Title
Molecular and shell biology – examining the biochemistry and physiology of prokaryotic nanocompartments

Author
Nichols, Robert James

Publication Date
2020

Peer reviewed|Thesis/dissertation

Molecular and shell biology – examining the biochemistry and physiology of prokaryotic nanocompartments

By

Robert James Nichols

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:

Associate Professor David Savage, Chair
Professor Arash Komeili
Professor Michael Marletta
Professor Krishna Niyogi

Summer 2020

Abstract

Molecular and shell biology – examining the biochemistry and physiology of prokaryotic nanocompartments

By

Robert James Nichols

Doctor of Philosophy in Molecular and Cell Biology

University of California, Berkeley

Associate Professor David Savage, Chair

For many decades, it was thought that the phenomenon of cellular compartmentalization was exclusive to eukaryotes. Advances in cell imaging techniques have revealed that the presence of subcellular compartments is widespread throughout prokaryotes and that there exists a great diversity of organelles in bacteria and archaea. Recently, a new class of prokaryotic organelle has been discovered – the protein-bounded prokaryotic nanocompartments, also called encapsulins. Encapsulins are among the simplest of the prokaryotic organelles, often comprised of a two gene system; a shell-encoding gene and a cargo-encoding gene. Despite their simplicity, little is known about the biochemistry and physiological function of the encapsulins.

Here we investigate the structure and function of the encapsulin nanocompartments from various prokaryotes. First, we demonstrate that the existence of prokaryotic nanocompartments extends beyond the sequence homologs of the previously characterized encapsulins and that there exist additional, evolutionarily distinct families of encapsulins with unique cargo systems. We have characterized one of these new encapsulin families, which we name Family 2A, from the cyanobacteria *Synechococcus elongatus* PCC 7942. This encapsulin system hosts a cysteine desulfurase cargo enzyme that is directed to the interior of the nanocompartment via an N-terminal targeting sequence. We have determined the structure of the Family 2A encapsulin using cryo-electron microscopy to 2.2 Å resolution. This structure has yielded insights into the cargo-binding site within the organelle and the potential size and charge selectivity of the compartment. Additionally, we have characterized the cysteine desulfurase activity of the Family 2A encapsulin and show an increase in cargo enzyme activity upon encapsulation.

Finally, we examine the biochemistry and physiological role of a previously discovered encapsulin from *Mycobacterium tuberculosis* (Mtb). We demonstrate that the Mtb encapsulin is stable and active under the harsh conditions of the phagolysosome – the ecological niche we implicate the Mtb encapsulin is localized. Furthermore, we show Mtb encapsulin plays a functional role in Mtb's resistance to oxidative stress.

Taken together, our work provides critical insight as to how the structure of the prokaryotic nanocompartments informs their function *in vivo* and show that this emerging class of protein-bounded organelle represents a new paradigm within the field of prokaryotic cell biology.

Dedication

I dedicate this dissertation to my mother and father, Marianne and Robert Nichols. Your sacrifices and support made this possible.

Acknowledgements

I want to begin by thanking my family. An infinite thank you to my mother Marianne, my father Robert, my grandmother Lois, and my two sisters Andrea and Jillian. Each of you were so important to my education, especially in the early years which are the most critical. You have been extremely supportive of my journey to get here, and it would have been impossible without your help. I love you all!

Thank you to my partner Madelaine Gaw who gave me critical emotional support in some of the most difficult times of graduate school. Thank you for allowing me to escape into social circles that were not comprised strictly of scientists. It truly allowed me to clear my head when I was stuck on difficult problems in lab and enabled me to reset my approach. I hope you learned a little bit about cyanobacteria and prokaryotic organelles along the way too!

Thank you to the special teachers I had in high school who went out of their way to be excellent instructors: Jason Smith, Shannon Mackes, and Paul Nale. Mr. Nale was the first to stoke my interest in biology, specifically, molecular biology. I know almost certainly I would have never continued my studies in biology if it were not for Mr. Nale's insanely well put-together high school biology courses. Thank you!

I am extremely grateful to the amazing faculty at Ithaca College. I would like to thank my undergraduate thesis advisor, Maki Inada, for allowing me into her lab as a freshman student and enabling me to develop my research skills both of the bench and of the mind. Maki guided me into the research world and showed me how rewarding a career in molecular biology research could be. Thank you Maki! The senior students in the Inada lab were also crucial to my development as a researcher. Thank you to Philip Feinberg and Bushra Amreen for teaching me the ins and outs of being in a research lab.

There are many Ithaca College faculty that I had as instructors who were critical in solidifying my strong foundation in molecular biology and biochemistry including Andy Smith, Marina Caillaud, Ed Cluett, David Gondek, Janet Hunting, DJ Robinson, and Scott Ulrich. I knew the faculty in the Biology and Chemistry departments at Ithaca College had created a unique learning community while I was attending the school, and I never took it for granted. Upon leaving Ithaca, however, I began to learn more about the "typical" experience of an undergrad studying these majors at other institutions. The more I learned, the more I appreciated how these two departments created such an extraordinary environment for students who really wanted to obtain more than just a textbook understanding of any subdiscipline between biology and chemistry.

I am thankful for Sandy Zinkel and her lab at Vanderbilt University for taking me on as an unpolished sophomore summer student and teaching me so many of the molecular and cell biology techniques that are in my arsenal today. A special thank you to Patrice Wagner who was the graduate student that took me under her wing that summer.

I have been tremendously lucky to have had Dave Savage as my graduate school mentor. I originally met Dave through the Amgen Scholars program when I worked in his lab as a summer student. My incredible experience that summer is what clinched my decision to attend graduate school and my goal of completing it at UC Berkeley. While I enjoyed the science in the lab and loved the idea of studying bacterial organelles, my research instincts and interests have always been more geared towards the molecular biology pertaining to the minutia of the 'central dogma' in eukaryotes. I knew, however, the Savage lab was the right place for my graduate work given my experience working with Dave when I was a summer student and Dave's proven track record of being an excellent mentor. Thank you so much Dave, for allowing me to shape my own project within the lab. I am so grateful for all of your guidance in my training and having you as an amazing resource for my development as a scientist. I am very fortunate to have had a PI that cares deeply about their students as individuals and realizes there is more to life than just the bands on a gel. Thank you!

Thank you to the other members of my thesis committee, Arash Komeili, Michael Marletta, and Krishna Niyogi for guiding me in my research and their assistance in helping me complete this dissertation.

Science is a collaborative process and I could not have completed this work without the training and companionship of my fellow lab members. Thank you to the Savage lab members who were only present for the very beginning of my time in lab, but whose work in setting up and organizing the lab to the state it was in upon me joining the lab, allowed me to "stand of the shoulders of giants": Dana Nadler, Stacy-Anne Morgan, Poh Teng, Rayka Yokoo, and Katie Kortright.

The lab mates I had for the majority of my time in graduate school made it all bearable, especially when I was going through a rough time of poor results. Thank you to Caleb Cassidy-Amstutz and Thawatchai Chaijaraphong for your hands-on training in many of the techniques I used for this work and for teaching me most of what I know about protein-bounded organelles. Thank you to Julia Borden who filled the void when these two left and for always being willing to talk with me about all of the crazy possibilities of the biology and biochemistry of encapsulins. Thank you to Rachel Hood, who set the bar high and demonstrated what it is like to be an excellent extremely organized scientist. Even though I did not become even a third of how thoroughly methodical you were, seeing that high standard showed I had much to improve in that area. I am exceedingly thankful for Avi Flamholz for being a great leader in the lab and his critical analysis of everyone's work in every lab meeting and in every huddle around a desk or whiteboard within the lab. My approach to science and answering new questions has improved greatly due to Avi's example. I am indebted to my baymate Luke Oltrogge for all of the vital information that was sprayed at me like a genius firehose from the other side of the bay. Many of your suggestions were critical to the success of my project including the 'Hail Mary' approach to purifying and refolding the protein I needed out of inclusion bodies.

Thank you to my step-baymate Tom Laughlin for everything you taught me about structural biology and being such a calm presence and great example of how to overcome

hardships with amazing success. Thanks to my other baymate Jack Desmarais, who was always willing to brainstorm with me, help me with a coding problem (I am very mediocre at coding), and generally being game to talk about fun and interesting random topics whenever we needed a break from the grind on the bench.

Thank you to Eli Dugan for being an amazing lab manager and keeping the lab organized despite the chaotic entropy of the lab. Shout outs to Shin Kim for following in Eli's footsteps and becoming an excellent lab manager as well. I hope the lab is ready to scream JENGA when one or both of you leave! You two have been an amazing support and resource for the lab.

Thank you Cissi Blikstad for joining our lab and showing me that I still had a lot to learn about biochemistry, and then teaching me to the point where I feel my skills are as refined as they could have been. Thank you to Noam Prywes who also joined the lab during an exciting time of my PhD when I finally started to catch fire and for further adding gas to that fire from our conversations and his poly-histidine-SUMO tag plasmid!

My love goes out to the protein engineering side of the lab: Sean Higgins, Ben Oakes, Arik Shams, and Emeric Charles. You all think you're so cool and that's because you actually are a little bit. I have somehow learned a lot about protein engineering without doing much in the way of benchwork because of the many lab meetings and long conversations I have had with Sean and Ben about their science. Arik, thanks for sharing my sense of humor and always being willing to send me memes or talk trash to lift my spirits. Emeric, thanks for sometimes giving me CRISPR side projects that provided a fun distraction from my own work and for showing the lab how to have a good work-life balance.

To all of the new students who recently joined the lab: Evan Groover, Naiya Phillips, Andrew Plebanek, and Luis Valentin-Alvarado, I am excited to watch you all you become top-notch "Savage cats." Thank you to Evan for, in general, being with the shits and doing things for the culture. I found I was the only person in lab since my summer internship in 2013 (that's seven years) who gave a damn about these things. Other people are going to read this and be confused, or think we are talking about some kind of inside joke that is just between us. They might attempt to understand by reading whatever the top results on urban dictionary are, but that's on them, and they need to do better. The real ones know.

Thank you to Naiya, who was an amazing rotation student with me and helped me work out the details of the cysteine desulfurase assay. I will miss our jaded conversations while we refolded and concentrated encapsulin shell protein. Thank you to Andrew for our great conversations about science fiction, for teaching me more about the lab's 3D printer, which was mostly accumulating dust, and the nice model of my encapsulin that you printed for me. Shout outs to Luis who taught me a lot about phylogenetic tree building and for being such a wonderful and caring person. I wish we could have been in lab together a bit longer, so I could do some fun Cas biochemistry with you.

Thank you so much to my amazing classmates who were by my side in this expedition. I especially want to thank my roommates who were a crucial emotional and social support throughout graduate school: Ben Adler, Eric Sawyer and George Otto. I love you guys. Thanks to my other classmates who have been amazing friends throughout graduate school including, but not limited to: Fernando Rodriguez-Perez, Dom Castanzo, Amy Eisenberg, Charlotte Nixon, Janice Chen, Eva Nichols, Tim Turkalo, Lydia Lutes, David McSwiggen, and Dan Kramer.

An obvious thank you goes out to all of my co-authors who made these papers possible. An extremely special thank you goes to Ben LaFrance who helped me obtain the structure of the nanocompartment we discovered. The project would have been a lot less interesting without you!

It is rare that people read theses, but if you've gotten this far, welcome to the journey. I myself have only read a handful of theses, but the ones I have read are very important to me. These include theses by lab mates whom I dearly miss working with. Others I have read are notorious for their prose and somehow have extreme entertainment value such as Kary Mullis' thesis on the schizokinen – if you have not read this yet, please walk to the UCB Valley Life Sciences Building, read it for ten minutes, then email me about your experience. Other theses I have read have been essential to the work in this dissertation and provided critical insight into the many mysteries of encapsulins including those by Markus Sutter, Devon Radford, and Caleb Cassidy-Amstutz. I thank them for paving the encapsulin world before me and making this work possible.

I cannot sit here and write my acknowledgments and not acknowledge that the world is in the middle of a global pandemic. Shout outs to SARS-CoV-2 for disrupting the world as we know it and forcing me to sit in my apartment and unproductively write this during 'shelter-in-place' orders. I would like to take a moment to recognize how a 30 kb RNA sequence has caused so much havoc and human suffering. I hope I can look back at this time in 50 years and see it as the most dreadful time of my stay on earth and not see any future calamities top this. Thank you in advance to the scientists who will make a vaccine possible, and thank you to the future scientists who will help prevent climate catastrophe and ensure this current time continues to hold the "most dreadful" title.

This pandemic highlights why we need to research all aspects of biology even though they might not be the current "hot thing to study." I am proud that I decided on a rather esoteric field for my graduate studies. This project would never have happened in any setting other than an academic one. Even though it has not been easy, I am glad I chose it.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction: Molecular biology of the shell

†The work presented in this chapter is adapted from the previously published article with permission from:
Nichols RJ*, Cassidy-Amstutz C*, Chaijarasphong T & Savage DF. (2017) Encapsulins: molecular biology of the shell. *Critical Reviews in Biochemistry and Molecular Biology. 52 (5): 583–94.*

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). Nanocompartments are icosahedral capsids that are smaller and less complex than microcompartments. 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 An overview of prokaryotic nanocompartments and the history of their discovery

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 reactions that would be thermodynamically unfavorable in the cytosol, 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[2–6]. These proteinaceous compartments, also known as bacterial microcompartments (BMCs) or prokaryotic microcompartments (MCPs), were first identified in 1956 and are large pseudo-icosahedral structures 80–400nm in diameter[2–4,7,8]. BMCs and MCPs have been identified in many bacterial phyla and compartmentalize a variety of metabolic reactions, ranging from catabolism of ethanolamine and propanediol to anabolic CO2 fixation[2–4,9].

Recently, a novel class of prokaryotic compartments, collectively called encapsulins, was identified. Encapsulin shell proteins share no sequence or structural similarity to BMC or MCP proteins and form significantly smaller capsids (25–42nm in diameter) than microcompartments[10–13]. 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*[14,15]. 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[14,16,17]. 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[15,18–21]. 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[10,17,22].

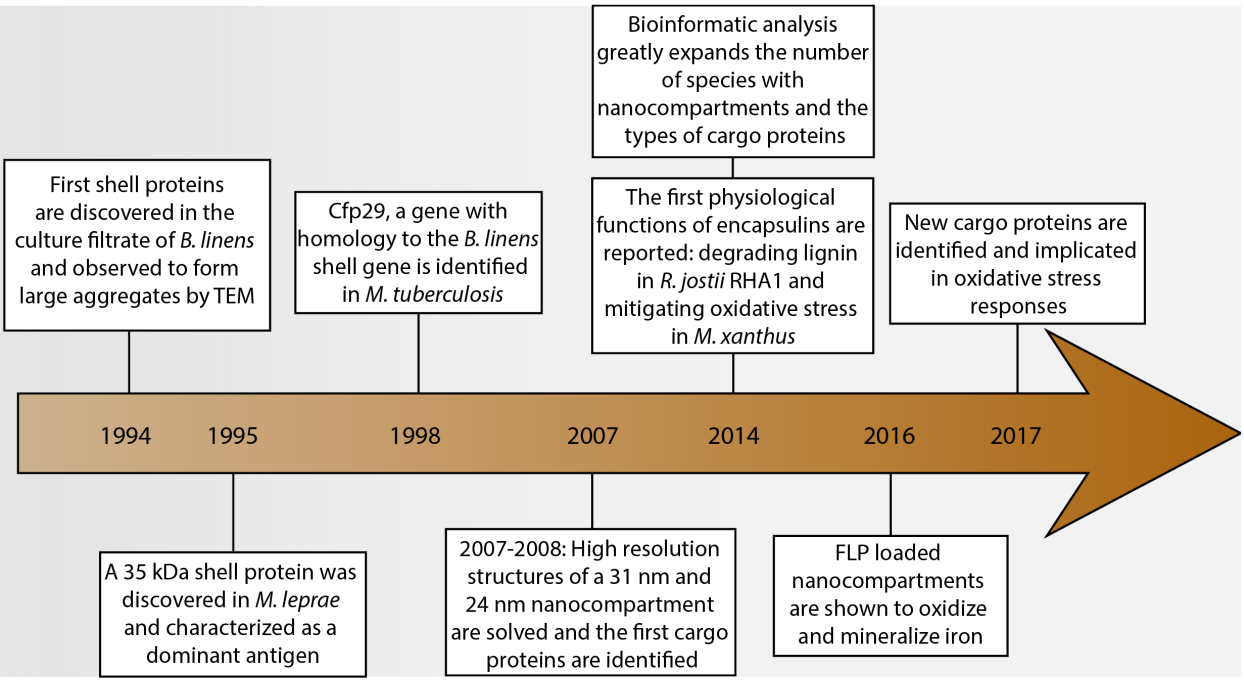

Figure 1.1 Timeline: A brief history of the discovery and characterization of prokaryotic nanocompartments

These structures provided significant insights into shell assembly and cargo loading and spurred new research into prokaryotic nanocompartments[10,17]. Recently, studies have focused on using encapsulins as a programmable nanoreactor or nanomaterial[23–34], while others have explored the physiological role of encapsulin and its cargo in their native bacterial context[11,35–38]. Bioinformatic analysis of sequenced genomes has identified thousands of nanocompartment systems in both Bacteria and Archaea with a diversity of putative cargo proteins[38,39] and recent reviews overview the potential of engineering nanocompartments for biotechnology applications[28,34]. Already the biotechnological potential of encapsulins is vast and promising. Various groups have begun engineering encapsulins to perform new functions including: i) targeting heterologous enzymes to repurpose encapsulins as synthetic nanoreactors[40], ii) repurposing encapsulins as a drug delivery platform by loading the compartment with a drug and fusing a tissue-specific targeting peptide to the encapsulin exterior[23], and iii) functionalizing the surface of encapsulin with an antigen to engineer the compartment for rational vaccine design[31].

While the bioengineering potential of encapsulins is promising and may become its own subfield within synthetic biology, the rest of this chapter and the majority of this

dissertation will focus on the advances in the basic sciences of encapsulin biology and highlight the many open questions related to bacterial physiology these discoveries raise.

1.3 Structure of the nanocompartment shell

Encapsulin shell proteins form nanocompartments that resemble viral capsid structures (Figure 1.2A). High resolution structures of nanocompartments from *P. furiosus*, *T. maritima*, and *Myxococcus xanthus* have revealed that their shells are icosahedral complexes formed by the self-assembly of a single species of protomer protein[10,11,17]. 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 homologous to the HK97 phage major capsid protein gp5 as evident by the structural similarity between gp5 and encapsulin protomers[41] and, like gp5, encapsulin shell proteins homo-oligomerize to form the complete nanocompartment. As is common with viral capsids, encapsulin shell proteins can assemble into icosahedrons of different sizes – encapsulins from *P. furiosus* and *M. xanthus* are composed of 180 protomers (31 nm in diameter)[10,11], while those of *T. maritima* are formed by 60 protomers (24 nm in diameter)[17]. More recently, a structure for a newly discovered encapsulin revealed that the nanocompartment from *Quasibacillus thermotolerans* forms a 240-subunit compartment (42 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 icosahedral structures[42]. 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[10]. Increasing in complexity, the Q. thermotolerans encapsulin forms a T=4 compartment with 12 pentameric vertices and 30 hexameric capsomer faces[12]. Notably, all known nanocompartments are significantly smaller than the HK97 phage capsid, which forms a 420 protomer T=7 capsid[41] (Figure 1.2B).

The encapsulin protomer adopts an HK97 fold, which has three conserved domains: a peripheral domain (P-domain), an axial domain (A-domain), an elongated loop (E-loop), and an N-terminal arm domain (N-arm)[10,13,17,41,43] (Figure 1.2C). While the *P. furiosus* and *M. xanthus* encapsulin protomers align well with the HK97 gp5 protomer[10,11], only the A and E domains from the *T. maritima* encapsulin align well with those of the former[17]. Specifically, the E-loop from *T. maritima* is shorter and rotated relative to the E-loops of HK97 phage, *P. furiosus*, and *M. xanthus*[17]. 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[17]. This feature is not observed in the *P. furiosus* and *M. xanthus* nanocompartments or the HK97 phage[10,11,41]. 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* and *M. xanthus* nanocompartments form 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[38], implicating the E-loop in determining triangulation number and the size of the nanocompartment.

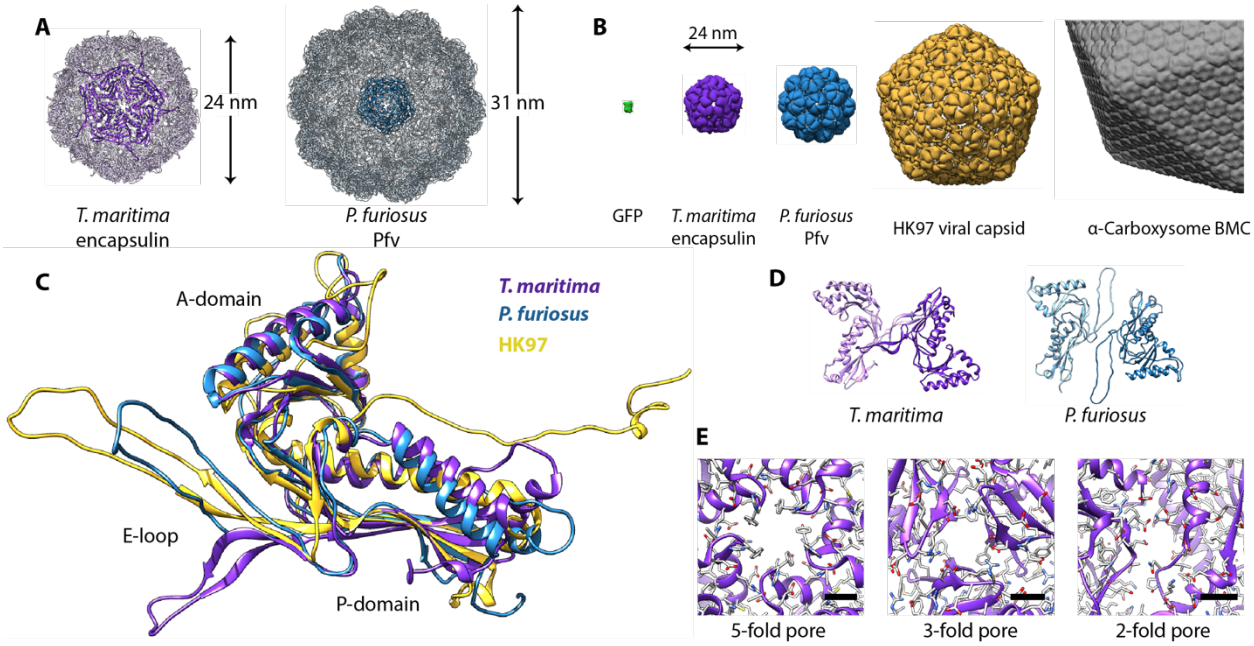

Figure 1.2 Structure of prokaryotic nanocompartments. (A) Nanocompartments from *T. maritima* [PDB: 3DKT] and *P. furiosus* [PDB: 2E0Z] with a vertex pentamer highlighted. (B) Size comparison of the nanocompartments to GFP [PDB: 1C4F], the HK97 viral capsid [PDB: 1OHG], and the carboxysome. Only a quarter 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 of 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 five-fold symmetry, a pore at the site of three-fold symmetry, and an opening at the dimer interface between two protomers. Black scale bars are 6 Å in length

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[10,17]. While only structures of FLP-packaging encapsulins have been solved so far, sequence similarity between the encapsulin shell proteins suggests encapsulins that package other cargo proteins will have similar pores. 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)[17] (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 permeability and affect the activity of the ferritin-like protein (FLP) cargo within[17]. 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[17]), 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 evidence that substrates of encapsulated enzymes, such as hydrogen peroxide or ferrous iron, are able to cross the shell, while proteins and other large molecules cannot[25,26,35,37]. 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[35,44]. 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[35]. Lastly, given that there is no known structure of a DyP-packaging encapsulin, these nanocompartments may have a pore architecture that permits transit of lignin across the shell. The pore openings in the shell likely play an important role in permitting encapsulated enzymes access to their substrates.

Recently, Williams and colleagues engineered the pores of the *T. maritima* encapsulin to understand how the compartment shell acts as a size selectivity barrier[45]. In this study, the shell residues at the five-fold pores were mutated to alanine or deleted entirely and replaced with a glycine linker. The pore mutations resulted in the diameter of the five-fold pore increasing from roughly 3 angstroms to 12 angstroms in diameter. The modified pore enabled the passage of terbium into the compartment, which is nearly double the atomic radius as the native iron substrate of the *T. maritima* pore[45]. Future work will hopefully determine the permeability of other encapsulins 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 FLP found next to the encapsulin shell gene in the *T. maritima* genome[17]. 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 FLP, DyP, hemerythrin, and rubrerythrin[17,38].

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[17,24–26,36,38]. 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 Section 1.5 “Physiology of Nanocompartments”[13,39].

There are, however, alternative 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 shell gene as a single polypeptide[10]. 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 ferredoxin protein with an N-terminal CLP[38]. 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[38]. 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.

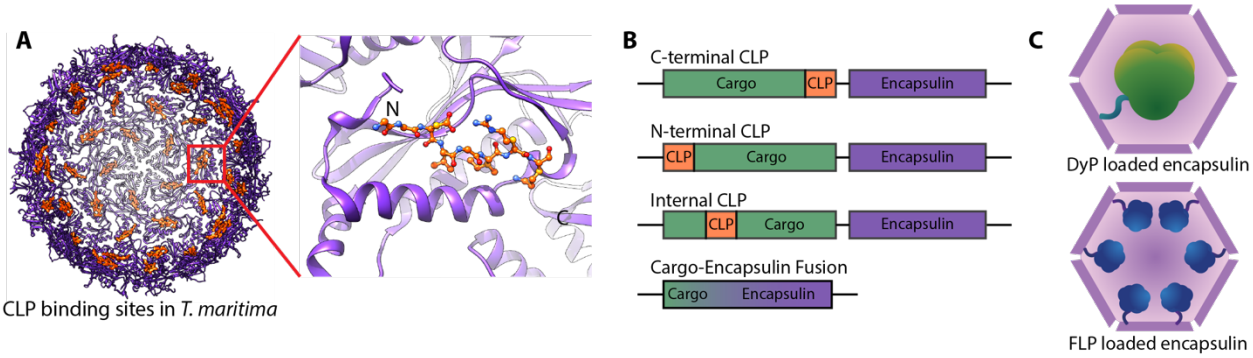

Figure 1.3 A cargo loading peptide, CLP, directs the encapsulation of cargo. (A) Each protomer has a lumenal hydrophobic pocket responsible for interacting with the CLP (B) The CLP coding sequence is located at either the 5’ or 3’ end of the cargo gene. In some cases, direct fusion of the cargo to the 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.

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[11]). 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[36]. 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 be 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[25,26]. 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[38]. 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[17,46]. 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[17]. 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[17], and measurements by native mass spectrometry, have found precisely six DyPs in the T=1 nanocompartment, yielding a 1:10 ratio of cargo to protomer[47]) (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[17,37,48]. 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[37]. This could allow for the packaging of up to 120 FLP molecules per nanocompartment, or a 2:1 ratio of cargo to protomer[37] (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[11].

Ultimately, although all characterized cargo proteins are oligomers, the significance of oligomerization on encapsulation is less well understood. For example, although FLP oligomerization is necessary for the formation of the ferroxidase active site, it may also play other roles, such as position the active site of FLP directly adjacent to the shell pore (He et al. 2016). Work with heterologous cargo has shown that high selectivity is achieved even when packaging just a monomer[26,46]. Given the diversity of cargo proteins and their different oligomeric states, it is likely that there are multiple 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.

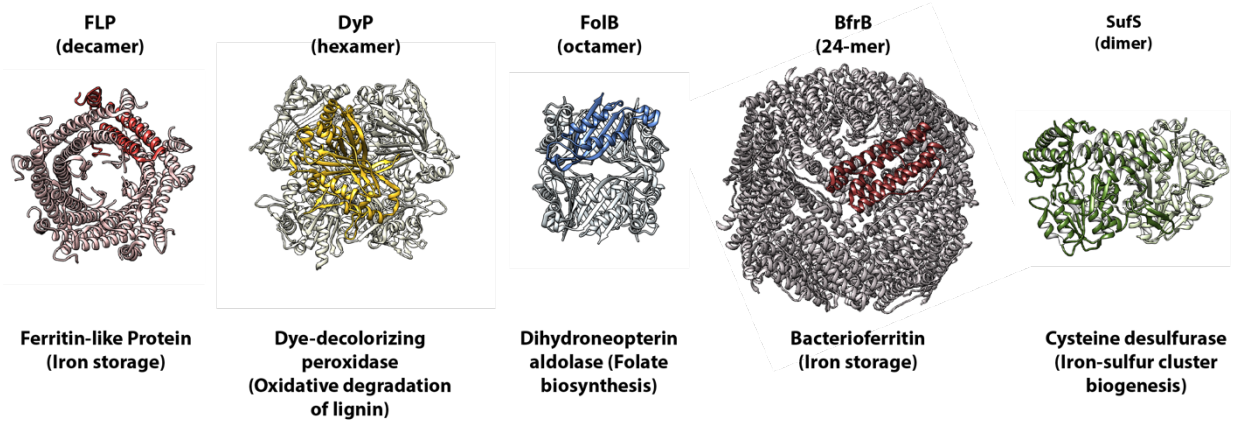

Figure 1.4 High-resolution structures of example oligomeric encapsulin cargo proteins: Ferritin-like protein [PDB: 5DA5], dye-decolorizing peroxidase [PDB: 3QNR], dihydroneopterin aldolase [PDB: 1NBU], bacterioferritin [PDB: 3UNO], and cysteine desulfurase [PDB: 1T3I].

1.5 Stability of encapsulins and a link to function

While a role for cargo encapsulation has been proposed in a few cases, the biological purpose of encapsulation in most organisms remains unknown[11,35,37]. 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 (Sutter et al. 2008, Radford 2015). 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[14,49]. 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* encapsulins, for example, display high tolerance to temperature and denaturant[17,26]. While nanocompartments from mesophilic organisms are not exceptionally thermostable[25], encapsulins are generally pH-stable and resistant to mechanical compression[17,26,35,47]. Like viral phage capsids, prokaryotic nanocompartments display minimal proteolytic degradation following treatment with nonspecific proteases[25,26]. Resistance to proteases also extends to the cargo proteins: firefly luciferase packaged into encapsulins from *Rhodococcus erythropolis* N771 showed no degradation following treatment with trypsin, while non-encapsulated luciferase was completely degraded[25].

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[35] or adding 7 Molar guanidine hydrochloride[26], suggesting that 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[11]. In *M. xanthus*, encapsulin protein levels are substantially reduced 48 h after sporulation and it is completely absent 96 hours after sporulation, suggesting degradation of the nanocompartment. Another theory is that the rate of nanocompartment turnover is slow compared to the rate of cell growth, so the intracellular concentration of encapsulins 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 are species-specific.

Stability may also be related to the hypothesized function of encapsulins in the extracellular environment[14,49]. Starting with their discovery in 1994, encapsulins have often been found in or purified from the culture supernatant[14,35,49]. This led to the hypothesis that nanocompartments are secreted, which was reinforced by the observation that encapsulins were membrane-localized[49]. However, no encapsulins have a known secretion signal[39], nor does there exist a known mechanism for secreting an intact 25–35 nm protein complex[50]. Given the chemical stability and protease resistance of nanocompartments, it is more probable that encapsulins accumulate in the culture supernatant following cell lysis[51]. 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 nanocompartment protein from *Streptomyces griseus* showed that it is located in the cytoplasm[21].

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[35]. 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*.

Additional evidence of encapsulin's putative physiological role in extracellular environments has been demonstrated in genetic studies of the *M. xanthus* encapsulin genes. Deletion of the encapsulin shell gene prevents *M. xanthus* cells from sporulating and forming fruiting bodies[52]. When WT *M. xanthus* cells were added to encapsulin mutant cells, the mutant cells were rescued and formed spores[53]. This extracellular complementation suggests the physiological connection between encapsulin and spore formation has some extracellular component. Whether the encapsulins are somehow up-taken by the mutant cells or the encapsulins perform their biochemical function *ex vivo*.

Future work is needed to investigate these possibilities in addition to ruling out alternative possibilities of gene transfer of the encapsulin genes from the WT to the mutant cells, or the production of a small molecule by the WT cells, which was not explored by Kim and colleagues[53].

