Title
Shellular biology – exploring the biochemistry and physiology of a protein nanocompartment

Author
Cassidy-Amstutz, Caleb

Publication Date
2017

Peer reviewed|Thesis/dissertation

Shellular biology – exploring the biochemistry and physiology of a protein nanocompartment

By

Caleb Cassidy-Amstutz

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Molecular and Cell Biology

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor David F. Savage, Chair
Professor Susan Marqusee
Professor Ming Hammond
Professor John Dueber

Spring 2017

Abstract

Shellular biology – exploring the biochemistry and physiology of a protein nanocompartment

by

Caleb Cassidy-Amstutz

Doctor of Philosophy in Molecular and Cellular Biology

University of California, Berkeley

Professor David F. Savage, Chair

Bacteria and Archaea use proteinaceous compartments for a variety of different reasons such as creating unique chemical environments or isolating pathways with toxic intermediates. Recently, a new class of protein compartment was discovered and named for its founding member, encapsulin. Encapsulins are structural similar to bacterial microcompartments but are smaller; hence, encapsulins are interchangeably called nanocompartments. Encapsulins assemble into 24-32 nm icosahedral shells and usually package a single type of cargo protein such as a ferritin-like protein or a peroxidase. Cargo proteins are directed to the interior of encapsulins through an interaction between the encapsulin lumen and the C-terminus of the cargo protein. With the recent discovery of encapsulins, there are many fundamental questions to answer in regards to both the biochemistry and physiology of nanocompartments. In this work, I began by heterologously expressing and then purifying the *T. maritima* encapsulin. I was able to show the subunits readily and robustly assemble into nanocompartments and these nanocompartments are resistant to both thermal and chemical insults. Through fusion of the C-terminus from a cargo protein to GFP, I was able to load encapsulins with GFP both *in vitro* and *in vivo*. However, *in vitro* loading was only accomplished by disassembling and reassembling encapsulin in the presence of targeting, suggesting that *in vivo* encapsulins are only competent to load cargo during assembly. While purifying *T. maritima* encapsulin, I discovered the shell is a flavoprotein that co-purifies with riboflavin and flavin mononucleotide. Cryo-electron microscopy identified the flavin binding site as being adjacent to a surface tryptophan and mutation of the tryptophan to alanine substantially reduced the amount of flavin that co-purified with encapsulins. Flavins are common cellular electron transporters and the presence of a flavin tightly bound to an encapsulin shell suggests it may be related to the functioning of the native cargo, a ferritin-like protein that oxidizes ferrous iron and forms an iron oxide mineral. I am currently investigating whether the bound flavins act as electron donors or acceptors for the reduction or oxidation of iron.

This work was one of the first biophysical characterizations of an encapsulin and the first suggestion that nanocompartments may not be passive barriers separating the encapsulin interior from the cytosol. Instead, encapsulins may play a direct and active role in the function of encapsulated cargo.

Acknowledgements
John Donne wrote, “No man is an island entire of itself” and that is especially true in graduate school. I would not have reached graduation if not for the professional and personal support from many individuals.

First, I would like to thank my advisor Dave Savage who is a wonderful mentor and boss. His positivity and attitude towards science has made the lab a wonderful place to research and grow as a scientist. I would also like to thank all the members of the Savage lab past and present for making it a fun and collaborative lab. In particular, I would like to thank Rob Nichols who is the other member of team nanocompartments and Avi Flamholz who kicked my butt at squash and pushed my science along through engaging conversations. I extend my thanks and gratitude to the members of my committee: Susan Marqusee, John Dueber, and Ming Hammond who helped direct and guide my scientific pursuits.

Earning a doctorate is the culmination of an interest in science that began over two decades ago. Along the way, I had numerous mentors who guided and refined my scientific interests. I thank my first mentors Jill Keeney and Catherine Stenson at Juniata College for believing in me and providing my first research opportunities. My path to a doctorate was full of turns and I’m thankful for Eric Stone and Traci Hall who helped me realize Bioinformatics wasn’t for me and supported my decision to pursue a different direction.

Graduate school, while ultimately rewarding, has had times of frustration and sadness. I am deeply thankful for all my friends, without whom I would not have persevered. Rachel Hood, who joined the Savage lab with me, has been my constant friend and wonderful lab mate for these past six years. It’s hard to praise her enough but one thing I’m particularly thankful for is getting to experience much of the beauty of the California wilderness with her and Brett Robinson. I will never forget our trip to the summit of Mt. Shasta. Having a weekly gathering to explore new movies and television shows was a delight and it was made even better by getting to share it with Rachel, Brett, Michelle Bloom, Kevin Tsui, and Carolyn Elya. Lastly, I must thank Rick Rune for being a truly amazing friend who was always there for me even at my lowest points.

I want to thank my family members who have always supported me such as my younger brother, Christian Cassidy-Amstutz. I owe a great deal to my older brother, Andrew Cassidy-Amstutz, and his wife, Julia Strasser, especially for feeding and sheltering me when I visit DC. I’m particularly thankful for my niece, Hannah, who has been an unending source of joy during my final year in graduate school. Going home after long nights and failed experiments was always made better by Mittens who was happy to see and ready to snuggle. Last and in no way least, I want to thank my mom, Pamela Whitenack, and dad, Marley Amstutz, who have provided me unconditional love and support and without whom I would not be here today.

Table of Contents

List of Figures

List of Tables

Chapter 1 Encapsulins: molecular biology of the shell

Adapted from Encapsulins: molecular biology of the shell

Robert J. Nichols*, Caleb Cassidy-Amstutz*, Thawatchai Chaijarasphong, David F. Savage

In Submission

1.1 Abstract

Compartmentalization is both a fundamental principle of cellular organization and an emerging theme in prokaryotic biology. Work in the past few decades has shown that protein-based organelles called microcompartments enhance the function of encapsulated cargo proteins. More recently, the repertoire of known prokaryotic organelles has expanded beyond microcompartments to include a new class of smaller proteinaceous compartments, termed nanocompartments (also known as encapsulins). Like microcompartments, nanocompartments are icosahedral capsids, but they are smaller and less complex. Encapsulins are formed by a single species of shell protein that self-assembles and typically encapsulates only one type of cargo protein. Significant progress has been made in understanding the structure of nanocompartment shells and the loading of cargo to the interior. Recent analysis has also demonstrated the prevalence of encapsulin genes throughout prokaryotic genomes and documented a large diversity of cargo proteins with a variety of novel functions, suggesting that nanocompartments play an important role in many microbes. Here we review the current understanding of encapsulin structure and function and highlight exciting open questions of physiological significance.

1.2 Introduction

Compartmentalization is a widespread feature within cells and serves a number of physiological roles[1]. Most commonly, compartmentalization is used for the spatial organization of metabolism and enables a variety of functions, including activation of thermodynamically unfavorable reactions, sequestering reactions to prevent off-target interactions, and isolating pathways that contain a toxic intermediate. While eukaryotic membrane-bound organelles represent the traditional view of cellular compartments, many prokaryotes rely on proteinaceous compartments that act in an analogous fashion[2–5]. These proteinaceous compartments, also known as bacterial microcompartments (BMCs), were first identified in 1956 and are large icosahedral structures 80 – 400 nm in diameter[2–4,6,7]. BMCs have been identified in many bacterial phyla and compartmentalize a variety of metabolic reactions, ranging from catabolism of ethanolamine and propanediol to anabolic CO_2 fixation[2–4,8].

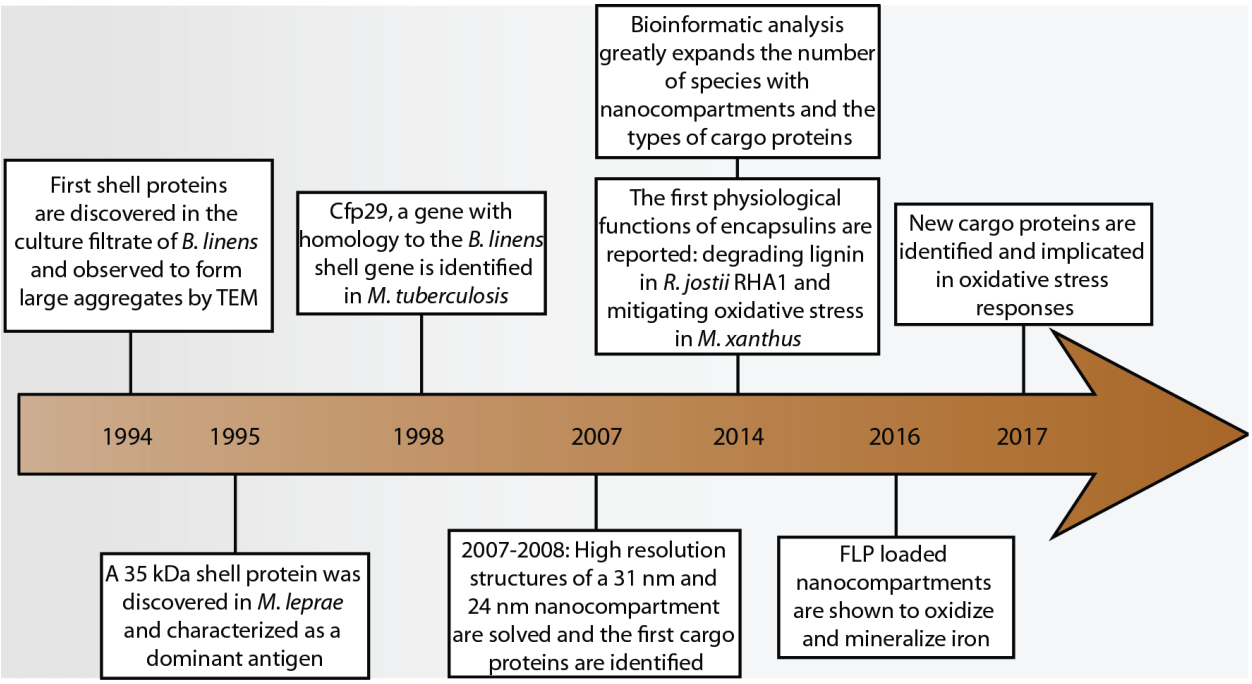

Figure 1-1: Timeline: A brief history of the discovery and characterization of prokaryotic nanocompartments.

Recently, a novel class of prokaryotic compartments, collectively called encapsulins, was identified. Encapsulin proteins share no sequence or structural homology to BMC proteins and form significantly smaller capsids (20 - 30 nm in diameter) than BMCs. Fittingly, encapsulins are also known as nanocompartments (hereafter, we use both terms interchangeably). The historical timeline (Figure 1-1) of nanocompartment biology begins in 1994, when the first nanocompartment was identified and characterized as a high molecular weight aggregate produced by *Brevibacterium linens*[9,10]. The nanocompartment was initially characterized as a bacteriocin and possible protease, but its bactericidal activity could be not validated once the complex was purified to homogeneity[9,11,12]. In other instances, nanocompartment proteins, although not realized as such at the time, were discovered during studies of pathogenesis in *Mycobacterium leprae*, and studies of sporulation-

specific proteins in *Streptomyces*, but remained largely uncharacterized[10,13–16]. In the mid-2000s, these early observations were clarified when crystal structures of the nanocompartments from *Pyrococcus furiosus* and *Thermotoga maritima* revealed that the observed high-molecular weight aggregates were capsid-like complexes loaded with additional proteins[12,17]. These structures provided significant insights into shell assembly and cargo loading and spurred new research into prokaryotic nanocompartments[12,17]. Recently, studies have focused on using encapsulin as a programmable nanoreactor or nanomaterial[18–24] while others have explored the physiological role of encapsulin and its cargo in their native bacterial context[25–29]. Bioinformatic analysis of sequenced genomes has identified thousands of nanocompartment systems in both Bacteria and Archaea with a diversity of putative cargo proteins[29,30] and a recent article reviewed the potential of engineering nanocompartments for biotechnology applications[31]. Here, we review the recent structural and biochemical advances in encapsulin biology and highlight the many open questions related to bacterial physiology these discoveries raise.

1.3 Structure of the nanocompartment shell

Encapsulin proteins form nanocompartments that resemble viral capsid structures (Figure 1-2A). High resolution structures of nanocompartments from *P. furiosus*, *T. maritima*, and *M. xanthus* have revealed that their shells are icosahedral complexes formed by the self-assembly of a single species of protomer protein[12,17,26]. While these structures provided a great deal of information about the shell, they were largely unable to resolve cargo proteins within the encapsulin lumen and cargo packing remains largely uncharacterized, as described below. Encapsulin shell proteins are structurally homologous to the HK97 phage major capsid protein gp5[32] and homo-oligomerize to form the complete nanocompartment. As is common with viral capsids, homologous proteins can assemble into icosahedrons of different sizes – encapsulins from *P. furiosus* and *M. xanthus* are composed of 180 protomers (31 nm in diameter), while those of *T. maritima* are formed by 60 protomers (24 nm in diameter)[12]. This structural organization can be formally classified by the triangulation number, a system for describing the ratio of pentameric to hexameric faces in icosohedral structures[33]. Following this classification, the encapsulin shell protomer from *T. maritima* forms a T=1 icosahedron comprised only of 12 pentameric vertices, while the *P. furiosus* and *M. xanthus* encapsulins form T=3 icosahedrons with 12 pentamers forming the vertices and 20 hexamers forming the faces of the icosahedron[17]. Notably, all known nanocompartments are significantly smaller than the HK97 phage capsid, which forms a 420 protomer T=7 capsid[32] (Figure 1-2B).

The encapsulin protomer adopts an HK97 fold, which has three conserved domains: a peripheral domain (P), an axial domain (A), and an elongated loop (E)[12,17,32] (Figure 1-2C). While the *P. furiosus* encapsulin protomer aligns well with the HK97 gp5 protomer[17], only the A and E domains from the *T. maritima* encapsulin align well with those of the former[12]. Specifically, the E loop from *T. maritima* is shorter and rotated relative to the E loops of HK97 phage and *P. furiosus*[12]. This rotation (Figure 1-2D) allows the E loop to form a beta sheet with the E loop of a neighboring protomer, creating a tight interaction[2]. This feature is not observed in the *P. furiosus* nanocompartment or the HK97 phage[17,32]. The difference in the rotation of the E loop may explain why the *T. maritima* nanocompartment forms a T=1 capsid while the *P. furiosus* nanocompartment forms a larger T=3 capsid. Recent work supports this idea, demonstrating that T=3 nanocompartments have an insertion in their E loop compared to T=1 nanocompartments[29], implicating the E loop in determining triangulation number and the size of the nanocompartment.

Structural analysis of nanocompartments has also revealed that the capsids have multiple pores that may control the exchange of small molecules between the cytosol and the encapsulin lumen[12,17]. All encapsulins have three major types of openings: pores formed from the interaction between two protomers (two-fold pores), pores at the sites of five-fold symmetry (five-fold pores), and pores at the sites of three-fold symmetry (three-fold pores)[12] (Figure 1-2E). The openings are all 5-6 Å in diameter, but vary in their chemical nature. Two-fold pores are lined with multiple negatively-charged residues, while the five-fold pores are uncharged and often display a ring of histidines along the cytosolic face. Intriguingly, in the crystal structure of *T. maritima* encapsulin, these histidines appear to coordinate a metal ion, suggesting they could alter shell

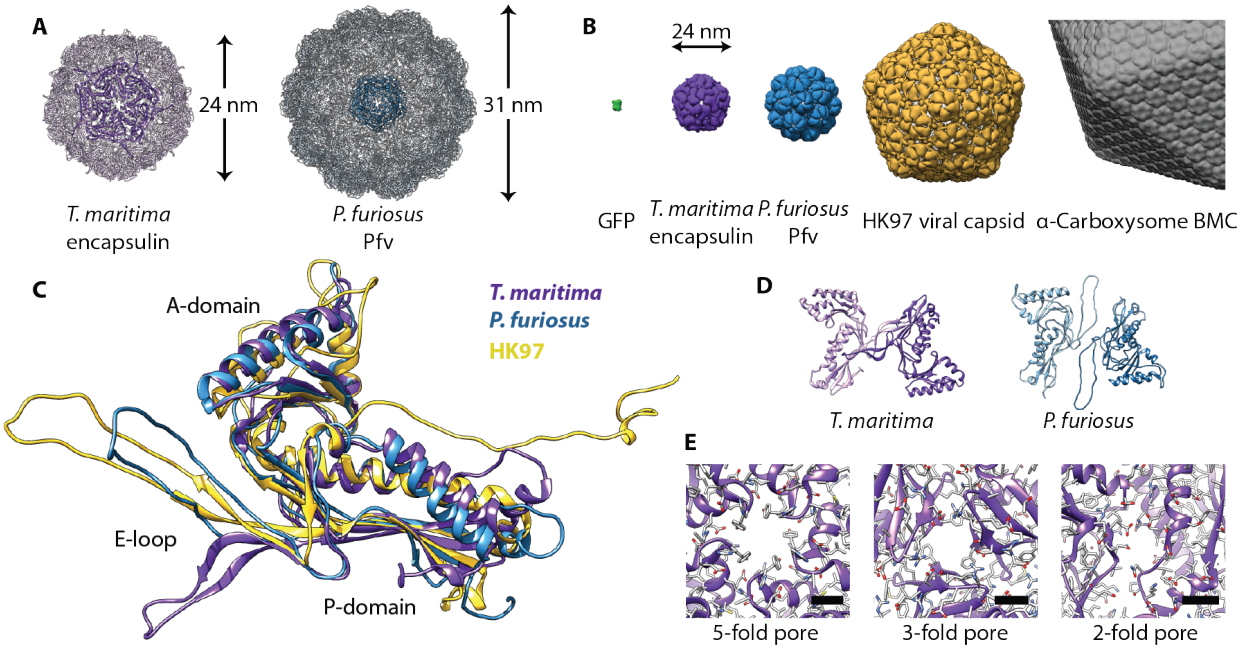

Figure 1-2: Structure of prokaryotic nanocompartments. (A). Nanocompartments from *T. maritima* [PDB: 3DKT] and *P. furiosus* [PDB: 2E0Z] with a vertex pentamer highlighted. (B). Size comparison of the nanocompartments to GFP [PDB: 1C4F], the HK97 viral capsid [PDB: 1OHG], and the carboxysome. Only a quarter of the carboxysome is displayed. (C). Alignment of a protomer from *T. maritima* encapsulin (purple), *P. furiosus* Pfv (blue), and the HK97 viral capsid (yellow). (D). A rotation of the E-loop in *T. maritima* encapsulin allows formation a beta-sheet interaction with a neighboring protomer while in *P. furiosus* the E-loop is not rotated and this interaction is not observed. (E). Three different types of openings are found in *T. maritima* encapsulin, a pore at the site of 5-fold symmetry, a pore at the site of 3-fold symmetry, and an opening at the dimer interface between 2 protomers. Black scale bars are 6 Å in length.

permeability and affect the activity of the ferritin-like protein cargo within[12]. The chemical nature of the three-fold pore is not conserved across species. In *T. maritima*, the three-fold pore is lined with positively charged residues[12], while in many other species it is predicted to be uncharged. Given that the chemical nature of the other two openings is well conserved across species, it is surprising the three-fold pore is not conserved. It is currently unknown whether the evolutionary variation at the three-fold pore has functional significance.

These openings likely serve as a permeability barrier to larger molecules, while permitting transit of small molecules and ions across the shell. There is ample evidence that substrates of encapsulated enzymes, such as hydrogen peroxide or ferrous iron, are able to cross the shell, while proteins are larger and cannot[21,22,25,28]. Whether molecules of intermediate size – e.g. larger metabolites – can readily enter nanocompartments remains an open question. For example, a nanocompartment from *Rhodococcus jostii* encapsulates a dye-decolorizing peroxidase (DyP) cargo and was reported to degrade nitrated lignin. Given that nitrated lignin (1-2 kDa) is substantially larger than the width of any of the openings in the nanocompartment, it is unclear how such transit is possible[25,34]. One possibility is that ends of the nitrated lignin chains are able to cross the pores. Alternatively, it has been proposed that *R. jostii* encapsulins disassemble upon recognition of lignin, thereby negating the need for lignin to cross the

shell, but this hypothesis remains untested[25]. The pore openings in the shell likely play an important role in permitting encapsulated enzymes access to their substrates. Future work will hopefully determine the permeability characteristics of model encapsulin's pores and elucidate how permeability affects the activity of cargo enzymes.

1.4 Loading of cargo proteins

Understanding the function of a nanocompartment begins with identifying its particular protein cargo. The first insights into the mechanism of cargo encapsulation came from the crystal structure of the *T. maritima* encapsulin (Figure 1-3A). A small amount of extra electron density was observed bound to a hydrophobic pocket on the lumenal face of the encapsulin shell, and corresponded to a short (~10 amino acid), C-terminal sequence of a ferritin-like protein (FLP) found next to to the encapsulin gene in the *T. maritima* genome[12]. Bioinformatic analysis found that this C-terminal sequence is conserved across species where the cargo protein and encapsulin genes are found together in a putative operon. Examples of such predicted cargo proteins include FLPs, dye decolorizing peroxidases (DyPs), hemerythrin, and rubrerythrin[12,29].

Multiple groups have since demonstrated that this C-terminal sequence, termed the cargo loading peptide (CLP), is sufficient for loading a cargo protein into the encapsulin complex (Figure 1-3B). Deletion of the CLP from a cargo protein abolishes encapsulation, while fusion of the CLP to the C-terminus of heterologous proteins, such as green fluorescent protein or luciferase, is sufficient for packaging[12,18,21,22,27,29]. Because these signature motifs are conserved for a given cargo type, they have aided in the identification of different families of nanocompartments, as discussed below in the Physiology of Nanocompartments section (Table 1-1)[30].

Encapsulin Family	Predicted Cargo
Family 1	Ferritin-like protein, bacterioferritin, rubrerythrin, hemerythrin, iron-mineralizing encapsulin-associated Firmicute cargo, ferredoxin, dye-decolorizing peroxidase, nitric oxide reductase-hydroxylamine oxireductase fusion
Family 2	Cysteine desulfurase, Crp-Fnr family transcriptional regulator-shell fusion
Family 3	Heme-dependent cytochrome P450-like monooxygenase hydroxylase, undecaprenyl-pyrophosphate binding protein
Family 4	Alpha ketoglutarate decarboxylase, 2-oxoglutarate dehydrogenase
Family 5	Inorganic pyrophosphatase
Family 6	Metal regulated compartment of unknown function
Family 7	Protease-shell fusion
Family 8	Divalent cation-dependent hydrolase

Table 1-1: Adapted from Radford 2015. List of characterized/ predicted encapsulin families and their predicted cargo proteins.

There are, however, alternatives modes of interaction between shell proteins and cargo (Figure 1-3B). In some cases, such as the *P. furiosus* encapsulin, the cargo protein lacks a CLP and is instead fused to the encapsulin gene as a single polypeptide[17]. In the encapsulin systems found in Firmicutes, loading may occur in part via an N-terminal CLP. This encapsulin system contains both an iron-mineralizing encapsulin-associated Firmicute (IMEF) protein with a standard C-terminal CLP, and a

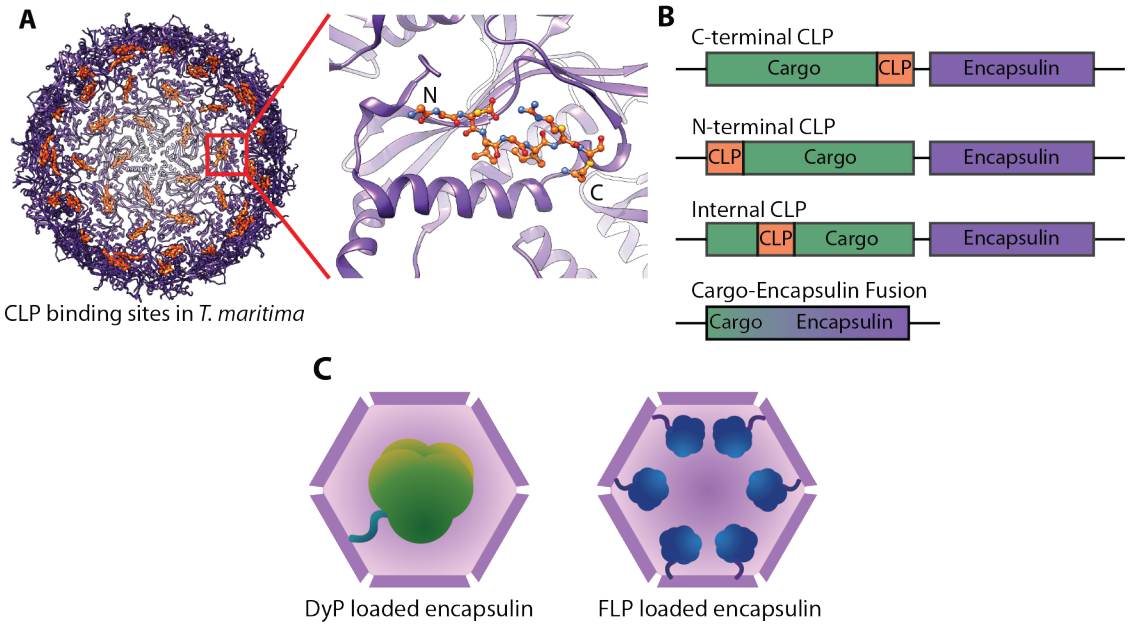

Figure 1-3: A cargo loading peptide (CLP) directs the encapsulation of cargo. (A).Each protomer has a lumenal hydrophobic pocket responsible for interacting with the CLP (B). While the CLP is usually located on one of the termini of the cargo, bioinformatics has identified cargo with an internal CLP sequence. In some cases, direct fusion of the cargo to the encapsulin gene negates the need for a CLP. (C). Model of nanocompartments with DyP or FLP bound. One DyP hexamer is bound per shell and up to 12 FLP decamers may be bound per shell.

ferredoxin protein with an N-terminal CLP[29]. Heterologous expression of shell, ferredoxin, and IMEF yielded nanocompartments containing only IMEF cargo, but co-expression of the shell and ferredoxin (without IMEF) demonstrated that ferredoxin can be loaded under these conditions[29]. This result raises questions – e.g., how prevalent are N-terminal CLPs? – and mechanistic possibilities, including modulating the stoichiometry of loading by employing CLPs with different affinities, and controlling encapsulin cargo loading by regulating the expression of cargo genes. Finally, it is also possible that CLPs located in the middle of a protein are functional. In a bioinformatic study, Radford detected putative cargo proteins that lacked N-terminal or C-terminal CLPs, yet contained internal CLP-like sequences[30] (Figure 1-3B). Whether these internal CLPs are sufficient for encapsulation is unknown.

There is also growing evidence that multiple cargo proteins can be loaded into a single encapsulin shell. In *M. xanthus*, three different proteins (EncB, EncC, and EncD) were found packaged into encapsulins purified from the native host[26]. Likewise, heterologous expression experiments, like the study of the IMEF Firmicute encapsulin above, indicate a broad potential for cargo targeting. Heterologous expression of an encapsulin from *Mycobacterium tuberculosis* with either DyP, a bacterioferritin (BfrB), or a folate biosynthesis enzyme (FolB) results in loaded nanocompartments[27]. Intriguingly, the cargo protein genes, BfrB, FolB, EncC, and EncD, are not located near their respective encapsulin gene, raising the question of how they are loaded. In many encapsulin systems, the shell gene and cargo gene form an operon suggesting encapsulins are loaded via co-translational assembly, but this has yet to experimentally proven (Figure 1-3B). Biochemical evidence supports encapsulins being loaded during assembly, so it is unclear how cells coordinate the co-expression of non-cistronic cargo

proteins to ensure efficient loading[21,22]. It should be noted that since assembly and loading are inherently sensitive to protein concentration, care should be taken when evaluating the results of heterologous expression experiments. Differences in protein levels and binding competition from other proteins for the CLP binding site may yield inconsistent results, as was seen for the ferredoxin and IMEF system previously mentioned[29]. Going forward, biochemical characterization of nanocompartments from their native hosts may be necessary to confirm the validity of potential cargo proteins.

Another important structural question related to loading is the occupancy of cargo within the nanocompartment. Each shell protein in an encapsulin has a CLP-binding site, but the volumetric constraints of the shell limit the amount of packaged cargo[12,19]. Modeling work has shown that for the larger cargo it is impossible to achieve a 1:1 ratio of cargo to protomer due to steric clashes[12]. However, there is no high resolution structure of a holo-encapsulin complex, and scant experimental data measuring the degree of cargo loading. The clearest indication to date comes from the *B. linens* nanocompartment, whose cargo is a DyP that assembles into a hexamer of diameter ~89 angstroms prior to encapsulation. Steric constraints were predicted to limit loading to one hexamer per nanocompartment[12], and measurements by native mass spectroscopy, have found precisely six DyPs in the T=1 nanocompartment, yielding a 1:10 ratio of cargo to protomer[35] (Figure 1-3C).

As this example demonstrates, loading is highly influenced by the oligomeric state of the cargo (Figure 1-4). For example, it is known that DyPs oligomerize into hexamers and FLPs oligomerize into decamers[12,28,36]. While the size of the DyP hexamer limits it to one hexamer per encapsulin, FLP decamers are small enough to bind to a pentamer without blocking any other CLP binding sites[28]. This could allow for the packaging of up to 120 FLP molecules per nanocompartment, or a 2:1 ratio of cargo to protomer[28] (Figure 1-3C). Cryo-electron microscopy on the *M. xanthus* encapsulin hints that the three encapsulated proteins, EncB, C, and D, may form a larger assembly[26].

Ultimately, although all characterized cargo proteins are oligomers, the significance of oligomerization is unknown. One theory is that oligomerization positions the active site of FLP directly adjacent to a pore in the shell facilitating iron oxidation[28]. Work with heterologous cargo has shown that high selectivity is achieved even when packaging just a monomer[19,22]. Given the diversity of cargo proteins and their different oligomeric states it is likely that there are multiples ways of assembling cargo within nanocompartments and with recent advances in electron microscopy, we expect that a variety of cargo assembly modalities will soon be described.

1.5 Stability of encapsulins and a link to function

Although the biological purpose of encapsulation largely remains unknown, one function may be to enhance the stability of cargo proteins. It should be noted that many homologs of cargo proteins are not encapsulated, suggesting encapsulation is not required for cargo enzyme function[12,30]. However, circumstantial observations support the hypothesis that encapsulation could increase the stability and/or lifetime of cargo proteins in their functional environment, such as by conferring resistance to proteases and other insults. Indeed, encapsulins were discovered due to their high stability, which is a general feature of nanocompartments that enabled their persistence in culture supernatant[9,37]. By characterizing the robust nature of prokaryotic nanocompartment shells, it may be possible to explain how the stability is imparted to the cargo protein and why these enzymes are encapsulated.

