TmE-SH3lig	24.1 ± 6.2

Table 2.1 – Dynamic light scattering assessments of TmE-PBD and TmE-PL variants



TmE-PDZlig : TmE$_{135}$-PDZ Ratio	Hydrodynamic Diameter (nm)	
	Day 1	Day 3
TmE-PDZlig only	23.5 ± 6.6 (0.254)	24.4 ± 7.0 (0.257)
20:1	30.3 ± 8.3 (0.274)	25.7 ± 7.5 (0.278)
10:1	33.2 ± 9.1 (0.278)	28.3 ± 8.4 (0.214)
5:1	40.1 ± 11.1 (0.183)	34.4 ± 10.4 (0.216)

Figure 2.4 – Dynamic light scattering profiles for *in vitro* sample mixtures containing variable mole ratios of TmE-PDZlig to TmE$_{135}$-PDZ. Average hydrodynamic diameter values at 1 and 3 days post mixing are displayed in the table below. Values in parentheses represent polydispersity indices for each sample.

Encouraged by this initial success, we sought to scale up our *in vitro* mixtures, adding variable molar ratios of TmE-PBDs and TmE-PLs up to equimolar 1:1 mixtures. As before, mixtures were prepared by combining TmE-PBDs and TmE-PLs in a common buffer solution and incubating them overnight at 4°C prior to DLS assessment. The scattering intensity data for the expanded *in vitro* mixtures showed that mixtures all of the TmE-PDB and TmE-PL mixtures tested exhibited similar trends. Specifically, mixtures of 100:1 and 1:100 of TmE-PDBs to TmE-PLs, regardless of PBD insertion position, showed minimal changes in their respective scattering intensity profiles relative to pure samples of the TmE-PBDs or TmE-PLs alone. However, as the molar ratios between the TmE-PBDs and TmE-PLs became smaller (i.e. ratios of 10:1 up to 1:1, or 1:10 up to 1:1), dramatic increases in scattering intensity for larger diameter assemblies became apparent, presumably due to rampant association of the two TmE variants. While this association of TmEs was the intent of our work, the apparent *in vitro* binding of TmE-PBDs and TmE-PLs was not an ordered process as visible aggregates emerged in our samples as the ratio of the two TmE variants approached 1:1 (**Figure 2.5A**). This undesirable aggregation was also observed spectrophotometrically by measuring incident light scattering at 350 nm (**Figure 2.5B**). Control experiments in which either TmE-PBDs or TmE-PLs were mixed at 1:1 molar ratios with wildtype TmE showed no changes in the scattering profiles relative to the individual TmE variants alone, indicating that the observed aggregation for mixtures of TmE-PBDs and TmE-PLs can be attributed to the protein-protein interactions mediated by the PDZ or SH3 domains with their corresponding ligand sequences (data not shown).

Reasoning that the formation of aggregates was likely the result of rapid, kinetically driven protein-protein interactions between the TmE-PBDs and TmE-PLs rather than thermodynamically favorable ones, we subsequent sought to optimize the *in vitro* mixing and sample incubation conditions to promote more thermodynamically driven, ordered assembly of supramolecular structures. However, attempts to optimize the association of TmE variants, including variations in buffer pH, ionic strength, incubation time, incubation temperature, and even attempts to reduce the net valency of the TmE-PBDs using chemically synthesized PDZ and SH3 linkers prior to mixing with TmE-PLs, proved ineffectual. Consequently, we decided to move on from our attempts to form higher-order structures in favor or functionalizing the exterior surfaces of TmEs with functional biocatalysts instead.

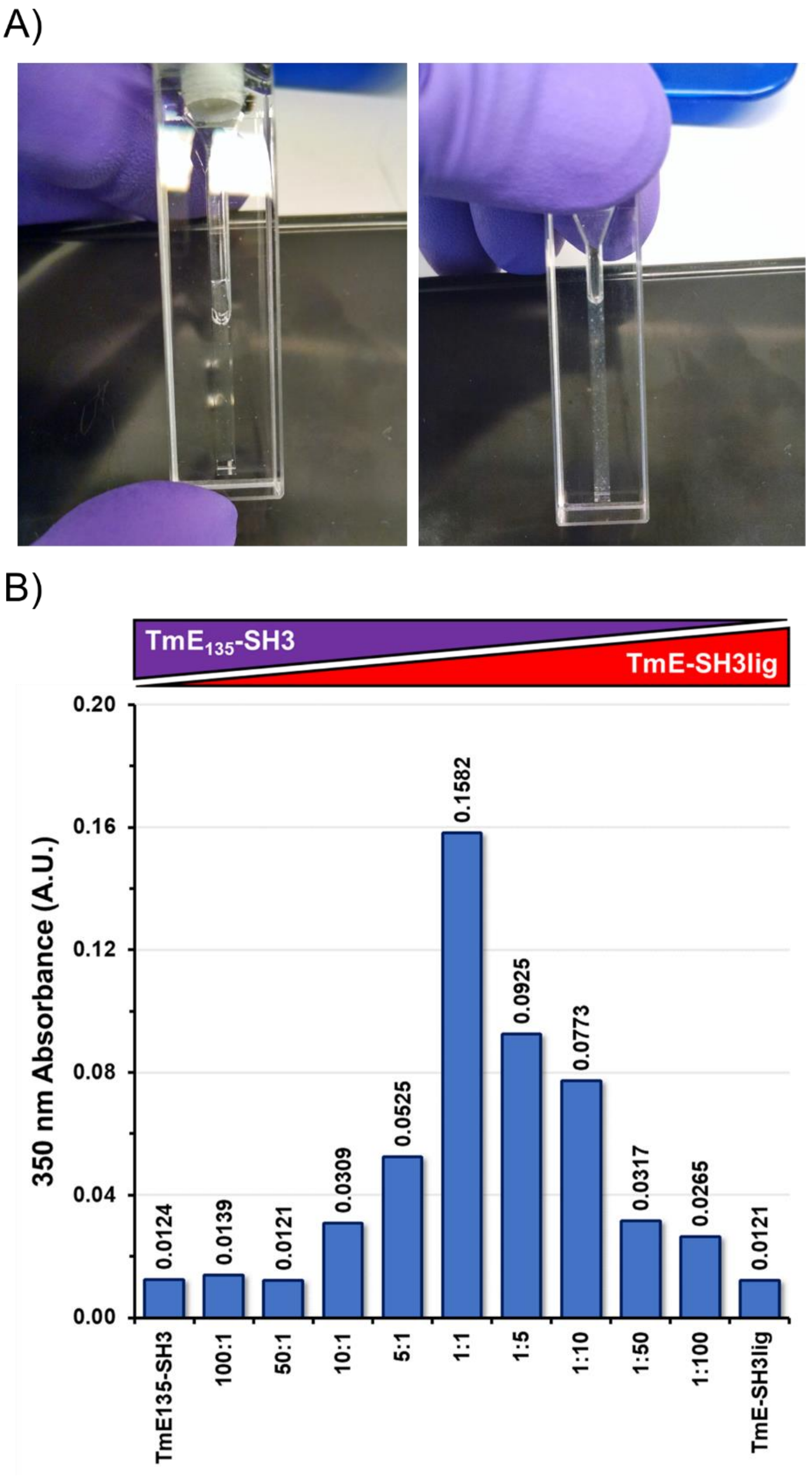


Figure 2.5 – A) The left-hand picture depicts a DLS cuvette containing TmE135-SH3 only. The right-hand picture depicts the same cuvette following the addition of a 1:1 molar ratio of TmE-SH3lig, resulting in visible, polydisperse aggregates. B) The chart represents a spectrophotometric light scattering assessment of aggregates formed upon mixing variable molar ratios of TmE135-SH3 and TmE-SH3lig. The values above each bar represent the background-corrected UV absorbance value at 350 nm for each sample.

2.2.2 – Covalent Display of Functional DHFR Enzymes

Following the problematic issues arising from the use of non-covalent PDZ and SH3 domains for encapsulin functionalization, we next sought to decorate the exterior of TmE encapsulins using the SpyCatcher/SpyTag bacterial superglue system as a means of covalently tethering biocatalysts to nanocontainer surfaces (30). External presentation of SpyCatcher domains was accomplished by fusing the SpyCatcher gene sequence onto the C-terminus of the TmE monomer gene to form the TmE-SpyCatcher (TmE-SC) hybrid (**Figure 2.6**). The TmE C-terminus was chosen for the SpyCatcher domain fusion to simplify the molecular cloning process as the generation of the PDZ and SH3 domain loop insertions used previously had been difficult and time

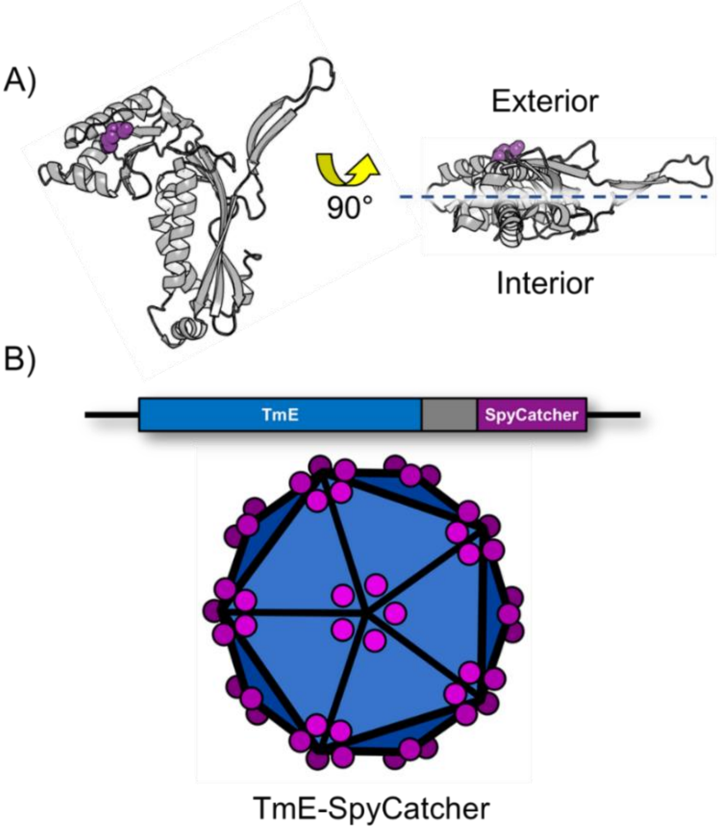


Figure 2.6 – A) TmE protomeric subunit (PDB: 3DKT) with the C-terminal residue depicted as purple spheres. B) Cartoon representation of the TmE-SpyCatcher fusion gene (top) with a 15 amino acid flexible linker sequence between the two genes depicted in grey. Below is a cartoon representation of the pentameric distribution of SpyCatcher domains across the nanocontainer surface.

consuming. Additionally, the C-termini of adjacent TmE monomers are spaced roughly 3 nm apart from one another in a pentagonal arrangement around the nanocontainer's 5-fold vertices upon assembly, which we reasoned should provide sufficient space for decoration with SpyTag-containing biocatalysts (30). Purified TmE-SC fusion proteins were morphologically assessed via TEM and were found to be visually indistinguishable from native TmE nanocontainers (**Figure 2.7**). This result is consistent with TEM images collected previously for the PDZ and SH3 decorated encapsulins, and with previous reports in which the SpyCatcher domain specifically was not visible on nanocontainer surfaces due to the combination of its small size (12.4 kDa), the negative staining conditions, and the accelerating voltages used for visual assessment during TEM imaging (36).

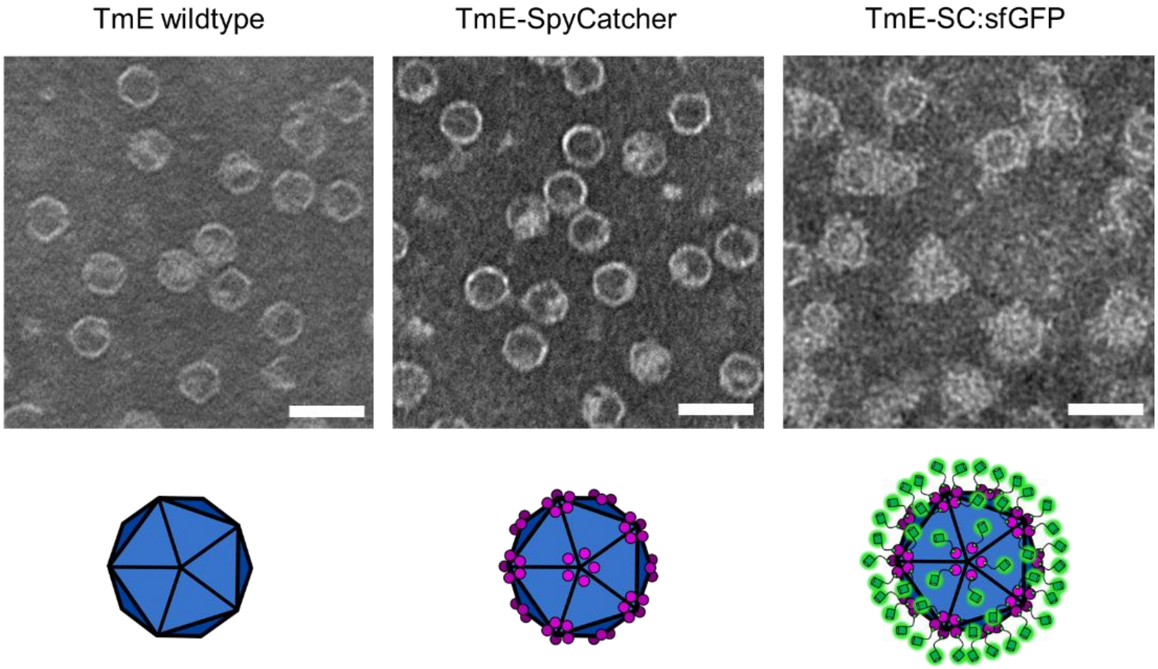


Figure 2.7 – Transmission electron microscopy images of wildtype TmE, TmE-SpyCatcher, and TmE-SpyCatcher with covalently tethered sfGFP-ST reporter proteins. Scale bars = 40 nm.

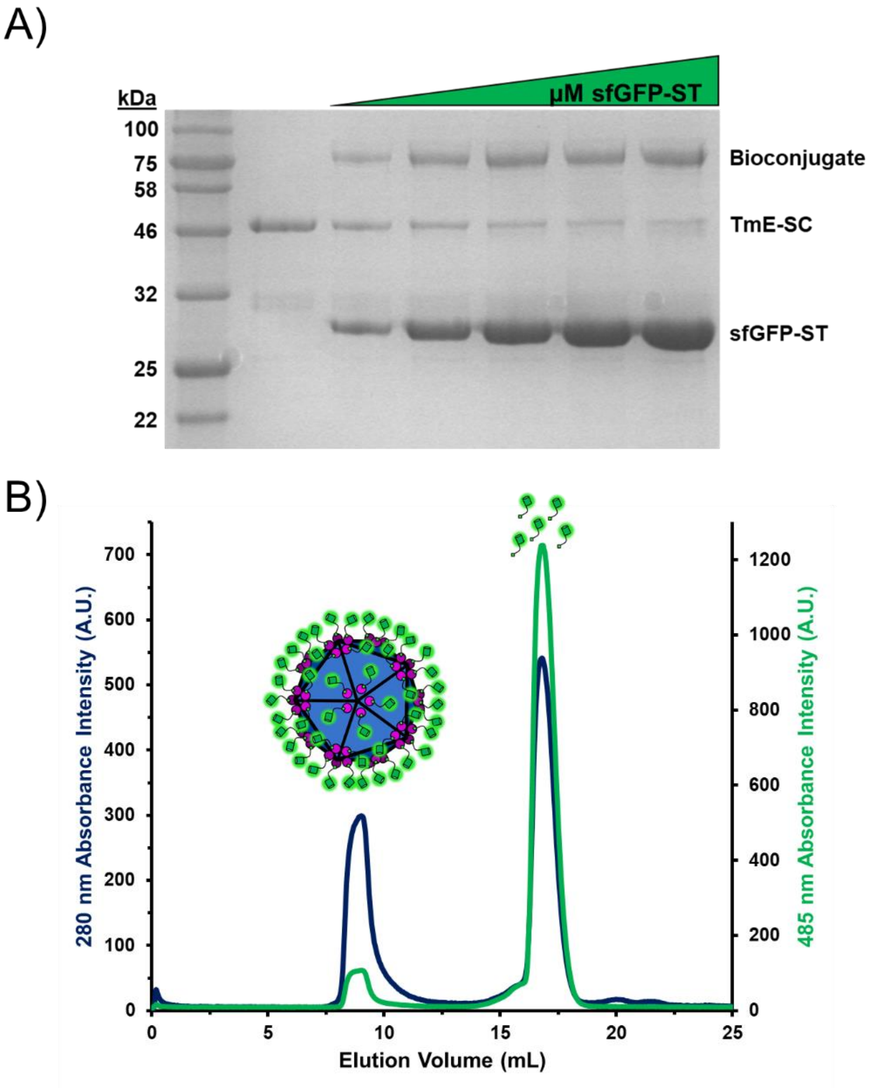


Figure 2.8 – A) SDS-PAGE analysis of *in vitro* titration experiments showing the progressive accumulation of a high molecular weight TmE-SC:sfGFP-ST bioconjugate when the concentration of sfGFP-ST is increased relative to a static concentration of TmE-SC. B) SEC chromatogram depicting the separation of bioconjugated nanocontainers formed from an *in vitro* reaction mixture containing excess sfGFP-ST. TmE-SC:sfGFP-ST conjugates elute near the void volume of the SEC column while excess sfGFP-ST elutes approximately 8 mL later.

Having shown that TmE-SpyCatcher hybrids retain the ability to self-assemble into macromolecular nanocontainers, we next sought to probe the functionality of the encapsulin-bound SpyCatcher domains by recombinantly fusing the AHIVMVDAYKPTK sequence of the corresponding SpyTag ligand onto the N-terminus of superfolder GFP (sfGFP-ST). **Figure 2.8A** shows that *in vitro* titration of the 44.0 kDa TmE-SC with purified 29.9 kDa sfGFP-ST leads to the formation of a high molecular weight bioconjugate product of approximately 74 kDa in SDS-PAGE tests, indicating that the surface-presented SpyCatcher domains retain their native ability to form isopeptide bonds with available SpyTag fusion proteins. Covalent capture of sfGFP-ST by TmE-SC was further verified during size exclusion chromatography analysis by the presence of a 485 nm absorbance signal, corresponding to the absorbance of the sfGFP chromophore (37), at the same elution volume as the bioconjugated protein product (**Figure 2.8B**). Subsequent TEM images collected from the SEC-purified TmE-SC:sfGFP-ST bioconjugates show that the nanocontainers lose much of the surface clarity previously seen for the wildtype and TmE-SC containers, indicative of sfGFP-ST surface attachment (**Figure 2.7**).

Following the successful covalent attachment of sfGFP-ST probes on encapsulin surfaces, the SpyTag sequence was next recombinantly fused onto the N-terminus of *E. coli* dihydrofolate reductase (DHFR-NST) in order to assess the catalytic behavior of nanocontainer-bound enzymes. A subsequent *in vitro* titration of purified TmE-SC with DHFR-NST showed the formation of a similar high molecular weight bioconjugate protein in SDS-PAGE gels as was observed previously during the titration of TmE-SC with sfGFP-ST (**Figure 2.9A**). To verify that the high molecular weight band indeed arose from the formation of a covalent bond between TmE-SC and DHFR-NST, a mutation was introduced into the SpyTag sequence which has been shown previously to abolish the capacity of the SpyTag to form isopeptide bonds (30). Specifically, the SpyTag's sole aspartate residue, whose sidechain γ-carbonyl is the target of nucleophilic attack during isopeptide bond formation with the SpyCatcher domain, was converted into an alanine. Repeating the titration experiment with the modified DHFR variant, referred to as DHFR-NST(DA), resulted in a complete lack of bioconjugate formation, confirming that the high molecular weight protein is a product of isopeptide bond formation between the SpyCatcher and SpyTag elements (**Figure**

2.9B). Likewise, TEM images collected for TmE-SC alone and TmE-SC mixed with DHFR-NST(DA) appear visually identical while images collected for TmE-SC:DHFR-NST bioconjugates show a similar reduction in clarity around the surface of the nanocontainers akin to the TEM images for TmE-SC:sfGFP-ST bioconjugates, though to a seemingly lesser extent (**Figure 2.9C**).

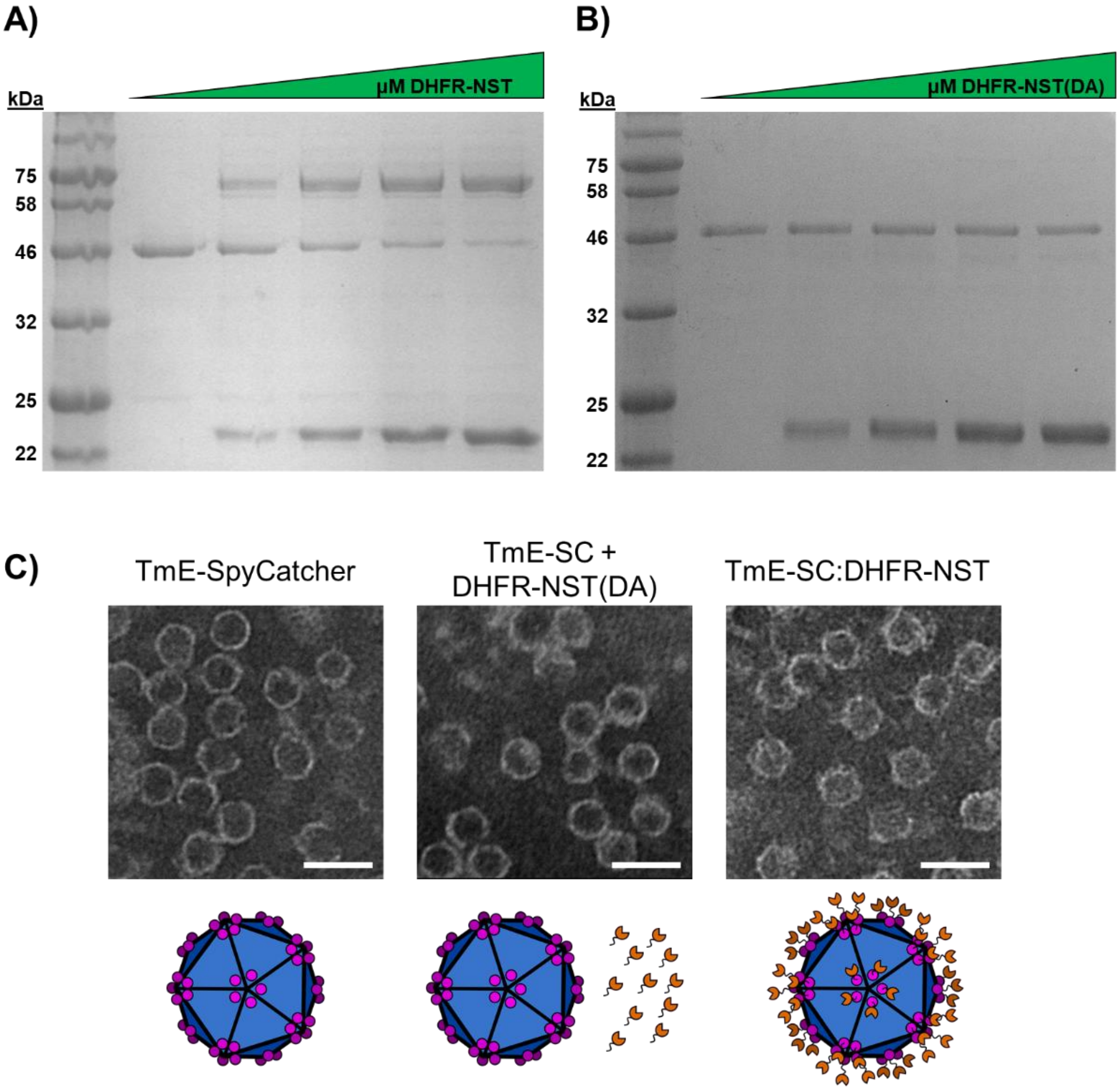


Figure 2.9 – *In vitro* titrations of a static concentration of TmE-SC with variable concentrations of A) DHFR-NST or B) the DHFR-NST(DA) non-binding variant. C) TEM images collected for TmE-SC, TmE-SC mixed with excess DHFR-NST(DA), or TmE-SC with bioconjugated DHFR-NST. Scale bars = 40 nm.

	Variant	K_m (μM)	k_{cat} (s^{-1})	k_{cat}/K_m ($\mu M^{-1}s^{-1}$)	Fold Change
	DHFR-NST	0.87 ± 0.04	4.17 + 0.02	4.79	--
	DHFR-M42W (i.e. DHFR*)	6.06 + 0.67	4.67 + 0.14	0.77	--
N-terminal SpyTag	DHFR*-NST	2.96 ± 0.33	3.98 ± 0.09	1.35	--
	MBP-SC immobilized DHFR*	3.95 ± 0.76	1.77 ± 0.08	0.45	↓3.0
	TmE-SC immobilized DHFR*	13.13 ± 1.73	4.49 ± 0.21	0.34	↓4.0
	Thrombin-cleaved TmE-SC/DHFR*	4.74 ± 0.65	3.93 ± 0.14	0.83	↓1.6
C-terminal SpyTag	DHFR*-CST	6.83 ± 0.83	2.79 ± 0.10	0.41	--
	MBP-SC immobilized DHFR*	6.25 ± 1.18	3.00 ± 0.16	0.48	↑1.2
	TmE-SC immobilized DHFR*	19.74 ± 2.57	2.27 ± 0.12	0.12	↓3.4

Table 2.2 – Michaelis-Menten kinetics values collected for all DHFR variants. The "Fold Change" values for immobilized DHFR*-NST and DHFR*-CST variants refer to the fold change relative to either free DHFR*-NST or DHFR*-CST in solution, respectively

Initial Michaelis-Menten kinetics tests of DHFR enzymatic activity monitoring the conversion of dihydrofolate (DHF) to tetrahydrofolate (THF) show that the presence of the SpyTag sequence on DHFR has no detrimental effect on the K_m for DHF, but the overall k_{cat} is reduced by approximately 3 to 5-fold relative to previously reported literature values (**Table 2.2**) (38-40). Following this initial kinetic assessment, we noticed that the enzyme's inherently low K_m for DHF (< 1 μM) necessitated that we conduct our kinetics experiments near the limit of detection for our spectrophotometric assay. In order to obtain more reliable and reproducible data, we introduced a single point mutation distal to the DHFR active site at residue 42, converting the native methionine to a tryptophan. This M42W variant of DHFR has been characterized previously and has been shown to possess a higher K_m value for DHF (39, 41-43). Michaelis-Menten kinetics tests with our DHFR(M42W)-SpyTag variant (herein referred to as DHFR*-NST) showed that the K_m value indeed increased approximately 3-fold to a value of 2.96 ± 0.33 μM. At this point, we also decided to generate an additional DHFR(M42W) variant with the SpyTag ligand fused onto the enzyme's C-terminus in order to probe whether the orientation of the biocatalyst would influence its catalytic performance upon immobilization onto encapsulin surfaces. This C-terminal variant (DHFR*-CST) was also kinetically characterized, exhibiting a K_m value of 6.83 ± 0.83 μM, which represents an approximate 7-fold increase over the initial DHFR-NST. Interestingly, the k_{cat} values for DHFR*-NST and DHFR*-CST remained relatively unchanged with the addition of the M42W mutation (**Table 2.2**). Collectively, the unchanged k_{cat} values coupled with the slightly increased K_m values for the M42W variants were sufficient for higher fidelity spectroscopic analyses.

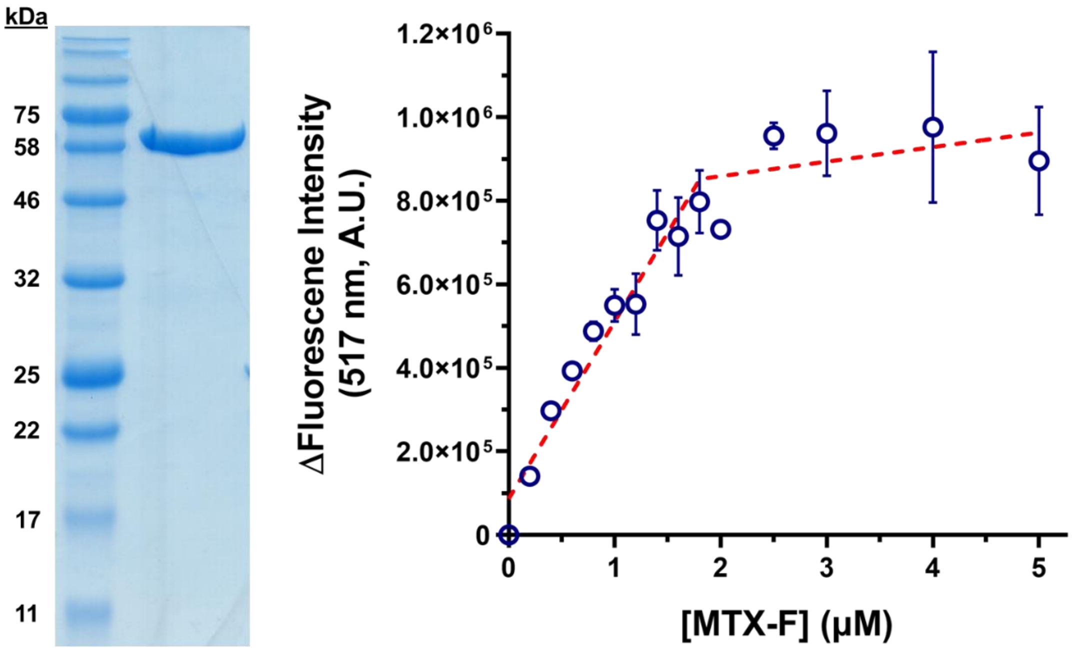


Figure 2.10 – SDS-PAGE image depicting a highly pure sample of TmE-SC:DHFR*-CST bioconjugate. A diluted sample of bioconjugate was titrated with increasing concentrations of MTX-F to obtain the binding curve shown on the right.

We next sought to functionalize TmE-SC nanocontainers with our DHFR*-ST variants and reassess their respective catalytic performances. However, accurately analyzing the kinetic profiles of encapsulins presenting surface-immobilized DHFR* necessitated that the amount of bound, functional enzyme in each bioconjugate sample be quantitatively determined. To this end, we found that co-expression of both the TmE-SC and the DHFR*-ST variants in *E. coli* cells led to the production of nanocontainers with seemingly complete occupancy of the available SpyCatcher domains due to higher *in vivo* overexpression of the DHFR*-ST variants relative to TmE-SC when the genes for both proteins were induced from T7 promoters using IPTG. Fully-decorated nanocontainers could subsequently be recovered in high purity from co-expression cultures, permitting the concentration of surface bound DHFR* to be determined spectrophotometrically using the combined theoretical extinction coefficients for the TmE-SC and DHFR*-ST variants since both proteins are present in a 1:1 molar ratio. However, since the spectrophotometric data cannot distinguish between functional and non-functional immobilized enzymes, we also performed fluorescence-based titrations of our protein samples using a fluorescein-labeled derivative of methotrexate (MTX-F), which is a well-established and potent competitive inhibitor of DHFR, whose fluorescein moiety experiences a 3 to 4-fold increase in emission intensity upon active site binding (44, 45). **Figure 2.10** shows a purified sample of TmE-SC:DHFR*-CST whose concentration was determined to be 4.6 μM from its 280 nm absorbance signal. Subsequently, a portion of the protein sample was diluted to a final concentration of 1.5 μM and was then titrated with MTX-F ranging from 0.2 to 5.0 μM. Segmental regression analysis of the resulting fluorescence titration data yielded an intersection of the two linear trendlines at 1.9 ± 0.3 μM, which corresponds to a back-calculated concentration of 5.5 ± 0.9 μM for the original protein stock. Comparatively, the titration data is in reasonably good agreement with the absorbance data and suggests that all of the encapsulin-bound DHFR* enzymes possess solvent accessible, functional active sites. Additional control reactions confirmed that neither the encapsulins nor the SpyCatcher domains themselves interfere with MTX-F binding during titration analyses. However, given that the spectrophotometric analysis is significantly faster and less error-prone than the MTX-F titration method, we elected to characterize all of the decorated nanocontainers in this manner going forward.

Kinetics data for both DHFR*-ST variants show that immobilization onto TmE surfaces via the SpyCatcher domains has no appreciable effect on the enzymes' k_{cat} values, though 4.4-fold and 2.9-fold increases in K_m for DHF were observed for the DHFR*-NST and DHFR*-CST enzymes, respectively (**Table 2.2**). To determine whether these decreases in substrate binding affinity resulted from immobilization of the DHFR* variants onto encapsulins themselves or whether the effect resulted from isopeptide bond formation with another protein in general, the

SpyCatcher domain was genetically fused onto the C-terminus of *E. coli* maltose binding protein (MBP) to generate the MBP-SpyCatcher hybrid (MBP-SC) as a generic covalent binding partner. Subsequent kinetics tests with MBP-SC:DHFR*-ST conjugates showed no significant increase in K_m for either DHFR* variant, indicating that the decrease in DHF binding affinity upon encapsulin immobilization is a phenomenon that is likely specific to the nanocontainer assemblies. Additionally, the DHFR*-NST variant possesses an encoded thrombin cleavage site interspersed between the SpyTag sequence and the beginning of the DHFR* gene. Encapsulins decorated with DHFR*-NST were kinetically assessed (**Table 2.2**) and were then treated with bovine thrombin to liberate the bound DHFR* enzymes (**Figure 2.11**). The cleaved DHFR* enzymes were then isolated via SEC and their kinetic parameters were reassessed for comparison (**Table 2.2**). The resulting data show that thrombolytic cleavage of DHFR*-NST from the encapsulin surface alleviates the observed decrease in binding affinity for DHF, further indicating that the K_m alterations are specific to encapsulin immobilization. We speculate that the decreased binding affinities likely arise from either steric crowding of the surface-bound DHFR* enzymes, localized increases in solvent viscosity near the encapsulin surface due to extensive ordering of solvent water molecules (though this effect is more likely to manifest as a decrease in k_{cat} rather than K_m), or due to some percentage of the enzyme population adopting orientations that occlude access to the active site as a result of the flexible nature of the linker sequences connecting the TmE and SpyCatcher domains (46, 47). However, given that the observed K_m increases for DHF are relatively minor, we did not investigate this effect further.