1.6 The physiological roles of encapsulins

Given the general architectural similarities between encapsulins and BMCs, drawing analogies between the two systems may provide clues to understanding the physiological roles of nanocompartments. Studies have revealed two major functions of microcompartments[2,54]: (i) to increase the local concentration of substrates near the compartmentalized enzymes, such as CO2 at the site of its fixation inside the carboxysome[55], and (ii) to sequester toxic pathway intermediates such as aldehydes (and possibly enzyme-generated radicals) from the cytosol in catabolic microcompartments[56,57]. 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[11]. 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[58,59]. 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[11]. Furthermore, this ΔencA 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[52].

A connection to oxidative stress is also seen in the encapsulin found in *M. tuberculosis*. Pathogenesis of *M. tuberculosis* is contingent upon an ability to evade the host immune response and the bactericidal ROS generated by the host[60,61]. The encapsulin from *M. tuberculosis* was found to sequester three different cargo proteins: bacterioferritin (BfrB), dihydroneopterin aldolase (FolB), and a dye DyP[36]. Each of these three cargo proteins

have independently been shown to have antioxidant activity[62–65]. For example, BfrB's role in protecting *M. tuberculosis* against oxidative stress was demonstrated by an increased sensitivity to cumene hydroperoxide in a ΔbrfB *M. tuberculosis* mutant[65]. 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 will be needed to more clearly delineate the role of compartmentalization in ROS tolerance.

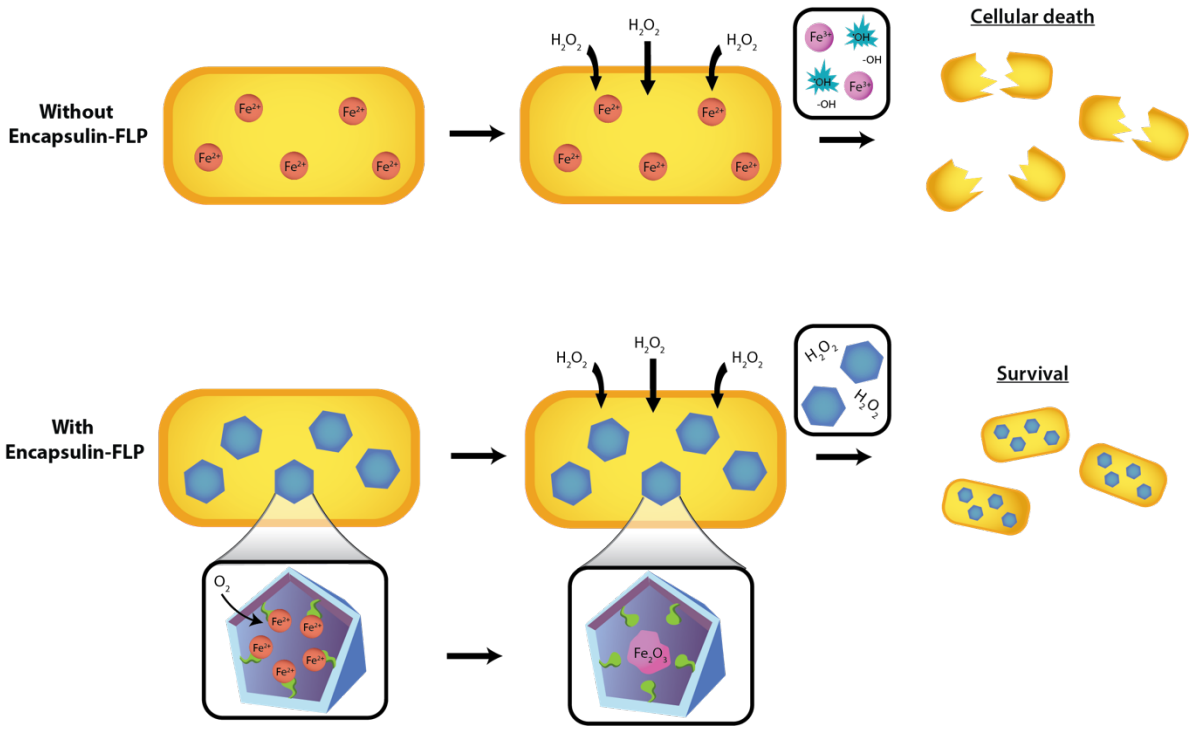

Figure 1.5 FLP-encapsulins protect the cell against oxidative stress. Cytosolic ferrous iron (red) is harmful to the cell under oxidative stress. FLP-encapsulins (blue) will sequester the ferrous iron and store it as mineralized iron oxide (magenta), thereby protecting the cell from reactive oxygen species.

In addition to their potential role in ameliorating oxidative stress, DyP-containing encapsulins have also been implicated in catabolic metabolism[35]. 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[44]. *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[35]. 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 BMCs[66]. 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 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[38,39]. 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 *Escherichia coli* heterologously expressing the hemerythrin-encapsulin system from Streptomyces sp. AA4 to survive treatment with hydrogen peroxide and nitric oxide[67–69]. 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 NIR-HAO and may explain how these bacteria protect themselves from toxic anammox metabolic products such as nitric oxide[70]. 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 sequence identity to known encapsulin shell proteins[39]. 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 encapsulin-like nanocompartments. The newly assigned nanocompartment systems were then grouped based on sequence identity of the shell protein and their shared predicted cargo types. This search identified several putative novel encapsulin shell families with no apparent sequence identity to previously-described encapsulins. As the field moves forward with a likely expanding repertoire of prokaryotic nanocompartments, it may be useful to refer to the various groups of nanocompartments using this family-based nomenclature. Under this system, all previously characterized and published encapsulins would be categorized as Family 1. Intriguingly, Family 1 was determined as the second-most abundant family in Radford's bioinformatic search. 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 detailed experimental characterization of the structure and biochemistry of Family 2 encapsulins[13] is explored further in Chapter 2.

While the predicted Family 2 encapsulins and the several other putative encapsulin families identified in the bioinformatic study were not experimentally characterized by Radford, the bioinformatic study brings clarity to older results from groups studying these genes by implicating them as prokaryotic nanocompartments. Specifically, the EshA protein in *S. 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[19,20]. 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 (Saito et al. 2003). Interestingly, genetic disruption of the *eshA* gene in *Streptomyces coelicolor* A3(2) abolishes the ability of this organism to produce the antibiotic actinorhodin[19], while *eshA* deletion in *S. griseus* interferes with both sporulation and streptomycin production[20,21]. 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[38,39,71]. 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 encapsulin 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.

1.7 References

1. Diekmann, Y. & Pereira-Leal, J. B. Evolution of intracellular compartmentalization. *Biochem. J* **449**, 319–331 (2013).
2. Kerfeld, C. A., Heinhorst, S. & Cannon, G. C. Bacterial microcompartments. *Annu. Rev. Microbiol.* **64**, 391–408 (2010).
3. Yeates, T. O., Thompson, M. C. & Bobik, T. A. The protein shells of bacterial microcompartment organelles. *Curr. Opin. Struct. Biol.* **21**, 223–231 (2011).
4. Chowdhury, C., Sinha, S., Chun, S., Yeates, T. O. & Bobik, T. A. Diverse bacterial microcompartment organelles. *Microbiol. Mol. Biol. Rev.* **78**, 438–468 (2014).
5. Cornejo, E., Abreu, N. & Komeili, A. Compartmentalization and organelle formation in bacteria. *Curr. Opin. Cell Biol.* **26**, 132–138 (2014).
6. Grant, C. R., Wan, J. & Komeili, A. Organelle Formation in Bacteria and Archaea. *Annu. Rev. Cell Dev. Biol.* **34**, 217–238 (2018).
7. Drews, G. & Niklowitz, W. *Beiträge zur Cytologie der Blaualgen. 1. Untersuchungen zur Substruktur von Phormidium uncinatum Gom*. (Springer, 1956).
8. Shively, J. M. Inclusion bodies of prokaryotes. *Annu. Rev. Microbiol.* **28**, 167–187 (1974).
9. Axen, S. D., Erbilgin, O. & Kerfeld, C. A. A taxonomy of bacterial microcompartment loci constructed by a novel scoring method. *PLoS Comput. Biol.* **10**, e1003898 (2014).
10. Akita, F. *et al.* 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 (2007).
11. McHugh, C. A. *et al.* A virus capsid-like nanocompartment that stores iron and protects bacteria from oxidative stress. *EMBO J.* **33**, 1896–1911 (2014).
12. Giessen, T. W. *et al.* Large protein organelles form a new iron sequestration system with high storage capacity. *Elife* **8**, (2019).
13. Nichols, R. J. *et al.* Discovery and characterization of a novel family of prokaryotic nanocompartments involved in sulfur metabolism. *bioRxiv* 2020.05.24.113720 (2020) doi:10.1101/2020.05.24.113720.
14. Valdés-Stauber, N. & Scherer, S. Isolation and characterization of Linocin M18, a bacteriocin produced by Brevibacterium linens. *Appl. Environ. Microbiol.* **60**, 3809–3814 (1994).
15. Winter, N. *et al.* Characterization of the gene encoding the immunodominant 35 kDa protein of Mycobacterium leprae. *Mol. Microbiol.* **16**, 865–876 (1995).
16. Hicks, P. M., Rinker, K. D., Baker, J. R. & Kelly, R. M. 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 (1998).
17. Sutter, M. *et al.* Structural basis of enzyme encapsulation into a bacterial nanocompartment. *Nat. Struct. Mol. Biol.* **15**, 939–947 (2008).
18. Triccas, J. A. *et al.* A 35-kilodalton protein is a major target of the human immune response to Mycobacterium leprae. *Infect. Immun.* **64**, 5171–5177 (1996).
19. Kawamoto, S. *et al.* 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 (2001).

20. Kwak, J., McCue, L. A., Trczianka, K. & Kendrick, K. E. Identification and characterization of a developmentally regulated protein, EshA, required for sporogenic hyphal branches in Streptomyces griseus. *J. Bacteriol.* **183**, 3004–3015 (2001).
21. Saito, N., Matsubara, K., Watanabe, M., Kato, F. & Ochi, K. 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 (2003).
22. Namba, K. *et al.* Expression and molecular characterization of spherical particles derived from the genome of the hyperthermophilic euryarchaeote Pyrococcus furiosus. *J. Biochem.* **138**, 193–199 (2005).
23. Moon, H., Lee, J., Min, J. & Kang, S. Developing genetically engineered encapsulin protein cage nanoparticles as a targeted delivery nanoplatform. *Biomacromolecules* **15**, 3794–3801 (2014).
24. Rurup, W. F., Snijder, J., Koay, M. S. T., Heck, A. J. R. & Cornelissen, J. J. L. M. Self-sorting of foreign proteins in a bacterial nanocompartment. *J. Am. Chem. Soc.* **136**, 3828–3832 (2014).
25. Tamura, A. *et al.* Packaging guest proteins into the encapsulin nanocompartment from Rhodococcus erythropolis N771. *Biotechnol. Bioeng.* **112**, 13–20 (2015).
26. Cassidy-Amstutz, C. *et al.* Identification of a Minimal Peptide Tag for in Vivo and in Vitro Loading of Encapsulin. *Biochemistry* **55**, 3461–3468 (2016).
27. Choi, B. *et al.* 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 (2016).
28. Giessen, T. W. & Silver, P. A. Converting a Natural Protein Compartment into a Nanofactory for the Size-Constrained Synthesis of Antimicrobial Silver Nanoparticles. *ACS Synth. Biol.* **5**, 1497–1504 (2016).
29. Sonotaki, S. *et al.* Successful PEGylation of hollow encapsulin nanoparticles from Rhodococcus erythropolis N771 without affecting their disassembly and reassembly properties. *Biomater Sci* **5**, 1082–1089 (2017).
30. Klem, R., de Ruiter, M. V. & Cornelissen, J. J. L. M. Protecting Encapsulin Nanoparticles with Cysteine-Knot Miniproteins. *Mol. Pharm.* **15**, 2991–2996 (2018).
31. Lagoutte, P. *et al.* Simultaneous surface display and cargo loading of encapsulin nanocompartments and their use for rational vaccine design. *Vaccine* **36**, 3622–3628 (2018).
32. Lee, T.-H., Carpenter, T. S., D'haeseleer, P., Savage, D. F. & Yung, M. C. Encapsulin carrier proteins for enhanced expression of antimicrobial peptides. *Biotechnol. Bioeng.* **117**, 603–613 (2020).
33. Diaz, D., Vidal, X., Sunna, A. & Care, A. Bioengineering a light-responsive encapsulin nanoreactor: a potential tool for photodynamic therapy. *bioRxiv* 2020.06.06.138305 (2020) doi:10.1101/2020.06.06.138305.
34. Demchuk, A. M. & Patel, T. R. The biomedical and bioengineering potential of protein nanocompartments. *Biotechnol. Adv.* **41**, 107547 (2020).
35. Rahmanpour, R. & Bugg, T. D. H. Assembly in vitro of Rhodococcus jostii RHA1 encapsulin and peroxidase DypB to form a nanocompartment. *FEBS J.* **280**, 2097–2104 (2013).

36. Contreras, H. *et al.* Characterization of a Mycobacterium tuberculosis nanocompartment and its potential cargo proteins. *J. Biol. Chem.* **289**, 18279–18289 (2014).
37. He, D. *et al.* Structural characterization of encapsulated ferritin provides insight into iron storage in bacterial nanocompartments. *Elife* **5**, (2016).
38. Giessen, T. W. & Silver, P. A. Widespread distribution of encapsulin nanocompartments reveals functional diversity. *Nat Microbiol* **2**, 17029 (2017).
39. Radford, D. Understanding the encapsulins: prediction and characterization of phage capsid-like nanocompartments in prokaryotes. (2015).
40. Lau, Y. H., Giessen, T. W., Altenburg, W. J. & Silver, P. A. Prokaryotic nanocompartments form synthetic organelles in a eukaryote. *Nat. Commun.* **9**, 1311 (2018).
41. Wikoff, W. R. *et al.* Topologically linked protein rings in the bacteriophage HK97 capsid. *Science* **289**, 2129–2133 (2000).
42. Caspar, D. L. & Klug, A. Physical principles in the construction of regular viruses. *Cold Spring Harb. Symp. Quant. Biol.* **27**, 1–24 (1962).
43. Duda, R. L. & Teschke, C. M. The amazing HK97 fold: versatile results of modest differences. *Curr. Opin. Virol.* **36**, 9–16 (2019).
44. Ahmad, M. *et al.* Identification of DypB from Rhodococcus jostii RHA1 as a lignin peroxidase. *Biochemistry* **50**, 5096–5107 (2011).
45. Williams, E. M., Jung, S. M., Coffman, J. L. & Lutz, S. Pore Engineering for Enhanced Mass Transport in Encapsulin Nanocompartments. *ACS Synth. Biol.* **7**, 2514–2517 (2018).
46. Snijder, J. *et al.* Defining the stoichiometry and cargo load of viral and bacterial nanoparticles by Orbitrap mass spectrometry. *J. Am. Chem. Soc.* **136**, 7295–7299 (2014).
47. Snijder, J. *et al.* Assembly and Mechanical Properties of the Cargo-Free and Cargo-Loaded Bacterial Nanocompartment Encapsulin. *Biomacromolecules* **17**, 2522–2529 (2016).
48. Zubieta, C. *et al.* Crystal structures of two novel dye-decolorizing peroxidases reveal a β-barrel fold with a conserved heme-binding motif. *Proteins: Struct. Funct. Bioinf.* **69**, 223–233 (2007).
49. Rosenkrands, I. *et al.* Identification and characterization of a 29-kilodalton protein from Mycobacterium tuberculosis culture filtrate recognized by mouse memory effector cells. *Infect. Immun.* **66**, 2728–2735 (1998).
50. Green, E. R. & Mecsas, J. Bacterial Secretion Systems: An Overview. in *Virulence Mechanisms of Bacterial Pathogens* (eds. Kudva, I. T. et al.) vol. 1778 213–239 (ASM Press, 2016).
51. Wang, P. *et al.* Robust growth of Escherichia coli. *Curr. Biol.* **20**, 1099–1103 (2010).
52. Kim, D. *et al.* Operon required for fruiting body development in Myxococcus xanthus. *J. Microbiol. Biotechnol.* **19**, 1288–1294 (2009).
53. Kim, D. *et al.* Mutants defective in the production of encapsulin show a tan-phase-locked phenotype in Myxococcus xanthus. *J. Microbiol.* **57**, 795–802 (2019).
54. Yeates, T. O., Crowley, C. S. & Tanaka, S. Bacterial microcompartment organelles: protein shell structure and evolution. *Annu. Rev. Biophys.* **39**, 185–205 (2010).

55. Mangan, N. M., Flamholz, A., Hood, R. D., Milo, R. & Savage, D. F. pH determines the energetic efficiency of the cyanobacterial CO2 concentrating mechanism. *Proc. Natl. Acad. Sci. U. S. A.* **113**, E5354–62 (2016).
56. Penrod, J. T. & Roth, J. R. Conserving a volatile metabolite: a role for carboxysome-like organelles in Salmonella enterica. *J. Bacteriol.* **188**, 2865–2874 (2006).
57. Sampson, E. M. & Bobik, T. A. Microcompartments for B12-dependent 1,2-propanediol degradation provide protection from DNA and cellular damage by a reactive metabolic intermediate. *J. Bacteriol.* **190**, 2966–2971 (2008).
58. Imlay, J. A., Chin, S. M. & Linn, S. Toxic DNA damage by hydrogen peroxide through the Fenton reaction in vivo and in vitro. *Science* **240**, 640–642 (1988).
59. Andrews, S. C. Iron Storage in Bacteria. in *Advances in Microbial Physiology* (ed. R.K. Poole) vol. 40 281–351 (Academic Press, 1998).
60. Manca, C., Paul, S., Barry, C. E., 3rd, Freedman, V. H. & Kaplan, G. Mycobacterium tuberculosis catalase and peroxidase activities and resistance to oxidative killing in human monocytes in vitro. *Infect. Immun.* **67**, 74–79 (1999).
61. Slauch, J. M. How does the oxidative burst of macrophages kill bacteria? Still an open question. *Molecular microbiology* vol. 80 580–583 (2011).
62. Goulding, C. W. *et al.* Regulation by oligomerization in a mycobacterial folate biosynthetic enzyme. *J. Mol. Biol.* **349**, 61–72 (2005).
63. Sugano, Y., Muramatsu, R., Ichiyanagi, A., Sato, T. & Shoda, M. 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 (2007).
64. Pandey, R. & Rodriguez, G. M. 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 (2012).
65. Reddy, P. V., Puri, R. V., Khera, A. & Tyagi, A. K. 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 (2012).
66. Bobik, T. A., Lehman, B. P. & Yeates, T. O. Bacterial microcompartments: widespread prokaryotic organelles for isolation and optimization of metabolic pathways. *Mol. Microbiol.* **98**, 193–207 (2015).
67. Chow, E. D., Liu, O. W., O'Brien, S. & Madhani, H. D. 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 (2007).
68. Kendall, J. J., Barrero-Tobon, A. M., Hendrixson, D. R. & Kelly, D. J. Hemerythrins in the microaerophilic bacterium Campylobacter jejuni help protect key iron–sulphur cluster enzymes from oxidative damage. *Environ. Microbiol.* **16**, 1105–1121 (2014).
69. Li, X. *et al.* A bacterial hemerythrin-like protein MsmHr inhibits the SigF-dependent hydrogen peroxide response in mycobacteria. *Front. Microbiol.* **5**, 800 (2014).
70. Kartal, B. *et al.* How to make a living from anaerobic ammonium oxidation. *FEMS Microbiol. Rev.* **37**, 428–461 (2013).
71. Heinemann, J. *et al.* Fossil record of an archaeal HK97-like provirus. *Virology* **417**, 362–368 (2011).

Chapter 2

Discovery and characterization of a novel family of prokaryotic nanocompartments involved in sulfur metabolism

†The work presented in this chapter is adapted from the previously published article with permission from:
Nichols RJ*, LaFrance B*. Phillips NR, Oltrogge LM, Valentin-Alvarado L, Bischoff AJ, Nogales E, & Savage DF (2020) Discovery and characterization of a novel family of prokaryotic nanocompartments involved in sulfur metabolism. *bioRxiv. 2020.05.24.113720*

2.1 Abstract

Prokaryotic nanocompartments, also known as encapsulins, are a recently discovered proteinaceous organelle in prokaryotes that compartmentalize cargo enzymes. While initial studies have begun to elucidate the structure and physiological roles of encapsulins, bioinformatic evidence suggests that a great diversity of encapsulin nanocompartments remains unexplored. Here, we describe a novel encapsulin in the freshwater cyanobacterium *Synechococcus elongatus* PCC 7942. This nanocompartment is upregulated upon sulfate starvation and encapsulates a cysteine desulfurase enzyme via an N-terminal targeting sequence. Using cryo-electron microscopy, we have determined the structure of the nanocompartment complex to 2.2 Å resolution. Lastly, biochemical characterization of the complex demonstrated that the activity of the cysteine desulfurase is enhanced upon encapsulation. Taken together, our discovery, structural analysis, and enzymatic characterization of this prokaryotic nanocompartment provide a foundation for future studies seeking to understand the physiological role of this encapsulin in various bacteria.

2.2 Introduction

Subcellular compartmentalization is an important strategy used by cells to facilitate metabolic pathways that are incompatible with the rest of the cytosol. Contrary to common misconceptions that organelles are exclusive to eukaryotes, even prokaryotes partition metabolic pathways into unique chemical environments using subcellular compartments[1]. For example, studies of the bacterial microcompartments called carboxysomes have shown how the complex sequesters the enzyme rubisco and facilitates substrate channeling by increasing the local concentration of CO_2[2–4]. In addition to modulating cargo enzyme activity, compartmentalization can also provide a means of sequestering toxic intermediates of metabolic pathways from the rest of the cell. For example, the propane-diol utilization (PDU) microcompartment sequesters a cytotoxic aldehyde intermediate from the cytoplasm and allows it to be subsequently converted by downstream, compartmentalized enzymes to efficiently generate the end products of the pathway[4–7].

Recently, another class of protein-bounded compartments, known as prokaryotic nanocompartments, has been discovered[8]. These nanocompartments, also called encapsulins, are smaller and less complex than microcompartments. They are typically found as a two-gene system, which encodes a shell protein that self-assembles into an icosahedral capsid-like compartment, and a cargo protein that becomes encapsulated by the shell through a targeting peptide sequence[9,10]. Many functionally diverse cargo proteins have been found to be associated with encapsulins, including ferritin-like proteins (FLP), iron mineralizing encapsulin cargo from firmicutes (IMEF), DyP-type peroxidases, and hydroxylamine oxidoreductase (HAO)[8,11–13]. The precise physiological role of these compartments remains elusive except for a few instances. Notably, the DyP-containing encapsulins from *Myxococcus xanthus* have been implicated in nutrient starvation and oxidative stress responses[14–16]. The FLP and IMEF containing encapsulins appear to be involved in iron storage and mitigation of toxic reactive oxygen species products of the

Fenton reaction due to free iron during oxidative stress[11,12,17]. Encapsulins are also thought to be integral to highly-specialized metabolism, such as that found in anammox bacteria, in which the HAO cargo has been hypothesized to reduce a cytotoxic hydroxylamine metabolic intermediate[11,13,18].

Based on the evidence accumulated thus far, it appears that encapsulins play diverse physiological roles. Despite this diversity, encapsulation of redox reactions is a recurring theme[10]. Thus far, study of this expansive repertoire of encapsulins has been limited to the homologs of closely related compartment shell proteins. Here we describe a new family of nanocompartment systems that are evolutionarily distinct from those previously reported. Specifically, we have implicated a role for this encapsulin family in the sulfur starvation response. Further, we have identified a unique cysteine desulfurase cargo enzyme and elucidated an N-terminal encapsulation targeting sequence that is necessary and sufficient for compartmentalization. Finally, we report a high-resolution structure (2.2 Å) of the complex and identify the cargo binding site within the compartment. This structure greatly informs our model for the biochemical function of this novel organelle.

2.3 Results

2.3a A novel family of predicted prokaryotic nanocompartments is widespread throughout bacterial phyla

A unifying feature of the encapsulin nanocompartments is the shared HK97 phage-like fold of the shell protein[10,11]. Bacteriophages belonging to the order Caudovirales also possess a major capsid protein that is structurally homologous to the HK97 fold. Given the shared homology of encapsulins and Caudovirales capsid proteins, an evolutionary relationship between the two has been proposed[19–21], and the existence of other bacteriophage-related nanocompartments beyond the close relatives of known encapsulins has been postulated[10,22,23]. Recently, a bioinformatic study explored this possibility by searching prokaryotic genomes for phage capsid genes that are unlikely to be functional phages, but may actually be putative encapsulins[22]. This search suggested that previously published encapsulins, hereafter referred to as Family 1 encapsulins, comprise a minor fraction of all encapsulin systems. Here we report the first characterization of a novel encapsulin family, which we term Family 2. This novel family is even more prevalent than Family 1 encapsulins (Table 2.1) and is present in many model organisms. Despite the prevalence of Family 2 encapsulins, the experimental characterization of this family as a prokaryotic nanocompartment has never been explored.

Phylogenetic analysis of the encapsulin shell proteins revealed that Family 2 encapsulins are distinct from the previously published Family 1 systems (Figure 2.1A). Family 2 further divides into what we propose as two distinct subfamilies, Family 2a and Family 2b. The subfamilies can be distinguished from each other by their phylogenetic clustering (Figure 2.2). Notably, the two subfamilies are found in distinct genomic contexts. Family 2a is found adjacent to sulfur metabolism genes whereas Family 2b neighbors genes involved in terpenoid synthesis (Figure 2.1B; Figure 2.3). Most prevalent among the Family 2a subfamily was the co-occurrence of a neighboring cysteine desulfurase gene, while the

Family 2b shell genes were found to most often neighbor a polyprenyl diphosphate synthase gene (Table 2.2 and Table 2.3). The individual subfamilies can also be defined by the CRP/FNR cyclic nucleotide-binding domain that is predominantly found in the Family 2b shell sequences but not Family 2a. Our focus for the remainder of the paper will be on the Family 2a subfamily, which is widespread and found distributed in a polyphyletic fashion throughout 13 bacterial phyla (Figure 2.1A).

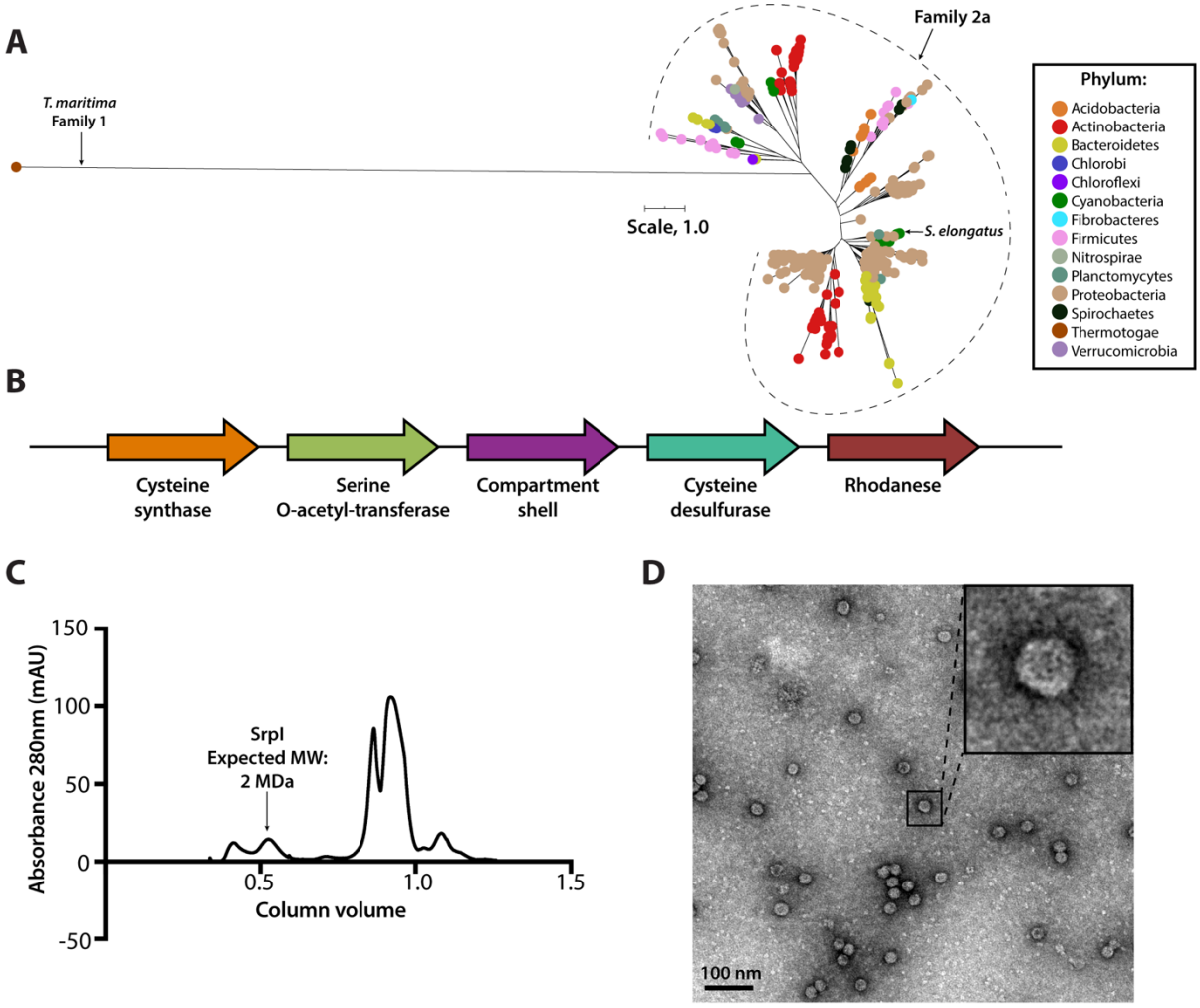

Figure 2.1 SrpI is a bacterial nanocompartment that is widespread throughout bacterial phyla and found neighboring sulfur metabolism genes. (A) Maximum-likelihood phylogenetic tree of Family 2a encapsulin shell proteins using the *T. maritima* Family 1 encapsulin shell protein (WP_004080898.1) as a Family 1 representative. Scale bar, one substitution per site. (B) Genomic neighborhood of the Family 2a encapsulin shell gene from *S. elongatus* PCC 7942. (C) Size exclusion-chromatogram of purified SrpI shell protein using a Superose™ 6 Increase column (GE Life Sciences). Expected molecular weight was determined using the previously characterized *T. maritima* encapsulin and Bio-Rad gel filtration calibration standard (D) Negative stain TEM micrograph of resulting SrpI encapsulin-containing fraction post size-exclusion chromatography. Scale bar, 100 nm.

One such occurrence of Family 2a is in the model cyanobacterium *Synechococcus elongatus* PCC 7942 (henceforth *S. elongatus*) and we sought to validate whether the

predicted encapsulin shell gene (Synpcc7942_B2662, SrpI) was indeed part of a nanocompartment complex. Expression of the shell gene in *Escherichia coli* BL21 (DE3) cells, followed by purification and size-exclusion chromatography, revealed that the protein eluted with an estimated molecular weight of ≈2 MDa (Figure 2.1C), the typical size for many previously characterized encapsulins[24,25]. Consistent with the previously characterized Family 1 encapsulin from *Thermotoga maritima*, a high molecular weight band was detected with SDS-PAGE analysis for non-heat denatured samples. Boiling the sample yielded a band at 35 kDa, the expected weight of the monomeric shell protein[24] (Figure 2.4A). Negative stain transmission electron microscopy (TEM) of the purified sample indicated the complex forms a nanocompartment with an average diameter of 25±1 nm (n=180) (Figure 2.1D, Figure 2.4B).