Prokaryotic nanocompartments are robust structures with resistance to a number of physical and biological insults. *T. maritima* encapsulin, for example, displays high tolerance to temperature and denaturant[12,22]. While nanocompartments from mesophilic organisms are not exceptionally thermostable[21], encapsulins are generally pH-stable and resistant to mechanical compression[12,22,25,35]. Like viral phage capsids, prokaryotic nanocompartments display minimal proteolytic degradation following treatment with nonspecific proteases[21,22]. Resistance to proteases also extends to the cargo proteins: firefly luciferase packaged into encapsulin from *Rhodococcus erythropolis* N771 showed no degradation following treatment with trypsin, while non-encapsulated luciferase was completely degraded[21]. Purified encapsulins also display long-term stability, with minimal degradation observed for *T. maritima* encapsulins following long-term storage (Savage Lab unpublished data).

The robust nature of nanocompartments raises the question of how they are disassembled and degraded *in vivo*. Disassembly of nanocompartments *in vitro* required lowering of the pH to strongly acidic conditions[25] or adding 7 M guanidinium chloride[22], suggesting cells may have special mechanisms for degrading encapsulins. One hypothetical model posits currently unknown accessory factors induce the disassembly of the shell, after which the disassembled shell protomers are targeted for proteolytic degradation[26]. In *M. xanthus*, encapsulin protein levels are substantially reduced 48 hours after sporulation and it is completely absent 96 hours after sporulation, suggesting degradation of the nanocompartment. Another theory is that kinetics of nanocompartment turnover is slow compared to the rate of cell growth, so the cellular load of encapsulin is reduced by dilution from cell growth and division. More work is needed to determine the validity of each model; it may be that strategies for turnover of these robust nanocompartments is species-specific.

Stability may also be related to the hypothesized function of encapsulins in the extracellular environment[9,37]. Starting with their discovery in 1994, encapsulins have often been found in or purified from culture supernatant[9,25,37]. This led to the hypothesis that nanocompartments are secreted, which was reinforced by the observation that encapsulins were membrane-localized[37]. However, no encapsulins have a known secretion signal[30], nor does there exist a known mechanism for secreting an intact 20-30 nm protein complex[38]. Given the chemical stability and protease resistance of nanocompartments, it is more probable that encapsulins accumulate in the culture

supernatant following cell lysis[39]. Relatedly, we hypothesize that the historical observations of nanocompartments associating with the membrane is a function of the size and density of encapsulin. Centrifugation methods used for membrane fractionation also sediment assembled nanocompartments, thus spuriously enriching these soluble complexes with membrane proteins. Notably, immunoelectron microscopy against the encapsulin protein from *Streptomyces griseus* showed that it is located in the cytoplasm[14].

Thus, contrary to the secretion hypothesis, these data suggest that encapsulins are normally found in the cytosol but persist and accumulate following cell lysis, leading to their detection in the culture supernatant. Whether the accumulation of nanocompartments in the extracellular environment is biologically relevant remains an open question. For example, as described below, it has been hypothesized (but not proven) that the presence of extracellular, encapsulated DyP is important for the degradation of lignin by *R. jostii* RHA1[25]. Despite the capacity of nanocompartments to impart stability to their cargo *in vitro*, further work will be required to determine the significance of this observation, both *in vivo* and *ex vivo*.

1.6 The physiological roles of encapsulins

The biochemical and structural similarities between encapsulins and proteinaceous bacterial microcompartments may provide clues to understanding the physiological roles of nanocompartments. Studies have revealed two major functions of microcompartments[2,40]: i.) to increase the local concentration of substrates near enzymes, such as CO_2 at the site of its fixation inside the carboxysome[41], and ii.) to sequester toxic pathway intermediates such as aldehydes (and possibly enzyme-generated radicals) from the cytosol in catabolic microcompartments[42,43]. However, the shell and cargo proteins of microcompartments are significantly more elaborate than those of nanocompartments. Whether functional parallels exist between these two systems thus remains an open question. We highlight here the small, but growing, number of examples in which a physiological function for nanocompartments has been determined.

The most informative results so far come from studies of the FLP-encapsulin system, which suggest that nanocompartments can function as iron storage containers and mitigate oxidative stress. While studying *M. xanthus* cells under nutrient (amino acid) starvation conditions, which in this organism induces a switch to fruiting body formation followed by spore development, McHugh and colleagues found that the encapsulin shell protomer (EncA) and its three FLP cargo proteins (EncB, EncC, and EncD) were significantly upregulated[26]. The authors hypothesized that the encapsulin could be acting as a secondary and larger capacity ferritin-like system to either store iron during nutrient stress or sequester free iron during oxidative stress (Figure 1-5).

When ferrous iron (Fe^{2+}) is exposed to reactive oxygen species (ROS), the Fenton reaction – in which Fe^{2+} is oxidized to Fe^{3+}, yielding a harmful hydroxyl radical by-product – takes place[44,45]. Ferritins protect the cell from the toxic Fenton reaction, raising the question of whether the FLP-encapsulin can protect the cell against oxidative stress in a similar manner. To test this, the authors measured the ability of a *M. xanthus* mutant lacking the encapsulin shell to survive oxidative stress with hydrogen peroxide, and found that the mutant was significantly more sensitive to oxidative stress than the wild type[26]. Furthermore, this Δ*encA* mutant is unable to develop fruiting bodies, suggesting that the presence of the nanocompartment is essential to the transition from its vegetative stage to its spore-forming stage when subjected to starvation[46].

A connection to oxidative stress is also seen in the encapsulin found in *Mycobacterium tuberculosis.* Pathogenesis of *M. tuberculosis* is contingent upon an ability to evade the host immune response and the bactericidal ROS generated by the host[47,48]. The encapsulin from *M. tuberculosis* was found to sequester three different cargo proteins: bacterioferritin (BfrB), dihydroneopterin aldolase (FolB), and a dye decolorizing peroxidase (DyP)[27]. Each of these three cargo proteins have independently been shown to have antioxidant activity[49–52]. For example, BfrB's role in protecting *M. tuberculosis* against oxidative stress was demonstrated by an increased sensitivity to cumene hydroperoxide in a Δ*brfB M. tuberculosis* mutant[50]. The DyP cargo acts as a heme-dependent peroxidase and consumes ROS (e.g. hydrogen peroxide) suggesting that the nanocompartment can ameliorate oxidative stress. The exact reason as to why these cargo proteins are encapsulated, however, remains mysterious, and future work

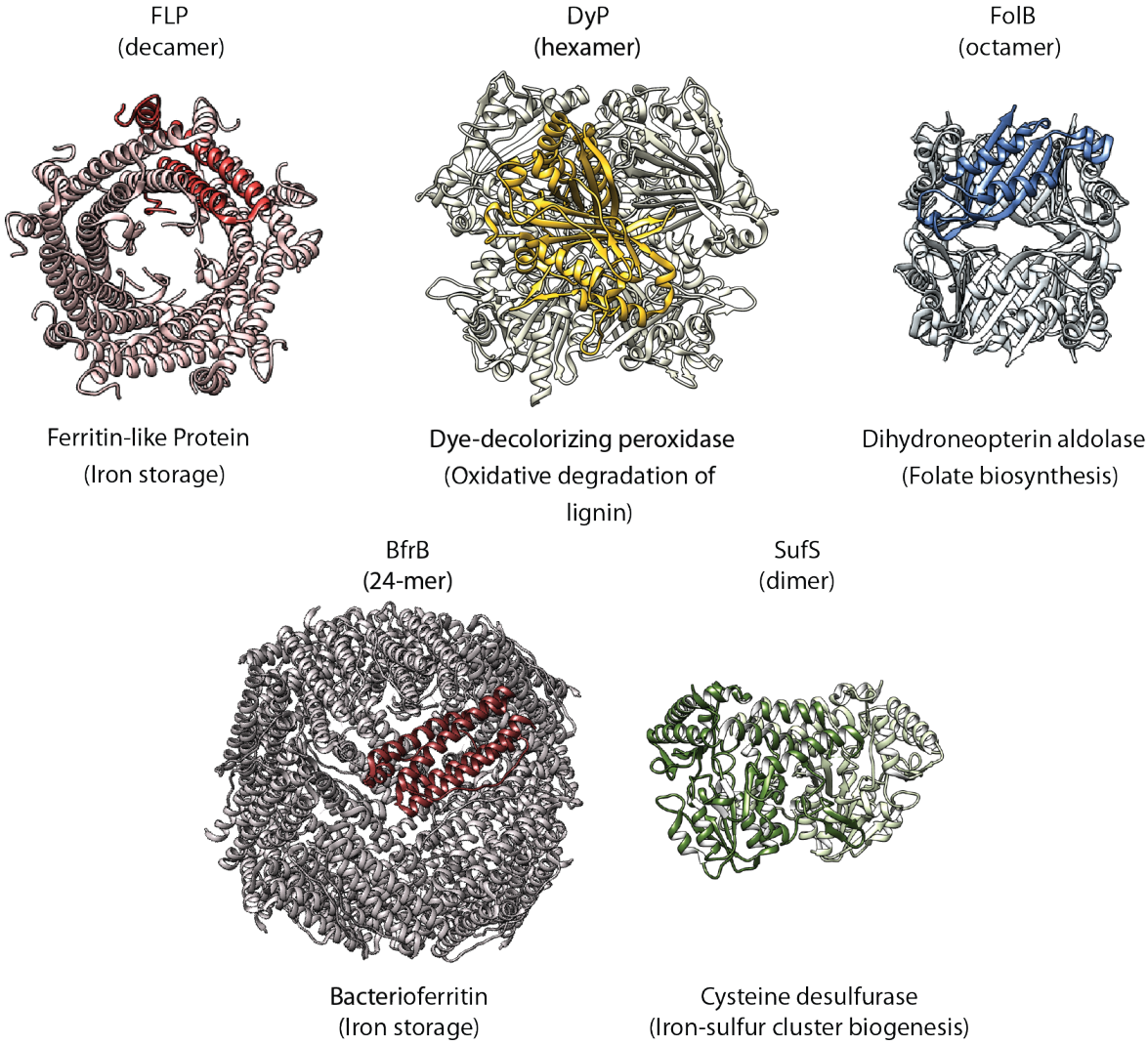

Figure 1-4: High-resolution structures of example oligomeric encapsulin cargo proteins: Ferritin-like protein [PDB: 5DA5], dye-decolorizing peroxidase [PDB: 3QNR], dihydroneopterin aldolase [PDB: 1NBU], bacterioferritin [PDB: 3UNO], and cysteine desulfurase [PDB: 1T3I].

will be needed to more clearly delineate the role of compartmentalization in ROS tolerance.

In addition to their potential role in ameliorating oxidative stress, DyP-containing encapsulins have also been implicated in catabolic metabolism[25]. The bacterium *R. jostii* RHA1 is capable of robust lignin degradation, and deletion of the DyP cargo gene (*dypB*) yields the mutant strain incapable of lignin degradation[53]. *In vitro* peroxidase activity assays showed that the encapsulin-DypB complex had an eight-fold increase in activity for degrading nitrated lignin substrate compared to the naked DypB enzyme[25]. An increase in enzymatic activity upon encapsulation suggests that the nanocompartment may be acting by either stabilizing the cargo protein or by increasing the local substrate concentration for the enzyme, thereby enhancing the enzymatic reaction in a manner similar to bacterial microcompartments[54]. These results raise many provocative mechanistic questions. Lignin is a highly heterogeneous polymer, and it remains unclear whether the shells possess specificity for this substrate. Additionally, the lignin degradation activity of *R. jostii* RHA1 is found in the extracellular fraction, yet it

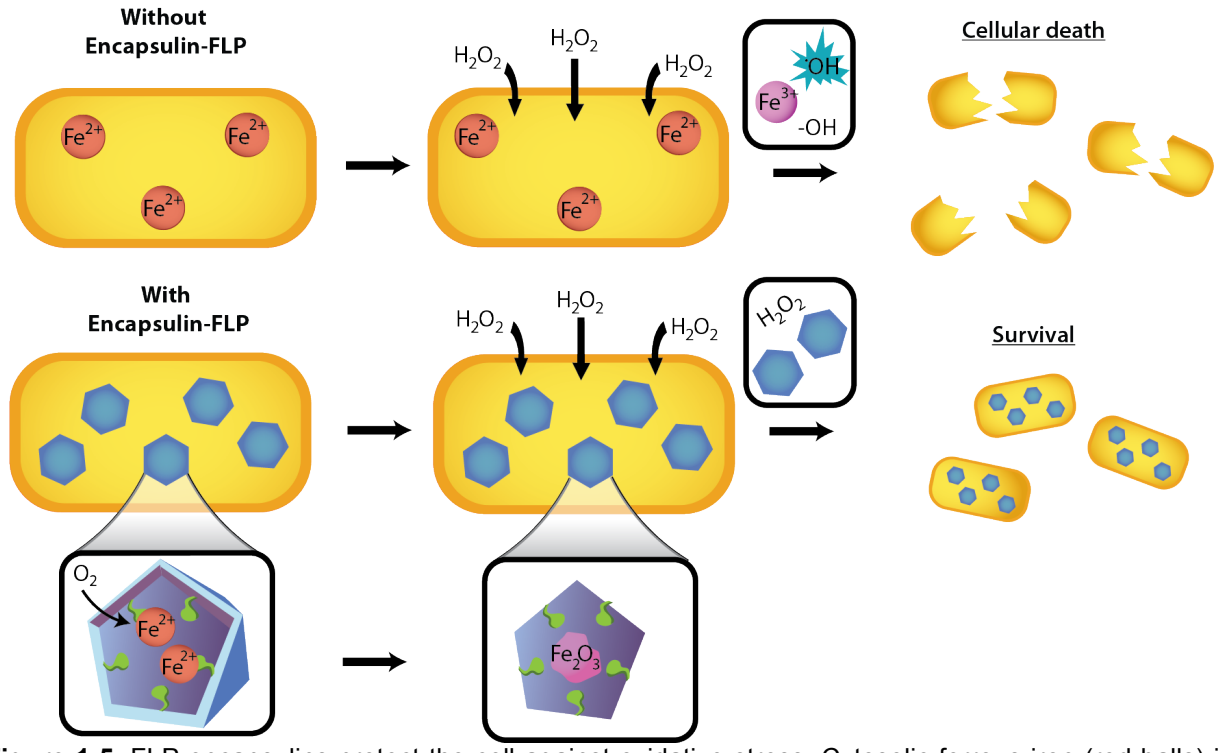

Figure 1-5: FLP-encapsulins protect the cell against oxidative stress. Cytosolic ferrous iron (red balls) is harmful to the cell under oxidative stress. FLP-encapsulins (blue hexagons) will sequester the ferrous iron and store it as mineralized iron oxide (magenta), thereby protecting the cell from reactive oxygen species.

is unknown how the DypB protein or encapsulin could be targeted for secretion. As described above, an untested yet intriguing hypothesis is that the stability of encapsulins facilitates their survival following parental cell lysis. In this case, the encapsulin-DypB complex could remain stable and active in the extracellular environment, conferring a fitness advantage to the bacterial community.

A broader understanding of the physiological role of bacterial nanocompartments could be gained by learning more about the diverse cargo types hosted by different encapsulin systems and which species possess them. Two recent studies have explored the prevalence of encapsulins across prokaryotes as well as the diversity of cargo proteins sequestered by various encapsulin systems[29,30]. Giessen and Silver performed an extensive search for homologs of the handful of characterized encapsulins and found over 900 occurrences of putative encapsulin systems. These newly identified homologs were widespread throughout prokaryotes (15 bacterial and 2 archaeal phyla) with a majority of the hits being found in Actinobacteria and Proteobacteria. In addition to the previously-characterized cargo proteins such as FLP, DyP, and FolB, new cargo proteins were identified in this search, including hemerythrins, a four-helix bundle protein of unknown function termed IMEF, and a nitrite reductase-hydroxylamine oxidoreductase (NIR-HAO) fusion cargo. Notably, the authors demonstrated that members of each of these systems do in fact form nanocompartments, and that the encapsulin shells co-purify with their predicted cargo protein(s).

In addition to validating that these predicted encapsulin systems were true cargo-bearing nanocompartments, Giessen and Silver tested possible functional roles for each

of the encapsulins. Given prior experimental evidence of hemerythrin's role in oxidative and nitric oxide stress responses, they tested the ability of *E. coli* heterologously expressing the hemerythrin-encapsulin system from *Streptomyces* sp. AA4 to survive treatment with hydrogen peroxide and nitric oxide[55–57]. Compared to *E. coli* strains expressing the hemerythrin cargo alone or the encapsulin shell alone, the strain expressing both demonstrated increased protection from oxidative and nitrosative stress. This suggests that both components of the bacterial organelle are important for cell survival during oxidative and nitrosative stress. Similar results were also observed for the encapsulin-IMEF system (from *Bacillaceae bacterium* MTCC 10057) in protecting *E. coli* from the dangerous byproducts of the Fenton reaction. Furthermore, a novel encapsulin system unique to anaerobic ammonium oxidizing (anammox) bacteria was experimentally validated. This encapsulin system was shown to possess a nitrite reductase-hydroxylamine oxidoreductase fusion cargo (NIR-HAO) and may explain how these bacteria protect themselves from toxic anammox metabolic products such as nitric oxide[58]. The exact role of the NIR-HAO type encapsulins in anammox metabolism will likely require the development of genetic tools in an encapsulin-bearing anammox bacterium such as *Candidatus Kuenenia stuttgartiensis* and subsequent genetic manipulation of the encapsulin genes.

Finally, additional evidence suggests that the characterized encapsulins discussed here represent just a small fraction of extant nanocompartments found in microbial genomes. This evidence comes from Radford and colleagues, who used a comprehensive bioinformatic approach in order to identify novel nanocompartments with limited homology to known encapsulins[30]. Since it is likely that other types of nanocompartments would display homology to bacteriophage capsids, prokaryotic genomes were searched against a database of bacteriophage capsid proteins. Putative shell genes that neighbored other phage-like genes were discarded as likely prophages, while those that did not were identified as predicted encapsulins. The newly assigned nanocompartment systems were then grouped into families based on sequence homology of the shell protein and shared predicted cargo types. This search identified seven putative novel encapsulin shell families with no apparent sequence homology to previously-described encapsulins (Table 1). All previously-published nanocompartments fall into the family 1 encapsulins, which is the second largest family. The largest family of encapsulins, family 2, is a novel shell gene family that has not yet been appreciated as a prokaryotic nanocompartment. Many of the family 2 encapsulin systems neighbor a cysteine desulfurase gene, which is the predicted cargo for this nanocompartment system. Family 2 encapsulins that do not neighbor cysteine desulfurase genes contain a putative Crp/Fnr (cyclic AMP receptor protein/ fumarate and nitrate reductase regulator protein) transcriptional regulatory domain that is an internal fusion with the shell gene. The other novel shell families identified were significantly less prevalent and are thought to encapsulate a variety of enzymes, including a cytochrome P450-like monooxygenase and an alpha-ketoglutarate decarboxylase.

While these predicted novel encapsulin families were not experimentally characterized, the bioinformatic study brings clarity to older results from groups studying these genes by implicating them as prokaryotic nanocompartments. Specifically, the EshA protein in *Streptomyces griseus,* identified by Radford as a family 2 encapsulin with a Crp/Fnr fusion domain, was shown to form large multimers more than a decade

ago[15,16]. Transmission electron microscopy of the purified EshA complex showed that these homomultimers form 27 nm particles that look similar in morphology to family 1 encapsulins[14]. Interestingly, genetic disruption of the *eshA* gene in *Streptomyces coelicolor* A3(2) abolishes the ability of this organism to produce the antibiotic actinorhodin[16], while *eshA* deletion in *S. griseus* interferes with both sporulation and streptomycin production[14,15]. The specific functions of EshA in these contexts, and why it possesses the Crp/Fnr fusion, remain mysterious. Further biochemical investigation of the protein complex, and elucidation of the metabolites and proteins with which it interacts, will hopefully provide mechanistic answers to these compelling observations.

These encapsulin systems are widespread throughout prokaryotic phyla and are not restricted within distinct phylogenetic lineages. This led Radford to conclude that encapsulins are inherited via horizontal gene transfer as suggested by others[29,30,59]. A unifying theme among the nanocompartment families that have been experimentally studied so far is their importance in response to stresses including nutrient starvation or oxidative stress. Future research will likely uncover more shell families and cargo types with novel and diverse functions. Lastly, additional biochemical and genetic studies of known encapsulin systems will help us understand their structural architecture and the unique contribution of compartmentalization to facilitating and enhancing microbial metabolism.

In this dissertation, I asked how heterologous cargo can be loaded into *T. maritima* encapsulins and how differences in the length of the CLP affected the degree of cargo loading. Answering this questions required biophysical characterization of empty nanocompartments and led to the striking observation that *T. maritima* encapsulins are flavoproteins. Next I sought to determine if there was any functional relevance to encapsulins being flavoproteins especially in regards to the enzymatic activity of the packaged FLPs. Preliminary experiments did not reveal a role for the bound flavins; however, they also did not rule out a role for the flavins. Finally, I explored how encapsulins from *M. tuberculosis* help the host survive in the oxidizing environment of the phagolysosome. Overall, this dissertation expands the field of nanocompartment research with one of the first biophysical characterizations of an encapsulin and novel physiological roles for *T. maritima* and *M. tuberculosis* encapsulins.

1.7 References

(1) Diekmann, Y., and Pereira-Leal, J. B. (2013) Evolution of intracellular compartmentalization. *Biochem. J 449*, 319–331.
(2) Kerfeld, C. A., Heinhorst, S., and Cannon, G. C. (2010) Bacterial microcompartments. *Annu. Rev. Microbiol. 64*, 391–408.
(3) Chowdhury, C., Sinha, S., Chun, S., Yeates, T. O., and Bobik, T. A. (2014) Diverse bacterial microcompartment organelles. *Microbiol. Mol. Biol. Rev. 78*, 438–468.
(4) Yeates, T. O., Thompson, M. C., and Bobik, T. A. (2011) The protein shells of bacterial microcompartment organelles. *Curr. Opin. Struct. Biol. 21*, 223–231.
(5) Cornejo, E., Abreu, N., and Komeili, A. (2014) Compartmentalization and organelle formation in bacteria. *Curr. Opin. Cell Biol. 26*, 132–138.
(6) Shively, J. M. (1974) Inclusion bodies of prokaryotes. *Annu. Rev. Microbiol. 28*, 167–187.
(7) Niklowitz, W., and Drews, G. (1956) BEITRAGE ZUR CYTOLOGIE DER BLAUALGEN. 1. UNTERSUCHUNGEN ZUR SUBSTRUKTUR VON PHORMIDIUM UNCINATUM GOM. *Arch. Mikrobiol. 24*, 134–146.
(8) Axen, S. D., Erbilgin, O., and Kerfeld, C. A. (2014) A taxonomy of bacterial microcompartment loci constructed by a novel scoring method. *PLoS Comput. Biol. 10*, e1003898.
(9) Valdés-Stauber, N., and Scherer, S. (1994) Isolation and characterization of Linocin M18, a bacteriocin produced by Brevibacterium linens. *Appl. Environ. Microbiol. 60*, 3809–3814.
(10) Winter, N., Triccas, J. A., Rivoire, B., Pessolani, M. C., Eiglmeier, K., Lim, E. M., Hunter, S. W., Brennan, P. J., and Britton, W. J. (1995) Characterization of the gene encoding the immunodominant 35 kDa protein of Mycobacterium leprae. *Mol. Microbiol. 16*, 865–876.
(11) Hicks, P. M., Rinker, K. D., Baker, J. R., and Kelly, R. M. (1998) Homomultimeric protease in the hyperthermophilic bacterium Thermotoga maritima has structural and amino acid sequence homology to bacteriocins in mesophilic bacteria. *FEBS Lett. 440*, 393–398.
(12) Sutter, M., Boehringer, D., Gutmann, S., Günther, S., Prangishvili, D., Loessner, M. J., Stetter, K. O., Weber-Ban, E., and Ban, N. (2008) Structural basis of enzyme encapsulation into a bacterial nanocompartment. *Nat. Struct. Mol. Biol. 15*, 939–947.
(13) Triccas, J. A., Roche, P. W., Winter, N., Feng, C. G., Butlin, C. R., and Britton, W. J. (1996) A 35-kilodalton protein is a major target of the human immune response to Mycobacterium leprae. *Infect. Immun. 64*, 5171–5177.
(14) Saito, N., Matsubara, K., Watanabe, M., Kato, F., and Ochi, K. (2003) Genetic and biochemical characterization of EshA, a protein that forms large multimers and affects developmental processes in Streptomyces griseus. *J. Biol. Chem. 278*, 5902–5911.
(15) Kwak, J., McCue, L. A., Trczianka, K., and Kendrick, K. E. (2001) Identification and characterization of a developmentally regulated protein, EshA, required for sporogenic hyphal branches in Streptomyces griseus. *J. Bacteriol. 183*, 3004–3015.
(16) Kawamoto, S., Watanabe, M., Saito, N., Hesketh, A., Vachalova, K., Matsubara, K., and Ochi, K. (2001) Molecular and Functional Analyses of the Gene (eshA) Encoding

the 52-Kilodalton Protein of Streptomyces coelicolor A3 (2) Required for Antibiotic Production. *J. Bacteriol. 183*, 6009–6016.
(17) Akita, F., Chong, K. T., Tanaka, H., Yamashita, E., Miyazaki, N., Nakaishi, Y., Suzuki, M., Namba, K., Ono, Y., Tsukihara, T., and Nakagawa, A. (2007) The Crystal Structure of a Virus-like Particle from the Hyperthermophilic Archaeon Pyrococcus furiosus Provides Insight into the Evolution of Viruses. *J. Mol. Biol. 368*, 1469–1483.
(18) Rurup, W. F., Snijder, J., Koay, M. S. T., Heck, A. J. R., and Cornelissen, J. J. L. M. (2014) Self-sorting of foreign proteins in a bacterial nanocompartment. *J. Am. Chem. Soc. 136*, 3828–3832.
(19) Snijder, J., van de Waterbeemd, M., Damoc, E., Denisov, E., Grinfeld, D., Bennett, A., Agbandje-McKenna, M., Makarov, A., and Heck, A. J. R. (2014) Defining the stoichiometry and cargo load of viral and bacterial nanoparticles by Orbitrap mass spectrometry. *J. Am. Chem. Soc. 136*, 7295–7299.
(20) Moon, H., Lee, J., Min, J., and Kang, S. (2014) Developing genetically engineered encapsulin protein cage nanoparticles as a targeted delivery nanoplatform. *Biomacromolecules 15*, 3794–3801.
(21) Tamura, A., Fukutani, Y., Takami, T., Fujii, M., Nakaguchi, Y., Murakami, Y., Noguchi, K., Yohda, M., and Odaka, M. (2015) Packaging guest proteins into the encapsulin nanocompartment from Rhodococcus erythropolis N771. *Biotechnol. Bioeng. 112*, 13–20.
(22) Cassidy-Amstutz, C., Oltrogge, L., Going, C. C., Lee, A., Teng, P., Quintanilla, D., East-Seletsky, A., Williams, E. R., and Savage, D. F. (2016) Identification of a Minimal Peptide Tag for in Vivo and in Vitro Loading of Encapsulin. *Biochemistry 55*, 3461–3468.
(23) Giessen, T. W., and Silver, P. A. (2016) Converting a Natural Protein Compartment into a Nanofactory for the Size-Constrained Synthesis of Antimicrobial Silver Nanoparticles. *ACS Synth. Biol. 5*, 1497–1504.
(24) Choi, B., Moon, H., Hong, S. J., Shin, C., Do, Y., Ryu, S., and Kang, S. (2016) Effective Delivery of Antigen–Encapsulin Nanoparticle Fusions to Dendritic Cells Leads to Antigen-Specific Cytotoxic T Cell Activation and Tumor Rejection. *ACS Nano 10*, 7339–7350.
(25) Rahmanpour, R., and Bugg, T. D. H. (2013) Assembly in vitro of Rhodococcus jostii RHA1 encapsulin and peroxidase DypB to form a nanocompartment. *FEBS J. 280*, 2097–2104.
(26) McHugh, C. A., Fontana, J., Nemecek, D., Cheng, N., Aksyuk, A. A., Heymann, J. B., Winkler, D. C., Lam, A. S., Wall, J. S., Steven, A. C., and Hoiczyk, E. (2014) A virus capsid-like nanocompartment that stores iron and protects bacteria from oxidative stress. *EMBO J. 33*, 1896–1911.
(27) 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., and Goulding, C. W. (2014) Characterization of a Mycobacterium tuberculosis nanocompartment and its potential cargo proteins. *J. Biol. Chem. 289*, 18279–18289.
(28) He, D., Hughes, S., Vanden-Hehir, S., Georgiev, A., Altenbach, K., Tarrant, E., Mackay, C. L., Waldron, K. J., Clarke, D. J., and Marles-Wright, J. (2016) Structural characterization of encapsulated ferritin provides insight into iron storage in bacterial nanocompartments. *Elife 5*. e18972.