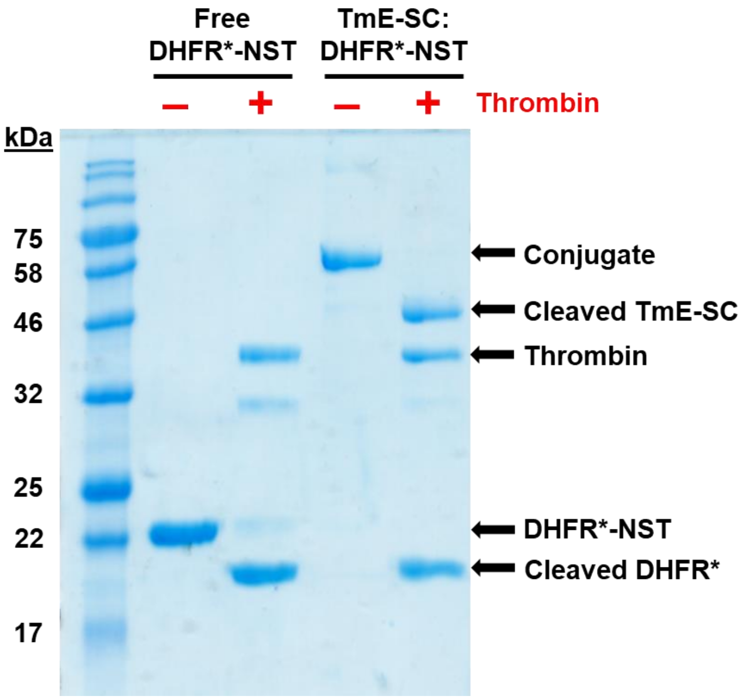
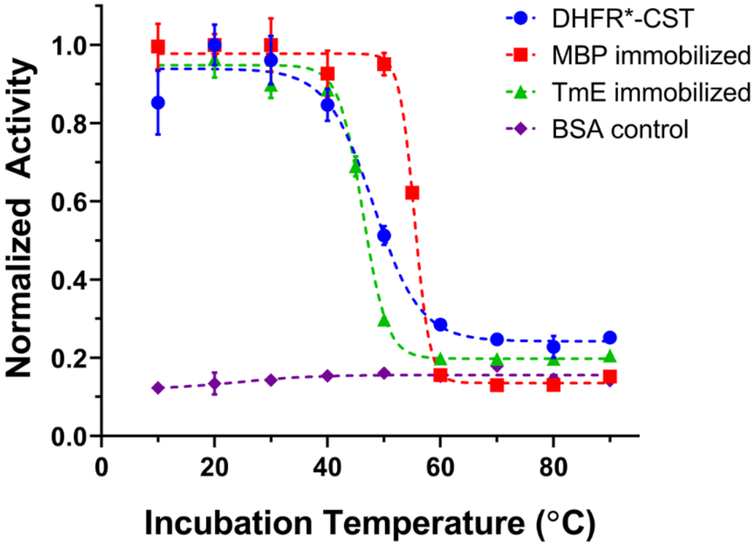


Figure 2.11 – SDS-PAGE analysis of thrombin cleavage tests. The N-terminal thrombin cleavage sites remains accessible in TmE-SC:DHFR*-NST samples, as determined by the disappearance of the initial conjugate protein band and the formation of a new band corresponding to the size of the cleaved DHFR* formed in a parallel control digestion of free DHFR*-NST.

Having shown that the enzymatic activities of both DHFR*-NST and DHFR*-CST are largely unaffected by surface immobilization, we subsequently investigated whether tethering DHFR* to encapsulins conferred any observable effects with respect to enzyme stability. To this end, the enzymatic activities of purified free DHFR*-CST, DHFR*-CST covalently conjugated to MBP-SC, and DHFR*-CST conjugated to TmE-SC were assessed following brief incubations at increasing temperatures. DHFR*-CST was assayed in particular as this variant was subsequently used for the generation of encapsulin-scaffolded multienzyme nanoreactors (discussed in Chapter 3). The resulting activity profiles (**Figure 2.12**) show that untethered DHFR*-CST exhibits an apparent melting temperature (T_m) value of 48°C. Covalent fusion of DHFR*-CST to MBP-SC increases the T_m value by approximately 7°C to 55°C, which is well within the previously reported range of T_m values determined for the fairly thermostable MBP (48, 49). Surprisingly, however, covalent fusion of DHFR*-CST to the hyperthermophilic TmE nanocontainers yielded no such stabilizing benefits as the T_m observed value is approximately equal to that of the free enzyme in solution.



Sample	T_m (°C)
DHFR*-CST	48.1 ± 1.3
MBP-immobilized	55.4 ± 0.2
TmE-immobilized	46.2 ± 0.4

Figure 2.12 – Thermal stability of free DHFR*-CST, MBP-immobilized enzyme, and TmE-immobilized enzyme reported as retention of enzymatic activity following 30 minutes of incubation at each temperature. Bovine serum albumin (BSA) was also assessed as a control. The T_m values depicted in the right-hand chart were obtained by applying a Boltzmann fit to the thermal stability data.

2.3 – Conclusion

Our primary goal in this study was to introduce either non-covalent or covalent peptide binding domains into the *T. maritima* encapsulin monomeric subunit in order to generate engineered encapsulin nanocontainers whose exterior surfaces could be functionalized for biocatalytic applications via non-native protein-protein interactions. In our initial experiments, we successfully grafted the PDZ and SH3 non-covalent peptide binding domains into solvent-exposed flexible loops in the TmE monomer sequence located between residues 59/60, residues 124/125, and residues 135/136. Subsequent IMAC purification and morphological characterizations using TEM and DLS showed that wildtype-sized assembled nanocontainers could be isolated for all of the variants, indicating that the incorporation of either PBD at these loop positions was well tolerated by the TmEs. However, *in vitro* mixing experiments showed that association of the PDZ- or SH3-functionalized TmEs with corresponding nanocontainers presenting the respective PDZ or SH3 ligand sequences generated amorphous aggregates of varying sizes rather than the desired ordered assemblies. Previous works involving the formation of ordered supramolecular arrays using VLPs and other protein-based nanocontainers have shown that careful balancing of buffer conditions, particularly solution pH and ionic strength, are often required to bring nanocontainers into weakly attractive regimes where thermodynamic assembly is favored over disordered kinetically-driven associations (50-53). Attempts to modulate buffer conditions in a similar manner were unsuccessful with our TmE-PBD and TmE-PL variants, however, as aggregates were continuously detected in optimization trials. As an alternative to the presented research design, several recent studies have shown that proteinaceous nanocontainers, particularly those being assembled using programmed protein-protein interactions similar to the PDZ and SH3 domains used here, could be assembled in a hierarchical fashion using a "layer-by-layer" technique in which iterative strata of nanocontainers presenting opposing surface binding elements are constructed in a bottom-up manner from a starting solid support (54, 55). Such a strategy might be more amenable to our TmE-PBD and TmE-PL variants as the solid support could limit the growth of container supramolecular lattices to two dimensions using a single component at a time rather than attempting to construct assemblies three dimensionally in solvent containing both TmE variants simultaneously.

As an alternative to the non-covalent peptide binding domains, we transitioned to the usage of the covalent SpyCatcher/SpyTag system to functionalize the exterior surfaces of our encapsulin nanocontainers, thus using the TmE shell as a scaffold for the dense display of active biocatalysts. As with the TmE-PBDs, the stability and plasticity of the encapsulin shells was exemplified as fusion of the SpyCatcher domain onto the encapsulin monomer sequence was well tolerated, yielding assembled nanocontainers shown to adopt wildtype-like morphologies via TEM. Further, we have shown that the SpyCatcher domain permits both *in vitro* and *in vivo* decoration of TmEs, with TmEs decorated *in vivo* achieving complete occupancy of all 60 SpyCatcher domains present per container. DHFR* variants tethered to encapsulin surfaces were largely unaffected by isopeptide immobilization, displaying only modest increases in K_m that could subsequently be alleviated following proteolytic liberation from the nanocontainer surfaces. While these kinetics results are promising, DHFR is a relatively small, monomeric enzyme that tolerated the dense surface display well. The same may not be true for larger and/or multimeric proteins, thus additional engineering may be needed to address such added complexities in future scaffolding designs. Several methods to address this issue could include the redistribution of the SpyCatcher domains across the encapsulin surface, generating linker sequences between the TmE surface and the covalently fused SpyCatcher domain (or between the SpyTag sequence and the protein of interest) of sufficient length and stability to increase the radial spacing between surface-tethered proteins, or the co-expression of TmE-SC and wildtype TmE monomers *in vivo* to achieve assembled nanocontainers with reduced surface valency.

Our final analytical test in which we assessed the thermal stability of bioconjugated DHFR* variants yielded perhaps the most surprising results as DHFR*s exhibited increased stability when tethered to monomeric MBP-SC fusion proteins, but not when immobilized on the surface of hyperthermophilic TmE-SC nanocontainers. This result could perhaps be explained by the nature of the thermostable behaviors intrinsic to MBP and encapsulins, respectively. Namely, whereas encapsulins containers are known to exhibit thermal tolerance in their assembled state, MBP exhibits thermostable behavior as a singular protein monomer with reported T_m values exceeding 60°C under certain buffer conditions (48, 49). To the best of our knowledge, no group has previously tested the inherent thermal stability of non-assembled encapsulin monomers, presumably because encapsulins assemble readily *in vivo* and have only been purified as whole nanocontainer structures. As such, it is feasible that the thermal tolerance of encapsulins is an ensemble property dictated by the intersubunit protein-protein interactions of the assembled shells rather than a native property of the individual monomers, and thus enzymes covalently fused to TmE surfaces may not experience associative increases in thermal tolerance themselves.

2.4 – Materials and Methods

2.4.1 – Reagents and Materials

Fluroescein methotrexate triammonium salt was purchased from Biotium (Fremont, CA). All other chemicals and reagents were purchased from Sigma Aldrich (St. Louis, MO). Carbon film 200 mesh copper electron microscopy grids were purchased from Electron Microscopy Sciences (Hatfield, PA). All pre-packed chromatography columns were purchased from GE Healthcare (Marlborough, MA). Free TALON affinity resin was purchased from Genesee Scientific (San Diego, CA). Free amylose affinity resin was purchased from New England Biolabs (Ipswich, MA). Electrocompetent *E. coli* DH5α cells were purchased from ThermoFisher

Scientific (Waltham, MA) while chemically competent *E. coli* BL21(DE3) cells were acquired from New England Biolabs (Ipswich, MA). All commercial enzymes used for molecular cloning were purchased from New England Biolabs unless otherwise noted. Bovine thrombin was purchased from Millipore Sigma (Burlington, MA). All DNA primers were ordered from Integrated DNA Technologies (Coralville, IA). Isolated genomic DNA from *Thermotoga maritima* was acquired from ATCC (Manassas, VA). The pJD757 vector containing the genes encoding the PDZ and SH3 peptide binding domains was a generous gift from Dr. John Dueber (University of California, Berkeley, CA). The pD441-SR:H6_sfGFP_SC plasmid containing both of the wildtype gene sequences encoding superfolder GFP (sfGFP) and SpyCatcher was a generous gift from Dr. Vincent Conticello (Emory University, Atlanta, GA).

2.4.2 – Molecular Cloning

The wildtype TmE gene was PCR amplified from isolated *T. maritima* genomic DNA (ATCC®43589™) using primers encoding NdeI and HindIII restriction sites within the forward and reverse primers, respectively. The amplified gene was digested with NdeI and HindIII restriction enzymes and was then ligated into linear pET23b vector that had been digested with the same enzymes using T4 DNA ligase. The final pET23b:TmE vector assembly was confirmed via DNA sequencing. TmE variants containing embedded PDZ of SH3 peptide binding domains were generated by first PCR amplifying the respective PBD from the pJD757 vector using terminal primers encoding ~15 to 20 bp of sequence homology to the region of insertion within the TmE gene. Subsequently, the PCR-amplified PBDs were used in two sequential overlap extension PCR reactions to add the 5’ end of the TmE gene onto the amplified PBD, followed by the 3’ end. The full fusion genes were then digested with NdeI and HindIII restriction enzymes, ligated into linearized pET23b vector using T4 DNA ligase, and confirmed via DNA sequencing. TmE variants containing the PDZ or SH3 ligand sequences as C-terminal fusions were generated in a series of iterative PCR reactions in which a common forward primer was used in conjunction with a collection of reverse primers containing successive portions of the respective ligand sequence. The final amplified fusion genes were digested with NdeI and HindIII restriction enzymes, ligated into linearized pET23b using T4 DNA ligase, and confirmed via DNA sequencing. The pET23b:TmE-SC vector was created by first amplifying the SpyCatcher gene and an upstream flexible linker (-GSGGGTGGGSGGGTS-) from the pD441-SR:H6_sfGFP-SC vector using primers encoding flanking regions at both ends designed to have sequence complementarity to the 3’ end of the TmE gene and a section of the pET23b vector downstream from the TmE gene. This amplified PCR product was isolated and used as a megaprimer to perform whole-plasmid amplification of the pET23b:TmE vector, thereby inserting the flexible linker and the SpyCatcher domain into the vector sequence at the C-terminal end of the TmE gene. The final plasmid product was confirmed via DNA sequencing.

The sfGFP gene was initially amplified from the pD441-SR:H6_sfGFP-SC vector using a series of iterative PCR reactions. Specifically, a common reverse primer containing an encoded XhoI restriction site was used in conjunction with a series of forward primers designed to sequentially construct the full SpyTag gene sequence, a hexahistidine affinity tag, and a NcoI restriction site on the 5’ end of the sfGFP gene. The final PCR product was digested with NcoI and XhoI, and was then inserted into pCDFDuet-1 vector that had been linearized with the same restriction enzymes in order to generate the pCDF:sfGFP-ST plasmid. The final plasmid product was confirmed via DNA sequencing.

The *E. coli* DHFR gene was PCR amplified from isolated genomic DNA using primers encoding NdeI and SpeI restriction sites at the 5’ and 3’ ends of the amplified gene, respectively. Prior to the current work, DNA sequences encoding the SpyTag ligand and a short flexible linker (-GSGSS-) had been inserted into the pET14b vector (Novagen) directly downstream from the native NcoI site, putting the SpyTag ligand in the same reading frame as the plasmid’s N-terminal hexahistidine affinity tag with the flexible linker joining the two tag sequences (unpublished work). The pET14b vector had also been previously modified to contain a non-native SpeI restriction site directly upstream of the plasmid’s native XhoI site. The amplified DHFR gene and the pET14b-SpyTag vector were digested in tandem with NdeI and SpeI restriction endonucleases, then the two ligated DNA products were covalently joined using T4 DNA ligase to form the pET14b:DHFR-NST vector. Successful incorporation of the DHFR gene was confirmed via DNA sequencing. The M42W variant of DHFR was produced using site-directed mutagenesis primers in conjunction with whole-plasmid amplification. Template pET14b:DHFR-ST plasmid was removed via DpnI digestion, and then the linear mutant DNA was transformed directly into DH5α chemically competent cells for plasmid propagation and subsequent isolation. Successful incorporation of the M42W mutation was confirmed via DNA sequencing. The final plasmid was deemed pET14b:DHFR*-NST.

For *in vivo* co-expression with the pET23b:TmE-SC plasmid, the DHFR*-NST gene was moved to the pCDFDuet-1 vector to avoid issues arising from the use of vectors containing identical antibiotic resistances and origins of replication. Consequently, the DHFR*-NST gene was excised from pET14b by digestion with NcoI and XhoI restriction enzymes and was then ligated into pCDFDuet-1 vector that had been digested with the same enzymes. Subsequently, the pCDF:DHFR*-CST variant was generated in a similar manner to the sfGFP-ST gene by sequentially PCR amplifying the DHFR* gene several times with a common forward primer containing a NcoI restriction site and a series of reverse primers designed to successively add a hexahistidine affinity tag, the SpyTag sequence, and a terminal XhoI restriction site to the 3' end of the amplified gene. The final DHFR*-CST gene product was digested with NcoI and XhoI and then ligated into fresh pCDFDuet-1 vector that had been digested with the same enzymes. All constructs were verified via DNA sequencing.

To generate the maltose binding protein-SpyCatcher (MBP-SC) hybrid gene, the SpyCatcher domain was first PCR amplified from the pD441-SR:H6_sfGFP-SC vector using primers encoding NdeI and XbaI restriction sites at the 5' and 3' ends of the amplified gene, respectively. The amplified gene was subsequently digested with NdeI and XbaI, and it was then inserted into linear pMAL-c2x vector that had been digested with the same enzymes to form the pMAL-c2x:SC vector. The final plasmid product was confirmed via DNA sequencing.

2.4.3 – Protein Overexpression and Purification

All recombinant proteins were overexpressed in BL21(DE3) *E. coli* cells. For wildtype TmE, TmE-PDZ variants, TmE-SH3 variants, and TmE-SpyCatcher variants, *E. coli* cells from a single colony containing the corresponding pET23b:TmE vectors were inoculated into LB medium and grown at 37°C in the presence of 0.1 mg/mL ampicillin to an OD_{600} value between 0.5 – 0.7. Protein overexpression was induced with the addition of a final concentration of 0.3 mM isopropyl β-D-thiogalactopyranoside (IPTG). Cells were harvested via centrifugation following 18 hours of expression time at 37°C. For DHFR variants, cells containing the respective pCDF:DHFR vector were grown in LB medium at 37°C in the presence of 0.05 mg/mL streptomycin to an OD_{600} value between 0.5 – 0.7, and then protein expression was induced with the addition of a final concentration of 0.3 mM IPTG. The cells were harvested via centrifugation after 16 hours of expression time at 30˚C. For sfGFP-SpyTag, cells containing the pCDF:sfGFP-ST vector were grown in LB medium at 37°C in the presence of 0.05 mg/mL streptomycin to an OD_{600} value between 0.5 – 0.7, and then protein expression was induced with a final concentration of 0.5 mM IPTG. The cells were harvested via centrifugation after 18 hours of expression time at 30°C. For MBP-SC, cells containing the pMAL-c2x:SC vector were grown in LB medium at 37°C in the presence of 0.1 mg/mL ampicillin and 10 mM glucose to an OD_{600} value between 0.5 – 0.7. Protein overexpression was induced with the addition of a final concentration of 0.3 mM IPTG. The cells were harvested via centrifugation after 6 hours of expression time at 30°C.

Wildtype TmE lacking a C-terminal hexahistidine affinity tag was purified according to previously established procedures (56). IMAC affinity chromatography was abandoned for purification of TmE-SC nanocontainers decorated with DHFR-SpyTag variants due to recent literature reports showing that hexhistidine affinity tags fused to TmE have the potential to promote non-specific association of proteins with TmE surfaces (34). All TmE-SC nanocontainers decorated with DHFR-SpyTag variants via *in vivo* co-expression were subsequently purified in an identical manner. Cell pellets were resuspended on ice in 50 mM Tris-HCl (pH 7.5) containing Protease Inhibitor Cocktail (Sigma Aldrich, P8849), 0.1 μg/mL PMSF, and 5 μg/mL DNaseI and then the cells were lysed by sonication (10 second pulses, 20 seconds rest between pulses) for 3.5 minutes. Insoluble debris was removed via centrifugation at 4000 rpms for 40 minutes at 4°C. The clarified supernatant was then loaded into a HiTrap Q-FF anion exchange column (pre-equilibrated with 5 column volumes of lysis buffer) and the column flow-through was collected. Solid ammonium sulfate was added to the flow-through sample to reach a final concentration of 20%

(w/v), and then the sample was incubated for 20 minutes at 4°C on an orbital rocking table to promote precipitation of unwanted contaminant proteins. Following the incubation period, the sample tube was centrifuged at 4000 rpms for 20 minutes at 4°C to isolate precipitated contaminant proteins. The resulting supernatant was decanted into a fresh Falcon tube, and additional solid ammonium sulfate was added to reach a final concentration of 40% (w/v). The same 4°C incubation and centrifugation steps were repeated to isolated precipitated nanocontainers. The resulting solid pellet was resuspended in 5 to 10 mL of 50 mM Tris-HCl (pH 7.5) and was immediately loaded into a cellulose dialysis bag (14 kDa MWCO). The dialysis bag was immersed sequentially into two beakers containing 1 L of 50 mM HEPES-KOH (pH 7.5), 100 mM NaCl each with a minimum of 4 hours dialysis at 4°C with gentle stirring in each beaker to remove residual ammonium sulfate. The dialyzed sample was then recovered, passed through a 0.2 μm syringe filter, and then loaded into a HiPrep 16/60 Sephacryl S-500 HR size exclusion column pre-equilibrated in 50 mM HEPES-KOH (pH 7.5), 100 mM NaCl and connected to an NGC Chromatography FPLC system (Bio-Rad, Hercules, CA). The same buffer was used as the mobile phase during this second chromatography step. Assembled encapsulins were isolated from the size exclusion column and stored at 4°C until use in experiments. Protein concentrations for isolated TmE, TmE-SC and TmEΔ9Gly2-SC were determined spectrophotometrically at 280 nm using theoretical extinction coefficient values: 36,440 $M^{-1}cm^{-1}$, 47,900 $M^{-1}cm^{-1}$, and 46,410 $M^{-1}cm^{-1}$, respectively. For decorated TmEs, protein concentrations were obtained by using the combined extinction coefficients for the nanocontainer and DHFR-SpyTag variant used.

For IMAC purifications of sfGFP-SpyTag, all TmE-PBD variants, and both TmE-PL variants, cell pellets were resuspended in 50 mM Tris-HCl (pH 7.5), 150 mM NaCl, 10 mM Imidazole and then the cells were lysed via sonication in the same manner as described above for the TmE-SC variants. Insoluble debris was removed via centrifugation at 4000 rpms for 40 minutes at 4°C, then the clarified lysate was loaded into a 30 mL polypropylene disposable column containing 1 mL of TALON affinity resin (pre-equilibrated with 5 column volumes of lysis buffer). The column was sealed and placed on a rocking table at 4˚C for 20 minutes. The column was then unsealed and the flow-through was collected. Subsequently, the resin bed was washed with 15 column volumes of the resuspension buffer, then bound protein was eluted by washing the resin bed with 4 column volumes of 50 mM Tris-HCl (pH 7.5), 150 mM NaCl, 250 mM Imidazole. The eluted protein was then loaded into a cellulose dialysis bag (14 kDa MWCO) and dialyzed overnight at 4°C into 50 mM Tris-HCl (pH 7.5), 150 mM NaCl to remove excess imidazole. For sfGFP-ST, the final concentration of the isolated protein was determined via the absorbance of the native fluorophore at 485 nm using the previously reported extinction coefficient value of 8.33 x 10^4 $M^{-1}cm$-^{1} (37). Purified sfGFP-SpyTag was kept at 4°C until use in experiments. For TmE-PBD and TmE-PL variants, proteins concentrations were determined via their respective 280 nm absorbance signals using the following theoretical extinction coefficients: $TmE_{59/124/135}$-PDZ = 39,545 $M^{-1}cm^{-1}$; TmE-PDZlig = 36,565 $M^{-1}cm^{-1}$; $TmE_{59/124/135}$-SH3 = 52,035 $M^{-1}cm^{-1}$; TmE-SH3lig = 36,565 $M^{-1}cm^{-1}$.

Purification of DHFR-NST and DHFR-CST variants was performed following the same TALON resin purification protocol as for sfGFP-SpyTag with the exception that the lysis buffer consisted of 50 mM Tris-HCl (pH 7.5), 150 mM NaCl, 2.5 mM β-mercaptoethanol, 10 mM Imidazole. Bound enzyme was eluted from the resin using the same buffer solution with the final concentration of imidazole increased to 250 mM. The eluted protein was then loaded into a cellulose dialysis bag (14 kDa MWCO) and dialyzed overnight at 4°C into 50 mM K-Phosphate (pH 7.0), 100 mM NaCl, 2.5 mM β-mercaptoethanol. Purified DHFR-SpyTag was flash frozen in liquid nitrogen and stored at -80°C until use in experiments. Protein concentrations for DHFR-NST, DHFR*-NST, and DHFR*-CST were determined spectrophotometrically at 280 nm using theoretical extinction coefficient values: 34,950 $M^{-1}cm^{-1}$ for DHFR-NST and 40,450 $M^{-1}cm^{-1}$ for both variants containing the M42W mutation.

Purification of MBP-SC was performed by first resuspending cell pellets in 50 mM Tris-HCl (pH 7.5), 500 mM NaCl containing Protease Inhibitor Cocktail, 0.1 μg/mL PMSF, and 5 μg/mL DNaseI. The cell solution was then lysed via sonication in the same manner as described above. Insoluble debris was removed via centrifugation at 4000 rpms for 40 minutes at 4°C, then the clarified lysate was loaded into a 30 mL polypropylene disposable column containing 1 mL of amylose affinity resin (pre-equilibrated with 5 column volumes of lysis buffer). The column was sealed and placed on a rocking table at 4˚C for 20 minutes. The column was then unsealed and the flow-through was collected. Subsequently, the resin bed was washed with 15 column volumes of the resuspension buffer, then bound MBP-SC was eluted by washing the resin bed with 4 column volumes of 50 mM Tris-HCl (pH 7.5), 500 mM NaCl, 10 mM maltose. Purified MBP-SC was flash frozen in liquid nitrogen and stored at -80°C until use in experiments. Protein concentrations for purified MPB-SC were determined spectrophotometrically at 280 nm using the theoretical extinction coefficient 77,810 $M^{-1}cm^{-1}$.

2.4.4 – Dynamic Light Scattering Analysis

In vitro mixtures of purified TmE nanocontainers presenting corresponding PBDs and PLs were mixed in 50 mM Tris-HCl (pH 7.5), 150 mM NaCl buffer in varying mass or concentration ratios. Final protein concentrations were kept to a maximum of no more than 10 μM for each TmE variant. Samples were incubated at 4°C overnight prior to being subjected to DLS assessment using a NanoPlus DLS Particle Analyzer (Micromeritics Instrument Corporation, Norcross, GA). Data collection was performed at 20°C, and all samples were passed through 0.2 μm syringe filters prior to assessment to remove any dust.

2.4.5 – *In vitro* SpyCatcher/SpyTag Conjugation Reactions

In vitro conjugations between purified proteins with fused SpyCatcher domains and partner proteins presenting fused SpyTags were performed by mixing a given final concentration of the SpyCatcher fusion protein with a minimum of a 5 μM excess of the SpyTag-fusion partner. All conjugation reactions were performed in 50 mM K-Phosphate (pH 7.0), 100 mM NaCl, 2.5 mM β-mercaptoethanol. Unless otherwise stated, conjugations were allowed to proceed overnight at 4°C prior to purification of the resulting bioconjugate. Purification of bioconjugates from unreacted starting proteins was performed by loading the conjugation reaction mixture into a Superdex 200 Increase 10/300 GL column that had been pre-equilibrated with 50 mM K-Phosphate (pH 7.5), 100 mM NaCl, 2.5 mM β-mercaptoethanol and was connected to an ÄKTA Explorer 10 FPLC system (GE Life Sciences, Little Chalfont, Buckinghamshire, UK). The same buffer was used as the mobile phase during purification chromatography, and protein elution was monitored spectrophotometrically at 280 nm. For samples containing sfGFP-ST, sample elution was also monitored at 485 nm to detect sfGFP chromophores. Purified bioconjugates were kept at 4°C until needed for experimentation.

2.4.6 – Fluorescein Methotrexate Titrations

Fluorescein methotrexate (MTX-F) titrations were performed by titrating purified sample proteins with known concentrations of MTX-F in 50 mM K-Phosphate, 0.1 M NaCl, 2.5 mM β-mercaptoethanol and monitoring the change in fluorescence intensity at 517 nM upon MTF-F binding. The range of MTX-F tested spanned from 0.2 – 5.0 μM. All fluorometric readings were collected using a HORIBA Jobin Yvon FluoroMax®-3 fluorimeter (Edison, NJ). Samples were excited at 496 nm and the resulting emission data was collected with both the excitation and emission slits of the instrument set at 2 nm and the data integration time set to 0.2 seconds. Segmented regression analysis of titration data was performed using GraphPad Prism 8.4.1 (La Jolla, CA).

2.4.7 – Michaelis-Menten Kinetics Analyses

Steady-state kinetics of free and bioconjugated DHFR variants were performed at 22°C by monitoring the depletion of NADPH spectrophotometrically at 340 nm on a Varian Cary 50 Bio UV-visible spectrophotometer (Santa Clara, CA). The reaction mixture used for all kinetics runs consisted of 100 mM K-Phosphate (pH 7.0), 100 mM NaCl, 1 mM dithiothreitol, 20 mM sodium ascorbate, and 0.1 mM NADPH. A final concentration of free DHFR* or bioconjugated DHFR* between 2.5 – 50 nM was added into the reaction mixture and allowed to equilibrate for at least 10 minutes to prevent hysteresis effects. Kinetics tests were initiated upon the addition of dihydrofolate (DHF) into the reaction mixture. A combined extinction coefficient for NADPH and

DHF of 11,800 $M^{-1}cm^{-1}$ (57) was used to calculate the rate of NADPH consumption during the course of each reaction. All reactions were conducted in triplicate with error values representing one standard deviation from the mean. Kinetics data were processed using GraphPad Prism 8.4.1 (La Jolla, CA).

2.4.8 – Thermal Stability Assays

Assessments of DHFR* thermal stability in relation to bioconjugation state were performed by incubating 2.5 μM stocks of enzyme in 100 mM K-Phosphate (pH 7.0), 100 mM NaCl, 2.5 mM β-mercaptoethanol within the heating blocks of a Bio-Rad C1000 Touch Thermal Cycler (Hercules, CA) at temperatures ranging from 10°C up to 90°C for 30 minutes. Immediately following each incubation, enzyme stocks were removed and used to prepare reaction mixtures containing 100 mM K-Phosphate (pH 7.0), 100 mM NaCl, 1 mM dithiothreitol, 20 mM sodium ascorbate, 0.1 mM NADPH, and a final concentration of 5 nM enzyme. Reactions were initiated upon the addition of a final concentration of 60 μM DHF and were monitored spectrophotometrically at 340 nm as described in Section 2.4.7 above for 20 minutes at 22°C. All reactions were conducted in triplicate with error values representing one standard deviation from the mean. Thermal activity plots were analyzed using GraphPad Prism 8.4.1 (La Jolla, CA).

2.4.9 – Transmission Electron Microscopy Imaging

Individual TEM grids were prepared by applying 4 μL of biological sample onto the carbon surface of carbon-copper grids for five minutes. Samples were diluted to a final protein concentration between 0.1 – 0.3 mg/mL to prevent overcrowding of proteins on the grid surface. After five minutes had elapsed, the edge of the grid was gently blotted against a sheet of Whatman 1 filter paper. The grid was then inverted and quickly immersed sequentially into two drops of deionized water. The edge of the grid was again blotted against filter paper, and then negative staining was accomplished by applying 4 μL of 1% (w/v) phosphotungstic acid (pH 6.5, prepared in double distilled water and pH adjusted with KOH) onto the surface of the grid and allowing it to incubate for 20 seconds. The edge of the grid was then blotted against filter paper one final time, and then the grids were allowed to air dry for five minutes before being placed into a vacuum desiccator for five additional minutes prior to imaging. TEM imaging was performed on a Hitachi HT7700 transmission electron microscope (Hitachi, Tokyo, Japan) operating at an accelerating voltage of 80.0 kV.

2.5 – References

1. Douglas, T.; Young, M., Viruses: making friends with old foes. *Science* **2006**, *312* (5775), 873-875.
2. Andrews, S. C., The Ferritin-like superfamily: Evolution of the biological iron storeman from a rubrerythrin-like ancestor. *Biochimica et Biophysica Acta* **2010**, *1800* (8), 691-705.
3. Yeates, T. O.; Crowley, C. S.; Tanaka, S., Bacterial microcompartment organelles: protein shell structure and evolution. *Annual Review of Biophysics* **2010**, *39*, 185-205.
4. Cornejo, E.; Abreu, N.; Komeili, A., Compartmentalization and organelle formation in bacteria. *Current Opinion in Cell Biology* **2014**, *26*, 132-138.

5. Bobik, T. A.; Lehman, B. P.; Yeates, T. O., Bacterial microcompartments: widespread prokaryotic organelles for isolation and optimization of metabolic pathways. *Molecular Microbiology* **2015**, *98* (2), 193-207.

6. Giessen, T. W.; Silver, P. A., Widespread distribution of encapsulin nanocompartments reveals functional diversity. *Nature Microbiology* **2017**, *2*, 17029.

7. Schwarz, B.; Uchida, M.; Douglas, T., Biomedical and Catalytic Opportunities of Virus-Like Particles in Nanotechnology. *Advances in Virus Research* **2017**, *97*, 1-60.

8. Aumiller, W. M.; Uchida, M.; Douglas, T., Protein cage assembly across multiple length scales. *Chemical Society Reviews* **2018**, *47* (10), 3433-3469.

9. Diaz, D.; Care, A.; Sunna, A., Bioengineering Strategies for Protein-Based Nanoparticles. *Genes* **2018**, *9* (7), 370.

10. Schoonen, L.; van Hest, J. C. M., Functionalization of protein-based nanocages for drug delivery applications. *Nanoscale* **2014**, *6* (13), 7124-7141.

11. Eiben, S.; Koch, C.; Altintoprak, K.; Southan, A.; Tovar, G.; Laschat, S.; Weiss, I. M.; Wege, C., Plant virus-based materials for biomedical applications: Trends and prospects. *Advanced Drug Delivery Reviews* **2019**, *145*, 96-118.

12. Sutter, M.; Boehringer, D.; Gutmann, S.; Guenther, S.; Prangishvili, D.; Loessner, M. J.; Stetter, K. O.; Weber-Ban, E.; Ban, N., Structural basis of enzyme encapsulation into a bacterial nanocompartment. *Nature Structural & Molecular Biology* **2008**, *15* (9), 939-947.

13. Akita, F.; Chong, K. T.; Tanaka, H.; Yamashita, E.; Miyazaki, N.; Nakaishi, Y.; Suzuki, M.; Namba, K.; Ono, Y.; Tsukihara, T.; Nakagawa, A., The Crystal Structure of a Virus-

like Particle from the Hyperthermophilic Archaeon Pyrococcus furiosus Provides Insight into the Evolution of Viruses. *Journal of Molecular Biology* **2007**, *368* (5), 1469-1483.

14. McHugh, C. A.; Fontana, J.; Nemecek, D.; Cheng, N. Q.; Aksyuk, A. A.; Heymann, J. B.; Winkler, D. C.; Lam, A. S.; Wall, J. S.; Steven, A. C.; Hoiczyk, E., A virus capsid-like nanocompartment that stores iron and protects bacteria from oxidative stress. *The EMBO Journal* **2014**, *33* (17), 1896-1911.

15. Giessen, T. W.; Orlando, B. J.; Verdegaal, A. A.; Chambers, M. G.; Gardener, J.; Bell, D. C.; Birrane, G.; Liao, M.; Silver, P. A., Large protein organelles form a new iron sequestration system with high storage capacity. *eLIFE* **2019**, *8*, e46070.

16. Nichols, R. J.; Cassidy-Amstutz, C.; Chaijarasphong, T.; Savage, D. F., Encapsulins: molecular biology of the shell. *Critical Reviews in Biochemistry and Molecular Biology* **2017**, 1-12.

17. Gilcrease, E. B.; Winn-Stapley, D. A.; Hewitt, F. C.; Joss, L.; Casjens, S. R., Nucleotide sequence of the head assembly gene cluster of bacteriophage L and decoration protein characterization. *Journal of Bacteriology* **2005**, *187* (6), 2050-2057.

18. Lander, G. C.; Evilevitch, A.; Jeembaeva, M.; Potter, C. S.; Carragher, B.; Johnson, J. E., Bacteriophage Lambda Stabilization by Auxiliary Protein gpD: Timing, Location, and Mechanism of Attachment Determined by Cryo-EM. *Structure* **2008**, *16* (9), 1399-1406.

19. Sathaliyawala, T.; Islam, M. Z.; Li, Q.; Fokine, A.; Rossmann, M. G.; Rao, V. B., Functional analysis of the highly antigenic outer capsid protein, Hoc, a virus decoration protein from T4-like bacteriophages. *Molecular Microbiology* **2010**, *77* (2), 444-455.

20. Qin, L.; Fokine, A.; O'Donnell, E.; Rao, V. B.; Rossmann, M. G., Structure of the small outer capsid protein, Soc: a clamp for stabilizing capsids of T4-like phages. *Journal of Molecular Biology* **2010**, *395* (4), 728-741.

21. Strable, E.; Prasuhn, D. E., Jr.; Udit, A. K.; Brown, S.; Link, A. J.; Ngo, J. T.; Lander, G.; Quispe, J.; Potter, C. S.; Carragher, B.; Tirrell, D. A.; Finn, M. G., Unnatural amino acid incorporation into virus-like particles. *Bioconjugate Chemistry* **2008**, *19* (4), 866-875.

22. Pokorski, J. K.; Steinmetz, N. F., The Art of Engineering Viral Nanoparticles. *Molecular Pharmaceutics* **2011**, *8* (1), 29-43.

23. Dueber, J. E.; Wu, G. C.; Malmirchegini, G. R.; Moon, T. S.; Petzold, C. J.; Ullal, A. V.; Prather, K. L. J.; Keasling, J. D., Synthetic protein scaffolds provide modular control over metabolic flux. *Nature Biotechnology* **2009**, *27* (8), 753-U107.

24. Moon, T. S.; Dueber, J. E.; Shiue, E.; Prather, K. L., Use of modular, synthetic scaffolds for improved production of glucaric acid in engineered E. coli. *Metabolic Engineering* **2010**, *12* (3), 298-305.