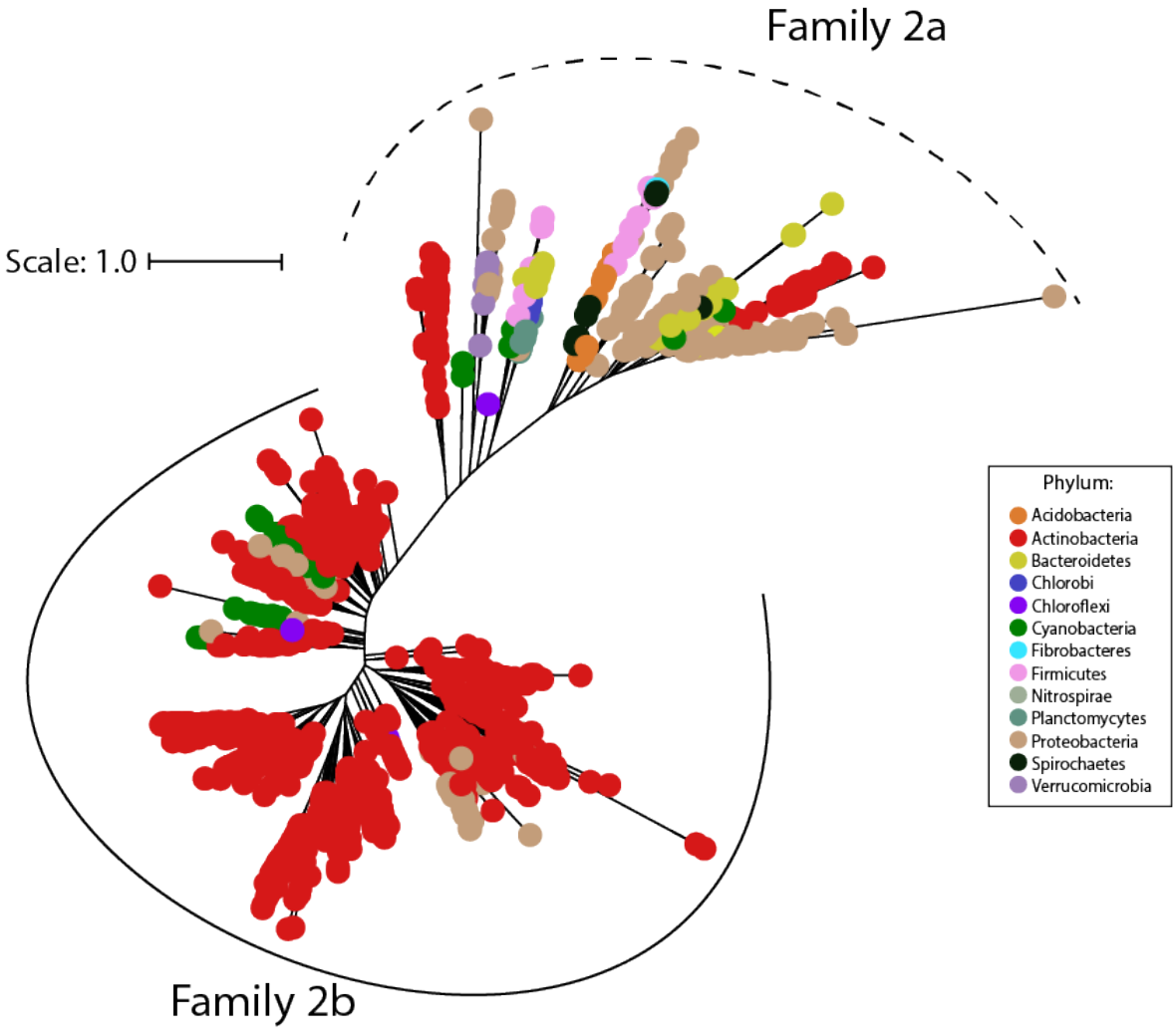

Figure 2.2 Family 2 encapsulins can be divided into two phylogenetically distinct subfamilies. Maximum likelihood phylogenetic tree of 1383 members of Family 2 encapsulins. Family 2a (dashed line grouping) and Family 2b (solid line grouping) represent two subfamilies. Family 2 encapsulins are found throughout 13 different bacterial phyla and are distributed in a polyphyletic fashion. Scale bar, one substitution per site.

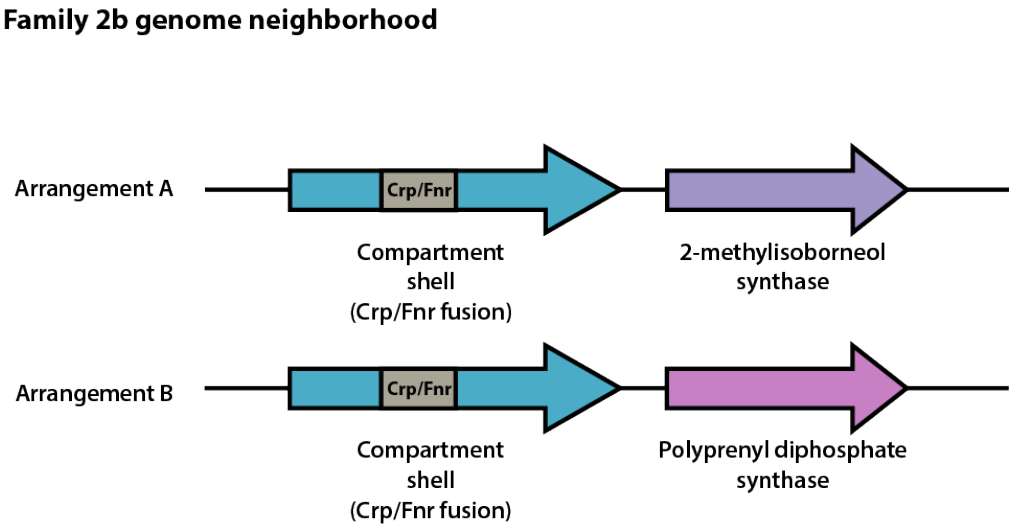

Figure 2.3 Family 2b shell genes neighbor 2-methylisoborneol synthase or polyprenyl diphosphate synthase. Schematic of representative genomic arrangements for Family 2b shell genes determined using the EFI-GNT web tool. One or both arrangements (A and B) may be found within a given genome. Family 2b shell genes possess a Crp/Fnr transcriptional regulator domain (grey) within the shell gene.

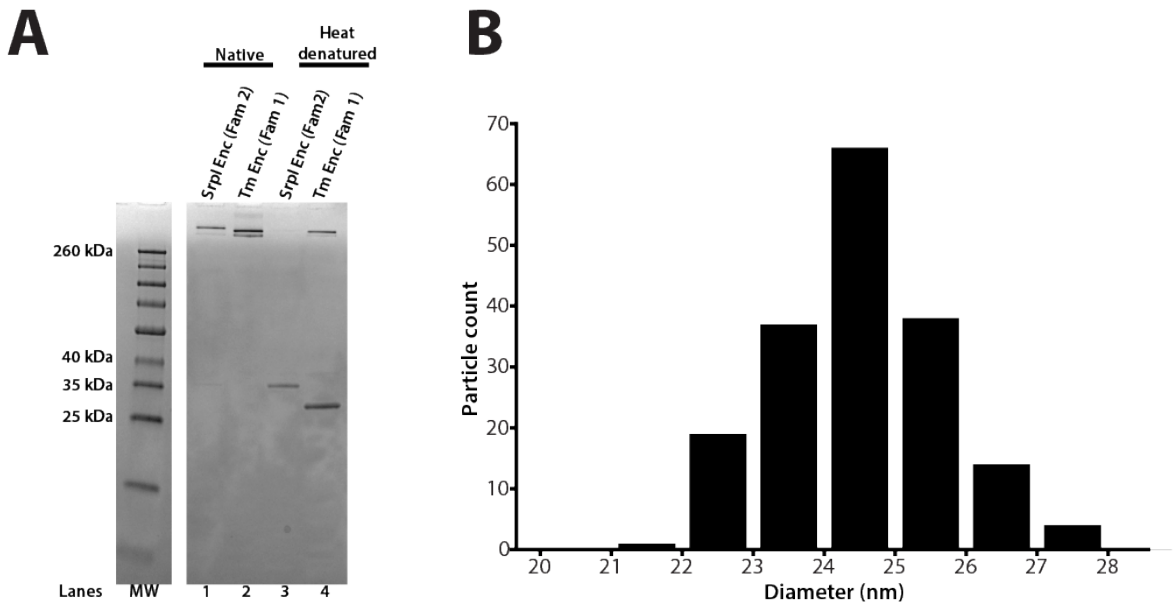

Figure 2.4 The Family 2 encapsulin, SrpI, forms a high molecular weight complex similar to the Family 1 encapsulin from *T. maritima*. A) SDS-PAGE analysis of purified *S. elongatus* PCC 7942 Family 2 encapsulin (SrpI Enc) and *T. maritima* Family 1 encapsulin (Tm Enc). Lanes 1 and 2 correspond to samples that were not heat-denatured (Native). Lanes 3 and 4 correspond to samples that were heat denatured at 95°C for 15 minutes. Molecular weight marker (MW): Spectra™ Multicolor Broad Range Protein Ladder. B) Size distribution of SrpI encapsulin. Diameter of purified SrpI encapsulin determined by negative stain transmission electron microscopy. Quantification of 180 particles from micrographs performed using FIJI image processing package.

2.3b SrpI encapsulin is upregulated under sulfur starvation and hosts a cysteine desulfurase cargo protein

Previous work by Nicholson and colleagues in *S. elongatus* demonstrated that the encapsulin shell gene (Synpcc7942_B2662) is one of many whose mRNA expression level is upregulated upon sulfur starvation[26,27]. Thus, this gene, which is found on a plasmid encoding many sulfur-related genes, was termed SrpI for Sulfur regulated plasmid-encoded gene-I[27,28]. In order to validate this result at the protein level, we sulfur starved wild-type *S. elongatus* cells for the duration of a 48-hour time course to detect the upregulation of the nanocompartment and, potentially, identify additional cargo via mass spectrometry.

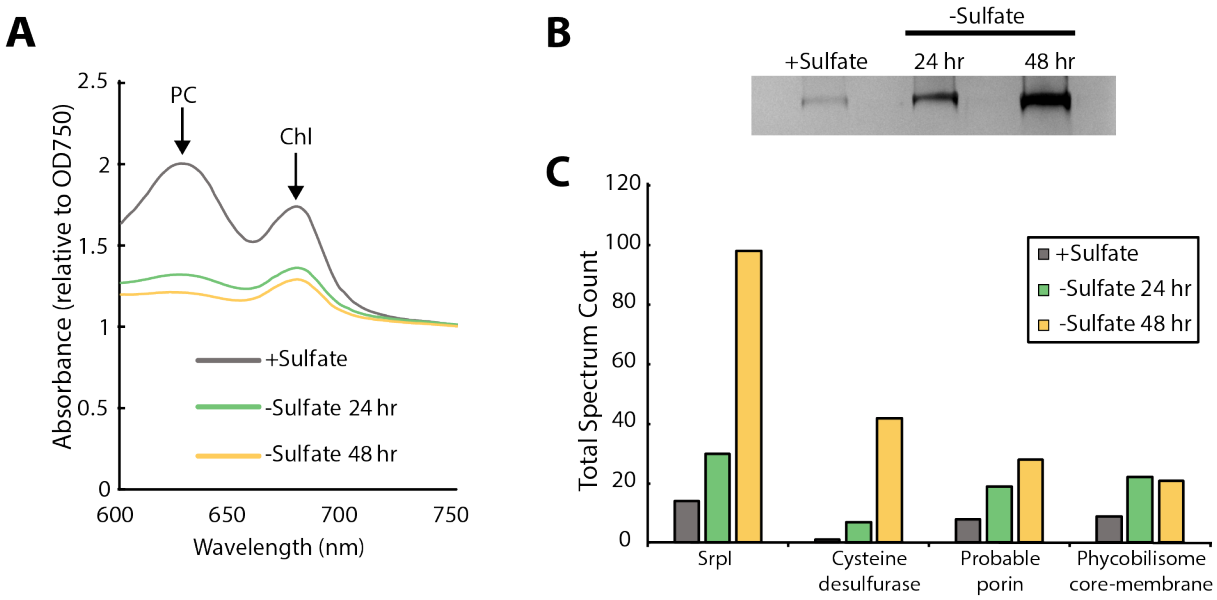

Figure 2.5 SrpI encapsulin is upregulated in S. elongatus upon sulfate starvation. (A) Absorbance spectra of *S. elongatus* liquid cultures under nutrient-replete conditions (+Sulfate) and sulfur starvation (-Sulfate) for 24 and 48 hours. Absorbance maxima of phycocyanin (PC) at 620 nm and chlorophyll (Chl) at 680 nm are indicated. Absorbance spectra are normalized to the same optical density at 750 nm. (B) Non-denaturing SDS-PAGE analysis of lysates from nutrient-replete and sulfur starved *S. elongatus* cultures visualized by silver stain. Inputs were normalized using absorbance at 280 nm. (C) Liquid chromatography-mass spectrometry of excised high molecular weight bands from SDS-PAGE analysis. Top protein hits from each condition are represented by total spectrum counts.

Consistent with previous studies of sulfur starvation in cyanobacteria, we observed the expected chlorosis phenotype due to the degradation of phycobilisomes[29]. Chlorosis was confirmed by loss of phycocyanin absorbance at 620 nm (Figure 2.5A, Figure 2.6). During this time-course, SDS-PAGE analysis of cell lysates also indicated upregulation of a high molecular weight complex (Figure 2.5B). Bands were excised, proteolytically digested, and analyzed via liquid chromatography–mass spectrometry. After 48 hours of sulfur starvation, the top hits, as determined by total spectral counts, were the putative encapsulin shell protein, SrpI, and the product of the neighboring gene

(Synpcc7942_B2661), a cysteine desulfurase (Figure 2.5C). Taken together, these results suggest that the cysteine desulfurase, which neighbors the SrpI shell gene, is the encapsulated cargo protein[10] (Table 2.2) .

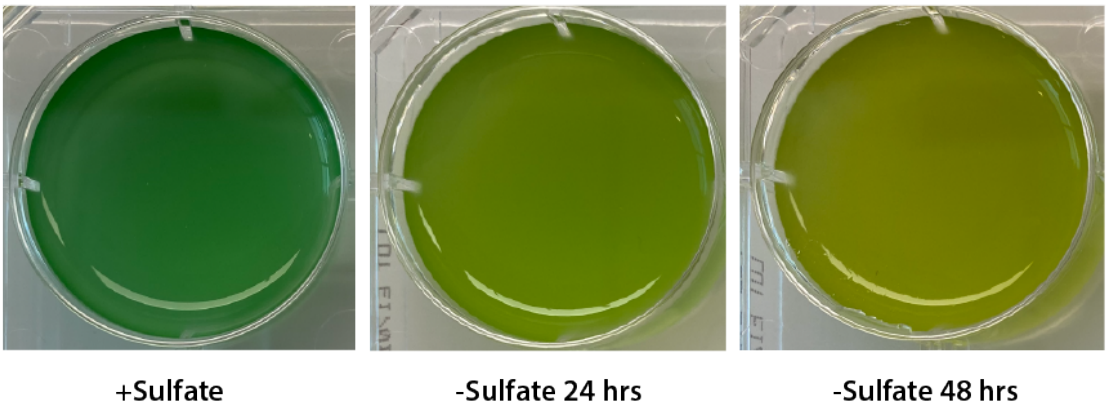

Figure 2.6 Chlorosis phenotype of sulfur-starved S. elongatus PCC 7942. Liquid cultures of *S. elongatus* PCC 7942 grown in nutrient replete medium (+Sulfate) or under sulfate starvation for 24 hours and 48 hours.

2.3c A disordered N-terminal domain targets cargo for SrpI encapsulation *in vivo*

Sequence alignment of the five cysteine desulfurases found in the *S. elongatus* genome revealed that the SrpI-associated cysteine desulfurase (Synpcc7942_B2661), hereafter named CyD, possesses a unique N-terminal domain in addition to the canonical cysteine desulfurase domain (Figure 2.7A). This N-terminal domain is shared by cysteine desulfurases found adjacent to SrpI homologs in species possessing this encapsulin system. Structural prediction using the primary sequence of the SrpI-associated CyD revealed that the N-terminal domain (NTD) is highly disordered (Figure 2.7A). Intrinsically disordered domains are known to evolve rapidly, preserving bulk chemical characteristics even as the sequence diverges greatly (Moesa et al., 2012; Varadi et al., 2015). Accordingly, sequence conservation throughout the N-terminal domain is sparse. However, two motifs, 'LARLANEFFS' and 'AASPYYFLDG', can be found in most SrpI-associated CyD sequences (Figure 2.7B; Figure 2.8).

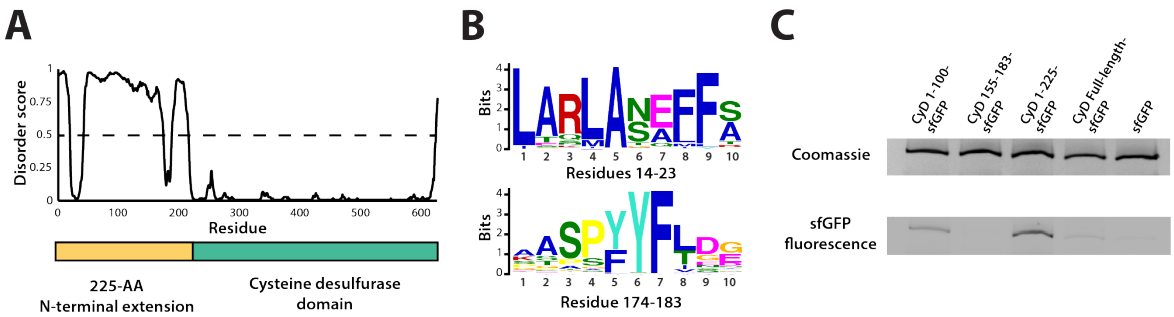

Figure 2.7 An N-terminal signal sequence directs cargo loading in vivo. (A) Domain organization of cysteine desulfurase (CyD; Synpcc7942_B2661) and the predicted disorder scores calculated using DISOPRED3. CyD can be split into two domains – a highly disordered N-terminal domain and an ordered cysteine desulfurase domain. (B) Sequence WebLogos of conserved motifs

found within the N-terminal domain of CyD calculated using the MEME suite motif discovery server. (C) SDS-PAGE analysis of CyD constructs fused to sfGFP. Loading of fusion cargo and untagged-sfGFP control was determined by fluorescence of the nanocompartment band prior to Coomassie staining.

We next sought to confirm that CyD is the cargo protein by using the N-terminal domain to target heterologous cargo to the compartment. To identify the minimal sequence necessary for encapsulation, truncated sequences of the CyD cargo gene were fused to the superfolder green fluorescent protein variant (sfGFP) and co-expressed with the shell protein in *E. coli*. This same approach has been applied to identify targeting sequences for the Family 1 encapsulins (Cassidy-Amstutz et al., 2016). Examination of these expressed constructs via SDS-PAGE and Coomassie stain showed that all constructs formed the nanocompartment complex, as indicated by the presence of the signature high molecular weight band that also served as a loading control. Encapsulation of the heterologous sfGFP-fusion cargo was assayed by measuring GFP fluorescence of the high molecular weight band (Figure 2.7C). The entire 225 amino acid N-terminal domain from CyD fused to sfGFP (CyD 1-225-sfGFP) yielded the highest loading signal. Targeting with the first 100 amino acids of the NTD also functioned, albeit not as efficiently as the full N-terminus. The entire cysteine desulfurase fused to sfGFP was also encapsulated, yet again not as well as the 225-NTD. This discrepancy may be due to steric hindrance resulting in fewer copies of the larger full-length CyD-sfGFP construct physically packed inside the compartment. Lastly, CyD 155-183-sfGFP, containing the conserved motif of 'AASPYYFLDG', was not sufficient to sequester cargo within the compartment, nor was the non-tagged sfGFP construct.

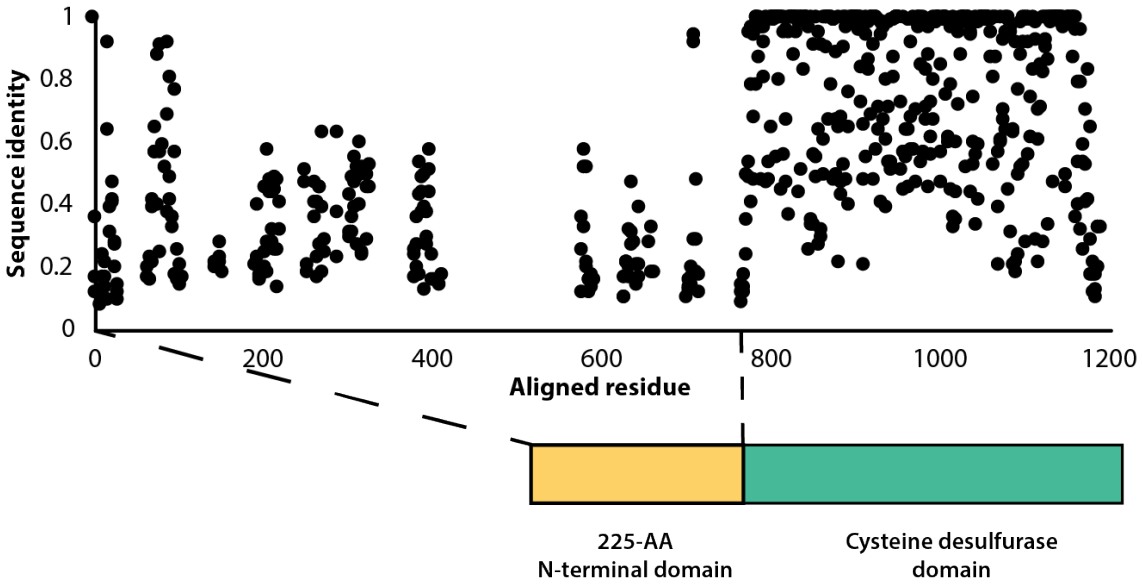

Figure 2.8 Sequence conservation of the F2A encapsulin-associated cysteine desulfurase is sparse throughout the disordered N-terminal domain. Percent sequence identity from multiple sequence alignment of 997 F2A encapsulin-associated cysteine desulfurase sequences.

2.3d The CyD NTD is necessary and sufficient for loading heterologous cargo *in vitro*

Prior work demonstrated that it is possible to assay cargo loading by disassembling the shell protein with a chaotrope, such as guanidine hydrochloride (GuHCl), and re-folding the shell protein in the presence of cargo protein[24]. In this manner, we can control the amount of cargo protein and ensure that loading is due to the targeting sequence rather than mass-action. The shell protein was purified, unfolded in GuHCl, and then refolded by dilution in the presence of purified 225NTD-sfGFP or untagged sfGFP. After refolding and concentration of the sample, the loaded compartment fraction was separated from un-loaded cargo via size exclusion chromatography (Figure 2.9A-C). Again, encapsulation was assayed via SDS-PAGE analysis. Only the 225NTD-sfGFP construct displayed GFP fluorescence in the high molecular weight band, indicative of sfGFP loading (Figure 2.10A). Furthermore, analysis of the compartment and cargo fractions using denaturing SDS-PAGE showed the presence of cargo protein in the compartment fraction only for the 225NTD-sfGFP construct whereas the presence of sfGFP lacking the NTD was only found in the cargo fraction (Figure 2.10B).

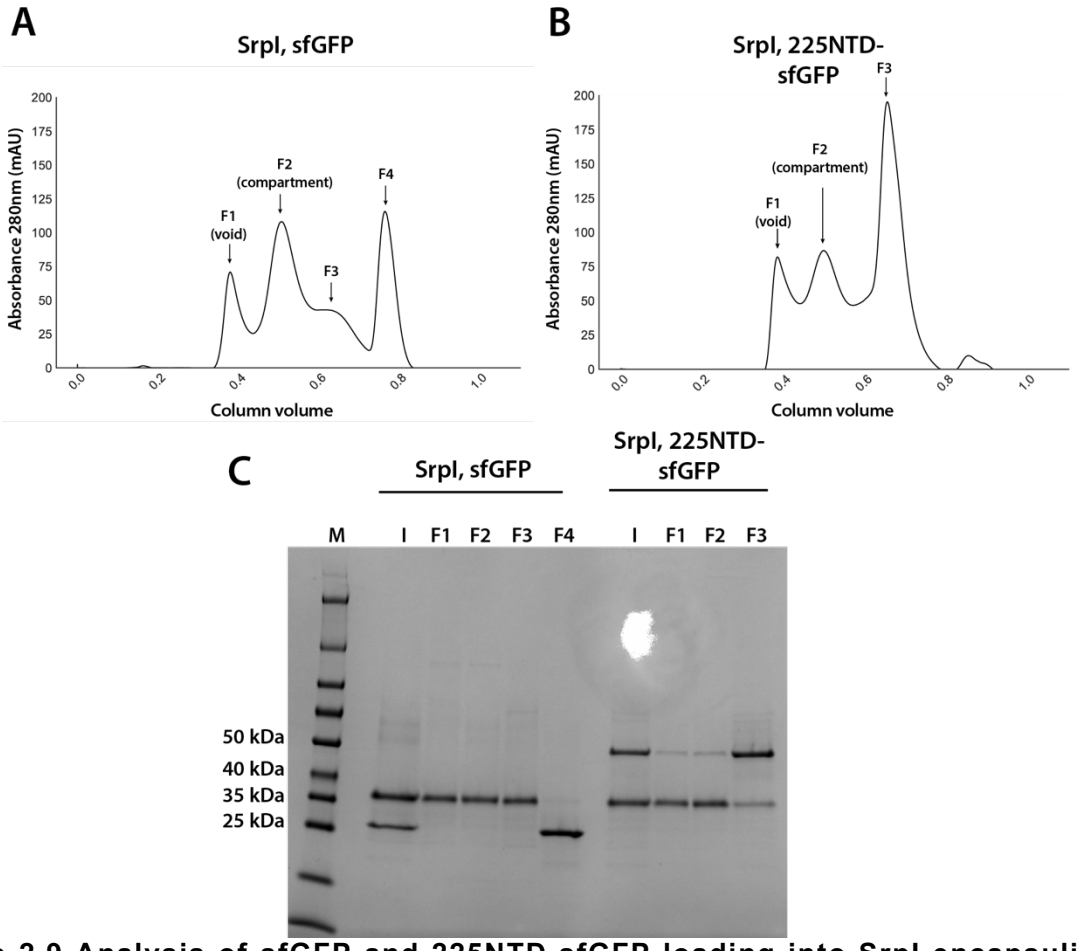

Figure 2.9 Analysis of sfGFP and 225NTD-sfGFP loading into SrpI encapsulin. (A) and (B) size exclusion chromatogram of SrpI encapsulin shell protein (35 kDa) refolded in the presence of untagged sfGFP (27kDa) and 225NTD-sfGFP (50kDa) respectively using a Superose™ 6 Increase column (GE Life Sciences). Fractions are labeled F1-F4 (C) SDS-PAGE analysis of fractions from size exclusion (F1-4), the pre-size exclusion input (I), and a molecular weight marker (M).

We were also able to load the full-length cysteine desulfurase *in vitro* using the same procedure. We observed encapsulation of the native cargo as indicated by the co-elution of cargo in the compartment fraction (Figure 2.10C). Lastly, we found that the disordered NTD is essential for cargo loading. A mutant CyD lacking the entire N-terminal domain (ΔNTD-CyD) was not measurably encapsulated, as evidenced by separate elution of compartment and truncated cargo (Figure 2.10C).

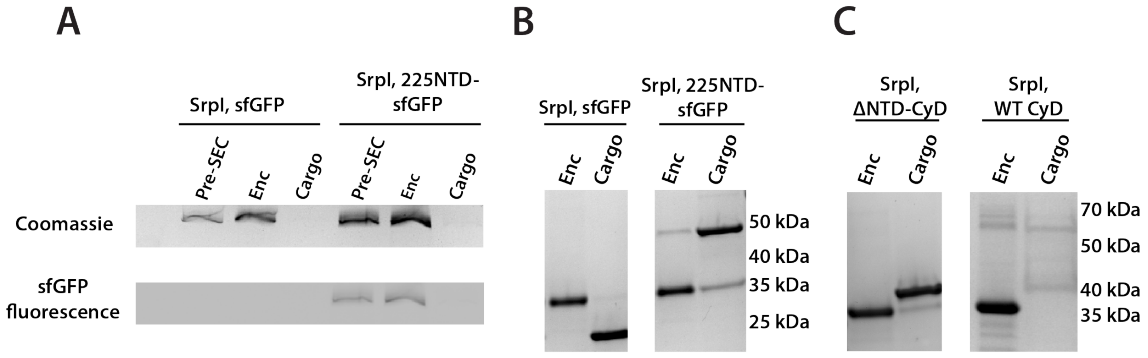

Figure 2.10 The 225-NTD of CyD is necessary and sufficient for cargo loading in vitro. (A) Non-denaturing SDS-PAGE of sfGFP or the CyD N-terminal domain-sfGFP fusion (225NTD-sfGFP) loaded *in vitro* into SrpI encapsulin. sfGFP fluorescence followed by Coomassie staining of the encapsulin (Enc), cargo, and pre size-exclusion chromatography (Pre-SEC) fractions was performed to determine cargo loading and presence of nanocompartment respectively. (B) Denaturing SDS-PAGE of *in vitro* loaded sfGFP and 225NTD-sfGFP samples to determine presence of SrpI shell monomer (35kDa), sfGFP (27kDa), and 225NTD-sfGFP (50kDa) in the encapsulin and cargo fractions from size-exclusion chromatography. (C) Denaturing SDS-PAGE of *in vitro* loaded native cysteine desulfurase (WT CyD) and cysteine desulfurase with the N-terminal domain removed (ΔNTD-CyD) to determine presence of SrpI shell monomer (35kDa), WT CyD (68kDa), and ΔNTD-CyD (45kDa) in the encapsulin and cargo fractions from size-exclusion chromatography.

2.3e Visualization of cargo density by transmission electron microscopy

To further characterize the interaction between SrpI encapsulin and the CyD cargo, we first used negative stain TEM to understand the spatial organization of the CyD cargo within the nanocompartment. Even in raw micrographs, a clear difference in density on the shell interior was observed when comparing the apo-SrpI encapsulin (no cargo) to the holo-SrpI encapsulin (CyD loaded) (Figure 2.11A,C). The apo-SrpI encapsulin structure revealed the shell without any internal cargo density, with modest features clearly visible as expected from a ≈20Å negative stain reconstruction (Figure 2.11B, Methods). The lack of density within the interior of the apo-SrpI capsid also demonstrates that the shell is permissive to uranyl formate stain, which would also allow definition of internal features for the cargo-loaded sample. For the cargo-loaded structure, clear density for the CyD exists at one of the sites of three-fold symmetry within the complex (Figure 2.11D). A difference map, created by subtracting apo-SrpI from the holo-SrpI structure, revealed density corresponding to 1-2 copies of the cysteine desulfurase (Figure 2.11E).

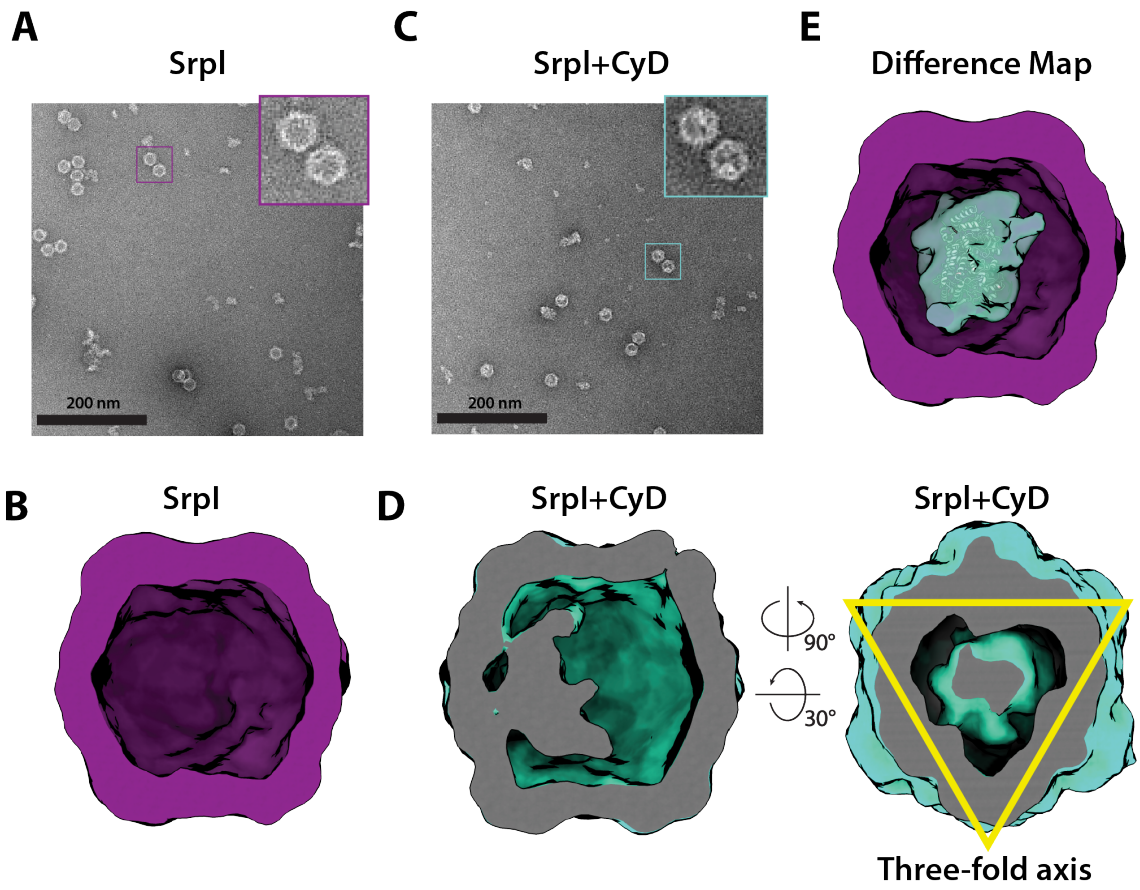

Figure 2.11 Negative stain analysis indicates CyD loading into SrpI encapsulin (A) Negative stain micrograph of an apo-SrpI shell in contrast with (C) the holo-SrpI shell that includes the CyD cargo. (B) and (D) 3D reconstruction of apo-SrpI and holo-SrpI, respectively. (E) Difference map showing additional density for the holo-SrpI with a homologous cysteine desulfurase dimer docked in (pdb:6c9e).