(29) Giessen, T. W., and Silver, P. A. (2017) Widespread distribution of encapsulin nanocompartments reveals functional diversity. *Nat Microbiol 2*, 17029.
(30) Radford, D. (2015) Understanding the Encapsulins: Prediction and Characterization of Phage Capsid-like Nanocompartments in Prokaryotes. UNIVERSITY OF TORONTO (CANADA).
(31) Giessen, T. W. (2016) Encapsulins: microbial nanocompartments with applications in biomedicine, nanobiotechnology and materials science. *Curr. Opin. Chem. Biol. 34*, 1–10.
(32) Wikoff, W. R., Liljas, L., Duda, R. L., Tsuruta, H., Hendrix, R. W., and Johnson, J. E. (2000) Topologically linked protein rings in the bacteriophage HK97 capsid. *Science 289*, 2129–2133.
(33) Caspar, D. L., and Klug, A. (1962) Physical principles in the construction of regular viruses. *Cold Spring Harb. Symp. Quant. Biol. 27*, 1–24.
(34) Ahmad, M., Taylor, C. R., Pink, D., Burton, K., Eastwood, D., Bending, G. D., and Bugg, T. D. H. (2010) Development of novel assays for lignin degradation: comparative analysis of bacterial and fungal lignin degraders. *Mol. Biosyst. 6*, 815–821.
(35) 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., and Heck, A. J. R. (2016) Assembly and Mechanical Properties of the Cargo-Free and Cargo-Loaded Bacterial Nanocompartment Encapsulin. *Biomacromolecules 17*, 2522–2529.
(36) Zubieta, C., Krishna, S. S., Kapoor, M., Kozbial, P., McMullan, D., Axelrod, H. L., Miller, M. D., Abdubek, P., Ambing, E., Astakhova, T., Carlton, D., Chiu, H.-J., Clayton, T., Deller, M. C., Duan, L., Elsliger, M.-A., Feuerhelm, J., Grzechnik, S. K., Hale, J., Hampton, E., Han, G. W., Jaroszewski, L., Jin, K. K., Klock, H. E., Knuth, M. W., Kumar, A., Marciano, D., Morse, A. T., Nigoghossian, E., Okach, L., Oommachen, S., Reyes, R., Rife, C. L., Schimmel, P., van den Bedem, H., Weekes, D., White, A., Xu, Q., Hodgson, K. O., Wooley, J., Deacon, A. M., Godzik, A., Lesley, S. A., and Wilson, I. A. (2007) Crystal structures of two novel dye-decolorizing peroxidases reveal a beta-barrel fold with a conserved heme-binding motif. *Proteins 69*, 223–233.
(37) Rosenkrands, I., Rasmussen, P. B., Carnio, M., Jacobsen, S., Theisen, M., and Andersen, P. (1998) Identification and characterization of a 29-kilodalton protein from Mycobacterium tuberculosis culture filtrate recognized by mouse memory effector cells. *Infect. Immun. 66*, 2728–2735.
(38) Green, E. R., and Mecsas, J. (2016) Bacterial Secretion Systems: An Overview. *Microbiol Spectr 4*.
(39) Wang, P., Robert, L., Pelletier, J., Dang, W. L., Taddei, F., Wright, A., and Jun, S. (2010) Robust growth of Escherichia coli. *Curr. Biol. 20*, 1099–1103.
(40) Yeates, T. O., Crowley, C. S., and Tanaka, S. (2010) Bacterial microcompartment organelles: protein shell structure and evolution. *Annu. Rev. Biophys. 39*, 185–205.
(41) Mangan, N. M., Flamholz, A., Hood, R. D., Milo, R., and Savage, D. F. (2016) pH determines the energetic efficiency of the cyanobacterial CO2 concentrating mechanism. *Proc. Natl. Acad. Sci. U. S. A. 113*, E5354–62.
(42) Penrod, J. T., and Roth, J. R. (2006) Conserving a volatile metabolite: a role for carboxysome-like organelles in Salmonella enterica. *J. Bacteriol. 188*, 2865–2874.

(43) Sampson, E. M., and Bobik, T. A. (2008) Microcompartments for B12-dependent 1,2-propanediol degradation provide protection from DNA and cellular damage by a reactive metabolic intermediate. *J. Bacteriol. 190*, 2966–2971.
(44) Imlay, J. A., Chin, S. M., and Linn, S. (1988) Toxic DNA damage by hydrogen peroxide through the Fenton reaction in vivo and in vitro. *Science 240*, 640–642.
(45) Andrews, S. C. (1998) Iron Storage in Bacteria, in *Advances in Microbial Physiology* (R.K. Poole, Ed.), pp 281–351. Academic Press.
(46) Kim, D., Chung, J., Hyun, H., Lee, C., Lee, K., and Cho, K. (2009) Operon required for fruiting body development in Myxococcus xanthus. *J. Microbiol. Biotechnol. 19*, 1288–1294.
(47) Slauch, J. M. (2011) How does the oxidative burst of macrophages kill bacteria? Still an open question. *Mol. Microbiol. 80*, 580–583.
(48) Manca, C., Paul, S., Barry, C. E., 3rd, Freedman, V. H., and Kaplan, G. (1999) Mycobacterium tuberculosis catalase and peroxidase activities and resistance to oxidative killing in human monocytes in vitro. *Infect. Immun. 67*, 74–79.
(49) Pandey, R., and Rodriguez, G. M. (2012) A ferritin mutant of Mycobacterium tuberculosis is highly susceptible to killing by antibiotics and is unable to establish a chronic infection in mice. *Infect. Immun. 80*, 3650–3659.
(50) Reddy, P. V., Puri, R. V., Khera, A., and Tyagi, A. K. (2012) Iron storage proteins are essential for the survival and pathogenesis of Mycobacterium tuberculosis in THP-1 macrophages and the guinea pig model of infection. *J. Bacteriol. 194*, 567–575.
(51) Goulding, C. W., Apostol, M. I., Sawaya, M. R., Phillips, M., Parseghian, A., and Eisenberg, D. (2005) Regulation by oligomerization in a mycobacterial folate biosynthetic enzyme. *J. Mol. Biol. 349*, 61–72.
(52) Sugano, Y., Muramatsu, R., Ichiyanagi, A., Sato, T., and Shoda, M. (2007) DyP, a unique dye-decolorizing peroxidase, represents a novel heme peroxidase family: ASP171 replaces the distal histidine of classical peroxidases. *J. Biol. Chem. 282*, 36652–36658.
(53) Ahmad, M., Roberts, J. N., Hardiman, E. M., Singh, R., Eltis, L. D., and Bugg, T. D. H. (2011) Identification of DypB from Rhodococcus jostii RHA1 as a lignin peroxidase. *Biochemistry 50*, 5096–5107.
(54) Bobik, T. A., Lehman, B. P., and Yeates, T. O. (2015) Bacterial microcompartments: widespread prokaryotic organelles for isolation and optimization of metabolic pathways. *Mol. Microbiol. 98*, 193–207.
(55) Kendall, J. J., Barrero-Tobon, A. M., Hendrixson, D. R., and Kelly, D. J. (2014) Hemerythrins in the microaerophilic bacterium Campylobacter jejuni help protect key iron–sulphur cluster enzymes from oxidative damage. *Environ. Microbiol. 16*, 1105–1121.
(56) Li, X., Tao, J., Hu, X., Chan, J., Xiao, J., and Mi, K. (2014) A bacterial hemerythrin-like protein MsmHr inhibits the SigF-dependent hydrogen peroxide response in mycobacteria. *Front. Microbiol. 5*, 800.
(57) Chow, E. D., Liu, O. W., O'Brien, S., and Madhani, H. D. (2007) Exploration of whole-genome responses of the human AIDS-associated yeast pathogen Cryptococcus neoformans var grubii: nitric oxide stress and body temperature. *Curr. Genet. 52*, 137–148.
(58) Kartal, B., de Almeida, N. M., Maalcke, W. J., den Camp, H. J. M. O., Jetten, M. S.

M., and Keltjens, J. T. (2013) How to make a living from anaerobic ammonium oxidation. *FEMS Microbiol. Rev. 37*, 428–461.
(59) Heinemann, J., Maaty, W. S., Gauss, G. H., Akkaladevi, N., Brumfield, S. K., Rayaprolu, V., Young, M. J., Lawrence, C. M., and Bothner, B. (2011) Fossil record of an archaeal HK97-like provirus. *Virology 417*, 362–368.

Chapter 2 Identification of a minimal peptide tag for *in vivo* and *in vitro* loading of encapsulin

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

2-1 Abstract

The encapsulation of enzymes and other proteins within a proteinaceous shell has been observed in many bacteria and archaea, but the function and utility of many such compartments is enigmatic. Efforts to study these functions have been complicated by the size and complexity of traditional protein compartments. One potential system for investigating the affect of compartmentalization is encapsulin, a large and newly discovered class of protein shells, that are typically composed of 2 proteins: a protomer that assembles into the icosahedral shell and a cargo protein packaged inside. Encapsulins are one of the simplest known protein shell systems and readily self-assemble *in vivo*. Systematic characterization of the effects of compartmentalization requires the ability to load a wide range of cargo proteins. Here, we demonstrate that foreign cargo can be loaded into the encapsulin from *Thermotoga maritima* both *in vivo* and *in vitro* by fusion of the cargo protein with a short C-terminal peptide present in the native cargo. To facilitate biochemical characterization, we also develop a simple and rapid purification protocol and demonstrate the thermal and pH stability of the shell. Efforts to study the biophysical effects of protein encapsulation have been problematic in complex compartments, but the simplicity of assembling and loading encapsulin make it an ideal system for future experiments exploring the effects of encapsulation on proteins.

2-2 Introduction

A variety of protein-based organelles have been identified among prokaryotes. One family of such organelles – the bacterial microcompartments (BMCs) – are 80-400 nm in diameter and formed of thousands of shell protein subunits assembled in an icosahedral geometry. [1-3] These compartments have been implicated in a variety of metabolic processes ranging from carbon fixation to catabolism of ethanolamine and propanediol.[1,2,4] Recently, a novel class of bacterial protein compartments called encapsulins has been identified and characterized.[5-9] These so-called nanocompartments are generally much smaller and simpler than BMCs, with typical diameters ranging from 25-32 nm.[5,7] In contrast to BMCs, whose shells are generally comprised of multiple homologous proteins, encapsulins form from a single shell protein that is homologous to the HK97 bacteriophage capsid protein.[5,7,10]

Encapsulins are typically expressed from the same operon as their cargo protein, which can include ferritin-like proteins (FLP) as in *Thermatoga maritima* or dye decolorizing peroxidases (DypB) as in *Brevibacterium linens* and *Rhodococcus jostii* RHA1.[5,11] FLP and DypB sequences share a C-terminal extension found only in sequences from species with encapsulins[5], and it has been demonstrated that tagging of proteins with the C-terminal extension of DypB leads to incorporation with the DypB-specific encapsulin.[5,12,13] While a few cargo proteins have been implicated in protecting from oxidative stress[7,8], the physiological role of most encapsulated FLP or DypB is unclear. Additionally, recent bioinformatic work has shown that nanocompartments are more widely dispersed among the Bacteria and Archaea than previously thought.[14]

Little is known about how protein encapsulation affects the cargo proteins in microcompartments due to the technical challenges of working with complex, multi-protein BMCs.[15,16] Artificial compartments have been made from viral capsids to create nanoreactors[17], which have been used to study the effects of encapsulation on single enzymes or entire pathways.[18,19] Unlike viral capsids which naturally package nucleic acid, encapsulins are attractive model for the study of natural protein encapsulation because they resemble BMCs but are much simpler to express heterologously, purify, and characterize biochemically.[9] Here, we focus our efforts on the *T. maritima* encapsulin, which is approximately 25 nm in diameter and natively encapsulates an FLP.[5] We identify a minimal C-terminal targeting peptide that enables efficient encapsulin loading *in vivo*. We further show that *T. maritima* encapsulins can be loaded with protein *in vitro* by refolding the complex in the presence of a cargo protein. This work lays a biochemical foundation for future studies of the biophysical effects of encapsulation on cargo proteins.

2-3 Results

2.3.1 Cargo Peptide Targets Heterologous Cargo to T. maritima Encapsulin

In a crystal structure of the *T. maritima* encapsulin, Sutter et al. serendipitously observed that the C-terminus of FLP binds to the encapsulin lumen.[5] Each protomer has a binding site for the cargo loading peptide (CLP), which implies a maximal loading of 60 cargo proteins per nanocompartment.[5] We sought to confirm that the C-terminus of FLP acts as a CLP by co-expressing encapsulin with the last 30 amino acids of FLP fused to superfolder green fluorescent protein (sfGFP) (Figure 2-1A). Nanocompartments were purified by 10% to 60% sucrose gradient centrifugation followed by size exclusion chromatography on a Superose 6 column. Purified nanocompartments appeared green by eye, and sodium dodecyl sulfate (SDS)-polyacrylamide gel electrophoresis (PAGE) confirmed the presence of both encapsulin and tagged sfGFP (Figure 2-2A). Transmission electron microscopy confirmed that purified encapsulin had properly assembled into 24 nm compartments (Figure 2-2B,C). Purified encapsulin expressed alone or co-expressed with tagged sfGFP both migrated as the assembled complex by native PAGE; however, only the co-expressed sample displayed GFP fluorescence (Figure 2-1B). This suggests that the CLP derived from FLP is sufficient for cargo loading *in vivo*. We also observed a minor high molecular weight contaminant that co-purified with encapsulated sfGFP. Mass spectrometry identified this band as GroEL. While each protomer contains a binding site for the CLP, steric constraints likely limit the maximum number of tagged sfGFP to ~20 per encapsulin. To investigate the heterogeneity of cargo encapsulation, sfGFP-loaded encapsulins were imaged by total internal reflection fluorescence (TIRF) microscopy. Small green fluorescent punctae were observed, and quantification of individual particle fluorescence revealed a roughly symmetric distribution with minimal outliers (Figure 2-1C). In previous work, encapsulins from *Brevibacterium linens* loaded with a CLP derived from DypB were analyzed by native mass spectrometry and found to have 8 to 14 cargo proteins loaded in a distribution similar to the one we observed for loading of sfGFP in *T. maritima* encapsulin.[24] This suggests co-expression of encapsulin with a cargo protein fused to a CLP derived from FLP results in encapsulins where each compartment has a similar degree of cargo loaded.

The length of the CLP necessary for efficient cargo loading is an open question. We therefore sought to determine the minimal CLP length necessary for signal sequence function. Previous studies used a 37 amino acid cargo peptide to package foreign cargo in encapsulin.[13] However, a sequence alignment of CLP from different cargo proteins showed that only the last ~10 residues have a high degree of homology (Figure 2-3).[5] Previous heterologous loading results using a CLP derived from DypB cargo protein were achieved using a 30 amino acid or longer CLP. [12,13] To determine the length necessary for efficient loading, constructs of sfGFP with a cargo peptide of 0, 5, 15, and 30 amino acids were generated. As a negative control, a construct was generated with a CLP sequence formed by randomly shuffling the 30 amino acid CLP sequence. Analysis of the amount of sfGFP loaded in purified

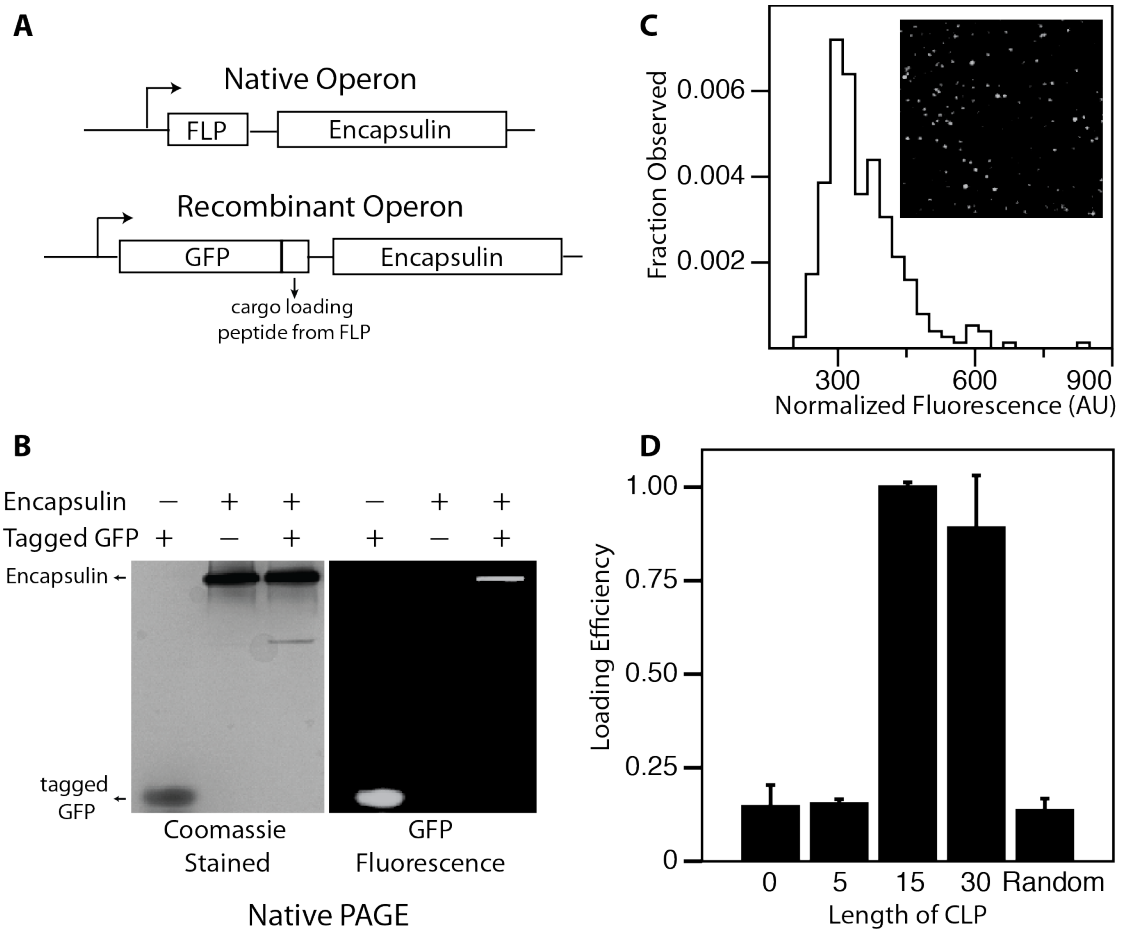

Figure 2-1: Heterologous cargo (sfGFP) can be targeted to *T. maritima* encapsulin. (A) Native *T. maritima* nanocompartment operon and operon used for recombinant co-expression of tagged sfGFP and encapsulin. (B) Native PAGE of purified tagged sfGFP, encapsulin, and co-expressed tagged sfGFP and encapsulin. (left) Coomassie stained; (right) GFP fluorescence. (C) Distribution of GFP particle fluorescence observed from TIRF microscopy of purified sfGFP loaded encapsulin (n=278). (Inset) TIRF microscopy image of purified particles. (D) Amount of normalized sfGFP fluorescence that co-migrated with encapsulin on native PAGE. Error bars represent the calculated standard deviation (n = 3).

nanocompartments from *in vivo* co-expression demonstrated that the randomized sequence and 0 and 5 amino acid peptide constructs had minimal cargo loading. The full 30 amino acid cargo peptide and the 15 amino acid peptide demonstrated robust and equivalent cargo loading (Figure 2-1D). Thus, a minimal fragment of the C-terminus of FLP 15 amino acids or longer can act as a CLP for foreign proteins into encapsulin.

2.3.2 A Simplified Purification Strategy for Empty Encapsulins

We developed a rapid purification strategy for empty encapsulins to facilitate downstream biochemical characterization (Figure 2-4A). Previous purification strategies relied on time consuming differential centrifugation spins to

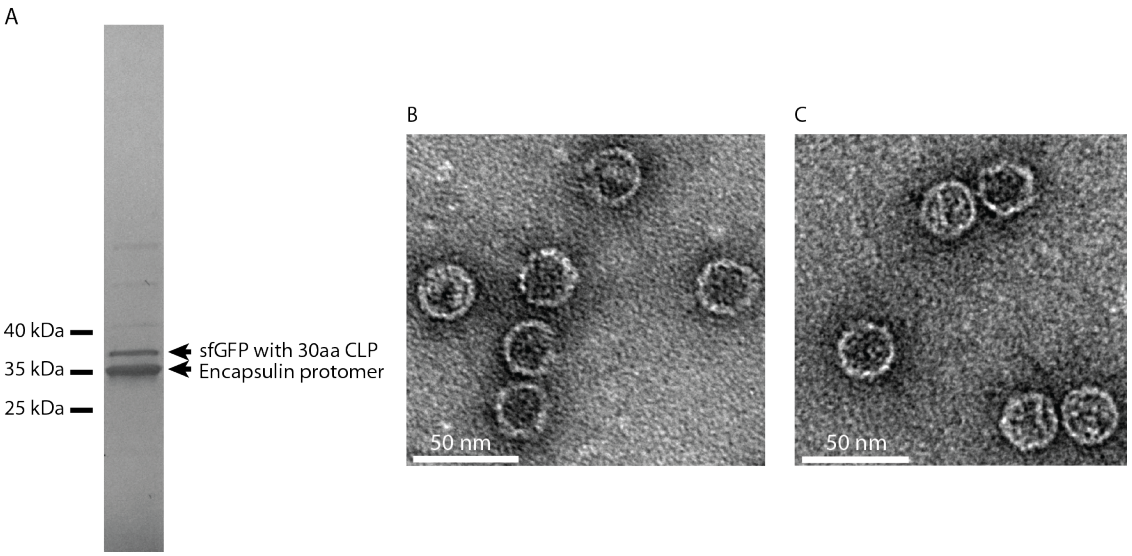

Figure 2-2: Purified *T. maritima* encapsulins (A) SDS-PAGE of *in vivo* purified encapsulin co-expressed with a 30 amino acid CLP tagged sfGFP Purified recombinant encapsulins form a nanocompartment. (B) Encapsulin purified from *E. coli* imaged by TEM. (C) Encapsulin co-expressed with CLP tagged sfGFP purified from *E. coli* imaged by TEM

isolate assembled nanocompartments.[5,6] Initially, lysates were treated at 80˚ C for 1.5 hours to denature the majority of *E. coli* proteins. Following centrifugation to eliminate precipitant, the soluble fraction was further enriched using an ammonium sulfate precipitation (encapsulin in the 50% to 75% fraction) in order to remove substantial nucleic acid contamination. Finally, properly assembled nanocompartments were isolated by size exclusion chromatography. Encapsulin eluted from a Superose 6 size exclusion column slightly overlapping with the void and before a 670 kDa thyroglobulin standard suggesting encapsulin is purified as the assembled 60-mer nanocompartment (Figure 2-4B). Analysis by SDS-PAGE revealed the purified encapsulin protomers were highly enriched and we were able to achieve a yield of >50 mg of encapsulin protomer per liter of media with a purification strategy that takes less than a day to complete (Figure 2-6).

We next tested if cargo loading altered the formation of the nanocompartment complex. We characterized empty encapsulin using dynamic light scattering (DLS) and transmission electron microscopy (TEM). Both techniques revealed that empty encapsulin forms monodisperse particles with an average diameter of 24 nm (Figure 2-4C,D). We further assayed empty encapsulin assembly with native mass spectrometry (MS). A distribution corresponding to encapsulin was centered at m/z ~ 18,000 with a measured mass of 1,835,100 ± 500 Da (Figure 2-5). The measured mass was less than the expected mass of a 60-mer nanocompartment, 1,842,540 Da, but greater than that of 59-mer nanocompartment, 1,811,831 Da. A minor signal centered at m/z ~ 11,500 was observed that likely corresponds to contamination from co-purified tetradecameric GroEL. In order to determine the reason for the mass

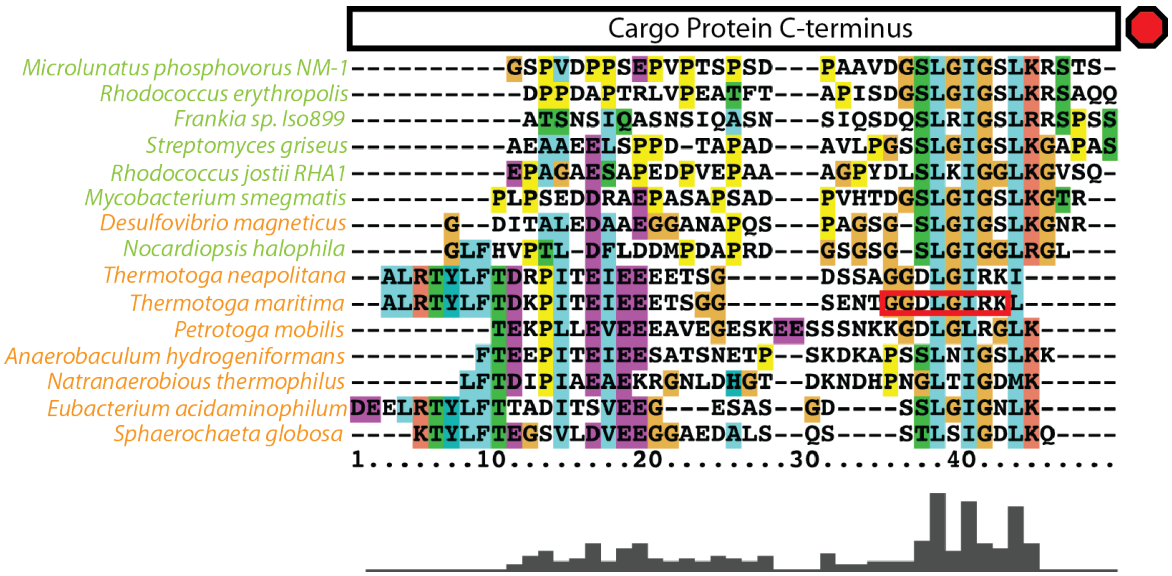

Figure 2-3: Multiple sequence alignment of CLP sequences from encapsulated DypB and FLP proteins. The peptide found bound to the lumen of *T. maritima* encapsulin is shown by a red rectangle.

discrepancy, encapsulin was fragmented into protomers by collision-induced dissociation (CID) tandem MS. Only one protomer species was observed with a mass of 30,579 ± 1 Da; however, the expected protomer mass is 30,709 Da. The 130 Da difference from the expected mass is likely due to cleavage of the N-terminal methionine, a feature not seen in natively expressed encapsulin[5] but commonly found in *E. coli*-expressed proteins. A 60-mer nanocompartment in which all protomers lack the N-terminal methionine would have a mass of 1,834,740, which is within the error of the observed mass. Interestingly, species corresponding to the protomer (30,579 ± 1 Da), the dimer (61,220 ± 20 Da), the trimer (91,800 ± 20 Da), and the tetramer (122,500 ± 100 Da) were present in low abundance along with the assembled nanocompartment. No distributions corresponding to the pentameric capsomer or any higher-order incompletely assembled species were observed. This may be due to the higher order species having a concentration below the detection limit. It is unclear if these species are from disassembly of the nanocompartments during mass spectrometry or if these species co-purified with the assembled nanocompartments. Empty and cargo-loaded encapsulin therefore form compartments with the same dimensions from the same number of protomers, suggesting that cargo is not necessary for nanocompartment formation.

2.3.3 Reversible Disassembly of Encapsulin

Having the capability to generate large quantities of empty encapsulins, we investigated the possibility of *in vitro* cargo loading. The first step in this process was to identify a set of conditions to reversibly disassemble encapsulin. A previous study demonstrated reversible complete disassembly of encapsulin from *Rhodococcus jostii* RHA1 using acetate buffer at pH 3.[11] To identify suitable conditions for the disassembly of the *T. maritima* encapsulin, we assayed native PAGE behavior as a function of pH (Figure 2-7A). Encapsulin was robust to insult

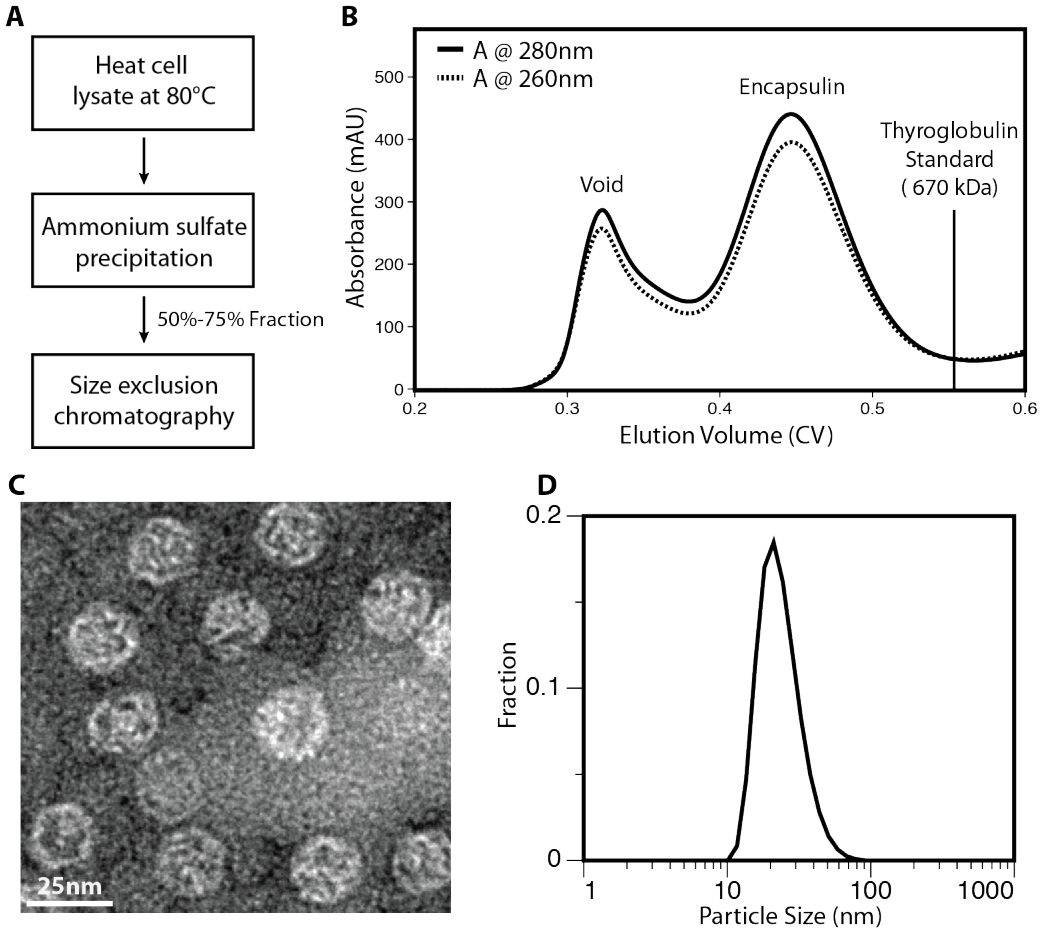

Figure 2-4: A simplified purification strategy for producing empty encapsulin. (A) Simplified protocol for purification of recombinant *T. maritima* encapsulin. (B) Chromatogram of the elution of encapsulin from a Superose 6 size exclusion column in column volumes (CV). (C) TEM of purified empty encapsulin. (D) Particle size of purified empty encapsulins as measured by DLS.

and no apparent disruption of the structure was observed for pH 4 through 12. At pH 2 and 13, two populations – the completely assembled nanocompartment and protomer – were observed, while at pH 0, 1 and 14 only the protomer was observed. Minimal precipitant was observed for any of the conditions tested. Thus, we find that strongly acidic or alkaline conditions induce nanocompartment disassembly without leading to aggregation, making them ideal conditions to pursue for *in vitro* cargo loading.