25. Schwarz, B.; Madden, P.; Avera, J.; Gordon, B.; Larson, K.; Miettinen, H. M.; Uchida, M.; LaFrance, B.; Basu, G.; Rynda-Apple, A.; Douglas, T., Symmetry Controlled, Genetic Presentation of Bioactive Proteins on the P22 Virus-like Particle Using an External Decoration Protein. *ACS Nano* **2015**, *9* (9), 9134-9147.

26. Giessen, T. W.; Silver, P., A catalytic nanoreactor based on in vivo encapsulation of multiple enzymes in an engineered protein nanocompartment. *ChemBioChem* **2016**, *17* (20), 1931-1935.

27. Cayetano-Cruz, M.; Coffeen, C. F.; Valadez-Garcia, J.; Montiel, C.; Bustos-Jaimes, I., Decoration of virus-like particles with an enzymatic activity of biomedical interest. *Virus Research* **2018**, *255*, 1-9.

28. Nguyen, J. T.; Turck, C. W.; Cohen, F. E.; Zuckermann, R. N.; Lim, W. A., Exploiting the basis of proline recognition by SH3 and WW domains: design of N-substituted inhibitors. *Science* **1998**, *282* (5396), 2088-2092.

29. Harris, B. Z.; Hillier, B. J.; Lim, W. A., Energetic determinants of internal motif recognition by PDZ domains. *Biochemistry* **2001**, *40* (20), 5921-5930.

30. Zakeri, B.; Fierer, J. O.; Celik, E.; Chittock, E. C.; Schwarz-Linek, U.; Moy, V. T.; Howarth, M., Peptide tag forming a rapid covalent bond to a protein, through engineering a bacterial adhesin. *Proceedings of the National Academy of Sciences* **2012**, *109* (12), E690-E697.

31. Schoene, C.; Fierer, J. O.; Bennett, S. P.; Howarth, M., SpyTag/SpyCatcher cyclization confers resilience to boiling on a mesophilic enzyme. *Angewandte Chemie International Edition* **2014**, *53* (24), 6101-6104.

32. Nguyen, P. Q.; Botyanszki, Z.; Tay, P. K. R.; Joshi, N. S., Programmable biofilm-based materials from engineered curli nanofibres. *Nature Communications* **2014**, *5* (1), 4945.

33. Tamura, A.; Fukutani, Y.; Takami, T.; Fujii, M.; Nakaguchi, Y.; Murakami, Y.; Noguchi, K.; Yohda, M.; Odaka, M., Packaging Guest Proteins into the Encapsulin Nanocompartment from Rhodococcus erythropolis N771. *Biotechnology and Bioengineering* **2015**, *112* (1), 13-20.

34. Cassidy-Amstutz, C.; Oltrogge, L.; Going, C. C.; Lee, A.; Teng, P.; Quintanilla, D.; East-Seletsky, A.; Williams, E. R.; Savage, D. F., Identification of a Minimal Peptide Tag for in Vivo and in Vitro Loading of Encapsulin. *Biochemistry* **2016**, *55* (24), 3461-3468.

35. Sigmund, F.; Massner, C.; Erdmann, P.; Stelzl, A.; Rolbieski, H.; Desai, M.; Bricault, S.; Worner, T. P.; Snijder, J.; Geerlof, A.; Fuchs, H.; Hrabe de Angelis, M.; Heck, A. J. R.; Jasanoff, A.; Ntziachristos, V.; Plitzko, J.; Westmeyer, G. G., Bacterial encapsulins as orthogonal compartments for mammalian cell engineering. *Nature Communications* **2018**, *9* (1), 1990.

36. Kim, H.; Choi, H.; Bae, Y.; Kang, S., Development of target-tunable P22 VLP-based delivery nanoplatforms using bacterial superglue. *Biotechnology and Bioengineering* **2019**, *116* (11), 2843-2851.

37. Pedelacq, J. D.; Cabantous, S.; Tran, T.; Terwilliger, T. C.; Waldo, G. S., Engineering and characterization of a superfolder green fluorescent protein. *Nature Biotechnology* **2006**, *24* (1), 79-88.

38. Shaw, D.; Odom, J. D.; Dunlap, R. B., High expression and steady-state kinetic characterization of methionine site-directed mutants of Escherichia coli methionyl- and selenomethionyl-dihydrofolate reductase. *Biochimica et Biophysica Acta* **1999**, *1429* (2), 401-410.

39. Ohmae, E.; Fukumizu, Y.; Iwakura, M.; Gekko, K., Effects of mutation at methionine-42 of Escherichia coli dihydrofolate reductase on stability and function: implication of hydrophobic interactions. *Journal of Biochemistry* **2005**, *137* (5), 643-652.

40. Loveridge, E. J.; Hroch, L.; Hughes, R. L.; Williams, T.; Davies, R. L.; Angelastro, A.; Luk, L. Y. P.; Maglia, G.; Allemann, R. K., Reduction of Folate by Dihydrofolate Reductase from Thermotoga maritima. *Biochemistry* **2017**, *56* (13), 1879-1886.

41. Rajagopalan, P. T.; Zhang, Z.; McCourt, L.; Dwyer, M.; Benkovic, S. J.; Hammes, G. G., Interaction of dihydrofolate reductase with methotrexate: ensemble and single-molecule kinetics. *Proceedings of the National Academy of Sciences* **2002**, *99* (21), 13481-13486.

42. Wang, L.; Goodey, N. M.; Benkovic, S. J.; Kohen, A., The role of enzyme dynamics and tunnelling in catalysing hydride transfer: studies of distal mutants of dihydrofolate reductase. *Philosophical Transactions of the Royal Society B Biological Sciences* **2006**, *361* (1472), 1307-1315.

43. Wang, L.; Goodey, N. M.; Benkovic, S. J.; Kohen, A., Coordinated effects of distal mutations on environmentally coupled tunneling in dihydrofolate reductase. *Proceedings of the National Academy of Sciences* **2006**, *103* (43), 15753.

44. Gapski, G. R.; Whiteley, J. M.; Rader, J. I.; Cramer, P. L.; Henderson, G. B.; Neef, V.; Huennekens, F. M., Synthesis of a fluorescent derivative of amethopterin. *Journal of Medicinal Chemistry* **1975**, *18* (5), 526-528.

45. Degan, P.; Carpano, P.; Cercignani, G.; Montagnoli, G., A fluorescence study of substrate and inhibitor binding to bovine liver dihydrofolate reductase. *The International Journal of Biochemistry* **1989**, *21* (3), 291-295.

46. Jasensky, J.; Ferguson, K.; Baria, M.; Zou, X.; McGinnis, R.; Kaneshiro, A.; Badieyan, S.; Wei, S.; Marsh, E. N. G.; Chen, Z., Simultaneous Observation of the Orientation and Activity of Surface-Immobilized Enzymes. *Langmuir* **2018**, *34* (31), 9133-9140.

47. Hoarau, M.; Badieyan, S.; Marsh, E. N. G., Immobilized enzymes: understanding enzyme – surface interactions at the molecular level. *Organic & Biomolecular Chemistry* **2017**, *15* (45), 9539-9551.

48. Novokhatny, V.; Ingham, K., Thermodynamics of maltose binding protein unfolding. *Protein Science: A Publication of the Protein Society* **1997**, *6* (1), 141-146.

49. Ganesh, C.; Shah, A. N.; Swaminathan, C. P.; Surolia, A.; Varadarajan, R., Thermodynamic characterization of the reversible, two-state unfolding of maltose binding protein, a large two-domain protein. *Biochemistry* **1997**, *36* (16), 5020-5028.

50. Kostiainen, M. A.; Hiekkataipale, P.; Laiho, A.; Lemieux, V.; Seitsonen, J.; Ruokolainen, J.; Ceci, P., Electrostatic assembly of binary nanoparticle superlattices using protein cages. *Nature Nanotechnology* **2012**, *8*, 52.

51. Liljeström, V.; Mikkilä, J.; Kostiainen, M. A., Self-assembly and modular functionalization of three-dimensional crystals from oppositely charged proteins. *Nature Communications* **2014**, *5*, 4445.

52. Asor, R.; Ben-Nun-Shaul, O.; Oppenheim, A.; Raviv, U., Crystallization, Reentrant Melting, and Resolubilization of Virus Nanoparticles. *ACS Nano* **2017**, *11* (10), 9814-9824.

53. Uchida, M.; McCoy, K.; Fukuto, M.; Yang, L.; Yoshimura, H.; Miettinen, H. M.; LaFrance, B.; Patterson, D. P.; Schwarz, B.; Karty, J. A.; Prevelige, P. E., Jr.; Lee, B.; Douglas, T., Modular Self-Assembly of Protein Cage Lattices for Multistep Catalysis. *ACS Nano* **2018**, *12* (2), 942-953.

54. Uchida, M.; LaFrance, B.; Broomell, C. C.; Prevelige, P. E., Jr.; Douglas, T., Higher order assembly of virus-like particles (VLPs) mediated by multi-valent protein linkers. *Small* **2015**, *11* (13), 1562-1570.

55. Choi, H.; Choi, B.; Kim, G. J.; Kim, H.-u.; Kim, H.; Jung, H. S.; Kang, S., Fabrication of Nanoreaction Clusters with Dual-Functionalized Protein Cage Nanobuilding Blocks. *Small* **2018**, *14* (35), 1801488.

56. Williams, E. M.; Jung, S. M.; Coffman, J. L.; Lutz, S., Pore Engineering for Enhanced Mass Transport in Encapsulin Nanocompartments. *ACS Synthetic Biology* **2018**, *7* (11), 2514-2517.

57. Stone, S. R.; Morrison, J. F., Kinetic mechanism of the reaction catalyzed by dihydrofolate reductase from Escherichia coli. *Biochemistry* **1982**, *21* (16), 3757-3765.

Chapter 3:

Development of a Bi-enzymatic Synthetic Metabolon Using Encapsulin-based Nanoreactors

3.1 – Introduction

In nature, both eukaryotic and prokaryotic organisms rely on multistep enzymatic cascade reactions to rapidly facilitate the numerous cellular processes necessary to sustain life and growth (1). Enzymatic cascades are inherently beneficial as they allow organisms to maintain lower net concentrations of metabolic intermediates, thus preventing the formation of futile byproducts or the release of toxic species into the larger cellular environment (2, 3). Further, metabolic cascades can often enhance the individual reactivities and selectivities of pathway enzymes through a process known as "substrate channeling" in which normal reaction equilibria are subverted through the direct transfer of intermediates between sequentially acting enzyme catalysts (2, 4). For efficient substrate channeling to occur, however, functionally related enzymes must typically be co-localized within macromolecular complexes or organelle structures to minimize the spatial distances chemical intermediates must traverse between active sites (5).

Chemists and synthetic biologists alike have devoted intense research efforts towards mimicking the channeling effects of natural systems using synthetically crafted multienzyme metabolons. Generally, synthetic enzyme complexes have been constructed using a number of related strategies including the generation of non-native fusion proteins (6, 7), using protein or DNA-based scaffolds as selective enzyme templates (8-13), the immobilization of enzymes onto solid supports (14-16), or by sequestering enzymes within lipid-bound micelles or proteinaceous cages (17-24). In recent decades, naturally occurring protein-based cage assemblies, including bacterial microcompartments, encapsulins, and virus-like particles, have been increasingly utilized for the development of artificial metabolons due to their uniform sizes, engineerability and their capacity to serve as both nanoscale containers and as functionalizable templates (25-27).

Accordingly, several exciting examples in the last few years have illustrated either the covalent scaffolding of multienzyme systems on nanocontainer surfaces for the generation of high sensitivity biosensors (28, 29), or the encapsulation of multiple enzymes within container luminal spaces for the biocatalytic production of biofuel precursors and commercially relevant fabric dyes (30-32).

Chapter 2 of this thesis details the polyvalent exterior surface decoration of *T. maritima* encapsulin nanocontainers with functional DHFR biocatalysts isolated from *E. coli*. The following chapter describes our efforts to expand upon this nanocontainer design in order to create a synthetic two-enzyme metabolon by coupling the enzymatic activity of DHFR to that of the THF-dependent aryl-*O*-demethylase enzyme LigM abstracted from the soil bacterium *Sphingomonas paucimobilis* SYK-6 (33, 34). Specifically, we seek to utilize both the encapsulin's native role as a biological container for protein cargoes alongside its engineered function as a high-density symmetrical template. We plan to achieve this feat by covalently tethering DHFR enzymes to the container exterior surface while simultaneously performing the non-covalently packaging of LigM enzymes within the encapsulin lumen. LigM was chosen as the downstream enzyme in our metabolon design as it naturally participates in the catabolic utilization of aryl substrates derived from the degradation of lignin biopolymers as a carbon source for cellular growth (33, 34). The depolymerization and valorization of lignin's heterogeneous phenylpropanoid constituents represents an important current challenge in industry and biotechnology as lignin is both the second most abundant polymer in nature and one of the largest untapped carbon sources on the planet due to its insolubility and its intractability (35-38). As such, nanocontainer-scaffolded synthetic metabolons, such as the one we propose herein, represent valuable methodologies for the ecologically friendly and sustainable production of biofuels and high value chemical synthons.

3.2 – Results and Discussion

Building off of our successful high-density display of active *E. coli* DHFR-M42W variant biocatalysts (denoted DHFR* for brevity) on the exterior surface of TmE nanocontainers using the covalent SpyCatcher/SpyTag system (described in Chapter 2), we subsequently sought to expand our encapsulin-based architecture into a multienzyme nanoreactor capable of performing sequentially-acting catalytic steps. To this end, we decided to employ LigM from the soil bacterium *Sphingomonas paucimobilis* SYK-6 (33, 34), as the second biocatalyst in our system following the design scheme detailed in **Figure 3.1**. Specifically, our plan was to decorate the exterior surface of TmE-SC nanocontainers with DHFR* enzymes, as described in Chapter 2, while simultaneously loading LigM enzymes into the TmE interior, thus artificially bringing the two catalysts into close spatial proximity. In so doing, THF generated *in situ* from the reduction of DHF by surface-tethered DHFR* was expected to drive the THF-dependent demethylation of the aryl substrate vanillate by encapsulated LigM. The corresponding protocatechuate (PCA) formed from vanillate demethylation represents an industrially significant precursor used to generate the high value synthons *cis,cis*-muconic acid and 2-pyrone-4,6-discarboxylate, which in turn are used in the mass production of nylon and biodegradable polymers (38-40). Similarly, the N^5-methyltetrahydrofolate (5-CH_3-THF) produced upon methylation of THF represents an important biological supplement for the prevention of neonatal neural tube defects (41-43). Metabolite exchange was expected to occur between the two biocatalysts in our nanocontainer-based system via the native pores located at the three distinct symmetry axes uniformly arrayed throughout the encapsulin shell. This overall nanoreactor design was chosen to examine whether the co-localization of DHFR* and LigM would permit substrate channeling-like effects for the scaffolded metabolon enzymes, thereby yielding enhanced catalytic flux rates relative to the observed throughput for the two enzymes free in solution.

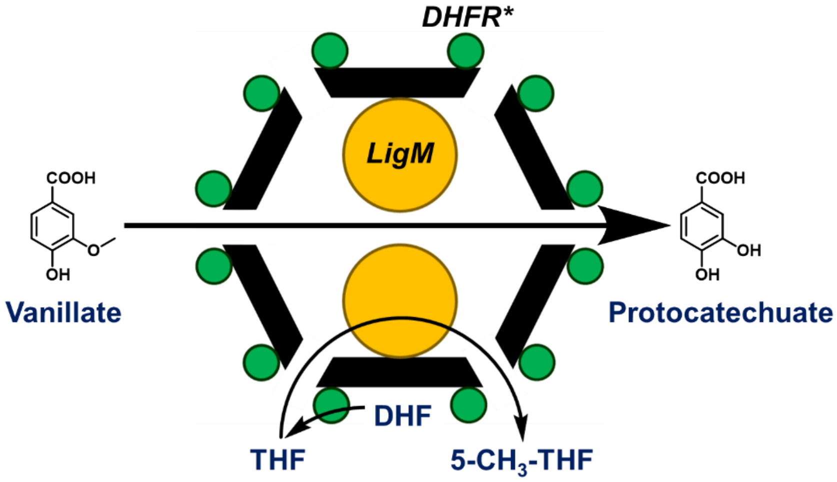


Figure 3.1 – Cartoon representation of the proposed multienzyme nanoreactor design. Surface-bound DHFR* enzymes are depicted as green spheres while encapsulated LigM enzymes are depicted as yellow spheres. The encapsulin shell is depicted as a black

Previous works have shown that non-native cargoes can be directed into the TmE lumen using short aliphatic targeting peptides derived from the C-terminal region of the native ferritin-like protein (FLP) cargo (44-46). As an initial test for our design strategy, we recombinantly fused the FLP’s C-terminal cargo localization peptide (Clp) sequence onto the C-terminus of LigM and purified the resulting fusion protein to homogeneity. The demethylation activity of the purified LigM-Clp was subsequently assessed using exogenously added THF as the methyl acceptor, yielding the activity profile depicted in **Figure 3.2A**. The resulting profile shows that the LigM-Clp fusion protein (referred to as simply “LigM” from this point onward) retains enzymatic activity as approximately 20% conversion of the starting 2 mM vanillate is shown to occur over the 5 hour reaction period. In accordance with the inherent stoichiometry of LigM, the final PCA and 5-CH_3-THF reaction products emerge in the expected 1:1 molar ratio. Interestingly, we also noticed a discrepancy in the cofactor pools over the course of the reaction as approximately 3-fold more

THF was consumed relative to the amount of 5-CH_3-THF that was produced. This phenomenon will be discussed in greater detail later.

Using this activity data with LigM alone as our baseline for comparison, we next performed another activity assay, this time employing equimolar concentrations of purified DHFR*-CST and LigM with exogenously added DHF as the starting cofactor to examine if the *in situ* generation of THF by DHFR*-CST could effectively drive the demethylase activity of LigM. The resulting activity profile shows that the formation of THF is rapid as the starting DHF pool is completely depleted within the first hour of the reaction period **Figure 3.2B**. However, while both the consumption of DHF and the corresponding formation of THF could be readily observed among the 259 nm absorbance signals detected during the LC/MS analysis of the various reaction components, direct quantitation of DHF proved difficult as DHF was found to rapidly degrade in response to the acidic quenching conditions used when collecting time point samples. This susceptibility of DHF to lower pH values has been documented previously and hampered us from effectively generating standard curves necessary for accurate cofactor quantitation (47-48).

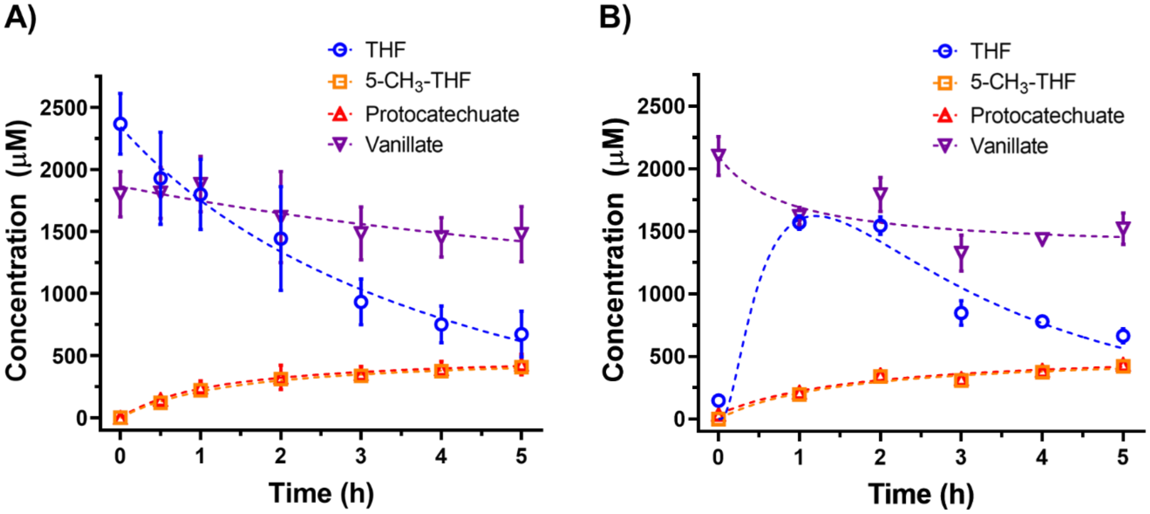


Figure 3.2 – Enzymatic activity assays monitoring the production of PCA and 5-CH_3-THF using either A) purified LigM alone in conjunction with THF, or B) equimolar concentrations of purified DHFR*-CST and LigM in conjunction with DHF.

Despite this inability to report DHF concentrations, standard curves for all other reaction components, including THF, could be generated with high accuracy, and thus we elected not to alter the assay parameters.

Interestingly, though the conversion of DHF to THF appeared to be rapid in the coupled enzyme reaction, the downstream conversion of vanillate to PCA by LigM appeared identical to the previous reaction containing free LigM alone as roughly 20% total conversion was observed overall. This result was not entirely unexpected as a Michaelis-Menten assessment conducted previously with wildtype LigM reported a k_{cat} value 5.76 ± 0.25 s^{-1} and of K_M value 0.72 ± 0.11 mM for THF (49). While the k_{cat} value of LigM is nearly equivalent to that of our DHFR* variant, the K_M value is more than 100-fold and 35-fold higher than the K_M values exhibited for DHF by free DHFR*-CST (K_M = 6.83 ± 0.83 μM) or TmE-immobilized DHFR*-CST (K_M = 19.74 ± 2.57 μM), respectively. We thus initially concluded that the nearly identical activity profiles observed

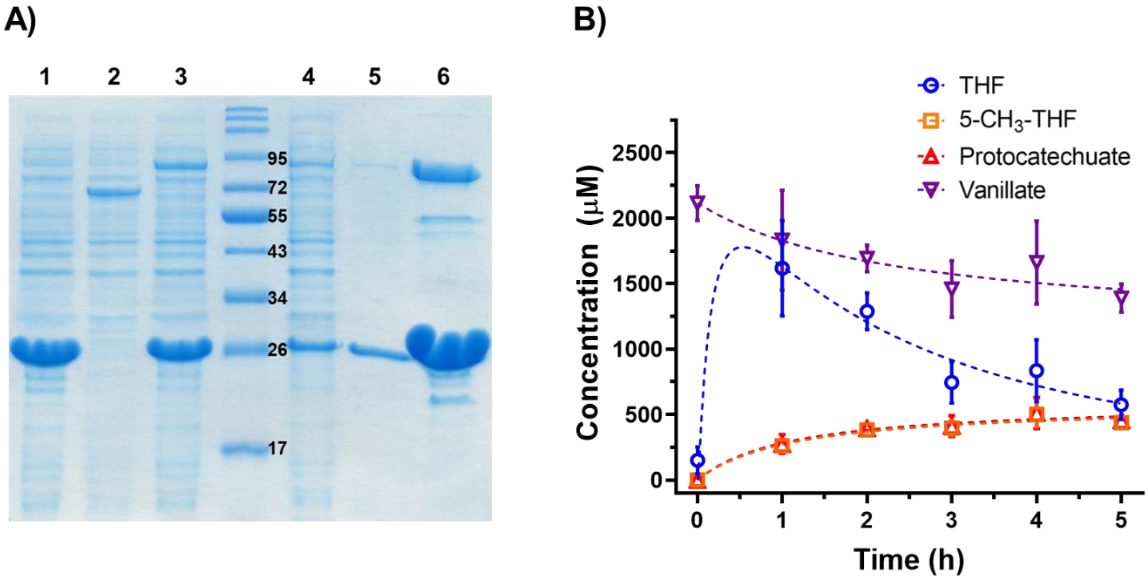


Figure 3.3 – A) SDS-PAGE gel depicting the *in vitro* bioconjugation of DHFR*-CST (20.8 kDa) and LigM-SC (65.2 kDa). Lane 1: DHFR*-CST clarified lysate; Lane 2: LigM-SC clarified lysate; Lane 3: combined clarified lysate samples following 60 minute incubation; Lane 4: IMAC resin flow-through; Lane 5: IMAC resin buffer wash; Lane 6: elution of proteins bound to IMAC resin using 250 mM imidazole wash. Bioconjugated LigM-SC:DHFR*-CST (86.0 kDa) visible in lanes 3 and 6. Numerical values in the middle lane represent the molecular weights (in kDa) of the adjacent protein ladder bands. B) Activity profile generated using LigM-SC:DHFR*-CST bioconjugates with DHF as the starting cofactor.

for the LigM only and coupled enzyme assays was likely due to the significantly weaker substrate binding of LigM serving as the primary rate-limiting factor in the overall cascade reaction.

Having validated the feasibility of coupling the catalytic activities of DHFR*-CST and LigM into a complimentary reaction pathway, our next goal was to determine whether reducing the spatial proximity between the two metabolon biocatalysts could influence the overall flux through the two-step cascade by generating sufficiently high local concentrations of THF to compensate for its weaker binding affinity with LigM. Consequently, we genetically fused the SpyCatcher domain onto the C-terminus of LigM to generate a LigM-SpyCatcher (LigM-SC) variant that would permit us the ability to covalently tether LigM directly to DHFR*-CST via isopeptide bond formation. Clarified cell lysates derived from separate heterologous expression cultures of LigM-SC and DHFR*-CST were mixed *in vitro* to effect covalent bond formation (**Figure 3.3A**). Following a brief incubation period, the conjugated LigM-SC:DHFR*-CST complex was isolated as a >95% pure product through sequential IMAC and SEC chromatography steps. Interestingly, a subsequent activity assay performed with these covalently-tethered biocatalysts yielded identical results as were observed for the previous two reactions with approximately 20% conversion of the initial vanillate pool occurring over the 5 hour incubation period (**Figure 3.3B**). These results seemingly indicate that the spatial proximity of the two enzymes has no effect on the corresponding flux through the multistep catalytic pathway.

Given the good agreement of the preceding three activity assay datasets, we subsequently sought to probe the relative efficiency of our bi-enzymatic nanoreactor scaffold design. Based on the published crystal structures for both LigM (49, 50) and TmE (51), we estimated that a theoretical maximum of 32 LigM monomers could be packaged within each nanocontainer given an approximate volume of 128 nm^3 for each LigM enzyme and a luminal void volume of 4189

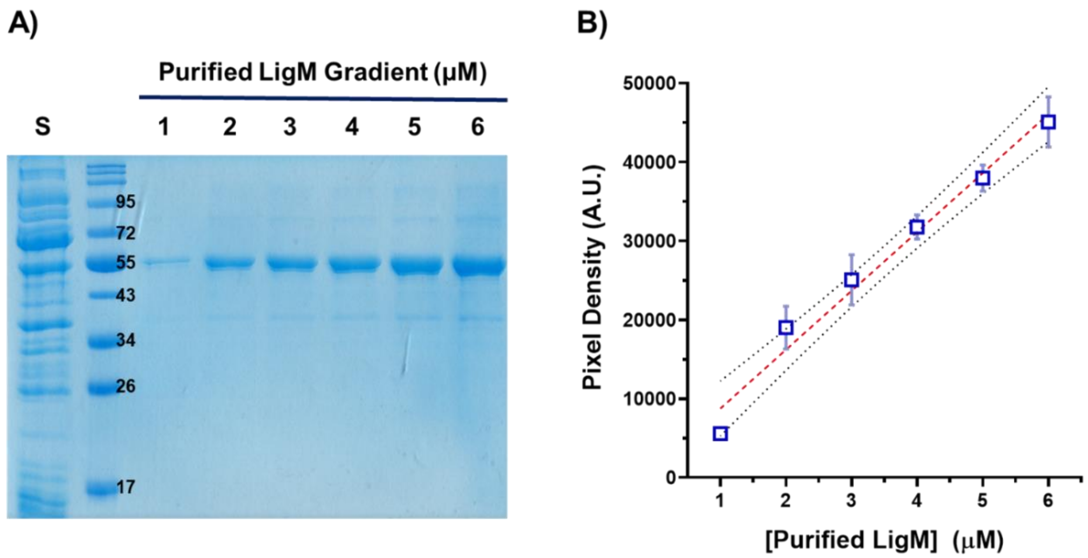


Figure 3.4 – A) SDS-PAGE gel image depicting the semi-purified TmE-SC●LigM:DHFR*CST nanoreactors (Lane S, left) and the corresponding gradient of purified LigM used for gel densitometry analysis. Numerical values in the second lane represent the molecular weights (in kDa) of the adjacent protein ladder bands. B) Gel densitometry plot containing the linear regression analysis derived from the purified LigM gradient SDS-PAGE data (R^2 value = 0.979).

nm^3 for the assembled encapsulins. Though this theoretical maximum would likely never be reached in practice due to steric constraints within the encapsulin lumen, the combination of these estimates and our ability to achieve full covalent decoration of TmE-SC nanocontainers with 60 immobilized DHFR* enzymes per container indicate that our multienzyme system should have an approximate minimum ratio of 2:1 DHFR* to LigM enzymes upon assembly. Nanoreactors were subsequently constructed by first co-expressing TmE-SC and LigM (TmE-encapsulated LigM referred to herein as TmE-SC●LigM), and then decorating the exterior surface of the nanocontainers *in vitro* by mixing the resulting clarified lysate from the TmE-SC●LigM culture with clarified lysate derived from an expression culture containing DHFR*-CST to avoid the metabolic burdens of expressing all three proteins in a single host. Due to the large culture sizes needed to achieve sufficient yields of the three protein components, the nanoreactors were only semi-purified following several chromatography and salt-mediated precipitation steps. However, distinct protein bands corresponding to the sizes of the encapsulated LigM enzyme and the TmE-

SC:DHFR*-CST fusion proteins were detectable via SDS-PAGE analysis (**Figure 3.4A**). The concentration of the encapsulated LigM population was subsequently determined by gel densitometry analysis (**Figure 3.4B**) and served as the basis for the preparation of our activity assays given that LigM represents the rate-limiting enzyme in the cascade reaction.

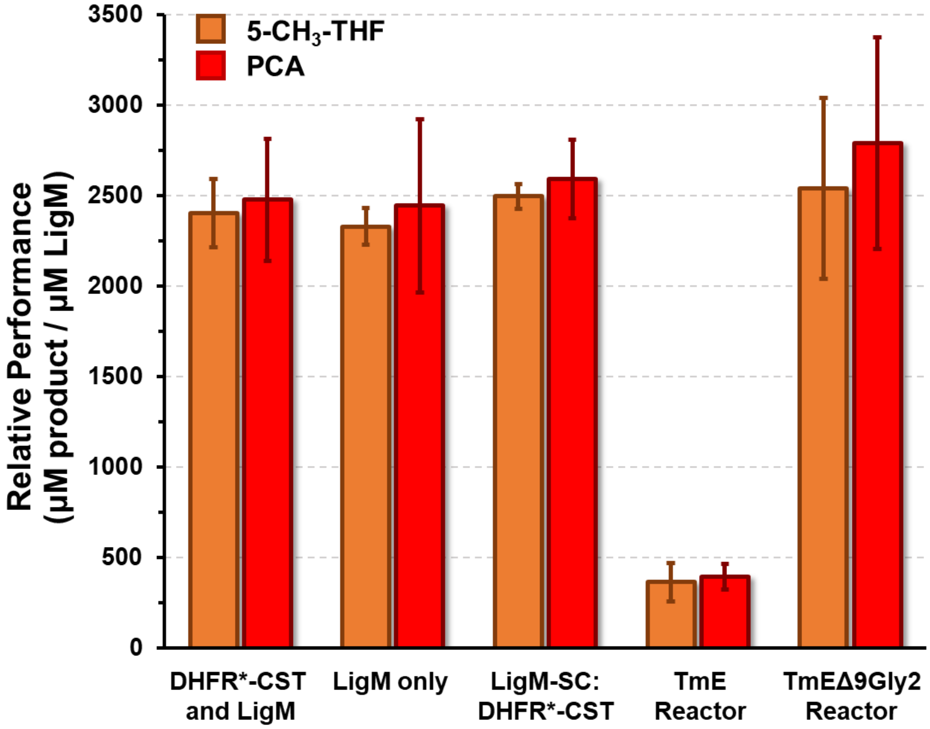


Figure 3.5 – Relative performance of different activity assays assessed by normalizing the endpoint concentrations of the terminal PCA and 5-CH_3-THF cascade products to the concentration of LigM used in each assay.

Activity assays performed with DHFR*-decorated TmE-SC●LigM nanoreactors (i.e. TmE-SC●LigM:DHFR*-CST) show that the bi-enzymatic cascade is indeed functional as DHF was readily converted to THF, and the concentrations of both PCA and 5-CH_3-THF were observed to gradually increase in a linear fashion over the 5 hour reaction timeframe. However, in comparison to the preceding reaction profiles, encapsulation of LigM leads to an apparent 5-fold reduction in cascade efficiency when the final amounts of PCA and 5-CH_3-THF are normalized relative to the concentration of LigM present (**Figure 3.5**). We reasoned that this inefficiency was likely due to either some percentage of the encapsulated LigM enzymes being inactive (either from protein denaturation or due to crowding within the encapsulin lumen), or due to restricted diffusion of the bulky substrates and folate cofactors through the narrow 3 Å pores of the native encapsulin shell. The latter phenomenon has been reported previously for encapsulin-based systems, such as a ~1000-fold reduction in turnover observed for firefly luciferase encapsulated within an encapsulin nanocontainer isolated from *Rhodococcus erythropolis* N771 due to the restricted diffusion of the required adenosine triphosphate and D-luciferin substrates into the encapsulin lumen (52). Fortuitously, our lab has previously generated an encapsulin variant, referred to as TmEΔ9Gly2, possessing artificially enlarged pores located at the container's 5-fold symmetry axes which have been shown to allow for the enhanced diffusion of small cation probes across the encapsulin shell relative to the wildtype TmE containers (46). An atomic level structure was recently obtained for the TmEΔ9Gly2 variant via cryogenic transmission electron microscopy, showing that the engineered pores of the variant are approximately 18 Å in diameter, or 6-fold wider than the 3 Å pores natively located at the 5-fold axes (53). We thus recombinantly introduced the same Δ9Gly2 pore mutations into the TmE-SC gene sequence (new variant referred to as TmEΔ9Gly2-SC) to examine if the diffusion of substrates and/or cofactors was indeed the source of the poor performance exhibited by our nanoreactors.

DHFR*-decorated TmEΔ9Gly2-SC●LigM nanoreactors were prepared and purified in the same fashion as the preceding nanoreactors containing wildtype-sized pores. Excitingly, activity assay data collected for the TmEΔ9Gly2-SC●LigM:DHFR*-CST scaffolds showed a complete recovery of cascade efficiency as the LigM-normalized endpoint concentrations of PCA and 5-CH_3-THF matched those obtained for all three of the assays performed with non-encapsulated LigM enzymes (**Figure 3.5**). These results indicate that restricted metabolite exchange was likely the cause of the inefficiency exhibited by the initial TmE-SC●LigM nanoreactor scaffolds. Additionally, the recovery of cascade efficiency further indicated that the entire population of LigM enzymes within the TmEΔ9Gly2-SC●LigM nanoreactors are catalytically competent, and thus the physical sequestration of LigM enzymes within the encapsulin lumen does not negatively impact enzymatic function. However, as with the activity assay performed using conjugated LigM-SC:DHFR*-CST fusion proteins, the co-localization of the two biocatalysts in our nanoreactor-based metabolon design appears to offer no evidence of pathway flux enhancements resulting from cascade catalysis relative to the reactions performed with both enzymes freely dispersed in solution.