2.3f Structural details of the SrpI shell revealed by cryo-EM

Motivated by our negative stain TEM results, we sought to obtain a high-resolution structure of the nanocompartment complex by single-particle cryo-electron microscopy (cryo-EM). All encapsulin structures published thus far have belonged to the Family 1 encapsulins. These structures are all icosahedral, vary in size from 24-42 nm in diameter, and have a triangulation number of T=1, T=3, or more recently T=4[8,12,16,30]. Cryo-EM analysis was performed on purified holo-SrpI to resolve the shell structure at 2.2 Å resolution. This represents the first Family 2 encapsulin structure and is the highest resolution structure for an encapsulin to date, allowing for accurate atomic model building (Figure 2.12A; Figure 2.13, Figure 2.14). The SrpI encapsulin is 24.5 nm in diameter and has a T=1 icosahedral capsid formed by the self-assembly of 60 SrpI monomeric subunits (Figure 2.12A), similar to previously reported encapsulin structures. Given the structural similarity of the entire shell, it is unsurprising that the SrpI monomer also shares the canonical HK97 fold found in Family 1 encapsulins and Caudovirales bacteriophages (Figure 2.12B; Figure 2.15C).

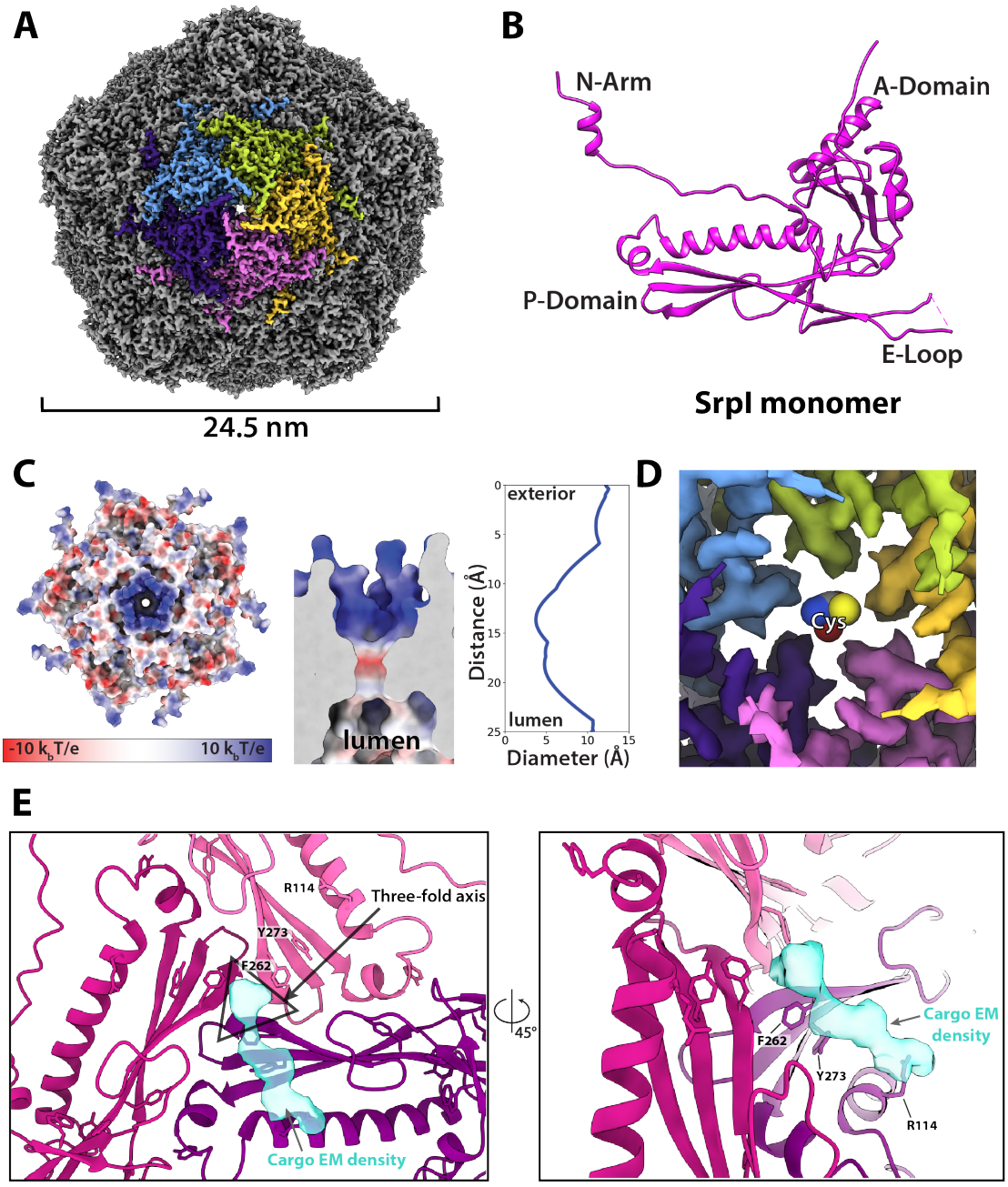

Figure 2.12 CryoEM structure of the SrpI encapsulin reveals a common HK97 fold, a potential mechanism for cysteine selectivity, and a cargo binding site (A) The SrpI encapsulin structure at 2.2 Å resolution. This SrpI encapsulin forms a T=1 icosahedral structure 24.5 nm in diameter. Five subunits around a five-fold axis are shown in distinct colors. (B) SrpI monomer subunits have a HK97 fold with the characteristic A-Domain, E-Loop, P-Domain, and N-Arm. (C) Electrostatic surface potential at the five-fold symmetry axis indicates a relatively neutral pore with an electropositive exterior (left). At its constriction point, the pore is 3.7 Å in diameter. (D) Modeling of a cysteine amino acid at the five-fold axis illustrates the possible mechanism of substrate selection (permissivity to cysteines) by the pore. (E) Unassigned density (turquoise) near the 3-fold axis (grey triangle) revealed by symmetry expansion and focused classification of the holo-SrpI cryoEM map shown in two different orientations.

Most of the topological elements are shared between the monomeric subunits of encapsulins and Caudovirales shells, including the A-domain (axial domain), E-loop (extended loop), and P-domain (peripheral domain). However, unlike other Family 1 encapsulin shell proteins, the Family 2a SrpI possesses an extended N-terminal arm (N-arm) that is more characteristic of bacteriophage structures[31] (Figure 2.15C). Similar to other HK97 bacteriophage capsids, the N-arm of the SrpI shell intertwines with the neighboring subunit to create a chainmail-like topology (Figure 2.16). The most striking differences in quaternary structure between the SrpI encapsulin and the previously studied Family 1 encapsulins can be observed at the major vertices that form the 5-fold axis of symmetry. The vertices of the SrpI encapsulin protrude from the capsid yielding a raised “spike” morphology not found in the Family 1 encapsulins (Figure 2.15D). This difference is due to an extended C-terminus in the SrpI shell protomer that is found near the A-domain, whereas the C-terminus of the Family 1 encapsulins is located farther away from the 5-fold axis.

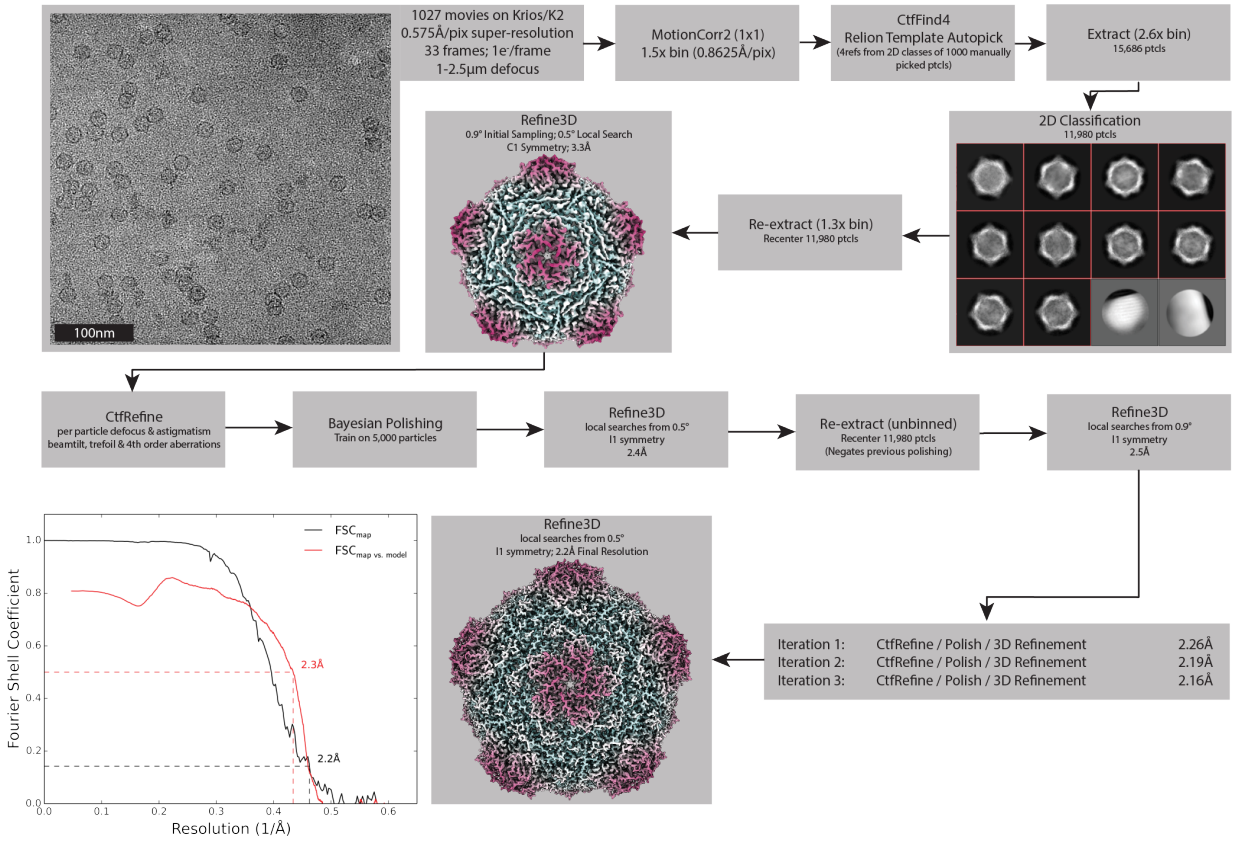

Figure 2.13 Processing Pipeline for the SrpI encapsulin. Processing workflow within RELION that used to reconstruct the holo-SrpI structure. An identical approach was used for the apo-SrpI encapsulin. Once both these structures were determined, symmetry expansion and focused classification was used to compare subtle differences in the density (not shown, see Figure 2.12E).

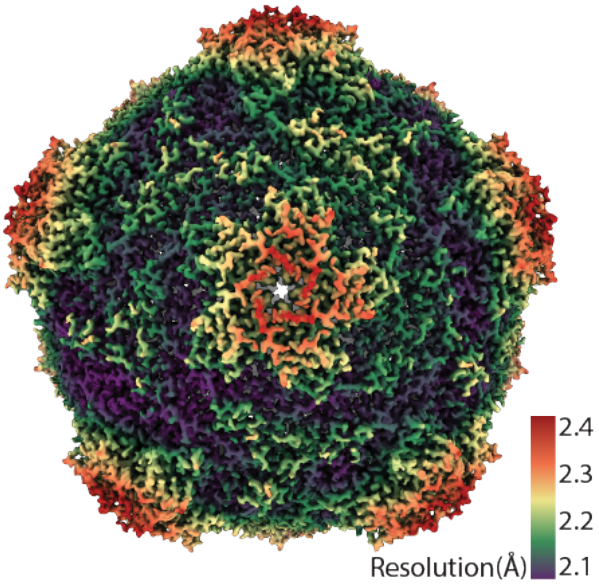

Figure 2.14 CryoEM resolution map of SrpI encapsulin.

Interestingly, the A-domain of SrpI that forms a pore at the 5-fold axis is composed of residues that are positively charged (Figure 2.12C; Figure 2.17), in contrast to the negatively charged five-fold axis pore of Family 1 encapsulins[12]. It is thought that the 5-fold pore at the capsid vertices creates a selective barrier to allow encapsulin substrates into the compartment lumen[8,10,12]. Cysteine is a likely substrate for the SrpI encapsulin given its cysteine desulfurase cargo enzyme. At physiological pH for *S. elongatus* growing in light, pH 8 - 8.4[3], roughly 30-54% of free cysteine will have a net charge of -1 and therefore could traverse the positively charged pore exterior[32]. Likewise, the size of the pore is also an important constraint for limiting the spectrum of substrates that can enter the compartment[33]. The SrpI encapsulin 5-fold pore is 3.7 Å in diameter at its most restrictive point (Figure 2.12C) as calculated by HOLE[34]. Furthermore, modeling of cysteine in the pore demonstrates it is likely small enough to enter the nanocompartment (Figure 2.12D).

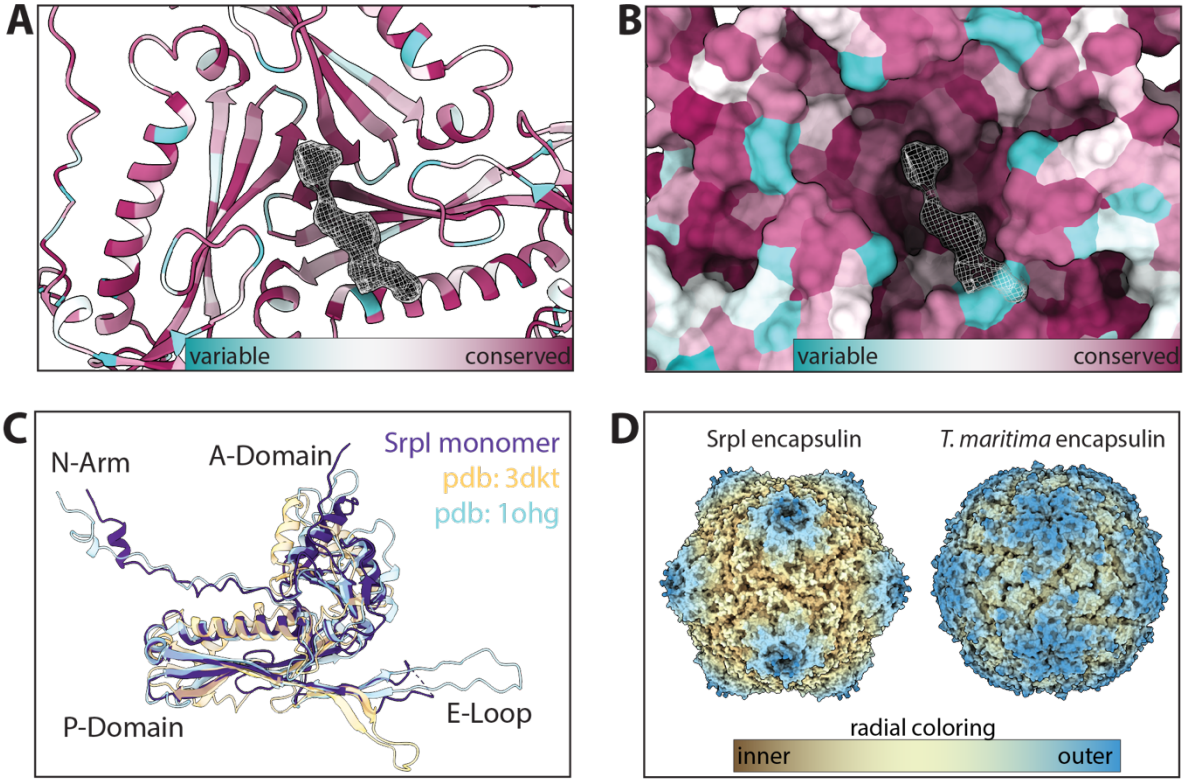

Figure 2.15 Secondary, Tertiary, and Quaternary homology between SrpI and other known encapsulins. (A) and (B) Sequence conservation of Family 2a encapsulins

mapped onto the atomic model and surface display for the Srpl shell respectively (conservation was calculated via ConSurf). (C) single subunit of Srpl compared to other proteins with a known HK97 structure: The Family 1 encapsulin from *T. maritima* (pdb: 3DKT) and the HK97 bacteriophage capsid (pdb: 1OHG) (D) Radial coloring to show the spikes of the Srpl encapsulin morphology in comparison with the *T. maritima* encapsulin.

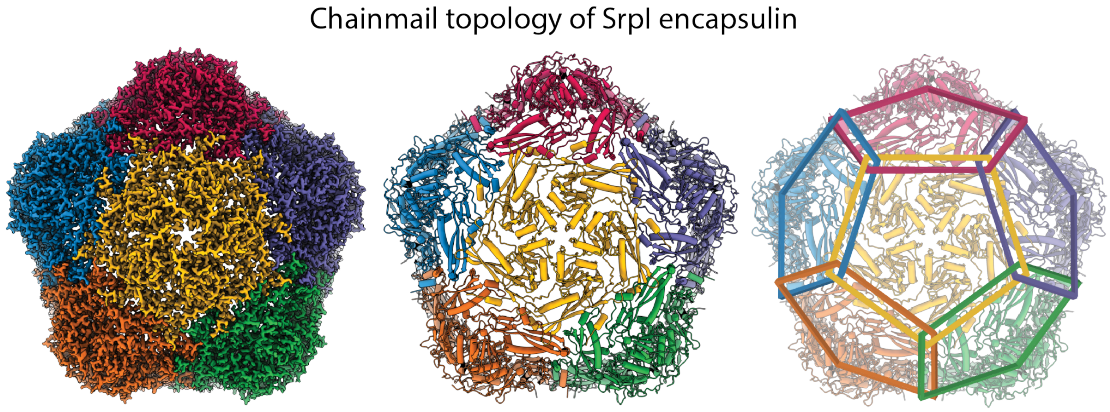

Figure 2.16 Chainmail-like topography of Srpl. Coloring of individual subunits highlights the chainmail overlapping topology of the Srpl encapsulin.

Unfortunately, during processing and classification the holo-Srpl encapsulin proved nearly indistinguishable from the apo-Srpl control (Figure 2.12). The inability to resolve significant portions of the cargo density may be due to low occupancy, or conformational flexibility of the cargo, which disappears at higher resolutions when many particles are averaged together. The inability to fully resolve cargo protein within the nanocompartment has been observed for previously published encapsulin structures[8]. However, symmetry expansion and focused classification of the holo-Srpl encapsulin revealed additional EM density at the interior surface of the shell that was localized to the three-fold symmetry axis (Figure 2.12E). This corroborates our findings from the holo-Srpl structure obtained via negative stain TEM, which also demonstrated cargo interfacing with the shell at the axis of three-fold symmetry (Figure 2.11D).

While the cargo EM density was too weak to accurately build an atomic model of the cysteine desulfurase cargo residues, we were able to determine which shell residues likely interact with cargo density based on their proximity to the putative cargo EM density (Figure 2.12E). Of note, shell residues at the three-fold axis neighboring the cargo EM density are highly conserved, which suggests that the interaction between the encapsulin cargo and shell may be conserved (Figure 2.15 A,B). Namely, residues F262, and Y273 are located near the suggested cargo density.

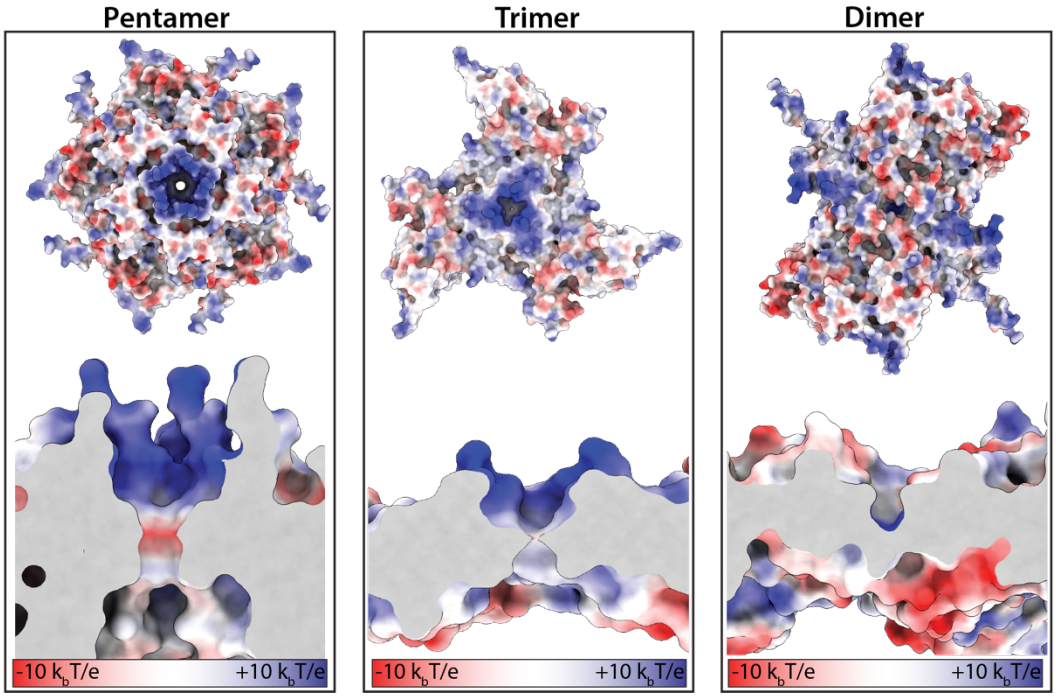

Figure 2.17 Electrostatic surface charges at the symmetry axes of the SrpI shell. Close up and slice-through views of the five-fold, three-fold, and two-fold axes. Surfaces colored according to electrostatic potential (red=negative, blue=positive).

2.3g Encapsulation of CyD modulates enzymatic activity

Finally, we wanted to understand the enzymatic activity of CyD and assess whether encapsulation affects the cargo. Enzyme activity was monitored via NADH fluorescence using an assay coupling cysteine desulfurase, which produces alanine, to alanine dehydrogenase[35]. Unencapsulated CyD utilized cysteine as a substrate and exhibited a k_{cat} of 10 ± 4 s^{-1} (Figure 7). In accordance with our hypothesis that free cysteine could enter the SrpI encapsulin pore, we found that the encapsulated CyD was roughly 7-fold more active than unencapsulated CyD, with a k_{cat} of 67 ± 5 s^{-1} (Figure 7). Importantly, unloaded SrpI encapsulin did not exhibit any cysteine desulfurase activity. Rate enhancement of cargo enzymes by Family 1 encapsulins has been observed previously and is discussed below[36].

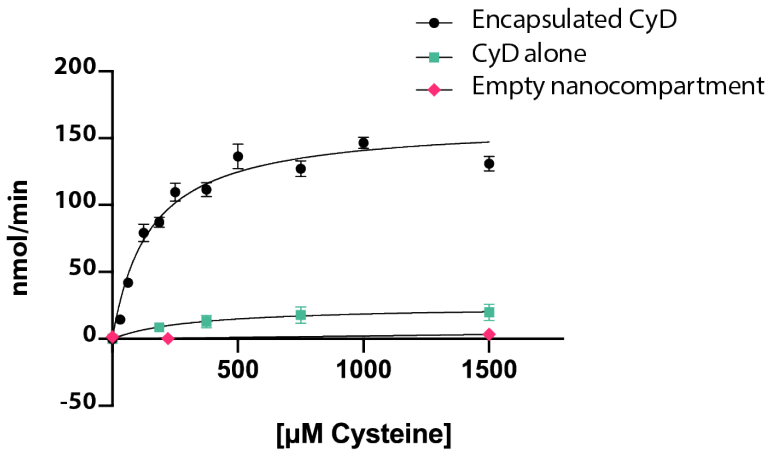

Figure 2.18: Cysteine desulfurase activity is enhanced upon encapsulation. Substrate-dependent activity of encapsulated cysteine desulfurase (encapsulated CyD), unencapsulated cysteine desulfurase (CyD alone), and empty nanocompartment using a coupled-enzyme assay with alanine dehydrogenase and production of NADH as a readout of cysteine desulfurase activity. Error bars are SD of 3-6 replicate experiments.

2.4 Discussion

Here we have identified a unique bacterial nanocompartment and established it as a member of a distinct family of encapsulins, which we name Family 2, that have thus far evaded characterization as a protein-bounded organelle.

2.4a Structural analysis of SrpI encapsulin reveals a potential role as an organelle

We report here the first high-resolution structure for Family 2 encapsulins and find that it shares the HK97 fold found in the Family 1 encapsulins. While there are many structural similarities between the Family 1 and Family 2 encapsulin shell proteins, there are notable differences in the structural properties of the individual domains within the HK97 fold that likely confer distinct functions. One considerable difference between the two encapsulin families is the nature of the pentameric vertex of the capsid that forms the major pore of the compartment. The five-fold pore is likely crucial to the organellular function of encapsulins, as it provides a selective barrier for the entrance of compartment substrates based on size and charge[10,12,33]. Structures determined for the Family 1 encapsulins reveal pores that are negatively charged[8,12,16]. Many of the Family 1 encapsulins possess cargo that binds iron, such as the IMEF and FLP containing encapsulins that have been demonstrated to be capable of iron storage[12,17]. The negative charge of the Family 1 pores is therefore noteworthy because it may allow passage of the positively charged iron ions to enter the encapsulin lumen.

In the Family 2 encapsulin structure reported here, we find the electrostatic charge of the pore to be positive - opposite to what is observed in Family 1. The positive charge of the Family 2 encapsulin is consistent with its likely substrate, cysteine, which will have a net negative charge at physiological pH[3,32]. This contrast between pore charges suggests an overarching theme that may be shared among the encapsulins: the electrostatic charge of the pore is likely reflective of the charge characteristics of the cargo substrate. Furthermore, the size of the pore also appears to be an important parameter as it selects for the entry of substrate molecules while still maintaining a partitioned barrier from larger molecules in the surrounding environment. Initial work has begun towards dissecting how the size of the pore affects mass transport of substrates varying in size. One such study has engineered the pore of the *T. maritima* encapsulin to allow for the diffusion of metals such as terbium, which is nearly double the atomic radius of the native iron substrate[33]. Further mutational studies of pore residues will be needed to better understand how the properties of encapsulin pores affect permeability and function of nanocompartments both *in vitro* and *in vivo*.

2.4b The effect of encapsulation on cargo protein function

Our enzymatic activity data of the CyD cargo provide evidence that encapsulation of the enzyme is important for its activity. We found that the k_{cat} for the encapsulated CyD was almost 7-fold higher than that of the naked cargo *in vitro*. Others have also reported changes in cargo activity upon encapsulation for the Family 1 encapsulins. One such example is the DyP-type peroxidase-containing encapsulin from *Rhodococcus jostii,* which was shown to exhibit an 8-fold higher activity on a lignin substrate compared to the unencapsulated peroxidase[36]. The biochemical function of the FLP-containing encapsulin

is also impeded in the absence of shell protein, as it loses the ability to properly store and mineralize iron[16,17]. The exact mechanism by which encapsulation affects cargo activity remains unknown and may differ for the various cargo types found in encapsulin systems. In the case of the CyD cargo, there are precedents for the enhancement of cysteine desulfurase activity in the presence of accessory proteins. For example, the protein SufE from *E. coli* has been shown to increase the activity of the cysteine desulfurase, SufS, 8 to 50-fold[37,38]. SufE binding stimulates an allosteric change in SufS and enables faster regeneration of the SufS active site by the removal of persulfide from the SufS active site, thereby allowing additional reaction cycles[38,39]. It is possible that SrpI may be acting as an accessory protein for the activity of its cysteine desulfurase cargo analogous to what has been observed for SufE and SufS.

2.4c SrpI may be linked to the canonical sulfur starvation pathway in cyanobacteria

The physiological response to sulfur starvation in cyanobacteria has been studied for decades, yet much is still unknown about the interplay of the known components of the pathway. Genetic and biochemical approaches to study sulfur starvation response in *S. elongatus* have shown that photosynthesis is halted as phycobiliproteins are disassembled from the thylakoid membrane and proteolyzed by the Clp protease complex to generate free amino acids such as cysteine[29,40,41]. Among the other known responses to sulfur starvation are the upregulation of proteins involved in sulfate transport: CysA, CysT, CysW, and SbpA[42,43]. Cyanobacteria are assimilatory sulfate reducers and thus, sulfate is sequentially reduced to sulfite and sulfide by the APS/PAPS pathway, which is then proceeded by the synthesis of L-cysteine via serine-O-acetyltransferase and cysteine synthase[44–46].

Here we report that *S. elongatus* cells deprived of sulfate dramatically upregulate the SrpI encapsulin and its cysteine desulfurase cargo. Furthermore, it is interesting to note that cyanobacteria that possess SrpI encapsulin genes are found in freshwater or brackish water, but not marine environments (Table 2.4). Sulfate is often a limiting nutrient in freshwater environments compared to marine habitats; therefore, the presence of SrpI exclusively in freshwater cyanobacteria suggests SrpI could play a role in sulfur starvation response[47–49]. Our biochemical characterization of this complex showed it is capable of using free L-cysteine as a substrate. Given that the previously characterized facets of the sulfur starvation pathway have been demonstrated to yield free cysteine, our findings suggest a potential link between the SrpI encapsulin and the rest of the known pathway. The physiological role of the SrpI encapsulin may be elucidated by determining the fate of the sulfide group from cysteine after conversion to alanine by the cysteine desulfurase. For example, the sulfur may be mobilized to a downstream sulfide carrier such as NifU or SufE to be incorporated into iron-sulfur clusters or thio-cofactors[50]. Alternatively, the sulfide from cysteine may remain within the compartment in the form of polysulfide. In this second model, the encapsulin may act as a storage cage for sulfur, similar to how Family 1 encapsulins are thought to act as iron stores[12,17].

2.4d Homologs of SrpI are found in pathogens

While this report is the first to recognize SrpI as an encapsulin, previous work has observed a role for SrpI homologs in mycobacterial pathogenesis. Previous work on a SrpI homolog from both *M. leprae* and *M. avium* has shown SrpI is the most antigenic protein in human leprosy patients[51]. Because of its ability to elicit a proliferative T-cell response in leprosy patients, SrpI homologs have been proposed as a useful antigen for vaccine development in disease-causing *Mycobacteria*[52–54]. In their search for candidate antigens from *M. avium* subsp. paratuberculosis, Leite and colleagues enriched for a complex that was identified as MAP2121c and MAP2120c, the SrpI shell and cysteine desulfurase homologs respectively[54]. Our results here validate the finding that these two proteins share a direct biochemical interaction. Moreover, we have identified the N-terminal domain of the cysteine desulfurase to be essential for its interaction with the shell and have presented structural evidence for the encapsulin shell residues with which the cysteine desulfurase may interact. Hopefully our structural and biochemical characterization of the compartment aids in future studies of the role of SrpI encapsulin in pathogenicity and host immune response.

2.4e The evolutionary origins of encapsulins and the prospect of additional undiscovered families

An evolutionary relationship between the encapsulins and Caudovirales bacteriophages is clear given the shared HK97 fold of the capsid proteins. Exactly how the encapsulins and Caudovirales bacteriophages are related, however, remains an open question[19]. It is possible that the HK97 fold derives from a cellular ancestor and was then recruited by a virus-like ancestor of Caudovirales phage[21]. Alternatively, the HK97 fold was of viral origin and cellular hosts co-opted the compartments as organelles that enabled some fitness benefit[19,21]. These two scenarios may not be mutually exclusive, and it is possible that the ancestry of the HK97 fold is intermingled between the two systems with the interconversion of capsids functioning as Caudovirales phage or prokaryotic encapsulins repeatedly over evolutionary history[22]. Our identification of SrpI and its homologs as members of an evolutionarily distinct encapsulin family may provide further insights into the divergence and origin of prokaryotic nanocompartments. Already, the breadth and diversity of known encapsulin systems is vast, yet it is likely that more await discovery.

