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

Authors: Robert J. Nichols[1,2], Benjamin LaFrance[1,2], Naiya R. Phillips[2], Luke M. Oltrogge[2], Luis E. Valentin-Alvarado[4], Amanda J. Bischoff[3], Eva Nogales[2,5,6]*, and David F. Savage[2,]*

[1]these authors contributed equally
[2]Department of Molecular and Cell Biology, University of California Berkeley, Berkeley, CA, USA.
[3]Department of Chemistry, University of California Berkeley, Berkeley, CA, USA.
[4]Department of Plant and Microbial Biology, University of California, Berkeley, CA, USA.
[5]Howard Hughes Medical Institute, University of California, Berkeley, CA, USA.
[6]Molecular Biophysics and Integrated Bioimaging Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA.
*email: enogales@lbl.gov and savage@berkeley.edu

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.

Introduction

Subcellular compartmentalization is an essential 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 (Grant et al., 2018). 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 (Oltrogge et al., 2020; Mangan et al., 2016; Kerfeld et al., 2018). In addition to modulating cargo 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 (Sampson and Bobik, 2008; Crowley et al., 2010; Chowdhury et al., 2015; Kerfeld et al., 2018).

Recently, another class of protein-bounded compartments, known as prokaryotic nanocompartments, has been discovered (Sutter, M., et al., 2008). 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 (Giessen, 2016; Nichols et al., 2017). 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) (Sutter et al., 2008; Giessen et al., 2019; Giessen and Silver, 2017; Xing et al., 2020). 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 (Kim et al., 2009; Kim et al., 2019; McHugh et al., 2014). 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 (Giessen et al., 2019; Giessen and Silver, 2017; He et al., 2016). 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 (Giessen and Silver, 2017; Kartal et al., 2013; Xing et al., 2020).

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 (Nichols et al., 2017). 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 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.

Results

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 (Giessen and Silver, 2017; Nichols et al., 2017). 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 (Koonin and Krupovic, 2018; Krupovic et al., 2019; Krupovic and Koonin, 2017), and the existence of other bacteriophage-related nanocompartments beyond the close relatives of known encapsulins has been postulated (Radford, 2015; Giessen, 2016; Nichols et al., 2017). 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 (Radford, 2015). 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 (Supp. Table 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 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 (Supp. Figure 1-1). 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 1B; Supp. Figure 1-2). 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 (Supp. Table 2 and 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 1A).

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 1C), the typical size for many previously characterized encapsulins (Cassidy-Amstutz et al., 2016; Snijder et al., 2016). 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 (Supp. Figure 1-3; Cassidy-Amstutz et

al., 2016). 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 1D, Supp. Figure 1-4).

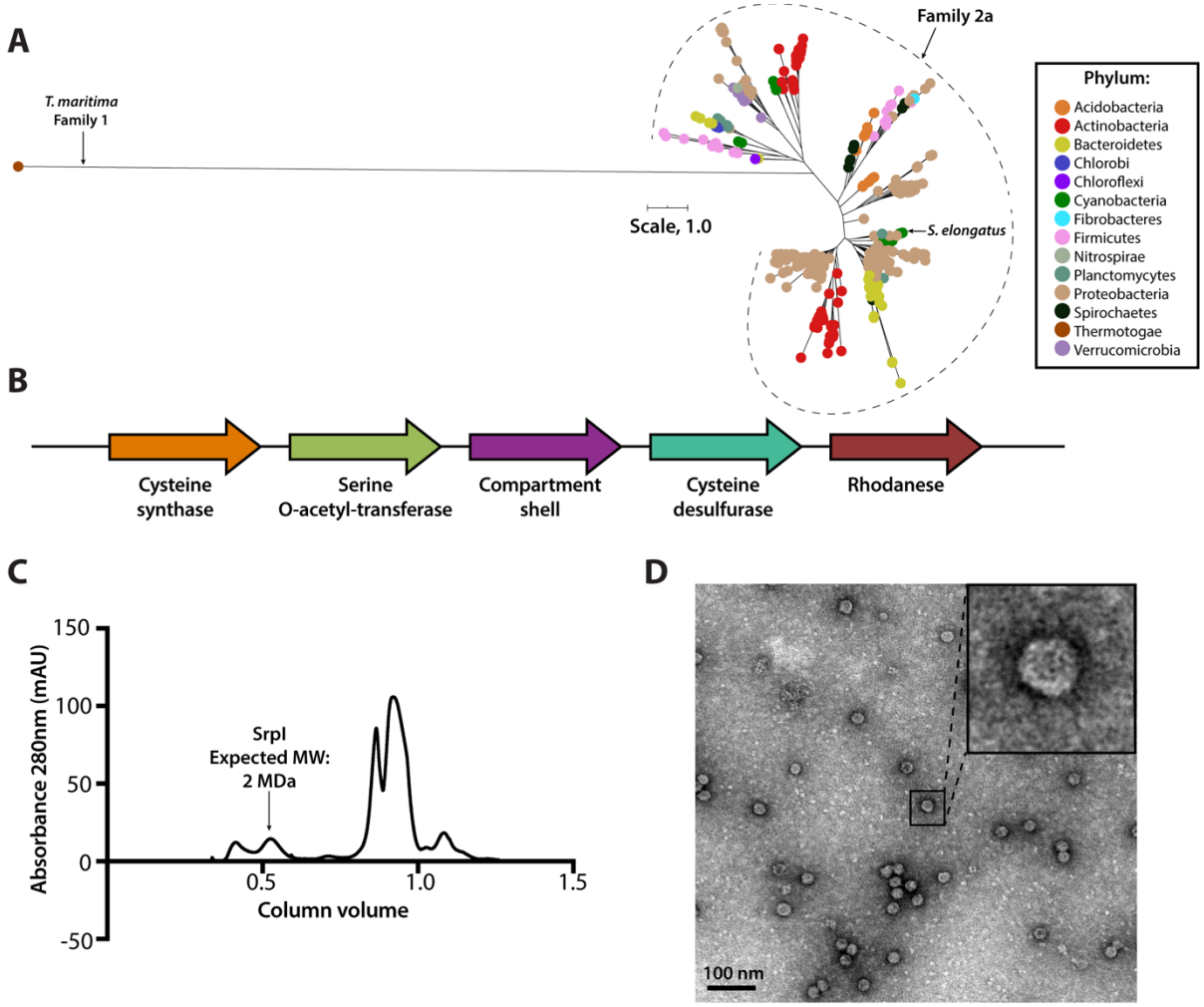


Figure 1: SrpI is a bacterial nanocompartment that is widespread throughout bacterial phyla and found neighboring sulfur metabolism genes. (A) Maximum-likelihood phylogenetic tree of Family 2a encapsulin shell proteins using the *T. maritima* Family 1 encapsulin shell protein (WP_004080898.1) as a Family 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.

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 (Nicholson et al., 1995; Nicholson and Laudenbach, 1995). Thus, this gene, which is found on a plasmid encoding many sulfur-related genes, was termed SrpI for Sulfur regulated plasmid-encoded gene-I (Chen et al., 2008; Nicholson and Laudenbach, 1995). 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.

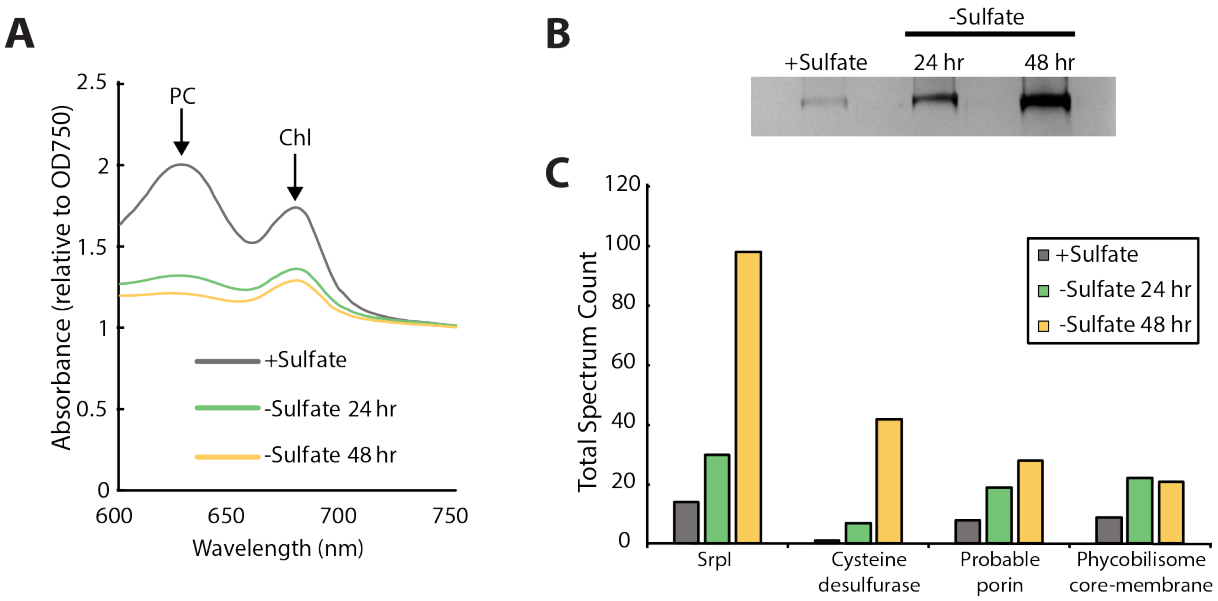


Figure 2: SrpI encapsulin is upregulated in *S. elongatus* upon sulfate starvation. (A) Absorbance spectra of *S. elongatus* liquid cultures under nutrient-replete conditions (+Sulfate) and sulfur starvation (-Sulfate) for 24 and 48 hours. Absorbance maxima of phycocyanin (PC) at 620 nm and chlorophyll (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 (Collier and Grossman, 1992). Chlorosis was confirmed by loss of phycocyanin absorbance at 620 nm (Figure 2A). During this time-course, SDS-PAGE analysis of cell lysates also indicated upregulation of a high molecular weight complex (Figure 2B). 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 2C). Taken together, these results suggest that the cysteine desulfurase, which neighbors the SrpI shell gene, is the encapsulated cargo protein (Supp. Table 2, Nichols et al., 2017).

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 3A, Supp. Figure 3-1). 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 3A). 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 (Supp Figure 3-2; Figure 3B).

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 3C). 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.

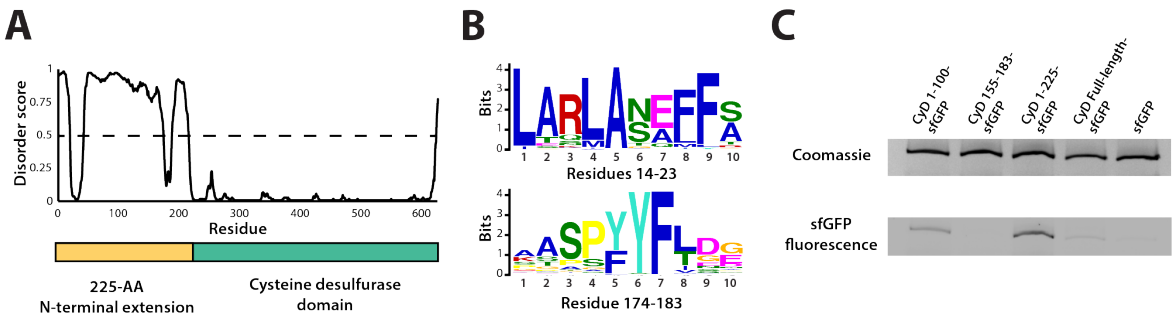


Figure 3: An N-terminal signal sequence directs cargo loading *in vivo*. (A) Domain organization of cysteine desulfurase (CyD; Synpcc7942_B2661) and the predicted disorder scores calculated using DISOPRED3. CyD can be split into two domains – a highly disordered N-terminal domain and an ordered cysteine 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.