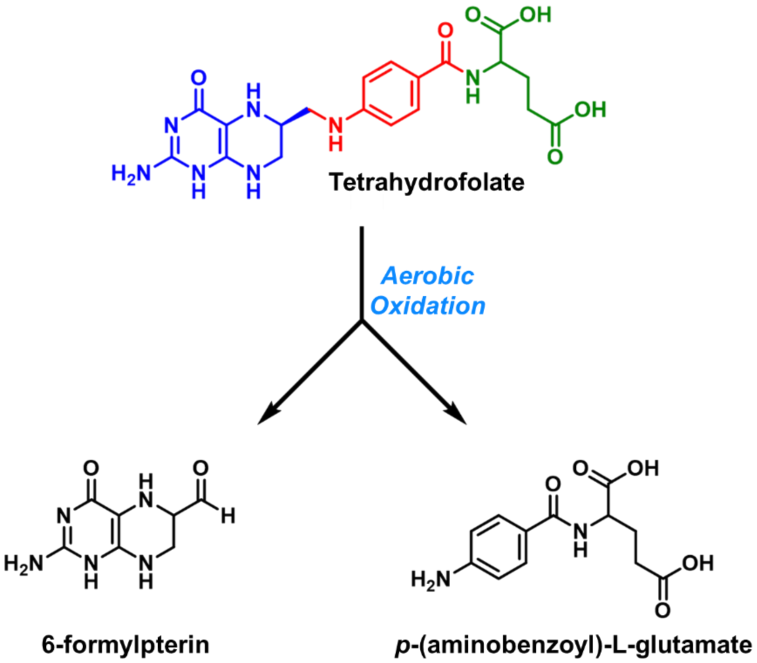


Figure 3.6 – Scheme depicting the initial products generated from the aerobic oxidative degradation of tetrahydrofolate. The 2-amino-4-oxo-6-methylpteridine, *p*-aminobenzoate, and L-glutamate portions of the tetrahydrofolate cofactor are colored in blue, red, and green, respectively, in the chemical structure above.

Following the collection of our activity assay datasets for both freely dispersed and encapsulin-scaffolded enzymes, we began to investigate different strategies to enhance the catalytic performance of our cascade systems beyond the seemingly persistent 20% conversion threshold observed in most of our activity tests. Upon closer examination of our enzymatic activity data, we noticed that the decline in THF concentration during the 5 hour reaction period was consistently greater than the amount of 5-CH_3-THF produced, indicating that the THF pool was being prematurely depleted and was perhaps hampering the demethylase activity of LigM. THF is well known to be susceptible to oxidative degradation by molecular oxygen under aerobic conditions, resulting in cleavage of the C^9-N^{10} bond joining the cofactor's pterin and p-aminobenzoic acid moieties to form 6-formylpterin and *p*-(aminobenzoyl)-L-glutamate (*p*-ABG) as degradation byproducts (**Figure 3.6**) (54, 55). Though excess sodium ascorbate salts were included in our reactions to serve as sacrificial oxidants intended to preserve the folate cofactor pools, we reassessed the raw LC/MS chromatograms from our activity assay datasets and located a product peak with a retention time of 5.1 minutes presenting positive and negative mode ion products directly corresponding to the molecular weight of *p*-ABG (**Figure 3.7A**). We subsequently performed parallel activity assays under aerobic and anaerobic conditions using a 1:1 molar ratio of purified DHFR*-CST and LigM, the results of which show that the THF pool appears to be much more stable and the accumulation of *p*-ABG is significantly reduced under anaerobic conditions (**Figure 3.7B**). Overall, approximately 14% conversion was achieved during the 5 hour reaction period as a result of having to perform the anaerobic assay at 22°C within our anaerobic chamber, whereas all previous reactions had been incubated at 37°C.

In conjunction with our aerobic versus anaerobic assessment of THF stability, we also located literature sources reporting that LigM is susceptible to product inhibition by 5-CH_3-THF with an apparent K_i value of 100 ± 1 μM, representing a product binding affinity more than 7-fold stronger than the enzyme's K_M value for the THF cofactor (49, 56). To verify if product inhibition was indeed contributing to the overall throughput limitations exhibited by our coupled enzyme cascade, we subsequently prepared an activity assay under aerobic conditions, again containing a 1:1 molar ratio of DHFR*-CST and LigM, using starting concentrations of 150 µM vanillate as the substrate and 2 mM DHF as the cofactor. This assay was designed to mimic reported conditions in which an excess of THF was able to overcome the higher binding affinity of the 5-CH_3-THF generated by LigM during the reaction period (56). The resulting reaction profile generated with the reduced concentration of vanillate showed better conversion over the previous trials with more than 60% of the vanillate pool converted to PCA and 5-CH_3-THF within 5 hours.

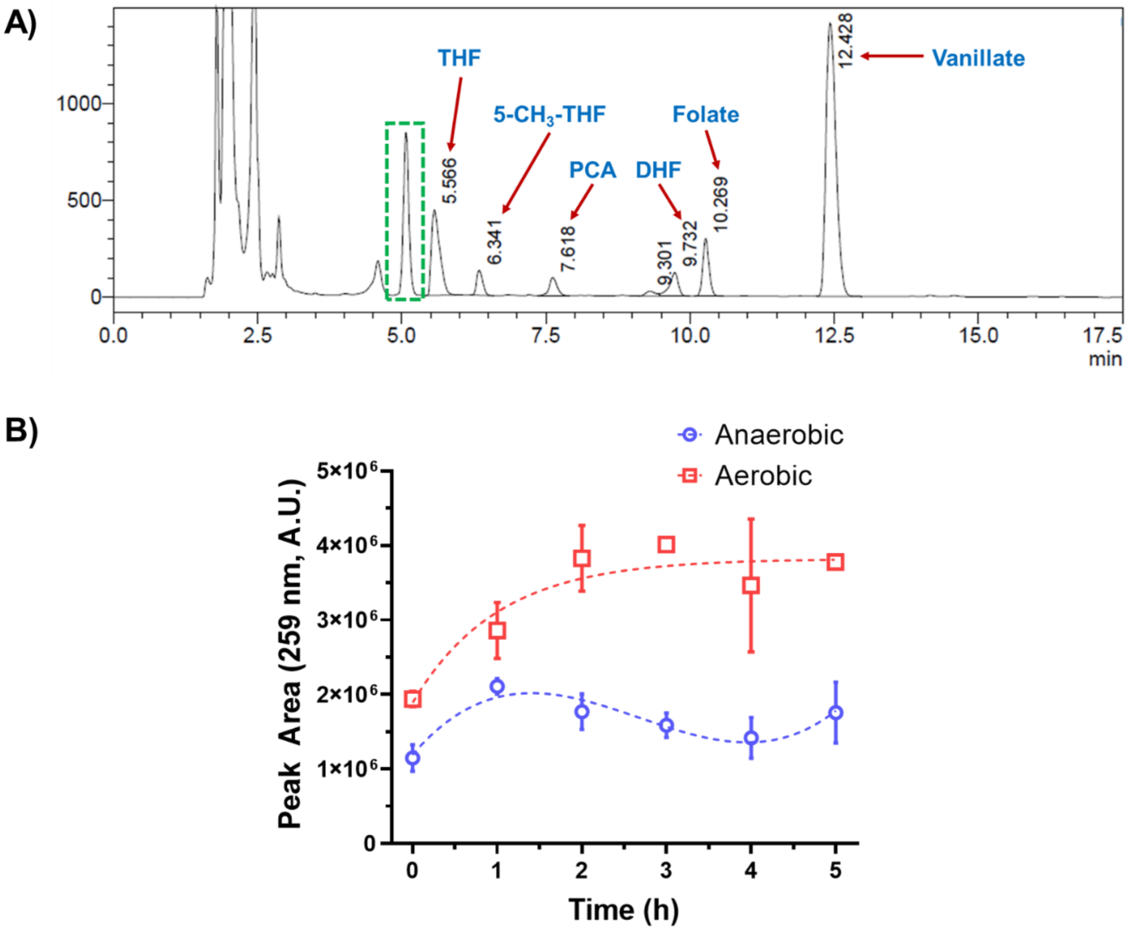


Figure 3.7 – A) LC/MS chromatogram depicting the substrates and products of an activity assay. The *p*-ABG peak arising from the degradation of THF is highlighted within the dashed green box. B) Comparison of *p*-ABG formation under aerobic and anaerobic assay conditions.

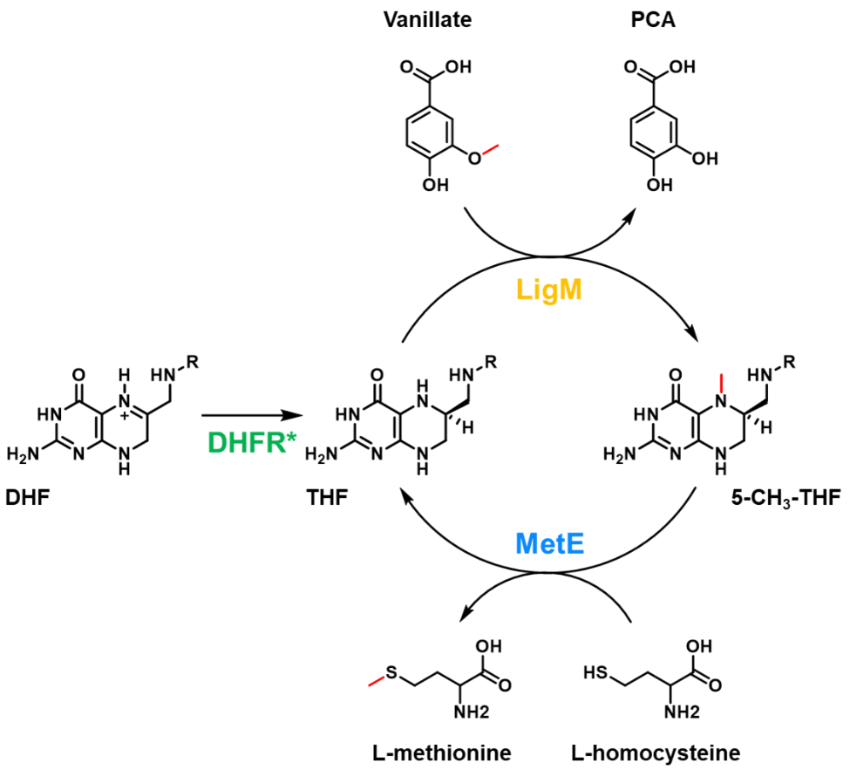


Figure 3.8 – Expanded nanoreactor cascade scheme with the incorporated the *C. roseus* MetE to serve as a THF-recycling system.

While increasing the *in situ* concentration of THF proved to be a viable option for improving the efficiency of our cascade when tested with enzymes heterogeneously dispersed in solution, this strategy is less attractive from an industrial perspective as it requires maintaining a persistently high concentration of a labile cofactor. As an alternative strategy, we began to explore the possibility of coupling our bi-enzymatic metabolon with a third biocatalyst designed to recycle the THF consumed by LigM, thereby permitting the solution concentration of THF to be kept at lower levels. Similar coupled-enzyme recycling systems have been effectively utilized in the past to recycle expensive and labile cofactors such as NAD(P)H using inexpensive feedstock metabolites (57). In living organisms, 5-CH_3-THF is used by methionine synthase (MetE) enzymes as a methyl donor for the generation of L-methionine from L-homoserine (58-60). Accordingly, Rosini *et al.* recently isolated the plant MetE gene native to *Catharanthus roseus* and successfully coupled the 5-CH_3-THF-dependent methylation of L-homoserine to the demethylation of a series of methylated aryl substrates accepted by LigM (56). In light of these results, we obtained the *C. roseus* MetE gene for ourselves and sought to employ it towards THF recycling in our own nanoreactor cascade following the scheme depicted in **Figure 3.8**. However, while we were able to successfully heterologously express and purify the MetE gene, we were unable to proceed further with our studies due to the unforeseen closure of Emory University in response to the emerging public health threat of COVID-19 epidemic.

3.3 – Conclusion

In this study, we have successfully constructed a multienzyme nanoreactor system utilizing the *T. maritima* encapsulin as a bifunctional scaffold for the simultaneous surface display of one enzyme in tandem with the specific encapsulation of another. Polyvalent decoration of the encapsulin surface with DHFR variant enzymes (described in Chapter 2) thus permitted the *in situ* generation of THF, which was then used to drive the LigM-catalyzed demethylation of an industrially-relevant aryl precursor molecule within the encapsulin lumen derived from the catabolism of plant-based lignin biopolymers. However, while the sequential catalysis steps were shown to function as intended with the initial TmE-SC●LigM:DHFR*-CST nanoreactor design, the overall throughput of the scaffolded metabolon cascade was approximately 5-fold less efficient than when the same cascade was performed with non-scaffolded enzymes freely distributed in solvent. We subsequently eliminated the observed reduction in cascade efficiency by recombinantly enlarging the 5-fold symmetry pores intrinsic to the encapsulin architecture using a reported mutagenesis strategy previously employed by members of our lab (46). This second nanoreactor design, deemed TmEΔ9Gly2-SC●LigM:DHFR*-CST, exhibited cascade efficiency on par with the non-scaffolded cascade enzymes, indicating that diffusion of the necessary substrates and cofactors through the encapsulin shell was the primary factor limiting the efficiency of the first nanoreactor chassis.

The initial intent of our encapsulin-based multienzyme platforms was to achieve faster cascade flux, akin to the metabolic substrate channeling observed in nature, by co-localizing our DHFR* and LigM biocatalysts on a common proteinaceous scaffold. Though our nanoreactor scaffolds were shown to effectively facilitate sequential catalytic steps, no evidence of cascade flux accelerations was observable with our system due to product inhibition of LigM by 5-CH_3-THF masking any such benefits. However, flux enhancements resulting from enzyme co-localization may still be revealed upon the elimination of the 5-CH_3-THF inhibition. To that end, we have begun attempts to couple the encapsulin-scaffolded cascade to the methyltransferase activity of plant-based MetE, which utilizes 5-CH_3-THF as a methyl donor for the production of L-methionine. Successful incorporation of MetE into our cascade design would also offer the added benefit of recycling the THF cofactor needed to drive further rounds of catalytic turnover with the encapsulated LigM.

In biological systems, multistep cascade reactions are often designed to maintain lower concentrations of pathway intermediates as a means of mitigating metabolite loss in order to better maximize atom economy (1, 2). In this light, our nanoreactor design, though functional, requires further optimization to behave more like a biologically-inspired enzymatic cascade. Specifically, the kinetic imbalance between the catalytic efficiencies of DHFR* and LigM were shown to lead to a rapid accumulation of THF as an intermediate cofactor, which was subject to deleterious oxidative degradation under the aerobic assay conditions. Transitioning from aerobic to anaerobic assay conditions resulted in better retention of the THF pool, which could be especially beneficial when coupled with the THF-recycling activity of MetE. Additionally, the design of the actual nanoreactor itself could be further optimized as the encapsulin scaffolds are inherently imbalanced, possessing a theoretical minimum of a 2-fold higher concentration of the faster DHFR* relative to the slower LigM catalysts. Future designs could account for the higher catalytic efficiency of DHFR* relative to LigM by switching the orientation of the two enzymes (i.e. surface tethered LigM in conjunction with encapsulated DHFR*). Given that each DHFR* monomer is approximately 3-fold smaller than a corresponding LigM monomer, however, this strategy might not be as effective since higher concentrations of DHFR* could theoretically be packaged within the lumen of each nanocontainer. Alternatively, varying concentrations of purified DHFR*-CST could be used during *in vitro* decoration of the TmE-SC surfaces to determine an optimal DHFR*-LigM cascade ratio.

3.4 – Materials and Methods

3.4.1 – Reagents and Materials

Vanillate and protocatechuate were purchased from Oakwood Chemical (West Columbia, SC). Tetrahydrofolate and 5-methyltetrahydrofolate were purchased from Cayman Chemical (Ann Arbor, MI). All other chemicals and reagents were purchased from Sigma Aldrich (St. Louis, MO). All pre-packed chromatography columns were purchased from GE Healthcare (Marlborough, MA). DH5α competent *E. coli* cells were purchased from ThermoFisher Scientific (Waltham, MA) and BL21(DE3) and competent *E. coli* cells were purchased from New England Biolabs (Ipswich, MA). All commercial enzymes used for molecular cloning were purchased from New England

Biolabs unless otherwise noted. The *Sphingomonas paucimobilis* SYK-6 gene (GenBank: BAD61059.1) was ordered as two gBlock gene fragments from Integrated DNA Technologies (Coralville, IA), which were recombinantly assembled and cloned into the pMATT2 vector using standard molecular cloning techniques. All DNA primers were ordered from Integrated DNA Technologies. The TA cloning kit was purchased from Invitrogen (Waltham, MA). The pET24b:MetE plasmid was a generous gift from Dr. Elena Rosini and Dr. Loredano Pollegioni (Università degli Studi dell'Insurbia, Varese, Italy).

3.4.2 – Molecular Cloning

The pET23b:TmE-SC and pCDF:DHFR*-CST plasmids were prepared as described in Chapter 2. The pET23b:TmE(Δ9Gly2)-SC variant plasmid was generated using mutagenic primers to perform whole-plasmid amplification with the pET23b:TmE-SC plasmid serving as the amplification template. Successful incorporation of the Δ9Gly2 pore mutation was confirmed via DNA sequencing. The *Sphingomonas paucimobilis* SYK-6 LigM gene was ordered as two gBlock fragments and assembled into a single continuous gene by overlap extension PCR. The LigM sequence was modified slightly from the wildtype sequence to reduce certain G/C-rich regions of the DNA sequence in order to facilitate the chemical synthesis of the gBlocks. The assembled overlap extension product, containing encoded HindIII and SpeI restriction sites at the 5' and 3' ends, respectively, was generated using Taq polymerase. The overlap extension product was then directly ligated into linear pCR™2.1 vector using a T/A cloning kit. Gene incorporation was confirmed by standard colony PCR. Subsequently, the full-length LigM gene was excised from the pCR™2.1 vector via digestion with HindIII and SpeI, and it was then transferred to pMATT2 vector that had been linearized with the same enzymes. The final pMATT2:LigM plasmid construct was confirmed via DNA sequencing. The pCDF:LigM-SC vector was prepared by first fusing the LigM and SpyCatcher gene sequences using overlap extension PCR and terminal primers encoding HindIII and AvrII restriction sites within the forward and reverse primers, respectively. The full length gene hybrid was then digested with HindIII and AvrII, and it was then ligated into linear pCDFDuet-1 vector that had been digested with the same enzymes. The final vector product was confirmed via DNA sequencing.

3.4.3 – Protein Overexpression and Purification

All recombinant proteins were overexpressed in *E. coli* BL21(DE3) cells. Cells containing the pCDF:DHFR*-CST vector were grown in LB medium at 37°C in the presence of 0.05 mg/mL streptomycin to an OD_{600} value between 0.5 – 0.7, and then protein expression was induced with the addition of a final concentration of 0.3 mM IPTG. The cells were harvested via centrifugation after 16 hours of expression time at 30˚C. For LigM and LigM-SC, cells containing the respective vector were grown in LB medium at 37°C in the presence of 0.05 mg/mL streptomycin to an OD_{600} value between 0.5 – 0.7, and then protein expression was induced with the addition of a final concentration of 0.1 mM IPTG. The cells were harvested via centrifugation after 20 hours of expression time at 18°C. *In vivo* cargo loading of LigM into TmE nanocontainers was performed by co-expression of the pET23b:TmE-SC or pET23b:TmE(Δ9Gly2)-SC vectors and the pMATT2:LigM vector. Cells containing the respective encapsulin and LigM vectors were grown in LB medium at 37°C in the presence of 0.1 mg/mL ampicillin and 0.05 mg/mL streptomycin to an OD_{600} value between 0.5 – 0.7, and then protein expression was induced with the addition of a final concentration of 0.05 mM IPTG. The cells were harvested via centrifugation after 20 hours of expression time at 18˚C. All cell pellets were frozen and stored at -20°C until purification.

Both LigM and LigM-SC were purified in an identical manner by first resuspending the respective cell pellets in 50 mM Tris-HCl (pH 7.5), 20 mM NaCl containing Protease Inhibitor Cocktail (Sigma Aldrich, P8849), 0.1 μg/mL PMSF, and 5 μg/mL DNaseI. The cell solution was then lysed via sonication (10 second pulses, 20 seconds rest between pulses) for 3.5 minutes. Insoluble debris was removed via centrifugation at 4000 rpms for 40 minutes at 4°C, then the clarified lysate was loaded into a 5 mL HiTrap Q-FF anion exchange chromatography column (pre-equilibrated with 5 column volumes of lysis buffer). The resin bed was washed with 3 column volumes of lysis buffer to remove non-specifically bound proteins, and then bound LigM or LigM-SC was eluted with 3 column volumes the same buffer containing 200 mM NaCl. The elution fraction was then gently concentrated in an Amicon Ultra centrifugal filter (10 kDa MWCO) at 4°C. The concentrated protein product was then recovered and loaded into a Superdex 200 Increase 10/300 GL column that had been pre-equilibrated with 50 mM K-Phosphate (pH 7.5), 100 mM NaCl. The same buffer was used subsequently as the mobile phase during size exclusion chromatography. Protein elution was monitored spectrophotometrically at 280 nm. Purified LigM or LigM-SC collected from the chromatography fractions was flash frozen in liquid nitrogen and stored at -80°C until use in experiments. Protein concentrations for LigM and LigM-SC were determined spectrophotometrically at 280 nm using theoretical extinction coefficient values: 108,750 $M^{-1}cm^{-1}$ and 120,210 $M^{-1}cm^{-1}$, respectively.

Bi-enzymatic nanoreactors were generated by resuspending cell pellets from a > 800 mL of TmE-SC (or TmEΔ9Gly2-SC) and LigM co-expression cell culture and cell pellets from a 100 – 200 mL of DFR*-CST expression culture in 50 mM Tris-HCl (pH 7.5) containing Protease

Inhibitor Cocktail, 0.1 µg/mL PMSF, and 5 µg/mL DNaseI. The cell solutions were then lysed separately by sonication in the same manner as described above. Insoluble debris was removed via centrifugation at 4000 rpms for 40 minutes at 4°C. Subsequently, the clarified lysates derived from the TmE-SC●LigM and DHFR*-CST expression cultures were pooled into a fresh Falcon tube and *in vitro* surface decoration of the encapsulins' SpyCatcher domains was carried out for 60 minutes at 4°C on an orbital rocking table. The pooled clarified supernatants were then loaded into a HiTrap Q-FF anion exchange column (pre-equilibrated with 5 column volumes of lysis buffer) and the column flow-through was collected. Solid ammonium sulfate was added to the flow-through sample to reach a final concentration of 20% (w/v), and then the sample was incubated for 20 minutes at 4°C on an orbital rocking table to promote precipitation of unwanted contaminant proteins. Following the incubation period, the sample tube was centrifuged at 4000 rpms for 20 minutes at 4°C to isolate precipitated contaminant proteins. The resulting supernatant was decanted into a fresh Falcon tube, and additional solid ammonium sulfate was added to reach a final concentration of 40% (w/v). The same 4°C incubation and centrifugation steps were repeated to isolated precipitated nanoreactors. The resulting solid pellet was resuspended in 5 to 10 mL of 50 mM Tris-HCl (pH 7.5) and was immediately loaded into a cellulose dialysis bag (14 kDa MWCO). The dialysis bag was immersed sequentially into two beakers containing 1 L of 50 mM HEPES-KOH (pH 7.5), 100 mM NaCl each with a minimum of 4 hours dialysis at 4°C with gentle stirring in each beaker to remove residual ammonium sulfate. The dialyzed sample was then recovered, passed through a 0.2 µm syringe filter, and then loaded into a HiPrep 16/60 Sephacryl S-500 HR size exclusion column that had been pre-equilibrated in 50 mM HEPES-KOH (pH 7.5), 100 mM NaCl and was connected to an NGC Chromatography FPLC system (Bio-Rad, Hercules, CA). The same buffer was used as the mobile phase during this second chromatography step. Assembled encapsulins were isolated from the size exclusion column and stored at 4°C until use in experiments. Concentrations of encapsulated LigM were determined via 12% SDS-PAGE gel densitometry analyses using purified LigM to generate standard curves. All densitometry analyses were performed in triplicate and were processed using the FIJI image analysis software (61).

3.4.4 – LC/MS Analyses

Multienzyme cascade activity assays were performed at either 37°C under aerobic conditions or at 22°C under anaerobic conditions using a reaction mixture containing 100 mM K-Phosphate (pH 7.0), 100 mM NaCl, 2 mM dithiothreitol, 5 mM sodium ascorbate, 3 mM NADPH, 2 mM DHF, and 2 mM vanillic acid. For the conversion reactions performed with LigM in the absence of DHFR*-CST, the reaction mixture was identical except that the NADPH was excluded and 2 mM THF was added in place of the DHF. Time point samples (35 μL each) were collected every hour for 5 hours. Each time point was immediately quenched by adding an equal volume of 240 mM HCl to the sample and mixing thoroughly. Samples were subsequently subjected to LC/MS analysis by injecting 50 μL of each quenched mixture into a Shimadzu LCMS-2020 Single Quadrupole Liquid Chromatograph Mass Spectrometer equipped with an Agilent Zorbax Eclipse Plus C-18 column (5 μm particle size; 4.6 x 250 mm; Santa Clara, CA) and were separated at a flow rate of 1.5 mL·min^{-1}. The mobile phase consisted of a mixture of (**A**) acetonitrile with 0.1% (v/v) formic acid and (**B**) water with 0.1% (v/v) formic acid. For compound separation, the following mobile phase gradient was used: 0 minutes, 5% **A**; 0.1 to 15 minutes, linear gradient from 5% to 20% **A**; 15.1 to 25 minutes, 5% **A** to re-equilibrate the column for subsequent injections. Elution of reaction components were monitored at 259 nm, and the retention times for the components were as follows: THF (5.6 min); N^5-CH_3-THF (6.4 min); protocatechuic acid (7.7 min); DHF (9.9 min); vanillic acid (12.5 min). For mass spectroscopy analysis, the interface temperature was set at 350°C, the drying line was set at 250°C, and the heat block was set at 200°C. Nitrogen gas was used as both the nebulizing and drying gas, and the respective flow rates for the nebulizing and drying lines were 1.5 and 3.0 $L \cdot min^{-1}$, respectively. All reactions were conducted in triplicate with error values representing one standard deviation from the mean. Curve fits for all activity assays were generated using GraphPad Prism 8.4.1 (La Jolla, CA).

3.5 – References

1. Lopez-Gallego, F.; Schmidt-Dannert, C., Multi-enzymatic synthesis. *Current Opinion in Chemical Biology* **2010**, *14* (2), 174-183.
2. Muschiol, J.; Peters, C.; Oberleitner, N.; Mihovilovic, M. D.; Bornscheuer, U. T.; Rudroff, F., Cascade catalysis – strategies and challenges en route to preparative synthetic biology. *Chemical Communications* **2015**, *51* (27), 5798-5811.
3. Lee, H.; DeLoache, W. C.; Dueber, J. E., Spatial organization of enzymes for metabolic engineering. *Metabolic Engineering* **2012**, *14* (3), 242-251.
4. Schoffelen, S.; van Hest, J. C. M., Multi-enzyme systems: bringing enzymes together in vitro. *Soft Matter* **2012**, *8* (6), 1736-1746.
5. Quin, M. B.; Wallin, K. K.; Zhang, G.; Schmidt-Dannert, C., Spatial organization of multi-enzyme biocatalytic cascades. *Organic & Biomolecular Chemistry* **2017**, *15* (20), 4260-4271.

6. Ljungcrantz, P.; Carlsson, H.; Mansson, M. O.; Buckel, P.; Mosbach, K.; Buelow, L., Construction of an artificial bifunctional enzyme, beta-galactosidase/galactose dehydrogenase, exhibiting efficient galactose channeling. *Biochemistry* **1989**, *28* (22), 8786-8792.

7. Seo, H. S.; Koo, Y. J.; Lim, J. Y.; Song, J. T.; Kim, C. H.; Kim, J. K.; Lee, J. S.; Choi, Y. D., Characterization of a Bifunctional Enzyme Fusion of Trehalose-6-Phosphate Synthetase and Trehalose-6-Phosphate Phosphatase of Escherichia coli. *Applied and Environmental Microbiology* **2000**, *66* (6), 2484.

8. Dueber, J. E.; Wu, G. C.; Malmirchegini, G. R.; Moon, T. S.; Petzold, C. J.; Ullal, A. V.; Prather, K. L. J.; Keasling, J. D., Synthetic protein scaffolds provide modular control over metabolic flux. *Nature Biotechnology* **2009**, *27* (8), 753-U107.

9. Fu, J.; Yang, Y. R.; Johnson-Buck, A.; Liu, M.; Liu, Y.; Walter, N. G.; Woodbury, N. W.; Yan, H., Multi-enzyme complexes on DNA scaffolds capable of substrate channelling with an artificial swinging arm. *Nature Nanotechnology* **2014**, *9* (7), 531-536.

10. Moon, T. S.; Dueber, J. E.; Shiue, E.; Prather, K. L., Use of modular, synthetic scaffolds for improved production of glucaric acid in engineered E. coli. *Metabolic Engineering* **2010**, *12* (3), 298-305.

11. Wilner, O. I.; Weizmann, Y.; Gill, R.; Lioubashevski, O.; Freeman, R.; Willner, I., Enzyme cascades activated on topologically programmed DNA scaffolds. *Nature Nanotechnology* **2009**, *4* (4), 249-254.

12. McConnell, S. A.; Cannon, K. A.; Morgan, C.; McAllister, R.; Amer, B. R.; Clubb, R. T.; Yeates, T. O., Designed Protein Cages as Scaffolds for Building Multienzyme Materials. *ACS Synthetic Biology* **2020**, *9* (2), 381-391.

13. Kim, S.; Hahn, J. S., Synthetic scaffold based on a cohesin-dockerin interaction for improved production of 2,3-butanediol in Saccharomyces cerevisiae. *Journal of Biotechnology* **2014**, 192 Pt A, 192-6.

14. Krzek, M.; van Beek, H. L.; Permentier, H. P.; Bischoff, R.; Fraaije, M. W., Covalent immobilization of a flavoprotein monooxygenase via its flavin cofactor. *Enzyme and Microbial Technology* **2016**, *82*, 138-143.

15. Velasco-Lozano, S.; Benítez-Mateos, A. I.; López-Gallego, F., Co-immobilized Phosphorylated Cofactors and Enzymes as Self-Sufficient Heterogeneous Biocatalysts for Chemical Processes. *Angewandte Chemie International Edition* **2017**, *56* (3), 771-775.

16. Hwang, E. T.; Lee, S., Multienzymatic Cascade Reactions via Enzyme Complex by Immobilization. *ACS Catalysis* **2019**, *9* (5), 4402-4425.

17. Matsumoto, R.; Kakuta, M.; Sugiyama, T.; Goto, Y.; Sakai, H.; Tokita, Y.; Hatazawa, T.; Tsujimura, S.; Shirai, O.; Kano, K., A liposome-based energy conversion system for accelerating the multi-enzyme reactions. *Physical Chemistry Chemical Physics* **2010**, *12* (42), 13904-13906.

18. Myhrvold, C.; Polka, J. K.; Silver, P. A., Synthetic Lipid-Containing Scaffolds Enhance Production by Colocalizing Enzymes. *ACS Synthetic Biology* **2016**, *5* (12), 1396-1403.

19. Alves, N. J.; Turner, K. B.; Daniele, M. A.; Oh, E.; Medintz, I. L.; Walper, S. A., Bacterial Nanobioreactors--Directing Enzyme Packaging into Bacterial Outer Membrane Vesicles. *ACS Applied Materials and Interfaces* **2015**, *7* (44), 24963-72.

20. Vriezema, D. M.; Garcia, P. M. L.; Sancho Oltra, N.; Hatzakis, N. S.; Kuiper, S. M.; Nolte, R. J. M.; Rowan, A. E.; van Hest, J. C. M., Positional Assembly of Enzymes in Polymersome Nanoreactors for Cascade Reactions. *Angewandte Chemie International Edition* **2007**, *46* (39), 7378-7382.

21. Azuma, Y.; Zschoche, R.; Tinzl, M.; Hilvert, D., Quantitative Packaging of Active Enzymes into a Protein Cage. *Angewandte Chemie International Edition* **2016**, *55* (4), 1531-1534.

22. Liang, M.; Frank, S.; Lunsdorf, H.; Warren, M. J.; Prentice, M. B., Bacterial microcompartment-directed polyphosphate kinase promotes stable polyphosphate accumulation in E. coli. *Biotechnology Journal* **2017**, *12* (3).

23. O'Neil, A.; Prevelige, P. E.; Douglas, T., Stabilizing viral nano-reactors for nerve-agent degradation. *Biomaterials Science* **2013**, *1* (8), 881-886.

24. Patterson, D. P.; Prevelige, P. E.; Douglas, T., Nanoreactors by Programmed Enzyme Encapsulation Inside the Capsid of the Bacteriophage P22. *ACS Nano* **2012**, *6* (6), 5000-5009.

25. Douglas, T.; Young, M., Viruses: making friends with old foes. *Science* **2006**, *312* (5775), 873-5.

26. Plegaria, J. S.; Kerfeld, C. A., Engineering nanoreactors using bacterial microcompartment architectures. *Current Opinion in Biotechnology* **2018**, *51*, 1-7.

27. Renggli, K.; Baumann, P.; Langowska, K.; Onaca, O.; Bruns, N.; Meier, W., Selective and Responsive Nanoreactors. *Advanced Functional Materials* **2011**, *21* (7), 1241-1259.

28. Aljabali, A. A. A.; Barclay, J. E.; Steinmetz, N. F.; Lomonossoff, G. P.; Evans, D. J., Controlled immobilisation of active enzymes on the cowpea mosaic virus capsid. *Nanoscale* **2012**, *4* (18), 5640-5645.

29. Koch, C.; Wabbel, K.; Eber, F. J.; Krolla-Sidenstein, P.; Azucena, C.; Gliemann, H.; Eiben, S.; Geiger, F.; Wege, C., Modified TMV Particles as Beneficial Scaffolds to Present Sensor Enzymes. *Frontiers in Plant Science* **2015**, *6* (1137).

30. Giessen, T. W.; Silver, P., A catalytic nanoreactor based on in vivo encapsulation of multiple enzymes in an engineered protein nanocompartment. *ChemBioChem* **2016**, *17* (20), 1931-1935

31. Lawrence, A. D.; Frank, S.; Newnham, S.; Lee, M. J.; Brown, I. R.; Xue, W.-F.; Rowe, M. L.; Mulvihill, D. P.; Prentice, M. B.; Howard, M. J.; Warren, M. J., Solution Structure of a Bacterial Microcompartment Targeting Peptide and Its Application in the Construction of an Ethanol Bioreactor. *ACS Synthetic Biology* **2014**, *3* (7), 454-465.

32. Patterson, D. P.; Schwarz, B.; Waters, R. S.; Gedeon, T.; Douglas, T., Encapsulation of an Enzyme Cascade within the Bacteriophage P22 Virus-Like Particle. *ACS Chemical Biology* **2014**, *9* (2), 359-365.

33. Abe, T.; Masai, E.; Miyauchi, K.; Katayama, Y.; Fukuda, M., A tetrahydrofolate-dependent O-demethylase, LigM, is crucial for catabolism of vanillate and syringate in Sphingomonas paucimobilis SYK-6. *Journal of Bacteriology* **2005**, *187* (6), 2030-2037.

34. Nishikawa, S.; Sonoki, T.; Kasahara, T.; Obi, T.; Kubota, S.; Kawai, S.; Morohoshi, N.; Katayama, Y., Cloning and sequencing of the Sphingomonas (Pseudomonas) paucimobilis gene essential for the O demethylation of vanillate and syringate. *Applied and Environmental Microbiology* **1998**, *64* (3), 836-842.

35. Becker, J.; Wittmann, C., A field of dreams: Lignin valorization into chemicals, materials, fuels, and health-care products. *Biotechnology Advances* **2019**, *37* (6), 107360.

36. Boerjan, W.; Ralph, J.; Baucher, M., Lignin biosynthesis. *Annual Review of Plant Biology* **2003**, *54*, 519-46.

37. Sjöström, E., Chapter 4 - LIGNIN. In Wood Chemistry (Second Edition), Sjöström, E., Ed. Academic Press: San Diego, **1993**; pp 71-89.

38. Zhang, R.; Zhao, C.-H.; Chang, H.-C.; Chai, M.-Z.; Li, B.-Z.; Yuan, Y.-J., Lignin valorization meets synthetic biology. *Engineering in Life Sciences* **2019**, *19* (6), 463-470.