2.5 Methods

2.5a Phylogenetic analysis of encapsulin genes

Homologs of SrpI were compiled using NCBI BLASTp with query sequence WP_011055154.1 (*Synechococcus elongatus*). BLASTp searches were carried out in February 2020 and hits with an E-value < 0.01 were collected and used in subsequent phylogenetic analysis. Sequences were aligned using MAFFT v7.453 along with the outgroup sequence WP_004080898.1 (*Thermotoga maritima*). The phylogenetic tree was generated from the MAFFT alignment using IQ-TREE[55] with LG model, 4 gamma categories and allowing for invariant sites and ultrafast bootstrap with 1000 replicates. Taxonomy metadata for the encapsulin sequences were compiled using the NCBI protein database. Phylogenetic trees were visualized and annotated using 'The Interactive Tree of Life v4' online server[56].

Genome neighborhood analysis of the Family 2 encapsulin sequences was performed using the Enzyme Function Initiative suite of web tools[57,58]. Sequences were compiled with the Enzyme Similarity Tool (ESI) using WP_011055154.1 (*S. elongatus*) as the query sequence. A Uniprot BLAST search was performed using ESI with an E value of 1E-5 and a maximum of 10,000 sequences. The resulting dataset was then submitted to the Genome Neighborhood Tool to identify the 10 genes upstream and downstream of every Family 2 encapsulin hit.

Secondary structure and disorder prediction of the encapsulin-associated cysteine desulfurase, Synpcc7942_B2661, was performed using PsiPred4 and Disopred3[59,60]. Sequence identity scores for homologs of Synpcc7942_B2661, were determined by aligning sequences using Clustal Omega and analyzing results with Geneius Prime Version 2019.2.1[61]. Conserved motifs within the cysteine desulfurase sequence homologs were determined using the MEME Suite 5.1.1[62]. Using Clustal Omega, the compiled cysteine desulfurase sequences were aligned and truncated to only include the N-terminal domain sequence. These truncated sequences were then analyzed using MEME to create sequence logos of the top occurring motifs.

2.5b Molecular cloning, protein expression and purification

All plasmids were constructed using either Gibson Assembly (NEB) or SLiCE[63] homology-based cloning strategies. Each construct was cloned into a pET-14-based destination vector with a T7 promoter. These constructs were transformed into *E. coli* BL21 (DE3) LOBSTR cells for protein expression[64]. Cells were grown in LB media containing 60 µg/mL kanamycin at 37°C, shaking at 250 rpm to an optical density (OD_{600} = 0.5-0.6) before lowering the temperature to 18°C, inducing with 0.5 mM IPTG, and growing overnight. Liquid cultures were harvested via centrifugation (4000 x *g*, 20 min, 4°C), flash frozen in liquid nitrogen, and stored at -80°C for future use.

All cysteine desulfurase constructs used in enzyme activity and electron microscopy experiments were purified using an N-terminal SUMO tag containing a poly-histidine sequence. Pellets were resuspended in lysis buffer (25 mM Tris-HCl pH 7.5, 150 mM NaCl, 2 0mM imidazole) supplemented with 0.1 mg/ml lysozyme and 1 U/mL

Benzonase® Endonuclease (Millipore Sigma). Sample was lysed with three passages through an Avestin EmulsiFlex-C3 homogenizer and clarified via centrifugation (15,000 x *g*, 30 min, 4°C). The resulting supernatant was then bound to HisPur™ Ni-NTA resin (ThermoFisher Scientific) for 1 hour at 4°C, followed by application of the sample to a gravity column. The resin was then washed with 30 resin volumes of wash buffer (25 mM Tris-HCl pH 7.5, 150mM NaCl, 40mM imidazole) prior to eluting with 25mM Tris-HCl pH 7.5, 150 mM NaCl, 350 mM imidazole. The eluate was then concentrated and desalted into 25 mM Tris-HCl pH 8, 150 mM NaCl using Econo-Pac® 10DG desalting columns (Bio-Rad). The SUMO tag was removed by adding SUMO-protease to the purified sample at a 1:200 (protease: purified protein) molar ratio and allowing cleavage overnight at 4°C. The sample was then further purified by size exclusion chromatography using a Superose™ 6 Increase column (GE Life Sciences) and fractions were analyzed by SDS-PAGE using 4-20% Criterion™ polyacrylamide gels (Bio-Rad) and visualized with GelCode Blue stain (ThermoFisher).

2.5c *In vitro* loading of cargo into SrpI encapsulin

To obtain sufficient quantities of the SrpI shell for *in vitro* loading experiments, the protein was purified from *E. coli* inclusion bodies. Serendipitously, adding a C-terminal 6X-His tag to SrpI yielded high quantities of insoluble shell protein. Purification, denaturation, and folding of shell protein was performed as previously described[65]. Briefly, cell pellets were resuspended in solution buffer (50 mM Tris-HCl pH 8, 1% Triton-X100, 100 mM NaCl, 10 mM DTT, 0.1 mg/ml lysozyme, and 1 U/mL Benzonase® Endonuclease) and lysed with an Avestin EmulsiFlex-C3 homogenizer. The lysate was then centrifuged at 11,000 x *g* for 20 min at 4°C, and the resulting pellet was resuspended in washing buffer A (50mM Tris-HCl pH 8, 0.5% Triton-X100, 100mM NaCl, 10mM DTT) followed by sonication and centrifugation at 11,000 x *g* for 20 min, 4°C. The resulting pellet was then resuspended in washing buffer B (50 mM Tris-HCl pH 8, 100 mM NaCl, 10 mM DTT) followed by sonication and centrifugation again at 11,000 x *g* for 20 min, 4°C. The pellet containing SrpI shell protein was then solubilized with extraction buffer (50 mM Tris-HCl pH 7.4, 6 M guanidine hydrochloride, 50 mM DTT), flash frozen and stored at -80°C for future use.

Refolding was performed by 100-fold dilution in refolding buffer (50 mM CAPS pH 10, 250 mM arginine, 150 mM NaCl, 20% Glycerol, 10 mM DTT). For *in vitro* cargo loading, refolding was performed by adding cargo protein prior to shell protein in a 10:1, cargo: compartment ratio. Sample was concentrated in an Amicon® stirred cell (Millipore Sigma) using a 10kDa MWCO filter, followed by desalting into 50 mM CAPS pH 10, 250 mM arginine, 150 mM NaCl. Subsequent purification was performed using either a Superose™ 6 Increase column or a HiPrep™ 16/60 Sephacryl® S-500 HR (GE Life Sciences).

2.5d *S. elongatus* PCC 7942 growth and sulfate deprivation

S. elongatus PCC 7942 was grown in BG-11 media[66] at 30°C with shaking (185 rpm) under white fluorescent lights at 60-100 µE. After the liquid culture reached log phase (OD_{750} = 0.4-0.5), sulfate starvation was performed by centrifugation of liquid culture (5,000 x *g*, 20 min, 25°C), resuspension of cells in BG-11 media lacking sulfate[29], and repeated for a total of three washes. Control samples were washed and resuspended

using normal BG-11. Samples were then returned to the above growth conditions for continued growth in sulfate dropout BG-11 media. Phycocyanin and chlorophyll levels were quantified by removing 1mL of culture at predetermined times and measuring 620 nm and 680 nm absorbance levels respectively, normalized to cell density at 750 nm.

2.5e Identification of protein complex upregulated under sulfate starvation

Sulfate-starved and control *S. elongatus* PCC 7942 liquid cultures (50mL) were harvested via centrifugation (4000 x *g*, 20 min, 4°C), flash-frozen in liquid nitrogen and stored at -80°C for future processing. Pellets were lysed via sonication and clarified by centrifugation (15,000 x g, 20 min, 4°C). Clarified lysates were analyzed using 4-20% Criterion™ polyacrylamide gels (Bio-Rad) and visualized by silver staining using Pierce™ Silver Stain Kit (ThermoFisher Scientific). Gel bands were excised and sent to UC Davis Proteomics Core Facility. In-gel proteolytic digestion of the samples was performed followed by LC/MS analysis with a Q Exactive™ Hybrid Quadrupole-Orbitrap. Spectra were searched against the *S. elongatus* PCC 7942 proteome and analyzed using Scaffold 4.

2.5f Cysteine desulfurase activity

Cysteine desulfurase activity was performed using a coupled enzyme assay with alanine dehydrogenase. Reactions were carried out at 25°C in 25 mM Tris-HCl pH 8, 150 mM NaCl, 5 mM NAD^+, 8 nM alanine dehydrogenase, 200 nM cysteine desulfurase, and varying L-cysteine concentrations. Cysteine desulfurase concentration for compartment loaded and non-loaded samples was determined by PLP absorbance at 420 nm. Activity was monitored by production of NADH using an Infinite® M1000 plate reader (Tecan) with excitation at 340 nm and emission at 460 nm. Reactions excluding cysteine desulfurase were used as negative controls for background subtraction. Activity data is reported as the initial rate of product formation over substrate concentration and fitted with the Michaelis–Menten equation using GraphPad Prism 8.

2.5g Negative Stain TEM

Holo-SrpI and apo-SrpI samples were diluted to 300 nM in TEM Buffer (50 mM CAPS pH 10, 250 mM arginine, 150 mM NaCl). 4 µL of each sample was placed onto a 400-mesh continuous carbon grid that had been glow discharged (Tergeo, Pie Scientific). After adsorption of the sample onto the grid (2min at room temperature), the sample was stained in five successive rinses with 40 µL droplets Uranyl Formate (UF; 2% w/v in water). To ensure stain was thick, and penetrated the shell interior, the grids sat for 30 sec with a droplet of UF. Grids were briefly side-blotted with Whatman filter paper for 1 seconds, leaving a very thin, but still visible, amount of stain still on the grid, followed by air drying for 10 min. Grids were visualized with an FEI Tecnai F20 electron microscope operating at 120 keV. For each construct, ~100 micrographs were collected on a Gatan Ultrascan 4k CCD camera, at a magnification of 80,000x (1.37Å/pix) and a defocus range from -0.5 to -2.0 µm defocus.

Each dataset was processed identically using RELION[67]. Briefly, CTF estimation was performed using CTFFIND4[68], and particles were picked using RELION's built-in LoG autopicker. Roughly 5,000 particles were extracted for each dataset, binned 4-fold,

followed by 2D classification into 20 classes. All classes that resembled particles (~80% of the initial particles picked), were selected for a final refinement without symmetry imposed using a hollow sphere of 25 nm as a reference. Both apo and holo constructs gave a final resolution estimate of ~18Å.

2.5h CryoEM Sample Preparation, Data Acquisition and Processing

Samples were prepared on UltrAuFoil 1.2/1.3 gold grids (Quantifoil). Grids were initially washed with 2 drops of chloroform and allowed to air dry. Of note, no glow discharge step was performed. A 2 mg/mL solution of the SrpI encapsulin in TEM Buffer supplemented with 0.05% NP-40 was applied to the grid, and immediately plunge-frozen in liquid ethane using a Vitrobot Mark IV (blot force 5, 3 sec blot, 100% humidity, 4°C, 1 sec drain time). For microscope and collection parameters see Table 2.5. Briefly, the holo-SrpI sample was collected on a Titan Krios, and the apo- and sfGFP- samples on Talos Arctica. Data was processed within the RELION pipeline[67], with defocus estimation using CTFFIND4[68]. Particles were automatically picked with LoG-picker and processed in accordance with Supplemental Figure 6-1.

2.5i Atomic model Building and Refinement

The final map for holo-SrpI was the highest resolution of all the states, and was therefore used for model building. An atomic model was built into the density for one asymmetric unit of the SrpI shell with COOT[69], and then refined with nearest neighbors using the real space refinement tool in PHENIX[70]. The MTRIAGE program within PHENIX was used to compute the model vs. map FSC, and HOLE was used to analyze the pores at the 5-fold symmetry axis[34]. For the apo-SrpI and sfGFP-SrpI structures, the shell density was indistinguishable compared to the holo-SrpI, so no model refinement was performed. Instead, these maps were used to guide interpretation of additional density within the shell, and to calculate difference maps between the holo- and apo- states using UCSF Chimera[71]. Additionally, symmetry expansion and focused alignment-free 3D classification were performed with RELION for all states. All coordinates and maps were visualized with UCSF ChimeraX and Pymol[71]; The PyMOL Molecular Graphics System, Version 2.3.2 Shrödinger, LLC).

2.6 Tables

Encapsulin Family	Count
Family 1	3023
Family 2	5540

Table 2.1 Total counts of Family 1 and Family 2 encapsulins found in prokaryotic genomes. Number of total Family 1 and Family 2 homologs compiled using NCBI BLASTp (E-value < 0.01). Accession IDs WP_004080898.1 (*T. maritima* encapsulin shell) and WP_011055154.1 (*S. elongatus* PCC 7942) encapsulin shell genes were used as Family 1 and Family 2 queries respectively. Results based on NCBI's Genome Information resource (February 2020).

Family 2a co-occurrence					
Shared name	**Description**	**pfam**	**Co-occurrence**	**Median gene distance**	**% Occurrence**
aminotran_5	Amino transferase class-V	PF00266	1208 / 1273	1	94.8939513
Hexapep	Bacterial transferase hexapeptide	PF00132	715 / 1273	2	56.16653574
Rhodanese	Rhodanese-like domain	PF00581	620 / 1273	3	48.70384918
BPD_transp_1	Binding-protein-dep	PF00528	483 / 1273	7	37.9418696
abc tran	ABC transporter	PF00005	379 / 1273	6	29.77219167
HTH_1-LysR	Bacterial regulatory helix turn helix	PF00126-PF03466	338 / 1273	4	26.55145326
Hexapep-SATase	Bacterial transferase hexapeptide	PF00132-PF06426	297 / 1273	1	23.33071485
HTH_3	Helix-turn-helix	PF01381	287 / 1273	1	22.54516889
PALP	Pyridoxal-phosphate dependent enzyme	PF00291	287 / 1273	2	22.54516889
N/A	N/A	No assigned PFAM	1058 / 1273	6	82.48232522

Table 2.2 Genome neighborhood analysis of Family 2a shell genes. Co-occurrence and median gene distance of genes found to neighbor Family 2a shell genes using the EFI-GNT web tool. Open reading frames neighboring Family 2a shell genes in the European Nucleic Acid (ENA) database are grouped by shared pfam values.

Family 2b co-occurance table					
shared name	**description**	**pfam**	**co-occurrence**	**median gene distance**	**% occurrence**
acetyltransf_1	Acetyltransferase (GNAT) family	PF00583	463 / 2130	2	21.7370892
cNMP_binding	cyclic nucleotide-binding domain	PF00027	768 / 2130	1	36.05633803
Amindase_2	N-acetylmuramoyl-L-alanine amidase	PF01510	440 / 2130	1	20.657277
polyprenyl_synt	Polyprenyl synthase	PF00348	869 / 2130	1	40.79812207
PALP	Pyridoxal-phosphate dependent enzyme	PF00291	568 / 2130	3	26.66666667
sigma70 r3-Sigma70	Sigma-70 region3	PF04539-PF04542-PF04545	503 2130	6	23.66197183
Na_H_Exchanger	Sodium/hydrogen exchanger	PF00999	463 / 2130	3	21.7370892
az-UBP	Zn-finger in ubiquitin-hydrolases	PF02148	434 / 2130	4	20.37558685
N/A	N/A	No assigned PFAM	2088 / 2130	5	98.02816901

Table 2.3 Genome neighborhood analysis of Family 2b shell genes. Co-occurrence and median gene distance of genes found to neighbor Family 2b shell genes using the EFI-GNT web tool. Open reading frames neighboring Family 2b shell genes in the European Nucleic Acid (ENA) database are grouped by shared pfam values.

NCBI Accession ID	Species	Evironment
AJD58949.1	Synechococcus sp. UTEX2973	Freshwater
WP_011055154.1	Synechococcus sp.	Freshwater
QFZ92646.1	Synechococcus elongatus PCC11802	Freshwater
AZB72703.1	Synechococcus elongatus PCC11801	Freshwater
WP_015125275.1	Synechococcus sp.PCC6312	Freshwater
WP_002735851.1	Microcystis aeruginosa	Freshwater
TRU86312.1	Microcystis novacekii Mn_MB_F_20050700_S1	Freshwater
TRV47998.1	Microcystis panniformis Mp_GB_SS_20050300_S99	Freshwater
WP_104397895.1	Microcystis aeruginosa	Freshwater
GBE76399.1	Microcystis aeruginosaNIES-87	Freshwater
WP_061431234.1	Microcystis aeruginosa	Freshwater
REJ53903.1	Microcystis aeruginosaDA14	Freshwater
WP_002750434.1	Microcystis sp.	Freshwater
NCQ89664.1	Microcystis aeruginosaLG13-13	Freshwater
WP_002786081.1	Microcystis aeruginosa	Freshwater
WP_014276787.1	Arthrospira platensis	Freshwater
WP_006670576.1	Limnospira	Freshwater
WP_152088332.1	Arthrospira platensis	Freshwater
WP_006619903.1	Arthrospira platensis	Freshwater
WP_054465623.1	Planktothricoides sp.SR001	Freshwater
WP_006625603.1	Arthrospira platensis	Freshwater
WP_130756769.1	Microcystis aeruginosa	Freshwater
KEF42600.1	Cyanobium sp.CACIAM14	Freshwater
WP_106222825.1	Aphanothece minutissima	Freshwater
WP_015108669.1	Cyanobium sp.	Freshwater
WP_094554187.1	Synechococcus sp.1G10	Freshwater
WP_048016926.1	Synechococcus sp.GFB01	Freshwater
GAL91719.1	Microcystis aeruginosaNIES-44	Freshwater
WP_008180793.1	Moorea producens	Brackish water
WP_071103433.1	Moorea producens	Brackish water
WP_070391731.1	Moorea producens	Brackish water
WP_075897528.1	Moorea bouillonii	Brackish water
WP_008180789.1	Moorea producens	Brackish water
WP_071103431.1	Moorea producens	Brackish water

Table 2.4 Family 2a shell genes are found in freshwater and brackish water cyanobacteria, but not marine cyanobacteria. NCBI BLASTp results of Family2a shell homologs found in cyanobacteria. Results based on NCBI's Genome Information resource (February 2020). Environment annotations based on the Joint Genome Institute (JGI) Integrated Microbial Genomes and Microbiomes (IMG/M) database and (Shih et al., 2013).

Data collection, 3D reconstruction, and refinement statistics.

Dataset	**Holo-SrpI**	**Apo-SrpI**	**SrpI-sfGFP**	**Holo-SrpI (neg stain)**	**Apo-SrpI (neg stain)**
Microscope	Titan Krios	Talos Arctica	Talos Arctica	Tecnai F20	Tecnai F20
Stage type	Autoloader	Autoloader	Autoloader	Side entry	Side entry
Voltage (kV)	300	200	200	120	120
Detector	Gatan K2	Gatan K3	Gatan K3	Gatan UltraScan	Gatan UltraScan
Data Collection Software	SerialEM	SerialEM	SerialEM	Leginon	Leginon
Acquisition mode	Super-resolution	Super-resolution	Super-resolution	CCD	CCD
Physical pixel size (Å)	0.575	0.569	0.569	1.37	1.37
Defocus range (μm)	1.0-2.5	1.0-2.5	1.0-2.5	0.4-1.5	0.4-1.5
Electron exposure (e^-/A^2)	35	40	40	50	50
Reconstruction	**EMD-XXXX**				
Session	19Jan30	19Sep22	19Apr09	19Jul09	19Jul09
Software	RELION 3.1	RELION 3.1	RELION 3.1	RELION 3.1	RELION 3.1
Particles picked	15,686	57,341	60,913	5,901	4,807
Particles final	11,980	34,425	36,618	5,901	4,807
Extraction box size (pixels)	512x512x512	360x360x360	360x360x360	256x256x256	256x256x256
Final pixel size (Å)	0.8625	1.14	1.14	5.48	5.48
Accuracy rotations (°)	0.24	0.36	0.34	1.1	1.3
Accuracy translations (Å)	0.29	0.4	0.4	1.35	1.5
Map resolution (Å)	2.2	3.1	3.3	~18	~18
Map sharpening B-factor ($Å^2$)	-35	-70	-70	NA	NA
Coordinate refinement					
Software	PHENIX				
Refinement algorithm	REAL SPACE				
Resolution cutoff (Å)	2.2				
$FSC_{model\text{-}vs\text{-}map}$=0.5 (Å)	2.3				
Model	**PDB-XXXX**				
Number of residues	280				
B-factor overall	20				
R.m.s. deviations					
Bond lengths (Å)	0.004				
Bond angles (°)	0.598				
Validation					
Molprobity clashscore	4.1				
Rotamer outliers (%)	0.8				
C_β deviations (%)	0.0				
Ramachandran plot					
Favored (%)	96.7				
Allowed (%)	3.3				
Outliers (%)	0.0				

Table 2.5 Data collection, reconstruction, and processing statistics.

2.7 Acknowledgements

We thank Avi Flamholz and Cecilia Blikstad for helpful discussions and critical reading of the manuscript. We acknowledge the UC Davis Proteomics Core Facility for mass spectrometry data collection and the UC Berkeley Electron Microscope Laboratory for training with TEM. We thank Patricia Dos Santos for helpful discussions regarding cysteine desulfurase biochemistry. We thank Patricia Grob and Daniel Toso for microscope support and Abhiram Chintangal for computational support. This work was supported by a grant from the US Department of Energy (no. DE-SC00016240) to D.F.S., B.L. was supported by an NSF-GRFP grant (no. 1106400), and E.N. is a Howard Hughes Medical Institute Investigator.

2.8 References

1. Grant, C. R., Wan, J. & Komeili, A. Organelle Formation in Bacteria and Archaea. *Annu. Rev. Cell Dev. Biol.* **34**, 217–238 (2018).
2. Oltrogge, L. M. *et al.* Multivalent interactions between CsoS2 and Rubisco mediate α-carboxysome formation. *Nat. Struct. Mol. Biol.* (2020) doi:10.1038/s41594-020-0387-7.
3. Mangan, N. M., Flamholz, A., Hood, R. D., Milo, R. & Savage, D. F. pH determines the energetic efficiency of the cyanobacterial CO2 concentrating mechanism. *Proc. Natl. Acad. Sci. U. S. A.* **113**, E5354–62 (2016).
4. Kerfeld, C. A., Aussignargues, C., Zarzycki, J., Cai, F. & Sutter, M. Bacterial microcompartments. *Nat. Rev. Microbiol.* **16**, 277–290 (2018).
5. Sampson, E. M. & Bobik, T. A. Microcompartments for B12-dependent 1,2-propanediol degradation provide protection from DNA and cellular damage by a reactive metabolic intermediate. *J. Bacteriol.* **190**, 2966–2971 (2008).
6. Crowley, C. S. *et al.* Structural insight into the mechanisms of transport across the Salmonella enterica Pdu microcompartment shell. *J. Biol. Chem.* **285**, 37838–37846 (2010).
7. Chowdhury, C. *et al.* Selective molecular transport through the protein shell of a bacterial microcompartment organelle. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 2990–2995 (2015).
8. Sutter, M. *et al.* Structural basis of enzyme encapsulation into a bacterial nanocompartment. *Nat. Struct. Mol. Biol.* **15**, 939–947 (2008).
9. Giessen, T. W. & Silver, P. A. Converting a Natural Protein Compartment into a Nanofactory for the Size-Constrained Synthesis of Antimicrobial Silver Nanoparticles. *ACS Synth. Biol.* **5**, 1497–1504 (2016).
10. Nichols, R. J., Cassidy-Amstutz, C., Chaijarasphong, T. & Savage, D. F. Encapsulins: molecular biology of the shell. *Crit. Rev. Biochem. Mol. Biol.* **52**, 583–594 (2017).
11. Giessen, T. W. & Silver, P. A. Widespread distribution of encapsulin nanocompartments reveals functional diversity. *Nat Microbiol* **2**, 17029 (2017).
12. Giessen, T. W. *et al.* Large protein organelles form a new iron sequestration system with high storage capacity. *Elife* **8**, (2019).
13. Xing, C.-Y. *et al.* A self-assembled nanocompartment in anammox bacteria for resisting intracelluar hydroxylamine stress. *Sci. Total Environ.* **717**, 137030 (2020).
14. Kim, D. *et al.* Operon required for fruiting body development in Myxococcus xanthus. *J. Microbiol. Biotechnol.* **19**, 1288–1294 (2009).
15. Kim, D. *et al.* Mutants defective in the production of encapsulin show a tan-phase-locked phenotype in Myxococcus xanthus. *J. Microbiol.* **57**, 795–802 (2019).
16. McHugh, C. A. *et al.* A virus capsid-like nanocompartment that stores iron and protects bacteria from oxidative stress. *EMBO J.* **33**, 1896–1911 (2014).
17. He, D. *et al.* Structural characterization of encapsulated ferritin provides insight into iron storage in bacterial nanocompartments. *Elife* **5**, (2016).
18. Kartal, B. *et al.* How to make a living from anaerobic ammonium oxidation. *FEMS Microbiol. Rev.* **37**, 428–461 (2013).
19. Koonin, E. V. & Krupovic, M. The depths of virus exaptation. *Curr. Opin. Virol.* **31**, 1–8 (2018).

20. Krupovic, M., Dolja, V. V. & Koonin, E. V. Origin of viruses: primordial replicators recruiting capsids from hosts. *Nat. Rev. Microbiol.* **17**, 449–458 (2019).
21. Krupovic, M. & Koonin, E. V. Multiple origins of viral capsid proteins from cellular ancestors. *Proc. Natl. Acad. Sci. U. S. A.* **114**, E2401–E2410 (2017).
22. Radford, D. Understanding the encapsulins: prediction and characterization of phage capsid-like nanocompartments in prokaryotes. (2015).
23. Giessen, T. W. Encapsulins: microbial nanocompartments with applications in biomedicine, nanobiotechnology and materials science. *Curr. Opin. Chem. Biol.* **34**, 1–10 (2016).
24. Cassidy-Amstutz, C. *et al.* Identification of a Minimal Peptide Tag for in Vivo and in Vitro Loading of Encapsulin. *Biochemistry* **55**, 3461–3468 (2016).
25. Snijder, J. *et al.* Assembly and Mechanical Properties of the Cargo-Free and Cargo-Loaded Bacterial Nanocompartment Encapsulin. *Biomacromolecules* **17**, 2522–2529 (2016).
26. Nicholson, M. L., Gaasenbeek, M. & Laudenbach, D. E. Two enzymes together capable of cysteine biosynthesis are encoded on a cyanobacterial plasmid. *Mol. Gen. Genet.* **247**, 623–632 (1995).
27. Nicholson, M. L. & Laudenbach, D. E. Genes encoded on a cyanobacterial plasmid are transcriptionally regulated by sulfur availability and CysR. *J. Bacteriol.* **177**, 2143–2150 (1995).
28. Chen, Y., Holtman, C. K., Magnuson, R. D., Youderian, P. A. & Golden, S. S. The complete sequence and functional analysis of pANL, the large plasmid of the unicellular freshwater cyanobacterium Synechococcus elongatus PCC 7942. *Plasmid* **59**, 176–192 (2008).
29. Collier, J. L. & Grossman, A. R. Chlorosis induced by nutrient deprivation in Synechococcus sp. strain PCC 7942: not all bleaching is the same. *J. Bacteriol.* **174**, 4718–4726 (1992).
30. Akita, F. *et al.* 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 (2007).
31. Duda, R. L. & Teschke, C. M. The amazing HK97 fold: versatile results of modest differences. *Curr. Opin. Virol.* **36**, 9–16 (2019).
32. Cameselle, J. C., Ribeiro, J. M. & Sillero, A. Derivation and use of a formula to calculate the net charge of acid-base compounds. Its application to amino acids, proteins and nucleotides. *Biochem. Educ.* **14**, 131–136 (1986).
33. Williams, E. M., Jung, S. M., Coffman, J. L. & Lutz, S. Pore Engineering for Enhanced Mass Transport in Encapsulin Nanocompartments. *ACS Synth. Biol.* **7**, 2514–2517 (2018).
34. Smart, O. S., Neduvelil, J. G., Wang, X., Wallace, B. A. & Sansom, M. S. HOLE: a program for the analysis of the pore dimensions of ion channel structural models. *J. Mol. Graph.* **14**, 354–60, 376 (1996).
35. Dos Santos, P. C. Chapter Seven - B. subtilis as a Model for Studying the Assembly of Fe–S Clusters in Gram-Positive Bacteria. in *Methods in Enzymology* (ed. David, S. S.) vol. 595 185–212 (Academic Press, 2017).

36. Rahmanpour, R. & Bugg, T. D. H. Assembly in vitro of Rhodococcus jostii RHA1 encapsulin and peroxidase DypB to form a nanocompartment. *FEBS J.* **280**, 2097–2104 (2013).
37. Loiseau, L., Ollagnier-de-Choudens, S., Nachin, L., Fontecave, M. & Barras, F. Biogenesis of Fe-S cluster by the bacterial Suf system: SufS and SufE form a new type of cysteine desulfurase. *J. Biol. Chem.* **278**, 38352–38359 (2003).
38. Outten, F. W., Wood, M. J., Munoz, F. M. & Storz, G. The SufE protein and the SufBCD complex enhance SufS cysteine desulfurase activity as part of a sulfur transfer pathway for Fe-S cluster assembly in Escherichia coli. *J. Biol. Chem.* **278**, 45713–45719 (2003).
39. Singh, H., Dai, Y., Outten, F. W. & Busenlehner, L. S. Escherichia coli SufE sulfur transfer protein modulates the SufS cysteine desulfurase through allosteric conformational dynamics. *J. Biol. Chem.* **288**, 36189–36200 (2013).
40. Baier, A., Winkler, W., Korte, T., Lockau, W. & Karradt, A. Degradation of phycobilisomes in Synechocystis sp. PCC6803 evidence for essential formation of an NblA1/NblA2 heterodimer and its codegradation by a Clp protease complex. *J. Biol. Chem.* **289**, 11755–11766 (2014).
41. Karradt, A., Sobanski, J., Mattow, J., Lockau, W. & Baier, K. NblA, a key protein of phycobilisome degradation, interacts with ClpC, a HSP100 chaperone partner of a cyanobacterial Clp protease. *J. Biol. Chem.* **283**, 32394–32403 (2008).
42. Green, L. S., Laudenbach, D. E. & Grossman, A. R. A region of a cyanobacterial genome required for sulfate transport. *Proc. Natl. Acad. Sci. U. S. A.* **86**, 1949–1953 (1989).
43. Laudenbach, D. E. & Grossman, A. R. Characterization and mutagenesis of sulfur-regulated genes in a cyanobacterium: evidence for function in sulfate transport. *J. Bacteriol.* **173**, 2739–2750 (1991).
44. Kopriva, S., Patron, N. J., Keeling, P. & Leustek, T. Phylogenetic Analysis of Sulfate Assimilation and Cysteine Biosynthesis in Phototrophic Organisms. in *Sulfur Metabolism in Phototrophic Organisms* (eds. Hell, R., Dahl, C., Knaff, D. & Leustek, T.) 31–58 (Springer Netherlands, 2008).
45. Schmidt, A. 21 - Sulphur Metabolism: D. Cysteine Synthase. in *Methods in Plant Biochemistry* (ed. Lea, P. J.) vol. 3 349–354 (Academic Press, 1990).
46. Schmidt, A. & Christen, U. A factor-dependent sulfotransferase specific for 3'-phosphoadenosine-5'-phosphosulfate (PAPS) in the Cyanobacterium Synechococcus 6301. *Planta* **140**, 239–244 (1978).
47. Giordano, M., Norici, A. & Hell, R. Sulfur and phytoplankton: acquisition, metabolism and impact on the environment. *New Phytol.* **166**, 371–382 (2005).
48. Pilson, M. E. Q. *An Introduction to the Chemistry of the Sea*. (Cambridge University Press, 2012).
49. Tipping, E. *et al.* Reversal of acidification in upland waters of the English Lake District. *Environ. Pollut.* **103**, 143–151 (1998).
50. Black, K. A. & Dos Santos, P. C. Shared-intermediates in the biosynthesis of thio-cofactors: Mechanism and functions of cysteine desulfurases and sulfur acceptors. *Biochim. Biophys. Acta* **1853**, 1470–1480 (2015).
51. Winter, N. *et al.* Characterization of the gene encoding the immunodominant 35 kDa protein of Mycobacterium leprae. *Mol. Microbiol.* **16**, 865–876 (1995).