In vitro cargo loading depends on the protomers' ability to both refold and reassemble after exposure to a denaturing condition. Using circular dichroism (CD), we assayed the ability of encapsulin protomers to refold following exposure to a variety of denaturing conditions. Untreated encapsulin displayed a strong alpha-helical component, evident by the negative peaks at 222 nm and 210 nm in the CD spectrum (Figure 2-8A). By monitoring changes in the 222 nm signal, we determined the degree of protomer unfolding in each denaturing condition and measured the completeness of protomer refolding. Under strongly alkaline conditions (pH 13), 63 ± 1.4 % of the encapsulin protomers were unfolded; following a return to pH 7.4, a near complete (88 ± 0.6 %) recovery of

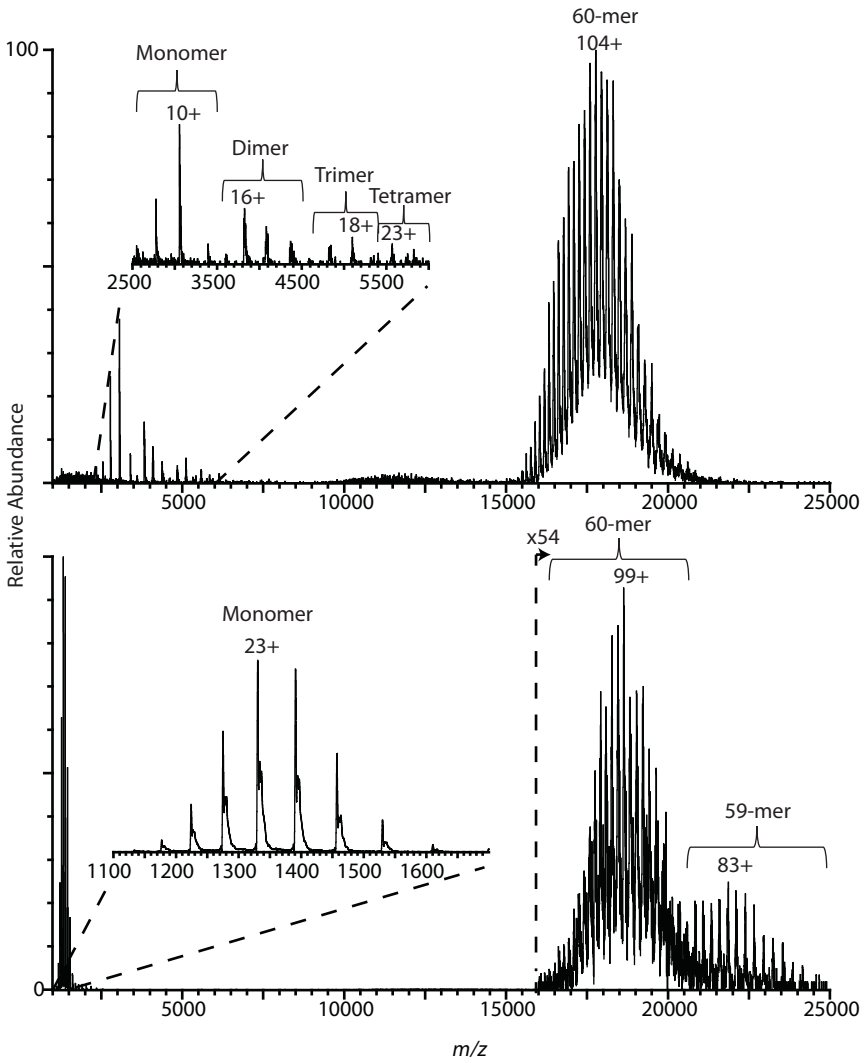

Figure 2-5: T. maritima encapsulin native spray mass spectrometry (Top) Native MS of 5 μM encapsulin in 150 mM ammonium acetate. (Bottom) Product ions after collision induced dissociation of the 60mer encapsulin complex.

the native fold was observed (Figure 2-7B). Interestingly, under strong acidic conditions (pH 1), only a minimal change in 222 nm signal was observed. To probe how acidic conditions affected protomer secondary structure, we collected the full pH 1 CD spectrum, which showed a negative peak at 215 nm, consistent with a secondary structure dominated by beta sheets. Surprisingly, after returning to a neutral pH (7.4), a CD spectrum almost identical to the pre-treatment spectrum was recovered (Figure 2-8B). This suggests that under strongly acidic conditions, protomers reversibly adopt an alternative structure, which may contain beta sheets

Following proper refolding, we hypothesized the protomers would self-assemble into nanocompartments. In order to confirm the completeness of treatment, samples from each condition were visualized using TEM. As expected, there was a significant reduction in the number of nanocompartment

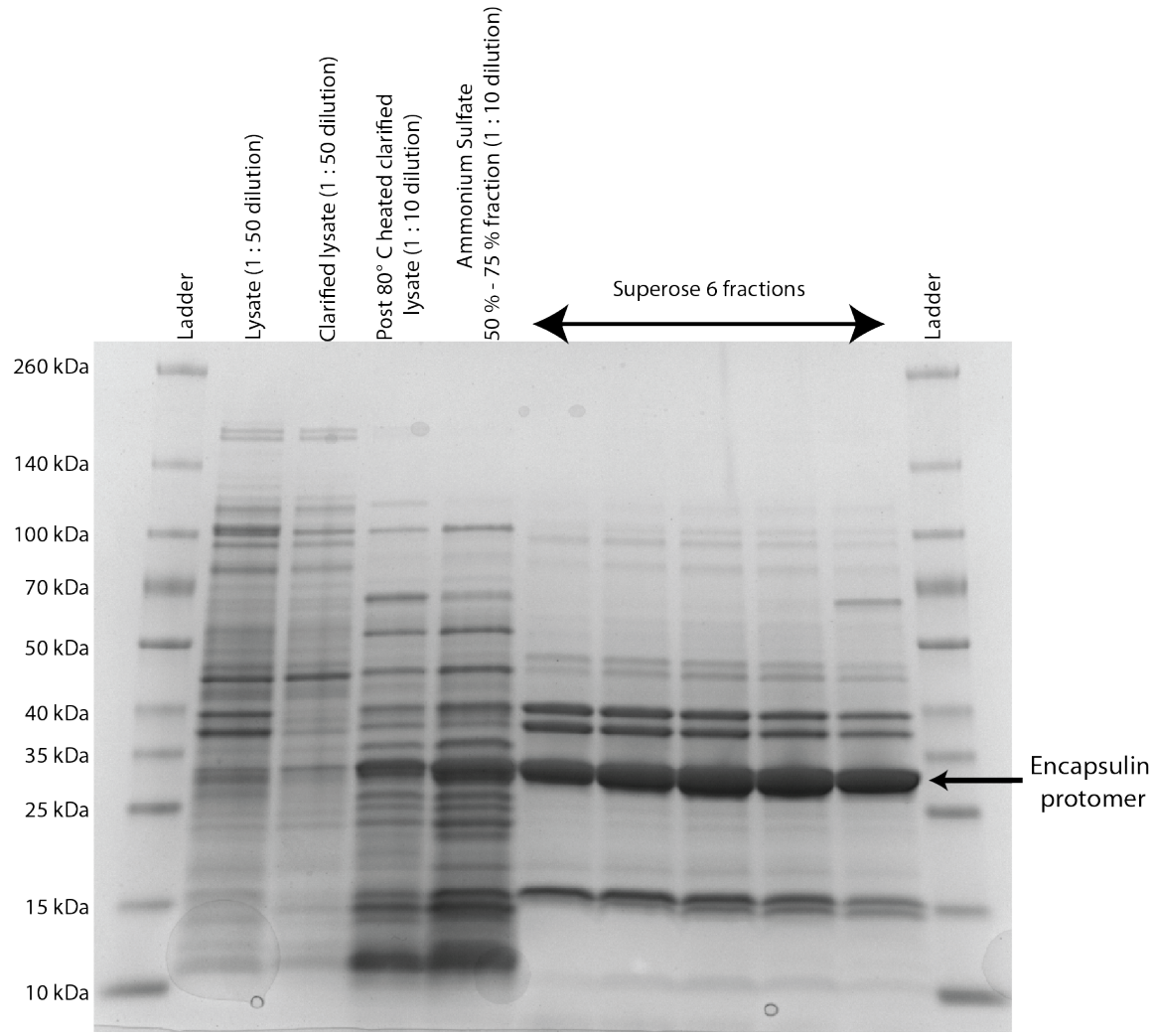

Figure 2-6: SDS-PAGE of the void peak and encapsuiln peak of purified empty encapsulins.

structures for samples at pH 13 or pH 1, while the refolded sample contained a majority of properly reassembled 24 nm nanocompartments (Figure 2-7D). These results demonstrate that strongly acidic or alkaline conditions reversibly disassemble and denature encapsulin, and make them potential conditions for *in vitro* loading of cargo proteins.

We also tested the ability of the chaotropes urea and guanidine hydrochloride (GuHCl) to reversibly denature encapsulin. In 12 M urea, encapsulin was partially unfolded, while treatment with 7 M GuHCl led to complete unfolding (Figure 2-7C). However, after unfolding with GuHCl a smaller fraction was able to refold (60 ± 3 %) than following alkaline or acidic denaturation (88% and 95% respectively). The remaining fraction of protomers misfolded and misassembled into larger aggregates. We screened a number of additives for their ability to increase the fraction of properly reassembled encapsulin, but unfortunately none of the compounds screened had a significantly positive effect on the yield of properly assembled encapsulins (Figure 2-9). Again we confirmed proper reassembly by visualizing samples in each condition by TEM (Figure 2-7E). Note that we observe occasional

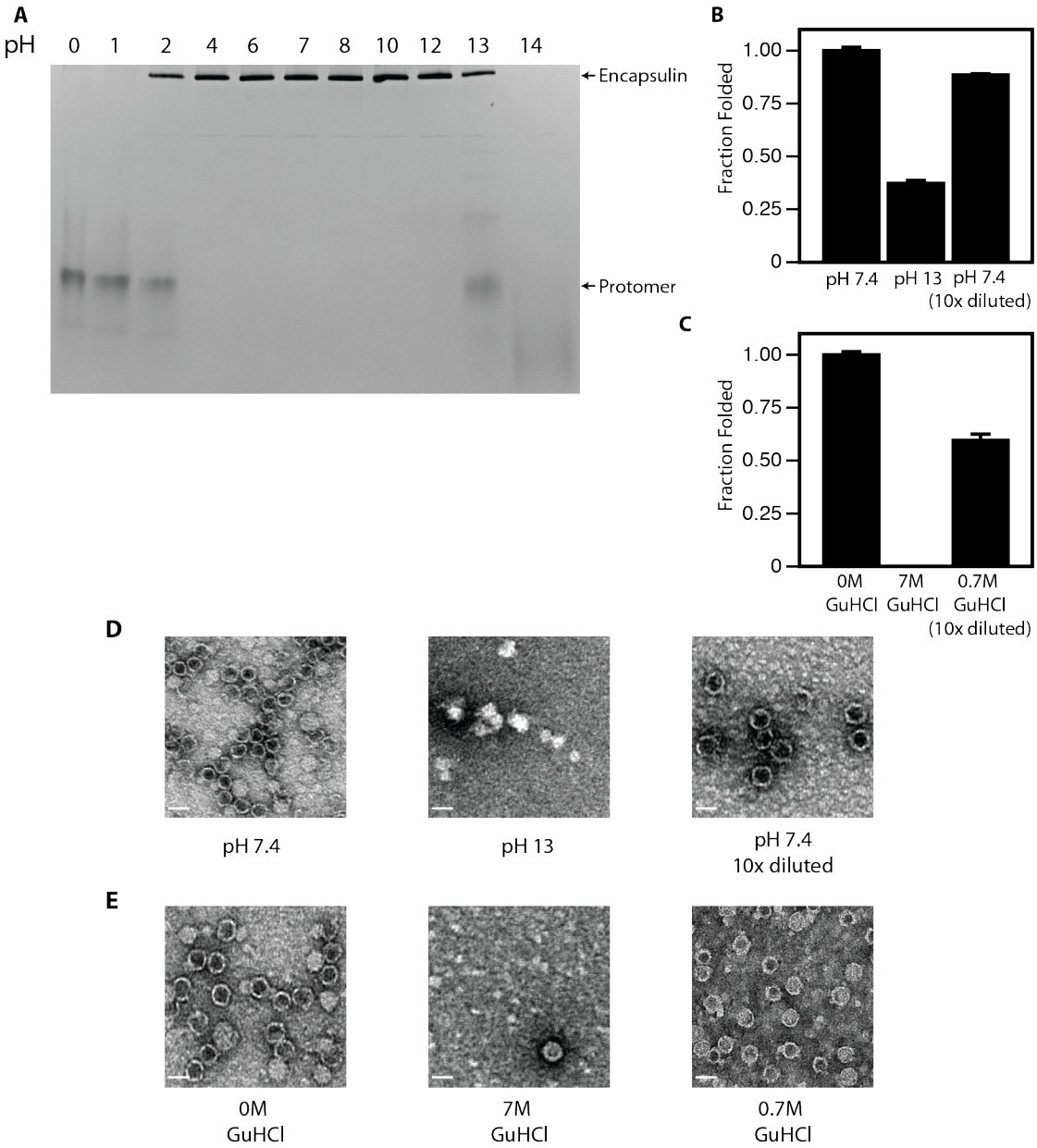

Figure 2-7: Encapsulin can be reversibly disassembled by acidic or alkaline pH or by unfolding with guanidine hydrochloride (GuHCl). (A) Native PAGE of encapsulin after incubation at the indicated pH. (B, C) Fraction of folded encapsulin protomers (as determined by the CD signal at 222 nm) before unfolding, during unfolding, and after refolding for (B) alkaline disassembly/reassembly and (C) for GuHCl disassembly/reassembly. Error bars represent the calculated standard deviation (n = 3). (D, E) TEM of samples before disassembly, during disassembly, and after reassembly for (D) alkaline disassembly/reassembly and (E) for GuHCl disassembly/reassembly. All scale bars are 25 nm.

structures in the 7 M GuHCl condition that, based on the CD data, are likely to have reassembled during the preparation of samples for TEM. Despite the lower refolding yield, the complete unfolding observed in GuHCl makes it a promising candidate for *in vitro* loading.

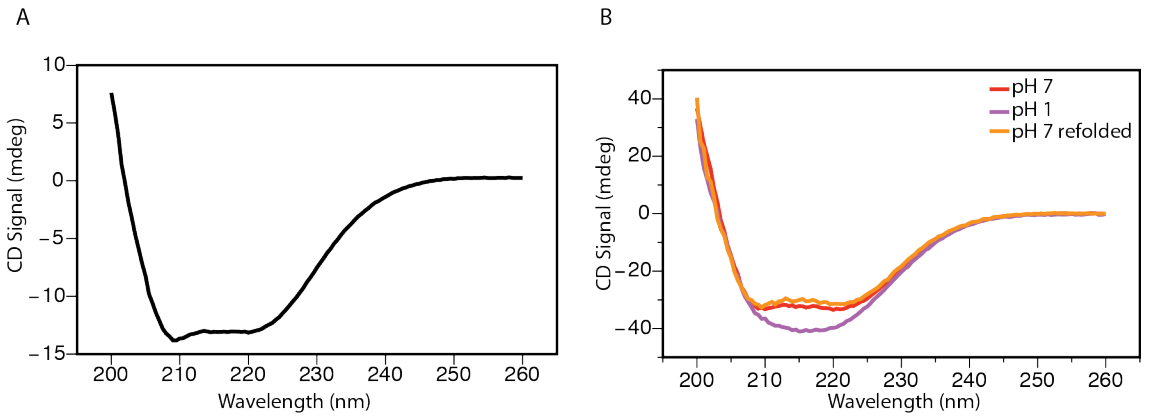

Figure 2-8: CD spectra of encapsulin (A) Spectra of purified empty encapsulin (8 μM) at pH 7.4 (B) Overlay of the CD spectrum of empty encapsulins (16 μM) before acid disassembly, at pH 1, and following reassembly

2.3.4 In Vitro Loading

Having developed three strategies – acidic, alkaline, or chaotropic conditions – for reversibly disassembling encapsulin, we sought to identify which method was the most effective for *in vitro* reloading. Empty encapsulins were disassembled and denatured by each of the established conditions above and allowed to reach equilibrium. Next, we induced refolding and reassembly by 10-fold dilution into a sample containing sfGFP tagged with a 30 amino acid CLP (Figure 2-10A). A 10-fold dilution was chosen as both cargo and shell proteins were fully folded in 0.7 M GuHCl (Figure 2-11A).[25] For all experiments, tagged cargo was added in two-fold molar excess to maximize the amount of cargo loaded. We observed that hexa histidine-tagged sfGFP with a 30 amino acid CLP was able to bind to the exterior of encapsulin. The nonspecific interaction between hexa histidine-tagged sfGFP and encapsulin was due to the hexa histidine tag and no nonspecific interaction was observed when the hexa histidine tag was removed prior to incubation of CLP tagged sfGFP and encapsulin (Figure 2-11B). Additionally, assembled encapsulins were highly resistance to the protease subtilisin (Figure 2-12A-C). Following reassembly, treatment with subtilisin disrupted the nonspecific interaction between encapsulin and tagged sfGFP but had no affect on loaded sfGFP (Figure 2-12D). Reassembled loaded encapsulins were purified by size exclusion to separate them from excess sfGFP and misassembled aggregates (Figure 2-10B).

We sought to measure the effectiveness of each disassembly condition for *in vitro* loading. The amount of cargo loaded *in vitro* was assayed by determining the amount of GFP fluorescence that co-migrated with encapsulin in native PAGE. Given that the disassembly conditions yield significantly different fractions of reassembled nanocompartments, GFP fluorescence was normalized to the density of the encapsulin band following Coomassie staining. While minimal cargo loading was observed for pH-based disassembly methods, robust loading was observed after GuHCl-based disassembly (Figure 2-10C). It is possible that the Coomassie dye may penetrate the encapsulin shell and stain the cargo proteins as well, but because the size of this effect should scale with the extent of cargo

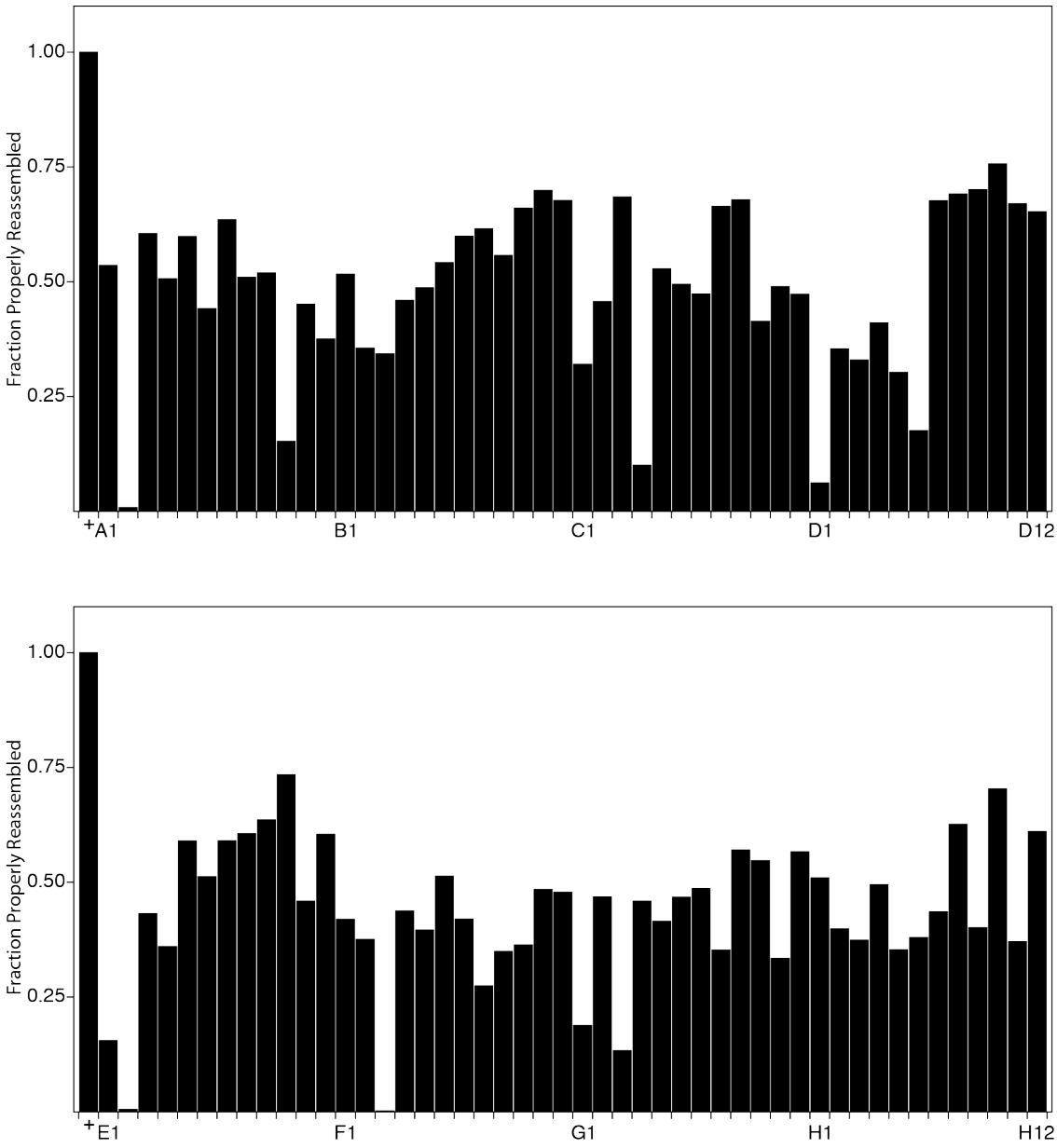

Figure 2-9: Additive Screen to Improve Reassembly Efficiency. Empty encapsulin was disassembled by GuHCl and reassembled by 10-fold dilution into buffer containing the additive. The (+) sample was not disassembled and served as the reference point for complete reassembly. The additive screen for each condition is listed in Table 2-3.

loading, it should not affect the overall trend. Correlating the GFP fluorescence at 511 nm to the absorbance at 280 nm, it is possible to determine the amount of cargo loaded. For sfGFP with a 30 amino acid CLP loaded *in vitro*, we calculated that 7.2 ± 2 sfGFP were loaded per nanocompartment. Previous work has shown that this method may not accurately measure loading for encapsulin.[13]

We next investigated if the signal sequence requirements *in vitro* differed from the minimal 15 amino acid CLP required *in vivo*. Empty encapsulins were disassembled by GuHCl and then refolded and reassembled in the presence of sfGFP with varying CLP lengths (Figure 2-10D). sfGFP lacking a CLP or with a

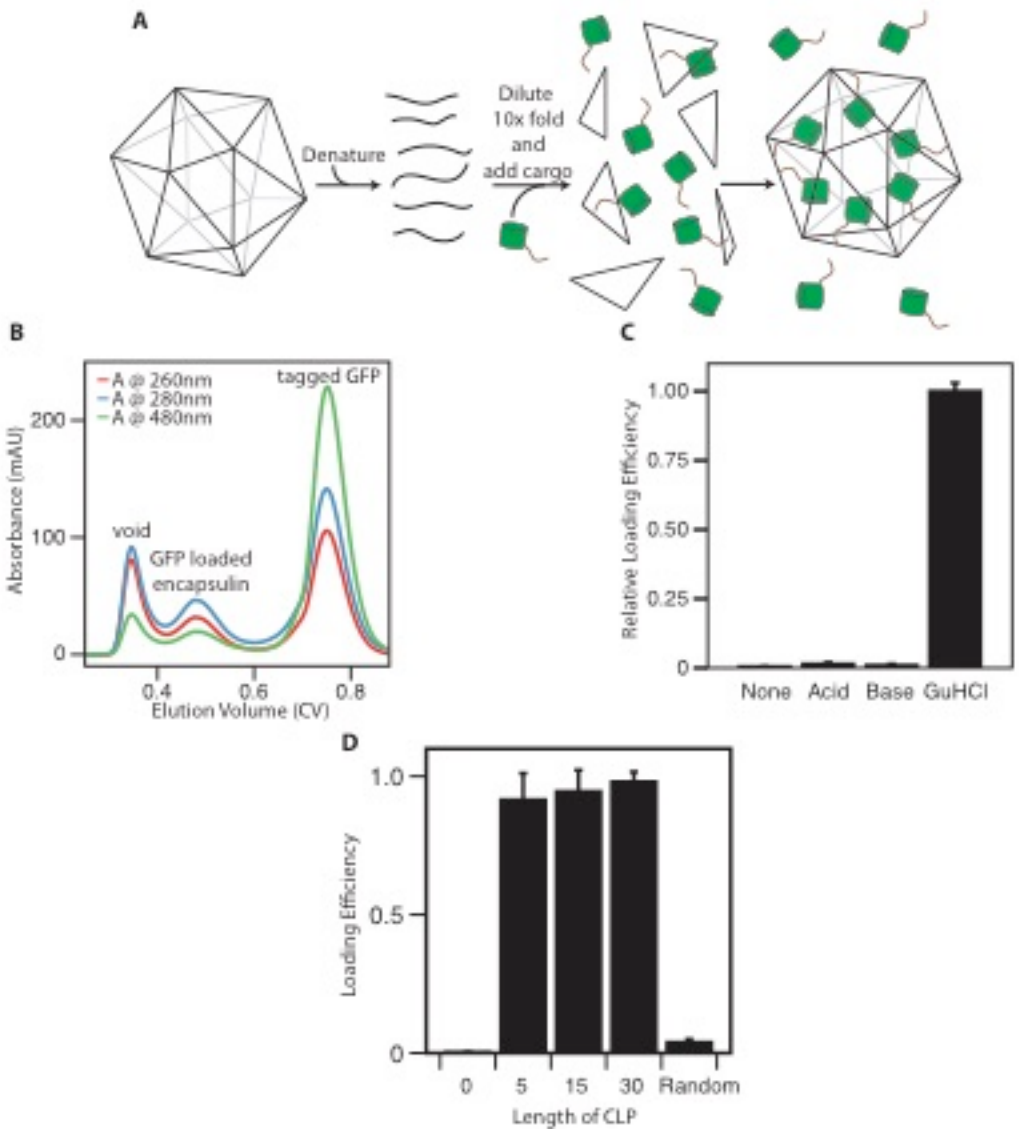

Figure 2-10: *In vitro* loading of sfGFP-CLP into encapsulin. (A) Model for disassembling encapsulin and then loading tagged sfGFP during reassembly. Nanocompartment reassembly is induced by diluting disassembled encapsulin into buffer containing the tagged sfGFP. (B) Size exclusion chromatogram of sfGFP-loaded encapsulin after GuHCl *in vitro* loading. (C) Amount of normalized sfGFP fluorescence that co-migrated with encapsulin on native PAGE following reversed disassembly by the indicated condition. Error bars represent the calculated standard deviation (n=3). (D) Degree of in vitro cargo loading for different CLP lengths. Cargo loading was measured by the amount of normalized sfGFP fluorescence that co-migrated with encapsulin by native PAGE following GuHCl-based disassembly. Reassembly occurred in the presence of sfGFP with the indicated CLP. Error bars represent the calculated standard deviation (n=3).

randomized CLP were not loaded into encapsulin *in vitro*, as seen in the *in vivo* results. When tagged with a 15 or a 30 amino acid CLP, sfGFP was efficiently loaded *in vitro* as these proteins were *in vivo*. Interestingly, the 5 amino acid cargo peptide functioned equivalently to the 15 and 30 amino acid CLP *in vitro,* while this CLP was non-functional *in vivo*. This suggests that the 5 amino acid

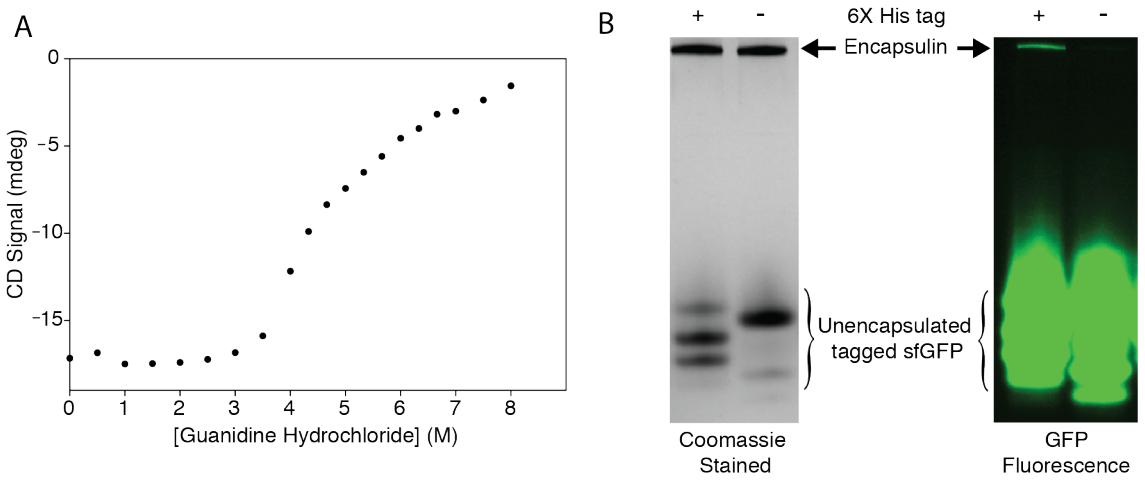

Figure 2-11: Guanidine hydrochloride melt of encapsulin (A) Equilibrium guanidine HCl melt curve of empty encapsulin. (B) A hexa-histidine affinity tag nonspecifically interacts with encapsulin. Coomassie stained (left) and GFP fluorescence (right) from native PAGE of encapsulin incubated with CLP tagged sfGFP that either has or lacks a hexa-histidine affinity tag.

tag is capable of binding to encapsulin but *in vivo* conditions limit successful incorporation. Thus, taken altogether, we have developed a GuHCl-based method for efficient encapsulin cargo loading *in vitro*.