39. Wu, W.; Dutta, T.; Varman, A. M.; Eudes, A.; Manalansan, B.; Loque, D.; Singh, S., Lignin Valorization: Two Hybrid Biochemical Routes for the Conversion of Polymeric Lignin into Value-added Chemicals. *Scientific Reports* **2017**, *7* (1), 8420.

40. Mori, K.; Kamimura, N.; Masai, E., Identification of the protocatechuate transporter gene in Sphingobium sp. strain SYK-6 and effects of overexpression on production of a value-added metabolite. *Applied Microbiology and Biotechnology* **2018**, *102* (11), 4807-4816.

41. Scaglione, F.; Panzavolta, G., Folate, folic acid and 5-methyltetrahydrofolate are not the same thing. *Xenobiotica* **2014**, *44* (5), 480-8.

42. Pietrzik, K.; Bailey, L.; Shane, B., Folic acid and L-5-methyltetrahydrofolate: comparison of clinical pharmacokinetics and pharmacodynamics. *Clinical Pharmacokinetics* **2010**, *49* (8), 535-48.

43. Obeid, R.; Holzgreve, W.; Pietrzik, K., Is 5-methyltetrahydrofolate an alternative to folic acid for the prevention of neural tube defects? *Journal of Perinatal Medicine* **2013**, *41* (5), 469-83.

44. Cassidy-Amstutz, C.; Oltrogge, L.; Going, C. C.; Lee, A.; Teng, P.; Quintanilla, D.; East-Seletsky, A.; Williams, E. R.; Savage, D. F., Identification of a Minimal Peptide Tag for in Vivo and in Vitro Loading of Encapsulin. *Biochemistry* **2016**, *55* (24), 3461-3468.

45. Künzle, M.; Eckert, T.; Beck, T., Metal-Assisted Assembly of Protein Containers Loaded with Inorganic Nanoparticles. *Inorganic Chemistry* **2018**, *57* (21), 13431-13436.

46. Williams, E. M.; Jung, S. M.; Coffman, J. L.; Lutz, S., Pore Engineering for Enhanced Mass Transport in Encapsulin Nanocompartments. *ACS Synthetic Biology* **2018**, *7* (11), 2514-2517.

47. Gazzali, A. M.; Lobry, M.; Colombeau, L.; Acherar, S.; Azaïs, H.; Mordon, S.; Arnoux, P.; Baros, F.; Vanderesse, R.; Frochot, C., Stability of folic acid under several parameters. *European Journal of Pharmaceutical Sciences* **2016**, *93*, 419-430.

48. De Brouwer, V.; Zhang, G. F.; Storozhenko, S.; Straeten, D. V.; Lambert, W. E., pH stability of individual folates during critical sample preparation steps in prevision of the analysis of plant folates. *Phytochemical Analysis* **2007**, *18* (6), 496-508.

49. Kohler, A. C.; Mills, M. J. L.; Adams, P. D.; Simmons, B. A.; Sale, K. L., Structure of aryl-O-demethylase offers molecular insight into a catalytic tyrosine-dependent mechanism. *Proceedings of the National Academy of Sciences USA* **2017**, *114* (16), E3205.

50. Harada, A.; Kamimura, N.; Takeuchi, K.; Yu, H. Y.; Masai, E.; Senda, T., The crystal structure of a new O-demethylase from Sphingobium sp. strain SYK-6. *The FEBS Journal* **2017**, *284* (12), 1855-1867.

51. Sutter, M.; Boehringer, D.; Gutmann, S.; Guenther, S.; Prangishvili, D.; Loessner, M. J.; Stetter, K. O.; Weber-Ban, E.; Ban, N., Structural basis of enzyme encapsulation into a bacterial nanocompartment. *Nature Structural & Molecular Biology* **2008**, *15* (9), 939-947.

52. Tamura, A.; Fukutani, Y.; Takami, T.; Fujii, M.; Nakaguchi, Y.; Murakami, Y.; Noguchi, K.; Yohda, M.; Odaka, M., Packaging Guest Proteins into the Encapsulin Nanocompartment from Rhodococcus erythropolis N771. *Biotechnology and Bioengineering* **2015**, *112* (1), 13-20.

53. Williams, E.; Zhao, H.; Jenkins, M.; Juneja, P.; Lutz, S., One-pot dynamic kinetic resolution via chemo-biocatalytic cascade reaction in protein nano-compartments. *Journal of the American Chemical Society* **2020**, Manuscript in preparation.

54. Reed, L. S.; Archer, M. C., Oxidation of tetrahydrofolic acid by air. *Journal of Agricultural and Food Chemistry* **1980**, *28* (4), 801-805.

55. Chippel, D.; Scrimgeour, K. G., Oxidative degradation of dihydrofolate and tetrahydrofolate. *Canadian Journal of Biochemistry* **1970**, *48* (9), 999-1009.

56. Rosini, E.; D'Arrigo, P.; Pollegioni, L., Demethylation of vanillic acid by recombinant LigM in a one-pot cofactor regeneration system. *Catalysis Science & Technology* **2016**, *6* (21), 7729-7737.

57. Hummel, W.; Gröger, H., Strategies for regeneration of nicotinamide coenzymes emphasizing self-sufficient closed-loop recycling systems. *Journal of Biotechnology* **2014**, *191*, 22-31.

58. Ferrer, J.-L.; Ravanel, S.; Robert, M.; Dumas, R., Crystal Structures of Cobalamin-independent Methionine Synthase Complexed with Zinc, Homocysteine, and Methyltetrahydrofolate. *Journal of Biological Chemistry* **2004**, *279* (43), 44235-44238.

59. Gonzalez, J. C.; Banerjee, R. V.; Huang, S.; Sumner, J. S.; Matthews, R. G., Comparison of cobalamin-independent and cobalamin-dependent methionine synthases from Escherichia coli: two solutions to the same chemical problem. *Biochemistry* **1992**, *31* (26), 6045-6056.

60. Pejchal, R.; Ludwig, M. L., Cobalamin-Independent Methionine Synthase (MetE): A Face-to-Face Double Barrel That Evolved by Gene Duplication. *PLOS Biology* **2005**, *3* (2), e31.

61. Schindelin, J.; Arganda-Carreras, I.; Frise, E.; Kaynig, V.; Longair, M.; Pietzsch, T.; Preibisch, S.; Rueden, C.; Saalfeld, S.; Schmid, B.; Tinevez, J.-Y.; White, D. J.; Hartenstein, V.; Eliceiri, K.; Tomancak, P.; Cardona, A., Fiji: an open-source platform for biological-image analysis. *Nature Methods* **2012**, *9* (7), 676-682.

Chapter 4:

User-triggered Assembly of Encapsulin Nanocontainers for the *in vitro* Loading of Non-native Cargo Under Mild Conditions

4.1 – Introduction

Naturally occurring proteinaceous cages are often inherently designed to encapsulate specific cargoes related to their respective biological functions, including genomic materials, proteins, and metal ions, to name a few (1). For nucleic acid and protein-based cargoes, encapsulation is typically directed *in vivo* using prescribed intermolecular interactions between the cargo in question and the luminal faces of the individual coat protein subunits from which the cage is constructed. In recent decades, the ability to specifically charge both naturally occurring and engineered protein containers with non-native cargoes by exploiting these pre-programmed interactions has gained increasing interest as a powerful tool for the generation of biocompatible nanomaterials and nano-vehicles with diverse applications in medicine, industry, and biotechnology. Commonly employed non-native cargoes include fluorogenic reporter proteins (2-6), enzymes (3, 7-11), nucleic acids (12-14), metal nanoparticles (15-18), small molecule drugs or cytotoxins (12, 19-22), and chemocatalysts (23-25). Similar to the process of native cargo loading, many foreign cargoes can be packaged within protein nanocages during the natural cage assembly process *in vivo* by simply co-expressing the cargo of interest and the cage’s coat protein subunits heterologously (26-29). However, *in vivo* cargo loading has certain limitations, such as difficulties with regard to controlling the amount of packaged cargo per container, non-specific encapsulation of unwanted cargoes, and the inability to load cargoes that cannot be synthesized by the host organism, such as chemocatalysts and small molecule drugs (6, 30).

As an alternative to the *in vivo* packaging of foreign cargoes, many protein cage assemblies can be artificially loaded *in vitro* by prompting the cage to first disassemble into its individual protomeric subunits (or into smaller oligomeric structures such as dimers and pentamers) and then subsequently reassemble in the presence of the cargo of interest. This strategy presents several advantages over *in vivo* methods as it allows for the encapsulation of non-biocompatible cargoes, and in many cases, containers can be quantitatively packaged with distinct cargo concentrations by modulating the cargo to protomer stoichiometry prior to prompting container reassembly (5, 31-33). For some cage architectures, mild modulation of solution pH and ionic strength is sufficient to induce reversible disassembly of the nanocage structure, with the virus-like particle derived from cowpea chlorotic mottle virus (CCMV) serving as perhaps the best studied example of this phenomenon (34-36). For more recalcitrant nanocontainers, however, harsher treatments such as extreme pH values and the use of chaotropic salts are needed to incite container disassembly (37-39). Container reassembly is subsequently accomplished using dialysis or dilution to neutralize the solution pH or reduce the solution concentration of chaotropes.

Following the resolution of several atomic level structures in the mid to late 2000s, encapsulin nanocontainers have been increasingly utilized as modular chassis for the encapsulation and/or delivery of foreign cargoes due to their uniform sizes, their ability to natively encapsulate proteinaceous cargoes using known cargo loading peptide (Clp) sequences, and their generally robust natures (40). However, no native disassembly mechanisms have been identified for any of the encapsulins assemblies characterized to date. Recent accounts have instead shown that encapsulin nanocontainers, specifically those originating from the hyperthermophilic bacterium *Thermotoga maritima* (TmE), are among the ranks of other recalcitrant cage assemblies, requiring extreme pH values (i.e. ≤ 2 or ≥ 13) or high concentrations of either guanidine hydrochloride (GdnHCl, 7 M) or urea (12 M) denaturants in order to undergo reversible disassembly (18, 38). Using these methods, Cassidy-Amstutz *et al.* were able to artificially load GdnHCl-disassembled TmE nanocontainers with superfolder GFP-Clp fusion proteins (sfGFP-Clp) (38). Several years later, Kunzle *et al.* followed this same general strategy to charge acid-disassembled TmEs with gold nanoparticles possessing chemically synthesized Clp peptides covalently appended to their surfaces (18).

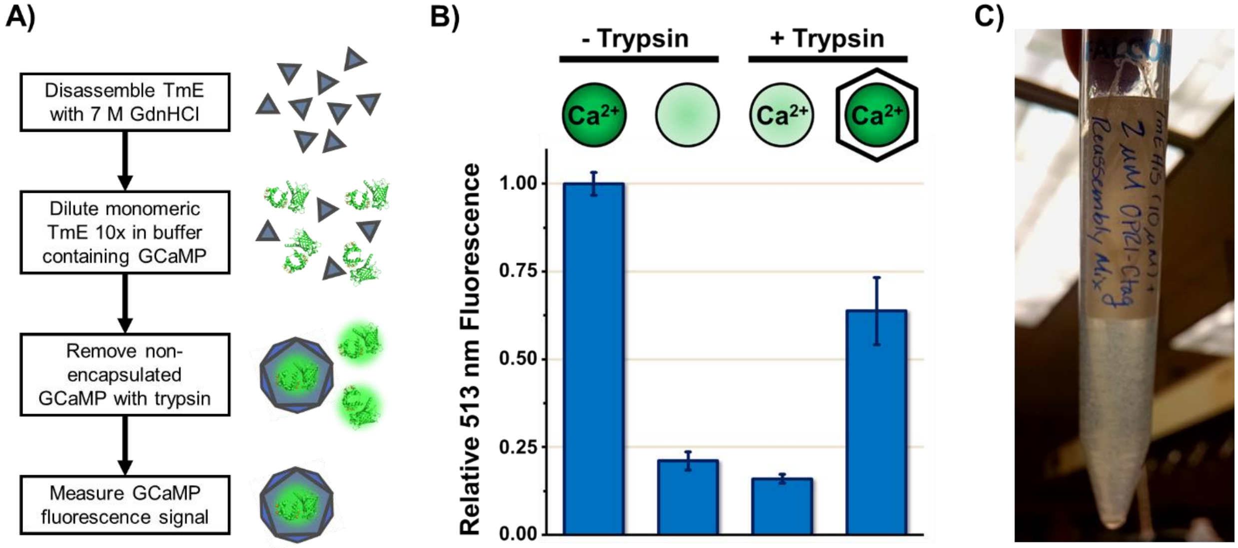


Figure 4.1 – Non-native cargo loading of TmE nanocontainers using GdnHCl-mediated disassembly. A) Schematic of the process used to perform *in vitro* loading of a GCaMP reporter protein. B) Assessment of retained GCaMP fluorescence with and without trypsin digestion. Addition of the calcium trigger used to initiate GCaMP fluorescence is indicated by the " Ca^{2+} " labels within the circles representing the GCaMP reporters above the chart. GCaMP proteins loaded into TmE via GdnHCl-mediated disassembly/reassembly are indicated by the hexagonal black shell around the GCaMP label (right-hand data bar). C) Attempted GdnHCl-mediated cargo loading of an enzyme, OPR1, from the OYE family of flavoenzyme, which resulted in rapid precipitation of the enzyme.

While both the pH and denaturant-based methods used to charge encapsulins with foreign cargoes previously proved effective, between 5 to 40% of the starting nanocontainer populations were lost during the different disassembly and reassembly processes, presumably due to misfolding and aggregation of the encapsulin protomers (38). Further, residual denaturants remain present during the container reassembly phase, which can be detrimental to certain cargoes of interest. Initial attempts to recapitulate the reported *in vitro* cargo loading of TmEs in our own lab using GdnHCl to prompt container disassembly proved successful when GCaMP, a green fluorescent protein variant exhibiting Ca^{2+}-dependent fluorescence (41), was used as a non-native reporter cargo (**Figure 4.1A**, see **Materials and Methods** for full details). Following the disassembly and reassembly of TmEs in the presence of excess GCaMP-Clp fusion proteins, approximately 60% of the original GCaMP population was found to be protected from proteolytic cleavage, indicating successful encapsulation within assembled TmEs in accordance with previous reports (**Figure 4.1B**) (42), while non-encapsulated GCaMP was completely degraded under identical exposure to protease. However, when the same *in vitro* loading procedure was repeated using several flavoprotein enzymes originating from the Old Yellow Enzyme family, rampant precipitation of the cargo proteins was repeatedly observed during the reassembly phase (**Figure 4.1C**). In light of both the previously published results and the experiments conducted in our own lab, the ability to perform the *in vitro* loading of non-native cargoes under benign solvent conditions thus represents an unmet challenge for encapsulin-based nanocontainer systems.

As an alternative to the process of sequentially disassembling and reassembling protein cage structures, herein we report a novel methodology of preventing the native *in vivo* assembly of heterologously expressed TmE nanocontainers by recombinantly grafting the *E. coli* maltose binding protein (MBP) onto the N-terminus of the TmE protomer gene (i.e. MBP-TmE). Specifically, fused MBP proteins were introduced to exploit the limited volumetric constraints within the interior of assembled encapsulin cages, thus generating sufficiently strong steric repulsion between adjacent MBP-TmE fusion proteins to prevent full container assembly due to the N-terminus of each TmE protomer being natively oriented into the container lumen (**Figure 4.2A**) (43). Subsequent proteolytic liberation of the MBP proteins using protease cleavage sites encoded in a flexible linker interspaced between the MBP and TmE proteins was used to alleviate the steric obstructions, permitting user-defined temporal control over nanocontainer assembly (**Figure 4.2B**). This new method, referred to as cleavage-prompted assembly (CPA), was investigated as an alternative method for the facile loading of non-native protein cargoes under benign solvent conditions.

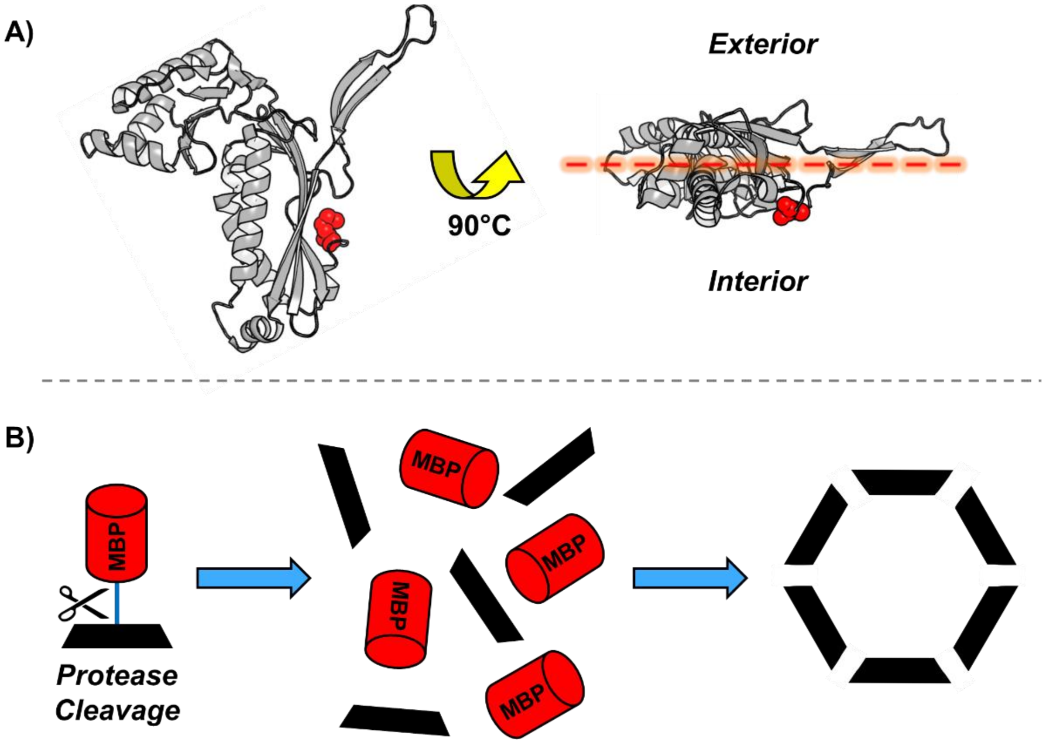


Figure 4.2 – A) TmE protomer atomic structure with the N-terminal methionine residue depicted as red spheres (PDB ID: 3DKT). B) Schematic representation of the proposed CPA method of user-prompted nanocontainer assembly *in vitro*. Encapsulin protomers are represented as black tiles while tethered MBP proteins are represented as red cylinders.

4.2 – Results and Discussion

4.2.1 – TmE Critical Assembly Concentration Determination

Though potential dynamic behavior has been speculated for certain encapsulin species (44), no native disassembly or reassembly pathway has been formally established for the subset of currently characterized encapsulin nanocontainers, as stated previously. This lack of dynamic behavior is in stark contrast to that of the structurally homologous viral capsids presumed to share a common ancestor with encapsulin cages, such as the HK97 bacteriophage (43, 45), which readily undergo assembly and disassembly *in vivo* as part of the infectious viral lifecycle. In nature, viral capsids are assembled in a concentration-dependent manner, and in many cases a distinct minimum concentration of viral coat proteins must be achieved in order to initiate the polymerization-based assembly of full capsids structures (46-49). This minimum concentration, referred to as the apparent critical assembly concentration ($\mathrm{CAC_{app}}$), is an intrinsic property characteristic of different viral species (50). Based off of these studies, we attempted to determine the $\mathrm{CAC_{app}}$ value for TmE nanocontainers both to better understand the native assembly behavior of encapsulins in general and to enhance the effectiveness of our proposed CPA strategy.

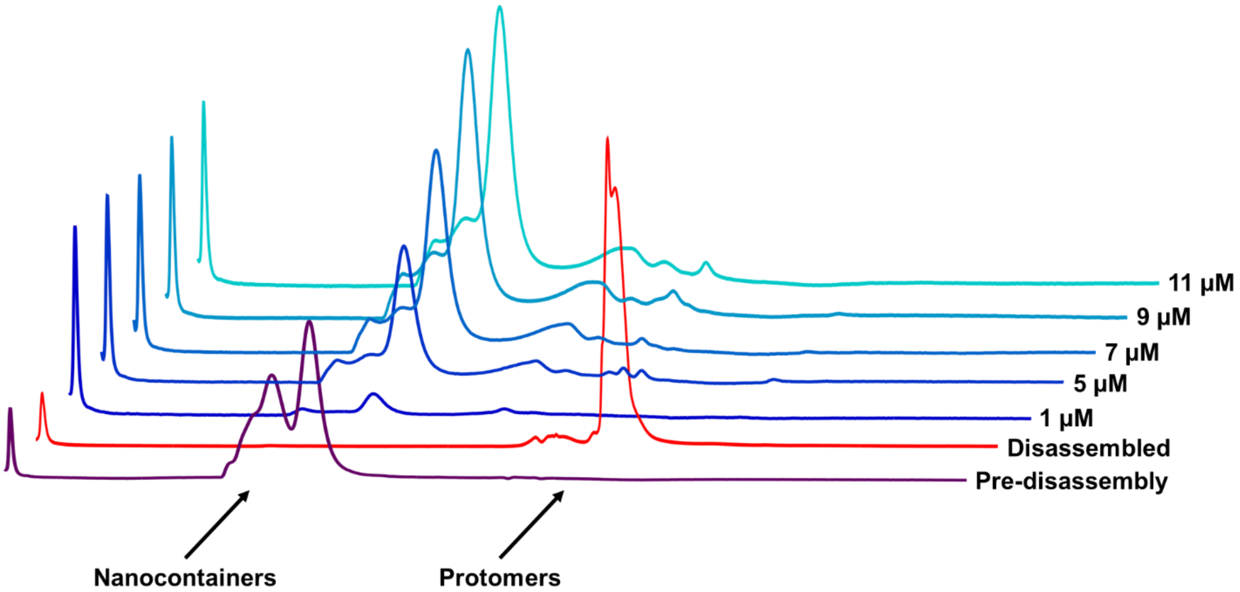


Figure 4.3 – Attempted determination of the $\mathrm{CAC_{app}}$ value for TmE via analytical HPLC-SEC. All of the reassembly mixtures are normalized to one another while the "Pre-disassembly" and "Disassembled" samples were not normalized to prevent them from overshadowing the reassembly mixture data.

Determination of the $\mathrm{CAC_{app}}$ value for TmE encapsulins was attempted by following a recently reported methodology used to experimentally elucidate a similar value in the low micromolar range for VLPs derived from the MS2 bacteriophage (50). Specifically, purified TmE nanocontainers were completely disassembled in low pH buffer and the resulting protomers were diluted to varying final concentrations between 1 and 11 μM. Reassembly of encapsulin containers in the various samples was then initiated by dialyzing the samples into a buffered solution at pH 7.5. The dialyzed samples were then assessed via HPLC-based analytical SEC (**Figure 4.3**). Interestingly, untreated encapsulin nanocontainers presented two distinct peaks in the resulting chromatograms, possibly due to aggregation or due to the presence of a subpopulation of incompletely assembled full container shells. Acid-based treatment of the TmEs subsequently caused a dramatic downfield shift in the sample elution volume, indicating successful disassembly of the nanocontainers. However, diluted samples of encapsulin protomers exhibited elution volumes directly corresponding to the size of fully assembled nanocontainers for all concentrations tested upon dialysis into pH 7.5 buffer. These results indicate that TmE nanocontainers have a $\mathrm{CAC_{app}}$ value less than 1 μM (below the limit of detection for our HPLC-SEC assay), which in turn suggests that the intermolecular interactions that govern the assembly of the encapsulin cages are likely quite strong. This strong preference for nanocontainer formation at low protomer concentrations seems logical given that encapsulins are designed to function as nanoscale organelles *in vivo*, and as such do not necessarily require the ability to undergo reversible disassembly and reassembly to fulfill their specific biological roles.

4.2.2 – CPA Proof-of-Principle

To determine if the introduction of steric bulk within the TmE lumen could prevent the formation of fully assembled nanocontainers, the TmE protomer gene was recombinantly fused to the C-terminal end of the *E. coli* MBP gene. A 24 amino acid flexible linker sequence was encoded between the MBP and TmE genes to provide a small degree of separation between the translated proteins. Additionally, a Factor Xa protease recognition site (recognition sequence "…IEGR↓…") was encoded within the flexible linker 4 amino acids upstream of TmE gene's N-terminal methionine residue (**Figure 4.4A**). Following heterologous expression in *E. coli* BL21(DE3) cells, amylose resin affinity chromatography was performed as an initial test to assess whether the MBP-TmE fusion proteins could still adopt fully assembled cage structures. Specifically, the proper assembly of nanocages was expected to effectively sequester the fused MBP proteins within the encapsulin lumen, thereby abrogating the ability of the MBP domains to interact and bind to the amylose resin. **Figure 4.4B** shows, however, that high molecular weight MBP-TmE fusion protein (73.4 kDa) was highly enriched upon amylose chromatography and could be easily isolated in high yield with excess MBP (40.3 kDa) as the predominant co-eluted species. These results indicate both that the fused MBP domains are solvent exposed and that encapsulin assembly was thus interrupted.

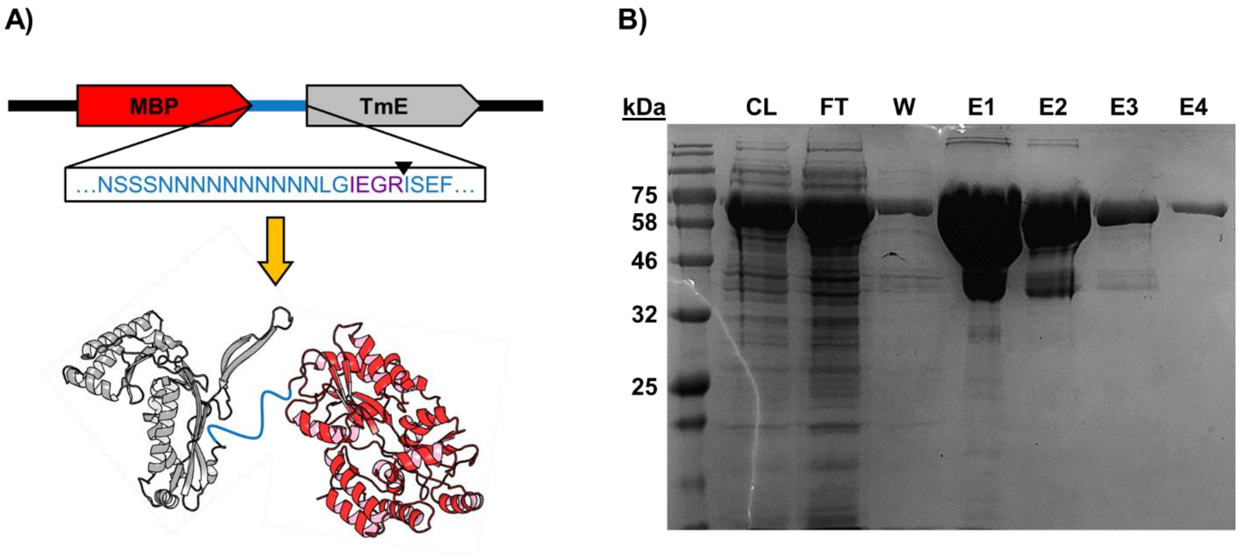
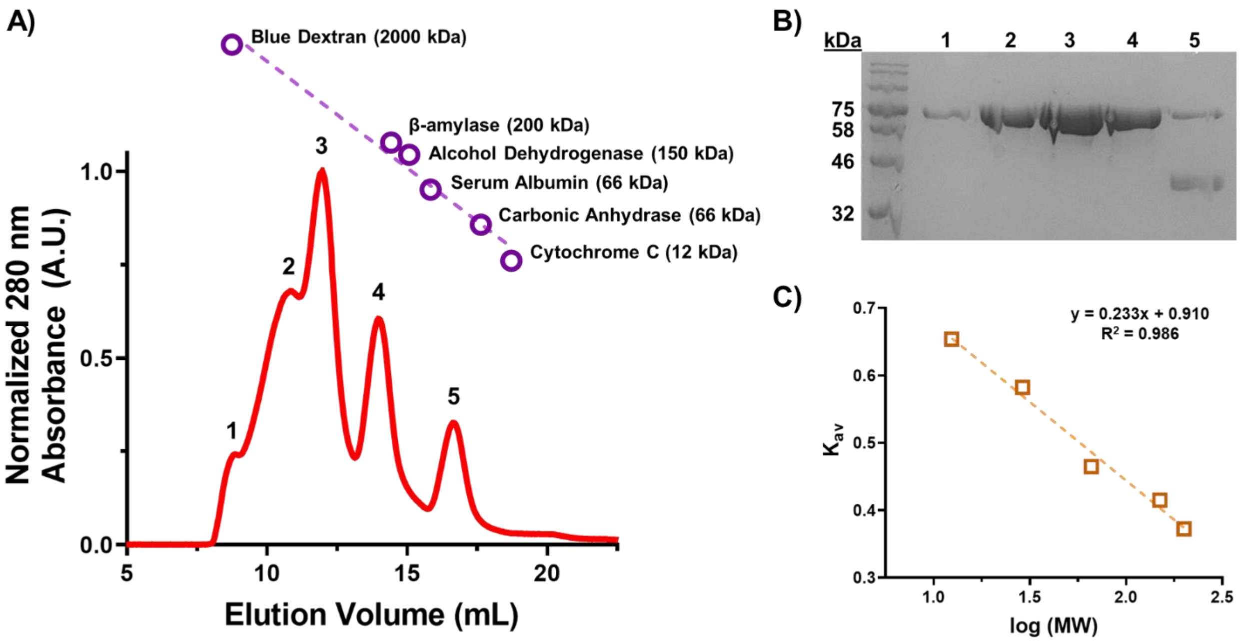


Figure 4.4 – A) Cartoon representations of the MBP-TmE tandem genes and the corresponding translated protein products (TmE PDB ID: 3DKT; MBP PDB ID: 1ANF). The 24 amino acid flexible linker joining the MBP and TmE genes is depicted in the pop-out text box. The Factor Xa "IGER" cleavage site is colored in purple text with the site of proteolytic cleavage following the terminal arginine residue indicated by the black wedge. B) SDS-PAGE image of fractions collected during amylose-affinity purification of the expressed MBP-TmE protein fusion. The lane labels represent the clarified cell lysate (CL), the column flow-through (FT), a subsequent wash of the resin bed (W), and sequential elution fractions with buffer containing 10 mM maltose (E1-E4).



Number	Volume (mL)	Calculated MW (kDa)	Predicted Oligomeric State	Relative Ratio in State
1	8.9	7362 ± 4556	100 ± 62	8%
2	10.8	2089 ± 1116	28 ± 15	24%
3	12.0	1000 ± 485	14 ± 7	35%
4	14.0	273 ± 110	3.7 ± 1.5	21%
5	16.7	48 ± 15	0.7 ± 0.2	11%

Figure 4.5 and **Table 4.1** – A) Analytical SEC profile of MBP-TEV. A plot of the retention times for the protein collection from the SEC molecular weight standard kit are plotted above the MBP-TEV chromatogram. B) SDS-PAGE analysis of the protein composition of each numbered peak from the SEC chromatogram. C) A plot of the distribution coefficient (K_{av}) versus the logarithm of the corresponding molecular weight for each of the proteins from the standard kit. The equation derived from the linear fit of the K_{av} versus log (MW) was used to generate the molecular weight interpretations of the MBP-TmE peaks presented in **Table 4.1**.

While the amylose affinity chromatography results promisingly imply that encapsulin assembly was disrupted due to the steric bulk of the fused MBP domains, the extent of the disruption remains unclear. To further determine whether MBP-TmE fusion proteins could still self-assemble to some degree or whether they existed as entirely monomeric proteins in solution, purified fusion protein was subjected to size exclusion chromatography analysis. The resulting chromatogram, along with the corresponding SDS-PAGE fractions generated from SEC profiling

(**Figure 4.5A** and **4.5B**), shows multiple distinct elution peaks spanning a broad range of molecular weights. Comparison of the elution peak retention times with those of protein standards run over the same chromatography column indicate that the MBP-TmE fusion proteins self-assemble into a series of higher order oligomeric states, though the large nature of these states coupled with the intrinsic error of the standard curve made precise characterization of the oligomeric states' molecular weight values difficult (**Figure 4.5C** and **Table 4.1**). The highest molecular weight peak shares a similar retention time with fully assembled encapsulin nanocontainers. Given that these protein samples were capable of binding to the amylose affinity resin prior to SEC analysis, this shared retention time indicates that the MBP-TmE fusion proteins are possibly dynamic, partially constructed cage structures, undergoing equilibrium-driven assembly and disassembly processes in solution.

Having shown that the MBP-TmE fusion proteins exist in a series of incompletely assembled oligomeric states, we next sought to determine whether treatment of the fusion protein with protease could specifically trigger CPA *in vitro*. As stated above, a Factor Xa cleavage site was encoded 4 amino acids upstream of the TmE's initial methionine residue. We subsequently tested the cleavage efficiency of both commercial Factor Xa and bovine trypsin and found that both proteases were capable of cleaving our fusion proteins *in vitro*. However, we subsequently elected to treat our MBP-TmE fusions with bovine trypsin in order to initiate CPA as an alternative to Factor Xa for several practical reasons: 1) previous studies from our own laboratory have shown TmE to be resistant to trypsin protease treatment (42); 2) *in vitro* trypsin cleavage is typically rapid with non-encapsulated test proteins often being degraded in less than one hour at room temperature while Factor Xa treatments tend to require longer incubation periods **(42)**; 3) commercially

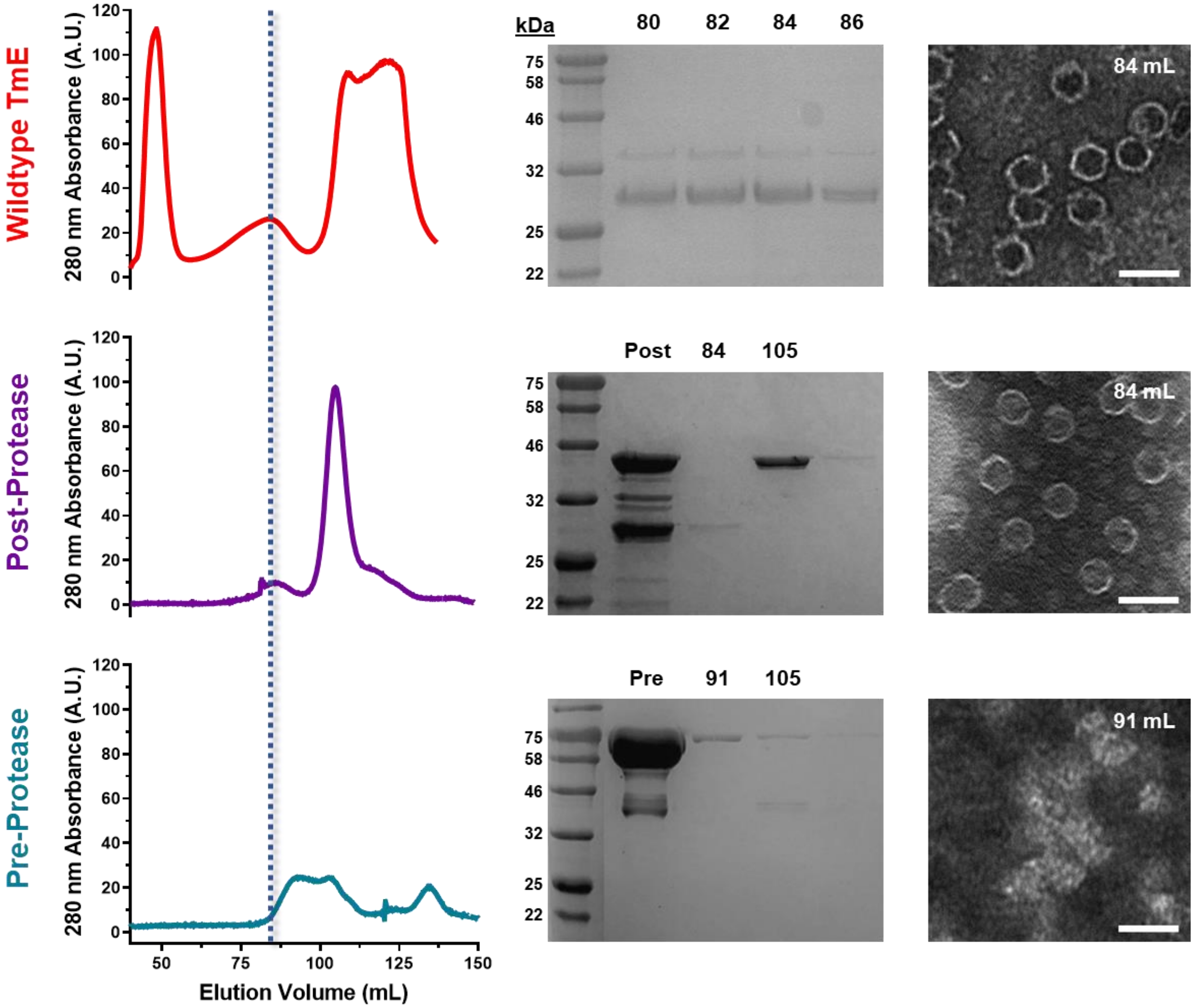


Figure 4.6 – Comparison of SEC chromatograms generated from MBP-TmE prior to (blue trace, bottom) and after (purple trace, middle) treatment with trypsin. A chromatogram of wildtype TmE (red trace, top) is included for reference. The peak maximum corresponding to the elution of assembled TmE nanocontainers is denoted by the vertical dashed blue line intersecting all three chromatograms. SDS-PAGE gel images depicting relevant fractions are immediately to the right of each chromatogram. The numerical values above each SDS-PAGE lane represent the elution volume (in mL) corresponding to sample lane below. The "Pre" and "Post" labels represent the undigested and digested MBP-TmE samples, respectively, prior to SEC analysis. TEM images corresponding to samples collected at specific elution volumes from each SEC run are depicted to the far right of each chromatogram. Each TEM image is labeled in the upper right corner with the elution fraction from which it was taken. Scale bars = 40 nm.

purchased bovine trypsin (~24 kDa) is approximately half the size of commercial bovine Factor Xa (~43 kDa, activated form), which we reasoned would likely be beneficial for achieving effective cleavage of the MBP domains given the semi-preassembled nature of the TmE containers (51, 52). SEC analysis using a larger, preparative scale chromatography column shows that purified MBP-TmE presents a series of closely overlapping absorbance peaks between the 85 and 135 mL elution volume range (**Figure 4.6**). Corresponding SDS-PAGE assessment of these peaks shows that the majority of the MBP-TmE fusion protein elutes in the initial peak whose maximum is centered around ~91 mL while a combination of residual fusion protein and contaminant MBP eluted in the adjacent peak centered around ~105 mL. However, the resulting chromatogram generated upon treatment of MBP-TmE fusion protein with trypsin for 15 minutes at 37°C shows several dramatic shifts as a new peak emerges centered around the ~84 mark while the peak as the ~105 mL mark grows significantly in intensity. SDS-PAGE assessment of these fractions reveals that the former peak contains a faint protein band consistent with the size of wildtype TmE while the latter peak presents a band consistent with the size of free MBP, indicating that successful scission of the two proteins has occurred. Further, the ~84 mL retention peak directly corresponds to the established retention volume observed for purified wildtype TmE (**Figure 4.6**).