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 (Cassidy-Amstutz et al., 2016). 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 (Supp. Fig. 4-1). 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 4A). 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 4B).

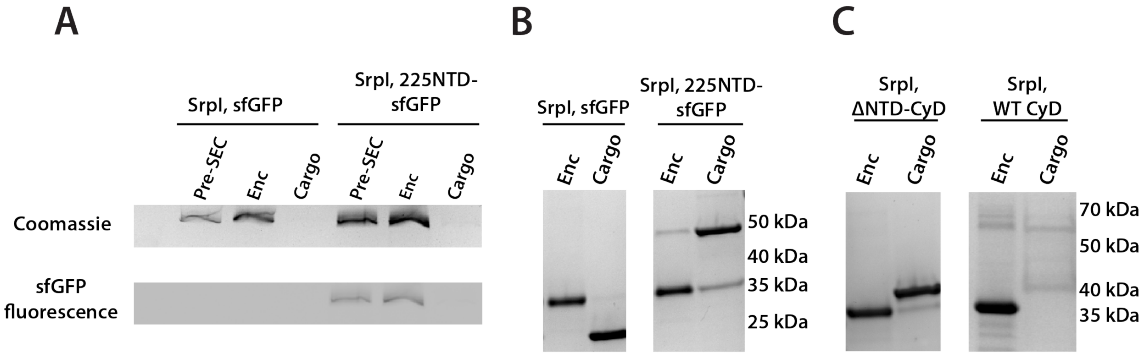


Figure 4: The 225-NTD of CyD is necessary and sufficient for cargo loading *in vitro*. (A) Non-denaturing SDS-PAGE of sfGFP or the CyD N-terminal domain-sfGFP fusion (225NTD-sfGFP) loaded *in vitro* into SrpI encapsulin. sfGFP fluorescence followed by Coomassie staining of the encapsulin (Enc), cargo, 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.

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 4C). 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 4C).

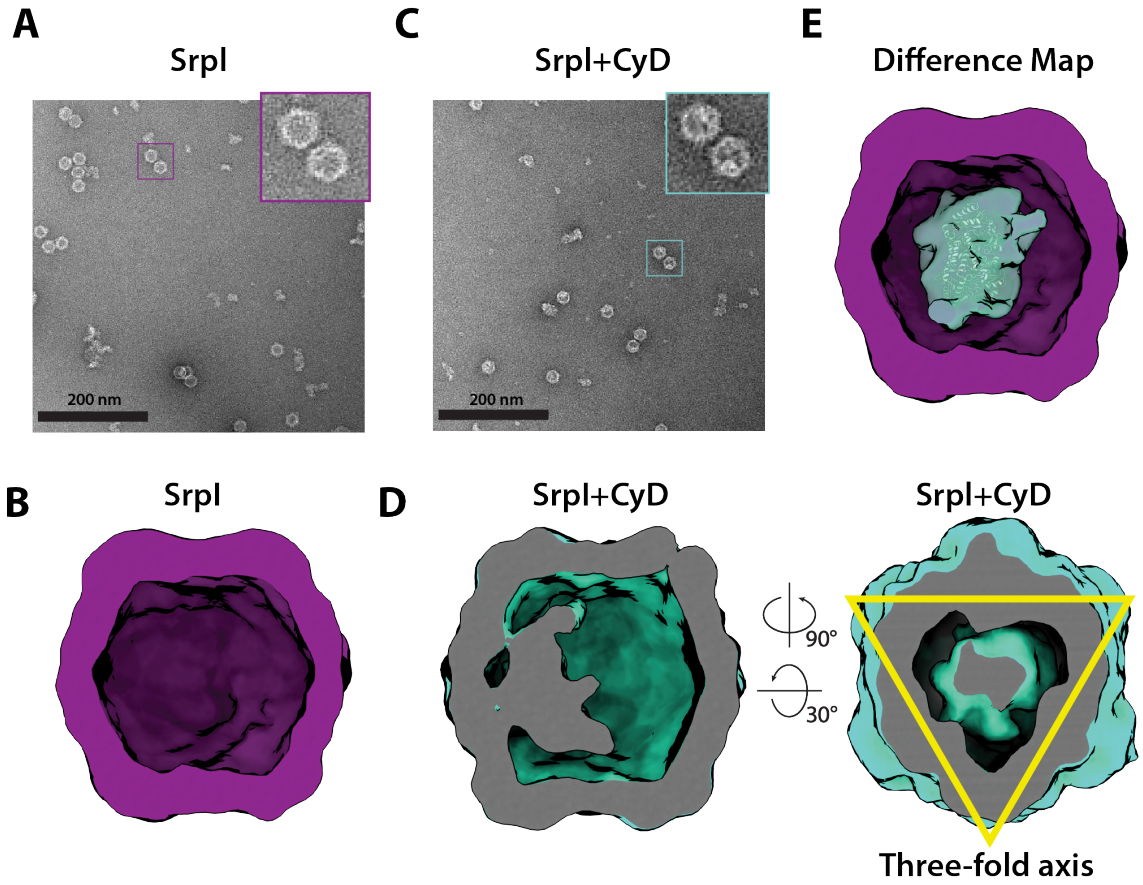


Figure 5: Negative stain analysis indicates CyD loading into SrpI encapsulin (A) Negative stain micrograph of an apo-SrpI shell in contrast with (C) the holo-SrpI shell that includes the CyD cargo. (B) and (D) 3D reconstruction of apo-SrpI and holo-SrpI, respectively. (E) Difference map showing additional density for the holo-SrpI with a homologous cysteine desulfurase dimer docked in (pdb:6c9e).

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 5A,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 5B, 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 5D). 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 5E).

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 (Sutter et al., 2008; Akita et al., 2007; McHugh et al., 2014; Giessen et al., 2019). 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 6A, Supp. Figure 6-1, Supp. Figure 6-2). 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 6A), 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 6B, Supp. Figure 6-3 C).

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 (Supp. Figure 6-3C; Duda and Teschke, 2019). Similar to other HK97 bacteriophage capsids, the N-arm of the SrpI shell intertwines with the neighboring subunit to create a chainmail-like topology (Supp. Figure 6-4). 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 (Supp. Figure 6-3 D). 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.

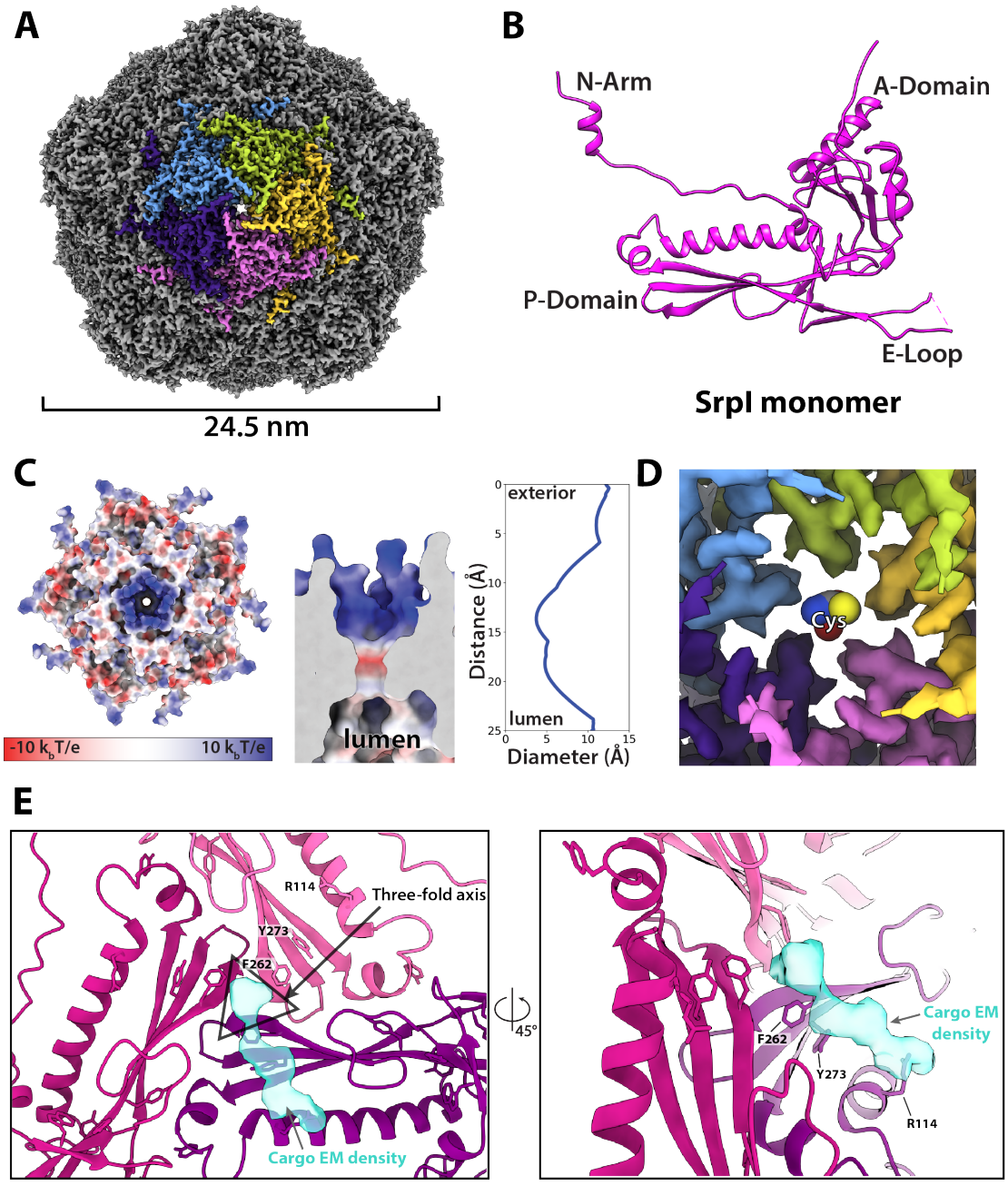


Figure 6 CryoEM structure of the SrpI encapsulin reveals a common HK97 fold, a potential mechanism for cysteine selectivity, and a cargo binding site (A) The SrpI encapsulin structure at 2.2Å resolution. This SrpI encapsulin forms a T=1 icosahedral structure 24.5 nm in diameter. Five subunits around 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 3.7Å 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.

Interestingly, the A-domain of SrpI that forms a pore at the 5-fold axis is composed of residues that are positively charged (Figure 6C, Supp. Figure 6-5), in contrast to the negatively charged five-fold axis pore of Family 1 encapsulins (Giessen et al., 2019). It is thought that the 5-fold pore at the capsid vertices creates a selective barrier to allow encapsulin substrates into the compartment lumen (Giessen et al., 2019; Nichols et al., 2017; Sutter et al., 2008). 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 (Mangan et al., 2016), roughly 30-54% of free cysteine will have a net charge of -1 and therefore could traverse the positively charged pore exterior (Cameselle et al., 1986). Likewise, the size of the pore is also an important constraint for limiting the spectrum of substrates that can enter the compartment (Williams et al., 2018). The SrpI encapsulin 5-fold pore is 3.7Å in diameter at its most restrictive point (Figure 6C) as calculated by HOLE (Smart et al., 1996). Furthermore, modeling of cysteine in the pore demonstrates it is likely small enough to enter the nanocompartment (Figure 6D).