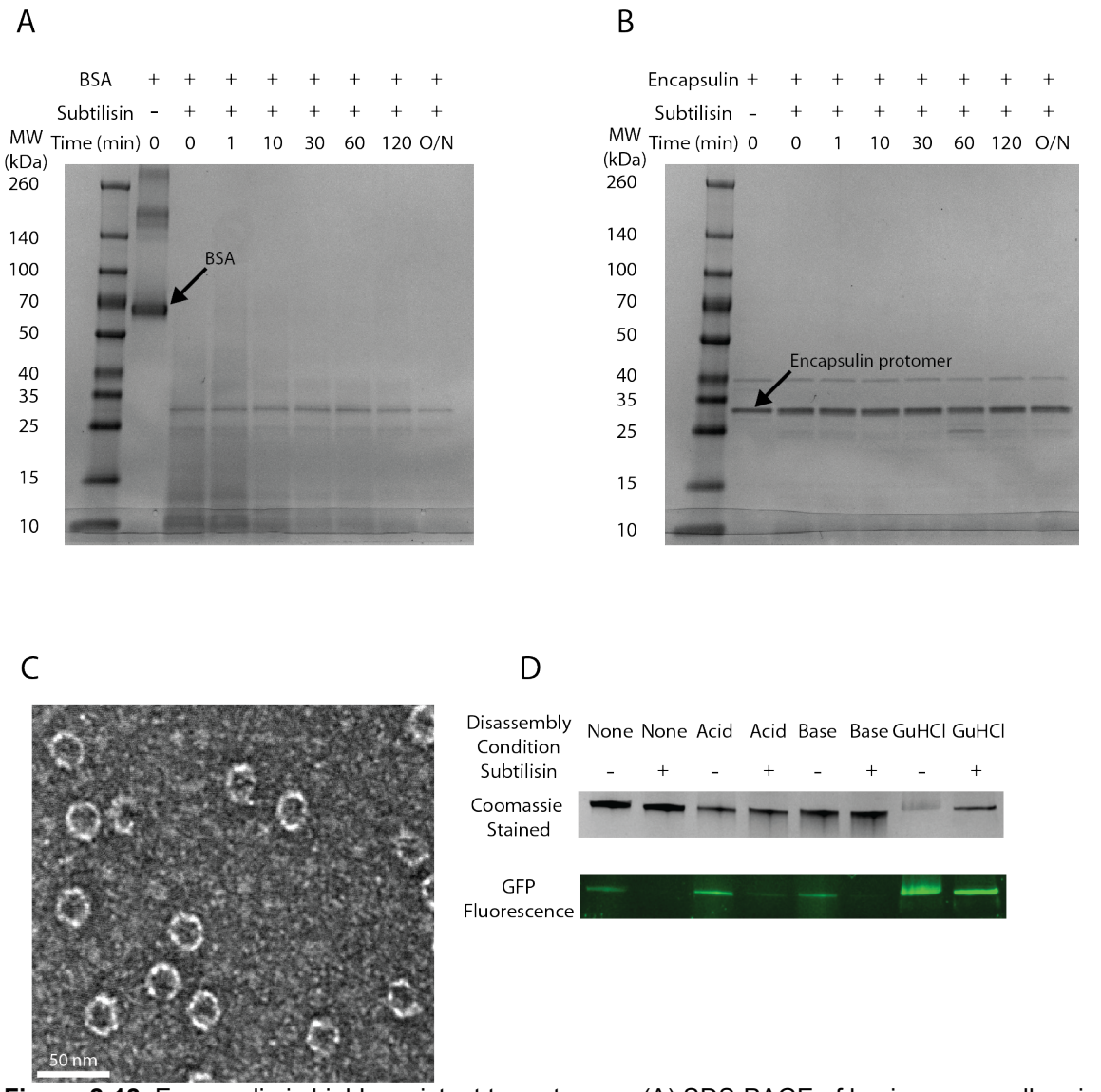

Figure 2-12: Encapsulin is highly resistant to proteases. (A) SDS-PAGE of bovine serum albumin treated with subtilisin. (B) SDS-PAGE of encapsulin treated with subtilisin. (C) TEM of encapsulin after 48 hours of subtilisin treatment. (D) Coomassie and GFP fluorescence from native PAGE of empty encapsulin *in vitro*-loaded with a 30 amino acid CLP following treatment with subtilisin.

2-4 Discussion

Encapsulins are a powerful and simple tool for studying the biological impacts of compartmentalization. Properly assembled encapsulin nanocompartments are readily purified following heterologous expression. Foreign proteins can be targeted to encapsulin by a C-terminal cargo loading peptide *in vitro* and *in vivo.*[12,13] Despite the ease of expression and loading, there has been minimal biochemical characterization of encapsulin and its interaction with the cargo loading peptide. In this study, we demonstrate that *T. maritima* encapsulin is robust to thermal and pH insults, and identify a shorter CLP that targets cargo efficiently *in vitro* and *in vivo*.

Cargo loading was previously achieved with CLPs longer than 30 amino acids; however, robust loading of sfGFP was observed with substantially shorter CLPs. A 15 amino acid CLP is fully functional for targeting sfGFP to encapsulin *in vivo*, suggesting that only a portion of the longer tags are making specific contacts with encapsulin. Interestingly, the 5 amino acid CLP was functional *in vitro* but not *in vivo*. We suspect that the 5 amino acid tag has reduced affinity for encapsulin, and that the concentration of cargo *in vivo* does not favor binding, while the higher concentration of cargo present during *in vitro* loading shifts the equilibrium to favor loading. Alternatively, binding of the 5 amino acid CLP may be disfavored *in vivo* due to competition from other native proteins or metabolites while these are not a factor for *in vitro* loading. We suspect that cargo loading peptides shorter than 15 amino acids will be functional *in vivo*, but more work would be needed to identify the minimal CLP and characterize the importance of individual residues in the sequence.

Encapsulin reversibly disassembles under a variety of different conditions, but we observed that the completeness of reassembly is highly dependent on the disassembly condition. When protomers were denatured due to treatment at alkaline pH or with GuHCl, a fraction of protomers misassembled into larger structures after reassembly was induced. Disassembly by acidic conditions resulted in the formation of a structure that is likely composed of beta sheets, which almost completely reassembled into properly sized nanocompartments. Refolding from a completely unfolded state, as with alkaline pH or GuHCl, may allow protomers to sample an off-pathway conformation that is not sampled when refolding from the beta sheet conformation.

The majority of studied encapsulins assemble into dodecahedrons (i.e. triangulation number = 1) formed from 12 pentamers; however, a few encapsulins have been shown to assemble into larger icosahedrals formed from both pentamers and hexamers.[7,10] Altered assembly following denaturation was previously observed in *Rhodococcus jostii* RHA1 encapsulin, which assembles into a 24nm capsid *in vivo*. Following disassembly at pH 3, however, *R. jostii* encapsulin completely reassembled into a 31 nm structure.[11] During refolding, it is possible that a subset of protomers assemble into hexameric capsomers and associate with pentamers to form larger structures. The high concentration of protomers *in vitro* following disassembly may support formation of hexamers during reassembly. This pool of hexamers and the refolded pentamers could assemble into a stable larger nanocompartment. This would be unlikely *in vivo*, where the pool of free protomer is kept low due to the constant assembly into nanocompartments and the number of misassembled encapsulins would be minimized.

Cargo loading requires the CLP binding site to be accessible during assembly. The three tested denaturation treatments – acid, alkaline, or GuHCl, result in vastly different amounts of *in vitro* cargo loading, which are likely due to differences in accessibility of the cargo peptide binding site during reassembly and differences in reassembly kinetics. Robust *in vitro* cargo loading was observed only for GuHCl-based denaturation, and this is also the only method that yielded complete unfolding of the protomer. We hypothesize that reassembly

proceeds in a path similar to *in vivo* assembly, where the CLP binding site is accessible while the nanocompartment is assembling. The inability to load cargo after pH-based disassembly may be due to an alternative reassembly pathway where the CLP binding site is not formed until after assembly has occurred. Under acidic conditions, protomers adopted an alternative beta-sheet conformation, which suggest there may be an alternative assembly pathway following acid denaturation. Cargo loading occurs during assembly, but questions remain regarding the order of and interplay between CLP binding and nanocompartment assembly.

In summary, we have demonstrated the cargo peptide derived from *T. maritima* FLP is functional for targeting heterologous proteins to encapsulin *in vitro* and *in vivo*. We found that shorter cargo peptides than have been used previously achieve equal degrees of cargo loading. To facilitate biochemical analysis of the nanocompartment shell, we developed a rapid, high-yield purification method and identified conditions under which encapsulin would reversibly disassemble. These conditions can be used to load cargo *in vitro*, negating the need to purify encapsulins that have been co-expressed with the cargo of interest. Although open questions surround the reason for and effects of protein encapsulation[26-28], encapsulins are a simple and tractable system that are well positioned to interrogate these questions.

Plasmid Name	Description
pDEST	pET 14 based Golden Gate destination plasmid
pDEST-Nhis	pET 14 based Golden Gate destination plasmid with a N terminal 6X His tag
pDEST-Chis	pET 14 based Golden Gate destination plasmid with a C terminal 6X His tag
pGFP30_Encap	Operon of sfGFP with a 30 amino acid CLP and Encapsulin in pDEST-Chis
pGFP15_Encap	Operon of sfGFP with a 15 amino acid CLP and Encapsulin in pDEST-Chis
pGFP5_Encap	Operon of sfGFP with a 5 amino acid CLP and Encapsulin in pDEST-Chis
pGFPRand_Encap	Operon of sfGFP with a randomized 30 amino acid CLP and Encapsulin in pDEST-Chis
pEmpty_Encap	Encapsulin in pDEST
pGFP30	sfGFP with a 30 amino acid CLP in pDEST-Nhis
pGFP15	sfGFP with a 15 amino acid CLP in pDEST-Nhis
pGFP5	sfGFP with a 5 amino acid CLP in pDEST-Nhis
pGFPRand	sfGFP with a randomized 30 amino acid CLP in pDEST-Nhis

Table 2-1: Plasmids used Chapter 2

2-5 Methods

2.5.1 Molecular biology and cell culture

The *T. maritima* encapsulin gene was synthesized and codon optimized for expression in *E. coli* by Genewiz (South Plainfield, NJ). The cargo peptide was amplified from *T. maritima* genomic DNA (ATCC 43589D-2) obtained from the ATCC (Manassas, VA). All constructs were assembled using golden gate cloning into a pET 14b expression vector.[20] All primers were ordered from IDT (Coralville, Iowa). Superfolder green fluorescent protein (sfGFP) with the randomized tag was ordered from IDT as a gBlock. A full list of plasmids and primers used in this work are listed in the supplement (Table 4-1 and Table 4-2).

All constructs for purification were transformed into C43 (DE3) cells and grown at 37° C to an optical density (600 nm) of 0.3-0.4. The cells were cooled to 18° C and induced with 500 μM isopropyl β-D-1-thiogalactopyranoside and allowed to grow overnight. Cells were harvested by centrifugation and frozen at -20° C until needed.

2.5.2 Protein Purification

The purification for the tagged sfGFP/encapsulin co-expression construct was adapted from a previous study.[5] Briefly, the cell pellet was resuspended in 20 mM sodium phosphate, pH 7.4, 50 mM NaCl, and 6 mM 2-mercaptoethanol and lysed by sonication. Lysate was clarified by centrifugation in a JA-20 rotor at 30,000 g for 25 minutes. The clarified lysate was layered onto a 32% (w / v) sucrose cushion and centrifuged in a type 45 Ti rotor at 100,000 g for 18 hours. The pellet was resuspended in 20 mM sodium phosate, pH 7.4, 50 mM NaCl, and 6 mM 2-mercaptoethanol. The resuspended pellet was layered onto a 10% - 50% (w / v) sucrose gradient at centrifuged in a SW 41 Ti rotor at 100,000 g for 18 hours. The fractions from the sucrose gradient containing encapsulin were buffer exchanged into buffer A (20 mM sodium phosate, pH 7.4 and 50 mM NaCl) and loaded onto a Superose 6 10/300 GL column (GE Healthcare, Little

Primer Name	Primer Sequence
Encap_f	catgacGGTCTCTgaggaacaatATGAGCGAGTTTCTGAAACGCA
Encap_r	catgacGGTCTCTcgctGAACTTCAGCAGGATCAGGG
FLP_30aa_tag_f	cacaccGGTCTCcCAAAgaagaagaaacgtccgg
FLP_30aa_tag_r	catgacGGTCTCTcctcCAATtcagagcttccttatgccgagg
sfGFP_f	catgacGGTCTCTGTCCAAAGGTGAAGAACTGTTCACCGGTGTTG
sfGFP_r	catgacGGTCTCTtttgtagagctcatccatgccgt
FLP_15aa_tag_f	catgacggtctctcaaaggtggatcagaaaacacgg
FLP_5aa_tag_f	catgacggtctctcaaaggcataaggaagctctgaATTG
FLP_0aa_tag_f	catgacggtctctcaaaATTGgaggaacaatATGAGCGAG
Empty_Encap_f	catgacGGTCTCTgtccGAGTTTCTGAAACGCAGCTTCGC
Encap_removal_r	catgacGGTCTCTcgctCGCTCCTattgttcctcCAATtca

Table 2-2: Primers used Chapter 2

Chalfont, UK). Unless otherwise indicated, all Superose 6 columns were run at 0.5 mL / min. Fractions containing assembled encapsulin were collected and concentrated with a Vivaspin 6 (100 kDa molecular weight cutoff (MWCO)) concentrator.

Constructs for determining the *in vivo* cargo peptide interaction length were purified by nickel immobilized metal ion affinity chromatography (IMAC). Cell pellets were resuspended in buffer B (20 mM sodium phosphate, pH 7.4, 50 mM NaCl, 10 mM imidazole) and lysed by sonication. Cleared lysate was incubated with Ni-nitrilotriacetic acid (NTA) resin from Qiagen (Hilden, Germany) for 1 hour at 4° C. The resin was washed with 20 resin volumes of buffer C (20 mM Sodium Phosphate, pH 7.4, 50 mM NaCl, 30 mM imidazole) and bound encapsulin was eluted with 3 resin volumes of buffer D (20 mM sodium phosphate, pH 7.4, 50 mM NaCl, 350 mM imidazole). The Ni-NTA elute was concentrated and loaded onto a Superose 6 column and fractions containing assembled encapsulin were collected and concentrated.

Empty encapsulin cell pellets were resuspended in buffer A and lysed by homogenization with an Avestin EmulsiFlex-C3 (Ottawa, Canada). Cleared lysate was incubated at 80° C for 90 minutes and then clarified by centrifugation at 30,000 g for 25 minutes. Ammonium sulfate was added to a final 50% (w / v) concentration and precipitant was removed. Ammonium sulfate was added to the soluble fraction to reach a final 75% (w / v) concentration. The precipitant was resuspended in buffer A and loaded onto a Superose 6 column. Fractions containing encapsulin were collected and concentrated.

sfGFP constructs tagged with different CLP lengths used for *in vitro* loading were purified by Ni IMAC. For each construct, the cell pellet was resuspended in buffer B (20 mM sodium phosphate, pH 7.4, 50 mM NaCl, 10 mM Imidazole) and lysed by sonication. Clarified lysate was incubated with Ni-NTA resin for 1 hour at 4° C. The resin was washed with 20 resin volumes of buffer C (20 mM sodium phosphate, pH 7.4, 50 mM NaCl, 30 mM imidazole) and tagged sfGFP was eluted with 3 resin volumes of buffer D (20 mM Sodium Phosphate, pH 7.4, 50 mM NaCl, 350 mM Imidazole). The Ni column elute was concentrated

and loaded onto a Superose 6 column. Fractions containing assembled tagged sfGFP were collected and concentrated with Vivaspin 6 (10 kDA MWCO) concentrators.

2.5.3 Total Internal Reflection (TIRF) Microscopy

Encapsulin loaded with sfGFP (5 µL at 160 nM) was spotted on a 1% (w / v) agarose pad, allowed to dry for 5 minutes, and imaged on an inverted TIRF microscope previously described.[21] GFP fluorophores were excited with the 488 nm laser and images were collected for 300 seconds with a 0.1 second exposure per image. Image files were analyzed with FIJI.[22] A composite image was made by projecting the max intensity from the first 50 images. This cutoff was chosen due to substantial photobleaching observed during the course of the experiment. The fluorescence of each punctae was normalized to the area of the punctae and plotted with DataGraph.

2.5.4 Fluorescence Measurements

Fluorescence measurements were taken in a Tecan M1000 PRO (San Jose, CA) plate reader. Samples were excited from the top at 488 nm with a 5.0 nm bandwith and emission was measured at 511 nm with a 5.0 nm bandwith. At total of 50 flashes were made per well each with an integration time of 20 µs.

2.5.5 Biophysical Characterization

Polyacrylamide gel electrophoresis (PAGE): Samples were run on Criterion TGX 4-20% gels (BioRad, Hercules, CA) and stained with GelCode Blue for 30 minutes at room temperature (RT) and destained with diH2O for two hours at RT. Gels were imaged with a ChemiDoc MP Imaging System (BioRad, Hercules, CA) and densitometry analysis was performed with Image Lab (BioRad, Hercules, CA).

Electron Microscopy: Empty encapsulin (8 µM) was adsorbed onto a thin carbon grid and stained with 1% (w/v) uranyl acetate solution. Films were washed with water twice and dehydrated. Grids were examined with a JOEL 12000EX or Tecnai 12 transmission electron microsocope and images were captured with a charge-coupled device (CCD) camera.

Dynamic Light Scattering (DLS): Data was collected with a Zetasizer Nano ZS from 32 µM empty encapsulin in 50 mM PBS buffer. Signal was averaged over 15 trials each lasting 30 seconds.

Native Mass Spectroscopy: Purified empty encapsulin was unfolded with 7 M GuHCl in buffer A. The unfolded encapsulin was loaded onto a Superose 6 column equilibrated with 7 M GuHCl in buffer A and the fractions corresponding to encapsulin were collected. Encapsulin reassembly was induced by diluting the disassembled encapsulin 10-fold into A buffer. Samples were allowed reassemble for 18 hours at 4° C. Reassembled encapsulin was loaded onto a Superose 6 column. Fractions containing assembled encapsulin were collected, buffer exchanged into 150 mM ammonium acetate, and then concentrated to encapsulin protomer concentration of 300 µM (5 µM 60-mer nanocompartment). Data was acquired on a Waters Quadrupole-Time-of-Flight (Q-TOF) Premier

(Waters, Milford, MA) mass spectrometer. Nanoelectrospray from borosilicate capillaries (1.0 mm outer diameter /0.78 mm inner diameter, Sutter Instruments, Novato, CA, USA) pulled to a tip inner diameter of ~1 μm with a Flaming/Brown micropipette puller (Model P-87, Sutter Instruments, Novato, CA, USA) was used to form protein ions. Nanoelectrospray was initiated by applying a 1.6 kV potential to a platinum wire in contact with the sample solution in the capillary. The backing pressure in the source of the instrument was raised to ~6.6 mbar and the sampling cone voltage set to 200 V to aid in protein complex desolvation. Measured masses were adjusted for peak broadening due to solvent adduction using the method of Robinson and coworkers.[23]

Encapsulin pH Stability: Empty encapsulin (8 μM) was incubated with 90 mM phosphate buffer at the specified pH for 90 minutes at room temperature. Samples from each pH were mixed with 4X native loading dye (200 mM Tris-HCl pH 6.8, 40 % glycerol, and 0.08 % bromophenol blue) and analyzed by native PAGE.

Circular Dichroism (CD): Data was collected on a J-815 circular dichroism spectrometer (JASCO) from 8 μM empty encapsulin in buffer A with 1 mM dithiothreitol (DTT) in a 1 mm quartz cell. Spectra were collected for 200 nm – 260 nm in 0.5 nm steps with one second of signal averaging for each step. Five accumulations were averaged for each spectra. The GuHCl equilibrium melt curve was determined by incubating 8 μM empty encapsulin with GuHCl in buffer A plus 1 mM DTT overnight. Signal at 222 nm was collected for 120 seconds with 2 seconds of signal averaging at each time point for each indicated concentration of GuHCl. For measuring fraction folded and fraction unfolded, signal at 222 nm was collected for 120 seconds with two seconds of signal averaging for each time point. The completely folded point was defined as the signal measured from 8 μM empty encapsulin in buffer A. The completely unfolded point was defined as the signal measured from 8 μM empty encapsulin in 7 M GuHCl.

2.5.6 Disassembly / Reassembly

Acid-based disassembly was induced incubating 80 μM encapsulin in buffer A plus 1mM DTT with HCl (150 mM final concentration) for 1 hour at 4° C. The pH of the solution was determined to be approximately 1 by pH strips.

Base-based disassembly was induced incubating 80 μM encapsulin in buffer A plus 1mM DTT with NaOH (150 mM final concentration) for 1 hour at 4° C. The pH of the solution was determined to be approximately 13 by pH strips.

For guanidine hydrochloride (GuHCl)-based disassembly, empty encapsulin was mixed with 8 M GuHCl to reach 7 M GuHCl in buffer A with 1mM DTT. Encapsulin was allowed to unfold for 1 hour at 4° C.

For all conditions, reassembly was initiated by diluting the sample 10-fold into buffer A with 1 mM DTT and allowing the sample to incubate at RT for 16 hours. For *in vitro* cargo loading, the dilution buffer contained 16 μM tagged sfGFP. Protease treatment was performed by incubating samples with 100 units subtilisin (Sigma-Aldrich, St. Louis, MO) at RT for 3 hours.

2.5.7 Multiple Sequence Alignment

A BLAST search of the non-redundant database against T. maritima FLP sequence revealed other encapsulated FLP proteins. Encapsulated DypB sequences were identified similarly. The multiple sequence alignment was generated and viewed with Clustal X[29].

2.5.8 Reassembly Additive Screen

20 μM empty encapsulin was disassembled in 7M GuHCl. Refolding was initiated by diluting 10-fold into buffer containing the additive. Additives were taken from the Solublity & Stability Screen (Hampton Research, Aliso Viejo CA).

2.5.9 Protease Sensitivity

2 μM empty encapsulin or 2 μM bovine serum albumin (BSA) (Sigma-Aldrich) was incubated with 5 μg of subtilisin for the indicated time at RT. The reaction was stopped by addition of 4 X SDS loading dye and heating at 90° C for 10 minutes. For TEM, a sample was taken after 48 hours at RT and imaged by negative stain.

2.5.10 Hexa histidine affinity tag cleavage

Purified hexa histidine tagged sfGFP was incubated with 100 : 1 (w / w) protease from the tobacco etch virus overnight at 4° C. Protease and uncleaved sfGFP was removed by running the reaction over Ni-NTA resin equilibrated with Buffer A. The flow-through was concentrated using a Vivaspin concentrator.

Table 2-3: Additives Screened to Improve Reassembly Efficiency

Well	[conc]	units	Additive 1	[conc]	units	Additive 2
A01	10	% v/v	Water			
A02	7.5	% w/v	Trichloroacetic acid			
A03	25	mM	L-Arginine			
A04	25	mM	L-Arginine	25	mM	L-Glutamic acid
A05	50	mM	Glycine			
A06	50	mM	L-Proline			
A07	12	mM	L-Histidine			
A08	50	mM	β-Alanine			
A09	50	mM	L-Serine			
A10	50	mM	L-Arginine ethyl ester dihydrochloride			
A11	50	mM	L-Argininamide dihydrochloride			
A12	50	mM	6-Aminohexanoic acid			
B01	50	mM	Gly-gly			
B02	20	mM	Gly-gly-gly			
B03	0.5	% w/v	Tryptone			
B04	250	mM	Betaine monohydrate			
B05	75	mM	D-(+)-Trehalose dihydrate			
B06	200	mM	Xylitol			
B07	200	mM	D-Sorbitol			
B08	200	mM	Sucrose			
B09	50	mM	Hydroxyectoine			
B10	250	mM	Trimethylamine N-oxide dihydrate			
B11	200	mM	Methyl-α-D-glucopyranoside			
B12	1	% v/v	Triethylene glycol			
C01	50	mM	Spermine tetrahydrochloride			
C02	50	mM	Spermidine			
C03	50	mM	5-Aminovaleric acid			
C04	50	mM	Glutaric acid			
C05	8	mM	Adipic acid			
C06	50	mM	Ethylenediamine dihydrochloride			
C07	50	mM	Guanidine hydrochloride			
C08	50	mM	Urea			
C09	50	mM	N-Methylurea			
C10	20	mM	N-Ethylurea			
C11	3	% w/v	N-Methylformamide			
C12	0.3	% w/v	Hypotaurine			
D01	15	mM	TCEP hydrochloride			
D02	2	mM	GSH (L-Glutathione reduced)	2	mM	GSSG (L-Glutathione oxidized)
D03	0.5	% w/v	Benzamidine hydrochloride			
D04	5	mM	Ethylenediaminetetraacetic acid disodium salt dihydrate			
D05	10	mM	Magnesium chloride hexahydrate	10	mM	Calcium chloride dihydrate
D06	10	mM	Cadmium chloride hydrate	10	mM	Cobalt(II) chloride hexahydrate
D07	100	mM	Non Detergent Sulfobetaine 195 (NDSB-195)			

D08	100	mM	Non Detergent Sulfobetaine 201 (NDSB-201)			
D09	100	mM	Non Detergent Sulfobetaine 211 (NDSB-211)			
D10	100	mM	Non Detergent Sulfobetaine 221 (NDSB-221)			
D11	80	mM	Non Detergent Sulfobetaine 256 (NDSB-256)			
D12	50	mM	Taurine			
E01	50	mM	Acetamide			
E02	50	mM	Oxalic acid dihydrate			
E03	50	mM	Sodium malonate pH 7.0			
E04	50	mM	Succinic acid pH 7.0			
E05	0.5	% v/v	Tacsimate pH 7.0			
E06	2.5	% w/v	Tetraethylammonium bromide			
E07	2.5	% w/v	Cholin acetate			
E08	2.5	% w/v	1-Ethyl-3-methylimidazolium acetate			
E09	2.5	% w/v	1-Butyl-3-methylimidazolium chloride			
E10	2.5	% w/v	Ethylammonium nitrate			
E11	50	mM	Ammonium sulfate			
E12	50	mM	Ammonium chloride			
F01	50	mM	Magnesium sulfate hydrate			
F02	50	mM	Potassium thiocyanate			
F03	25	mM	Gadolinium(III) chloride hexahydrate			
F04	25	mM	Cesium chloride			
F05	25	mM	4-Aminobutyric acid (GABA)			
F06	50	mM	Lithium nitrate			
F07	50	mM	DL-Malic acid pH 7.0			
F08	50	mM	Lithium citrate tribasic tetrahydrate			
F09	25	mM	Ammonium acetate			
F10	25	mM	Sodium benzenesulfonate			
F11	25	mM	Sodium p-toluenesulfonate			
F12	100	mM	Sodium chloride			
G01	140	mM	Potassium chloride			
G02	70	mM	Sodium phosphate monobasic monohydrate	130	mM	Potassium phosphate dibasic
G03	100	mM	Sodium sulfate decahydrate			
G04	140	mM	Lithium chloride			
G05	100	mM	Sodium bromide			
G06	4	% v/v	Glycerol	40	mM	Lithium chloride
G07	5	% v/v	Glycerol			
G08	1	% v/v	Ethylene glycol			
G09	1	% v/v	Polyethylene glycol 200			
G10	0.5	% v/v	Polyethylene glycol monomethyl ether 550			
G11	0.5	% w/v	Polyethylene glycol monomethyl ether 750			
G12	5	% v/v	Formamide			
H01	5	% v/v	Polypropylene glycol P 400			
H02	2.5	% v/v	Pentaerythritol ethoxylate (15/4 EO/OH)			
H03	1	% w/v	1,2-Propanediol			
H04	0.3	% w/v	Polyethylene glycol monomethyl ether 1,900			
H05	0.3	% w/v	Polyethylene glycol 3,350			

H06	**0.3**	**% w/v**	**Polyethylene glycol 8,000**
H07	**0.2**	**% w/v**	**Polyvinylpyrrolidone K15**
H08	**10**	**mM**	**6-O-α-D-Maltosyl-β-cyclodextrin**
H09	**1**	**mM**	**(2-Hydroxypropyl)-β-cyclodextrin**
H10	**8**	**mM**	**α-Cyclodextrin**
H11	**1**	**mM**	**β-Cyclodextrin**
H12	**5**	**mM**	**Methyl-β-cyclodextrin**

2-6 References

(1) Yeates, T. O., Thompson, M. C., and Bobik, T. A. (2011) The protein shells of bacterial microcompartment organelles. *Curr. Opin. Struct. Biol. 21*, 223–231.
(2) Kerfeld, C. A., Heinhorst, S., and Cannon, G. C. (2010) Bacterial microcompartments. *Annu. Rev. Microbiol. 64*, 391–408.
(3) Chowdhury, C., Sinha, S., Chun, S., Yeates, T. O., and Bobik, T. A. (2014) Diverse bacterial microcompartment organelles. *Microbiol. Mol. Biol. Rev. 78*, 438–468.
(4) Axen, S. D., Erbilgin, O., and Kerfeld, C. A. (2014) A taxonomy of bacterial microcompartment loci constructed by a novel scoring method. *PLoS Comput. Biol. 10*, e1003898.
(5) Sutter, M., Boehringer, D., Gutmann, S., Günther, S., Prangishvili, D., Loessner, M. J., Stetter, K. O., Weber-Ban, E., and Ban, N. (2008) Structural basis of enzyme encapsulation into a bacterial nanocompartment. *Nat. Struct. Mol. Biol. 15*, 939–947.
(6) Valdés-Stauber, N., and Scherer, S. (1994) Isolation and characterization of Linocin M18, a bacteriocin produced by Brevibacterium linens. *Appl. Environ. Microbiol. 60*, 3809–3814.
(7) McHugh, C. A., Fontana, J., Nemecek, D., Cheng, N., Aksyuk, A. A., Heymann, J. B., Winkler, D. C., Lam, A. S., Wall, J. S., Steven, A. C., and Hoiczyk, E. (2014) A virus capsid-like nanocompartment that stores iron and protects bacteria from oxidative stress. *EMBO J. 33*, 1896–1911.
(8) 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., and Goulding, C. W. (2014) Characterization of a Mycobacterium tuberculosis nanocompartment and its potential cargo proteins. *J. Biol. Chem. 289*, 18279–18289.
(9) Moon, H., Lee, J., Min, J., and Kang, S. (2014) Developing genetically engineered encapsulin protein cage nanoparticles as a targeted delivery nanoplatform. *Biomacromolecules 15*, 3794–3801.
(10) Akita, F., Chong, K. T., Tanaka, H., Yamashita, E., Miyazaki, N., Nakaishi, Y., Suzuki, M., Namba, K., Ono, Y., Tsukihara, T., and Nakagawa, A. (2007) The crystal structure of a virus-like particle from the hyperthermophilic archaeon Pyrococcus furiosus provides insight into the evolution of viruses. *J. Mol. Biol. 368*, 1469–1483.
(11) Rahmanpour, R., and Bugg, T. D. H. (2013) Assembly in vitro of Rhodococcus jostii RHA1 encapsulin and peroxidase DypB to form a nanocompartment. *FEBS J. 280*, 2097–2104.
(12) Tamura, A., Fukutani, Y., Takami, T., Fujii, M., Nakaguchi, Y., Murakami, Y., Noguchi, K., Yohda, M., and Odaka, M. (2015) Packaging guest proteins into the encapsulin nanocompartment from Rhodococcus erythropolis N771. *Biotechnol. Bioeng. 112*, 13–20.
(13) Rurup, W. F., Snijder, J., Koay, M. S. T., Heck, A. J. R., and Cornelissen, J. J. L. M. (2014) Self-sorting of foreign proteins in a bacterial nanocompartment. *J. Am. Chem. Soc. 136*, 3828–3832.
(14) Radford, D. (2015) Understanding the Encapsulins: Prediction and

Characterization of Phage Capsid-like Nanocompartments in Prokaryotes.
(15) Zhou, H.-X., Rivas, G., and Minton, A. P. (2008) Macromolecular crowding and confinement: biochemical, biophysical, and potential physiological consequences. *Annu Rev Biophys 37*, 375–397.
(16) Kim, E. Y., Slininger, M. F., and Tullman-Ercek, D. (2014) The effects of time, temperature, and pH on the stability of PDU bacterial microcompartments. *Protein Sci. 23*, 1434–1441.
(17) Comellas-Aragonès, M., Engelkamp, H., Claessen, V. I., Sommerdijk, N. A. J. M., Rowan, A. E., Christianen, P. C. M., Maan, J. C., Verduin, B. J. M., Cornelissen, J. J. L. M., and Nolte, R. J. M. (2007) A virus-based single-enzyme nanoreactor. *Nat Nanotechnol 2*, 635–639.
(18) Patterson, D. P., Prevelige, P. E., and Douglas, T. (2012) Nanoreactors by programmed enzyme encapsulation inside the capsid of the bacteriophage P22. *ACS Nano 6*, 5000–5009.
(19) Patterson, D. P., Schwarz, B., Waters, R. S., Gedeon, T., and Douglas, T. (2014) Encapsulation of an enzyme cascade within the bacteriophage P22 virus-like particle. *ACS Chem. Biol. 9*, 359–365.
(20) Engler, C., Kandzia, R., and Marillonnet, S. (2008) A one pot, one step, precision cloning method with high throughput capability. *PLoS ONE 3*, e3647.
(21) Lee, S.-H., Shin, J. Y., Lee, A., and Bustamante, C. (2012) Counting single photoactivatable fluorescent molecules by photoactivated localization microscopy (PALM). *Proc. Natl. Acad. Sci. U.S.A. 109*, 17436–17441.
(22) 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., and Cardona, A. (2012) Fiji: an open-source platform for biological-image analysis. *Nature Publishing Group 9*, 676–682.
(23) McKay, A. R., Ruotolo, B. T., Ilag, L. L., and Robinson, C. V. (2006) Mass measurements of increased accuracy resolve heterogeneous populations of intact ribosomes. *J. Am. Chem. Soc. 128*, 11433–11442.
(24) Snijder, J., van de Waterbeemd, M., Damoc, E., Denisov, E., Grinfeld, D., Bennett, A., Agbandje-McKenna, M., Makarov, A., and Heck, A. J. R. (2014) Defining the stoichiometry and cargo load of viral and bacterial nanoparticles by Orbitrap mass spectrometry. *J. Am. Chem. Soc. 136*, 7295–7299.
(25) Andrews, B. T., Schoenfish, A. R., Roy, M., Waldo, G., and Jennings, P. A. (2007) The rough energy landscape of superfolder GFP is linked to the chromophore. *J. Mol. Biol. 373*, 476–490.
(26) Chowdhury, C., Chun, S., Pang, A., Sawaya, M. R., Sinha, S., Yeates, T. O., and Bobik, T. A. (2015) Selective molecular transport through the protein shell of a bacterial microcompartment organelle. *Proc. Natl. Acad. Sci. U.S.A. 112*, 2990–2995.
(27) Sanfelice, D., Politou, A., Martin, S. R., De Los Rios, P., Temussi, P., and Pastore, A. (2013) The effect of crowding and confinement: a comparison of Yfh1 stability in different environments. *Phys Biol 10*, 045002.
(28) Avalos, J. L., Fink, G. R., and Stephanopoulos, G. (2013) Compartmentalization of metabolic pathways in yeast mitochondria improves the

production of branched-chain alcohols. *Nature Biotechnology 31*, 335–341.
(29) Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J., and Higgins, D. G. (2007) Clustal W and Clustal X version 2.0. *Bioinformatics 23*, 2947–2948.