In conjunction with the SEC and SDS-PAGE analyses, select fractions collected from the separate SEC purifications were examined via TEM imaging. Images collected for samples originating from the ~91 mL elution fraction of the pre-digestion MBP-TmE purification consistently exhibited what appear to be amorphous protein assemblies of indeterminate structure (**Figure 4.6**). In contrast, image samples collected from the ~84 mL elution fractions for both the trypsin treated MBP-TmE and wildtype TmE purifications show clearly defined encapsulin assemblies corresponding to literature established sizes and morphologies (43). These data further indicate that CPA is a viable methodology for the user-initiated construction of whole encapsulin nanocages.

4.2.3 – Optimization of CPA Method Using TEV Protease

Following our initial tests proving the viability of the CPA methodology, we subsequently exchanged the Factor Xa cleavage site within the flexible linker interspersed between the MBP and TmE genes with a tobacco etch virus (TEV) protease cleavage site (**Figure 4.7A**). This alteration was deemed necessary as trypsin, while rapid and effective, is a non-specific protease which could easily degrade intended non-native cargo proteins during the CPA process. This point was exemplified experimentally as GCaMP, the reporter protein used previously in our *in vitro* loading tests, was rapidly degraded in the presence of trypsin (**Figure 4.7B**). By comparison, TEV protease is well known to exhibit strict selectivity for its ENLYFQ↓G/S cleavage recognition sequence, thus reducing the potential for cargo protein loss (53). Indeed, treatment of GCaMP with TEV protease led to no observable degradation of the protein by SDS-PAGE analysis nor a

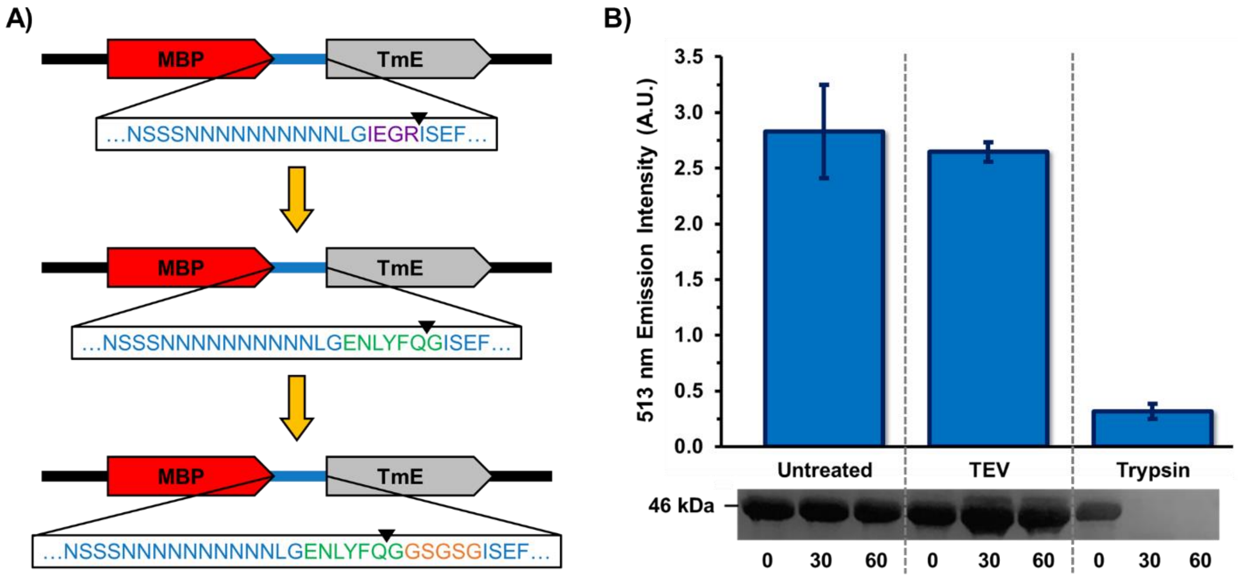


Figure 4.7 – A) Successive optimization of the flexible linker sequence between the MBP and TmE genes. MBP-*TEV*-TmE was formed by exchanging the Factor Xa cleavage (purple IGER, top) site for a TEV cleavage site (green ENLYFQG, middle). MBP-*TEVext*-TmE was formed by adding 5 additional residues (orange GSGSG, bottom) after the TEV cleavage site. B) Comparison of the retained fluorescence signal for GCaMP following treatment with either TEV protease or trypsin. The SDS-PAGE gel below the chart depicts a time course assessment of GCaMP digestion at 0, 30, and 60 minutes following the addition of protease.

reduction in its fluorescence signal (**Figure 4.7B**). It is worth noting that while Factor Xa also offers higher selectivity relative to trypsin and could have been used with the corresponding cleavage site already present in our MBP-TmE fusion construct, TEV protease adopts a similar size and fold relative to trypsin (i.e. ~27 kDa), and it requires significantly less upstream processing compared to Factor Xa, thus making it easier and faster to produce via recombinant methods (51, 52, 54, 55).

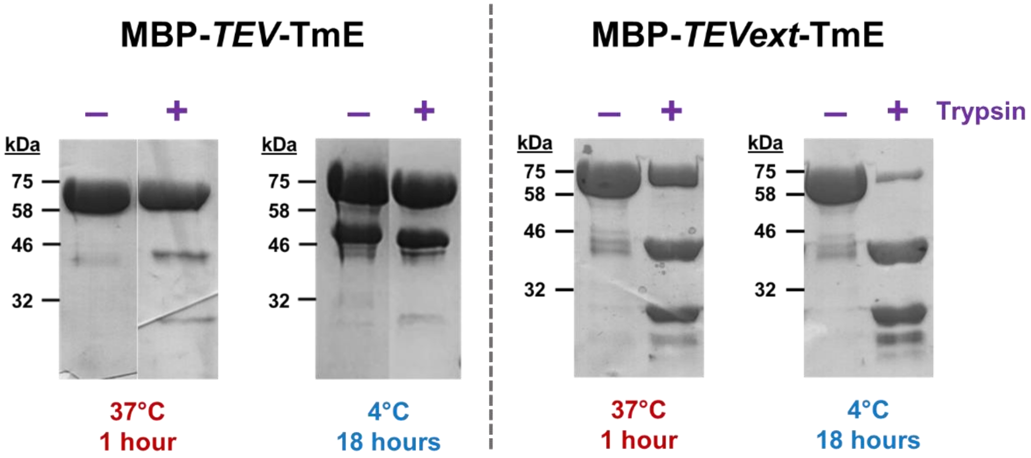


Figure 4.8 – SDS-PAGE assessments of proteolytic cleavage efficiency using TEV protease at varying temperatures and over varying timeframes. Protein molecular weights are as follows: MBP-*TEVext*-TmE (74.1 kDa); cleaved MBP (42.8 kDa); TmE (30.5 kDa).

Proteolytic cleavage tests performed on the new fusion variant containing the TEV cleavage site (i.e. MBP-*TEV*-TmE) with wildtype TEV protease show that while cleavage does occur, it is significantly slower with minimal liberation of TmE observed after either 1 hour at 37°C or 18 hours at 4°C (**Figure 4.8**). Increasing the concentration of TEV protease yielded no appreciable increases in the degree of fusion protein cleavage under the same reaction conditions (not shown). We attributed this lack of proteolytic activity to either the reported slow processivity of TEV protease (54, 56), or to the 4 amino acid spacer peptide between the C-terminal end of the TEV cleavage site and the N-terminal methionine of the TmE protein leaving insufficient physical space for TEV protease to effectively bind to its recognition site. To address these issues, we first redesigned our MBP-*TEV*-TmE fusion protein by recombinantly adding a 5 amino acid GSGSG spacer following the TEV recognition site to generate an extended MBP-*TEV*-TmE variant (i.e. MBP-*TEVext*-TmE, **Figure 4.7A**). Subsequent cleavage tests with the MBP-*TEVext*-TmE variant showed enhanced scission, with > 50% cleavage occurring after a 1 hour incubation at 37°C and > 70% cleavage occurring after an 18 hour incubation at 4°C (**Figure 4.8**).

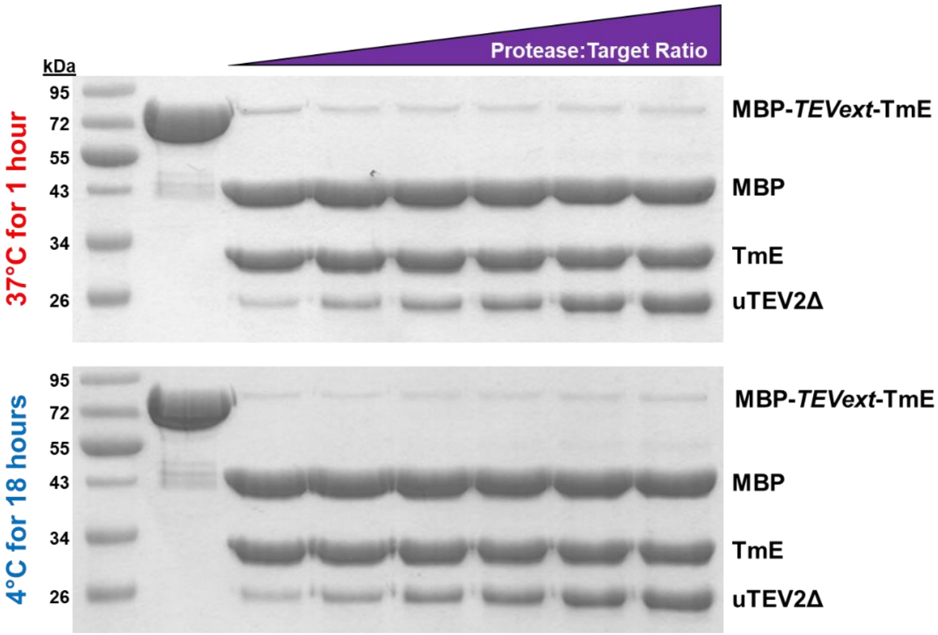


Figure 4.9 - SDS-PAGE assessments of proteolytic cleavage efficiency using the uTEV2Δ protease variant. Cleavage efficiency was assessed at varying temperatures, over varying timeframes, and at varying ratios of protease relative to the MBP-*TEVext*-TmE fusion protein target (10 μM). The leftmost sample lane (directly to the right of the protein ladder lane) contains no uTEV2Δ, the second sample lane contains 2.5 μM uTEV2Δ (0.25:1.0 ratio), and the rightmost sample lane 20 μM uTEV2Δ (2.0:1.0 ratio).

Sanchez and Ting recently reported a series of engineered TEV protease variants displaying accelerated cleavage rates relative to wildtype TEV (56). One TEV variant containing two point mutations and a C-terminal truncation following residue 219, referred to as uTEV2Δ, yielded turnover rates approximately 6-fold faster than wildtype TEV. We thus recombinantly introduced the same point mutations and C-terminal truncation into the wildtype TEV protease gene to recreate the uTEV2Δ variant ourselves. Subsequent cleavage tests of MBP-*TEVext*-TmE using heterologously expressed and purified uTEV2Δ show greatly enhanced cleavage with >90% of the initial fusion protein target degraded after 1 hour at 37°C or 18 hours at 4°C (**Figure 4.9**). Collectively, the combination of the extended linker region in MBP-*TEVext*-TmE and the accelerated cleavage rates of uTEV2Δ are able to adequately achieve TmE liberation on par with the trypsin protease used previously.

4.2.4 – Attempted Non-native Cargo Loading via CPA

We next sought to attempt *in vitro* cargo loading using the CPA method optimized for TEV protease in conjunction with the MBP-*TEVext*-TmE fusion protein. Using the GCaMP reporter protein containing a fused Clp peptide tag (48.4 kDa) as our non-native test cargo, we first attempted *in vitro* cargo loading by simply incubating MBP-*TEVext*-TmE, GCaMP, and uTEV2Δ in solution for 1 hour at 37°C before loading the sample mixture into an analytical SEC column for analysis. Comparison of chromatograms generated for untreated MBP-*TEVext*-TmE, MBP-*TEVext*-TmE treated with protease, and MBP-*TEVext*-TmE treated with protease in the presence of a 2-fold excess of GCaMP (**Figure 4.10**) show that the fusion protein was efficiently cleaved in both samples containing uTEV2Δ; however, no GCaMP appears to have been packaged into the containers in the third sample, as evidenced by the lack of a 495 nm absorbance signal (corresponding to the absorbance of the GCaMP chromophore) in the same 8-10 mL elution volume range as the assembled encapsulins. Instead, a strong 495 nm absorbance signal is observed later around the 16-18 mL range. SDS-PAGE analysis of fractions collected during each SEC run confirm that no other protein bands co-elute with the assembled encapsulins. Instead, a

48 kDa band corresponding to the molecular weight of GCaMP is observed to co-elute alongside the cleaved ~43 kDa monomeric MBP-linker protein in the fractions collected from the 16-18 mL range.

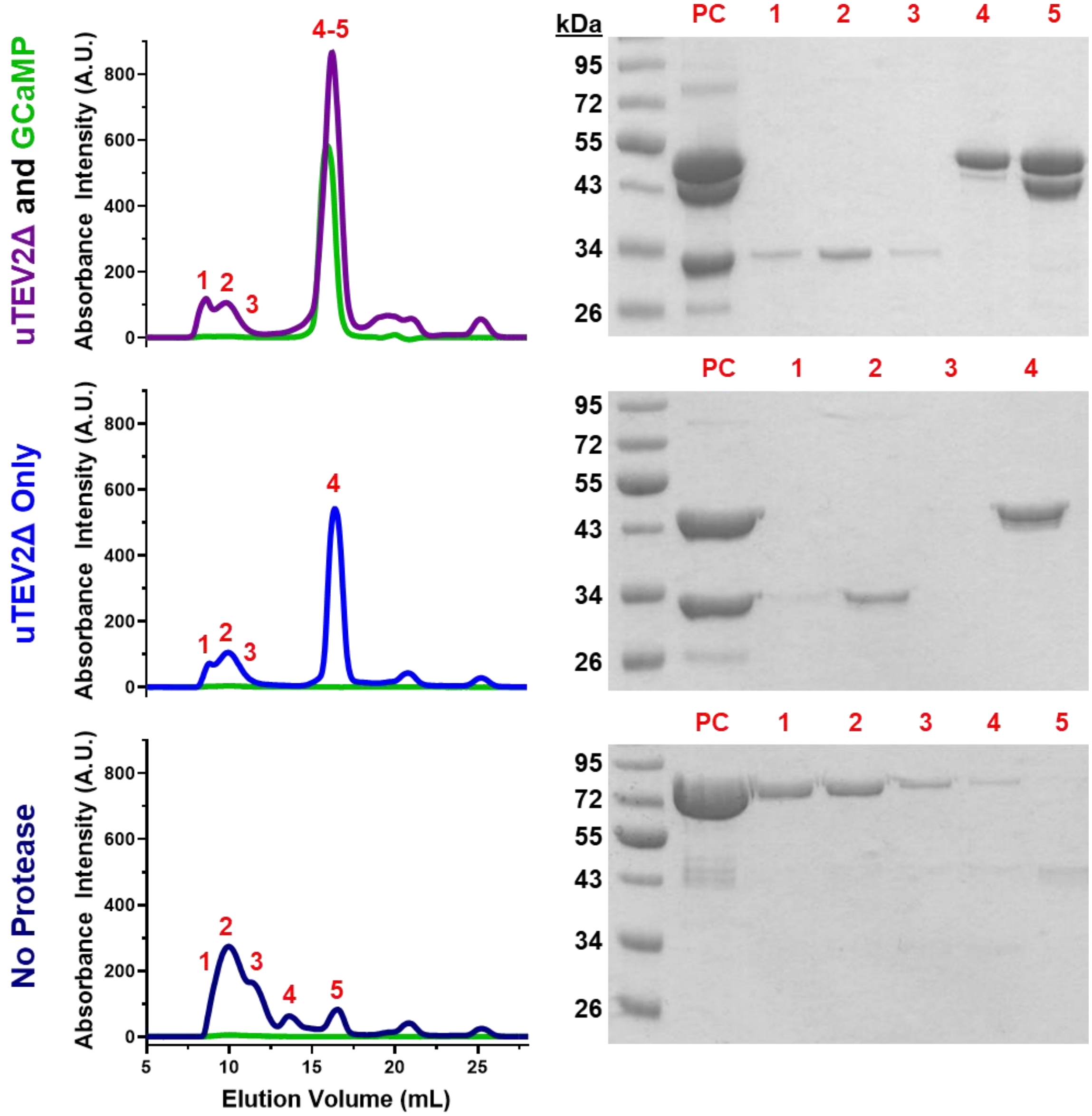


Figure 4.10 – Comparison of analytical SEC chromatograms generated for untreated MBP-*TEVext*-TmE (dark blue trace, bottom), MBP-*TEVext*-TmE treated with uTEV2Δ (blue trace, middle), and MBP-*TEVext*-TmE treated with uTEV2Δ in the presence of a 2-fold excess of GCaMP (purple trace, top). The absorbance at 485 nm is represented by the green trace in each chromatogram, showing that only the top chromatogram possesses a GCaMP signal. Fractions from each chromatogram were assessed via SDS-PAGE and are numbered on the SEC traces in red. The "PC" lanes in each SDS-PAGE gel show the sample composition "pre-column" (i.e. prior to SEC analysis).

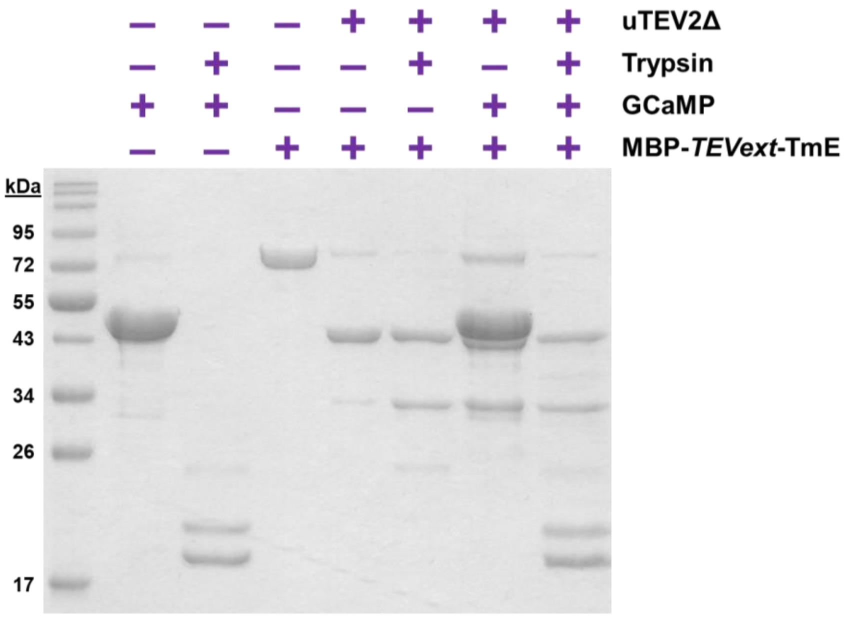


Figure 4.11 – SDS-PAGE assessment of revised attempt to perform *in vitro* cargo loading via CPA by increasing the excess of GCaMP relative to MBP-*TEVext*-TmE from 2-fold up to 5-fold. The contents of each lane in the gel image above are denoted by the plus (+) and minus (–) symbols in the same horizontal plane as the respective protein component in the right-hand legend.

In light of these results, we suspected that while CPA does liberate the majority of the TmE monomers from the MBP-*TEVext*-TmE fusion proteins, some population of whole MBP-*TEVext*-TmE proteins may not only survive the treatment with uTEV2Δ, but they may also become incorporated into the newly assembling encapsulin nanocontainers. In particular, our suspicions were based on the broadness of the encapsulin-containing peaks in the SEC chromatograms presented in **Figure 4.10**, the fact that this same 8-10 mL elution range contains multiple absorbance peaks, and due to some percentage of full length MBP-*TEVext*-TmE seeming to persist in SDS-PAGE gels from both the test cleavage trials with uTEV2Δ presented in **Figure 4.9** and the pre-SEC samples presented in **Figure 4.10**. If this is indeed the case, then the residual MBP-*TEVext*-TmE population could be filling the void space within the nascent nanocontainers, thus preventing effective GCaMP binding. In an attempt to eliminate this possible interference by full length MBP-*TEVext*-TmE, we repeated the *in vitro* loading experiment using a larger, 5-fold excess of GCaMP relative to the starting concentration of MBP-*TEVext*-TmE under the rationale that higher concentrations of GCaMP might better outcompete any residual fusion protein for encapsulation within the assembling nanocontainers. **Figure 4.11** shows, however, that excess GCaMP alone is insufficient to address this problem as the entire GCaMP population added into the *in vitro* mixture was rapidly degraded by trypsin treatment following a 2 hour prior incubation with uTEV2Δ, indicating that none of the reporter protein had been sequestered within encapsulin shells during the nanocontainer assembly phase.

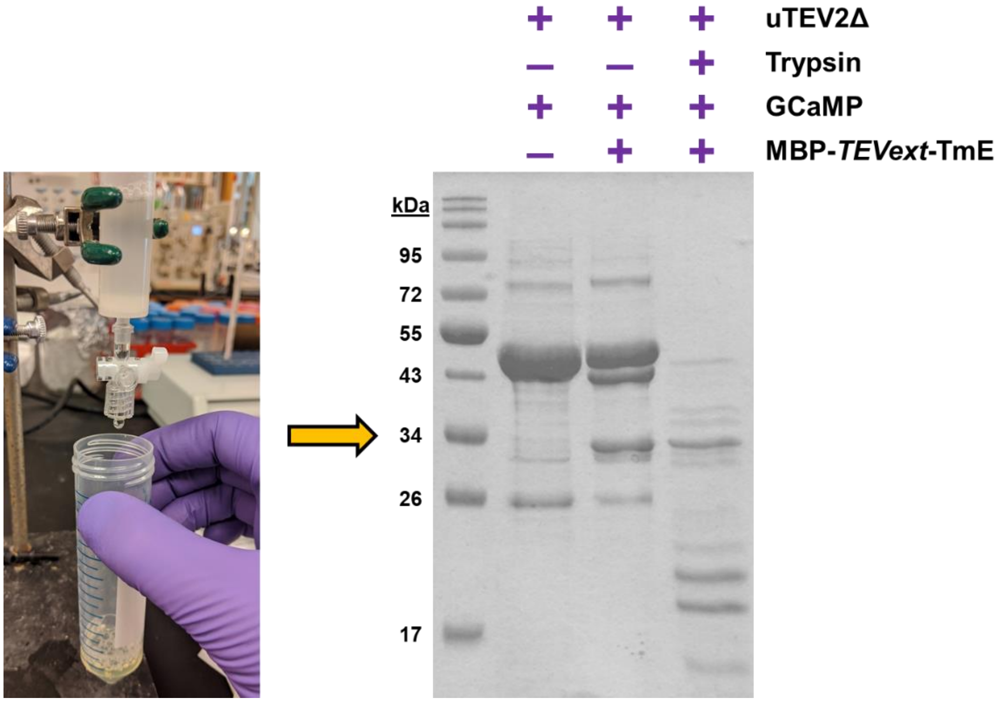


Figure 4.12 – Attempted *in vitro* cargo loading by dropwise elution of MBP-*TEVext*-TmE into a Falcon tube containing molar excesses of both GCaMP (48.4 kDa) and uTEV2Δ (24.8 kDa) to promote rapid cleavage upon the addition of each drop. The left-most sample lane in the SDS-PAGE gel image depicts the contents of the collection tube prior to the elution of any MBP-*TEVext*-TmE. The next sample lane depicts the contents of the collection tube following a brief incubation period after the dropwise addition of MBP-*TEVext*-TmE was completed (cleaved MBP = 42.8 kDa, TmE = 30.5 kDa). The final lane depicts the same sample following a brief incubation with trypsin.

As an alternative method intended to limit potential MBP-*TEVext*-TmE interference during *in vitro* loading of GCaMP, we next attempted to effectively increase the molar excess of GCaMP even further by first loading MBP-*TEVext*-TmE onto amylose affinity resin, and then slowly eluting the bound fusion protein in a dropwise fashion into a collection tube containing uTEV2Δ and concentrated GCaMP (**Figure 4.12**). In so doing, we hoped that sequential addition of fusion protein would allow the GCaMP to immediately bind to the luminal surfaces of freshly cleaved TmE protomers in order to ensure incorporation of the reporter proteins into the assembling cages. The concentration of uTEV2Δ was also increased 12-fold in the collection tube relative to previous tests as a means of promoting rapid cleavage of the fusion protein. After eluting the MBP-*TEVext*-TmE into the collection tube containing GCaMP and uTEV2Δ over the span of approximately 20 minutes, the collection tube was sealed and incubated at room temperature (~22°C) for 1 hour. Cleavage of the fusion protein was clearly evident by SDS-PAGE analysis (**Figure 4.12**). Despite the alterations to the *in vitro* loading procedure, however, no GCaMP appears to have been loaded into the newly formed encapsulins as trypsin treatment was again shown to completely eliminate the GCaMP protein band in the final sample.

4.3 – Conclusion

In the works presented in this chapter, we have effectively introduced a novel method for preventing the complete assembly of *T. maritima* encapsulin nanocontainers *in vivo* through the simple genetic fusion of the globular *E. coli* MBP protein to the N-terminus of the encapsulin monomeric coat protein. This methodology, which exploits the interior-facing orientation of the encapsulin protomers' N-termini, relies on the generation of unfavorable steric interactions within the limited volume of the nanocontainer lumen as the key factor preventing assembly of the full cage architecture. We subsequently showed that proteolytic separation of the MBP and TmE proteins using specific cleavage sites encoded within a flexible linker peptide connecting the two proteins led to the full assembly of nanocontainers presenting wildtype-like morphologies and SEC profiles. We thus refer to this technique as “cleavage-promoted assembly.” However, the original intent of this study was to utilize this novel technique as a means to artificially load non-native cargoes into encapsulin nanocontainers *in vitro* without the need for dramatic shifts in pH or the use of highly concentrated denaturing chaotropic salts. Unfortunately, attempts to load a modified GFP derivative reporter protein using CPA were unsuccessful whereas successful *in vitro* loading of the same protein can be achieved using high concentrations of GdnHCl to reversibly disassemble TmE nanocontainers in accordance with previously published procedures (38).

Several results from our experiments presented herein provide possible explanations for the current failures of CPA as an *in vitro* cargo loading technique. When using uTEV2Δ to perform CPA, a persistent amount of full-length MBP-*TEVext*-TmE remained present in all of our tests. We thus suspected that un-cleaved MBP-*TEVext*-TmE starting proteins were being incorporated into the newly-forming nanocontainer cages and preventing the productive binding of GCaMP reporters to TmE protomers due to the residual fused MBP proteins occupying a majority of the available luminal space within the cages. Several attempts to outcompete any residual MBP-*TEVext*-TmE fusion protein from being incorporated into nascent encapsulin shells by increasing the concentrations of both GCaMP and uTEV2Δ were unsuccessful. The failure of these optimization tests implies that alternative of additional factors may be contributing to the inability to perform non-native cargo loading using the CPA technique.

Analysis of the original MBP-TmE fusion protein via analytical SEC showed that the steric bulk generated by the MBP proteins did not completely abolish the native oligomerization of TmE protomers as multiple peaks corresponding to several oligomeric assemblies were observed. Similarly, the modified MBP-*TEVext*-TmE variant displayed a nearly identical SEC profile with multiple oligomeric states present at largely the same retention volumes (not shown). Given that the majority of the observed oligomeric states for both fusion protein variants seem to possess approximate molecular weights above 1 MDa, it appears that the encapsulin protomers are still able to adopt fairly extensive semi-assembled states in spite of the artificially introduced unfavorable steric interactions. These data collectively suggest that the persistence of some un-cleaved MBP-*TEVext*-TmE in our CPA experiments may be the result of the largely pre-assembled nanocontainers needing only a few wildtype-sized encapsulin protomers to fill in the remaining gaps in order to form completed cage structures. If such is the case, exogenous GCaMP may not have sufficient time to enter the container interior prior to the completion of cage assembly. Additionally, then the presence of un-cleaved MBP proteins occupying the cages' luminal spaces is again implicated as a possible cause of *in vitro* cargo loading failure.

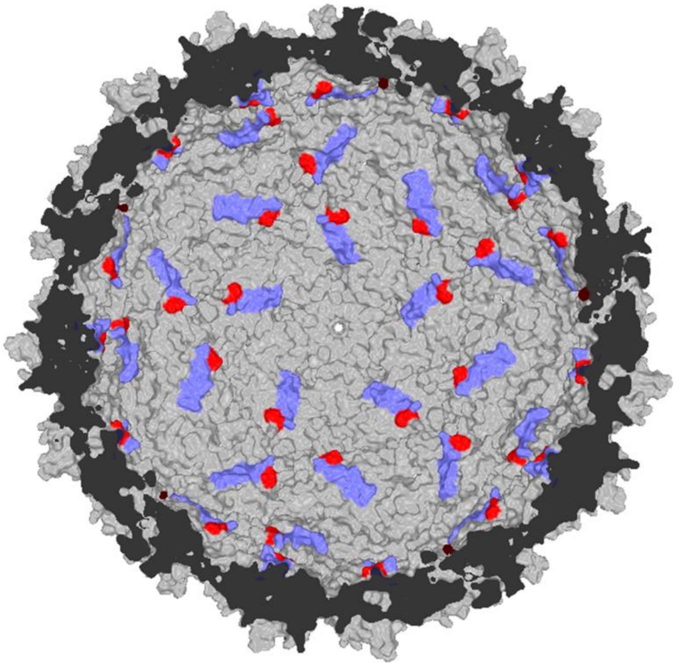


Figure 4.13 – Cross-sectional view of the assembled TmE nanocontainer (PDB ID: 3DKT). The N-terminal methionine residue of each shell monomer is colored in red while the adjacent Clp binding site it colored in slate blue.

Lastly, we reexamined the reported *T. maritima* atomic structure to look for further physical insights that might explain the failure of *in vitro* cargo loading via CPA. We subsequently noticed that the N-termini of the TmE monomers are directly adjacent to the luminal hydrophobic binding clefts where the Clp in known to dock during the native *in vivo* encapsulation process (**Figure 4.13**). Given that N-terminally fused MBP proteins are thus in close spatial proximity to the Clp binding clefts, it is likely that GCaMP is unable to effectively bind to the TmE monomers until scission of the MBP proteins occurs. If the resulting rate of encapsulin assembly following scission is rapid and exceeds the binding rate of GCaMP-Clp fusion proteins, then effective encapsulation of GCaMP would likely not occur. Alternatively, both the original MBP-TmE and the MBP-*TEVext*-TmE fusion designs leave behind 4 and 9 amino acid peptide segments, respectively, attached to the protomers' N-termini. It is also possible that these residual peptide segments could interfere with access to the Clp binding clefts in a manner which precludes effective *in vitro* cargo loading.

Ultimately, CPA is a facile and benign means of interrupting the native *in vivo* assembly of TmE nanocontainers. Furthermore, it represents a general strategy which could easily be applied to other protein-based nanocage architectures presenting luminally-oriented termini. However, further refinement of the technique is currently necessary to eliminate the pre-assembly of individual protomers and to ensure that residual fusion proteins do not become incorporated into nascent cages. A more detailed understanding of a given container's assembly kinetics may be beneficial in addressing these practical concerns as such information may afford researchers new insights into the temporal constraints that ultimately affect *in vitro* cargo loading methods.

4.4 – Materials and Methods

4.4.1 – Reagents and Materials

All chemicals and reagents were purchased from Sigma Aldrich (St. Louis, MO) unless otherwise noted. Carbon film 200 mesh copper electron microscopy grids were purchased from Electron Microscopy Sciences (Hatfield, PA). All pre-packed chromatography columns were purchased from GE Healthcare (Marlborough, MA). Free TALON affinity resin was purchased from Genesee Scientific (San Diego, CA). Free amylose affinity resin was purchased from New England Biolabs (Ipswich, MA). Electrocompetent *E. coli* DH5α cells were purchased from ThermoFisher Scientific (Waltham, MA) while chemically competent *E. coli* BL21(DE3) cells were acquired from New England Biolabs (Ipswich, MA). All commercial enzymes used for molecular cloning were purchased from New England Biolabs. All DNA primers were ordered from Integrated DNA Technologies (Coralville, IA). The pMAL-c2x vector was a gift from Paul Riggs (Addgene plasmid # 75286; http://n2t.net/addgene:75286; RRID:Addgene_75286), and the pRK793 vector was a gift from David Waugh (Addgene plasmid # 8827; http://n2t.net/addgene:8827; RRID:Addgene_8827). The pD434-SR vector containing a copy of the *T. maritima* encapsulin gene that had been codon-optimized for recombinant expression in *E. coli* host cells was purchased from ATUM (Newark, CA).