Unfortunately, during processing and classification the holo-SrpI encapsulin proved nearly indistinguishable from the apo-SrpI control (Supp. Figure 6-1). 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 (Sutter et al., 2008). However, symmetry expansion and focused classification of the holo-SrpI encapsulin revealed additional EM density at the interior surface of the shell that was localized to the three-fold symmetry axis (Figure 6E). This corroborates our findings from the holo-SrpI structure obtained via negative stain TEM, which also demonstrated cargo interfacing with the shell at the axis of three-fold symmetry (Figure 5D).

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 6E). 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 (Supp Figure 6-3 A,B). Namely, residues F262, and Y273 are located near the suggested cargo density.

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 (Dos Santos, 2017). Unencapsulated CyD was active towards a cysteine 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 (Rahmanpour and Bugg, 2013).

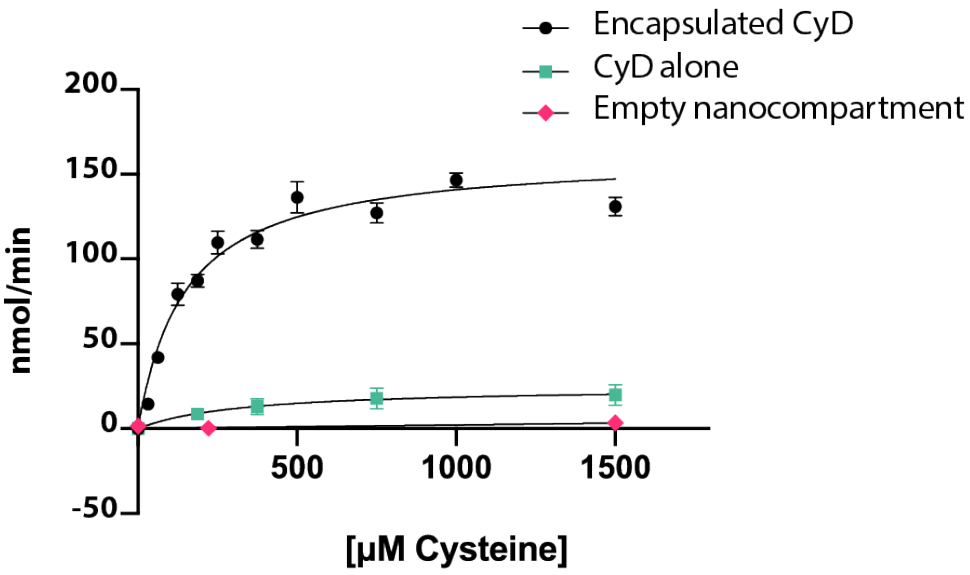


Figure 7: Cysteine desulfurase activity is enhanced upon encapsulation. Substrate-dependent activity of encapsulated cysteine desulfurase (encapsulated CyD), unencapsulated cysteine desulfurase (CyD alone), and empty nanocompartment using a coupled-enzyme assay with alanine dehydrogenase and production of NADH as a readout of cysteine desulfurase activity. Error bars are SD of 3-6 replicate experiments.

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.

Structural analysis of Srpl 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 (Nichols et al., 2017; Williams et al., 2018; Giessen et al., 2019). Structures determined for the Family 1 encapsulins reveal pores that are negatively charged (Giessen et al., 2019; McHugh et al., 2014; Sutter et al., 2008). 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 (Giessen et al., 2019; He et al., 2016). 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 (Cameselle et al., 1986; Mangan et al., 2016). 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 (Williams et al., 2018). 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*.

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 (Rahmanpour and Bugg, 2013). 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 (He et al., 2016; McHugh et al., 2014). 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 (Loiseau et al., 2003; Outten et al., 2003). 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 (Outten et al., 2003; Singh et al., 2013). 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.

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 (Baier et al., 2014; Collier and Grossman, 1992; Karradt et al., 2008). Among the other known responses to sulfur starvation are the upregulation of proteins involved in sulfate transport CysA, CysT, CysW, and SbpA (Green et al., 1989; Laudenbach and Grossman, 1991). 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 (Kopriva et al., 2008; Schmidt, 1990; Schmidt and Christen, 1978).

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 (Supp. Table 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 (Giordano et al., 2005; Pilson, 2012; Tipping et al., 1998). 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 (Black and Dos Santos, 2015). 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 (Giessen et al., 2019; He et al., 2016).

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 (Winter et al., 1995). 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* (Abdellrazeq et al., 2018; Abdellrazeq et al., 2020; Leite et al., 2015). 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 (Leite et al., 2015). 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.

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 (Koonin and Krupovic, 2018). It is possible that the HK97 fold derives from a cellular ancestor and was then recruited by a virus-like ancestor of Caudovirales phage (Krupovic and Koonin, 2017). Alternatively, the HK97 fold was of viral origin and cellular hosts co-opted the compartments as organelles that enabled some fitness benefit (Koonin and Krupovic, 2018; Krupovic and Koonin, 2017). 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 (Radford, 2015). 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.

Methods

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 (Kalyaanamoorthy et al., 2017) 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 (Letunic and Bork, 2019).

Genome neighborhood analysis of the Family 2 encapsulin sequences was performed using the Enzyme Function Initiative suite of web tools (Gerlt, 2017; Zallot et al., 2019). 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 (Buchan and Jones, 2019; Jones and Cozzetto, 2015). 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 (Sievers et al., 2011). Conserved motifs within the cysteine desulfurase sequence homologs were determined using the MEME Suite 5.1.1 (Bailey et al., 2009). 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.

Molecular cloning, protein expression and purification

All plasmids were constructed using either Gibson Assembly (NEB) or SLiCE (Zhang et al., 2012) 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 (Andersen et al., 2013). 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 0.5mM 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 (25mM Tris-HCl pH 7.5, 150mM NaCl, 20mM 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 (25mM Tris-HCl pH 7.5, 150mM NaCl, 40mM imidazole) prior to eluting with 25mM Tris-HCl pH 7.5, 150mM NaCl, 350mM imidazole. The eluate was then concentrated and desalted into 25mM Tris-HCl pH 8, 150mM 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).

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 (Palmer and Wingfield, 2012). Briefly, cell pellets were resuspended in solution buffer (50mM Tris-HCl pH 8, 1% Triton-X100, 100mM NaCl, 10mM 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 (50mM Tris-HCl pH 8, 100mM NaCl, 10mM 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 (50mM Tris-HCl pH 7.4, 6M guanidine hydrochloride, 50mM DTT), flash frozen and stored at -80°C for future use.

Refolding was performed by 100-fold dilution in refolding buffer (50mM CAPS pH 10, 250mM arginine, 150mM NaCl, 20% Glycerol, 10mM 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 50mM CAPS pH 10, 250mM arginine, 150mM NaCl. Subsequent purification was performed using either a Superose™ 6 Increase column or a HiPrep™ 16/60 Sephacryl® S-500 HR (GE Life Sciences).

S. elongatus PCC 7942 growth and sulfate deprivation

S. elongatus PCC 7942 was grown in BG-11 media (Allen, 1968) 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 (Collier and Grossman, 1992), 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.

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.

Cysteine desulfurase activity

Cysteine desulfurase activity was performed using a coupled enzyme assay with alanine dehydrogenase. Reactions were carried out at 25°C in 25mM Tris-HCl pH 8, 150mM NaCl, 5mM 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.

Negative Stain TEM

Holo-SrpI and apo-SrpI samples were diluted to 300 nM in TEM Buffer (50mM CAPS pH 10, 250mM arginine, 150mM NaCl). 4uL 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 40uL droplets Uranyl Formate (UF; 2% w/v in water). To ensure stain was thick, and penetrated the shell interior, the grids sat for 30sec with a droplet of UF. Grids were briefly side-blotted with Whatman filter paper for 1 sec, 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 (Scheres, 2012). Briefly, CTF estimation was performed using CTFFIND4 (Rohou and Grigorieff, 2015), 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.

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 2mg/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 Supplemental Table 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 (Scheres, 2012), with defocus estimation using CTFFIND4 (Rohou and Grigorieff, 2015). Particles were automatically picked with LoG-picker and processed in accordance with Supplemental Figure 6-1.

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 (Emsley et al., 2010), and then refined with nearest neighbors using the real space refinement tool in PHENIX (Adams et al., 2010). 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 (Smart et al., 1996). 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 (Goddard et al., 2018). 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 (Goddard et al., 2018); The PyMOL Molecular Graphics System, Version 2.3.2 Shrödinger, LLC).

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 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.

References:

Abdellrazeq GS, Elnaggar MM, Bannantine JP, Park KT, Souza CD, Backer B, Hulubei V, Fry LM, Khaliel SA, Torky HA, Schneider DA, Davis WC. 2018. 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.

Abdellrazeq GS, Fry LM, Elnaggar MM, Bannantine JP, Schneider DA, Chamberlin WM, Mahmoud AHA, Park K-T, Hulubei V, Davis WC. 2020. 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*.

Adams PD, Afonine PV, Bunkóczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung L-W, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson JS, Terwilliger TC, Zwart PH. 2010. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**:213–221.

Akita F, Chong KT, Tanaka H, Yamashita E, Miyazaki N, Nakaishi Y, Suzuki M, Namba K, Ono Y, Tsukihara T, Nakagawa A. 2007. The crystal structure of a virus-like particle from the hyperthermophilic archaeon Pyrococcus furiosus provides insight into the evolution of viruses. *J Mol Biol* **368**:1469–1483.

Allen MM. 1968. SIMPLE CONDITIONS FOR GROWTH OF UNICELLULAR BLUE-GREEN ALGAE ON PLATES 1, 2. *J Phycol* **4**:1–4.

Andersen KR, Leksa NC, Schwartz TU. 2013. Optimized E. coli expression strain LOBSTR eliminates common contaminants from His-tag purification. *Proteins: Struct Funct Bioinf* **81**:1857–1861.

Baier A, Winkler W, Korte T, Lockau W, Karradt A. 2014. 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.

Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS. 2009. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res* **37**:W202–8.

Black KA, Dos Santos PC. 2015. Shared-intermediates in the biosynthesis of thio-cofactors: Mechanism and functions of cysteine desulfurases and sulfur acceptors. *Biochim Biophys Acta* **1853**:1470–1480.

Buchan DWA, Jones DT. 2019. The PSIPRED Protein Analysis Workbench: 20 years on. *Nucleic Acids Res* **47**:W402–W407.

Cameselle JC, Ribeiro JM, Sillero A. 1986. 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.

Cassidy-Amstutz C, Oltrogge L, Going CC, Lee A, Teng P, Quintanilla D, East-Seletsky A, Williams ER, Savage DF. 2016. Identification of a Minimal Peptide Tag for in Vivo and in Vitro Loading of Encapsulin. *Biochemistry* **55**:3461–3468.

Chen Y, Holtman CK, Magnuson RD, Youderian PA, Golden SS. 2008. The complete sequence and functional analysis of pANL, the large plasmid of the unicellular freshwater cyanobacterium Synechococcus elongatus PCC 7942. *Plasmid* **59**:176–192.