52. Abdellrazeq, G. S. *et al.* A Mycobacterium avium subsp. paratuberculosis relA deletion mutant and a 35 kDa major membrane protein elicit development of cytotoxic T lymphocytes with ability to kill intracellular bacteria. *Vet. Res.* **49**, 53 (2018).
53. Abdellrazeq, G. S. *et al.* Simultaneous cognate epitope recognition by bovine CD4 and CD8 T cells is essential for primary expansion of antigen-specific cytotoxic T-cells following ex vivo stimulation with a candidate Mycobacterium avium subsp. paratuberculosis peptide vaccine. *Vaccine* (2020).
54. Leite, F. L., Reinhardt, T. A., Bannantine, J. P. & Stabel, J. R. Envelope protein complexes of Mycobacterium avium subsp. paratuberculosis and their antigenicity. *Vet. Microbiol.* **175**, 275–285 (2015).
55. Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A. & Jermiin, L. S. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* **14**, 587–589 (2017).
56. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res.* **47**, W256–W259 (2019).
57. Gerlt, J. A. Genomic Enzymology: Web Tools for Leveraging Protein Family Sequence–Function Space and Genome Context to Discover Novel Functions. *Biochemistry* **56**, 4293–4308 (2017).
58. Zallot, R., Oberg, N. & Gerlt, J. A. The EFI Web Resource for Genomic Enzymology Tools: Leveraging Protein, Genome, and Metagenome Databases to Discover Novel Enzymes and Metabolic Pathways. *Biochemistry* **58**, 4169–4182 (2019).
59. Buchan, D. W. A. & Jones, D. T. The PSIPRED Protein Analysis Workbench: 20 years on. *Nucleic Acids Res.* **47**, W402–W407 (2019).
60. Jones, D. T. & Cozzetto, D. DISOPRED3: precise disordered region predictions with annotated protein-binding activity. *Bioinformatics* **31**, 857–863 (2015).
61. Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, (2011).
62. Bailey, T. L. *et al.* MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* **37**, W202–8 (2009).
63. Zhang, Y., Werling, U. & Edelmann, W. SLiCE: a novel bacterial cell extract-based DNA cloning method. *Nucleic Acids Res.* **40**, e55 (2012).
64. Andersen, K. R., Leksa, N. C. & Schwartz, T. U. Optimized E. coli expression strain LOBSTR eliminates common contaminants from His-tag purification. *Proteins: Struct. Funct. Bioinf.* **81**, 1857–1861 (2013).
65. Palmer, I. & Wingfield, P. T. Preparation and Extraction of Insoluble (Inclusion-Body) Proteins from Escherichia coli. *Curr. Protoc. Protein Sci.* (2012).
66. Allen, M. M. SIMPLE CONDITIONS FOR GROWTH OF UNICELLULAR BLUE-GREEN ALGAE ON PLATES 1, 2. *J. Phycol.* **4**, 1–4 (1968).
67. Scheres, S. H. W. A Bayesian view on cryo-EM structure determination. *J. Mol. Biol.* **415**, 406–418 (2012).
68. Rohou, A. & Grigorieff, N. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192**, 216–221 (2015).
69. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010).

70. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221 (2010).
71. Goddard, T. D. *et al.* UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Sci.* **27**, 14–25 (2018).
72. Shih, P. M. *et al.* Improving the coverage of the cyanobacterial phylum using diversity-driven genome sequencing. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 1053–1058 (2013).

Chapter 3

A nanocompartment system containing the peroxidase DypB contributes to defense against oxidative stress in *M. tuberculosis*.

†The work presented in this chapter is adapted from a manuscript in preparation with permission from coauthors:
Lien KL, Nichols RJ, Cassidy-Amstutz C, Dinshaw K, Knight M, Savage DF, Stanley SA (2020) A nanocompartment system containing the peroxidase DypB contributes to defense against oxidative stress in *M. tuberculosis*. *(In preparation)*

3.1 Abstract

Encapsulin nanocompartments are an emerging class of prokaryotic protein-based organelles consisting of an encapsulin protein shell that encloses a protein cargo[1]. Genes encoding nanocompartments are widespread in bacteria and archaea, and recent works have characterized the biochemical function of several cargo enzymes[2]. However, the importance of these organelles to the physiology of the endogenous host is poorly understood. Here, we report that the human pathogen *Mycobacterium tuberculosis* (Mtb) produces a nanocompartment that contains the dye-decolorizing peroxidase DypB. We show that this nanocompartment is important for the ability of Mtb to resist oxidative stress in low pH environments, including during infection of host cells and upon treatment with a clinically relevant antibiotic. Our findings are the first to implicate a nanocompartment in bacterial pathogenesis and reveal a new mechanism that Mtb uses to combat oxidative stress.

3.2 Introduction:

Tuberculosis is an infectious disease that has plagued humans for millennia[3,4]. Today tuberculosis persists as a public health problem and continues to be a world-wide epidemic. In 2018, 1.5 million people died from tuberculosis and the World Health Organization estimates that one quarter of the world population is infected with latent tuberculosis[5]. The bacterium *Mycobacterium tuberculosis* (Mtb) is the causative agent of tuberculosis. While the incidence of tuberculosis infection is decreasing world-wide by 2% per year, there is growing concern that the drugs used to clinically treat tuberculosis will rapidly become ineffective[5]. One in every four recorded cases of tuberculosis exhibit resistance to one of the four common antibiotics used to target Mtb: rifampicin, isoniazid, ethambutal, or pyrazinamide[6]. Furthermore, Mtb has evolved numerous evasion strategies in the ongoing cellular arms-race between the host immune system and the intracellular pathogen[7].

The growing need for new solutions to treat Mtb highlights the importance of studying the physiology and cell biology of Mtb as well its interplay with the host immune system. Recently, it was discovered that Mtb possess a prokaryotic organelle known as encapsulin[8]. Encapsulins, also known as prokaryotic nanocompartments, are a protein-bounded subcellular structure that hosts an enzymatic reaction excluded from the cytosol of the rest of the cell[1]. Nanocompartments are comprised of 60-240 shell subunits and assemble into a capsid-like structure roughly 24-45 nm in diameter[9–11]. This shell protein encapsulates an enzymatic cargo protein as directed by site-specific binding of the cargo to the lumenal face of the shell that forms the compartment interior[2,9,12].

The shell forming protein in Mtb, CFP29, has been studied for many years in absence of the knowledge that it was a subcellular compartment. Upon studying proteins found in the culture filtrate of Mtb growth media, CFP29 (culture filtrate protein, 29 kDa), was identified as an antigen that elicited an immune response in mouse memory effector cells exposed to Mtb[13,14]. Additionally, it was shown that memory of CFP29 exposure in mice, stimulated interferon-gamma signaling in the mouse memory effector cells and provided a form of

protective immunity to Mtb infection[14]. It was not until 2014, however, that CFP29 became appreciated as an encapsulin. Contreras and colleagues were able to demonstrate that CFP29 formed an icosahedral nanocompartment similar to previously characterized encapsulins[8]. A homolog of CFP29 in *Brevibacterium linens* had previously been shown to encapsulate a dye-decolorizing peroxidase DypB via a C-terminal aliphatic targeting sequence. It was found that three proteins in the Mtb proteome possessed similar C-termini; the shell-neighboring gene encoding DypB, and two genes found outside of the shell-encoding operon – the iron storing ferritin protein (BfrB) and the folate biosynthesis protein 7,8-dihydroneopterin aldolase (FolB)[8].

While it is clear that CFP29 is an encapsulin and may play an important role in host immune response, the exact mechanism of the Mtb encapsulin's interaction with the host immune system remains unclear. The physiology of prokaryotic nanocompartments in general is poorly understood. Although putative encapsulin systems have been identified in >11,00 bacterial and archaeal genomes[1,2,12,15], very little is known about their endogenous function. Based on genomic organization, encapsulin systems are often predicted to compartmentalize enzymes involved in oxidative stress defense, iron storage, and anaerobic ammonium oxidation[2]. Surprisingly, the endogenous function of encapsulin systems has only been well-characterized in a single bacterial species, *Myxococcus xanthus*[10], and the physiological role of nanocompartments is largely unexplored.

Here we use a genetics and biochemistry-based approach to understand the relationship between the Mtb encapsulin and the host immune system.

3.3 Results

3.3a Structural characterization of encapsulins purified natively from Mtb

The *Mycobacterium tuberculosis* genome encodes the predicted encapsulin gene *Rv0798c/*Cfp29[8] within a two-gene operon with *Rv0799c*, the dye-decolorizing peroxidase DypB (Figure 3.1A). Overexpression of the predicted Mtb encapsulin gene in *Escherichia coli* was previously shown to result in the formation of nanocompartment-like structures[8]. Three potential cargo proteins for the nanocompartment were proposed based upon a putative shared encapsulation targeting sequence: DypB, FolB, and BrfB. In *E. coli*, overexpression of each protein with Cfp29 resulted in encapsulation. However, the previous study did not address whether Mtb produces endogenous nanocompartments or identify the specific function of these compartments in Mtb biology. Interestingly, Cfp29 was previously identified as an immunodominant T cell antigen in both mice and human TB patients[13]. Furthermore, Cfp29 was identified in a transposon screen as being required for growth in mice[16]. Taken together, these results suggest that Mtb might produce an encapsulin that is important for infection.

To confirm the previous finding that *Rv0798c* and *Rv0799c* form an encapsulin system that can assemble when heterologously expressed in a host species, we expressed these genes in *E. coli* and isolated nanocompartments. Protein lysates were clarified by centrifugation and further purified by ultracentrifugation and size exclusion

chromatography. Assembled encapsulin nanocompartments are distinguishable by their high-molecular weight[13]. Indeed, a fraction from the sucrose purification was found to contain a high-molecular weight species >260 kDa observable on an SDS-PAGE gel (Figure 3.2A). Fractions containing putative nanocompartments were pooled and imaged using transmission electron microscopy, which revealed the presence of nanocompartment-like icosahedral structures with the expected diameter of ~25 nm (Figure 3.1B,C).

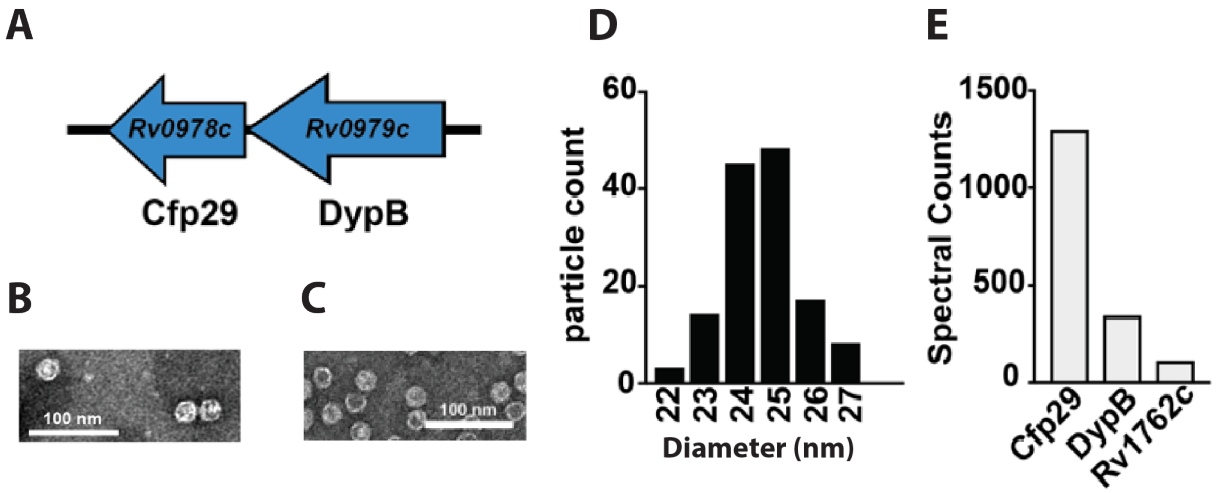

Figure 3.1 Natively purified Mtb encapsulin and heterologously purified CFP encapsulin form a prokaryotic nanocompartment. A) Schematic of nanocompartment genomic locus in Mtb, which encodes the encapsulin shell protein (Cfp29) and the dye-decolorizing peroxidase cargo protein (DypB). B) Negative-stain TEM of Mtb encapsulin protein purified natively from Mtb cells. C) Negative-stain TEM of Mtb encapsulin proteins purified following heterologous expression of the Mtb nanocompartment operon in *E. coli*. D) Histogram of size distribution of Mtb nanocompartments from *E. coli* measured using Negative-stain TEM. E) Total spectrum counts of Mtb encapsulin complex purified natively from Mtb cells.

To determine whether Mtb produces endogenous nanocompartments, we performed an ultracentrifugation based nanocompartment isolation strategy using wild-type H37Rv strain bacteria grown to mid-log phase (Figure 3.2B). Mass spectrometry analysis of the nanocompartment fraction identified both the putative encapsulin protein Cfp29 and DypB (Figure 3.1E). TEM analysis confirmed the presence of ~25 nm nanocompartment particles (Figure 3.1D). Interestingly, Rv1762c, a protein of unknown function, was also consistently identified in purified nanocompartment preparations from Mtb (Figure 3.1E). We were unable to identify either FolB or BrfB in nanocompartments from Mtb, suggesting that these proteins are not endogenous substrates for encapsulation under normal laboratory growth conditions for Mtb.

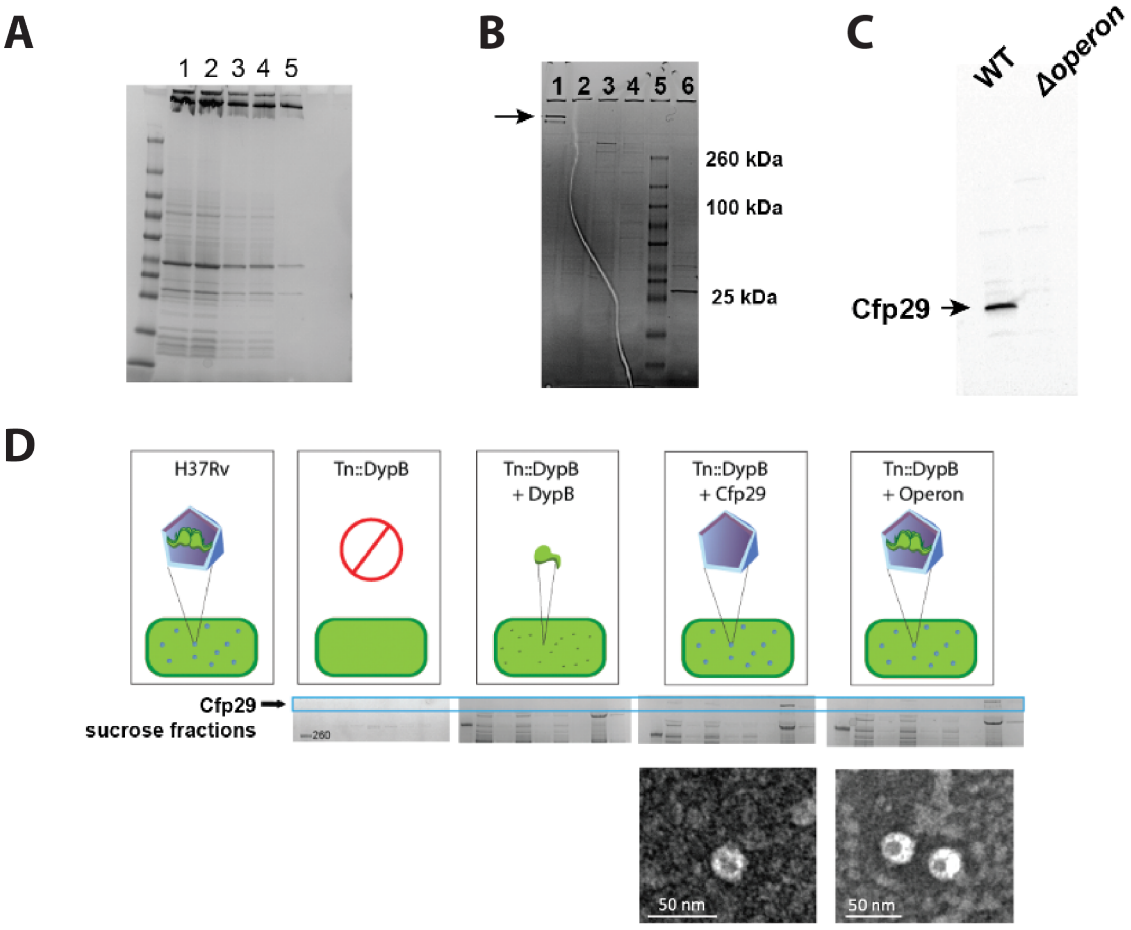

Figure 3.2 Analysis of purified Mtb encapsulin A) SDS-PAGE analysis of Mtb encapsulin purified from *E. coli*, wild type Mtb, and mutant Mtb. Coomassie stained SDS-PAGE of fractions collected during purification of nanocompartments heterologously expressed in E. coli: (1) Ultracentrifugation pellet post CellLytic B solubilization (2) Size exclusion chromatography input (3) lane 1 diluted (4) lane 2 diluted (5) encapsulin fraction from size exclusion B) sucrose fractions collected during purification of nanocompartments from wild-type Mtb lysates. Lane 5 is ladder; lane 6 is fraction 1 boiled in SDS for 30 minutes. C) Western blot for Cfp29 from wild-type Mtb (lane 1) and Δoperon mutant (lane 2) lysates. D) Complementation strategy schematic for DypB::Tn mutants (top). DypB::Tn mutants were transformed with ATc-inducible complementation constructs encoding the unencapsulated cargo protein (DypB), the encapsulin shell protein (Cfp29), or the nanocompartment operon (Operon). Lysates from each strain were used for nanocompartment purification. Sucrose fractions containing high molecular weight Cfp29 protomers were identified in complemented strains expressing the encapsulin shell and the operon (middle) and were analyzed using TEM (bottom).

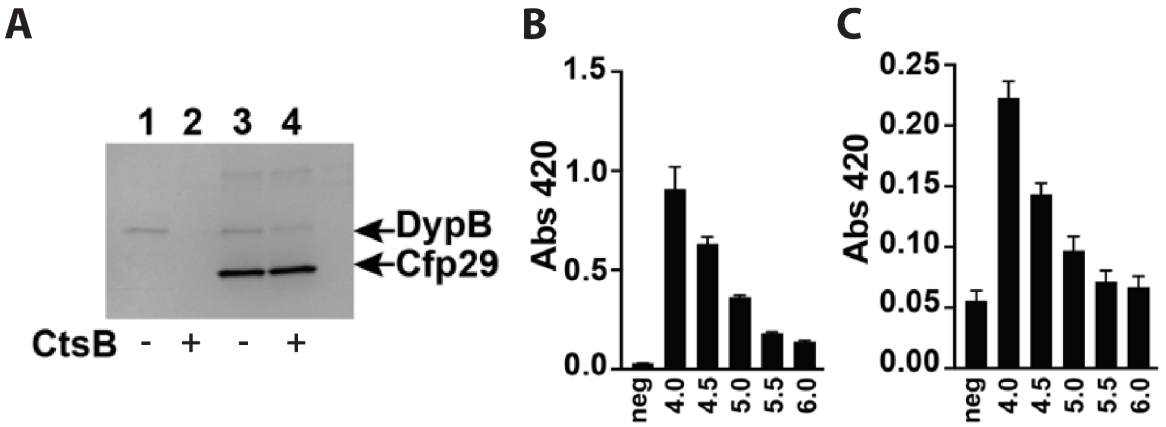

Figure 3.3 DypB-containing Mtb encapsulin is stable under phagolysosomal conditions. A) Coomassie-stained SDS-PAGE of unencapsulated (lanes 1 and 2) and encapsulated (lanes 3 and 4) DypB following 14 hours of Cathepsin B (CtsB) treatment. B) Unencapsulated and C) encapsulated DypB (5 nM) peroxidase activity using ABTS (480 nM) as a substrate in the presence of H2O2 (480 nM) at varying pH levels as reported by a change in the absorbance at 420 nm.

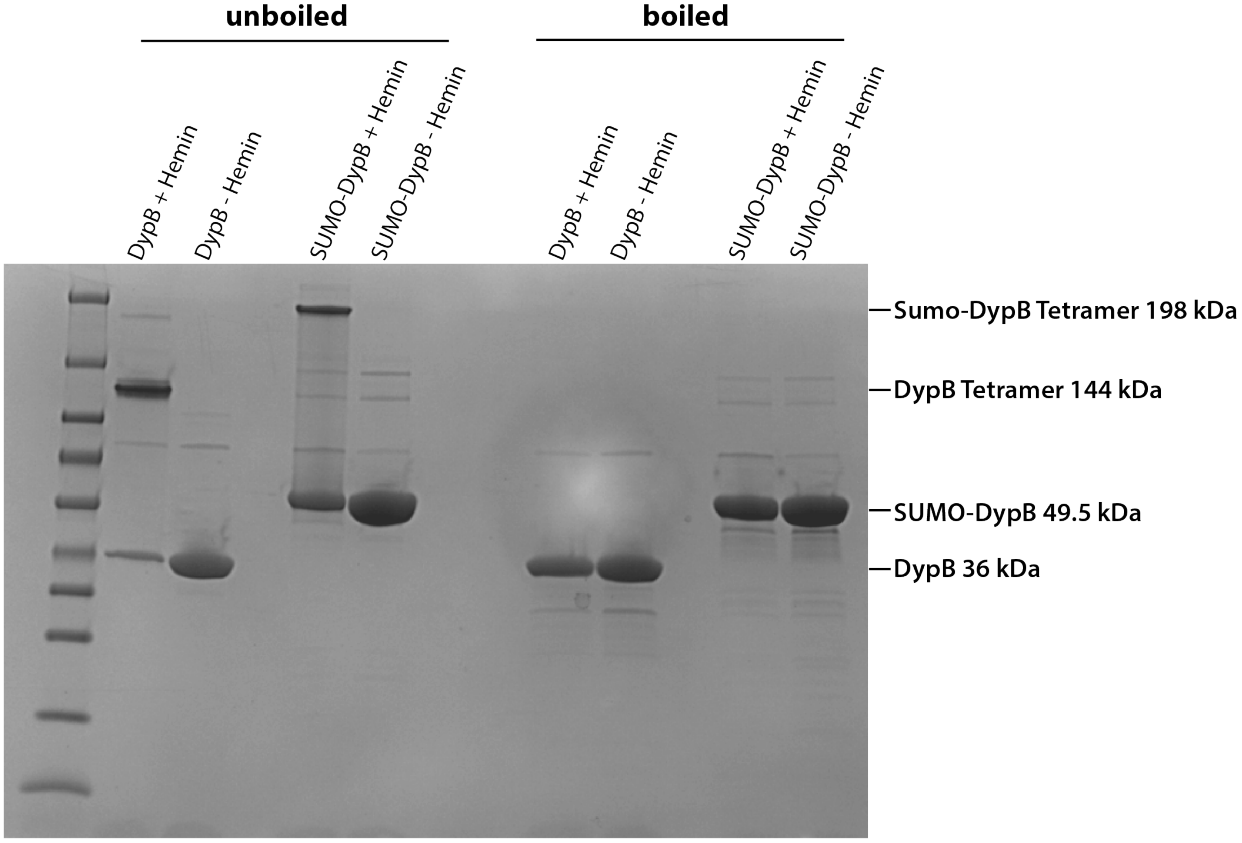

Figure 3.4 SDS-PAGE analysis of DypB purified from *E. coli*. DypB samples with or without addition of hemin were analyzed by SDS-PAGE. Samples were loaded either in their unboiled native state (left half) or heat-denatured by boiling at 95°C for 15 minutes. Addition of hemin yields a tetrameric DypB at 144 kDa.

3.3b Mtb encapsulins are active and stable under phagolysosomal conditions

Cfp29 (a.k.a. 'culture filtrate protein 29') was originally identified in the supernatants of Mtb cells grown in axenic culture[14]. As Cfp29 lacks a secretory signal sequence and is part of a large macromolecular complex, it is unclear how a nanocompartment could be actively secreted. However, nanocompartment structures are remarkably stable[17], and it is possible that nanocompartments released from dying bacteria accumulate in culture as they are highly resistant to proteolysis/degradation. To test this hypothesis, we purified unencapsulated DypB and DypB packaged within Mtb nanocompartments and exposed them to a lysosomal protease, cathepsin B, while monitoring proteolysis. Whereas unencapsulated DypB was completely degraded after 14 hours of exposure to cathepsin B, encapsulated DypB was protected from degradation, demonstrating the stability of nanocompartment structures in proteolytic environments (Figure 3.3A). Interestingly, DypB proteins are known to have low pH optimum[8,18]. To demonstrate that Mtb DypB has a similarly low pH optimum, SUMO-tagged unencapsulated DypB was purified from *E. coli* (Figure 3.4), and the peroxidase activity was evaluated at a range of pH values using the ABTS (2,2'-azino-bis(3-ethylbenzothiazoline-6-sulfonic acid) dye-decolorizing assay previously used to characterize DypB proteins[8,19]. Similar to the *Vibrio cholerae* DypB[18], purified Mtb DypB exhibited highest activity under acidic conditions, pH 4 (Figure 3.3B). We next sought to test the ability of the encapsulated Mtb to oxidize ABTS across a range of pH values. Similar to the free DypB, the encapsulated DypB displayed highest activity at pH 4.0 (Figure 3.3C). Taken together, these data demonstrate the stability and function of *M. tuberculosis* nanocompartments under proteolytic and acid stress, conditions that mimic the host lysosomal environment.

The Mtb DypB protein was previously shown to function as a bona fide dye-decolorizing peroxidase[8]. Because peroxidases consume H_2O_2, they often participate in defense against oxidative stress[20]. During infection of macrophages, Mtb encounters H_2O_2 produced by the oxidative burst initiated by phagocytes[21]. We therefore reasoned that DypB-containing nanocompartments might function to protect Mtb from H_2O_2 induced stress. To test this hypothesis, we created a mutant strain lacking both genes from the two-gene nanocompartment operon (Figure 3.1A), Δoperon. Lysates from Δoperon mutants were used for nanocompartment purification (data not shown) and Western blot analysis using an antibody for Cfp29 as a probe (Figure 3.2C), which revealed that viable nanocompartments were not produced. In addition, we isolated a transposon mutant (*DypB::Tn*) containing an insertion in DypB, a mutation that eliminated expression of both Cfp29 and DypB (Figure 3.2D). As both mutants lack expression of the two genes in the nanocompartment operon, we used them to assess the role of the DypB/encapsulin system in defense against oxidative stress.

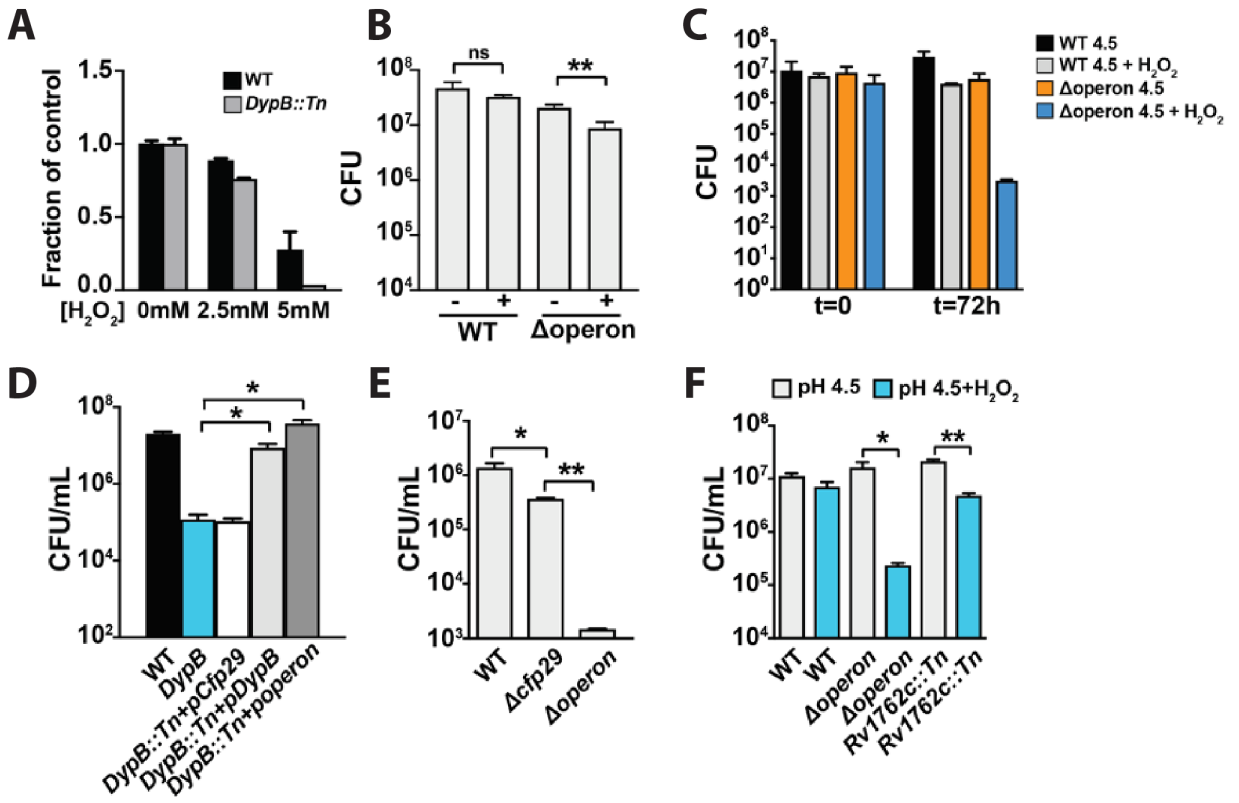

Figure 3.5 Nanocompartments protect Mtb from oxidative stress in acidic environments. A) OD_{600nm} measurements of wild-type and *DypB::Tn* mutants grown in 7H9 medium following exposure to 2.5 mM H_2O_2 for 96 hours. Values reported are normalized to the untreated controls. CFU enumeration of wild-type and Mtb nanocompartment mutants grown in B) standard 7H9 medium (pH 6.5) and C, E, F) acidified 7H9 medium (pH 4.5) following exposure to oxidative stress (2.5 mM H_2O_2) for 72 hours. D) CFU enumeration of wild-type, *DypB::Tn* mutants, and complemented mutants (pCfp29, pDypB, pOperon) following 24 hour exposure to oxidative stress (2.5 mM H_2O_2) in acidified 7H9 medium (pH 4.5). All figures are representative of at least 2 or 3 independent experiments. p-values were determined using unpaired t test. *p<0.05, **p<0.01.

3.3c Mtb encapsulins provide protection against oxidative stress

To test whether DypB containing nanocompartments are required for defense against H_2O_2, we exposed wild-type and *DypB::Tn* mutant Mtb to increasing concentrations of H_2O_2 and monitored bacterial survival. Mutants lacking DypB nanocompartments were more susceptible to H_2O_2 when compared with wild-type bacteria as measured by OD_{600} (Figure 3.5A) or by plating for CFU (Figure 3.5B). However, the phenotype was relatively modest. We next reasoned that DypB nanocompartments might be required to resist oxidative stress in acidic environments that more closely mimic the *in vivo* environment. During infection, Mtb bacilli encounter the low pH of the phagolysosome, and the ability to tolerate low pH is required for Mtb survival in both infected macrophages and mice[22]. We therefore tested the susceptibility of DypB mutants to a combination of H_2O_2 and acid stress (pH 4.5). Wild-type Mtb was able to withstand these conditions and did not significantly decrease in number over a three-day exposure period (Figure 3.5C). In contrast, the Δoperon mutant was resistant to each stressor individually, but was highly

susceptible to H_2O_2 at pH 4.5 (Figure 3.5C). Thus, DypB-containing nanocompartments are required to protect bacteria from oxidative stress at low pH.

We next sought to determine whether encapsulation of DypB is important for its activity in bacterial cells. To accomplish this, *DypB::Tn* mutants were complemented with constructs expressing DypB alone, Cfp29 encapsulin alone, or both proteins. As expected, expression of DypB did not restore production of nanocompartments, whereas expression of Cfp29 alone resulted in the formation of empty nanocompartment structures (Figure 3.2D). Co-expression of both proteins resulted in the formation of DypB containing nanocompartments in the *DypB::Tn* mutant (Figure 3.2D). We exposed the full set of complemented mutants to H_2O_2 at pH 4.5 and determined survival after 3 days by plating for CFU. As expected, the DypB mutants were attenuated under these conditions when compared to wild-type (Figure 3.5C). Restoring expression of the nanocompartment shell protein in the absence of DypB had no effect on bacterial survival (Figure 3.5D). Complementation by overexpression of unencapsulated DypB was sufficient to confer almost wild-type levels of resistance to oxidative and acid stress (Figure 3.5D). However, complementation with both Cfp29 and DypB resulted in enhanced resistance when compared with mutants complemented with DypB alone (Figure 3.5B), suggesting that encapsulation of DypB enhances its function.