4.4.2 – Molecular Cloning

The MBP-TmE fusion protein was generated by PCR amplifying the codon-optimized TmE gene with primers encoding EcoRI and HindIII restriction sites at the 5’ and 3’ ends, respectively, using the pD434-SR:TmE vector as template. The resulting amplification product was digested with EcoRI and HindIII restriction endonucleases, and then the digested DNA product was ligated into pMAL-c2x vector that had been linearized using the same two restriction enzymes. DNA ligation was performed using T4 DNA ligase, and the resulting plasmid containing the MBP-TmE fusion gene was verified via DNA sequencing. The MBP-*TEV*-TmE variant was generated by PCR amplifying the TmE gene in three sequential PCR reactions using a common 3’ primer and iterative 5’ primers designed to construct the TEV cleavage site 12 base pairs upstream from the TmE gene’s start codon. The common 3’ primer and the final 5’ primers also encoded a regions of sequence homology corresponding to regions directly upstream from the native Factor Xa site and directly downstream from the HindIII restriction site in the pMAL-c2x vector. The pMAL-c2x vector was then amplified with 5’ and 3’ primers containing encoded segments homologous to the 5’ and 3’ ends of the TEV-TmE amplification product. Both the TEV-TmE and pMAL-c2x linear amplification products were produced with 5’ overhangs using the polymerase incomplete primer extension (PIPE) method reported previously (57). Subsequently, the two amplifications were digested with DpnI restriction endonuclease for 1 hour at 37°C to remove residual template plasmid, and then the two products were mixed *in vitro* to promote annealing of homologous single-stranded overhangs. The combined DNA mixture was then used to transform aliquots of *E. coli* DH5α cells. Successful production of the pMAL-c2x:MBP-*TEV*-TmE plasmid was confirmed via DNA sequencing. The MBP-TEVext-TmE variant was also produced following an identical PIPE cloning method as used for MBP-*TEV*-TmE with sequential 5’ primers designed to insert a GSGSG peptide linker sequence directly downstream from the TEV cleavage site. Successful production of the final pMAL-c2x:MBP-*TEVext*-TmE variant plasmid was confirmed via DNA sequencing. The uTEV2Δ variant reported previously by Sanchez & Ting (56) was produced using a TEV-S219V variant encoded in the pRK793 vector (54) as an initial template. Specifically, three sequential whole-plasmid mutagenic PCR reactions were performed to introduce successive mutations into the TEV gene. The individual mutations included two point mutations (T30A and S153N), and the introduction of a new stop codon immediately after residue 219 to truncate the C-terminal end of the translated TEV protein. For each reaction 50 ng of template DNA (either the pRK793 vector or one of the downstream mutated vector products) were used for PCR amplification. The template DNA was then removed via 1 hour digestion with DpnI at 37°C. Subsequently, amplified linear product DNA was transformed into chemically competent DH5α cells. The successful incorporation of each mutation was verified by DNA sequencing prior to proceeding to the next mutagenesis step. The final vector containing the uTEV2Δ variant was also confirmed via DNA sequencing.

4.4.3 – Protein Overexpression and Purification

All MBP-TmE variants were overexpressed and purified in an identical manner. Briefly, electrocompetent *E. coli* BL21(DE3) cells were transformed with the corresponding pMAL-c2x vector containing a given MBP-TmE variant. Transformed cells were then grown overnight on LB agar selection plates containing a final concentration of 0.1 mg/mL ampicillin at 37°C. Single colonies from the selection plate were used to inoculate 5 mL of liquid LB media containing a final concentration of 0.1 mg/mL ampicillin. Following overnight growth at 37°C, 3 mL from the saturated starter culture were used to inoculate 300 mL of fresh LB media containing final concentrations of 10 mM glucose and 0.1 mg/mL ampicillin. The large culture was then grown at 37°C to an OD_{600} value between 0.5 – 0.7, and then protein expression was induced with a final concentration of 0.1 mM IPTG. Protein expression was conducted at 37°C for 18 hours prior to harvesting the cell pellets via centrifugation. Cell pellets were stored at -20°C until purification.

Wildtype TmE was overexpressed according to previously established procedures (42). Briefly, the pET23b-TmE vector was transformed into electrocompetent *E. coli* BL21(DE3) cells and the resulting transformed cells were grown overnight at 37°C on LB agar selection plates containing a final concentration of 0.1 mg/mL ampicillin. A single was then used to inoculate 600 mL of liquid LB media supplemented with 0.1 mg/mL ampicillin. The liquid culture was grown at 37°C to an OD_{600} value between 0.5 – 0.7, and then protein expression as induced with a final concentration of 0.3 mM IPTG. Protein expression was conducted at 37°C for 18 hours prior to harvesting the cell pellets via centrifugation. Cell pellets were stored at -20°C until purification.

The GCaMP-Clp fusion protein was overexpressed according to previously established procedures (42). Briefly, the pMATT2:GCaMP vector was transformed into electrocompetent *E. coli* BL21(DE3) and the resulting transformed cells were grown overnight at 37°C on LB agar selection plates containing a final concentration of 0.05 mg/mL streptomycin. Single colonies from the selection plate were used to inoculate 5 mL of liquid LB media containing a final concentration of 0.05 mg/mL streptomycin. Following overnight growth at 37°C, 3 mL from the saturated starter culture were used to inoculate 300 mL of fresh LB media containing a final concentration of 0.05 mg/mL streptomycin. The large culture was then grown at 37°C to an OD_{600} value between 0.5 – 0.7, and then protein expression was induced with a final concentration of 0.3 mM IPTG. Protein expression was conducted at 20°C for 18 hours prior to harvesting the cell pellets via centrifugation. Cell pellets were stored at -20°C until purification.

For uTEV2Δ, the pRK793:uTEV2Δ vector was transformed into electrocompetent *E. coli* BL21(DE3) and the resulting transformed cells were grown overnight at 37°C on LB agar selection plates containing a final concentration of 0.1 mg/mL ampicillin. Single colonies from the selection plate were used to inoculate 5 mL of liquid LB media containing a final concentration of 0.1 mg/mL ampicillin. Following overnight growth at 37°C, the entire saturated starter culture was used to inoculate 600 mL of fresh LB media containing final concentrations of 10 mM glucose and 0.1 mg/mL of ampicillin. The large culture was then grown at 37°C to an OD_{600} value between 0.5 – 0.7, and then protein expression was induced with a final concentration of 0.3 mM IPTG. Protein expression was conducted at 18°C for 18 hours prior to harvesting the cell pellets via centrifugation. Cell pellets were stored at -20°C until purification.

All MBP-TmE fusion proteins were purified following an identical protocol. Cell pellets containing a given MBP-TmE variant were resuspended in 5 -10 mL of 50 mM HEPES-KOH (pH 7.5), 300 mM NaCl containing Protease Inhibitor Cocktail (Sigma Aldrich, P8849), 0.1 µg/mL PMSF, and 5 µg/mL DNaseI. The resuspended cells were then lysed by sonication (10 second pulses, 20 second rest between pulses) for 3.5 minutes on ice. Insoluble cellular debris was removed via centrifugation at 4000 rpms for 40 minutes at 4°C. The clarified supernatant was then decanted into a 30 mL polypropylene disposable column containing 1 mL of amylose affinity resin (pre-equilibrated with 5 column volumes of lysis buffer) and the sealed column tube was then incubated on a rocking table at 4°C for 20 minutes. The column was then unsealed and the flow-through was collected. The resin bed was subsequently washed with 15 column volumes of the resuspension buffer, and then bound MBP-TmE proteins were eluted by washing the resin bed with 3 – 5 column volumes of 50 mM HEPES-KOH (pH 7.5), 300 mM NaCl, 10 mM maltose. Purified MBP-TmE proteins were stored at 4°C until use. Protein concentrations were determined spectrophotometrically at 280 nm using the following theoretical extinction coefficients: 102,915 $M^{-1}cm^{-1}$ for MBP-TmE and 104,405 $M^{-1}cm^{-1}$ for both MBP-*TEV*-TmE and MBP-*TEVext*-TmE.

Purification of wildtype TmE was according to previously established procedures (42, 58). Cell pellets were resuspended on ice in 10 mL of 50 mM Tris-HCl (pH 7.5) containing Protease Inhibitor Cocktail, 0.1 μg/mL PMSF, and 5 μg/mL DNaseI and then the cells were lysed by sonication (10 second pulses, 20 seconds rest between pulses) for 3.5 minutes. Insoluble debris was removed via centrifugation at 4000 rpms for 40 minutes at 4°C. The clarified supernatant was then loaded into a HiTrap Q-FF anion exchange column (pre-equilibrated with 5 column volumes of lysis buffer) and the column flow-through was collected. Subsequently, 1 g of PEG-800 and 0.2 g of NaCl were added to the flow-through fraction and dissolved at room temperature with gentle mixing. The sample was then incubated at 4°C for 1 hour to promote precipitation of the encapsulin nanocontainers. The solution was then centrifuged at 5000 rpms for 45 minutes at 4°C to collect all solid precipitates. Following centrifugation, the supernatant was decanted, and then the precipitated protein pellet was resuspended in 5 mL of 50 mM HEPES-KOH (pH 7.5), 100 mM NaCl. A final concentration of 0.1 mg/mL bovine trypsin was then added, and then the sample was incubated at 37°C for 10 minutes to proteolytically remove non-encapsulin contaminants. The sample was then passed through a 0.20 μm nylon syringe filter, and then loaded into a HiPrep 16/60 Sephacryl S-500HR size exclusion chromatography column pre-equilibrated in 50 mM HEPES-KOH (pH 7.5), 100 mM NaCl and connected to an NGC Chromatography FPLC system (Bio-Rad, Hercules, CA). The same buffer was used as the mobile phase during SEC. Fractions with assembled TmE nanocontainers were pooled and stored at 4°C until use. The concentration of isolated TmE was determined spectrophotometrically at 280 nm using the theoretical extinction coefficient 36,440 $M^{-1}cm^{-1}$.

Purification of GCaMP-Clp fusion proteins was performed as reported previously (42). Briefly, cell pellets were resuspended in 10 mL of 50 mM Tris-HCl (pH 7.5), 20 mM NaCl containing Protease Inhibitor Cocktail, 0.1 μg/mL PMSF, and 5 μg/mL DNaseI and then the cells were lysed by sonication (10 second pulses, 20 seconds rest between pulses) for 3.5 minutes. Insoluble debris was removed via centrifugation at 4000 rpms for 40 minutes at 4°C. The clarified supernatant was then loaded into a HiTrap Q-FF anion exchange column (pre-equilibrated with 5 column volumes of lysis buffer) connected to an ÄKTA Explorer 10 FPLC system (GE Life Sciences, Little Chalfont, Buckinghamshire, UK). The bound protein was eluted by increasing the NaCl content of the mobile phase linearly from 20 mM to 1 M. Fractions containing GCaMP were pooled and the protein concentration was determined spectrophotometrically at 280 nm using the extinction coefficient 31,860 $M^{-1}cm^{-1}$. The purified protein was divided into aliquots, which were then flash frozen in liquid nitrogen and stored at -70°C until use.

Purification of uTEV2Δ was by first resuspending cell pellets in 10 mL of 50 mM Tris-HCl (pH 7.5), 50 mM NaCl, 2.5 mM β-mercaptoethanol, 10 mM Imidazole containing 0.1 μg/mL PMSF and 5 μg/mL DNaseI. The cells were then lysed by sonication (10 second pulses, 20 seconds rest between pulses) for 3.5 minutes. Insoluble debris was removed via centrifugation at 4000 rpms for 40 minutes at 4°C, then the clarified lysate was loaded into a 30 mL polypropylene disposable column containing 1 mL of TALON affinity resin (pre-equilibrated with 5 column volumes of lysis buffer). The column was sealed and placed on a rocking table at 4˚C for 20 minutes. The column was then unsealed and the flow-through was collected. Subsequently, the resin bed was washed with 15 column volumes of the resuspension buffer, then bound protein was eluted by washing the resin bed with 3 column volumes of 100 mM Tris-HCl (pH 7.5), 50 mM NaCl, 2.5 mM β-mercaptoethanol, 250 mM Imidazole. The eluted protein was then loaded into a cellulose dialysis bag (14 kDa MWCO) and dialyzed overnight at 4°C into 100 mM K-Phosphate (pH 7.0), 100 mM NaCl, 2.5 mM β-mercaptoethanol to remove excess imidazole. The concentration of purified uTEV2Δ was determined spectrophotometrically using the theoretical extinction coefficient 31,970 $M^{-1}cm^{-1}$.

4.4.4 – *In vitro* Loading via GdnHCl-based Denaturation

To perform *in vitro* cargo loading following the methodology established by Cassidy-Amstutz *et al*. (38), purified TmE nanocontainers were first dialyzed into 25 mM HEPES-KOH (pH 7.5) overnight at 4°C. Following dialysis, the TmEs were concentrated using Amicon Ultra 0.5 mL centrifugal filters (10 kDa MWCO, Millipore Sigma, Burlington, MA) to reach a final concentration of approximately 0.5 mM. Subsequently, 8 μL of the concentrated TmE stock were mixed with 42 μL of concentrated 8 M GdnHCl to generate a final 50 μL sample solution containing 6.7 M GdnHCl and 80 μM TmE. The solution was incubated on ice for 1 hour to promote encapsulin disassembly, and then 450 μL of reassembly buffer containing 100 mM HEPES-KOH (pH 7.5), 1 mM DTT, and 16 μM GCaMP were added to the 50 μL disassembly mixture to initiate reassembly of the nanocontainers. The reassembly phase was allowed to proceed for 16 hours at 22°C, and then non encapsulated GCaMP was removed by adding a final concentration of 0.1 mg/mL bovine trypsin (Millipore Sigma, Burlington, MA) into the sample mixture. Trypsin digestion was carried out for 10 minutes at 37°C, and then 5 μL of 2 M $CaCl_2$ was spiked into the sample mixture to both inactivate the proteolytic activity of trypsin and to initiate the Ca^{2+}-dependent fluorescence of any surviving GCaMP reporter proteins. Fluorescence measurements from surviving GCaMP proteins were collected using a HORIBA Jobin Yvon FluoroMax®-3 fluorimeter (Edison, NJ). Samples were excited at 485 nm and the resulting emission data was collected over the range of 450 to 600 nm. Excitation and emission slits were set at 1 nm each, and the data integration time was set at 0.1 seconds.

4.4.5 – Encapsulin CAC_{app} Determination

Attempted determination of the CAC_{app} value for TmE was performed following the procedure described by Li *et al.* (50). To initiate encapsulin disassembly, a 10 mL solution containing 40 μM of purified TmEs in 50 mM HEPES-KOH (pH 7.5), 0.1 mM NaCl was mixed with a 10 mL solution of 2x Disassembly Mix containing 0.4 M HCl and 2 mM DTT. The resulting mixture had a pH of ~1.0, as determined using MColorpHast™ pH-indicator strips (Millipore Sigma, Burlington, MA). The disassembly mixture was incubated on ice for 90 minutes, and then individual samples containing final TmE concentrations between 1 – 11 μM were prepared by diluting disassembled TmEs in additional 1x Disassembly Mix solution. The individual TmE samples were then loaded into cellulose dialysis bags (14 kDa MWCO) and dialyzed overnight at 4°C against 3 L of 50 mM HEPES-KOH (pH 7.5), 100 mM NaCl. The samples were recovered from the dialysis bags the following morning and assessed via HPLC-based SEC by injecting 50 μL from each sample mixture sequentially into an Agilent 1200 HPLC (Agilent Technologies Inc., Santa Clara, CA) equipped with a Sepax SRT SEC-500 gel filtration column (4.6 x 300 mm, 5 μm). Sample elution was monitored spectrophotometrically at 280 nm at a flow rate of 0.4 mL/min using 50 mM HEPES-KOH, 100 mM NaCl as the isocratic mobile phase.

4.4.6 – Size Exclusion Chromatography Analyses

The SEC-based oligomeric state analysis of MBP-TmE was carried out by injecting 1 mL of purified fusion protein into an ÄKTA Explorer 10 FPLC system equipped with a Superdex 200 Increase 10/300 GL column that had been pre-equilibrated with 50 mM HEPES-KOH (pH 7.5), 100 mM NaCl. SEC was performed at a flow rate of 0.5 mL/min using the same pre-equilibration buffer as the isocratic mobile phase. Protein elution was monitored spectrophotometrically at 280 nm. Subsequently, a MWGF200 molecular weight standard kit (Sigma Aldrich, St. Louis, MO) containing horse heart cytochrome C, bovine carbonic anhydrase, bovine serum albumin, yeast alcohol dehydrogenase, sweet potato β-amylase, and blue dextran was loaded into the same column and SEC was performed under identical conditions as for MBP-TmE. The resulting elution volumes were used to determine the partition coefficients (K_{AV}) for the respective protein standards, and a plot of K_{AV} versus logMW was subsequently used to generate a linear calibration curve for the SEC column (59).

Comparisons of untreated MBP-TmE, samples of MBP-TmE treated with bovine trypsin only, and samples treated with bovine trypsin in the presence of GCaMP were performed by loading 5 mL of the respective MBP-TmE sample into an NGC Chromatography FPLC system equipped with a HiPrep 16/60 Sephacryl S-500HR column that had been pre-equilibrated in 50 mM HEPES-KOH (pH 7.5), 100 mM NaCl. SEC was carried out at speeds between 0.3 to 0.7 mL/min using the same pre-equilibration buffer as the isocratic mobile phase. Protein elution was monitored spectrophotometrically at 280 nm.

4.4.7 – Proteolytic Cleavage Assays

Proteolytic cleavage assays using bovine trypsin were performed at either 4, 22, or 37°C for a total of 15 minutes up to 18 hours. A final trypsin concentration of 0.1 mg/mL was used in all cleavage assays, and proteolytic cleavage was halted by the addition of a final concentration of 10-20 mM $CaCl_2$. For samples containing GCaMP, the large excess of $CaCl_2$ was simultaneously utilized to initiate Ca^{2+}-dependent fluorescence of the GCaMP chromophore. For proteolytic assays using TEV protease, variable concentrations of uTEV2Δ between 0.5 and 10 μM were used for cleavage assays performed at 4, 22 or 37°C.

4.4.8 – Transmission Electron Microscopy Imaging

Individual TEM grids were prepared by applying 4 μL of biological sample onto the carbon surface of carbon-copper grids for five minutes. Samples were diluted to a final protein concentration between 0.1 – 0.3 mg/mL to prevent overcrowding of proteins on the grid surface. After five minutes had elapsed, the edge of the grid was gently blotted against a sheet of Whatman 1 filter paper. The grid was then inverted and quickly immersed sequentially into two drops of deionized water. The edge of the grid was again blotted against filter paper, and then negative staining was accomplished by applying 4 μL of 1% (w/v) phosphotungstic acid (pH 6.5, prepared in double distilled water and pH adjusted with KOH) onto the surface of the grid and allowing it to incubate for 20 seconds. The edge of the grid was then blotted against filter paper one final time, and then the grids were allowed to air dry for five minutes before being placed into a vacuum desiccator for five additional minutes prior to imaging. TEM imaging was performed on a Hitachi

HT7700 transmission electron microscope (Hitachi, Tokyo, Japan) operating at an accelerating voltage of 80.0 kV.

4.5 – References

1. Aumiller, W. M.; Uchida, M.; Douglas, T., Protein cage assembly across multiple length scales. *Chemical Society Reviews* **2018**, *47* (10), 3433-3469.
2. Wörsdörfer, B.; Pianowski, Z.; Hilvert, D., Efficient in Vitro Encapsulation of Protein Cargo by an Engineered Protein Container. *Journal of the American Chemical Society* **2012**, *134* (2), 909-911.
3. Glasgow, J. E.; Capehart, S. L.; Francis, M. B.; Tullman-Ercek, D., Osmolyte-mediated encapsulation of proteins inside MS2 viral capsids. *ACS Nano* **2012**, *6* (10), 8658-8664.
4. Rurup, W. F.; Snijder, J.; Koay, M. S. T.; Heck, A. J. R.; Cornelissen, J., Self-Sorting of Foreign Proteins in a Bacterial Nanocompartment. *Journal of the American Chemical Society* **2014**, *136* (10), 3828-3832.
5. Rurup, W. F.; Verbij, F.; Koay, M. S. T.; Blum, C.; Subramaniam, V.; Cornelissen, J., Predicting the Loading of Virus-Like Particles with Fluorescent Proteins. *Biomacromolecules* **2014**, *15* (2), 558-563.
6. Dashti, N. H.; Abidin, R. S.; Sainsbury, F., Programmable In Vitro Coencapsidation of Guest Proteins for Intracellular Delivery by Virus-like Particles. *ACS Nano* **2018**, *12* (5), 4615-4623.

7. Patterson, D. P.; Prevelige, P. E.; Douglas, T., Nanoreactors by Programmed Enzyme Encapsulation Inside the Capsid of the Bacteriophage P22. *ACS Nano* **2012**, *6* (6), 5000-5009.

8. Patterson, D. P.; Schwarz, B.; Waters, R. S.; Gedeon, T.; Douglas, T., Encapsulation of an Enzyme Cascade within the Bacteriophage P22 Virus-Like Particle. *ACS Chemical Biology* **2014**, *9* (2), 359-365.

9. Sánchez-Sánchez, L.; Cadena-Nava, R. D.; Palomares, L. A.; Ruiz-Garcia, J.; Koay, M. S. T.; Cornelissen, J. J. M. T.; Vazquez-Duhalt, R., Chemotherapy pro-drug activation by biocatalytic virus-like nanoparticles containing cytochrome P450. *Enzyme and Microbial Technology* **2014**, *60*, 24-31.

10. Schoonen, L.; Nolte, R. J. M.; van Hest, J. C. M., Highly efficient enzyme encapsulation in a protein nanocage: towards enzyme catalysis in a cellular nanocompartment mimic. *Nanoscale* **2016**, *8* (30), 14467-14472.

11. Giessen, T. W.; Silver, P., A catalytic nanoreactor based on in vivo encapsulation of multiple enzymes in an engineered protein nanocompartment. *ChemBioChem* **2016**, *17*, 1931-1935

12. Ashley, C. E.; Carnes, E. C.; Phillips, G. K.; Durfee, P. N.; Buley, M. D.; Lino, C. A.; Padilla, D. P.; Phillips, B.; Carter, M. B.; Willman, C. L.; Brinker, C. J.; Caldeira, J. d. C.; Chackerian, B.; Wharton, W.; Peabody, D. S., Cell-Specific Delivery of Diverse Cargos by Bacteriophage MS2 Virus-like Particles. *ACS Nano* **2011**, *5* (7), 5729-5745.

13. Choi, K.-m.; Choi, S.-H.; Jeon, H.; Kim, I.-S.; Ahn, H. J., Chimeric Capsid Protein as a Nanocarrier for siRNA Delivery: Stability and Cellular Uptake of Encapsulated siRNA. *ACS Nano* **2011**, *5* (11), 8690-8699.

14. Li, L.; Muñoz-Culla, M.; Carmona, U.; Lopez, M. P.; Yang, F.; Trigueros, C.; Otaegui, D.; Zhang, L.; Knez, M., Ferritin-mediated siRNA delivery and gene silencing in human tumor and primary cells. *Biomaterials* **2016**, *98*, 143-151.

15. Suzuki, M.; Abe, M.; Ueno, T.; Abe, S.; Goto, T.; Toda, Y.; Akita, T.; Yamada, Y.; Watanabe, Y., Preparation and catalytic reaction of Au/Pd bimetallic nanoparticles in Apo-ferritin. *Chemical Communications* **2009**, (32), 4871-4873.

16. Kang, Y. J.; Uchida, M.; Shin, H.-H.; Douglas, T.; Kang, S., Biomimetic FePt nanoparticle synthesis within Pyrococcus furiosus ferritins and their layer-by-layer formation. *Soft Matter* **2011**, *7* (23), 11078-11081.

17. Künzle, M.; Eckert, T.; Beck, T., Metal-Assisted Assembly of Protein Containers Loaded with Inorganic Nanoparticles. *Inorganic Chemistry* **2018**, *57* (21), 13431-13436.

18. Künzle, M.; Mangler, J.; Lach, M.; Beck, T., Peptide-directed encapsulation of inorganic nanoparticles into protein containers. *Nanoscale* **2018**, *10* (48), 22917-22926.

19. Wu, M.; Brown, W. L.; Stockley, P. G., Cell-Specific Delivery of Bacteriophage-Encapsidated Ricin A Chain. *Bioconjugate Chemistry* **1995**, *6* (5), 587-595.

20. Wu, W.; Hsiao, S. C.; Carrico, Z. M.; Francis, M. B., Genome-Free Viral Capsids as Multivalent Carriers for Taxol Delivery. *Angewandte Chemie International Edition* **2009**, *48* (50), 9493-9497.

21. Moon, H.; Lee, J.; Kim, H.; Heo, S.; Min, J.; Kang, S., Genetically engineering encapsulin protein cage nanoparticle as a SCC-7 cell targeting optical nanoprobe. *Biomaterials Research* **2014**, *18* (1), 21.

22. Moon, H.; Lee, J.; Min, J.; Kang, S., Developing Genetically Engineered Encapsulin Protein Cage Nanoparticles as a Targeted Delivery Nanoplatform. *Biomacromolecules* **2014**, *15* (10), 3794-3801.

23. Ke, Z.; Abe, S.; Ueno, T.; Morokuma, K., Catalytic Mechanism in Artificial Metalloenzyme: QM/MM Study of Phenylacetylene Polymerization by Rhodium Complex Encapsulated in apo-Ferritin. *Journal of the American Chemical Society* **2012**, *134* (37), 15418-15429.

24. Maity, B.; Fukumori, K.; Abe, S.; Ueno, T., Immobilization of two organometallic complexes into a single cage to construct protein-based microcompartments. *Chemical Communications* **2016**, *52* (31), 5463-5466.

25. Hestericová, M.; Heinisch, T.; Lenz, M.; Ward, T. R., Ferritin encapsulation of artificial metalloenzymes: engineering a tertiary coordination sphere for an artificial transfer hydrogenase. *Dalton Transactions* **2018**, *47* (32), 10837-10841.

26. Fiedler, J. D.; Brown, S. D.; Lau, J. L.; Finn, M. G., RNA-directed packaging of enzymes within virus-like particles. *Angewandte Chemie International Edition* **2010**, *49* (50), 9648-9651.

27. O'Neil, A.; Reichhardt, C.; Johnson, B.; Prevelige, P. E.; Douglas, T., Genetically programmed in vivo packaging of protein cargo and its controlled release from bacteriophage P22. *Angewandte Chemie International Edition* **2011**, *50* (32), 7425-7428.

28. Patterson, D. P.; McCoy, K.; Fijen, C.; Douglas, T., Constructing catalytic antimicrobial nanoparticles by encapsulation of hydrogen peroxide producing enzyme inside the P22 VLP. *Journal of Materials Chemistry B* **2014**, *2* (36), 5948-5951.

29. Lau, Y. H.; Giessen, T. W.; Altenburg, W. J.; Silver, P. A., Prokaryotic nanocompartments form synthetic organelles in a eukaryote. *Nature Communications* **2018**, *9* (1), 1311.

30. Wilkerson, J. W.; Yang, S.-O.; Funk, P. J.; Stanley, S. K.; Bundy, B. C., Nanoreactors: Strategies to encapsulate enzyme biocatalysts in virus-like particles. *New Biotechnology* **2018**, *44*, 59-63.

31. Zschoche, R.; Hilvert, D., Diffusion-Limited Cargo Loading of an Engineered Protein Container. *Journal of the American Chemical Society* **2015**, *137* (51), 16121-16132.

32. Azuma, Y.; Zschoche, R.; Tinzl, M.; Hilvert, D., Quantitative Packaging of Active Enzymes into a Protein Cage. *Angewandte Chemie International Edition* **2016**, *55* (4), 1531-1534.

33. Sharma, J.; Douglas, T., Tuning the catalytic properties of P22 nanoreactors through compositional control. *Nanoscale* **2020**, *12* (1), 336-346.

34. Lavelle, L.; Gingery, M.; Phillips, M.; Gelbart, W. M.; Knobler, C. M.; Cadena-Nava, R. D.; Vega-Acosta, J. R.; Pinedo-Torres, L. A.; Ruiz-Garcia, J., Phase Diagram of Self-assembled Viral Capsid Protein Polymorphs. *The Journal of Physical Chemistry B* **2009**, *113* (12), 3813-3819.

35. Comellas-Aragones, M.; Engelkamp, H.; Claessen, V. I.; Sommerdijk, N. A.; Rowan, A. E.; Christianen, P. C.; Maan, J. C.; Verduin, B. J.; Cornelissen, J. J.; Nolte, R. J., A virus-based single-enzyme nanoreactor. *Nature Nanotechnology* **2007**, *2* (10), 635-639.

36. Minten, I. J.; Claessen, V. I.; Blank, K.; Rowan, A. E.; Nolte, R. J. M.; Cornelissen, J. J. L. M., Catalytic capsids: the art of confinement. *Chemical Science* **2011**, *2* (2), 358-362.

37. Lee, K. W.; Tan, W. S., Recombinant hepatitis B virus core particles: Association, dissociation and encapsidation of green fluorescent protein. *Journal of Virological Methods* **2008**, *151* (2), 172-180.

38. Cassidy-Amstutz, C.; Oltrogge, L.; Going, C. C.; Lee, A.; Teng, P.; Quintanilla, D.; East-Seletsky, A.; Williams, E. R.; Savage, D. F., Identification of a Minimal Peptide Tag for in Vivo and in Vitro Loading of Encapsulin. *Biochemistry* **2016**, *55* (24), 3461-3468.

39. Sharma, J.; Uchida, M.; Miettinen, H. M.; Douglas, T., Modular interior loading and exterior decoration of a virus-like particle. *Nanoscale* **2017**, *9* (29), 10420-10430.

40. Nichols, R. J.; Cassidy-Amstutz, C.; Chaijarasphong, T.; Savage, D. F., Encapsulins: molecular biology of the shell. *Critical Reviews in Biochemistry and Molecular Biology* **2017**, 1-12.

41. Nakai, J.; Ohkura, M.; Imoto, K., A high signal-to-noise Ca2+ probe composed of a single green fluorescent protein. *Nature Biotechnology* **2001**, *19* (2), 137-141.

42. Williams, E. M.; Jung, S. M.; Coffman, J. L.; Lutz, S., Pore Engineering for Enhanced Mass Transport in Encapsulin Nanocompartments. *ACS Synthetic Biology* **2018**, *7* (11), 2514-2517.

43. Sutter, M.; Boehringer, D.; Gutmann, S.; Guenther, S.; Prangishvili, D.; Loessner, M. J.; Stetter, K. O.; Weber-Ban, E.; Ban, N., Structural basis of enzyme encapsulation into a bacterial nanocompartment. *Nature Structural and Molecular Biology* **2008**, *15* (9), 939-947.

44. Rahmanpour, R.; Bugg, T. D. H., Assembly in vitro of Rhodococcus jostii RHA1 encapsulin and peroxidase DypB to form a nanocompartment. *The FEBS Journal* **2013**, *280* (9), 2097-2104.

45. Hendrix, R. W.; Johnson, J. E., Bacteriophage HK97 Capsid Assembly and Maturation. In Viral Molecular Machines, Rossmann, M. G.; Rao, V. B., Eds. Springer US: Boston, MA, 2012; pp 351-363.

46. Seifer, M.; Zhou, S.; Standring, D. N., A micromolar pool of antigenically distinct precursors is required to initiate cooperative assembly of hepatitis B virus capsids in Xenopus oocytes. *Journal of Virology* **1993**, *67* (1), 249.

47. Zlotnick, A.; Johnson, J. M.; Wingfield, P. W.; Stahl, S. J.; Endres, D., A Theoretical Model Successfully Identifies Features of Hepatitis B Virus Capsid Assembly. *Biochemistry* **1999**, *38* (44), 14644-14652.

48. Johnson, J. M.; Tang, J.; Nyame, Y.; Willits, D.; Young, M. J.; Zlotnick, A., Regulating Self-Assembly of Spherical Oligomers. *Nano Letters* **2005**, *5* (4), 765-770.

49. Kler, S.; Asor, R.; Li, C.; Ginsburg, A.; Harries, D.; Oppenheim, A.; Zlotnick, A.; Raviv, U., RNA Encapsidation by SV40-Derived Nanoparticles Follows a Rapid Two-State Mechanism. *Journal of the American Chemical Society* **2012**, *134* (21), 8823-8830.

50. Li, L.; Xu, C.; Zhang, W.; Secundo, F.; Li, C.; Zhang, Z.-P.; Zhang, X.-E.; Li, F., Cargo-Compatible Encapsulation in Virus-Based Nanoparticles. *Nano Letters* **2019**, *19* (4), 2700-2706.

51. Walsh, K. A., [4] Trypsinogens and trypsins of various species. In Methods in Enzymology, Academic Press: 1970; Vol. 19, pp 41-63.

52. Fujikawa, K.; Coan, M. H.; Legaz, M. E.; Davie, E. W., Mechanism of activation of bovine factor X (Stuart factor) by intrinsic and extrinsic pathways. *Biochemistry* **1974**, *13* (26), 5290-5299.

53. Phan, J.; Zdanov, A.; Evdokimov, A. G.; Tropea, J. E.; Peters, H. K.; Kapust, R. B.; Li, M.; Wlodawer, A.; Waugh, D. S., Structural Basis for the Substrate Specificity of Tobacco Etch Virus Protease. *Journal of Biological Chemistry* **2002**, *277* (52), 50564-50572.

54. Kapust, R. B.; Tözsér, J.; Fox, J. D.; Anderson, D. E.; Cherry, S.; Copeland, T. D.; Waugh, D. S., Tobacco etch virus protease: mechanism of autolysis and rational design of stable mutants with wild-type catalytic proficiency. *Protein Engineering, Design and Selection* **2001**, *14* (12), 993-1000.

55. Tans, G.; Rosing, J., Snake Venom Activators of Factor X: An Overview. *Pathophysiology of Haemostasis and Thrombosis* **2001**, *31* (3-6), 225-233.

56. Sanchez, M. I.; Ting, A. Y., Directed evolution improves the catalytic efficiency of TEV protease. *Nature Methods* **2020**, *17* (2), 167-174.

57. Klock, H. E.; Lesley, S. A., The Polymerase Incomplete Primer Extension (PIPE) Method Applied to High-Throughput Cloning and Site-Directed Mutagenesis. In High Throughput

Protein Expression and Purification: Methods and Protocols, Doyle, S. A., Ed. Humana Press: Totowa, NJ, 2009; pp 91-103.

58. Giessen, T. W.; Silver, P. A., Converting a Natural Protein Compartment into a Nanofactory for the Size-Constrained Synthesis of Antimicrobial Silver Nanoparticles. *ACS Synthetic Biology* **2016**, *5* (12), 1497-1504.

59. Tayyab, S.; Qamar, S.; Islam, M., Size exclusion chromatography and size exclusion HPLC of proteins. *Biochemical Education* **1991**, *19* (3), 149-152.

Chapter 5:

General Conclusions and Final Thoughts

5.1 – General Conclusion

In recent decades, naturally-derived proteinaceous nanocontainers have garnered increasing interest in the biological, chemical, and physical science communities due to the broad scope of their potential applications. Particularly, the inherent plasticity, biocompatibility, and the ease with which numerous protein cage architectures can be produced have led to their incorporation in diverse biotechnologically-related fields, such as the development of vaccines and tissue-specific drug delivery vehicles (1-6), the generation of robust scaffolds for single and multistep catalytic transformations (7-13), and the creation of novel biomaterials displaying unique higher-order physiochemical properties (14-18). While the variegated feats accomplished thus far are certainly impressive, we anticipate that both nanocontainer engineering strategies and the incorporation of proteinaceous scaffolds into manifold experimental designs will likely continue to flourish in the coming years as researchers, armed with both a mélange of well-established cage assemblies and an ever-expanding repertoire of atomic level structural data, will be able to more precisely tailor scaffolds in a site-specific manner through both chemical and genetic means. Furthermore, increasingly powerful bioinformatic technologies are continuing to routinely identify new biological nanocontainers derived from viruses and from metabolically-active assemblies abstracted from all other kingdoms of life (19-23). Many recently characterized nanocontainers have been shown to exhibit novel biophysical properties, which is unsurprisingly considering the ubiquitous prevalence of protein-based cage architectures native to every natural biome on Earth (23, 24). Accordingly, we envision that the continued expansion of established nanocontainers and structural data will facilitate improved experimental designs by permitting researchers the ability to rationally choose specific architectures displaying properties suited to the intended downstream application as starting scaffolds.