Chowdhury C, Chun S, Pang A, Sawaya MR, Sinha S, Yeates TO, Bobik TA. 2015. Selective molecular transport through the protein shell of a bacterial microcompartment organelle. *Proc Natl Acad Sci U S A* **112**:2990–2995.

Collier JL, Grossman AR. 1992. Chlorosis induced by nutrient deprivation in Synechococcus

690 sp. strain PCC 7942: not all bleaching is the same. *J Bacteriol* **174**:4718–4726.
691 Crowley CS, Cascio D, Sawaya MR, Kopstein JS, Bobik TA, Yeates TO. 2010. Structural
692 insight into the mechanisms of transport across the Salmonella enterica Pdu
693 microcompartment shell. *J Biol Chem* **285**:37838–37846.
694 Dos Santos PC. 2017. Chapter Seven - B. subtilis as a Model for Studying the Assembly of
695 Fe–S Clusters in Gram-Positive Bacteria In: David SS, editor. Methods in Enzymology.
696 Academic Press. pp. 185–212.
697 Duda RL, Teschke CM. 2019. The amazing HK97 fold: versatile results of modest
698 differences. *Curr Opin Virol* **36**:9–16.
699 Emsley P, Lohkamp B, Scott WG, Cowtan K. 2010. Features and development of Coot. *Acta*
700 *Crystallogr D Biol Crystallogr* **66**:486–501.
701 Gerlt JA. 2017. Genomic Enzymology: Web Tools for Leveraging Protein Family Sequence–
702 Function Space and Genome Context to Discover Novel Functions. *Biochemistry*
703 **56**:4293–4308.
704 Giessen TW. 2016. Encapsulins: microbial nanocompartments with applications in
705 biomedicine, nanobiotechnology and materials science. *Curr Opin Chem Biol* **34**:1–10.
706 Giessen TW, Orlando BJ, Verdegaal AA, Chambers MG, Gardener J, Bell DC, Birrane G,
707 Liao M, Silver PA. 2019. Large protein organelles form a new iron sequestration system
708 with high storage capacity. *Elife* **8**. doi:10.7554/eLife.46070
709 Giessen TW, Silver PA. 2017. Widespread distribution of encapsulin nanocompartments
710 reveals functional diversity. *Nat Microbiol* **2**:17029.
711 Giordano M, Norici A, Hell R. 2005. Sulfur and phytoplankton: acquisition, metabolism and
712 impact on the environment. *New Phytol* **166**:371–382.
713 Goddard TD, Huang CC, Meng EC, Pettersen EF, Couch GS, Morris JH, Ferrin TE. 2018.
714 UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Sci*
715 **27**:14–25.
716 Grant CR, Wan J, Komeili A. 2018. Organelle Formation in Bacteria and Archaea. *Annu Rev*
717 *Cell Dev Biol* **34**:217–238.
718 Green LS, Laudenbach DE, Grossman AR. 1989. A region of a cyanobacterial genome
719 required for sulfate transport. *Proc Natl Acad Sci U S A* **86**:1949–1953.
720 He D, Hughes S, Vanden-Hehir S, Georgiev A, Altenbach K, Tarrant E, Mackay CL,
721 Waldron KJ, Clarke DJ, Marles-Wright J. 2016. Structural characterization of
722 encapsulated ferritin provides insight into iron storage in bacterial nanocompartments.
723 *Elife* **5**. doi:10.7554/eLife.18972
724 Jones DT, Cozzetto D. 2015. DISOPRED3: precise disordered region predictions with
725 annotated protein-binding activity. *Bioinformatics* **31**:857–863.
726 Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. 2017. ModelFinder:
727 fast model selection for accurate phylogenetic estimates. *Nat Methods* **14**:587–589.
728 Karradt A, Sobanski J, Mattow J, Lockau W, Baier K. 2008. NblA, a key protein of
729 phycobilisome degradation, interacts with ClpC, a HSP100 chaperone partner of a
730 cyanobacterial Clp protease. *J Biol Chem* **283**:32394–32403.
731 Kartal B, de Almeida NM, Maalcke WJ, den Camp HJMO, Jetten MSM, Keltjens JT. 2013.
732 How to make a living from anaerobic ammonium oxidation. *FEMS Microbiol Rev* **37**:428–
733 461.
734 Kerfeld CA, Aussignargues C, Zarzycki J, Cai F, Sutter M. 2018. Bacterial
735 microcompartments. *Nat Rev Microbiol* **16**:277–290.
736 Kim D, Choi J, Lee S, Hyun H, Lee K, Cho K. 2019. Mutants defective in the production of
737 encapsulin show a tan-phase-locked phenotype in Myxococcus xanthus. *J Microbiol*
738 **57**:795–802.
739 Kim D, Chung J, Hyun H, Lee C, Lee K, Cho K. 2009. Operon required for fruiting body
740 development in Myxococcus xanthus. *J Microbiol Biotechnol* **19**:1288–1294.

741 Koonin EV, Krupovic M. 2018. The depths of virus exaptation. *Curr Opin Virol* **31**:1–8.
742 Kopriva S, Patron NJ, Keeling P, Leustek T. 2008. Phylogenetic Analysis of Sulfate
743 Assimilation and Cysteine Biosynthesis in Phototrophic Organisms In: Hell R, Dahl C,
744 Knaff D, Leustek T, editors. Sulfur Metabolism in Phototrophic Organisms. Dordrecht:
745 Springer Netherlands. pp. 31–58.
746 Krupovic M, Dolja VV, Koonin EV. 2019. Origin of viruses: primordial replicators recruiting
747 capsids from hosts. *Nat Rev Microbiol* **17**:449–458.
748 Krupovic M, Koonin EV. 2017. Multiple origins of viral capsid proteins from cellular
749 ancestors. *Proc Natl Acad Sci U S A* **114**:E2401–E2410.
750 Laudenbach DE, Grossman AR. 1991. Characterization and mutagenesis of sulfur-regulated
751 genes in a cyanobacterium: evidence for function in sulfate transport. *J Bacteriol*
752 **173**:2739–2750.
753 Leite FL, Reinhardt TA, Bannantine JP, Stabel JR. 2015. Envelope protein complexes of
754 Mycobacterium avium subsp. paratuberculosis and their antigenicity. *Vet Microbiol*
755 **175**:275–285.
756 Letunic I, Bork P. 2019. Interactive Tree Of Life (iTOL) v4: recent updates and new
757 developments. *Nucleic Acids Res* **47**:W256–W259.
758 Loiseau L, Ollagnier-de-Choudens S, Nachin L, Fontecave M, Barras F. 2003. Biogenesis of
759 Fe-S cluster by the bacterial Suf system: SufS and SufE form a new type of cysteine
760 desulfurase. *J Biol Chem* **278**:38352–38359.
761 Mangan NM, Flamholz A, Hood RD, Milo R, Savage DF. 2016. pH determines the energetic
762 efficiency of the cyanobacterial CO2 concentrating mechanism. *Proc Natl Acad Sci U S*
763 *A* **113**:E5354–62.
764 McHugh CA, Fontana J, Nemecek D, Cheng N, Aksyuk AA, Bernard Heymann J, Winkler
765 DC, Lam AS, Wall JS, Steven AC, Hoiczyk E. 2014. A virus capsid-like
766 nanocompartment that stores iron and protects bacteria from oxidative stress. *EMBO J*
767 **33**:1896–1911.
768 Moesa HA, Wakabayashi S, Nakai K, Patil A. 2012. Chemical composition is maintained in
769 poorly conserved intrinsically disordered regions and suggests a means for their
770 classification. *Mol Biosyst* **8**:3262–3273.
771 Nicholson ML, Gaasenbeek M, Laudenbach DE. 1995. Two enzymes together capable of
772 cysteine biosynthesis are encoded on a cyanobacterial plasmid. *Mol Gen Genet*
773 **247**:623–632.
774 Nicholson ML, Laudenbach DE. 1995. Genes encoded on a cyanobacterial plasmid are
775 transcriptionally regulated by sulfur availability and CysR. *J Bacteriol* **177**:2143–2150.
776 Nichols RJ, Cassidy-Amstutz C, Chaijarasphong T, Savage DF. 2017. Encapsulins:
777 molecular biology of the shell. *Crit Rev Biochem Mol Biol* **52**:583–594.
778 Oltrogge LM, Chaijarasphong T, Chen AW, Bolin ER, Marqusee S, Savage DF. 2020.
779 Multivalent interactions between CsoS2 and Rubisco mediate α-carboxysome formation.
780 *Nat Struct Mol Biol*. doi:10.1038/s41594-020-0387-7
781 Outten FW, Wood MJ, Munoz FM, Storz G. 2003. The SufE protein and the SufBCD
782 complex enhance SufS cysteine desulfurase activity as part of a sulfur transfer pathway
783 for Fe-S cluster assembly in Escherichia coli. *J Biol Chem* **278**:45713–45719.
784 Palmer I, Wingfield PT. 2012. Preparation and Extraction of Insoluble (Inclusion-Body)
785 Proteins from Escherichia coli. *Curr Protoc Protein Sci*.
786 Pilson MEQ. 2012. An Introduction to the Chemistry of the Sea. Cambridge University
787 Press.
788 Radford D. 2015. Understanding the encapsulins: prediction and characterization of phage
789 capsid-like nanocompartments in prokaryotes.
790 Rahmanpour R, Bugg TDH. 2013. Assembly in vitro of Rhodococcus jostii RHA1 encapsulin
791 and peroxidase DypB to form a nanocompartment. *FEBS J* **280**:2097–2104.

Rohou A, Grigorieff N. 2015. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J Struct Biol* **192**:216–221.

Sampson EM, Bobik TA. 2008. Microcompartments for B12-dependent 1,2-propanediol degradation provide protection from DNA and cellular damage by a reactive metabolic intermediate. *J Bacteriol* **190**:2966–2971.

Scheres SHW. 2012. A Bayesian view on cryo-EM structure determination. *J Mol Biol* **415**:406–418.

Schmidt A. 1990. 21 - Sulphur Metabolism: D. Cysteine Synthase In: Lea PJ, editor. Methods in Plant Biochemistry. Academic Press. pp. 349–354.

Schmidt A, Christen U. 1978. A factor-dependent sulfotransferase specific for 3'-phosphoadenosine-5'-phosphosulfate (PAPS) in the Cyanobacterium Synechococcus 6301. *Planta* **140**:239–244.

Shih PM, Wu D, Latifi A, Axen SD, Fewer DP, Talla E, Calteau A, Cai F, Tandeau de Marsac N, Rippka R, Herdman M, Sivonen K, Coursin T, Laurent T, Goodwin L, Nolan M, Davenport KW, Han CS, Rubin EM, Eisen JA, Woyke T, Gugger M, Kerfeld CA. 2013. Improving the coverage of the cyanobacterial phylum using diversity-driven genome sequencing. *Proc Natl Acad Sci U S A* **110**:1053–1058.

Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J, Thompson JD, Higgins DG. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* **7**.

Singh H, Dai Y, Outten FW, Busenlehner LS. 2013. Escherichia coli SufE sulfur transfer protein modulates the SufS cysteine desulfurase through allosteric conformational dynamics. *J Biol Chem* **288**:36189–36200.

Smart OS, Neduvelil JG, Wang X, Wallace BA, Sansom MS. 1996. HOLE: a program for the analysis of the pore dimensions of ion channel structural models. *J Mol Graph* **14**:354–60, 376.