It is possible that overexpression of DypB compensates for a lack of encapsulation in the mutant complemented with DypB alone. We therefore examined the impact of encapsulation of DypB in mutant strains in which we deleted *Cfp29* only. Deletion of Cfp29 alone resulted in a ~0.5 log decrease in bacterial viability after 3 days of exposure to H_2O_2 at pH 4.5, confirming that production of the shell protein is important for full resistance (Figure 3.5E). As expected, a more significant attenuation was observed in the mutant lacking both DypB and Cfp29 (Δoperon). Our mass spectrometry data suggested that an additional protein, Rv1762c, associates with the Mtb DypB nanocompartment (Figure 3.5E). To test whether Rv1762 has functional significance, we isolated a transposon mutant in this gene and exposed the mutant to H_2O_2 at pH 4.5. We found that the Rv1762c transposon mutant is moderately susceptible to the combination of low pH and H_2O_2 (Figure 3.5F). Taken together, these data demonstrate that encapsulation of DypB enhances protection against oxidative stress in low pH conditions and that Rv1762c may have a functional role in the DypB encapsulin system.

3.3d A genome-wide screen implicates Mtb encapsulin-associated genes and genes involved in lipid metabolism are important for oxidative stress response

The Mtb genome encodes ~18 putative peroxidase genes[23]. To understand whether other peroxidases participate in defense against oxidative stress at low pH, we performed a transposon sequencing (Tn-seq) screen in Mtb exposed to H_2O_2 at pH 4.5[24,25]. To do so, we created a transposon library containing ~100K individual transposon mutants and exposed this library to H_2O_2 at pH 4.5 for three days, at which time surviving bacteria were plated. Transposon gene junctions were amplified and sequenced from recovered bacteria and sequencing data were analyzed using TRANSIT[26]. These data revealed that of the 18 putative peroxidases, only 2 were significantly attenuated in the culture conditions – DypB and KatG (Figure 3.6A). KatG encodes a catalase-peroxidase gene

that is important for defense against oxidative stress in the host[27,28]. Coincidentally, KatG is also required for activation of isoniazid (INH), a central component of anti-TB therapy. Approximately 10% of TB cases are caused by INH-resistant bacteria, many of which have loss of function mutations in KatG[29]. Interestingly, both the high proportion of KatG mutant bacteria in the human population and studies of virulence in animals suggest that KatG mutants are still capable of growth *in vivo*[28,30]. The fact that both DypB and the KatG mutants are important for survival in oxidative stress at low pH suggests that the DypB nanocompartment may be particularly important in compensating for a loss of KatG in strains of drug resistant Mtb.

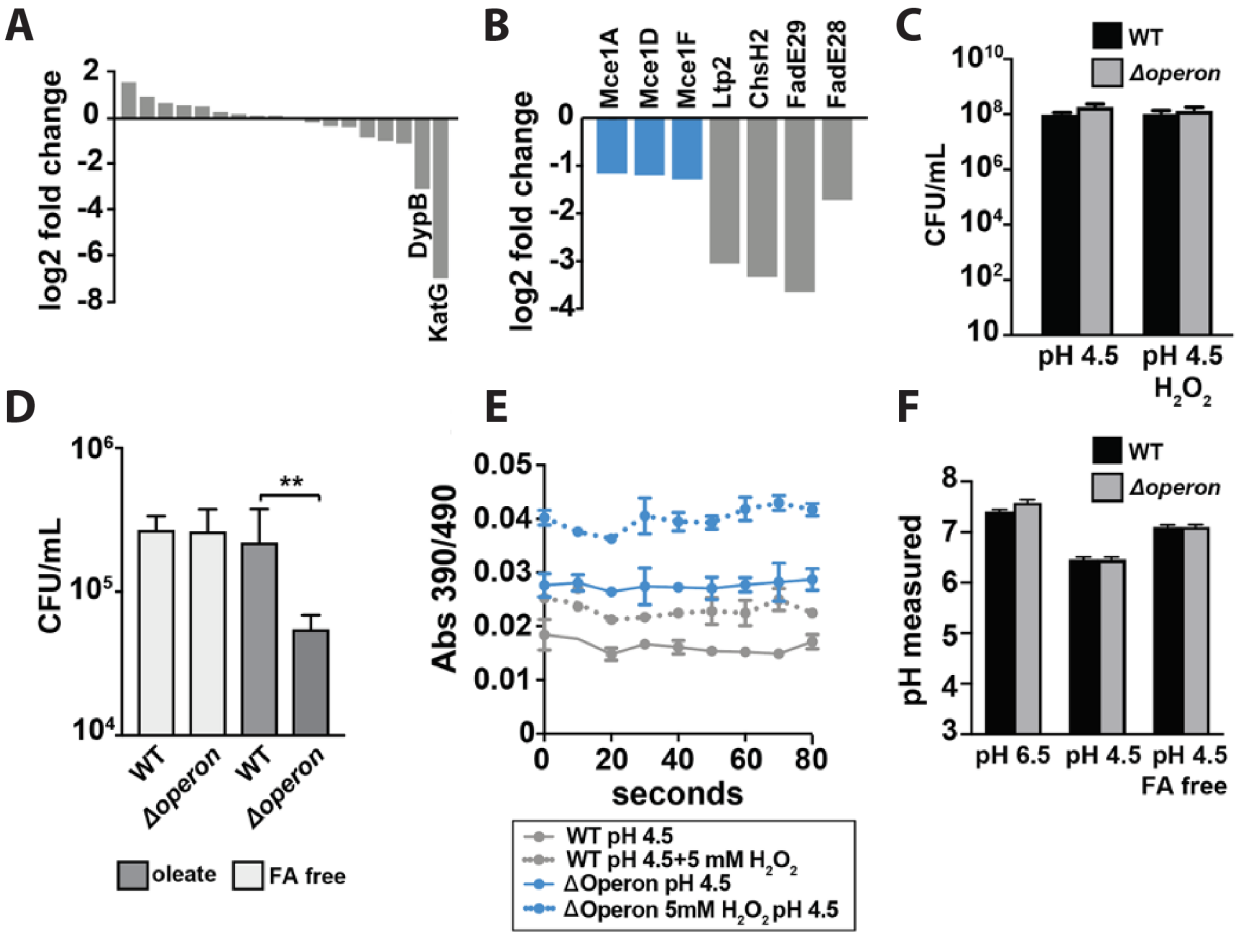

Figure 3.6 Susceptibility of Mtb nanocompartment mutants to oxidative and acid stress is mediated by free fatty acids. A) Tn-seq data showing normalized sequence reads per gene for all putative Mtb peroxidases and B) lipid and cholesterol metabolism Mtb mutants that were significantly attenuated following 72 hour exposure to 2.5 mM H_2O_2 at pH 4.5. C) CFU enumeration of wild-type Mtb and Δoperon mutants following 24 hour exposure to 2.5 mM H_2O_2 at pH 4.5 in Sauton's minimal medium and D) 72 hour exposure to 2.5 mM H_2O_2 at pH 4.5 in 7H9 medium prepared using lipid-free BSA +/- oleic acid (150 μM). E) Fluorescence emissions of wild-type Mtb and Δoperon mutants expressing mrx1-roGFP exposed to 5 mM H_2O_2 at pH 4.5 in 7H9 medium for 20 minutes. Data is reported as ratio of fluorescence emissions following excitation at 490 nm and 390 nm. Values are reported as a ratio to control for differences in plasmid expression. F) Intrabacterial pH measurements of wild-type Mtb and Δoperon expressing pUV15-pHGFP following 20 minute exposure to 5 mM H_2O_2 at pH 6.5 or pH 4.5 7H9 medium prepared with standard BSA or lipid free

BSA. A-C), E-F) are representative of at least 3 independent experiments and D) of two independent experiments. p-values were determined using an unpaired t test. *p<0.05, **p<0.01.

In the Tn-seq data, we observed that mutants involved in lipid or cholesterol metabolism were also attenuated when exposed to H_2O_2 at pH 4.5 (Figure 3.6B). *M. tuberculosis* biology is highly linked to lipid biology; *M. tuberculosis* granulomas are rich in free fatty acids[31], and *M. tuberculosis* utilizes lipids as a source of nutrition both during infection of macrophages *ex vivo* and during *in vivo* growth[32–34]. We hypothesized that lipids might mediate the sensitivity of Mtb DypB mutants to oxidative stress under acidic conditions. The standard Mtb culture medium 7H9 contains Bovine Serum Albumin (BSA), which has binding sites for fatty acids and might also serve as a cholesterol shuttle in serum[35,36]. We tested whether the susceptibility of Δoperon mutants to H_2O_2 at pH 4.5 was observable in Sauton's media, a minimal medium that lacks BSA. Δoperon mutants persisted to the same degree as wildtype in Sauton's medium (Figure 3.6C). To test whether lipids bound to BSA mediate the susceptibility of Mtb DypB mutants, we cultured wild-type and DypB mutants at pH 4.5 with H_2O_2 in media constituted with fatty acid-free BSA. In the absence of BSA bound lipids, we found that the Δoperon mutant was not susceptible to oxidative stress at low pH (Figure 3.6D). 7H9 medium prepared with lipid free BSA was reconstituted with oleic acid, an abundant lipid in mammalian systems[37]. Reconstitution of fatty acid free medium with oleic acid restored toxicity to DypB mutants under conditions of acid and oxidative stress (Figure 3.6D). Taken together, these data suggest that lipids found in the Mtb phagosome and that are utilized by Mtb as a carbon source may be toxic under conditions of oxidative stress in the absence of the DypB nanocompartment.

We next sought to test whether nanocompartment compartment mutants have altered redox status in the presence of oxidative stress. We transformed wild-type and mutant strains with Mrx1-roGFP, a fluorescent reporter of redox potential in mycobacteria[38]. Remarkably, the Δoperon mutant had an intracellular environment at baseline that was more oxidizing than wild-type bacteria exposed to H_2O_2 (Figure 3.6E), indicating a failure of redox homeostasis in mutants lacking the DypB encapsulin system. This oxidizing cellular environment was further increased by treatment with H_2O_2 (Figure 3.6E). Thus, mutants lacking DypB-containing nanocompartments exhibit a significant dysregulation of redox homeostasis.

We next considered the possibility that lipids might disrupt Mtb membranes, resulting in altered pH homeostasis, and a drop in cytosolic pH. A lowering of cytosolic pH would likely impact the function of the majority of peroxidases, but type B dye-decoloring peroxidases found in other bacterial species have the highest enzymatic activity at acidic pH levels[28] Interestingly, in acidified medium we found that the intracellular pH of wild-type bacteria dropped from ~pH 7.5 to ~ pH 6.2 and this decrease was dependent on the presence of albumin-bound fatty acids (Figure 3.6F). These data confirm published results demonstrating that the cytosolic pH of Mtb drops significantly during acid exposure[22].

3.3e Encapsulin mutants are sensitized to the antibiotic PZA

Our *in vitro* findings that DypB nanocompartments are important for defense against a combination of oxidative stress, low pH, and fatty acids suggested that this system may be important for surviving the phagosomal environment. Therefore, we next sought to determine whether DypB nanocompartments are required for bacterial growth in host cells. We found that both the *DypB::Tn* and Δoperon mutants had a defect in growth in bone-marrow derived macrophages that was manifest by two days after infection (Figure 3.7A,B). The importance of DypB for protection against oxidative stress at low pH suggested the possibility that this organelle might be important for Mtb resistance to the antibiotic pyrazinamide (PZA), which is most active at low pH levels including the phagolysosome environment[39]. Interestingly, a recent study demonstrated that in human TB patients, the pH of necrotic cavities is pH 5.5, a finding that may also explain the efficacy of pyrazinamide in treating Mtb infection[40]. To test bacterial susceptibility to PZA, we treated wild-type and Δoperon mutant Mtb with PZA in the presence of H_2O_2 at pH 5.5. We found that mutants lacking the DypB nanocompartment were sensitive to concentrations that did not impact growth of the wild-type bacteria (Figure 3.7C).

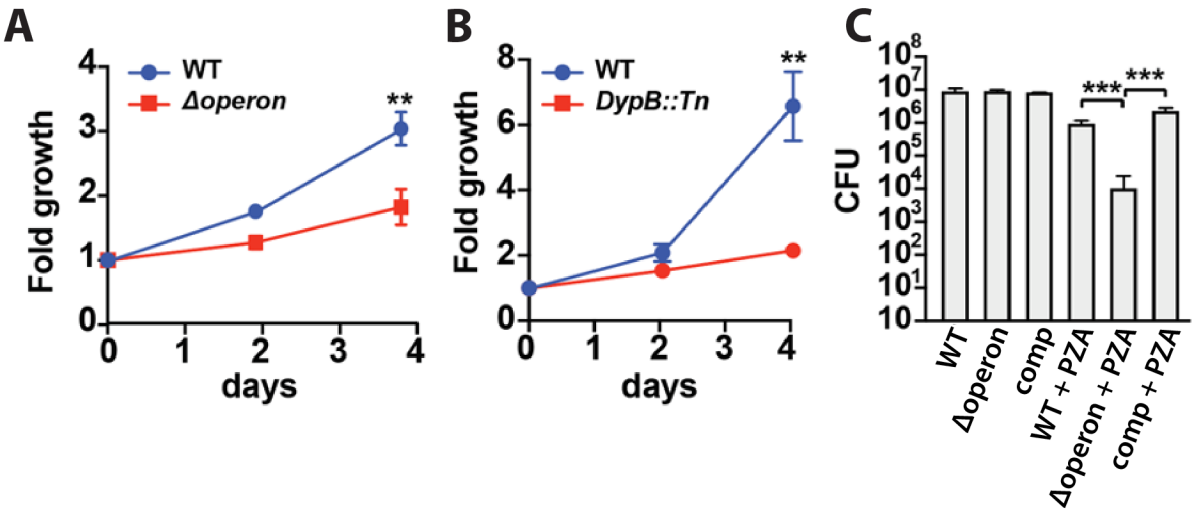

Figure 3.7. Nanocompartment mutants are attenuated for survival in macrophages and are more susceptible to pyrazinamide treatment. CFU enumeration of wild-type Mtb and A) Δoperon or B) *DypB::Tn* mutants during infection of murine bone marrow-derived macrophages. Macrophages were infected with a bacterial MOI of 1 and CFUs were enumerated immediately following phagocytosis and at days 2 and 4. Error bars are SD from 4 replicate wells. C) CFU enumeration of wild-type and Mtb Δoperon mutants following 72 hour exposure to pyrazinamide (24 μg/mL) and H_2O_2 (2.5 mM) in acidified 7H9 medium (pH 5.5).

3.4 Discussion

Encapsulin nanocompartments are an enigmatic proteinaceous organelle widespread in bacteria and archaea. Here we provide some of the first functional data demonstrating the physiological significance of a nanocompartment. In addition, our data is the first to specifically demonstrate the importance of encapsulation independent of the function of the encapsulated enzyme. Furthermore, we have demonstrated that an encapsulin

system is important for resisting oxidative stress, thereby expanding our understanding of how this globally significant pathogen resists host defenses to maintain infection.

3.4a Identification of cargo from natively purified encapsulin versus heterologously purified encapsulin

Previously, it was shown that Mtb encapsulin hosts three cargo types: BfrB, FolB, DypB[8]. These cargo proteins were identified by the putative C-terminal targeting sequence they shared with the previously characterized targeting sequence of the *B. linens* DypB encapsulin cargo. In our examination of the Mtb encapsulin purified natively from Mtb cells, we corroborated the finding that DypB was a cargo protein, but did not find BfrB nor FolB as hits in our mass spectrometry analysis. There are a few potential explanations for this discrepancy. One possibility is that the bioinformatic identification of BrfB and FolB containing a DypB-like targeting sequence did not use a stringent enough threshold for sequence similarity. This alone, however, does not explain how BfrB and FolB could be co-purified with the encapsulin upon heterologous expression in *E. coli*. It is possible that the overexpression of the cargo protein in *E. coli* yielded misincorporation of BfrB and FolB into the nanocompartment. There is precedent for this possibility as it has been shown that a randomized C-terminal sequence can yield some encapsulation of cargo protein without a proper targeting sequence, although not to the same extent as the wild type targeting sequence[17].

Alternatively, it is possible that the growth conditions used to grow Mtb in this study did not yield sufficient expression of BrfB or FolB such that their encapsulation could be measured by mass spectrometry of the purified compartments. Often the cargo gene for a given encapsulin system directly neighbors the shell gene with some existing as a shell-cargo fusion[1,9,12]. There is, however, evidence that encapsulins can sequester gene products that do not neighbor the shell gene such as the EncC and EncD proteins which were found in encapsulins natively purified from *M. xanthus*. In our identification of cargo from natively purified Mtb encapsulins, we found the presence of an additional non-neighboring gene product – the catalase/ peroxidase KatG. This protein was found in very low abundance, so it is possible that it is a contaminant from the preparation. However, KatG's shared function with DypB as a peroxidase in addition to its functional importance in antibiotic resistance make it an intriguing candidate cargo protein to investigate in the future[29].

3.4b Activity and stability of Mtb encapsulins in the phagolysosome

Mtb persist as intracellular pathogens within the phagolysosome of macrophages[41,42]. Given that Mtb encapsulins are be often found extracellularly in the culture filtrate when Mtb are cultured in media, it is likely that encapsulins are found outside of Mtb cells within the lumen of the phagolysosome of an infected macrophage. It is therefore possible that Mtb encapsulin is capable of performing its activity within the chemical environment of the phagolysosome. The phagolysosome is known to be an acidic environment with a measured pH around 4.5[43–45]. We have shown here that both the naked DypB and the encapsulated DypB exhibit increased activity under acidic conditions consistent with a predicted function of the peroxidase performing its biological activity primarily outside of the Mtb intracellular environment and within the lumen of the phagolysosome.

If a cargo protein were to be stable within the phagolysosome, it would need to be resistant to proteases in addition to being acid stable. The lysosome is a potent proteolytic environment encoding a family of proteases known as cathepsins[46]. We have shown that naked DypB is degraded upon exposure to the cathepsin protease Cathepsin B. Upon encapsulation, however, the DypB is more resistant to degradation. This protection from lysosomal proteases may provide an explanation as to why DypB needs to be encapsulated and why deletion of the encapsulin shell protein sensitizes Mtb cells to killing upon oxidative stress.

The phenomenon of protease resistance is not unique to the Mtb encapsulin but has been observed in encapsulin systems in other species. The ferritin-like protein (FLP)-containing encapsulin from *Thermotoga maritima* has been shown to resist subtilisin treatment *in vitro*[17]. Lau *et al.* demonstrated the encapsulin from *M. xanthus* protected a heterologous mNeonGreen protein from proteolytic degradation while unencapsulated mNeonGreen was degraded by endogenous yeast proteases[47]. More recently, it was shown that an encapsulin from a DypB-containing encapsulin from a relative of the same taxonomic family as Mtb, *Mycolicibacterium hassiacum*, remains fully intact upon challenge with proteinase K[48]. Resistance to proteases, therefore, appears to be a generalized function of encapsulin capsids. The biological importance of this remains unclear. It is likely that encapsulation of the cargo enzyme increases the stability and lifespan of the cargo. This conferred stability to cargo may be critical if the physiological function of encapsulins is extracellular to the cells that produced it, which has been hypothesized previously[1,15,49,50].

3.4c The role of Mtb encapsulin in survival of oxidative stress

Encapsulins have been shown previously to confer survival for bacteria under oxidative stress. The haemerythin encapsulin from *Streptomyces* sp. AA4 has been demonstrated to provide a fitness benefit upon hydrogen peroxide challenge, although this phenotype was acquired through the heterologous expression of the encapsulin system and not obtained from genetics in the native host[2]. Prior to this work, McHugh *et al*. provided the only genetic evidence demonstrating the biological importance of an encapsulin in its native organism[10]. They showed that loss of the encapsulin shell gene in *M. xanthus* decreased cell viability by roughly 50% when compared to wild type after hydrogen peroxide treatment. The genetic study in *M. xanthus*, however, did not examine the phenotypic effect of knocking out the cargo by itself or the viability of a shell-cargo double mutant[10].

Here we show evidence, similar to what was found for *M. xanthus*, that disruption of the shell gene yields a growth inhibition phenotype for Mtb exposed to hydrogen peroxide. The growth defect of Mtb is additive when the cargo gene is disrupted in addition to the shell protein. Additionally, complementation of DypB cargo in a *DypB::Tn* background restored CFU counts. This suggests that while the encapsulation of cargo is important for Mtb to survive oxidative stress, the cargo itself is critical for the biological function of Mtb evading oxidative stress. We showed that this growth phenotype is exacerbated by addition of fatty acids in the growth medium. Both wild type and Δoperon Mtb demonstrated similar growth upon hydrogen peroxide stress in fatty acid free media,

however, addition of oleic acid to the 7H9 growth medium yielded a significant growth defect for the Δoperon mutant.

The precise biological significance of this result remains to be further characterized. Macrophages are known to produce reactive oxygen species (ROS) in attempt to kill the intracellular pathogen[51,52]. Given the importance of Mtb encapsulin in Mtb survival of oxidative stress, it may be that the DypB peroxidase is acting to ameliorate the ROS that would otherwise be cytotoxic. While we used ABTS and hydrogen peroxide to study the activity of DypB *in vitro*, the specific substrate for DypB within the Mtb-infected macrophage is unknown. We have shown that the presence of fatty acid in the Mtb growth media is essential for detecting the growth phenotype of the encapsulin mutant. One possible substrate for DypB may be a fatty acid that has undergone lipid peroxidation. Further biochemical characterization of DypB in addition to metabolomic studies of wild type and encapsulin mutant Mtb will be needed to further elucidate the native substrate of the Mtb encapsulin.

3.5 Methods

3.5a Ethics statement. All procedures involving the use of mice were approved by the University of California, Berkeley Institutional Animal Care and Use Committee (protocol no. R353-1113B). All protocols conform to federal regulations, the National Research Council's *Guide for the Care and Use of Laboratory Animals*, and the Public Health Service's *Policy on Humane Care and Use of Laboratory Animals*.

3.5b *Mycobacterium tuberculosis* bacterial strains and plasmids. The *M. tuberculosis* strain H37Rv was used for all experiments. The transposon mutants *DypB::Tn* and Rv1762c::Tn were picked from an arrayed transposon mutant library generated at the Broad Institute. The Δoperon, ΔCfp29, and ΔDypB strains were made by homologous recombination using the pMSG361 vector[53]. For genetic complementation studies, the region encoding GFP and KanR in pUV15tetORm[54] was substituted via GoldenGate cloning with open reading frames for Rv0798c, Rv0799c, or the whole nanocompartment operon (Rv0798-99c). Expression of the complementation constructs was induced with anhydrotetracycline (200 ng/mL). To measure redox homeostasis, strains were transformed with pMV762-mrx1-roGFP2[38]. To measure intrabacterial pH, strains were transformed with pUV15-pHGFP (Addgene). The transposon mutant library for Tn-Seq was generated in *M. tuberculosis* using the ΦMycoMarT7 transposon donor plasmid.

3.5c *M. tuberculosis* bacterial cell culture. For infections*, M. tuberculosis* was grown to mid-log phase (OD_{600} = 0.5-1.0) in Middlebrook 7H9 liquid medium supplemented with 10% albumin-dextrose-saline, 0.4% glycerol, and 0.05% Tween-80 or on solid 7H10 agar plates supplemented with Middlebrook OADC (BD Biosciences) and 0.4% glycerol. When specified, Tween-80 was substituted with 0.05% Tyloxapol, and 10% albumin-dextrose-saline was prepared with fatty acid free BSA (Sigma-Aldrich). Sauton's media was prepared with tyloxapol as previously specified[54].

3.5d DypB activity assays. Activity of the encapsulated and unencapsulated DypB was performed using methods adapted from Contreras *et al.* 2014[8]. Briefly, DypB concentration for the encapsulated and unencapsulated DypB was determined by absorbance of the heme prosthetic group at 411 nm. Reactions were performed using 5 nM DypB, 480 nM H_2O_2, and 480 nM 2,2'-azino-bis (3-ethylbenzothiazoline-6-sulfonic acid) (ABTS) in 100mM sodium citrate buffer pH 4-6. Product formation was monitored over 20 minutes via absorbance at 420 nm using a Varian Cary® 50 UV-Vis Spectrophotometer (Agilent).

3.5e Nanocompartment Purification from Mtb. For each purification, 1.5 L of M. tuberculosis was grown to mid-log phase in standard 7H9 and washed with PBS. Bacteria were pelleted and lysed in buffer by bead beating (for 50 mL of buffer, PBS with 1 mM PMSF was supplemented with 50 mg lysozyme, 20 U DNaseI, and 100 µg RNase A). Lysates were passaged twice through 0.2 µm filters before removal from the BSL3. Clarified lysates were prepared by centrifugation at 20,000 x *g* for 20 minutes in a JA-20 rotor. Following clarification, lysates were layered onto top of a 38% sucrose cushion and

centrifuged for 18 hours at 100,000 x *g* in a type 50.2 Ti rotor. The supernatant was discarded and the pellet was re-suspended in 200 µL of PBS. Re-suspended pellets were layered on top of a 10-50% sucrose gradient and centrifuged for 21 hours at 100,000 x g in a SW 41 Ti rotor. The gradient was fractionated and aliquots from each fraction were analyzed by SDS-PAGE for the presence of CFP29.

3.5f Expression of Holo-nanocompartment and naked DypB in *E. coli*. Plasmids for the expression of the Holo-nanocompartment (DypB-loaded) and naked DypB constructs were designed using Gibson Assembly (NEB). Each construct was cloned into a pET-14-based destination vector containing a T7 promoter. The naked DypB construct contained an N-terminal poly-histidine tag for affinity purification. These constructs were transformed into *E. coli* BL21 (DE3) LOBSTR cells for protein overexpression. Cells were grown in LB media containing 60 µg/mL kanamycin at 37°C with shaking at 250 rpm until cultures reached an optical density (OD_{600} = 0.5-0.6). Samples were then induced with 0.5 mM IPTG and grown overnight at 18°C. Liquid cultures were harvested by centrifugation at 5000 x *g* for 20 minutes at 4°C, flash frozen in liquid nitrogen, and then stored at -80°C for future use.

3.5g Purification of Holo-nanocompartment complex from *E. coli*. Cell pellets (5g dry cell mass) were thawed at room temperature and resuspended in 50 mL of lysis buffer (20 mM Tris-HCl pH 8, 150 mM NH_4Cl, 20 mM $MgCl_2$) supplemented with 50 mg lysozyme, 20 U DNaseI, 100 µg RNase A. Samples were lysed by three passages through an Avestin EmulsiFlex-C3 homogenizer and clarified via centrifugation (15,000 x g, 30 min, 4°C). The clarified lysate was then spun at 110,000 x *g* for 3 hours at 4 °C. The supernatant was discarded and the resulting pellet was resuspended with wash buffer (20 mM Tris pH 8, 150 mM NH_4Cl, 20 mM $MgCl_2$ supplemented with 1X Cell Lytic B (Sigma-Aldrich). The sample was then spun at 4,000 x *g* at 4°C for 10 min followed by removing the supernatant and resuspension of the pellet in 4mL of 50 mM Tris-HCl pH 8, 300 mM NaCl. The sample was then incubated at room temperature for 10 minutes to allow for solubilization and then centrifuged at 4000 x *g* at 4°C for 10 minutes to remove insoluble material. The resulting supernatant was then concentrated using Vivaspin® 6 100,000 MWCO concentrator columns (Sartorius). The sample was then purified via size exclusion chromatography using a Superose™ 6 Increase column (GE Life Sciences) and fractions were analyzed by SDS-PAGE using 4-20% Criterion™ polyacrylamide gels (Bio-Rad) and visualized with GelCode Blue stain (ThermoFisher).

3.5h Purification of naked DypB from *E. coli*. Cell pellets (5g dry cell mass) were thawed at room temperature and resuspended in 50 mL of buffer A (25 mM Tris HCl pH 7.5, 150 mM NaCl, 20 mM imidazole) supplemented with 50 mg lysozyme, 20 U DNaseI, 100 µg RNase A. Samples were lysed by three passages through an Avestin EmulsiFlex-C3 homogenizer and clarified via centrifugation (15,000 x g, 30 min, 4°C). The resulting supernatant was then bound to HisPur™ Ni-NTA resin (ThermoFisher Scientific) for 90 minutes at 4°C and then applied to a gravity column. The nickel resin was then washed with 30 resin volumes of buffer B (25 mM Tris-HCl pH 7.5, 150 mM NaCl, 40mM imidazole) prior to eluting with buffer C (25 mM Tris-HCl pH 7.5, 150 mM NaCl, 350 mM imidazole). The eluate was then concentrated using Vivaspin® 20 10,000 MWCO

concentrator columns (Sartorius) and desalted into 25mM Tris pH 8, 300mM NaCl using Econo-Pac®10DG desalting columns (BioRad). The SUMO tag was removed upon addition of SUMO protease at a 1: 300 (SUMO protease: DypB) molar ratio and incubating overnight at 4°C. Purification was finished by size exclusion chromatography with a Superose™ 6 Increase column (GE Life Sciences).

3.5i Negative stain Transmission Electron Microscopy. Nanocompartment samples were diluted to 50 nM and applied to Formvar/ carbon-coated copper grids. The grids were then washed with MilliQ water three times followed by staining with 2% (w/v) uranyl acetate. Grids were examined using the FEI Tecnai 12, 120kV transmission electron microscope, and images were captured with a charge-coupled device (CCD) camera.

3.5j Cathepsin B proteolysis assay. Proteolysis was performed using 2 μM of encapsulated or unencapsulated DypB in proteolysis buffer (100mM Sodium acetate pH 5, 2 mM DTT) with or without the addition of 1 μM Cathepsin B (Sigma-Aldrich Cat# 219362). Reactions were run for 14 hours at 40°C followed by analysis of samples by SDS-PAGE using 4-20% Criterion™ polyacrylamide gels (Bio-Rad) and visualized with GelCode Blue stain (ThermoFisher).

3.5k Exposure to oxidative and pH stress. *M. tuberculosis* was grown to mid-log phase in 7H9 media. Bacteria were diluted to OD_{600} = 0.1 in 10 mL of specified media at pH 4.5-6.5 and H_2O_2 was added to bacterial cultures at specified concentrations. Bacteria were incubated with stressors for 24 or 72 hours. CFUs were enumerated by diluting bacteria in PBS with 0.05% Tween-80 and plating serial dilutions on 7H10 agar.

3.5l Measurement of redox homeostasis. *M. tuberculosis* strains were transformed with a plasmid expressing mrx1-roGFP2 and grown to mid-log phase in 7H9. Bacteria were diluted to OD_{600} = 0.25 in 200 μL of specified media and added to 96 well plates. Upon addition of H_2O_2 (5 mM), fluorescent emissions were recorded at 510 nm after excitation at 390 nm and 490 nm using a Spectramax M3 spectrophotometer. Values reported are emissions ratios (390 nm/ 490 nm) and were measured 20 minutes following addition of H_2O_2.

3.5m Measurement of intrabacterial pH. *M. tuberculosis* strains were transformed with a plasmid expressing pHGFP and grown to mid-log phase in 7H9. To prepare standards, 1.5 x 10^8 bacterial cells were pelleted and resuspended in 400 μL lysis buffer (50 mM Tris-HCl pH 7.5, 5 mM EDTA, 0.6% SDS, 1 mM PMSF) before bead beating. Cell debris were pelleted and clarified lysates were kept at 4°C until use, at which point 10 μL of clarified lysate were added to 200 μL of medium with varying pH levels (4.5-8.0). To prepare samples, 1.5 x 10^8 bacterial cells were pelleted and washed with PBS twice before being resuspended in specified media and diluted to OD_{600} = 0.5 in 200 μL of medium and added to 96 well plates. Upon addition of H_2O_2 (5 mM), fluorescent emissions were recorded at 510 nm following excitation at 395 nm and 475 nm. Values reported were interpolated from 395/475 ratios obtained from a standard curve.

3.5n Western blot analysis of Cfp29 expression. *M. tuberculosis* strains were grown to mid-log phase in 7H9 medium. Bacteria were pelleted and washed twice with PBS prior to being resuspended in lysis buffer (50 mM Tris-HCl pH 7.5, 5 mM EDTA, 0.6% SDS, 1 mM PMSF). Samples were lysed using a bead-beater, and cell debris were pelleted. Clarified lysates were heat-sterilized at 100°C for 15 minutes and frozen prior to use. Total protein lysates were analyzed by SDS-PAGE using precast Tris-HCl 4-20% criterion gels (Bio-Rad). Primary polyclonal antibodies for Cfp29 were generated by GenScript USA Inc via immunization of rabbits with three peptides from the protein sequence. HRP-conjugated goat anti-rabbit IgG secondary antibodies were used (sc-2030; Santa Cruz Biotechnology). Western Lightning Plus-ECL chemiluminescence substrate (Perkin Elmer) was used and blots were developed using a ChemiDoc MP System (Bio-Rad).