5.2 – Nanocontainer-based Superstructure Assemblies

In this dissertation, I have described several parallel strategies employed to rationally engineer encapsulin nanocontainers derived from the hyperthermophilic bacterium *T. maritima* with the intent of generating nanocontainer-based scaffolds designed to facilitate multistep catalysis through a synthetic metabolon. The first half of chapter 2 describes our initial efforts to take advantage of non-covalent protein-protein interactions intrinsic to several small, naturally-derived peptide-binding domains and their corresponding peptide ligands to construct organized supramolecular assemblies composed of heterogeneous TmE variants displaying PBDs or PLs on their exterior surfaces. The original intent of our efforts in this regard was to produce highly ordered encapsulin arrays in a bottom-up fashion such that we could differentially package separate biocatalysts into the TmE-PBD and TmE-PL pairs, ultimately generating superstructures with high density catalysts loads that could perform multistep catalysis via communication of metabolites through the native pores located at the various symmetry axes on each encapsulin nanocontainer. Excitingly, we were able to show that the encapsulin shell is tolerant to genetic insertion of two separate PBDs roughly 7-10 kDa in size within several solvent-exposed flexible loops, and that *in vitro* mixtures of TmE-PBDs with TmE-PLs produced observable association of the two nanocontainer populations. However, rather than generating ordered assemblies, we routinely witnessed the formation of disordered aggregates.

Three-dimensional nanocontainer arrays displaying highly ordered crystalline packing geometries have been successfully produced by several research groups in the last two decades. In the majority of these cases, such defined lattice structures have been accomplished by allowing containers to assemble using electrostatic interactions between a given nanocontainer and some form or polyvalent bridging element capable of binding to multiple nanocontainers simultaneously, such as multivalent ions (25, 26), synthetic dendrimers (27-29), or symmetrically anisotropic proteins (30, 31). In all of these cases, careful modulation of both the solution ionic strength and pH were necessary to reduce the degree of electrostatic attraction between nanocontainers and bridging elements as solution conditions favoring strong electrostatic attractions consistently led to the formation of kinetically-trapped disordered aggregates (16, 25, 31, 32). Maintaining containers near their pI values or using higher ionic strengths to effectively tune the solution Debye screening length has been shown to reduce the strength of electrostatic attractions, allowing containers to instead assemble into thermodynamically-controlled lattice structures mediated by the polyvalent interactions of the employed bridging element (16, 32). In an exciting work recently reported by Uchida *et al.*, the authors successfully generated two P22 VLP populations differentially loaded with either ketoisovalerate decarboxylase or alcohol dehydrogenase enzymes that were assembled into ordered superlattices using polycationic dendrimers and recombinantly engineered anionic peptide tags on their VLPs' exterior surfaces (16). The resulting superlattice structures were capable of producing isobutanol from ketoisovalerate in a multistep catalytic fashion, similar to the goal we sought to achieve with our encapsulin-based nanocontainer system.

While these electrostatic-based strategies have been successful, they are not immediately translatable to our TmE-PBD and TmE-PL design as the protein-protein interactions between these container variants are not necessarily controlled by electrostatic interactions. The SH3 domain in particular interacts with its ligand, which forms a unique proline helix structure, largely through the binding of the N-substituted amides of the ligand's proline residues within defined grooves located in the SH3 domain binding pocket that are shaped as to disfavor the binding of all other amino acids natively lacking substituted backbone amines (33). Recently, however, McCoy *et al.* reported a successful blending of the both the electrostatic and protein-protein interaction-based assembly methodologies (17). Specifically, P22 VLPs were first assembled into ordered three-dimensional lattices using weak electrostatic interactions mediated by polycationic dendrimers, and then engineered bifunctional protein domains that have been shown previously to non-covalently bind to P22 exterior surfaces with low nanomolar affinity were used to lock the lattice structures in place (17, 34). Subsequently, the dendrimers could be removed following the addition of the bifunctional protein linkers without any loss of lattice stability. However, attempts to construct the lattice arrays using only the bifunctional protein domains without prior electrostatic preassembly led to the formation of disordered aggregates, which the authors attribute to the fast on-rate and slow off-rate of the protein-protein interactions due to the linker's tight binding affinity (17, 35). It is likely that our own encapsulin-based system suffered from a similar problem due to the low to sub-micromolar binding affinities of both the PDZ and SH3 domains grafted into the TmE shells, and that revision of our strategy to incorporate a weak assembly method prior to the introduction of PBD/PL interactions would likely yield a similar outcome as that reported above. Rather than using heterogeneous populations of TmE-PBDs and TmE-PLs, we would likely be better suited using a single population of TmE-PBDs assembled via electrostatics that could then be locked in place using chemically synthesized peptide linkers possessing specific PL sequences at both the N- and C-termini.

5.3 – Synthetic Metabolons Using Nanocontainer Scaffolds

The generation of novel metabolic pathways represents an exciting avenue for the enhanced industrial production of value-added synthons, fine chemicals, and advanced biofuels in a green manner from renewable carbon sources such as feedstock-derived carbohydrates and lignin (36-39). Using modern recombinant DNA and microbial engineering technologies, synthetic multistep metabolic pathways (i.e. metabolons) can be constructed both *in vivo* within common host organisms, such as *E. coli* and baker's yeast, and *in vitro* for one-pot syntheses. Within the last decade, synthetic biologists have placed a special emphasis on the artificial co-localization of metabolon enzymes on or within protein, DNA, and lipid-based scaffolds in order to mimic the functional benefits observed in nature for biocatalysts similarly co-localized with organelles and macromolecular enzyme assemblies (40, 41). In particular, multiple experimental examples have shown that brining biocatalysts into close spatial proximity through both natural and synthetic scaffolding mechanisms can result in enhanced metabolon flux due to reductions in the diffusion loss of pathway intermediates, the establishment of local concentration gradients, and limiting the influence of competing substrates or ligands (41-45). Additionally, synthetic scaffolds permit researchers the ability to modulate the relative stoichiometries of specific metabolon enzymes in order to eliminate pathway bottlenecks resulting from kinetic imbalances between sequentially-acting biocatalysts (42, 43).

In recent years, engineered proteinaceous nanocontainers have been utilized as highly uniform and monodisperse scaffolds in several instances for the co-localization of multienzyme biosensor and metabolon systems (10, 11, 46, 47). However, in all of these cases, biocatalysts were functionalized onto only one face of the nanocontainer shell: either the exterior surface or the interior luminal wall. The second half of chapter 2 and chapter 3 of this dissertation collectively describe the generation of a functional nanocontainer-templated metabolon (i.e. a multienzyme nanoreactor) that utilizes both the exterior and luminal surfaces as unique scaffolding points for the simultaneous localization of complimentary biocatalysts. Specifically, we first showed that the covalent SpyCatcher/SpyTag bacterial superglue system (48) could be used to effectively and uniformly adhere up to 60 copies of an *E. coli* dihydrofolate reductase (DHFR) variant onto the exterior surface of TmE nanocontainers near the containers' 5-fold symmetry axes. Subsequently, a THF-dependent aryl-*O*-demethylase derived from *S. paucimobilis* SYK-6, referred to as LigM, was non-covalently packaged within the encapsulins' luminal spaces by taking advantage of a known protein-protein interaction between a C-terminal cargo-loading peptide (Clp) abstracted from the encapsulins' native ferritin-like protein cargo and the interior surface of the encapsulin shell. Nanoreactors constructed using this "inside and out" approach were shown to be metabolically competent as a THF cofactor pool generated by surface-immobilized DHFR enzymes was capable of entering the encapsulin lumen to fuel the demethylase activity of encapsulated LigM catalysts. While the overall turnover of our two-step metabolon was approximately 5-fold lower than the turnover observed for non-scaffolded DHFR and LigM biocatalysts *in vitro*, the introduction of previously-established mutations designed to enlarge the 5-fold symmetry pores within the encapsulin shell were found to increase the nanoreactors' efficiency to match that of the non-scaffolded enzymes (49). These results indicate that the individual performances of the two biocatalysts were not adversely affected through incorporation into the nanoreactor assembly, but rather the exchange of substrates through the narrow 3 Å pores within the native encapsulin shell was the likely bottleneck resulting in the suppressed flux rates observed for the initial nanoreactor design.

Frustratingly, we observed none of the metabolon flux enhancements reported for other synthetically scaffolded metabolons with our encapsulin-based system due to potent product inhibition of our LigM biocatalyst by 5-CH_3-THF. Additionally, we observed the linear aerobic degradation of the THF cofactor pool generated by surface-immobilized DHFR enzymes due to the slower relative activity of the LigM enzymes in our engineered metabolon. The loss of THF can be mitigated by performing reactions under anaerobic conditions, though the retention of the THF pool alone was insufficient to outcompete the observed 5-CH_3-THF inhibition of LigM. Recently, Rosini *et al.* showed that coupling a plant-based MetE enzyme, which natively demethylates 5-CH_3-THF in order to synthesize L-methionine from L-homocysteine, can serve as a viable THF-cofactor recycling system capable of eliminating the product inhibition that cripples LigM turnover (50). We have thus begun work to couple the same MetE enzyme to our own nanoreactors in the hopes that the removal of 5-CH_3-THF will allow us to both increase the turnover of our encapsulated LigM catalysts and to subsequently better understand if further nanoreactor modifications are needed to enhance pathway flux for our artificial metabolon. Specifically, previous works with scaffolded multienzyme systems have emphasized that careful balancing of the individual kinetic properties of metabolon enzymes is often necessary to achieve efficient pathway flux (10, 42, 43). Our nanoreactor system innately possesses a higher stoichiometry of DHFR enzymes relative to LigM enzymes due to the larger surface area of the encapsulin outer surface versus the limited interior volume of the encapsulin lumen. Both DHFR and LigM have nearly identical k_{cat} values, but the K_M for THF exhibited by LigM is two orders of magnitude higher than the corresponding K_M for DHF exhibited by our DHFR variant (51). We hope that removal of 5-CH_3-THF inhibition will allow us to determine if the excess of DHFR is in fact beneficial for providing sufficiently high THF concentrations to accommodate the lower inherent binding affinity of LigM, or if methods to alter the relative DHFR to LigM stoichiometry will be necessary to better optimize metabolon flux.

5.4 – *In vitro* Cargo Loading Strategies

The ability to reversibly disassemble purified protein nanocontainers *in vitro* represents a powerful strategy for the user-directed encapsulation of diverse non-native cargoes in predetermined stoichiometric ratios (7, 52-55). In many cases, nanocontainer disassembly and reassembly phases can be initiated by modulating solvent pH and ionic strength or by using mild levels of chemical denaturants (54, 56, 57). However, while several research groups have successfully directed the encapsulation of non-native protein and inorganic metal nanoparticle cargoes within TmE nanocontainers *in vitro* in recent years, the solvent conditions used to prompt TmE disassembly were significantly harsher, involving the use of extreme pH values or high concentrations of chaotropic salts (58, 59). Efforts to replicate these *in vitro* cargo loading methods for TmE in our own laboratory showed that while certain protein cargoes are tolerant of the harsh solution conditions, others are significantly less so as rapid aggregation was observed for several candidate enzyme cargoes we tested.

In light of these observations, Chapter 4 of this dissertation discusses our attempts to generate a novel methodology designed to simultaneously abrogate the native *in vivo* assembly of TmE nanocontainers and to introduce a user-triggerable "switch" that could allow the arrested container assembly to subsequently be re-initiated *in vitro*. More specifically, we sought to develop this methodology in order to allow researchers to perform *in vitro* cargo loading of encapsulins in a facile manner under benign solvent conditions. We subsequently showed that the native *in vivo* assembly of TmE nanocontainers could be arrested through the introduction of unfavorable steric clashes within the encapsulin lumen by genetically grafting a bulky 40.3 kDa *E. coli* MBP onto the N-terminus of the TmE gene. Further, encoded protease recognition sites located within a flexible linker sequence between the MBP and TmE genes could be used to liberate the TmE monomers and permit their subsequent assembly into wildtype-like nanocontainers. However, attempts to use this methodology, which we have deemed "cleavage-prompted assembly," to load GCaMP, a fluorescent reporter protein possessing a recombinantly-fused Clp affinity tag, into TmE nanocontainers *in vitro* were repeatedly unsuccessful. Attempts to optimize both the design of the MBP-TmE fusion proteins and the specific in-lab experimental processes were similarly ineffective.

Several of our experiments offer possible explanations as to our repeated failures to achieve effective *in vitro* cargo loading using our CPA method. Analytical size exclusion chromatography performed with both the original MBP-TmE fusion protein and the modified MBP-*TEVext*-TmE variant (the chromatogram for the latter variant was not shown in chapter 4 as it was highly similar to the one reported for MBP-TmE) showed the presence of multiple oligomeric states in both samples, indicating that the introduction of the MBP domains into the TmE lumen only somewhat abolishes the ability of TmE protomers to self-assembly during *in vivo* expression. Additionally, we noticed a persistent amount of un-cleaved fusion protein in many of our samples following treatment with protease, leading us to suspect that the liberated TmE monomers were able to rapidly fill in the gaps of the partially assembled fusion protein structures, effectively sealing them off without GCaMP ever gaining access to the TmE luminal spaces. Given the close proximity of the TmE N-terminus to the Clp binding cleft, it is also unlikely that GCaMP is able to pre-bind to the luminal faces of the TmE monomers prior to CPA, further hampering their encapsulation.

The native assembly pathways for encapsulins of any size (i.e. T = 1, 3, or 4) have, to the best of our knowledge, been largely underexplored to date. High resolution mass spectrometry analyses of the 60-mer encapsulins originating from *Brevibacterium linens* indicate that dimers of the encapsulin shell protomers likely represent the smallest functional shell units, and that assembly of the full shell structure occurs via the polymerization of dimeric subunits (60). While the affiliated kinetics of the encapsulin nanocontainer assembly process remain unknown, previous *in vitro* assembly studies with similarly sized icosahedral viral capsids have elucidated several common trends relating to the kinetic behaviors of capsid construction (61, 62). In particular, many icosahedral capsids undergo assembly according to a sigmoidal, "nucleation-and-growth" process in which the initial assembly of low number oligomeric subunits is slow (61, 63, 64). Once a sufficiently large base oligomer is established, however, it serves as a nucleation point for the rapid polymerization of the full capsid structure with few to no equilibrium intermediates detected in between the two oligomeric states, indicating that the capsid assembly process follows a steep downhill energy landscape under physiological conditions (61, 63, 65, 66). For some viral capsids, the rapid assembly phase is completed in under one second at rates exceeding the diffusion limit (65). Given that encapsulins are evolutionarily related to viral capsids and that several low number oligomeric states were detected in conjunction with the fully assembled *B. linens* encapsulins in the previously referenced high resolution mass spectrometry analyses, it is entirely possible that encapsulins undergo a similar rapid assembly process *in vitro* (60, 67). Such rapid assembly would further support our notion that proteolytically liberated TmE protomers in our CPA method could be assembling at rates that do not leave sufficient time for effective encapsulation of our GCaMP reporter proteins.

Another observed trait of viral capsids related to their "nucleation-and-growth" behavior is that many viral capsids will not assemble *in vitro* until a sufficiently high concentration of capsid shell proteins is reached, presumably due to the inability to form sufficiently large nucleation centers capable of triggering the rapid assembly phase (64, 68, 69). Recently, Li *et al.* were able to determine the apparent critical assembly concentration (CAC_{app}), which represents the minimum capsid monomer (or minimum biological oligomer) concentration in solution that will incite full capsid assembly, for VLPs derived from the bacteriophage MS2 (69). Adapting the reported methodologies, we attempted to determine the CAC_{app} value for TmE nanocontainers by disassembling the encapsulins under acidic conditions, diluting the resulting monomers to varying concentrations between 1 and 11 μM, and then gradually dialyzing the samples back into physiological-like buffer conditions to promote reassembly of the nanocontainers. However, while were able to show that we could both disassemble and reassemble the encapsulins using the acid treatment followed by a return to physiological pH using dialysis, we were unable to determine a CAC_{app} value as all of the concentrations tested showed complete reassembly of the encapsulin nanocontainers with no observable disassembled protomers remaining in analytical SEC traces. These data indicate that the CAC_{app} value for TmE is likely in the nanomolar range, and thus the TmE protomers have a strong affinity for one another. Taken collectively, these data also indicate that TmE nanocontainers likely assemble rapidly *in vitro*.

In summary, the CPA methodology described in chapter 4 does not currently represent a viable strategy for the encapsulation of non-native cargoes in TmE nanocontainers. However, we believe that it does represent a starting point from which a more refined system could be rationally engineered. For example, one could ask the question as to whether or not the sub-micromolar CAC_{app} value observed for TmE is inherent to all encapsulins, or whether this observed high affinity of the TmE protomers for one another is in fact the result of these encapsulins originating from an extremophilic organism. If so, then perhaps the CPA methodology is better suited to encapsulins arising from mesophilic organisms, such as *Brevibacterium linens* or *Mxyococcus xanthus* (67, 70), as these homologues may not experience the same degrees of container pre-assembly that we believe is a major factor hampering effective cargo loading with TmE. Alternatively, alterations to the steric-based obstruction of encapsulin assemblies by engineering new tethering points for the MBP (or another sterically bulky protein) could both reduce encapsulin pre-assembly while also granted the non-native cargoes greater access to the luminally-oriented Clp binding clefts prior to the initiation of encapsulin assembly.

5.5 – References

1. Ashley, C. E.; Carnes, E. C.; Phillips, G. K.; Durfee, P. N.; Buley, M. D.; Lino, C. A.; Padilla, D. P.; Phillips, B.; Carter, M. B.; Willman, C. L.; Brinker, C. J.; Caldeira, J. d. C.; Chackerian, B.; Wharton, W.; Peabody, D. S., Cell-Specific Delivery of Diverse Cargos by Bacteriophage MS2 Virus-like Particles. *ACS Nano* **2011**, *5* (7), 5729-5745.

2. Moon, H.; Lee, J.; Kim, H.; Heo, S.; Min, J.; Kang, S., Genetically engineering encapsulin protein cage nanoparticle as a SCC-7 cell targeting optical nanoprobe. *Biomaterials Research* **2014**, *18* (1), 21.

3. Moon, H.; Lee, J.; Min, J.; Kang, S., Developing Genetically Engineered Encapsulin Protein Cage Nanoparticles as a Targeted Delivery Nanoplatform. *Biomacromolecules* **2014**, *15* (10), 3794-3801.

4. Min, J.; Kim, S.; Lee, J.; Kang, S., Lumazine synthase protein cage nanoparticles as modular delivery platforms for targeted drug delivery. *RSC Advances* **2014**, *4* (89), 48596-48600.

5. Choi, B.; Moon, H.; Hong, S. J.; Shin, C.; Do, Y.; Ryu, S.; Kang, S., Effective Delivery of Antigen–Encapsulin Nanoparticle Fusions to Dendritic Cells Leads to Antigen-Specific Cytotoxic T Cell Activation and Tumor Rejection. *ACS Nano* **2016**, *10* (8), 7339-7350

6. Truffi, M.; Fiandra, L.; Sorrentino, L.; Monieri, M.; Corsi, F.; Mazzucchelli, S., Ferritin nanocages: A biological platform for drug delivery, imaging and theranostics in cancer. *Pharmacological Research* **2016**, *107*, 57-65.

7. Comellas-Aragones, M.; Engelkamp, H.; Claessen, V. I.; Sommerdijk, N. A.; Rowan, A. E.; Christianen, P. C.; Maan, J. C.; Verduin, B. J.; Cornelissen, J. J.; Nolte, R. J., A virus-based single-enzyme nanoreactor. *Nature Nanotechnology* **2007**, *2* (10), 635-639.

8. Minten, I. J.; Claessen, V. I.; Blank, K.; Rowan, A. E.; Nolte, R. J. M.; Cornelissen, J. J. L. M., Catalytic capsids: the art of confinement. *Chemical Science* **2011**, *2* (2), 358-362.

9. Patterson, D.; Edwards, E.; Douglas, T., Hybrid Nanoreactors: Coupling Enzymes and Small-Molecule Catalysts within Virus-Like Particles. *Israel Journal of Chemistry* **2015**, *55* (1), 96-101.

10. Patterson, D. P.; Schwarz, B.; Waters, R. S.; Gedeon, T.; Douglas, T., Encapsulation of an Enzyme Cascade within the Bacteriophage P22 Virus-Like Particle. *ACS Chemical Biology* **2014**, *9* (2), 359-365.

11. Giessen, T. W.; Silver, P., A catalytic nanoreactor based on in vivo encapsulation of multiple enzymes in an engineered protein nanocompartment. *ChemBioChem* **2016**, *17*, 1931-1935.

12. Wilkerson, J. W.; Yang, S.-O.; Funk, P. J.; Stanley, S. K.; Bundy, B. C., Nanoreactors: Strategies to encapsulate enzyme biocatalysts in virus-like particles. *New Biotechnology* **2018**, *44*, 59-63.

13. Ren, H.; Zhu, S.; Zheng, G., Nanoreactor Design Based on Self-Assembling Protein Nanocages. *International Journal of Molecular Sciences* **2019**, *20* (3).

14. Li, F.; Wang, Q., Fabrication of Nanoarchitectures Templated by Virus-Based Nanoparticles: Strategies and Applications. *Small* **2014**, *10* (2), 230-245.

15. Nguyen, T. K.; Ueno, T., Engineering of protein assemblies within cells. *Current Opinion in Structural Biology* **2018**, *51*, 1-8.

16. Uchida, M.; McCoy, K.; Fukuto, M.; Yang, L.; Yoshimura, H.; Miettinen, H. M.; LaFrance, B.; Patterson, D. P.; Schwarz, B.; Karty, J. A.; Prevelige, P. E., Jr.; Lee, B.; Douglas, T., Modular Self-Assembly of Protein Cage Lattices for Multistep Catalysis. *ACS Nano* **2018**, *12* (2), 942-953.

17. McCoy, K.; Uchida, M.; Lee, B.; Douglas, T., Templated Assembly of a Functional Ordered Protein Macromolecular Framework from P22 Virus-like Particles. *ACS Nano* **2018**, *12* (4), 3541-3550.

18. Zhang, G.; Quin, M. B.; Schmidt-Dannert, C., Self-Assembling Protein Scaffold System for Easy in Vitro Coimmobilization of Biocatalytic Cascade Enzymes. *ACS Catalysis* **2018**, *8* (6), 5611-5620.

19. Mokili, J. L.; Rohwer, F.; Dutilh, B. E., Metagenomics and future perspectives in virus discovery. *Current Opinion in Virology* **2012**, *2* (1), 63-77.

20. Giessen, T. W.; Silver, P. A., Widespread distribution of encapsulin nanocompartments reveals functional diversity. *Nature Microbiology* **2017**, *2*, 17029.

21. Munang'andu, H. M.; Mugimba, K. K.; Byarugaba, D. K.; Mutoloki, S.; Evensen, Ø., Current Advances on Virus Discovery and Diagnostic Role of Viral Metagenomics in Aquatic Organisms. *Frontiers in Microbiology* **2017**, *8* (406).

22. Tracey, J. C.; Coronado, M.; Giessen, T. W.; Lau, M. C. Y.; Silver, P. A.; Ward, B. B., The Discovery of Twenty-Eight New Encapsulin Sequences, Including Three in Anammox Bacteria. *Scientific Reports* **2019**, *9* (1), 20122.

23. Pasin, F.; Menzel, W.; Daròs, J.-A., Harnessed viruses in the age of metagenomics and synthetic biology: an update on infectious clone assembly and biotechnologies of plant viruses. *Plant Biotechnology Journal* **2019**, *17* (6), 1010-1026.

24. Suttle, C. A., Marine viruses — major players in the global ecosystem. *Nature Reviews Microbiology* **2007**, *5* (10), 801-812.

25. Asor, R.; Ben-Nun-Shaul, O.; Oppenheim, A.; Raviv, U., Crystallization, Reentrant Melting, and Resolubilization of Virus Nanoparticles. *ACS Nano* **2017**, *11* (10), 9814-9824.

26. Künzle, M.; Eckert, T.; Beck, T., Metal-Assisted Assembly of Protein Containers Loaded with Inorganic Nanoparticles. *Inorganic Chemistry* **2018**, *57* (21), 13431-13436.

27. Kostiainen, M. A.; Hiekkataipale, P.; de la Torre, J. Á.; Nolte, R. J. M.; Cornelissen, J. J. L. M., Electrostatic self-assembly of virus–polymer complexes. *Journal of Materials Chemistry* **2011**, *21* (7), 2112-2117.

28. Kostiainen, M. A.; Kasyutich, O.; Cornelissen, J. J. L. M.; Nolte, R. J. M., Self-assembly and optically triggered disassembly of hierarchical dendron–virus complexes. *Nature Chemistry* **2010**, *2* (5), 394-399.

29. Mikkilä, J.; Rosilo, H.; Nummelin, S.; Seitsonen, J.; Ruokolainen, J.; Kostiainen, M. A., Janus-Dendrimer-Mediated Formation of Crystalline Virus Assemblies. *ACS Macro Letters* **2013**, *2* (8), 720-724.

30. Künzle, M.; Eckert, T.; Beck, T., Binary Protein Crystals for the Assembly of Inorganic Nanoparticle Superlattices. *Journal of the American Chemical Society* **2016**, *138* (39), 12731-12734.

31. Liljeström, V.; Mikkilä, J.; Kostiainen, M. A., Self-assembly and modular functionalization of three-dimensional crystals from oppositely charged proteins. *Nature Communications* **2014**, *5*, 4445.

32. Kostiainen, M. A.; Hiekkataipale, P.; Laiho, A.; Lemieux, V.; Seitsonen, J.; Ruokolainen, J.; Ceci, P., Electrostatic assembly of binary nanoparticle superlattices using protein cages. *Nature Nanotechnology* **2012**, *8*, 52.

33. Nguyen, J. T.; Turck, C. W.; Cohen, F. E.; Zuckermann, R. N.; Lim, W. A., Exploiting the basis of proline recognition by SH3 and WW domains: design of N-substituted inhibitors. *Science* **1998**, *282* (5396), 2088-92.

34. Gilcrease, E. B.; Winn-Stapley, D. A.; Hewitt, F. C.; Joss, L.; Casjens, S. R., Nucleotide sequence of the head assembly gene cluster of bacteriophage L and decoration protein characterization. *Journal of Bacteriology* **2005**, *187* (6), 2050-7.

35. Uchida, M.; LaFrance, B.; Broomell, C. C.; Prevelige, P. E., Jr.; Douglas, T., Higher order assembly of virus-like particles (VLPs) mediated by multi-valent protein linkers. *Small* **2015**, *11* (13), 1562-70.

36. Peralta-Yahya, P. P.; Zhang, F.; del Cardayre, S. B.; Keasling, J. D., Microbial engineering for the production of advanced biofuels. *Nature* **2012**, *488* (7411), 320-328.

37. Jullesson, D.; David, F.; Pfleger, B.; Nielsen, J., Impact of synthetic biology and metabolic engineering on industrial production of fine chemicals. *Biotechnology Advances* **2015**, *33* (7), 1395-1402.

38. Narcross, L.; Fossati, E.; Bourgeois, L.; Dueber, J. E.; Martin, V. J. J., Microbial Factories for the Production of Benzylisoquinoline Alkaloids. *Trends in Biotechnology* **2016**, *34* (3), 228-241.

39. Wu, W.; Dutta, T.; Varman, A. M.; Eudes, A.; Manalansan, B.; Loque, D.; Singh, S., Lignin Valorization: Two Hybrid Biochemical Routes for the Conversion of Polymeric Lignin into Value-added Chemicals. *Scientific Reports* **2017**, *7* (1), 8420.

40. Chen, A. H.; Silver, P. A., Designing biological compartmentalization. *Trends in Cell Biology* **2012**, *22* (12), 662-670.

41. Lee, H.; DeLoache, W. C.; Dueber, J. E., Spatial organization of enzymes for metabolic engineering. *Metabolic Engineering* **2012**, *14* (3), 242-251.

42. Dueber, J. E.; Wu, G. C.; Malmirchegini, G. R.; Moon, T. S.; Petzold, C. J.; Ullal, A. V.; Prather, K. L. J.; Keasling, J. D., Synthetic protein scaffolds provide modular control over metabolic flux. *Nature Biotechnology* **2009**, *27* (8), 753-U107.

43. Moon, T. S.; Dueber, J. E.; Shiue, E.; Prather, K. L., Use of modular, synthetic scaffolds for improved production of glucaric acid in engineered E. coli. *Metabolic Engineering* **2010**, *12* (3), 298-305.

44. Wang, Y.; Yu, O., Synthetic scaffolds increased resveratrol biosynthesis in engineered yeast cells. *Journal of Biotechnology* **2012**, *157* (1), 258-260.

45. You, C.; Zhang, Y. H. P., Self-Assembly of Synthetic Metabolons through Synthetic Protein Scaffolds: One-Step Purification, Co-immobilization, and Substrate Channeling. *ACS Synthetic Biology* **2013**, *2* (2), 102-110.

46. Aljabali, A. A. A.; Barclay, J. E.; Steinmetz, N. F.; Lomonossoff, G. P.; Evans, D. J., Controlled immobilisation of active enzymes on the cowpea mosaic virus capsid. *Nanoscale* **2012**, *4* (18), 5640-5645.

47. Koch, C.; Wabbel, K.; Eber, F. J.; Krolla-Sidenstein, P.; Azucena, C.; Gliemann, H.; Eiben, S.; Geiger, F.; Wege, C., Modified TMV Particles as Beneficial Scaffolds to Present Sensor Enzymes. *Frontiers in Plant Science* **2015**, *6* (1137).

48. Zakeri, B.; Fierer, J. O.; Celik, E.; Chittock, E. C.; Schwarz-Linek, U.; Moy, V. T.; Howarth, M., Peptide tag forming a rapid covalent bond to a protein, through engineering a bacterial adhesin. *Proceedings of the National Academy of Sciences USA* **2012**, *109* (12), E690-7.

49. Williams, E. M.; Jung, S. M.; Coffman, J. L.; Lutz, S., Pore Engineering for Enhanced Mass Transport in Encapsulin Nanocompartments. *ACS Synthetic Biology* **2018**, *7* (11), 2514-2517.

50. Rosini, E.; D'Arrigo, P.; Pollegioni, L., Demethylation of vanillic acid by recombinant LigM in a one-pot cofactor regeneration system. *Catalysis Science & Technology* **2016**, *6* (21), 7729-7737.

51. Kohler, A. C.; Mills, M. J. L.; Adams, P. D.; Simmons, B. A.; Sale, K. L., Structure of aryl-O-demethylase offers molecular insight into a catalytic tyrosine-dependent mechanism. *Proceedings of the National Academy of Sciences USA* **2017**, *114* (16), E3205.

52. Rurup, W. F.; Verbij, F.; Koay, M. S. T.; Blum, C.; Subramaniam, V.; Cornelissen, J., Predicting the Loading of Virus-Like Particles with Fluorescent Proteins. *Biomacromolecules* **2014**, *15* (2), 558-563.

53. Sharma, J.; Douglas, T., Tuning the catalytic properties of P22 nanoreactors through compositional control. *Nanoscale* **2020**, *12* (1), 336-346.

54. Sharma, J.; Uchida, M.; Miettinen, H. M.; Douglas, T., Modular interior loading and exterior decoration of a virus-like particle. *Nanoscale* **2017**, *9* (29), 10420-10430.

55. Zschoche, R.; Hilvert, D., Diffusion-Limited Cargo Loading of an Engineered Protein Container. *Journal of the American Chemical Society* **2015**, *137* (51), 16121-16132.

56. Lee, K. W.; Tan, W. S., Recombinant hepatitis B virus core particles: Association, dissociation and encapsidation of green fluorescent protein. *Journal of Virological Methods* **2008**, *151* (2), 172-180.

57. Wu, W.; Hsiao, S. C.; Carrico, Z. M.; Francis, M. B., Genome-Free Viral Capsids as Multivalent Carriers for Taxol Delivery. *Angewandte Chemie International Edition* **2009**, *48* (50), 9493-9497.

58. Cassidy-Amstutz, C.; Oltrogge, L.; Going, C. C.; Lee, A.; Teng, P.; Quintanilla, D.; East-Seletsky, A.; Williams, E. R.; Savage, D. F., Identification of a Minimal Peptide Tag for in Vivo and in Vitro Loading of Encapsulin. *Biochemistry* **2016**, *55* (24), 3461-3468.

59. Künzle, M.; Mangler, J.; Lach, M.; Beck, T., Peptide-directed encapsulation of inorganic nanoparticles into protein containers. *Nanoscale* **2018**, *10* (48), 22917-22926.

60. Snijder, J.; Kononova, O.; Barbu, I. M.; Uetrecht, C.; Rurup, W. F.; Burnley, R. J.; Koay, M. S. T.; Cornelissen, J. J. L. M.; Roos, W. H.; Barsegov, V.; Wuite, G. J. L.; Heck, A. J. R., Assembly and Mechanical Properties of the Cargo-Free and Cargo-Loaded Bacterial Nanocompartment Encapsulin. *Biomacromolecules* **2016**, *17* (8), 2522-2529

61. Mateu, M. G., Assembly, stability and dynamics of virus capsids. *Archives of Biochemistry and Biophysics* **2013**, *531* (1), 65-79.

62. Porterfield, J. Z.; Zlotnick, A., An Overview of Capsid Assembly Kinetics. In Emerging Topics in Physical Virology, pp 131-158.

63. Hagan, M. F.; Elrad, O. M., Understanding the concentration dependence of viral capsid assembly kinetics--the origin of the lag time and identifying the critical nucleus size. *Biophysical Journal* **2010**, *98* (6), 1065-1074.

64. Zlotnick, A.; Johnson, J. M.; Wingfield, P. W.; Stahl, S. J.; Endres, D., A Theoretical Model Successfully Identifies Features of Hepatitis B Virus Capsid Assembly. *Biochemistry* **1999**, *38* (44), 14644-14652.

65. Kler, S.; Asor, R.; Li, C.; Ginsburg, A.; Harries, D.; Oppenheim, A.; Zlotnick, A.; Raviv, U., RNA Encapsidation by SV40-Derived Nanoparticles Follows a Rapid Two-State Mechanism. *Journal of the American Chemical Society* **2012**, *134* (21), 8823-8830.

66. Lutomski, C. A.; Lyktey, N. A.; Pierson, E. E.; Zhao, Z.; Zlotnick, A.; Jarrold, M. F., Multiple Pathways in Capsid Assembly. *Journal of the American Chemical Society* **2018**, *140* (17), 5784-5790.

67. Sutter, M.; Boehringer, D.; Gutmann, S.; Guenther, S.; Prangishvili, D.; Loessner, M. J.; Stetter, K. O.; Weber-Ban, E.; Ban, N., Structural basis of enzyme encapsulation into a bacterial nanocompartment. *Nature Structural and Molecular Biology* **2008**, *15* (9), 939-947.

68. Johnson, J. M.; Tang, J.; Nyame, Y.; Willits, D.; Young, M. J.; Zlotnick, A., Regulating Self-Assembly of Spherical Oligomers. *Nano Letters* **2005**, *5* (4), 765-770.

69. Li, L.; Xu, C.; Zhang, W.; Secundo, F.; Li, C.; Zhang, Z.-P.; Zhang, X.-E.; Li, F., Cargo-Compatible Encapsulation in Virus-Based Nanoparticles. *Nano Letters* **2019**, *19* (4), 2700-2706.

70. McHugh, C. A.; Fontana, J.; Nemecek, D.; Cheng, N. Q.; Aksyuk, A. A.; Heymann, J. B.; Winkler, D. C.; Lam, A. S.; Wall, J. S.; Steven, A. C.; Hoiczyk, E., A virus capsid-like nanocompartment that stores iron and protects bacteria from oxidative stress. *The EMBO Journal* **2014**, *33* (17), 1896-1911.