Snijder J, Kononova O, Barbu IM, Uetrecht C, Rurup WF, Burnley RJ, Koay MST, Cornelissen JJLM, Roos WH, Barsegov V, Wuite GJL, Heck AJR. 2016. Assembly and Mechanical Properties of the Cargo-Free and Cargo-Loaded Bacterial Nanocompartment Encapsulin. *Biomacromolecules* **17**:2522–2529.

Sutter M, Boehringer D, Gutmann S, Günther S, Prangishvili D, Loessner MJ, Stetter KO, Weber-Ban E, Ban N. 2008. Structural basis of enzyme encapsulation into a bacterial nanocompartment. *Nat Struct Mol Biol* **15**:939–947.

Tipping E, Carrick TR, Hurley MA, James JB, Lawlor AJ, Lofts S, Rigg E, Sutcliffe DW, Woof C. 1998. Reversal of acidification in upland waters of the English Lake District. *Environ Pollut* **103**:143–151.

Varadi M, Guharoy M, Zsolyomi F, Tompa P. 2015. DisCons: a novel tool to quantify and classify evolutionary conservation of intrinsic protein disorder. *BMC Bioinformatics* **16**:153.

Williams EM, Jung SM, Coffman JL, Lutz S. 2018. Pore Engineering for Enhanced Mass Transport in Encapsulin Nanocompartments. *ACS Synth Biol* **7**:2514–2517.

Winter N, Triccas JA, Rivoire B, Pessolani MC, Eiglmeier K, Lim EM, Hunter SW, Brennan PJ, Britton WJ. 1995. Characterization of the gene encoding the immunodominant 35 kDa protein of Mycobacterium leprae. *Mol Microbiol* **16**:865–876.

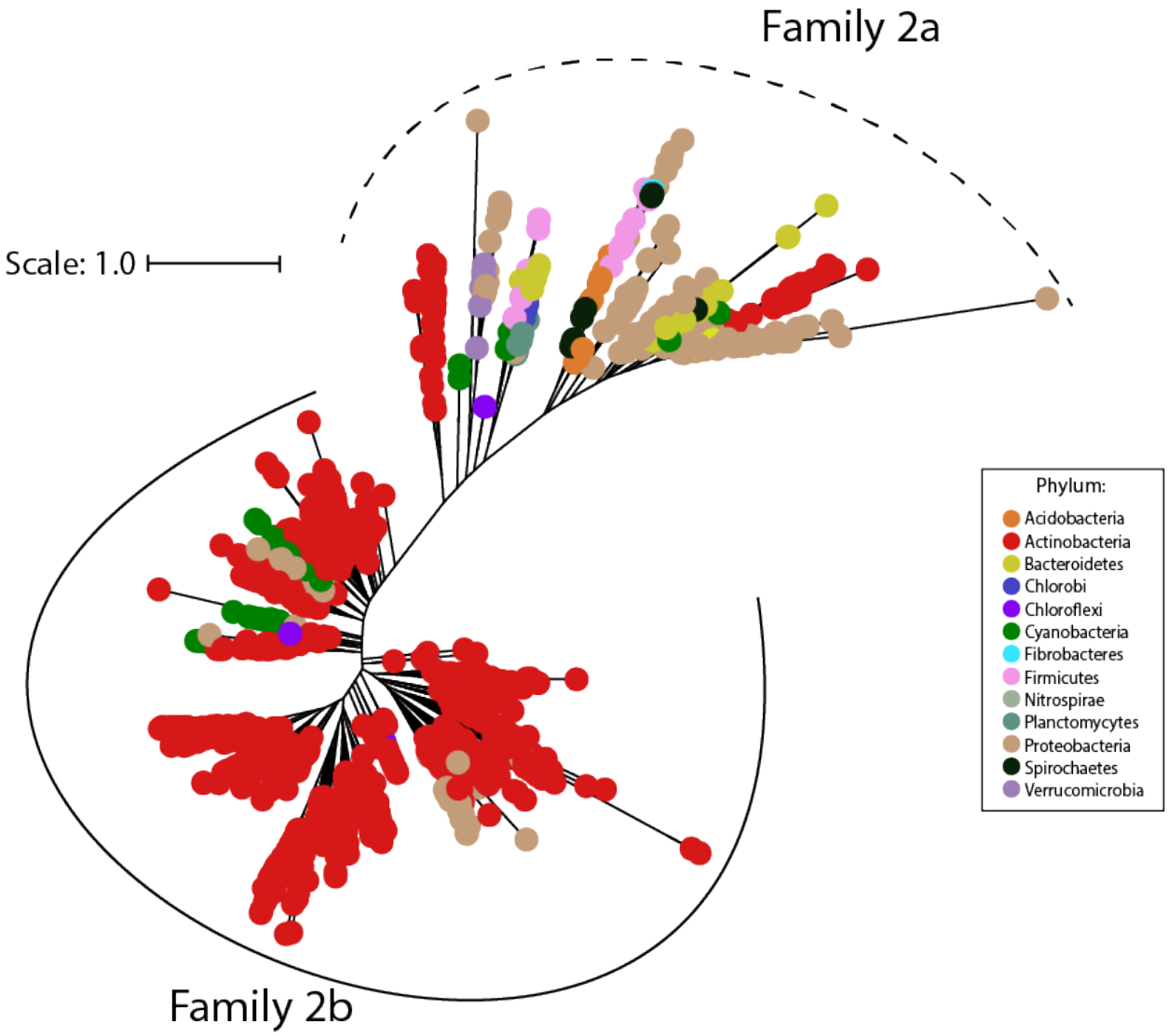
Xing C-Y, Fan Y-C, Chen X, Guo J-S, Shen Y, Yan P, Fang F, Chen Y-P. 2020. A self-assembled nanocompartment in anammox bacteria for resisting intracelluar hydroxylamine stress. *Sci Total Environ* **717**:137030.

Zallot R, Oberg N, Gerlt JA. 2019. 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.

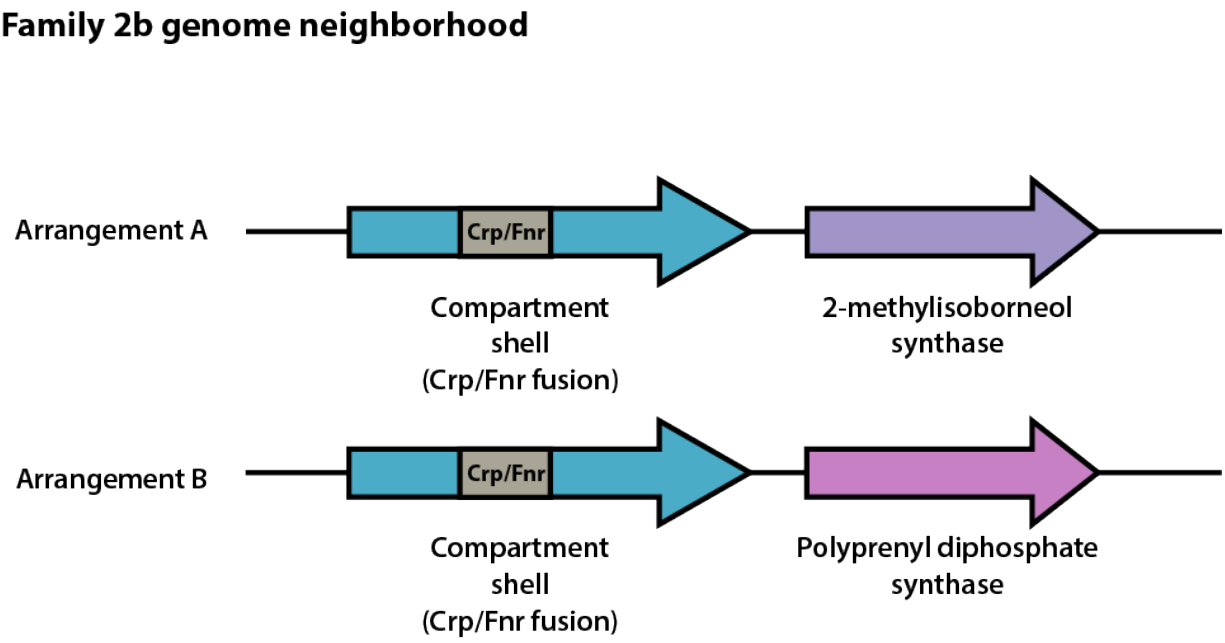
Zhang Y, Werling U, Edelmann W. 2012. SLiCE: a novel bacterial cell extract-based DNA

843 cloning method. *Nucleic Acids Res* **40**:e55.

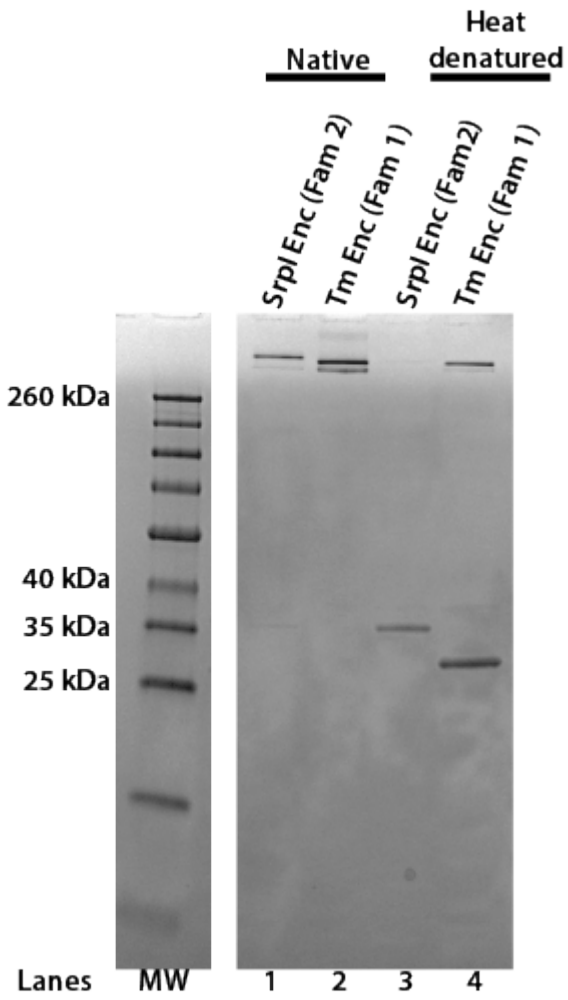
Supplementary Information:



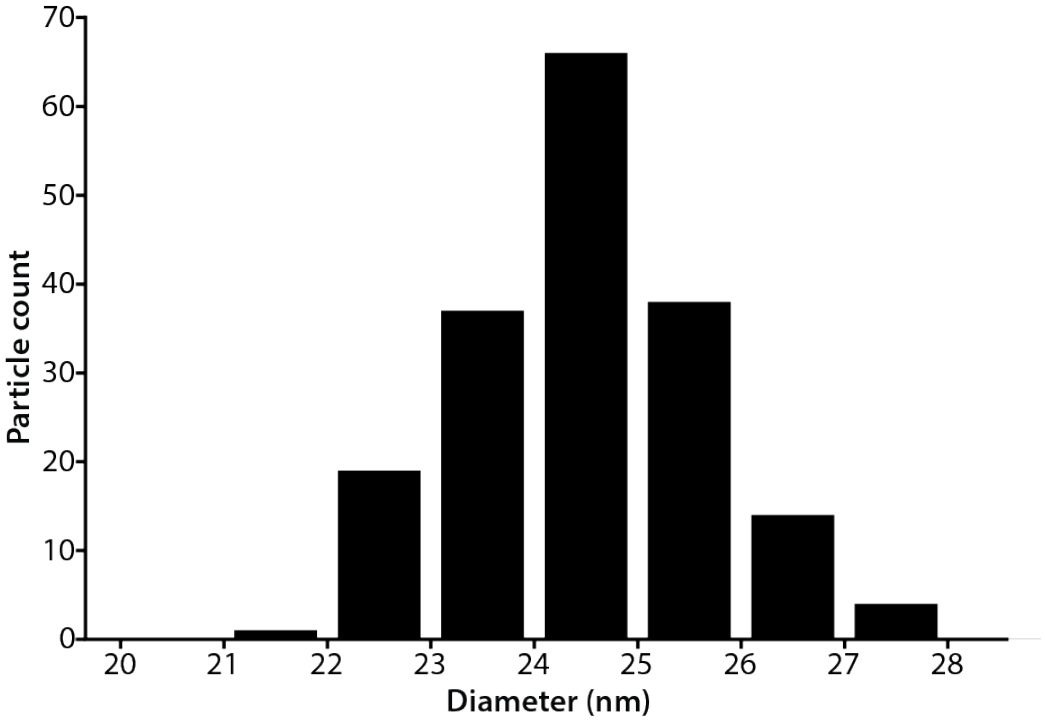
Supplementary Figure 1-1: Family 2 encapsulins can be divided into two phylogenetically distinct subfamilies. Maximum likelihood phylogenetic tree of 1383 members of Family 2 encapsulins. Family 2a (dashed line grouping) and Family 2b (solid line grouping) represent two subfamilies. Family 2 encapsulins are found throughout 13 different bacterial phyla and are distributed in a polyphyletic fashion. Scale bar, one substitution per site.



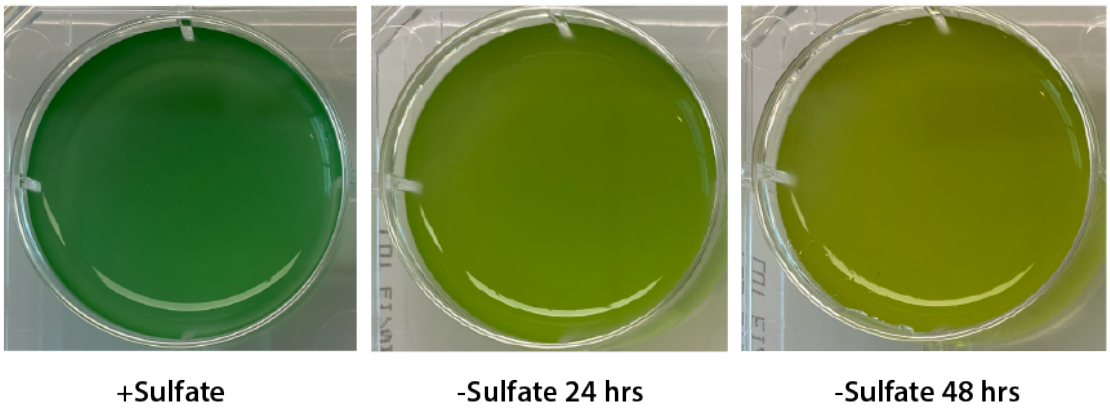
852
853 **Supplementary Figure 1-2: Family 2b shell genes neighbor 2-methylisoborneol synthase or polyprenyl**
854 **diphosphate synthase.** Schematic of representative genomic arrangements for Family 2b shell genes determined
855 using the EFI-GNT web tool. One or both arrangements (A and B) may be found within a given genome. Family 2b
856 shell genes possess a Crp/Fnr transcriptional regulator domain (grey) within the shell gene.



Supplementary Figure 1-3: The Family 2 encapsulin, SrpI, forms a high molecular weight complex similar to the Family 1 encapsulin from *T. maritima*. 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.



Supplementary Figure 1-4: 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.

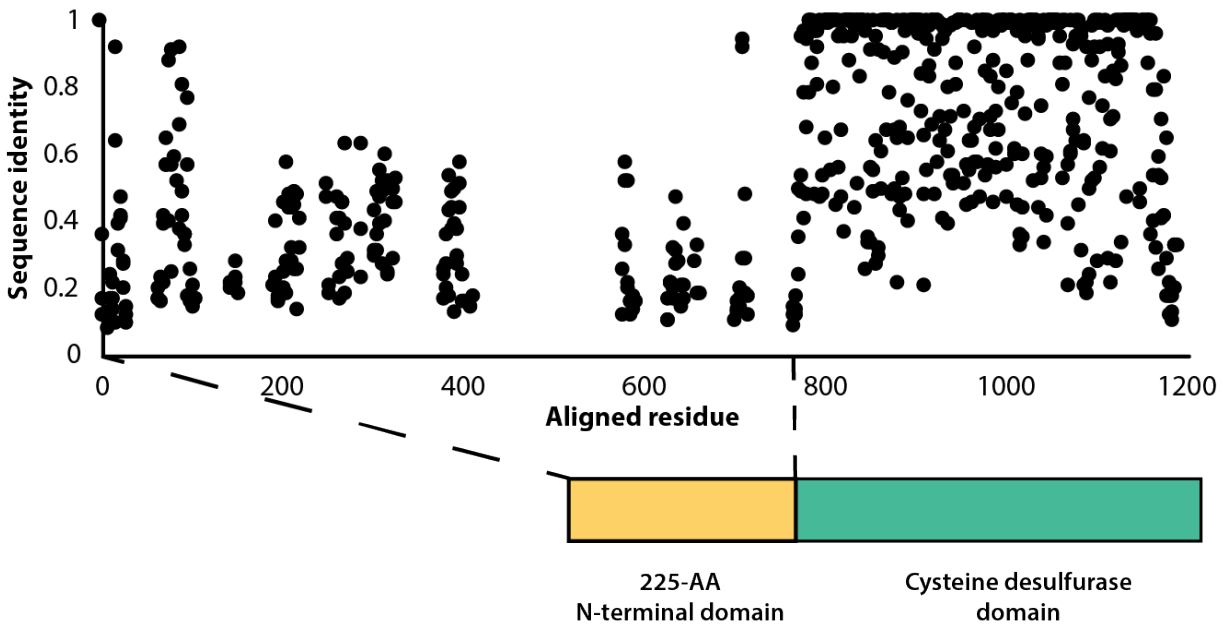


Supplementary Figure 2-1: Chlorosis phenotype of sulfur-starved *S. elongatus* PCC 7942. Liquid cultures of *S. elongatus* PCC 7942 grown in nutrient replete medium (+Sulfate) or under sulfate starvation for 24 hours and 48 hours.