3.5o Infection of murine macrophages. Macrophages were derived from bone marrow of C57BL/6 mice by flushing cells from femurs. Cells were cultured in DMEM supplemented with 10% FBS and 10% supernatant from 3T3-M-CSF cells for 6 days, with feeding on day 3. After differentiation, BMDMs continued to be cultured in BMDM media containing M-CSF. For infection, BMDMs were seeded at a density of 5 x 10^4 cells per well in a 96-well dish. BMDMs were allowed to adhere overnight and then infected with DMEM supplemented with 5% FBS and 5% horse serum (BMDMs) at a multiplicity of infection of 1. Following a 4-hour phagocytosis period, infection medium was removed and cells were washed with room temperature PBS before fresh, complete medium was added. For CFU enumeration, medium was removed and cells were lysed in water with 0.5% Triton-X and incubated at 37°C for 10 minutes. Following the incubation, lysed cells were re-suspended and serially diluted in PBS with 0.05% Tween-80. Dilutions were plated on 7H10 plates.

3.5p Transposon-sequencing screen. A transposon mutant library in H37Rv was grown to mid-log phase in 7H9. Bacteria were diluted to OD_{600} = 0.1 in 10 mL 7H9 at pH 4.5 with 2.5 mM H_2O_2. Mutants were exposed to these stressors for 72 hours and then diluted to 15,000 CFU/mL in PBS with 0.05% Tween-80. Approximately 30 thousand bacteria were plated onto six 245 mm x 245 mm 7H10 plates supplemented with 0.05% Tween-80 and Kanamycin (50 mg/mL). Control libraries were not exposed to low pH or H_2O_2 and were plated onto 7H10 plates. Colonies grew for 21 days and were collected for genomic DNA isolation. Samples for sequencing were prepared by the University of California, Davis Genome Center DNA Technologies Core by following the protocol outlined by Long et al[56]. PE100 reads were run on an Illumina HiSeq with ~20 million reads per sample. Sample alignment and TRANSIT pre-processing were performed by the University of California, Davis Bioinformatics Group as previously outlined[26]. TRANSIT analysis was performed as specified by DeJesus et al[26]. Resampling analysis was performed using the reference genome H37RvBD_prot and the following parameters: for global options, 0% of the N- and C- terminus were ignored; for resampling options, 10,000 samples were taken and normalized using the TTR function. Correction for genome positional bias was performed. Statistical significance was determined by p value ≤ 0.05 and log2 fold change ≤-1 or by p-adjusted value ≤ 0.05.

3.6 Acknowledgments

We thank Tom Ioerger and Michael DeJesus for assistance with TRANSIT analysis and Amit Singh for the kind gift of pMV762-mrx1-roGFP2. We thank Lindsay Eltis, Jeff Cox, and members of the Cox lab for helpful discussions. This work was supported by funding from the Center for Emerging and Neglected disease for funding to KAL, the National Institute of General Medical Sciences (grant no. R01GM129241) to D.F.S. and the National Institute of Allergy and Infectious Diseases (grant no. 1R01AI143722) to SAS.

3.7 References

1. Nichols, R. J., Cassidy-Amstutz, C., Chaijarasphong, T. & Savage, D. F. Encapsulins: molecular biology of the shell. *Crit. Rev. Biochem. Mol. Biol.* **52**, 583–594 (2017).
2. Giessen, T. W. & Silver, P. A. Widespread distribution of encapsulin nanocompartments reveals functional diversity. *Nat Microbiol* **2**, 17029 (2017).
3. Barberis, I., Bragazzi, N. L., Galluzzo, L. & Martini, M. The history of tuberculosis: from the first historical records to the isolation of Koch's bacillus. *J. Prev. Med. Hyg.* **58**, E9 (2017).
4. Zink, A. R. *et al.* Characterization of Mycobacterium tuberculosis complex DNAs from Egyptian mummies by spoligotyping. *J. Clin. Microbiol.* **41**, 359–367 (2003).
5. Tuberculosis (TB). https://www.who.int/news-room/fact-sheets/detail/tuberculosis.
6. Madrazo-Moya, C. F. *et al.* Whole genomic sequencing as a tool for diagnosis of drug and multidrug-resistance tuberculosis in an endemic region in Mexico. *PLoS One* **14**, e0213046 (2019).
7. Cambier, C. J., Falkow, S. & Ramakrishnan, L. Host evasion and exploitation schemes of Mycobacterium tuberculosis. *Cell* **159**, 1497–1509 (2014).
8. Contreras, H. *et al.* Characterization of a Mycobacterium tuberculosis nanocompartment and its potential cargo proteins. *J. Biol. Chem.* **289**, 18279–18289 (2014).
9. Sutter, M. *et al.* Structural basis of enzyme encapsulation into a bacterial nanocompartment. *Nat. Struct. Mol. Biol.* **15**, 939–947 (2008).
10. McHugh, C. A. *et al.* A virus capsid-like nanocompartment that stores iron and protects bacteria from oxidative stress. *EMBO J.* **33**, 1896–1911 (2014).
11. Giessen, T. W. *et al.* Large protein organelles form a new iron sequestration system with high storage capacity. *Elife* **8**, (2019).
12. Nichols, R. J. *et al.* Discovery and characterization of a novel family of prokaryotic nanocompartments involved in sulfur metabolism. *bioRxiv* 2020.05.24.113720 (2020) doi:10.1101/2020.05.24.113720.
13. Weldingh, K. & Andersen, P. Immunological evaluation of novel Mycobacterium tuberculosis culture filtrate proteins. *FEMS Immunol. Med. Microbiol.* **23**, 159–164 (1999).
14. Rosenkrands, I. *et al.* Identification and characterization of a 29-kilodalton protein from Mycobacterium tuberculosis culture filtrate recognized by mouse memory effector cells. *Infect. Immun.* **66**, 2728–2735 (1998).
15. Radford, D. Understanding the encapsulins: prediction and characterization of phage capsid-like nanocompartments in prokaryotes. (2015).
16. Zhang, Y. J. *et al.* Tryptophan biosynthesis protects mycobacteria from CD4 T-cell-mediated killing. *Cell* **155**, 1296–1308 (2013).
17. Cassidy-Amstutz, C. *et al.* Identification of a Minimal Peptide Tag for in Vivo and in Vitro Loading of Encapsulin. *Biochemistry* **55**, 3461–3468 (2016).
18. Uchida, T., Sasaki, M., Tanaka, Y. & Ishimori, K. A Dye-Decolorizing Peroxidase from Vibrio cholerae. *Biochemistry* **54**, 6610–6621 (2015).
19. Ahmad, M. *et al.* Identification of DypB from Rhodococcus jostii RHA1 as a lignin peroxidase. *Biochemistry* **50**, 5096–5107 (2011).

20. Mishra, S. & Imlay, J. Why do bacteria use so many enzymes to scavenge hydrogen peroxide? *Arch. Biochem. Biophys.* **525**, 145–160 (2012).
21. El-Benna, J., Dang, P. M.-C., Gougerot-Pocidalo, M. A., Marie, J. C. & Braut-Boucher, F. p47phox, the phagocyte NADPH oxidase/NOX2 organizer: structure, phosphorylation and implication in diseases. *Exp. Mol. Med.* **41**, 217–225 (2009).
22. Vandal, O. H., Pierini, L. M., Schnappinger, D., Nathan, C. F. & Ehrt, S. A membrane protein preserves intrabacterial pH in intraphagosomal Mycobacterium tuberculosis. *Nat. Med.* **14**, 849–854 (2008).
23. Meehan, C. J. *et al.* Whole genome sequencing of Mycobacterium tuberculosis: current standards and open issues. *Nat. Rev. Microbiol.* **17**, 533–545 (2019).
24. van Opijnen, T. & Camilli, A. Transposon insertion sequencing: a new tool for systems-level analysis of microorganisms. *Nat. Rev. Microbiol.* **11**, 435–442 (2013).
25. Griffin, J. E. *et al.* Cholesterol catabolism by Mycobacterium tuberculosis requires transcriptional and metabolic adaptations. *Chem. Biol.* **19**, 218–227 (2012).
26. DeJesus, M. A., Ambadipudi, C., Baker, R., Sassetti, C. & Ioerger, T. R. TRANSIT--A Software Tool for Himar1 TnSeq Analysis. *PLoS Comput. Biol.* **11**, e1004401 (2015).
27. Li, Z., Kelley, C., Collins, F., Rouse, D. & Morris, S. Expression of katG in Mycobacterium tuberculosis is associated with its growth and persistence in mice and guinea pigs. *J. Infect. Dis.* **177**, 1030–1035 (1998).
28. Pym, A. S., Saint-Joanis, B. & Cole, S. T. Effect of katG mutations on the virulence of Mycobacterium tuberculosis and the implication for transmission in humans. *Infect. Immun.* **70**, 4955–4960 (2002).
29. Seifert, M., Catanzaro, D., Catanzaro, A. & Rodwell, T. C. Genetic mutations associated with isoniazid resistance in Mycobacterium tuberculosis: a systematic review. *PLoS One* **10**, e0119628 (2015).
30. Nieto R, L. M. *et al.* Virulence of Mycobacterium tuberculosis after Acquisition of Isoniazid Resistance: Individual Nature of katG Mutants and the Possible Role of AhpC. *PLoS One* **11**, e0166807 (2016).
31. Kim, M.-J. *et al.* Caseation of human tuberculosis granulomas correlates with elevated host lipid metabolism. *EMBO Mol. Med.* **2**, 258–274 (2010).
32. Marrero, J., Rhee, K. Y., Schnappinger, D., Pethe, K. & Ehrt, S. Gluconeogenic carbon flow of tricarboxylic acid cycle intermediates is critical for Mycobacterium tuberculosis to establish and maintain infection. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 9819–9824 (2010).
33. Russell, D. G., Huang, L. & VanderVen, B. C. Immunometabolism at the interface between macrophages and pathogens. *Nat. Rev. Immunol.* **19**, 291–304 (2019).
34. Russell, D. G. Mycobacterium tuberculosis and the intimate discourse of a chronic infection. *Immunol. Rev.* **240**, 252–268 (2011).
35. van der Vusse, G. J. Albumin as fatty acid transporter. *Drug Metab. Pharmacokinet.* **24**, 300–307 (2009).
36. Sankaranarayanan, S. *et al.* Serum albumin acts as a shuttle to enhance cholesterol efflux from cells. *J. Lipid Res.* **54**, 671–676 (2013).
37. Abdelmagid, S. A. *et al.* Comprehensive profiling of plasma fatty acid concentrations in young healthy Canadian adults. *PLoS One* **10**, e0116195 (2015).

38. Bhaskar, A. *et al.* Reengineering redox sensitive GFP to measure mycothiol redox potential of Mycobacterium tuberculosis during infection. *PLoS Pathog.* **10**, e1003902 (2014).
39. Lamont, E. A. & Baughn, A. D. Impact of the host environment on the antitubercular action of pyrazinamide. *EBioMedicine* **49**, 374–380 (2019).
40. Kempker, R. R. *et al.* Lung Tissue Concentrations of Pyrazinamide among Patients with Drug-Resistant Pulmonary Tuberculosis. *Antimicrob. Agents Chemother.* **61**, (2017).
41. McDonough, K. A., Kress, Y. & Bloom, B. R. Pathogenesis of tuberculosis: interaction of Mycobacterium tuberculosis with macrophages. *Infect. Immun.* **61**, 2763–2773 (1993).
42. Flynn, J. L. & Chan, J. Immune evasion by Mycobacterium tuberculosis: living with the enemy. *Curr. Opin. Immunol.* **15**, 450–455 (2003).
43. Ohkuma, S. & Poole, B. Fluorescence probe measurement of the intralysosomal pH in living cells and the perturbation of pH by various agents. *Proc. Natl. Acad. Sci. U. S. A.* **75**, 3327–3331 (1978).
44. MacMicking, J. D., Taylor, G. A. & McKinney, J. D. Immune control of tuberculosis by IFN-gamma-inducible LRG-47. *Science* **302**, 654–659 (2003).
45. Huynh, K. K. & Grinstein, S. Regulation of Vacuolar pH and Its Modulation by Some Microbial Species. *Microbiol. Mol. Biol. Rev.* **71**, 452–462 (2007).
46. Müller, S., Dennemärker, J. & Reinheckel, T. Specific functions of lysosomal proteases in endocytic and autophagic pathways. *Biochim. Biophys. Acta* **1824**, 34–43 (2012).
47. Lau, Y. H., Giessen, T. W., Altenburg, W. J. & Silver, P. A. Prokaryotic nanocompartments form synthetic organelles in a eukaryote. *Nat. Commun.* **9**, 1311 (2018).
48. Lončar, N., Rozeboom, H. J., Franken, L. E., Stuart, M. C. A. & Fraaije, M. W. Structure of a robust bacterial protein cage and its application as a versatile biocatalytic platform through enzyme encapsulation. *Biochem. Biophys. Res. Commun.* **529**, 548–553 (2020).
49. Rahmanpour, R. & Bugg, T. D. H. Assembly in vitro of Rhodococcus jostii RHA1 encapsulin and peroxidase DypB to form a nanocompartment. *FEBS J.* **280**, 2097–2104 (2013).
50. Giessen, T. W. & Silver, P. A. Converting a Natural Protein Compartment into a Nanofactory for the Size-Constrained Synthesis of Antimicrobial Silver Nanoparticles. *ACS Synth. Biol.* **5**, 1497–1504 (2016).
51. Goyal, N., Kashyap, B., Singh, N. P. & Kaur, I. R. Neopterin and oxidative stress markers in the diagnosis of extrapulmonary tuberculosis. *Biomarkers* **22**, 648–653 (2017).
52. Jamaati, H. *et al.* Nitric Oxide in the Pathogenesis and Treatment of Tuberculosis. *Front. Microbiol.* **8**, 2008 (2017).
53. Rosenberg, O. S. *et al.* Substrates Control Multimerization and Activation of the Multi-Domain ATPase Motor of Type VII Secretion. *Cell* **161**, 501–512 (2015).
54. Ehrt, S. *et al.* Controlling gene expression in mycobacteria with anhydrotetracycline and Tet repressor. *Nucleic Acids Res.* **33**, e21 (2005).

55. Larsen, M. H., Biermann, K., Tandberg, S., Hsu, T. & Jacobs, W. R., Jr. Genetic Manipulation of Mycobacterium tuberculosis. *Curr. Protoc. Microbiol.* **Chapter 10**, Unit 10A.2 (2007).
56. Long J. E. *et al.* Identifying essential genes in Mycobacterium tuberculosis by global phenotypic profiling. *Methods Mol. Biol.* **1279**, 79–95 (2015).

Chapter 4

Conclusion

4.1 Summary

Our work highlights the growing evidence that the subcellular organization of prokaryotes is ubiquitous. Encapsulins, or prokaryotic nanocompartments, are among the most widespread of the prokaryotic organelles, yet their structure and functional importance are understudied[1]. This dissertation employs biochemistry, microbiology, and structural biology in attempt to better address the many unanswered questions about the functionality of encapsulins.

In Chapter 1, we overview the history of the research on encapsulins, even before they became appreciated as nanocompartments, and we outline what has been uncovered since their discovery as prokaryotic nanocompartments in 2008. We focus on the structural biology of the nanocompartments, which is the most well-characterized aspect of these systems. We also describe the biological importance of these encapsulins and underscore the many open questions pertaining to the physiological function of encapsulins as organelles.

In Chapter 2, we experimentally characterize a novel family of encapsulins that had previously been identified as a candidate nanocompartment in a bioinformatic study. We demonstrate that this novel encapsulin family, which we name Family 2A, is evolutionarily distinct from the previously characterized encapsulins, and show that it hosts a cysteine desulfurase cargo enzyme. In our study of this nanocompartment in the freshwater cyanobacteria, *Synechococcus elongatus* PCC 7942, we implicate a potential role for the nanocompartment in sulfur starvation stress response. We elucidated the signal sequence that targets cargo for encapsulation by Family 2A encapsulins. We obtained the structure of this nanocompartment via cryo-electron microscopy and were able to ascertain the site of cargo binding within the compartment, in addition to characterizing the potential size and charge selectivity of the organelle pores. Finally, we demonstrated that encapsulation of the cysteine desulfurase cargo enzyme enhances its activity.

In Chapter 3, we investigate the physiological role of a previously identified nanocompartment – the encapsulin from the human pathogen *Mycobacterium tuberculosis* (Mtb). Here we investigate the native enzyme cargo of this system by purifying encapsulins natively from Mtb. We show that the encapsulin and its cargo enzyme, DypB, are stable in environments that mimic the chemical milieu of the phagolysosome, and we suggest the biological function of the Mtb encapsulin may be to ameliorate reactive oxygen species within this subcellular niche of macrophages. Additionally, we demonstrate that Mtb encapsulin mutants are more susceptible to killing upon oxidative stress challenge. Furthermore, we implicate a link between Mtb encapsulins and lipid metabolism and identify it as a potential antibiotic target to be used in tandem with pyrazinamide treatment.

4.2 Discussion and future directions

4.2a Nanocompartment systems are widespread throughout prokaryotic genomes

Prior to the discovery of prokaryotic nanocompartments, much of the research on protein-bounded organelles has been primarily in the study of bacterial microcompartments such as the carboxysome and the propane-diol utilization microcompartment. These microcompartments are all evolutionary relatives and share a common BMC domain in the shell proteins that form the exterior of the compartment[2–4]. It is now evident that other protein-bounded compartments that are evolutionarily distinct from the microcompartments exist. These compartments, called prokaryotic nanocompartments are not close relatives of the bacterial microcompartments, but rather are more closely related to bacteriophage capsids[5–7].

When the first bacterial nanocompartment from *Thermotoga maritima* was characterized as an organelle by Sutter and colleagues in 2008, researchers soon began to study the close homologs of encapsulin systems that shared sequence homology with the *T. maritima* shell sequence. Using this approach many diverse encapsulin cargo systems were discovered[8–12]. Even among the close relatives of the *T. maritima* encapsulin, the associated cargo systems that have been shown to be encapsulated are diverse. These cargo types include the ferritin-like protein (FLP), the dye-decolorizing peroxidase DypB, the copper oxidase-hydroxylamine oxidoreductase (HAO), and the iron-mineralizing encapsulin-associated Firmicute (IMEF) protein. The physiological role of these systems appear to be varied as the predicted reactions for the various cargo types are widely different. For example the HAO cargo enzyme is predicted to detoxify hydroxylamine in anammox bacteria,[13,14] whereas the FLP-containing encapsulin has been shown to mineralize iron[11].

While the diversity of the close homologs of the first discovered encapsulin appeared vast, the search for novel encapsulin systems had been limited simply to the Basic Local Alignment Search Tool (BLAST)[12,15]. Potentially, even more unrelated encapsulin systems had yet to be discovered, but their identification as nanocompartments could not be achieved with a simple sequence homology search. Radford postulated that perhaps other bacteriophage-like compartments may exist in prokaryotes[6]. In this approach, all prokaryotic genomes were searched against a bacteriophage capsid database, likely prophages were discarded, and hits with putative neighboring cargo open reading frames were enriched to elucidate a larger set of candidate genes that may be prokaryotic nanocompartments.

This bioinformatic search revealed there exist many families of encapsulins that are evolutionarily distinct, with all of them sharing the phage-like HK97 fold. Prior to our work, these putative compartments had yet to be experimentally characterized. In Chapter 2, we demonstrate that one of these encapsulin families, Family 2A, is in fact an encapsulin system unique from the canonical previously studied Family 1 encapsulins. Our study, however, was limited to the cysteine desulfurase-containing Family 2A encapsulin system from *S. elongatus* PCC 7942, and we did not explore the physiology or biochemistry of

Family 2A systems from other species. Furthermore, we did not examine the closely related subfamily, Family 2B.

It is likely, as more and more prokaryotic genomes are sequenced that we uncover additional HK97-fold encapsulin systems with unique cargo enzymes and diverse physiological functions. Discovery of these systems will simply require the repeated bioinformatic approach that Radford has described[6] followed by the experimental methodology of characterizing these new systems, which we outline in Chapter 2[16]. Yet, even with this new expanded approach to finding novel nanocompartments, the search for other HK97-fold compartments may be narrow-minded. The array of protein-bounded compartments today is enough to suggest that protein compartments in prokaryotes likely extend beyond the compartment types we know today, such as bacterioferritins, microcompartments, and encapsulins. More recently, it was discovered that a phage-encoded microtubule-like protein formed a nucleus-like structure in *Pseudomonas chlororaphis* cells infected with the bacteriophage 201φ2-1[17,18]. We are likely just scratching the surface of the protein-bounded organelles that exist in prokaryotes. As the field moves forward in its search for additional protein-bounded compartments, it is important that we do not limit our search of new subcellular structures based on those that have already been characterized.

4.2b Encapsulin cargo enzyme activity is modulated by encapsulation

The effect of encapsulation on the activity of the sequestered enzyme is a fundamental question of encapsulin systems for which there are few answers in the literature. In our investigation of the Family 2A encapsulin, we observed an increase in the activity of the cysteine desulfurase cargo upon encapsulation. Specifically, we measured a roughly seven-fold increase for the k_{cat} of the enzyme. Changes in the activity of the cargo enzyme upon encapsulation have been measured before in the Family 1 encapsulins. For the DypB-containing encapsulin from *Rhodococcus jostii* RHA 1, the activity per milligram of DypB was shown to be eight-fold higher upon encapsulation[10]. The mechanism as to how the cargo activity is being modulated is unknown. It is possible that encapsulation of the cargo changes the conformational state of the enzyme into a more active form. Experimental validation of this would be difficult given the current methods available. Structural characterization of the conformational state of the encapsulated cargo will be challenging using cryo-electron microscopy. As we have reported in Chapter 2 using the Family 2A encapsulin, the cargo protein is not highly ordered and, therefore, a high resolution structure of the encapsulated cargo is difficult to obtain. Hydrogen-deuterium exchange is an alternative method that may provide insight on whether the cargo undergoes a conformational change upon encapsulation[19,20].

More recently, it has been reported that encapsulation of cargo increases the thermostability of the compartmentalized enzyme[21]. Using the encapsulin system from *Mycolicibacterium hassiacum*, it was found that when the purified protein was incubated at an increased temperature of 40°C, the unencapsulated DypB lost almost all activity after 30 minutes, whereas the encapsulated DypB was stable and active for over 3 hours[21]. Preliminary data from our lab suggests that non-native proteins targeted to the *T. maritima* (Tm) encapsulin interior exhibit a significant increase in thermostability. One

possible explanation for this result is the combination of macromolecular crowding and confinement stabilizing cargo proteins loaded inside the encapsulin[22]. Another possibility is the encapsulin acts as an 'Anifinsen cage' allowing the encapsulated cargo to fold while limiting aggregation[23]. To uncover how confinement within Tm encapsulin modulates thermostability, differential scanning calorimetry (DSC) could be employed to study whether the melting temperature of the cargo changes significantly upon encapsulation. The thermostable Tm encapsulin is an excellent model for this approach as the melting temperature of the Tm encapsulin, >100°C[24], should be substantially higher than any heterologously loaded cargo. Therefore, the unfolding temperatures of the two proteins should be easily distinguishable.

To better understand how the degree of crowding within the compartment effects cargo stability, one could tune the loading of cargo protein within the compartment to modulate the copy-number of cargo loaded and perform DSC to compare the melting temperature of cargo in both highly crowded (~10-30 copies of heterologous cargo) versus loosely packed conditions (~1-2 copies of cargo).

4.2c Fate of sulfur after catalysis by cysteine desulfurase

Our characterization of the cysteine desulfurase cargo found in the encapsulin from *S. elongatus* PCC 7942 was limited in that we were measuring the activity of cysteine desulfurase as a function of alanine production using a coupled enzyme assay with alanine dehydrogenase. We did not explore the ultimate fate of the sulfur group upon cysteine's catalysis by cysteine desulfurase. After the cysteine desulfurase active site is persulfurated, the enzyme will often interact with a downstream sulfide carrier protein which will shuttle the sulfur to downstream metabolites[25]. Sulfide carriers such as SufE and IscU have been shown to shuttle sulfur to protein complexes involved in the biosynthesis of iron-sulfur clusters or the synthesis of sulfur-containing metabolites such as thiamin and biotin[26,27]. If the cysteine desulfurase cargo is encapsulated within a nanocompartment, then what would be the downstream carrier protein? It is possible that the encapsulin shell itself may act as the downstream carrier. In this scenario the shell protein would require a lumenal-facing cysteine residue that would act as an acceptor for sulfide transfer. In our structural characterization of the nanocompartment, we find that two of the three shell cysteines are extrashellular-facing with the third cysteine found on a flexible loop near the exterior of the compartment. It is possible that this cysteine is able to transiently enter the lumen to access the cysteine desulfurase, although it is unlikely.

Alternatively, the sulfur group may remain within the compartment and a poly-sulfide chain may build off of the active site cysteine of the cysteine desulfurase. There is precedent for accumulation of polysulfide on the active site of sulfur transfer proteins, such as SufE, in absence of downstream proteins to accept the sulfur group[28]. If this were the case, the encapsulin may act as a storage compartment for sulfide. The physiological role for a sulfide store may serve to provide a source of sulfur after a sulfate starvation — similar to what has been suggested for the ferrosome compartment in *Desulfovibrio magneticus*' transition out of iron starvation[29,30]. There are many questions that arise if this model is to be considered, mainly: how is the accumulated sulfur released from the compartment upon the cell requiring access to the store?

Addressing the ultimate destination of the sulfur group, originating from the catalyzed cysteine, can be answered using multiple approaches. Radiolabeled L[^{35}S]-cysteine can be used as a substrate for the purified encapsulated or non-encapsulated cysteine desulfurase. Transfer of the sulfur group to the encapsulin shell protein can be assessed by non-reducing PAGE followed by autoradiography as has been done previously for studying sulfide transfer proteins[31]. An alternative approach would be to perform mass spectrometry of the cysteine desulfurase-loaded encapsulin with or without addition of the cysteine substrate to observe whether a polysulfide chain is accumulating on the cysteine desulfurase active site similar to what Ollagnier-de-Choudens and colleagues did in their studies of the SufS-SufE cysteine desulfurase[28].

4.3 References

1. Greening, C. & Lithgow, T. Formation and function of bacterial organelles. *Nat. Rev. Microbiol.* (2020) doi:10.1038/s41579-020-0413-0.
2. Yeates, T. O., Thompson, M. C. & Bobik, T. A. The protein shells of bacterial microcompartment organelles. *Curr. Opin. Struct. Biol.* **21**, 223–231 (2011).
3. Chowdhury, C., Sinha, S., Chun, S., Yeates, T. O. & Bobik, T. A. Diverse bacterial microcompartment organelles. *Microbiol. Mol. Biol. Rev.* **78**, 438–468 (2014).
4. Kerfeld, C. A., Aussignargues, C., Zarzycki, J., Cai, F. & Sutter, M. Bacterial microcompartments. *Nat. Rev. Microbiol.* **16**, 277–290 (2018).
5. Sutter, M. *et al.* Structural basis of enzyme encapsulation into a bacterial nanocompartment. *Nat. Struct. Mol. Biol.* **15**, 939–947 (2008).
6. Radford, D. Understanding the encapsulins: prediction and characterization of phage capsid-like nanocompartments in prokaryotes. (2015).
7. Nichols, R. J., Cassidy-Amstutz, C., Chaijarasphong, T. & Savage, D. F. Encapsulins: molecular biology of the shell. *Crit. Rev. Biochem. Mol. Biol.* **52**, 583–594 (2017).
8. McHugh, C. A. *et al.* A virus capsid-like nanocompartment that stores iron and protects bacteria from oxidative stress. *EMBO J.* **33**, 1896–1911 (2014).
9. Contreras, H. *et al.* Characterization of a Mycobacterium tuberculosis nanocompartment and its potential cargo proteins. *J. Biol. Chem.* **289**, 18279–18289 (2014).
10. Rahmanpour, R. & Bugg, T. D. H. Assembly in vitro of Rhodococcus jostii RHA1 encapsulin and peroxidase DypB to form a nanocompartment. *FEBS J.* **280**, 2097–2104 (2013).
11. He, D. *et al.* Structural characterization of encapsulated ferritin provides insight into iron storage in bacterial nanocompartments. *Elife* **5**, (2016).
12. Giessen, T. W. & Silver, P. A. Widespread distribution of encapsulin nanocompartments reveals functional diversity. *Nat Microbiol* **2**, 17029 (2017).
13. Kartal, B. *et al.* How to make a living from anaerobic ammonium oxidation. *FEMS Microbiol. Rev.* **37**, 428–461 (2013).
14. Xing, C.-Y. *et al.* A self-assembled nanocompartment in anammox bacteria for resisting intracelluar hydroxylamine stress. *Sci. Total Environ.* **717**, 137030 (2020).
15. Johnson, M. *et al.* NCBI BLAST: a better web interface. *Nucleic Acids Res.* **36**, W5–9 (2008).
16. Nichols, R. J. *et al.* Discovery and characterization of a novel family of prokaryotic nanocompartments involved in sulfur metabolism. *bioRxiv* 2020.05.24.113720 (2020) doi:10.1101/2020.05.24.113720.
17. Chaikeeratisak, V. *et al.* Viral Capsid Trafficking along Treadmilling Tubulin Filaments in Bacteria. *Cell* **177**, 1771–1780.e12 (2019).
18. Malone, L. M. *et al.* A jumbo phage that forms a nucleus-like structure evades CRISPR–Cas DNA targeting but is vulnerable to type III RNA-based immunity. *Nature Microbiology* **5**, 48–55 (2019).
19. Resetca, D. & Wilson, D. J. Characterizing rapid, activity-linked conformational transitions in proteins via sub-second hydrogen deuterium exchange mass spectrometry. *FEBS J.* **280**, 5616–5625 (2013).

20. Houde, D., Arndt, J., Domeier, W., Berkowitz, S. & Engen, J. R. Characterization of IgG1 Conformation and Conformational Dynamics by Hydrogen/Deuterium Exchange Mass Spectrometry. *Anal. Chem.* **81**, 5966 (2009).
21. Lončar, N., Rozeboom, H. J., Franken, L. E., Stuart, M. C. A. & Fraaije, M. W. Structure of a robust bacterial protein cage and its application as a versatile biocatalytic platform through enzyme encapsulation. *Biochem. Biophys. Res. Commun.* **529**, 548–553 (2020).
22. Zhou, H.-X., Rivas, G. & Minton, A. P. Macromolecular crowding and confinement: biochemical, biophysical, and potential physiological consequences. *Annu. Rev. Biophys.* **37**, 375–397 (2008).
23. Ellis, R. J. Protein folding: importance of the Anfinsen cage. *Curr. Biol.* **13**, R881–3 (2003).
24. Cassidy-Amstutz, C. *et al.* Identification of a Minimal Peptide Tag for in Vivo and in Vitro Loading of Encapsulin. *Biochemistry* **55**, 3461–3468 (2016).
25. Hidese, R., Mihara, H. & Esaki, N. Bacterial cysteine desulfurases: versatile key players in biosynthetic pathways of sulfur-containing biofactors. *Appl. Microbiol. Biotechnol.* **91**, 47–61 (2011).
26. Black, K. A. & Dos Santos, P. C. Shared-intermediates in the biosynthesis of thio-cofactors: Mechanism and functions of cysteine desulfurases and sulfur acceptors. *Biochim. Biophys. Acta* **1853**, 1470–1480 (2015).
27. Loiseau, L., Ollagnier-de-Choudens, S., Nachin, L., Fontecave, M. & Barras, F. Biogenesis of Fe-S cluster by the bacterial Suf system: SufS and SufE form a new type of cysteine desulfurase. *J. Biol. Chem.* **278**, 38352–38359 (2003).
28. Ollagnier-de-Choudens, S. *et al.* Mechanistic studies of the SufS--SufE cysteine desulfurase: evidence for sulfur transfer from SufS to SufE. *FEBS Lett.* **555**, 263–267 (2003).
29. Byrne, M. E. *et al.* Desulfovibrio magneticus RS-1 contains an iron- and phosphorus-rich organelle distinct from its bullet-shaped magnetosomes. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 12263–12268 (2010).
30. Grant, C. R. & Komeili, A. Ferrosomes are iron storage organelles formed by broadly conserved gene clusters in bacteria and archaea. *bioRxiv* 2020.01.10.902569 (2020) doi:10.1101/2020.01.10.902569.
31. Outten, F. W., Wood, M. J., Munoz, F. M. & Storz, G. The SufE protein and the SufBCD complex enhance SufS cysteine desulfurase activity as part of a sulfur transfer pathway for Fe-S cluster assembly in Escherichia coli. *J. Biol. Chem.* **278**, 45713–45719 (2003).