Chapter 3 The Role of Encapsulin Flavoproteins in Iron Storage and Mobilization

3-1 Introduction

Iron is an essential element for growth and survival in all organisms due to the ubiquitous role iron-sulfur clusters play in catalyzing reduction and oxidation (redox) reactions. Environmentally, iron is usually present as ferric, Fe(III), iron oxide minerals. The high insolubility of these minerals limits the bioavailability of iron and necessitates the need for cellular uptake and storage. Cellular uptake requires the solubilization of ferric iron. One method for solubilizing Fe(III) is reducing the extracellular insoluble Fe(III) to soluble ferrous iron, Fe(II), and transporting it into the cell. However, iron can not be stored intracellularly as ferrous iron due to its spontaneous reaction with reactive oxygen species (ROS) to form hydroxyl radicals that are highly damaging to cells[1]. To avoid this destructive reaction, cells reoxidize iron and store it as an insoluble mineral granule. Proteins in the ferritin superfamily form compartments, catalyze the oxidation of free intracellular iron, and deposit the iron inside the shell forming an iron oxide mineral. There are three specific iron-sequesting ferritin superfamily members known in the three domains of life: ferritin, bacterioferritin, and DNA-binding protein from starved cells (Dps)[2,3]. Ferritin, the most ubiquitous iron storage system, forms a ~12 nm cage from 24 homo-monomers[2–4]. Each monomer has a ferroxidase center (FOC) that catalyzes the oxidation of ferrous iron to ferric iron and each shell can hold ~5000 Fe(III) atoms[2–4]. Bacterioferritins are structurally similar to ferritins, forming a ~12 nm cage from 24 homo-monomers[2,5,6]. They differ from ferritins in their use of heme groups in the oxidation of iron and lower total iron storage capacity, ~2000 Fe(III) atoms per bacterioferritin[6]. The smallest iron storage cages are Dps proteins which form a ~9 nm cage from 12 homo-monomers and can store ~500 Fe(III) atoms per cage[2,3,7]. Dps proteins bind to chromosomal DNA and are thought to protect DNA from ROS damage by using the ROS to oxidize ferrous iron[7].

Iron oxidation for proteins in the ferritin superfamily is catalyzed at the ferroxidase center (FOC). In aerobic conditions, the oxidation of iron is coupled to the reduction of molecular oxygen[2]. Under anaerobic conditions, the oxidizing agent is less well established. Following oxidation, the ferric iron is released from the FOC and deposited in the central cavity as a ferrihydrite or iron phosphate core[2]. While iron oxidation and deposition are well characterized, the mechanism of *in vivo* iron release is not well understood. Iron can be released from ferritins *in vitro* with many non-specific reducing agents, such as glutathione, sodium dithionite, and hydroxylamine hydrochloride[8]. It is unlikely that iron is reduced indiscriminately *in vivo* and it has been proposed that electrons are shuttled to the iron core by ferrodoxins or flavodoxins[9]. Weeratunga et al. demonstrated that in *Pseudomonas aeruginosa* a specific ferrodoxin NADP reductase was responsible for mobilizing iron from bacterioferritin and was able to mobilize iron faster than sodium dithionite[10]. This confirmed that flavodoxins and ferrodoxins play an important role *in vivo* in mobilizing iron from the ferritins and related proteins.

Recently, a novel iron storage system with a distinct structural architecture was discovered. Unlike previous systems, such as ferritin, where each monomer formed both a part of the shell and a ferroxidase center, in this new class the cage and ferroxidase center are split into two proteins. The shell, termed encapsulin, is ~ 24 - 31 nm in diameter and has no sequence or structural homology to ferritins[11,12]. Instead it forms from 60 to 180 copies of a homo-monomer into an icosahedral structure similar to bacterial microcompartments[11,12]. For this reason, the shell is also referred to as a nanocompartment. Encapsulins are modular shells that have been found to package a variety of different proteins, such as peroxidases, hemerythrin, and rubrerythrin, but for this study we focus on encapsulins that package a ferritin-like protein (FLP)[11,13–16]. While FLPs are homologous to ferritins, they are smaller, ~ 12 kilodaltons (kDa) for FLP versus ~18 kDa for ferritin, and assemble into a decamer as opposed to a 24-mer[11,17]. Due to this truncation, each FLP does not have a ferroxidase center, instead the FOC is formed at the interface between FLP dimers[17]. Thus, the active iron storage system comprises encapsulin with packaged FLP.

Recent work suggests there is a diversity of encapsulin-cargo complexes that can facilitate iron storage. *Myxococcus xanthus* has a 31 nm encapsulin that packages not only FLP but also 2 addition proteins[14]. Instead of forming a single large iron granule, multiple ferric iron granules are formed near the encapsulin lumen. This results in an iron storage capacity equivalent to ferritin but with significantly faster kinetics of iron oxidation or reduction. McHugh et al. also showed that encapsulin was physiologically necessary for sporulation and response to oxidative stress in *M. xanthus*[14]. A smaller encapsulin, 24 nm diameter, that packages only FLP from *R. rubrum* has been suggested to store up to 20,000 iron atoms per shell, four times the amount of iron atoms as ferritin[17]. More recently, Giessen et al. identified encapsulins with non-FLP cargo that oxidize and store iron[15]. Heterologous expression of these encapsulins in *E. coli* protected the *E. coli* from oxidative stress[15]. The encapsulins are serving as iron storage systems in all of these species; but interestingly, all of these organisms also have ferritins or bacterioferritins that are expressed. It is unclear why these organisms require multiple iron storage mechanisms or how they functionally differ.

One class of FLP-encapsulin systems that have been identified but not characterized are those from anaerobic bacteria[16]. In this study, we characterize the FLP-encapsulin system from the anaerobic organism *T. maritima.* We show that the encapsulin binds riboflavin and flavin mononucleotide on the exterior of the shell. We propose that flavins are involved in electron transfer across the shell for either oxidation or reduction of iron. Additionally, we present the first high-resolution structure of a cargo loaded encapsulin. While some results are still preliminary, they suggest the shell may be playing an active role in iron oxidation and reduction.

3-2 Results and Discussion

3.2.1 Characterization of T. maritima encapsulin as a flavoprotein

During the purification of *T. maritima* encapsulins from *E. coli* for our previous study on loading heterologous cargo, we observed that encapsulin proteins can be tracked by the presence of a yellow coloration and purified encapsulins have a orange to yellow coloration depending on the encapsulin concentration (Figure 3-1A). The coloration was not due to the presence of stored iron as it occurred even when purifying empty encapsulin shells devoid of cargo protein. Yellow coloration is typical of a flavoprotein and indicates that *T. maritima* encapsulins may be flavoproteins[18]. In order to confirm that *T. maritima* encapsulins are flavoproteins, we compared the absorbance spectrum of encapsulin to flavin mononucleotide (FMN) and riboflavin (RF), specifically in the 300 nm to 500 nm wavelength regime. Both FMN and RF have a local maxima at ~350 nm and a global maxima at ~450 nm, absorption features shared by *T. maritima* encapsulins (Figure 3-1B). This result strongly supports *T. maritima* encapsulin being a flavoprotein.

The intracellular flavin pool is composed of three major species: riboflavin (RF), flavin mononucleotide (FMN), and flavin adenine dinucleotide (FAD), and we next sought to identify the specific co-factor bound to encapsulin. These molecules cannot be distinguished by UV/Vis absorbance data, and so we performed mass spectrometry on encapsulins to identify ions corresponding to RF, FMN, and FAD. Purified empty encapsulins in water were injected on a C18 column and denatured with acetonitrile, releasing the bound flavins. The flavins were eluted with a gradient of acetonitrile and ionized by electrospray ionization. The flavins RF and FMN, were present as the major and minor peaks respectively while no peak for FAD was observed (Figure 3-1C). Due to potential differences in ionization efficiency, it is not possible to make any conclusions about the relative levels of RF and FMN bound to encapsulins. Additionally, a small lumichrome peak, a flavin photodegradation product, was present suggesting some flavins were degraded during the encapsulin purification. From these results, it was clear that encapsulins bind to RF and FMN.

While free, oxidized flavins have a well-characterized fluorescence profile with maximal excitation at ~ 450 nm and maximal emission at ~ 530 nm, purified *T. maritima* encapsulins displayed no such fluorescence (Figure 3-1D). However, substantial flavin fluorescence was observed for encapsulins denatured with 7 M guanidinium chloride (GdmCl) demonstrating that flavins are quenched when bound to encapsulins (Figure 3-1E). In other flavoproteins, fluorescence is often quenched through a π stacking interaction with tyrosine, phenylalanine, or tryptophan suggesting that the flavin binding site is near an aromatic residue. Using the fluorescence from a known amount of RF or FMN, we estimated the number of flavins bound per nanocompartment in our purified sample (in the absence of exogenous flavins). Assuming all bound flavins are RF, there are 9.5 ± 0.1 riboflavins bound per nanocompartment while assuming all flavins are FMN there are 8.8 ± 0.1 FMNs bound per nanocompartment. From the symmetry of

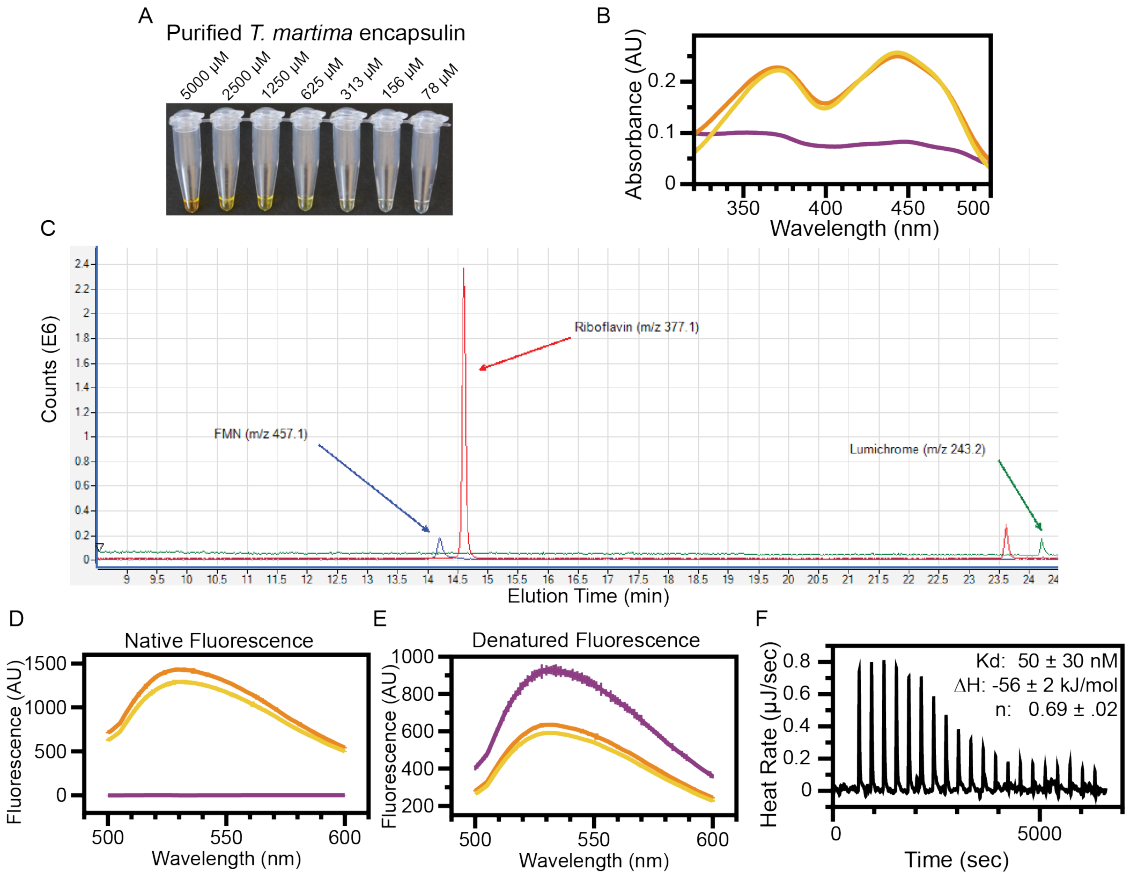

Figure 3-1: *T. maritima* encapsulin is a flavoprotein (A.) Purified *T. maritima* encapsulins at different concentrations. (B.) Absorbance spectra of *T. maritima* encapsulin (purple), FMN (orange), and RF (yellow) (C.) Extracted ion counts for RF (red), FMN (blue), and lumichrome (green) from mass spectroscopy on *T. maritima* encapsulin. (D.) Fluorescence emission after excitation at 450 nm for *T. maritima* encapsulin (purple), FMN (orange), and RF (yellow). Mean ± SD plotted. (E.) Fluorescence emission under denaturing conditions after excitation at 450 nm for *T. maritima* encapsulin (purple), FMN (orange), and RF (yellow). Mean ± SD plotted. (F.) Heat Rate for 20 injections of 75 µM FMN into 5 µM flavin-free encapsulin. Top right is the mean ± SD dissociation constant, enthalpy, and Hill coefficient calculated from triplicate runs.

encapsulin there are three different potential classes of flavin binding sites: at every protomer (60 per encapsulin), at sites of 3-fold symmetry (20 per encapsulin), or at the sites of 5-fold symmetry (12 per encapsulin)[11]. The purified encapsulin is thus not completely saturated at all binding sites, which is not unexpected given the observed propensity of flavins to degrade during purification.

To determine the number of flavin binding sites per encapsulin, we employed isothermal titration calorimetry (ITC). This required deflavination of encapsulins by unfolding and refolding of purified encapsulins. Even after this protocol, a minor amount of flavin absorbance signal indicated there was still residual binding. The affinity of flavin-free encapsulins for FMN, chosen for its higher solubility in water, was 50 nM ± 30 nM, a value typical of flavoproteins (Figure 3-1F)[19]. We also measured a Hill coefficient of .69 ± .02 suggesting there are 42 ± 1 flavin binding sites per encapsulin. Since the calculated number of

binding sites doesn't match any of the theoretical models, we are skeptical of the result and the quality of the flavin-free protein. Current efforts are underway to prepare fully deflavinated encapsulins. However, again, these results strongly support *T. maritima* encapsulin being a flavoprotein.

None of the previous studied nanocompartments were identified as flavoproteins. One possibility is that flavin binding is an artifact from heterologous expression of *T. maritima* encapsulins in *E. coli*; however, encapsulins from *M. xanthus*, *R. rubrum*, *Brevibacterium linens* M18, and *Rhodococcus erythropolis* N771 have been heterologously expressed in *E. coli* and none were reported as flavoproteins or having a yellow coloration[11,14,17,20]. Additionally for crystallization trials, Sutter et al. purified encapsulins from *T. maritima* cell pellets. While the encapsulins were not reported as flavoproteins, crystals of encapsulin from *T. maritima* have a distinct yellow coloration suggesting they bind flavins even in the native organism[11]. It is likely then that only a subset of encapsulins are flavoproteins.

3.2.2 Functional significance of bound flavins

In order to elucidate the role of flavin for encapsulins, we began by identifying the flavin binding site with cryo-electron microscopy (cryo-EM) single particle analysis. In the previously published *T. maritima* encapsulin crystal structure, there was no flavin electron density present despite evidence flavins were bound to encapsulin in the crystal[11]. It may have been that in the 3.5 Å resolution structure, it was not possible to distinguish the flavin density from neighboring density. Alternatively, collection and analysis of the x-ray scattering data relied heavily on encapsulin symmetry. If flavins were not bound in the majority of sites in the crystal then flavin density would be averaged out from the final electron density map. To avoid any issues with flavin occupancy, encapsulins were doped with five-fold molar FMN prior to cryo-EM. The excess FMN did not appear to affect assembly or structure as encapsulins formed hollow shells, ~24 nm in diameter on the cryo-EM grids (Figure 3-2A). Single particle analysis on 20,000 encapsulin particles yielded a 4 Å structure and the previous encapsulin crystal structure docked well into the cryo-EM structure (Figure 3-2B).

Despite the lower resolution of 4 Å, the cryo-EM single particle analysis, new insights into encapsulin structure were still observed. Previously, a ring of histidines on the cytosolic side of encapsulin was implicated in coordinating metal ions. In the cryo-EM reconstruction, it is clear from the density that the histidines are coordinating a molecule or ion (Figure 3-2C). Additionally, we noticed aberrant density not making interactions with any of the surrounding residues at a surface solvent-exposed tryptophan, W90 (Figure 3-2D). While FMN could not be unambiguously docked next to W90, the aberrant density and the involvement of an aromatic residue strongly supported this location as the flavin binding site.

To confirm the flavin binding site, encapsulins bearing the W90A mutation were expressed, purified, and characterized. This mutation did not appear to impact encapsulin stability or assembly as W90A encapsulins were purified using the same protocol as wild-type (WT) encapsulins and eluted from a size exclusion column at the same volume (Figure 3-2E). Examination by eye of the

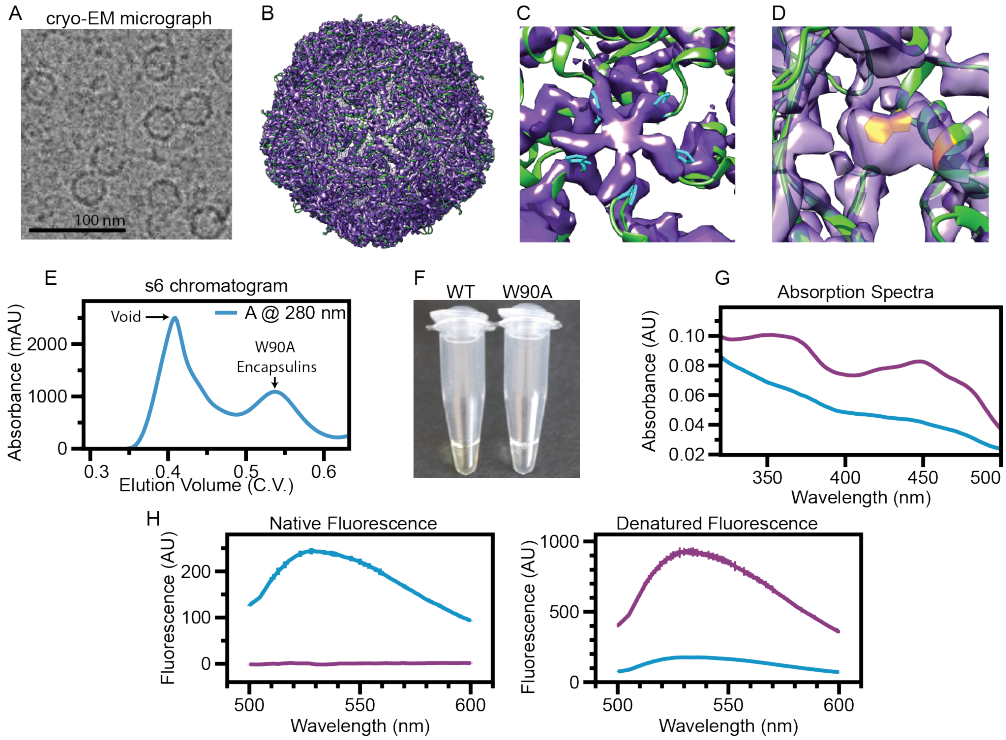

Figure 3-2: *T. maritima* encapsulins bind flavins adjacent to a surface tryptophan. (A.) Raw cryo-EM micrographs of empty encapsulins. (B.) *T. maritima* encapsulin crystal structure (green) (PDB: 3DKT) docked into the single particle analysis reconstruction of encapsulin (purple) (C.) Histidines (cyan) coordinating an unknown molecular at the encapsulin pentameric pore. (D.) Aberrant density around tryptophan 90 (orange). (E.) Size exclusion chromatogram from the purification of empty W90A encapsulins. (F.) Image of 250 μM WT encapsulin and 250 μM W90A encapsulin. (G.) Absorbance spectra of WT encapsulin (purple) and W90A encapsulin (blue). (H.) Left - Fluorescence emission after excitation at 450 nm for WT encapsulin (purple) and W90A encapsulin (blue). Right - Fluorescence emission under denaturing conditions after excitation at 450 nm for WT encapsulin (purple) and W90A encapsulin Mean ± SD plotted.

color of purified protein confirmed that flavin-binding is reduced with W90A encapsulins relative to WT, supporting the hypothesis of W90 forming the flavin binding pocket (Figure 3-2f). Strikingly, at high concentrations of W90A, the purified protein had a yellowish color indicating W90A encapsulins may still have some affinity for flavins. The absorption spectra of W90A encapsulins revealed flavins were still co-purifying but with significantly less abundance than for WT encapsulins (Figure 3-2G). Therefore, while W90 is partly responsible for forming the flavin binding pocket, its mutation to alanine is not sufficient to completely disrupt the binding pocket. It is unlikely that flavins are binding at a secondary location as W90A encapsulins do not show the quenched fluorescence seen with WT encapsulins (Figure 3-2H). This suggests the flavins are still binding in the same pocket yet without the tryptophan present the fluorescence cannot be quenched. Together, this data confirms our hypothesis that W90 forms a binding pocket for flavins in encapsulins.

Having identified the flavin binding site, we sought to determine the prevalence of flavin binding for other encapsulins. Previous studies, on

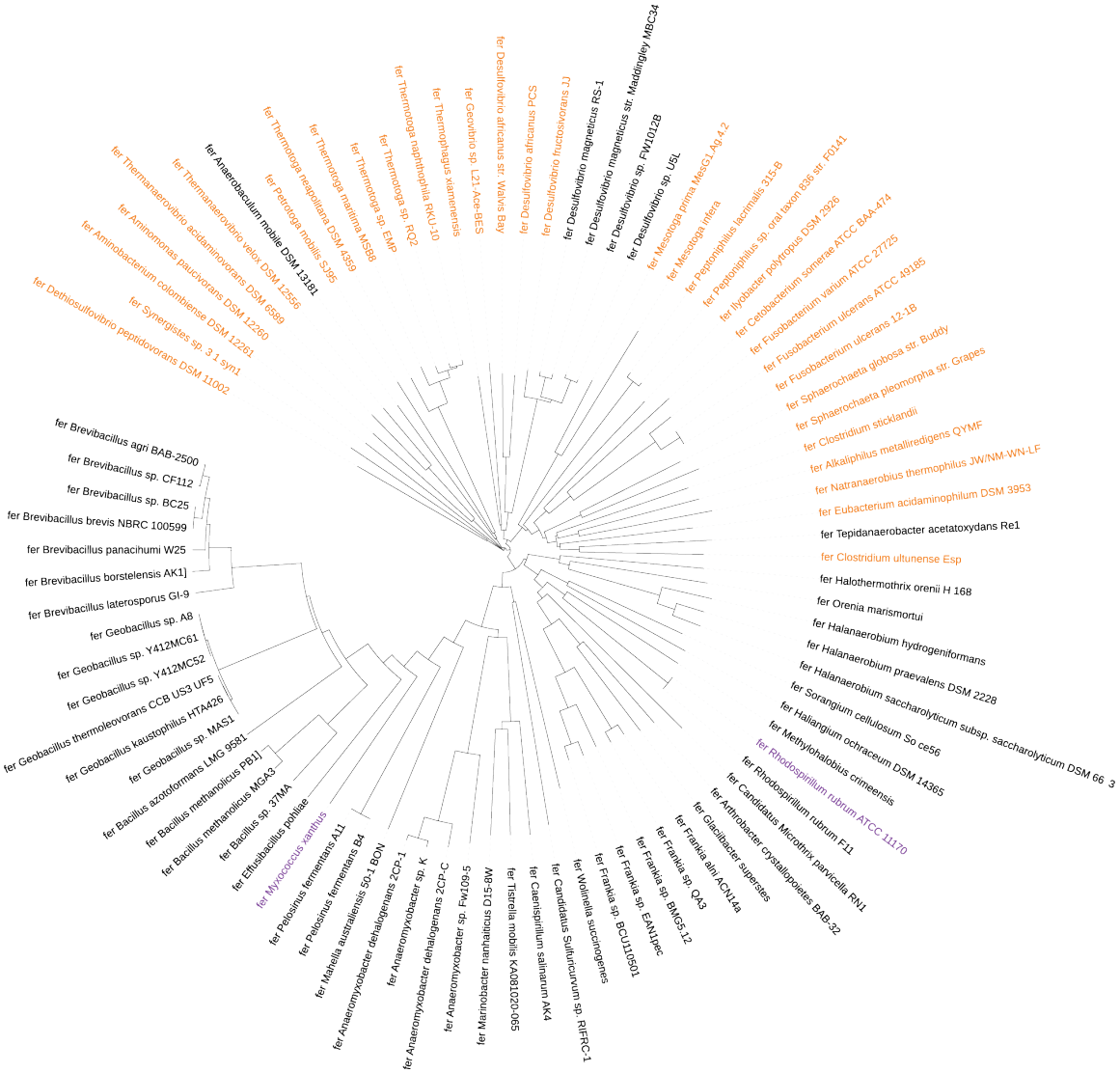

Figure 3-3: Tree of FLP packaging encapsulins. Encapsulins with a tryptophan at position 90 are colored in orange. Previously studied encapsulins are colored in purple.

encapsulins from other organisms, had not found flavins bound to the encapsulin shells[11,13–15]. Comparison of those encapsulin sequences to the *T. maritima* encapsulin sequence revealed that none of them have a tryptophan at position 90. Instead, *R. rubrum* has an alanine, *M. xanthus* has an arginine, *B. linens* has a glutamine. Therefore the presence of a tryptophan at this position may be a proxy for flavin binding. Potential flavin binding encapsulins were identified by generating a multiple sequence alignment of all known encapsulin and searching for a tryptophan at the equivalent of *T. maritima* encapsulin position 90. Interestingly, all potential flavin binding encapsulins package a ferritin-like protein (FLP) suggesting the flavin may somehow be involved in the functioning of the FLP[16]. Amongst FLP packaging encapsulins, ~35% are potential flavin-binding proteins. A tree of FLP packaging encapsulins generated from the multiple sequence alignment shows that the flavin binding encapsulins form the majority

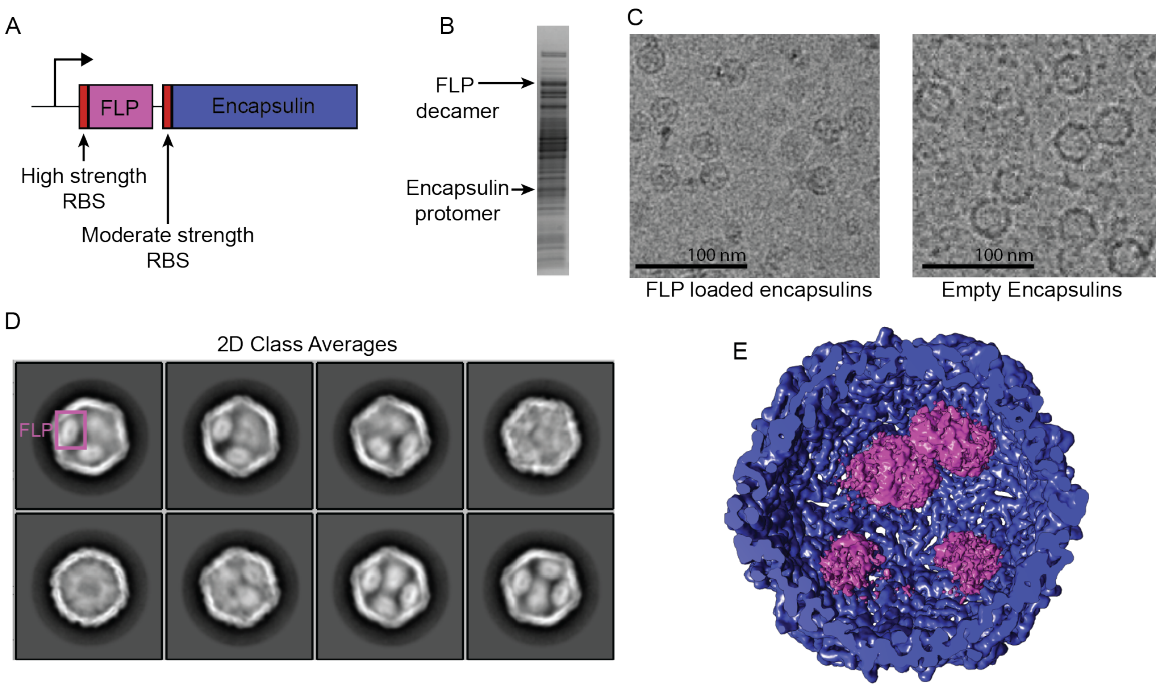

Figure 3-4: Cryo-EM of a FLP loaded encapsulin (A.) Expression system for purification of highly loaded encapsulin. (B.) SDS-PAGE of the lysate with bands for the FLP decamer and encapsulin protomer indicated. (C.) Left – Raw cryo-EM micrograph of FLP loaded encapsulins. Right – Raw cryo-EM micrograph of empty encapsulins. (D.) 2D class averages from single particle analysis on FLP loaded encapsulins. A FLP decamer is shown in the magenta box. (E.) Preliminary 3D reconstruction of FLP loaded encapsulin. FLPs shown in magenta and encapsulin shown in blue.

of an outgroup from the other sequences (Figure 3-3). It is unclear why only a subset of FLP packaging encapsulins have a flavin binding site. Most flavoproteins are involved in electron transfer reactions, and so we hypothesized that the flavin is connected to the redox role of FLP either by aiding in ferrous iron oxidation and mineralization or by reducing ferric iron and promoting iron mobilization[19].