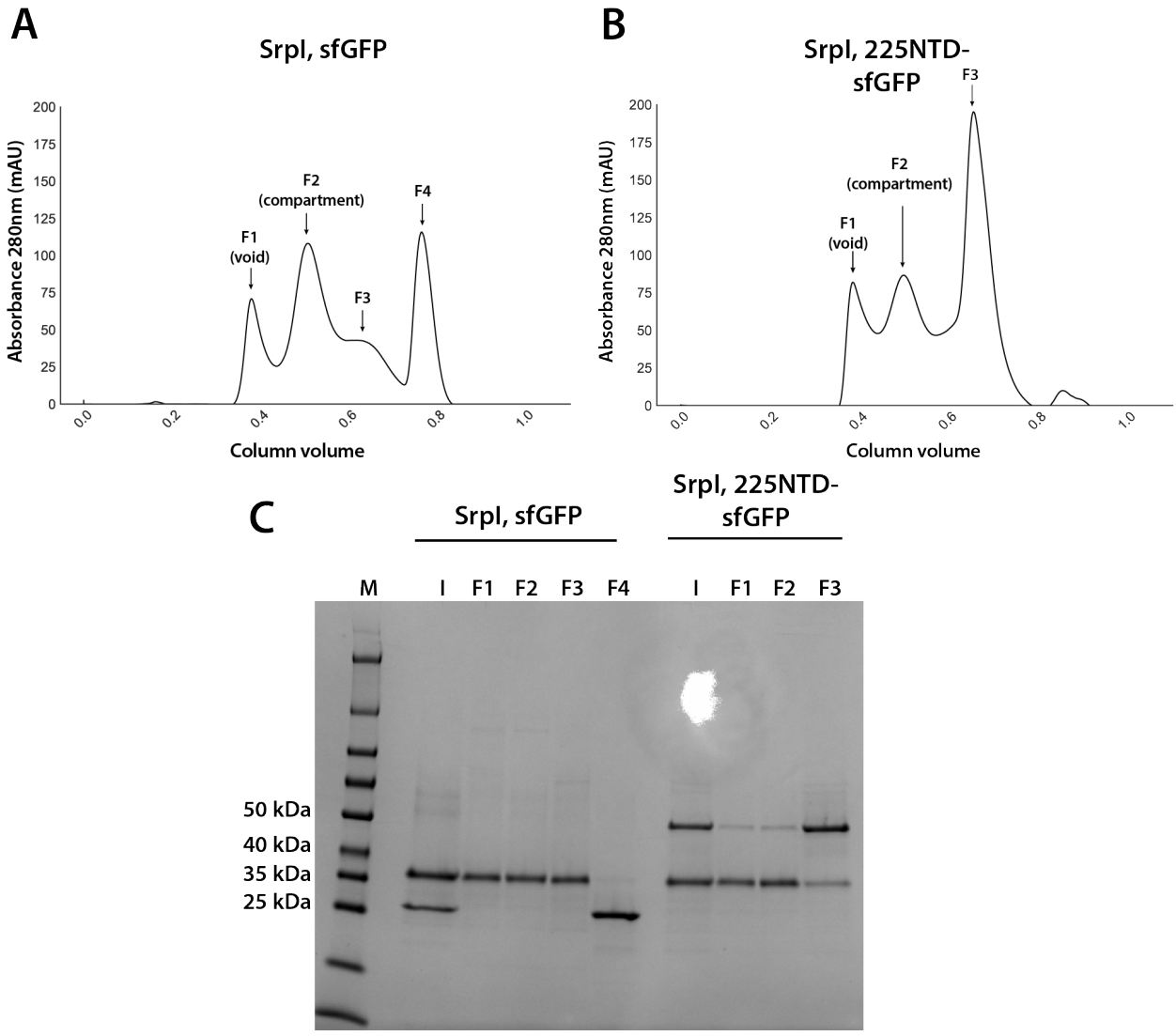
```
Synpcc7942_2558      ------------------------------------------------------------  0
Synpcc7942_1929      ------------------------------------------------------------  0
Synpcc7942_1255      ------------------------------------------------------------  0
Synpcc7942_B2661     MTNTVPSVPAVPNLPTQSDPFFNERSLEQLTQTVLQDLQQAGVSEAESAPTPLSVPTPAL  60
Synpcc7942_1738      ------------------------------------------------------------  0

Synpcc7942_2558      ------------------------------------------------------------  0
Synpcc7942_1929      ------------------------------------------------------------  0
Synpcc7942_1255      ------------------------------------------------------------  0
Synpcc7942_B2661     PTTSALAVPQSPTAIANVPAPPSSIDERSLAQLAQAVLQDPQLASAIASIFPSVTLPTSA  120
Synpcc7942_1738      ------------------------------------------------------------  0

Synpcc7942_2558      ------------------------------------------------------------  0
Synpcc7942_1929      ------------------------------------------------------------  0
Synpcc7942_1255      ------------------------------------------------------------  0
Synpcc7942_B2661     SVPRSVPVPPSFLPSLVPTAPPIHDEVGVIPHHQLPVPSQPTPAGLQQTASSKSGSGFYF  180
Synpcc7942_1738      ------------------------------------------------------------  0

Synpcc7942_2558      ------------------------------------------------------MSK  3
Synpcc7942_1929      ------------------------------------------------------------  0
Synpcc7942_1255      ---------------------------------MTRDRLLTHRQQFPALQNKA---  20
Synpcc7942_B2661     IDEQVETAIAALHSNLTVFPQLTTSSIPTLTGAHSAGAVGFDIHQVRRDFPILQERVNGR  240
Synpcc7942_1738      ------------------------------MTAIALTNLATEVRADFPILQQQVNGQ  27

Synpcc7942_2558      DPIYLDYHATTPVDSRVVTAMQPFWGEQFGNPASRSHRY--GLEAAAAVQVARETLATAI  61
Synpcc7942_1929      MSTYLDYSATTSMRPEVLERFTAVAQEQWGNAAS-LHQW--GNRAALVLERSRQQVAALI  57
Synpcc7942_1255      ---YFNYGGQGPLSTAALTAIQTAF---ADSQQLGPFSLATGQRVEQISDRLRTDLAALL  74
Synpcc7942_B2661     PLVWFDNAATTQKPQVVIDRLSHYYQHENSNIHRAAHELAA--RSTDAYEAAREQVRHFL  298
Synpcc7942_1738      PLVYLDNAATSQKPRAVIQSLVDYYEGYNSNVHRGVHTLSG--KATDAYEGARQKVARFI  85
                        :::  .       .:  :        ..     .      .     :  *  :   :

Synpcc7942_2558      AA-QPEELIFCSGATEANNLAIKGVAETQHRKGQHLITVATEHQAVLSPCRYLES-LGFQ  119
Synpcc7942_1929      QA-EPEAIAFSSGGTESDNWAILSPYLADPRP-GHLIISAVEHSAIARPAAWLEQ-RGWQ  114
Synpcc7942_1255      QT-QSENITLTENVSVGCN--IALWGVD-WQAGDRLLITDCEHPSVIATAQAIAQRFGVE  130
Synpcc7942_B2661     NAASTEEVVFVRGTTEAINLVAKSWGSQNLKEGDEIVITWLEHHANIVPWQQLSAETGAR  358
Synpcc7942_1738      NAKTEQEIVYTRNASEAINLVAYSFGMNFLQAGDEIILSAMEHHSNLIPWQFVAAKTGAA  145
                      :   : :    . : . *          :   .::    ** :       :    *

Synpcc7942_2558      LTVLPVNS---QGLISVEQVAAALRPDTILVSVMAANNEIGVLQPIAAIGALCRDRGVVF  176
Synpcc7942_1929      VTRLPVDR---SGRIQPADLASAVRPDTRLISVIWGQSEVGTIQPIAELAAIAREHGILF  171
Synpcc7942_1255      LDIWPLQAAVVAGQDLTEAFAAQLNSRTRLAAISHVLWNYGTVLPLAAIAEQAHRQDCLL  190
Synpcc7942_B2661     LRVVPVDD---YGQVRLDEYQKLLSDRTKIVSFTQVSNALGTITPAKEIIELAHRYGAKV  415
Synpcc7942_1738      LKFVGVTE---TGQFDLEQFRSLLSDRTKLVSVVHVSNTLGCCNPVTEICQLAHAKGARV  202
                      :    :      *          :   * : :.       *   *   :   .:  .  .

Synpcc7942_2558      HCDAAQALGRIPFNVSDLNVDLLSLSGHK-VYGPKGIGLLYRRSGVAI---APQLHGGSQ  232
Synpcc7942_1929      HTDAVQVAGRLPIDVQRLPIDLLSLSSHK-LYGPQGVGALYIRPGVEL---APLLQGGNQ  227
Synpcc7942_1255      LVDGAQSVGLLPLDLPATGVDFYAFTGHKWLCGPEGVGGLYSRPESRATSLPTWLGWRGV  250
Synpcc7942_B2661     LLDGAQSVSHLAVDVQALDCDWFVFSGHK-VFGPTGIGVLYGKQELLDAT-LPWQSGGNM  473
Synpcc7942_1738      LIDACQSLPHQAIDVQAIDCDWLVGSGHK-MCAPTGIGFLYGKLDLLRQM-PPFLGGGEM  260
                       *. *      .::     *    :.** : .* *:* ** :

Synpcc7942_2558      EQ---------------GLRAGTLAPALIVGFAQATELAIA--EQPTETLRLQTLRDRLWQ  276
Synpcc7942_1929      ES---------------GLRSGTPPIAAIAAFGEAAQLAAA--ELPHETARLQSLRDRLIA  271
Synpcc7942_1255      EKNTQGRPIRSKSDGRRYEIATSAYPLYAGLSAAIAQQADYGTIEDRYNRSRQLAQWLWQ  310
Synpcc7942_B2661     IADVTFEKTVYQPAPARFEAGTGNIADAVGLGAALEYVQKIGL-EAIAAYEHELLVHGTA  532
Synpcc7942_1738      IADVFLDHATYADLPHKFEAGTPAIGEAIALGAAIDYLTAIGM-DRIHAYEQQLTQHLFQ  319
                                    . .*       .:. *                 : *

Synpcc7942_2558      GLQAIGGLHLNGDPQQRLP---SNLNFSIE-GIDPSRLIQR----IRGAIAVSSGSACSS  328
Synpcc7942_1929      VLATEPRLRLTGDPIQRLP---HHASFIAR-GGTTGTSQQLVRAMNRLGFGISGGSACNS  327
Synpcc7942_1255      SLQALPKVRCLATTLPQ----AGLISFQIDSSQSP---FKIVSHLEGQGLQIRS------  357
Synpcc7942_B2661     LLSQIPGLRLIGTAPHK----AAVLSFVLE-GFSP---EAIGQALNREGIAVRAGHHCAQ  584
Synpcc7942_1738      RLAEIPELTVYGPTPEQDRDRAALAAFTAG-AVHP---HDLSTILDQSGIAIRAGHHCTQ  375
                      *     :   .    :         *    .                .: : .

Synpcc7942_2558      GNAEPSHVLTAIGRSPDLAHASLRIGLGRFTIAEEIDRTLEILIAAIQAER-SRLTRHKK  387
Synpcc7942_1929      GRSQPSPVLLAMGYSPQEALAGIRFSLGRSTQLAEVEAAAIALRSALHSLPQASLLSPA-  386
Synpcc7942_1255      -----------LVG-----PDCLRACCHYLTLETELESLVAAIAQI--------------  387
Synpcc7942_B2661     PILR------RFGL-----ETTVRPSLAFYNTFEELETLAAAIRRIQTGSL-A-L-----  626
Synpcc7942_1738      PLHR------ELQV-----QSTARASCYFYNTTDEIDRLIESLKEAVEFFG-A-IFS---  419
                                :           *      .   *::     :

Synpcc7942_2558      G 388
Synpcc7942_1929      - 386
Synpcc7942_1255      - 387
Synpcc7942_B2661     - 626
Synpcc7942_1738      - 419
```

873

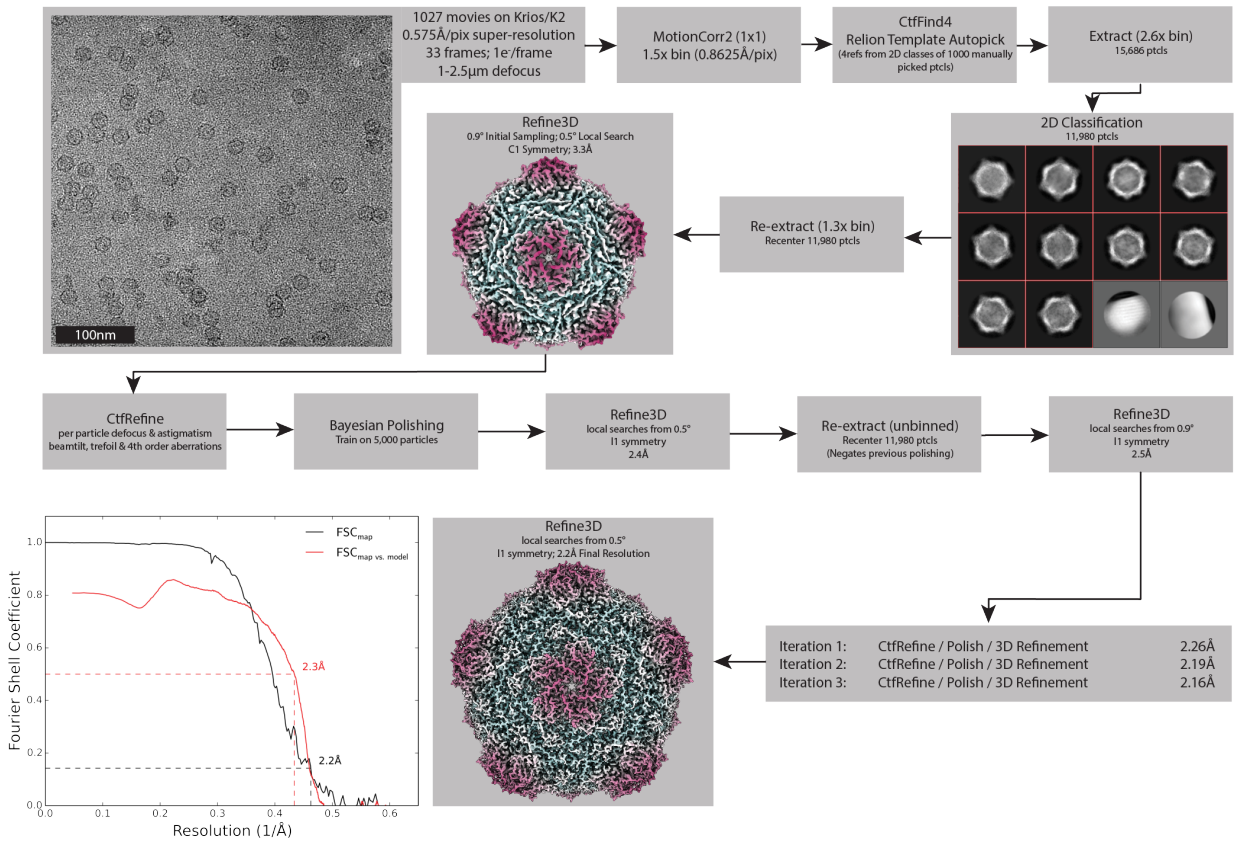
Supplementary Figure 3-1: SrpI-associated cysteine desulfurase, CyD, possesses a unique N-terminal domain that is absent in the four other cysteine desulfurase genes in the *S. elongatus* PCC 7942 genome. Multiple sequence alignment of the SrpI encapsulin-associated cysteine desulfurase (Synpcc7942_B2661) and the other genomic cysteine desulfurases (Synpcc7942_2558, Synpcc7942_1929, Synpcc7942_1255, Synpcc7942_1738). Alignment was performed using ClustalOmega. An asterisk (*) indicates a fully conserved residue, a colon (:) indicates conservation between amino acids of strongly similar properties, and a period (.) indicates conservation between amino acids of weakly similar properties.



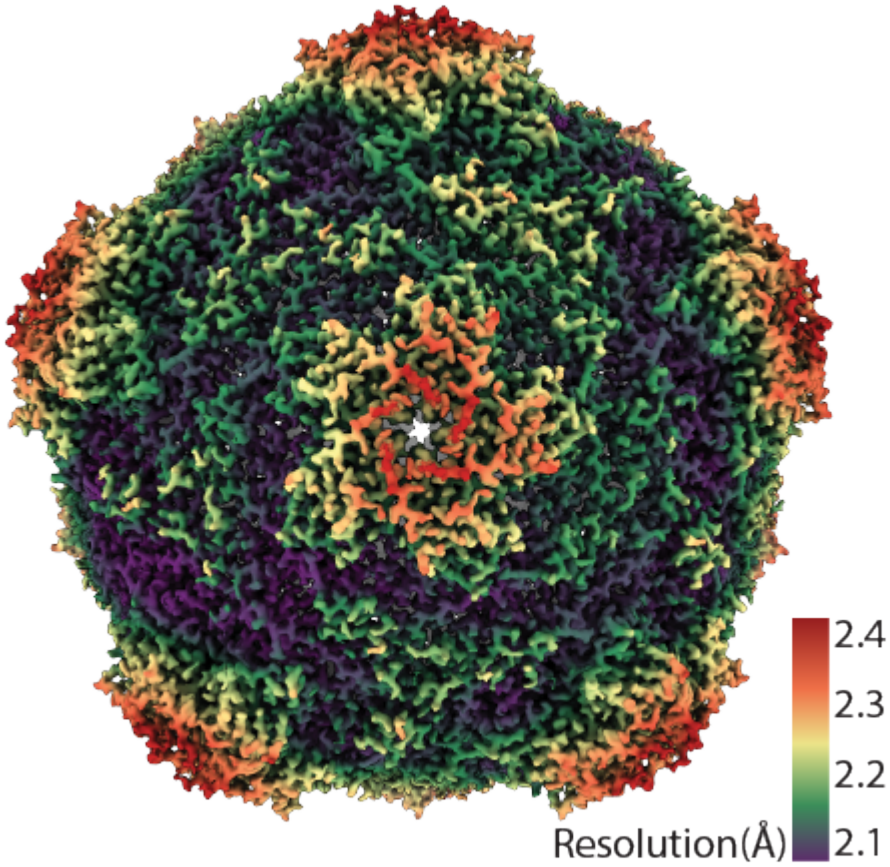
Supplementary Figure 3-2: Sequence conservation of the F2A encapsulin-associated cysteine desulfurase is sparse throughout the disordered N-terminal domain. Percent sequence identity from multiple sequence alignment of 997 F2A encapsulin-associated cysteine desulfurase sequences.



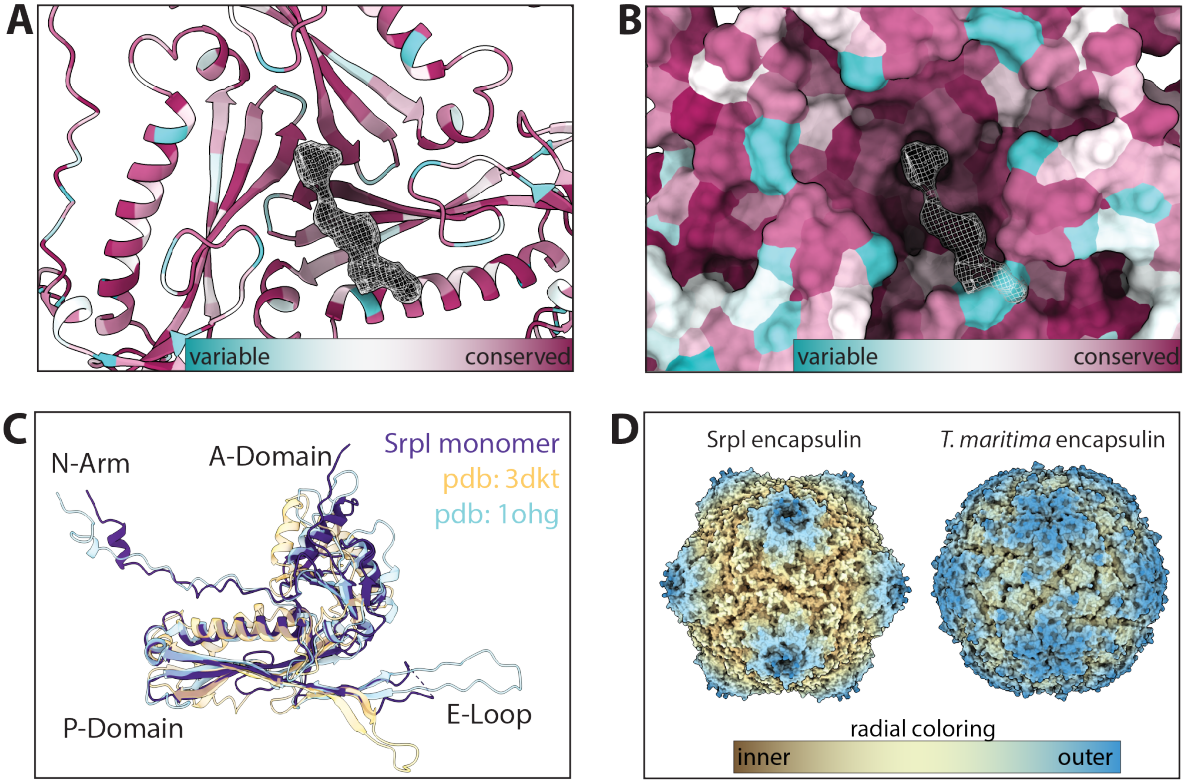
Supplementary Figure 4-1: Analysis of sfGFP and 225NTD-sfGFP loading into SrpI encapsulin. (A) and (B) size exclusion chromatogram of SrpI encapsulin shell protein (35 kDa) refolded in the presence of untagged sfGFP (27kDa) and 225NTD-sfGFP (50kDa) respectively using a Superose™ 6 Increase column (GE 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).



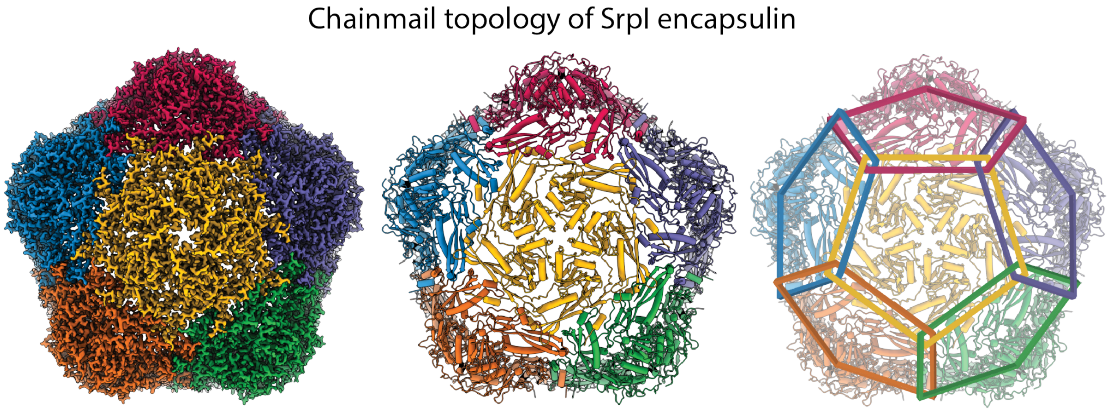
Supplementary Figure 6-1: Processing Pipeline for the SrpI encapsulin. Processing workflow within RELION that used to reconstruct the holo-SrpI structure. An identical approach was used for the apo-SrpI encapsulin. Once both these structures were determined, symmetry expansion and focused classification was used to compare subtle differences in the density (not shown, see Figure 6E).



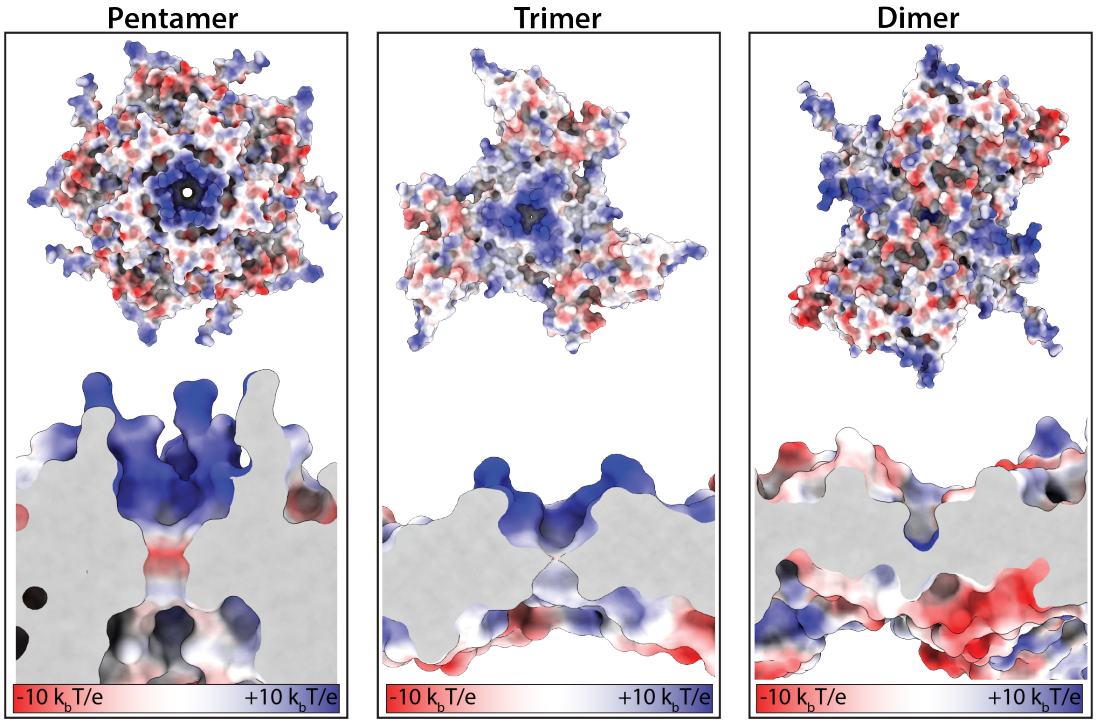
Supplementary Figure 6-2: CryoEM resolution map of SrpI encapsulin.



Supplementary Figure 6-3: 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 SrpI shell respectively (conservation was calculated via ConSurf). (C) single subunit of SrpI 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 SrpI encapsulin morphology in comparison with the *T. maritima* encapsulin.



Supplementary Figure 6-4: Chainmail-like topography of SrpI. Coloring of individual subunits highlights the chainmail overlapping topology of the SrpI encapsulin.



Supplementary Figure 6-5: 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).

Encapsulin Family	Count
Family 1	3023
Family 2	5540

Supplementary Table 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

Supplementary Table 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

Supplementary Table 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

Supplementary Table 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				

Supplementary Table 5: Data collection, reconstruction, and processing statistics.