Before investigating the interplay between flavins and enzymatic activity of FLP, we wanted to understand the structure of FLP loaded encapsulin. Our biochemical analysis identified the flavin binding site; however, to date that are no structures where FLP can be resolved inside the encapsulin. It has been suggested that a FLP decamer binds to each encapsulin pentamer aligning the FLP active site with the histidine-lined encapsulin pore and yielding encapsulins loaded with 120 FLP[17]. In order to address this question, we purified the FLP-loaded encapsulin and used cryo-EM single particle analysis to determine its structure. One potential issue was the heterogeneity of FLP loading. Potentially zero through 12 FLP decamers could be loaded per encapsulin, which would complicate the analysis of particles from cryo-EM. In order to minimize the loading heterogeneity, a construct was generated where FLP was more strongly expressed than encapsulin creating an intracellular environment where encapsulins should be saturated with FLPs during the assembly process (Figure 3-4A). This was confirmed by SDS-PAGE analysis where we saw a stronger

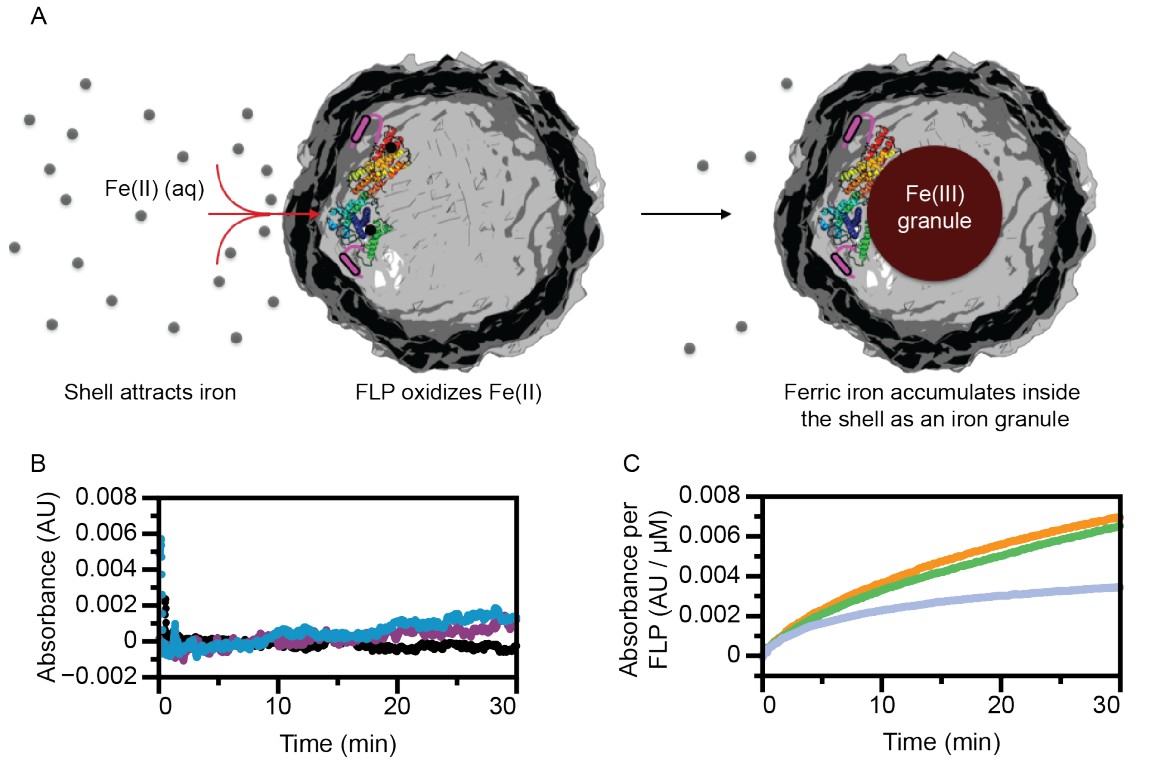

Figure 3-5: FLP as an iron storage system. (B.) Ferroxidase activity of empty WT encapsulin (purple), W90A encapsulin (blue), and no protein (black). (C.) Ferroxidase activity normalized to the FLP concentration for FLP loaded WT encapsulin (orange), FLP loaded W90A encapsulin (green), and unencapsulated FLP (gray).

band for the FLP decamer than for the encapsulin protomer (Figure 3-4B). Interestingly even under denaturing conditions, FLP ran as a decamer demonstrating a robust assembly much like the nanocompartment itself.

Purified FLP loaded encapsulins were spotted on grids and images of 1000 micrographs with 25,000 total particles were collected. Even on the raw micrographs, FLP was visible inside of encapsulin (Figure 3-4C). In micrographs from empty encapsulin, the interior is hollow while in the micrographs from FLP loaded encapsulin density can be observed close to the encapsulin lumen. The 25,000 particles were used to generate 2D class averages that showed most encapsulins loaded 3 or 4 FLP decamers (Figure 3-4D). In two cases, no FLP decamers were apparent inside the shells; it was unclear why no FLPs were loaded but these 2D classes averages represented a small fraction of the total particles. From the 2D class averages, an initial 3D reconstruction was generated that had 5 FLP decamers loaded inside. The 3D reconstruction is still a preliminary structure and the number of FLP decamers and their exact location will change as the structure is refined. However, it is clear from this data that encapsulins package substantially less than 12 FLP decamers. It remains to be seen what further insights about the degree of packaging and the interaction between the shell and FLPs can be gleaned from the final 3D reconstruction.

Encapsulins packaging FLPs are known to oxidize ferrous iron and store the ferric iron as an insoluble mineralized core (Figure 3-5A) [11,14,17]. Oxidation of iron involves transfer of an electron to an acceptor molecule, often molecular

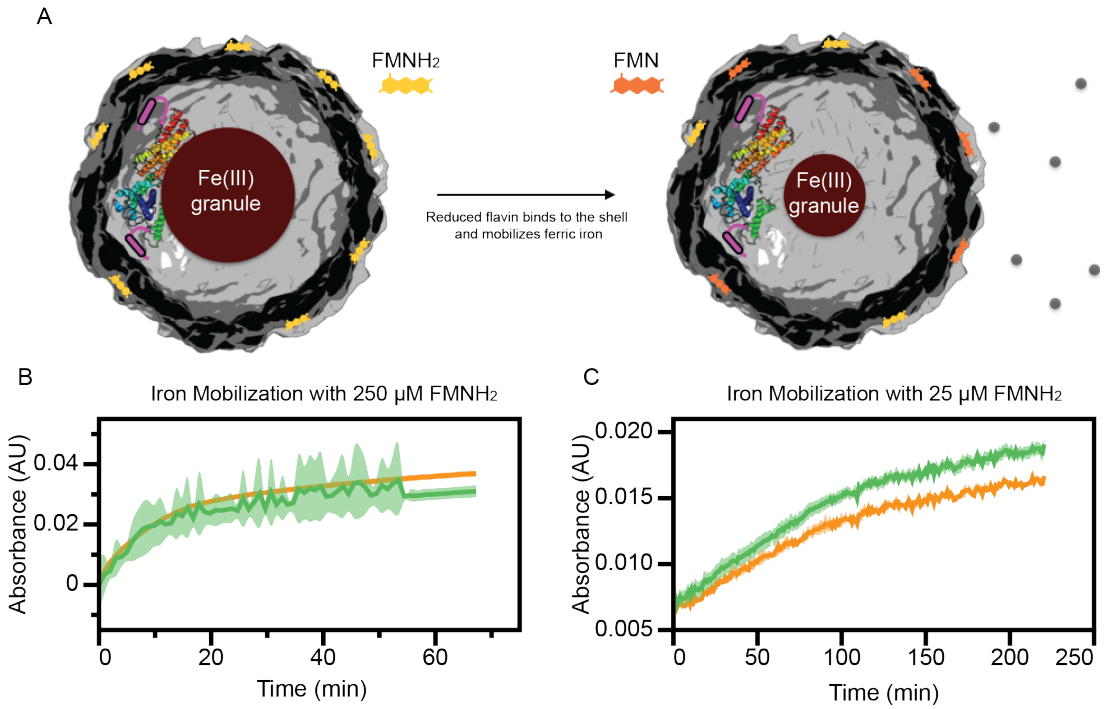

Figure 3-6: Proposed iron mobilization from *T. maritima* encapsulins (A.) Model for mobilization of iron from encapsulin by reduced flavins. (B.) Kinetics of iron mobilization by 250 μM $FMNH_2$ from FLP loaded WT encapsulin (orange) and FLP loaded W90A encapsulin (green). Mean ± SD envelope plotted. (C.) Kinetics of iron mobilization by 25 μM $FMNH_2$ from FLP loaded WT encapsulin (orange) and FLP loaded W90A encapsulin (green). Mean ± SD envelope plotted.

oxygen; however, a flavin bound to the shell could serve as the electron acceptor. The W90A encapsulins, with their reduced flavin affinity and fewer bound flavins may then have slower kinetics of iron oxidation. We tested this hypothesis by aerobically measuring the kinetics of iron oxidation for empty WT and W90A encapsulins, FLP loaded WT and W90A encapsulins, and unpackaged FLPs. Both empty shells displayed no iron oxidation (Figure 3-5B). Additionally, there was no discernable difference in kinetics for FLP loaded WT and W90A encapsulins but normalization to the amount of FLP revealed that encapsulation of FLP increases the rate of iron oxidation relative to unencapsulated FLP (Figure 3-5C). This encapsulation driven enhancement of FLP has been observed previously; yet, the cause for the increased kinetics is unknown[17]. One possibility is that the histidines on the encapsulin pentamer coordinate free iron and increase the local concentration near the FLPs. It was perhaps unsurprising that there is no difference between WT and W90A encapsulins, given that under aerobic conditions molecular oxygen is readily available and a stronger oxidizing agent than oxidized flavins. So under aerobic conditions, molecular oxygen should dominate as the iron-oxidizing agent. We are currently performing more experiments, both aerobic and anaerobic, to determine if any difference exist between the WT and W90A shell constructs.

The other possible role for the flavin co-factor is in iron mobilization. Essentially the inverse of iron acquisition, low reduction potential electrons are passed to the FLP active site and used to reduce ferric iron into soluble ferrous

iron. While intracellular iron mobilization is still not well understood, previous studies have shown that reduced flavins can mobilize iron under anaerobic conditions[21–23]. The flavins bound to encapsulin could serve as an electron donor for iron reduction and mobilization (Figure 3-6A). Reduced flavins have a short half-life in air, < 100 ms, so all experiments were performed anaerobically. In order to generate reduced flavins, a protocol from a previous study was adapted[21]. FMN was anaerobically incubated with excess NADH resulting in the formation of $FMNH_2$ and NAD and the remaining NADH. NADH could reduce ferric iron but on the timescale of hours compared to iron reduction on the timescale of minutes we observed for $FMNH_2$. We tested the ability of $FMNH_2$ to mobilize iron from WT and W90A encapsulins (Figure 3-6B,C). There was no difference in the rates of iron mobilization for either of the two $FMNH_2$ concentrations tested. One possibility for the lack of difference between WT and W90A encapsulins is that both shells still have flavins bound. While there are fewer flavins bound to W90A encapsulins, if the low reduction potential electrons from $FMNH_2$ are being passed through the bound flavins at a high flux then no difference would be observed. It may be necessary to test iron mobilization on encapsulins that have undergone complete deflavination. It is still too soon to conclude that the flavins bound to the encapsulin shell are uninvolved in iron mobilization. It is likely that slightly more nuanced experiments will be necessary to demonstrate the difference in iron mobilization from WT or W90A encapsulins.

For the first time, we solved a high-resolution structure of FLP loaded encapsulin and we have demonstrated that the T. maritima encapsulin is a flavoprotein. While still in a preliminary state, the structure has been useful for determining the number of FLP loaded per encapsulin. While previous work speculated that 12 FLP decamers are loaded per encapsulin, we have shown that only between 3 to 4 FLP decamers are loaded[17]. There does not appear to be any spatial constraints disallowing the loading of more FLPs, so it is unclear why more FLP decamers are not loaded. One possibility is that the diminished loading is an artifact of the heterologous expression system. *T. maritima* may control the loading and assembly of encapsulins through regulation, transcriptional, translational, or some accessory protein, that is lacking in the *E. coli* expression system. Another possibility is that it is inefficient to fully load encapsulins. In a larger 31 nm iron oxidizing encapsulin from *Myxococcus xanthu*s, multiple iron granules are formed in each encapsulin[14]. However, in the smaller FLP packaging encapsulins, it is believed that only one iron granule is formed per encapsulin[15,17]. In this case, increasing the number of FLPs would only serve to increase the kinetics of iron oxidation and mobilization, a feature *T. maritima* may not need for the physiological concentrations of iron it experiences. Further structural studies on other FLP packaging encapsulins may reveal that the degree of cargo loading is titrated to meet the needs of each organism.

T. maritima encapsulins bind riboflavin and flavin mononucleotide. Flavoproteins usually serve as electron carriers and we hypothesize that the *T. maritima* encapsulin may be involved with the functioning of its packaged enzyme cargo, FLP. This would be the first instance of a nanocompartment being directly involved with the functioning of a packaged enzyme. There are many

open questions in such a system such as how are the electrons shuttled from reduced flavins to FLPs and how are how are the oxidized flavins exchanged for reduced flavins to continue the mobilization of iron.

3-3 Methods

3.3.1 Molecular biology and cell culture

The construct for WT encapsulin was made in a previous study [24]. FLP was ordered codon optimized for *E. coli* as from IDT as a gBlock. FLP was cloned into a pET14 vector with encapsulin using a golden-gate reaction[25]. For expression of FLP by itself, the gBlock was cloned into a pET14 vector with a C-terminal hexa-histidine tag using golden gate reactions. The W90A mutation was introduced to encapsulin alone and the FLP-encapsulin construct using a golden-gate reaction. All constructs were transformed into DH5α cells and sequence verified.

All constructs for expression and purification were transformed into C43 (DE3) cells and grown in Terrific Broth (TB) media at 37° C and 250 rpm until an optical density (600 nm) of 0.3 was achieved. The cells were cooled to 18° C and induced with 500 μM isopropyl β-D-1-thiogalactopyranoside and allowed to grow overnight. Cells were harvested by centrifugation and frozen at -20° C until needed.

3.3.2 Protein Purification

The purification for all constructs except FLP was adapted from a previous study[24]. The cell pellet was thawed and resuspended in buffer A (50 mM Tris-HCl, pH 7.4, 150 mM NaCl, and 5 mM 2-mercaptoethanol). RNase A (100 μg), DNase (10 U), $MgCl_2$ (1 mM final concentration), and lysozyme (2 mg) were added and the resuspended cells were incubated at room temperature (RT) for 15 minutes. Cells were lysed by sonication and the lysate was clarified by centrifugation in a JA-20 rotor at 22500 g for 20 minutes. Following clarification, the supernatant was collected and placed in an 80° C incubator for 3 hours. The supernatant was centrifuged in a JA-20 rotor at 25000 g for 25 minutes to pellet aggregated protein. Following the spin, the supernatant was collected and ammonium sulfate was added to a final concentration of 35%. After a 30 minute RT incubation, precipitated protein was removed by centrifugation in a JS-5.3 rotor at 6800 g for 20 minutes. The supernatant was collected and ammonium sulfate was again added to reach a final concentration of 75%. The supernatant was incubated for 30 minutes at RT and precipitated proteins were pelleted by centrifugation in a JS-5.3 rotor at 6800 g for 25 minutes. The supernatant was discarded and the pellet was resuspended with buffer A. The resuspended protein was dialyzed into buffer B (20 mM Tris-HCl, pH 7.4, 150 mM NaCl and 5 mM 2-mercaptoethanol) to remove any remaining ammonium sulfate. Dialyzed protein was centrifuged in a type 50.2 Ti rotor at 125,000 g for 3 hours to pellet the nanocompartments. The pellet was resuspended in buffer A and loaded onto a Superose 6 10/300 GL column equilibrated with buffer C (50 mM Tris-HCl, pH 7.4 and 150 mM NaCl). Unless otherwise indicated, all Superose 6 columns were run at 0.5 mL / min. Fractions containing assembled encapsulins were collected and concentrated with a Vivaspin 6 (100 kDa molecular weight cutoff (MWCO)) concentrator. Purified protein was supplemented with glycerol (10% final concentration), flash frozen in ℓN_2, and stored at -80° C until needed.

FLP was purified by nickel immobilized metal ion affinity chromatography

(IMAC). Cell pellets were resuspended in 50 mM Tris-HCl, pH 7.4, 150 mM NaCl, and 10 mM imidazole. RNase A (100 µg), DNase (10 U), $MgCl_2$ (1 mM final concentration), and lysozyme (2 mg) were added and the resuspended cells were incubated at room temperature (RT) for 15 minutes. Cells were lysed by sonication and the lysate was clarified by centrifugation in a JA-20 rotor at 40000 g for 25 minutes. Cleared lysate was incubated with Ni-nitrilotriacetic acid (NTA) resin from Qiagen for 1 hour at RT. The resin was washed with 20 resin volumes of 50 mM Tris-HCl, pH 7.4, 500 mM NaCl, and 10 mM imidazole followed by 20 resin volumes of 50 mM Tris-HCl, pH 7.4, 150 mM NaCl, and 30 mM imidazole. Bound FLP was eluted with 3 resin volumes of 50 mM Tris-HCl, pH 7.4, 150 mM NaCl, and 350 mM imidazole. The Ni-NTA elute was dialyzed into buffer B, concentrated with a Vivaspin 6 (10 kDa MWCO), and loaded onto a Superose 6 column equilibrated with buffer C. Fractions containing FLP were collected and concentrated with a Vivaspin 6 (10 kDa molecular weight cutoff (MWCO)) concentrator. Purified protein was supplemented with glycerol (10% final concentration), flash frozen in ℓN_2, and stored at -80° C until needed.

3.3.3 Absorbance Spectrum Measurements

All absorption spectra were acquired in a Cary 50 Bio UV-Visible spectrophotometer. Each purified protein or metabolite was diluted to 20 µM with buffer C. Samples were placed in a Fisherbrand quartz cuvette and the absorbance was collected from 600 nm to 200 nm on the slowest scan speed.

3.3.4 Fluorescence Emission Spectrum

Fluorescence emission spectrums were collected on a Tecan Infinite M1000 PRO spectrophotometer. For native fluorescence, each protein was diluted to 20 µM and each metabolite was diluted to 2 µM with buffer C. Samples were excited at 450 nm with a 5 nm bandwidth. Emission data was collected from 500 nm to 600 nm in 1 nm steps with a 5 nm bandwidth. All samples were measured in triplicate and plotted with Datagraph as the mean value ± standard deviation. For denatured fluorescence, each protein was diluted to 20 µM and each metabolite was diluted to 2 µM in buffer C plus 7 M guanidine hydrochloride (GuHCl). Measurements parameters and plotting were the same as for the native fluorescence samples.

3.3.5 Mass Spectroscopy

Purified encapsulins were analyzed by LC-MS. The sample was buffer exchanged into water and then injected onto a Waters XBridge C18 column with an Agilent Infinity 1260 HPLC. The sample was eluted with a gradient of water to acetonitrile both with 0.1% formic acid. Mass analysis was performed with an Agilent 6500 Q-TOF operated in positive ion mode. The organic solvent denatures the encapsulin particles on the column thus releases the bound flavin species.

3.3.6 Isothermal Titration Calorimetry

Flavin-free encapsulins were generated as previously described in order

to measure the affinity of flavin mononucleotide (FMN) for encapsulins. Briefly, concentrated purified encapsulins were diluted with GuHCl and dithiothreitol (DTT) (final concentration 7M GuHCl and 10 mM DTT). Denatured encapsulins were loaded onto a Superose 6 column equilibrated with 20 mM phosphate, pH 7.4, 50 mM NaCl, and 7 M GuHCl. Fractions containing denatured encapsulins were refolded by ten-fold dilution into buffer D (20 mM phosphate, pH 7.4, 50 mM NaCl, 10 mM DTT). Refolded encapsulins were concentrated and loaded onto a Superose 6 column equilibrated with buffer E (20 mM phosphate, pH 7.4, 50 mM NaCl). Fractions containing properly assembled encapsulins were collected and concentrated.

Isothermal titration calorimetry (ITC) was performed on a Texas Instruments Nano ITC 2g. Flavin-free encapsulins were diluted to 5 μM in buffer D and 75 μM FMN was prepared by dissolving solid FMN into buffer E (FMN extinction coefficient: 12,020 M^{-1} cm^{-1}). The ITC run was set up as follows: 600 seconds initial baseline, 20 injections of FMN (each 4.92 μL) with 300 seconds of equilibration between each injection, and finally ends with 600 seconds final baseline. The binding experiment was repeated in triplicate.

3.3.7 Cryo-electron Microscopy

For structural determination of the flavin binding site by cryo-EM, flavin-free encapsulins were prepared as described above. Flavin-free encapsulins (60 μM) were incubated with FMN (300 μM) prior to spotting on a grid. Samples were frozen and imaged by Ben LaFrance.

Purified encapsulins with FLP were further enriched with a Q sepharose HP column. Purified protein was dialyzed into 20 mM Tris-HCl, pH 8.5 and 50 mM NaCl and loaded onto the column. FLP loaded nanocompartments were eluted with 20 mM Tris-HCl, pH 8.5 and 1000 mM NaCl, concentrated and buffer exchanged into buffer C. Prior to spotting on a grid, loaded encapsulins were incubated with a 5-fold molar excess of FMN. Samples were frozen and imaged by Ben LaFrance.

3.3.8 Multiple Sequence Alignment

Encapsulins sequences were downloaded from National Center for Biotechnology Information in fasta format. Multiple sequence alignment was generated with ClustalX and the images of the tree were generated with the interactive Tree of Life website.

3.3.9 Ferroxidase Assays

Purified protein was diluted in 50 mM Acetate, pH 6.0 and 300 mM NaCl buffer. FLP loaded WT encapsulin and FLP loaded W90A encapsulin were diluted to 20 μM of FLP. Ammonium ferrous sulfate (2.5 mM) was prepared anaerobically in 0.1% HCl. The ferrous iron was diluted ten-fold into the protein, mixed, and then immediately loaded into a Cary 50 Bio UV-Visible spectrophotometer. Absorbance at 310 nm was collected every second for 30 minutes. To normalize between all the samples, the initial 310 nm absorbance value was set as the 0 absorbance value. Data was plotted using Datagraph.

3.3.10 Iron Release Assays

Reduced FMN, $FMNH_2$, was generated by mixing FMN and NADH anaerobically and incubating overnight. Purified protein was diluted in buffer C and made anaerobic by purging with a H_2:N_2 gas mixture. Anaerobic 1,10 phenanthroline (750 μM final concentration) and $FMNH_2$ (25 μM or 250 μM) were added to the sample. Iron release was monitored by the gain in absorbance at 510 nm in a Tecan M200 plate reader.

3-4 References

(1) Lemire, J. A., Harrison, J. J., and Turner, R. J. (2013) Antimicrobial activity of metals: mechanisms, molecular targets and applications. *Nat. Rev. Microbiol. 11*, 371–384.
(2) Andrews, S. C., Robinson, A. K., and Rodríguez-Quiñones, F. (2003) Bacterial iron homeostasis. *FEMS Microbiol. Rev. 27*, 215–237.
(3) Zeth, K., Hoiczyk, E., and Okuda, M. (2016) Ferroxidase-Mediated Iron Oxide Biomineralization: Novel Pathways to Multifunctional Nanoparticles. *Trends Biochem. Sci. 41*, 190–203.
(4) Massover, W. H. (1993) Ultrastructure of ferritin and apoferritin: A review. *Micron 24*, 389–437.
(5) Honarmand Ebrahimi, K., Hagedoorn, P.-L., and Hagen, W. R. (2015) Unity in the biochemistry of the iron-storage proteins ferritin and bacterioferritin. *Chem. Rev. 115*, 295–326.
(6) Carrondo, M. A. (2003) Ferritins, iron uptake and storage from the bacterioferritin viewpoint. *EMBO J. 22*, 1959–1968.
(7) Calhoun, L. N., and Kwon, Y. M. (2011) Structure, function and regulation of the DNA-binding protein Dps and its role in acid and oxidative stress resistance in Escherichia coli: a review. *J. Appl. Microbiol. 110*, 375–386.
(8) Harrison, P. M. (1986) The structure and function of ferritin. *Biochem. Educ. 14*, 154–162.
(9) Topham, R., Goger, M., Pearce, K., and Schultz, P. (1989) The mobilization of ferritin iron by liver cytosol. A comparison of xanthine and NADH as reducing substrates. *Biochem. J 261*, 137–143.
(10) Weeratunga, S. K., Gee, C. E., Lovell, S., Zeng, Y., Woodin, C. L., and Rivera, M. (2009) Binding of Pseudomonas aeruginosa apobacterioferritin-associated ferredoxin to bacterioferritin B promotes heme mediation of electron delivery and mobilization of core mineral iron. *Biochemistry 48*, 7420–7431.
(11) Sutter, M., Boehringer, D., Gutmann, S., Günther, S., Prangishvili, D., Loessner, M. J., Stetter, K. O., Weber-Ban, E., and Ban, N. (2008) Structural basis of enzyme encapsulation into a bacterial nanocompartment. *Nat. Struct. Mol. Biol. 15*, 939–947.
(12) Akita, F., Chong, K. T., Tanaka, H., Yamashita, E., Miyazaki, N., Nakaishi, Y., Suzuki, M., Namba, K., Ono, Y., Tsukihara, T., and Nakagawa, A. (2007) The Crystal Structure of a Virus-like Particle from the Hyperthermophilic Archaeon Pyrococcus furiosus Provides Insight into the Evolution of Viruses. *J. Mol. Biol. 368*, 1469–1483.
(13) Ahmad, M., Roberts, J. N., Hardiman, E. M., Singh, R., Eltis, L. D., and Bugg, T. D. H. (2011) Identification of DypB from Rhodococcus jostii RHA1 as a lignin peroxidase. *Biochemistry 50*, 5096–5107.
(14) McHugh, C. A., Fontana, J., Nemecek, D., Cheng, N., Aksyuk, A. A., Heymann, J. B., Winkler, D. C., Lam, A. S., Wall, J. S., Steven, A. C., and Hoiczyk, E. (2014) A virus capsid-like nanocompartment that stores iron and protects bacteria from oxidative stress. *EMBO J. 33*, 1896–1911.

(15) Giessen, T. W., and Silver, P. A. (2017) Widespread distribution of encapsulin nanocompartments reveals functional diversity. *Nat Microbiol 2*, 17029.
(16) Radford, D. (2015) Understanding the Encapsulins: Prediction and Characterization of Phage Capsid-like Nanocompartments in Prokaryotes. University of Toronto (Canada).
(17) He, D., Hughes, S., Vanden-Hehir, S., Georgiev, A., Altenbach, K., Tarrant, E., Mackay, C. L., Waldron, K. J., Clarke, D. J., and Marles-Wright, J. (2016) Structural characterization of encapsulated ferritin provides insight into iron storage in bacterial nanocompartments. *Elife 5*.
(18) Williams, R. E., and Bruce, N. C. (2002) “New uses for an old enzyme”--the old yellow enzyme family of flavoenzymes. *Microbiology 148*, 1607–1614.
(19) Mayhew, S. G., and Ludwig, M. L. (1975) 2 Flavodoxins and Electron-Transferring Flavoproteins. *The Enzymes 12*, 57–118.
(20) Tamura, A., Fukutani, Y., Takami, T., Fujii, M., Nakaguchi, Y., Murakami, Y., Noguchi, K., Yohda, M., and Odaka, M. (2015) Packaging guest proteins into the encapsulin nanocompartment from Rhodococcus erythropolis N771. *Biotechnol. Bioeng. 112*, 13–20.
(21) Melman, G., Bou-Abdallah, F., Vane, E., Maura, P., Arosio, P., and Melman, A. (2013) Iron release from ferritin by flavin nucleotides. *Biochim. Biophys. Acta 1830*, 4669–4674.
(22) Coves, J., and Fontecave, M. (1993) Reduction and mobilization of iron by a NAD(P)H: flavin oxidoreductase from Escherichia coli. *Eur. J. Biochem. 211*, 635–641.
(23) Schröder, I., Johnson, E., and De Vries, S. (2003) Microbial ferric iron reductases. *FEMS Microbiol. Rev. 27*, 427–447.
(24) Cassidy-Amstutz, C., Oltrogge, L., Going, C. C., Lee, A., Teng, P., Quintanilla, D., East-Seletsky, A., Williams, E. R., and Savage, D. F. (2016) Identification of a Minimal Peptide Tag for in Vivo and in Vitro Loading of Encapsulin. *Biochemistry 55*, 3461–3468.
(25) Engler, C., Gruetzner, R., Kandzia, R., and Marillonnet, S. (2009) Golden gate shuffling: a one-pot DNA shuffling method based on type IIs restriction enzymes. *PLoS One 4*, e5553.

Chapter 4 Nanocompartments may ameliorate host oxidative stress

4-1 Introduction

Tuberculosis, caused by the etiological agent *Mycobacterium tuberculosis*, is a global health epidemic. In 2014, WHO reported > 9 million new infections and ~1.5 million deaths[1]. Administration of antibiotics is the standard course of treatment; however, recently the prevalence of drug-resistant tuberculosis has been increasing. This highlights the need for new treatments targeting novel mechanisms in *M. tuberculosis*.

One potential target is CFP29, originally purified from the short-term culture filtrate of *M. tuberculosis* strain H37Rv[2]. CFP29 has been found in the exosomes from *M. tuberculosis* infected cells and can trigger a strong reaction from the mouse adaptive immune system[2,3]. CFP29 is a protein nanocompartment, as termed an encapsulin, that forms an icosahedral structure ~24 nm in diameter[4,5]. Similar to the majority of encapsulins, *cfp29* is found in an operon with its upstream cargo gene, dye-decolorizing peroxidase (*dyP)*[5]. DyPs are a newly discovered class of bacterial and fungal haem containing peroxidases, which have been shown to be responsible for lignin degradation in *R. jostii* RHA1 but whose function is unknown in *M. tuberculosis*[6,7]. Contreras *et al.* identified two other cargo proteins of CLP29: bacterioferritin B (BfrB) and dihydroneopterin aldolase (FolB)[5]. Both of these proteins have been shown to protect *M. tuberculosis* from damage by reactive oxygen species (ROS)[8,9]. DyP has been shown *in vitro* to consume hydrogen peroxide to degrade guaiacol and other molecules and likely plays a role in ameliorating oxidative stress *in vivo*[5].

Unlike the cargo proteins, the role of CFP29 in *M. tuberculosis* is not well understood. The presence of an encapsulin gene in the genome of many members of the *Mycobacterium* genus hints at the importance of CFP29. Many different genera make up the list of encapsulin containing species suggesting horizontal gene transfer was responsible for spreading encapsulin genes between organisms[10]. However, *Mycobacterium* and *Streptomyces* are overrepresented, comprising ~15% and ~11% of all known species respectively. Biases in the population of sequenced genomes may account for the skew. Alternatively, encapsulins in these species may be playing an important physiological role that leads to them being vertically transferred at a much higher rate than in other species.

Here, we demonstrate that transposon disruption of *cfp29* or disruption of the cargo genes results in a growth defect in interferon gamma activated human cells. A strain of H37Rv with a transposition insertion in the *dyP* gene (Tn::DyP) shows reduced growth in murine bone marrow-derived macrophages (BMMs) and growth is rescued in phagocytic oxidase knockout BMMs. Additionally, we purify CFP29 nanocompartments from *M. tuberculosis* lysate and show that when grown in broth only DyP is loaded.

4-2 Results and Discussion:

The genes necessary for growth of *M. tuberculosis* in broth and murine BMMs have previously been identified using transposon knockout libraries[11,12].

Table 4-1: Fold change in read counts between initial and final library when grown in gamma-activated PBMCs or non-activated PBMCs

PBMCs + IFN-γ

ORF	*Gene*	*Log_2 Fold Change*
0798c	*cfp29*	-1.77
0799c	*dyP*	-1.07
3607c	*folB*	-0.28
3841	*bfrB*	-0.51

PBMCs

ORF	*Gene*	*Log_2 Fold Change*
0798c	*cfp29*	0.27
0799c	*dyP*	0.57
3607c	*folB*	-0.30
3841	*bfrB*	-1.06

Recently, the Stanley lab at University of California - Berkeley made a new *M. tuberculosis* transposon library to identify genes required for growth in human peripheral blood mononuclear cells (PBMCs) that either had or had not been activated with interferon-gamma (IFN-ɣ) (unpublished data). Cells with transposon insertions into *cfp29* grew half as well as wild-type *M. tuberculosis* in human PBMCs but only when the PMBCs had been IFN-ɣ activated (Table 4-1). This was quite surprising as an early transposon library had identified *cfp29* as an essential gene but future libraries did not replicate the finding or see a *cfp29* KO growth defect. Similarly, transposon disruption of the cargo genes showed growth defects in IFN-ɣ activated PBMCs (Table 4-1). Disruption of *dyP* had an even more severe growth defect, growing 30% as well as wild-type H37Rv while disruption of *bfrB* was less severe, growing 70% as well as wild-type. *folB* was shown to be essential for growth in broth but only a minor growth defect was seen in IFN-ɣ activated PBMCs. Disruptions of *bfrB* and *folB* were also growth impaired in non-activated PBMCs indicating they are important for growth even before activation of the intracellular immune response.

Cargo loading has been shown to be sensitive to cargo concentration and competition from other cargo proteins, especially under heterologous conditions[10,13]. DyP, BfrB, and FolB, were confirmed as cargo proteins by heterologous co-expression of one type of cargo protein with CFP29 in *E. coli*[5]. We were curious if these results would replicate in nanocompartments purified from the lysate of *M. tuberculosis* where all three cargo proteins should be in competition with each for loading. Previously CFP29 compartments were purified from the short-term culture filtrate and characterized as secreted membrane associated proteins[2]. As discussed in Chapter 1, this result was due to the size and stability of encapsulins and confounded the truth that encapsulins are soluble intracellular proteins. Using a series of centrifugation steps and analysis by SDS-PAGE, we were able to purify CFP29 nanocompartments from *M.*

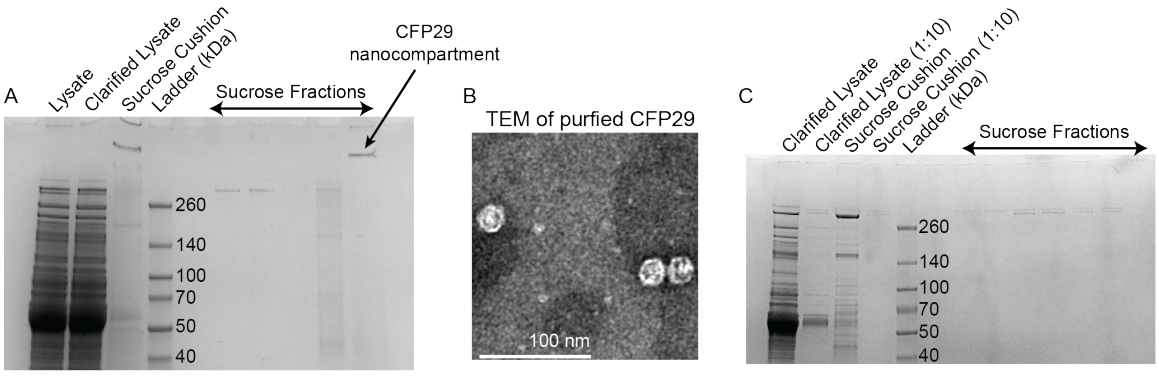

Figure 4-1: Purification of CFP29 nanocompartments (A.) Coomassie stained SDS-PAGE on purification of CFP29 from M. tuberculosis H37Rv. (B.) Negative stained TEM of purified CFP29s. (C.) Coomassie stained SDS-PAGE on attempted purification of CFP29 from Tn::DyP.

tuberculosis lysate (Figure 4-1A). Negative stain transmission electron microscopy confirmed purified CFP29 formed ~24 nm icosahedral structures (Figure 4-1B). To determine the loaded cargo, the band corresponding to assembled CFP29 nanocompartments was excised and analyzed by mass spectrometry (Table 4-2). After filtering out the human keratin results, four *M. tuberculosis* proteins were identified: CFP29, DyP, an protein of unknown function, and rpmA. This indicated that under broth growth conditions, only DyP was loaded into CFP29. DyP was the most likely cargo protein to be loaded under native conditions as encapsulins are usually found in an operon with their cargo protein, which is *dyP* in *M. tuberculosis.* Both the protein of unknown function and rpmA have previously been identified in the cell membrane fraction and are likely purification contaminants[14].

Table 4-2: Peptide counts from mass spectroscopy analysis of CFP29 nanocompartments

ORF	*Protein*	*Counts*
0798c	CFP29	1298
0799c	DyP	343
1762c	Unknown	113
2441c	rpmA	6

Despite not finding peptides for BfrB or FolB, it is not possible to rule them out as cargo proteins. Loading of BfrB and FolB may only occur under specific conditions that weren't present in our growth conditions. Alternatively, DyP may be able to outcompete BfrB or FolB for loading into the CFP29 nanocompartment. To determine if DyP was out competing BfrB and FolB, we attempted to purify encapsulins from Tn::DyP *M. tuberculosi*s. Using the same centrifugation based purification protocol, we were unable to enrich for encapsulins (Figure 4-1C). There are many potential reasons nanocompartments

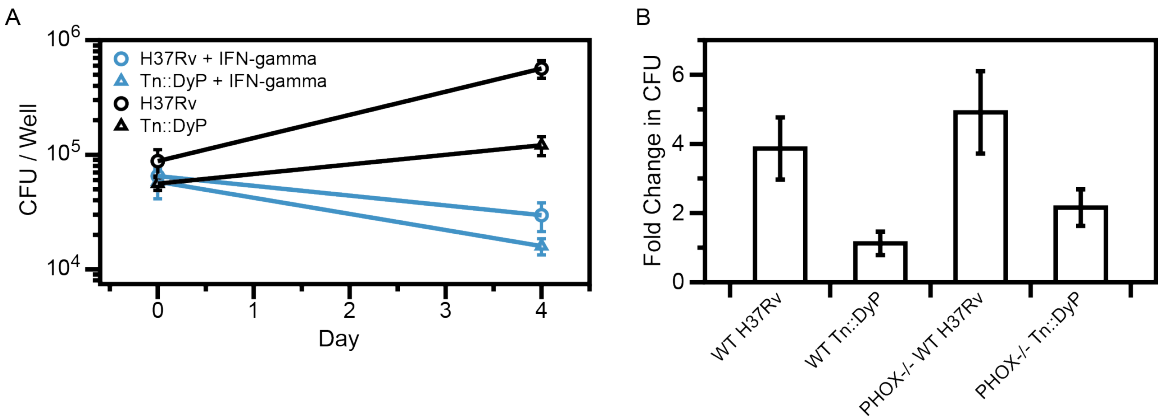

Figure 4-2: In Vivo Growth of Tn::DyP (A) M. tuberculosis CFU at day 0 and day 4 of H37Rv (circles) and Tn::DyP (triangles) with (blue) or without (black) IFN-ɣ activation. (B.) Fold change in CFU between day 0 or day 4 of H37Rv or Tn::DyP grown in wild-type BMMs or PHOX$^{-/-}$ BMMs.

could not be purified from Tn::DyP *M. tuberculosis*. One likely explanation arises for the nature of the *dyP* and *cfp29* operon. These two genes are tightly packed in the M. tuberculosis genome with the ribosome binding site of *cfp29* located in the 3' end of *dyP*. Disruption of *dyP* with a transposon would likely have polar effects on the transcription or translation of *cfp29*. Polar effects would also confound the results from the transposon library as insertions into *dyP* also disrupt *cfp29*. Complementation experiments with *dyP*, *cfp29* or the whole operon are ongoing to determine which components are important for growth.

Given the suggested role for encapsulins in mitigating oxidative stress, we were curious how the growth of Tn::DyP compared to *M. tuberculosis* H37Rv in murine BMMs. In unactivated BMMs, the growth of Tn::DyP is substantially reduced compared to H37Rv (Figure 4-2A,B). IFN-ɣ activation of the BMMs leads to killing of M. tuberculosis for both strains however more killing was observed for the Tn::DyP strain. This suggested DyP and possibly CFP29 are important for growth in macrophages and confirmatory experiments are currently in progress. Disruption of *dyP* like results in a growth defect by reducing the ability of Tn::DyP to degrade hydrogen peroxide. In macrophages, the phagocytic oxidase (PHOX) complex is the major source of the hydrogen peroxide found in the phagosomes. Tn::DyP growth was partially rescued in PHOX$^{-/-}$ BMMs suggesting DyP is playing an active role in degrading phagosomal hydrogen peroxide (Figure 4-2B). These preliminary results suggest that DyP and CFP29 play an important part in helping *M. tuberculosis* evade the host's immune response.

Encapsulins were purified from the *M. tuberculosis* lysate while previously they had been characterized as secreted proteins. The presence of encapsulins in the culture filtrate is an artifact from the stability of the assembled complex. They run as a high molecular band even under denaturing conditions (Figure 4-1a). This stability may allow the encapsulins to survive in a phagosome or phagolysome post cell lysis. If the DyP remains active, the encapsulins in the phagolysosome will act in an altruistic manner, reducing the oxidative stress for the surviving *M. tuberculosis*. This hypothesis still requires significant experimental validation but would be exciting reason for the encapsulation of DyP. While still preliminary, these results suggest a unique and interesting role

for encapsulins and DyPs in *M. tuberculosis*. Additionally, it expands the encapsulin field by providing, to date, the best-characterized physiological role for an encapsulin system.

4-3 Methods

4.3.1 Protein Purification

H37Rv lysate and Tn::DyP lysate were generous gifts from the Stanley lab. Encapsulins were purified using a modified previously published protocol[4]. Lysate was clarified by centrifugation at 20,000g for 20 minutes in a JA-20 rotor and then layered onto top of a 38% sucrose gradient. Samples were centrifuged in a type 50.2 Ti rotor for 18 hours at 100,000 rpm. The supernatant was discarded and the pellet was resuspended in 200 µL of PBS. The resuspended pellet was layered on top of a 10-50% sucrose gradient and centrifuged in SW 41 Ti rotor for 21 hours at 100,000g. The gradient was fractionated and aliquots from each fraction were analyzed by SDS-PAGE for the presence of CFP29.

4.3.2 Biophysical Characterization

Polyacrylamide Gel Electrophoresis (PAGE)

Samples were run on Criterion TGX 4 to 20% gels and stained with GelCode Blue for 30 min at room temperature (RT) and destained with diH2O for 2 h at RT. Gels were imaged with a ChemiDoc MP Imaging System.

4.3.3 Electron Microscopy

CFP29 was adsorbed onto a thin carbon grid and stained with a 1% (w / v) uranyl acetate solution. Films were washed with water twice and dehydrated. Grids were examined with a JEOL 12000EX transmission electron microscope, and images were captured with a charge-coupled device (CCD) camera.

4.3.4 In Vivo Growth Assays

Growth assays of H37Rv and Tn::DyP in BMMs were performed by Katie Lien in the Stanley lab and the results are reported here with her permission.

4-4 References

(1) [PDF]Global tuberculosis report 2015.
(2) Rosenkrands, I., Rasmussen, P. B., Carnio, M., Jacobsen, S., Theisen, M., and Andersen, P. (1998) Identification and characterization of a 29-kilodalton protein from Mycobacterium tuberculosis culture filtrate recognized by mouse memory effector cells. *Infect. Immun. 66*, 2728–2735.
(3) Giri, P. K., Kruh, N. A., Dobos, K. M., and Schorey, J. S. (2010) Proteomic analysis identifies highly antigenic proteins in exosomes from M. tuberculosis-infected and culture filtrate protein-treated macrophages. *Proteomics 10*, 3190–3202.
(4) Sutter, M., Boehringer, D., Gutmann, S., Günther, S., Prangishvili, D., Loessner, M. J., Stetter, K. O., Weber-Ban, E., and Ban, N. (2008) Structural basis of enzyme encapsulation into a bacterial nanocompartment. *Nat. Struct. Mol. Biol. 15*, 939–947.
(5) 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., and Goulding, C. W. (2014) Characterization of a Mycobacterium tuberculosis nanocompartment and its potential cargo proteins. *J. Biol. Chem. 289*, 18279–18289.
(6) Ahmad, M., Roberts, J. N., Hardiman, E. M., Singh, R., Eltis, L. D., and Bugg, T. D. H. (2011) Identification of DypB from Rhodococcus jostii RHA1 as a lignin peroxidase. *Biochemistry 50*, 5096–5107.
(7) Sugano, Y., Muramatsu, R., Ichiyanagi, A., Sato, T., and Shoda, M. (2007) DyP, a unique dye-decolorizing peroxidase, represents a novel heme peroxidase family: ASP171 replaces the distal histidine of classical peroxidases. *J. Biol. Chem. 282*, 36652–36658.
(8) Dántola, M. L., Vignoni, M., Capparelli, A. L., Lorente, C., and Thomas, A. H. (2008) Stability of 7,8-Dihydropterins in Air-Equilibrated Aqueous Solutions. *Health Care Anal. 91*, 411–425.
(9) Reddy, P. V., Puri, R. V., Khera, A., and Tyagi, A. K. (2012) Iron Storage Proteins Are Essential for the Survival and Pathogenesis of Mycobacterium tuberculosis in THP-1 Macrophages and the Guinea Pig Model of Infection. *J. Bacteriol. 194*, 567–575.
(10) Giessen, T. W., and Silver, P. A. (2017) Widespread distribution of encapsulin nanocompartments reveals functional diversity. *Nat Microbiol 2*, 17029.
(11) Sassetti, C. M., Boyd, D. H., and Rubin, E. J. (2003) Genes required for mycobacterial growth defined by high density mutagenesis. *Mol. Microbiol. 48*, 77–84.
(12) Rengarajan, J., Bloom, B. R., and Rubin, E. J. (2005) Genome-wide requirements for Mycobacterium tuberculosis adaptation and survival in macrophages. *Proc. Natl. Acad. Sci. U. S. A. 102*, 8327–8332.
(13) Cassidy-Amstutz, C., Oltrogge, L., Going, C. C., Lee, A., Teng, P., Quintanilla, D., East-Seletsky, A., Williams, E. R., and Savage, D. F. (2016) Identification of a Minimal Peptide Tag for in Vivo and in Vitro Loading of Encapsulin. *Biochemistry 55*, 3461–3468.
(14) Mawuenyega, K. G., Forst, C. V., Dobos, K. M., Belisle, J. T., Chen, J.,

Bradbury, E. M., Bradbury, A. R. M., and Chen, X. (2005) Mycobacterium tuberculosis functional network analysis by global subcellular protein profiling. *Mol. Biol. Cell 16*, 396–404.

Chapter 5 Conclusions and Future Directions

5-1 Conclusions

Compartmentalization, a fundamental feature of life, is often achieved in Bacteria and Archaea through proteinaceous shells. Encapsulins, a novel class of protein nanocompartments, were first discovered in 1994. Encapsulins assembled into ~24 to 31 nm icosahedral shells from 60 or 180 protomers respectively[1,2]. Cargo proteins are loaded into encapsulin via a cargo loading peptide (CLP) usually located at the C-terminus of the protein[2]. Since the structural characterization of encapsulins and cargo loading in 2008, a growing number of studies have investigated the physiological role of encapsulins and their utility as nanoreactors for facilitating catalysis of chemical reactions.

In chapter two, we show *T. maritima* encapsulins can serve as a modular nanoreactor. *T. maritima* encapsulins can be heterologously expressed in *E. coli* and purified in high yield. The nanocompartments display a high degree of thermal stability, only disassembling at > 100 C. Encapsulins are also robust to chemical insults, being stably assembled between pH 2 - 12 and only fully denatured at > 7 M guanidinium chloride. The yield, ease of purification, and stability of *T. maritima* encapsulins are all desirable traits for a protein nanoreactor. While T. maritima encapsulins normally load FLPs, we demonstrated it was possible to load heterologous cargo. sfGFP genetically fused to the CLP from FLP were readily loaded into encapsulins *in vitro* and *in vivo*. By fusion of the CLP to a protein of interest, it is possible to modulate the cargo loaded into encapsulin. We also found that while nature preferred a ~30 amino acid CLP, equivalent amounts of cargo loading were observed with a 15 amino acid CLP *in vivo* and a 15 or a 5 amino acid CLP *in vitro*. It is unclear why naturally occurring CLPs are ~ 30 amino acids in length when high affinity binding is seen with shorter CLPs. Perhaps the added flexibility from the longer CLP is important for functioning of the native cargo.

While working with encapsulins, we observed that purified encapsulins had a distinct yellow coloration, which was the basis for the experiments in chapter three. The yellow coloration was from co-purification of riboflavin and flavin mononucleotide indicating *T. maritima* is a flavoprotein. Using cryo-EM, we were able to identify the flavin binding which was adjacent to a surface solvent exposed tryptophan. Mutation of this tryptophan to alanine disrupted the flavin binding site and reduced the number of flavins that co-purified with encapsulins. As flavoproteins are often involved in electron transfer, we sought to determine if the encapsulin shell was involved with the enzymatic function of the loaded FLP[3]. Preliminary experiments demonstrated FLP loaded encapsulins can oxidize and deposit iron and reduced flavins are able to mobilize iron inside encapsulins. However, we have not yet seen a difference between in FLP function between wild type and alanine mutant encapsulins. Why some encapsulins would bind flavins remains an open question but further experiments should reveal if the flavin is connected to the functioning of FLP.

Additionally, we undertook cryo-EM single particle analysis on FLP-loaded encapsulins to determine a high resolution structure of a cargo loaded encapsulin. For peroxidase-encapsulin systems, it was known that only 6 DyP peroxidases could be loaded per compartment due to steric constraints[4]. For FLP-encapsulin systems, it had been speculated that 120 FLPs could be loaded per encapsulin[5]. In our 2D class

averages, we observed that between 3 or 4 FLP decamers were loaded, well below the speculated 12 decamers. Sterically, there should be room for more decamers per encapsulin and it is unclear why more aren't loaded or how the cells control loading. Molecular details are not apparent in the preliminary 3D reconstruction of FLP loaded encapsulins but with refinement we hopefully should be able to determine where the FLP decamers are binding to the shell and if there are any interactions between the FLP and encapsulins besides through the CLP.

In chapter four, we explore the physiological role of an encapsulin in *M. tuberculosis.* A transposon KO library, identified encapsulins and their cargo proteins as being important for growth of *M. tuberculosis* in IFN-ɣ activated human cells. Previous work had identified three different cargo proteins, however when nanocompartments were purified from *M. tuberculosis* grown in broth only the DyP peroxidase was packaged[6]. A DyP KO strain of *M. tuberculosis*, Tn::DyP, had reduced growth in murine derived macrophages compared to wild type *M. tuberculosis*. These preliminary results suggest a novel and important physiological role for encapsulins, helping pathogenic cells evade` the host's immune response. Peroxidase-encapsulin systems are present in a number of pathogenic bacteria, including pathogenic *E. coli*[7]. Ameliorating the oxidative stress created by the host cell in response to infection may be a general role for these encapsulin systems.

5-2 Future Directions

The study of protein nanocompartments is still a young field with many outstanding fundamental questions. Having started working on encapsulins when the field was nascent, I have felt the excitement and frustration that comes with working on a field where so many basic questions are unknown. While I could likely propose multiple theses worth of future directions, here I will focus on questions arising primarily from my research.

The majority of encapsulins are believed to package a single type of cargo protein, which is often located in an operon with the encapsulin gene. A handful of encapsulin systems have been shown to package other cargos that are transcribed from genomically distal locations[6,8,9]. However, loading of these cargo proteins usually occurred under artificial conditions and did not replicate under more native conditions[9]. Our observation that only DyP loaded encapsulins were purified from *M. tuberculosis* is one such example of this. This raises the question - how do encapsulins ensure loading of only the intended cargo protein when other competing intracellular proteins are present as well? The operonic structure of encapsulins systems suggests that co-expression of cargo and shell protomer is important for loading. I know that incubation of cargo with assembled encapsulins does not yield any loaded compartments, so cargo loading can only occur during assembly[10,11]. Is it important for the cargo to be co-expressed with the encapsulin or can one pre-express the cargo and still get efficient loading? Given the interest in moving heterologous enzymes into encapsulins, it is important to understand the mechanism of cargo loading[10]. Insights into the mechanism of loading multiple types of cargo may be gleaned from *M. xanthus*. The encapsulin in *M. xanthus* loads three different types of cargo with two located at distal genomic locations[8]. Understanding how this assembly is possible when most encapsulins load a single type of cargo would be useful for moving pathways of enzymes into encapsulins.

Current nanocompartment thinking suggests that in many cases, encapsulin loading occurs during co-expression of cargo and shell but this method of cargo encapsulation remains to be experimentally demonstrated.

Another fundamental question, broadly relevant to protein-based organelles, is to what extent the compartment itself participates in the reaction. In the case of ferritins it is well established that the shell controls the size of the iron granule[12]. Likewise, the role of the carboxysome is to create a compartment with a high concentration of carbon dioxide[13]. The reason for encapsulation into nanocompartments is more nebulous. Many of the species that have FLP-encapsulin systems also have other iron storage mechanisms and species with peroxidase-encapsulin systems have unencapsulated peroxidases. It has been suggested that FLP-encapsulin systems can store more iron per shell and peroxidase-encapsulin systems stabilize and protect the peroxidase but no definitive results exist[5]. In all of these cases, the shell is a passive participant in the role of the packaged enzyme. From our results with the flavoprotein *T. maritima* encapsulin, I have become interested in when the shell is more than a passive barrier. When and how is the shell actively involved with the functioning of the encapsulated enzyme and how widely spread is this phenomenon? Direct interaction with the packaged enzyme would be an exciting novel role for a protein compartment. On the practical end, the ability to modulate a reaction by binding of a molecule to the shell (such as electron transfer in the case of *T. maritima* encapsulin) is an exciting prospect for nanotechnology.

Towards the end of my research, I was fortunate to start a collaboration looking at the physiological role of encapsulins in *M. tuberculosis*. We showed these encapsulins and their packaged peroxidases may be involved in ameliorating oxidative stress from the host cell. This result while exciting in itself also highlighted how little the field knows about the physiological role for encapsulin systems. As the library of sequenced organisms grows and new tools are developed to find new types of encapsulins, the number of species with nanocompartments has grown[7,9]. In *M. tuberculosis*, *M. xanthus*, and *R. jostii* RHA1 encapsulins play important roles in growth, sporulation and lignin degradation respectively[8,14]. These three are the only known physiological roles for encapsulin while thousands of encapsulin systems are known. Further studies on encapsulins in their native hosts may yield exciting new roles for nanocompartments and further highlight the importance they play in growth and survival. These are only three of many fundamental questions about encapsulins and compartmentalization. I look forward to seeing the field evolve as more and more is understood about these small, novel, and important structures.

References:

(1) Valdés-Stauber, N., and Scherer, S. (1994) Isolation and characterization of Linocin M18, a bacteriocin produced by Brevibacterium linens. *Appl. Environ. Microbiol. 60*, 3809–3814.
(2) Sutter, M., Boehringer, D., Gutmann, S., Günther, S., Prangishvili, D., Loessner, M. J., Stetter, K. O., Weber-Ban, E., and Ban, N. (2008) Structural basis of enzyme encapsulation into a bacterial nanocompartment. *Nat. Struct. Mol. Biol. 15*, 939–947.
(3) Mayhew, S. G., and Ludwig, M. L. (1975) 2 Flavodoxins and Electron-Transferring Flavoproteins. *The Enzymes 12*, 57–118.
(4) 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., and Heck, A. J. R. (2016) Assembly and Mechanical Properties of the Cargo-Free and Cargo-Loaded Bacterial Nanocompartment Encapsulin. *Biomacromolecules 17*, 2522–2529.
(5) He, D., Hughes, S., Vanden-Hehir, S., Georgiev, A., Altenbach, K., Tarrant, E., Mackay, C. L., Waldron, K. J., Clarke, D. J., and Marles-Wright, J. (2016) Structural characterization of encapsulated ferritin provides insight into iron storage in bacterial nanocompartments. *Elife 5*.
(6) 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., and Goulding, C. W. (2014) Characterization of a Mycobacterium tuberculosis nanocompartment and its potential cargo proteins. *J. Biol. Chem. 289*, 18279–18289.
(7) Radford, D. (2015) Understanding the Encapsulins: Prediction and Characterization of Phage Capsid-like Nanocompartments in Prokaryotes. University of Toronto (Canada).
(8) McHugh, C. A., Fontana, J., Nemecek, D., Cheng, N., Aksyuk, A. A., Heymann, J. B., Winkler, D. C., Lam, A. S., Wall, J. S., Steven, A. C., and Hoiczyk, E. (2014) A virus capsid-like nanocompartment that stores iron and protects bacteria from oxidative stress. *EMBO J. 33*, 1896–1911.
(9) Giessen, T. W., and Silver, P. A. (2017) Widespread distribution of encapsulin nanocompartments reveals functional diversity. *Nat Microbiol 2*, 17029.
(10) Tamura, A., Fukutani, Y., Takami, T., Fujii, M., Nakaguchi, Y., Murakami, Y., Noguchi, K., Yohda, M., and Odaka, M. (2015) Packaging guest proteins into the encapsulin nanocompartment from Rhodococcus erythropolis N771. *Biotechnol. Bioeng. 112*, 13–20.
(11) Cassidy-Amstutz, C., Oltrogge, L., Going, C. C., Lee, A., Teng, P., Quintanilla, D., East-Seletsky, A., Williams, E. R., and Savage, D. F. (2016) Identification of a Minimal Peptide Tag for in Vivo and in Vitro Loading of Encapsulin. *Biochemistry 55*, 3461–3468.
(12) Schröder, I., Johnson, E., and De Vries, S. (2003) Microbial ferric iron reductases. *FEMS Microbiol. Rev. 27*, 427–447.
(13) Rae, B. D., Long, B. M., Whitehead, L. F., Förster, B., Badger, M. R., and Price, G. D. (2013) Cyanobacterial carboxysomes: microcompartments that facilitate CO2 fixation. *J. Mol. Microbiol. Biotechnol. 23*, 300–307.
(14) Rahmanpour, R., and Bugg, T. D. H. (2013) Assembly in vitro of Rhodococcus jostii RHA1 encapsulin and peroxidase DypB to form a nanocompartment. *FEBS J. 280*,

2097–2104.