

Distribution Agreement

In presenting this thesis as a partial fulfillment of the requirements for a degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis in whole or in part in all forms of media, now or hereafter now, including display on the World Wide Web. I understand that I may select some access restrictions as part of the online submission of this thesis. I retain all ownership rights to the copyright of the thesis. I also retain the right to use in future works (such as articles or books) all or part of this thesis.

Angel Gonzalez-Valero

April 2nd, 2019

Directed Compartment Self-Assembly and Modulation of Encapsulin Quaternary Structure

by

Angel Gonzalez-Valero

Stefan Lutz
Adviser

Emory University Department of Chemistry

Stefan Lutz
Adviser

Vincent Conticello
Committee Member

Yonggang Ke
Committee Member

2019

Directed Compartment Self-Assembly and Modulation of Encapsulin Quaternary Structure

By

Angel Gonzalez-Valero

Stefan Lutz

Adviser

An abstract of
a thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Sciences with Honors

Emory University Department of Chemistry

2019

Abstract

Directed Compartment Self-Assembly and Modulation of Encapsulin Quaternary Structure By Angel Gonzalez-Valero

Similar to membrane-based compartmentalization in eukaryotic cells, prokaryotes can establish physical boundaries within their cellular environment with the help of protein-based encapsulins. Individual encapsulin protomers can self-assemble to form nanocompartments; hollow spheres consisting of either 60 or 180 homomeric subunits. Structure analyses suggest that the E-loop, one of encapsulin's three distinct structural features, is responsible for conformational changes which give rise to the two distinct quaternary structures of these nanocompartments. The structural effects of swapped E-loops of encapsulins from *Thermotoga maritima* (TmE; native 60-mer) and *Myxococcus xanthus* (MxE; native 180-mer) are explored to determine the effect of chimeragenesis on protein quaternary structure. The study of encapsulin protein chimeras revealed conserved N-terminal glycine and C-terminal proline residues that flank the E-loop, yet formation of these chimeric constructs suffered from compartment stability issues. Encapsulin shell structure is further explored through the steric-based obstruction of compartment self-assembly via lumen-oriented fusion of maltose binding protein (MBP). MBP-TmE fusion protein provides a novel method of in vitro shell assembly, termed "directed compartment self-assembly" (DCSA). MBP-TmE fusion protein studies revealed the discovery of arrested oligomeric states that precede the formation of a fully-assembled nanocompartment. These studies also elucidated a novel mechanism for in vitro self-assembly of encapsulin nanocompartments triggered through fusion protein cleavage under physiological conditions.

Directed Compartment Self-Assembly and Modulation of Encapsulin Quaternary Structure

By

Angel Gonzalez-Valero

Stefan Lutz

Adviser

A thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Sciences with Honors

Emory University Department of Chemistry

2019

Acknowledgements

Thank you to all of the Lutz lab members that supported me and oriented me throughout my undergraduate career: Matt Jenkins, Elsie Williams, Sam Iamurri, Parisa Joud, Michelle Jung, Evy Kimbrough, and David White.

Thank you to all of the lab personnel involved in my scientific training: Matt Jenkins, Elsie Williams, Leann Quertinmont, and Stefan Lutz.

Thank you to my committee members for reviewing my works: Stefan Lutz, Vincent Conticello, and Yonggang Ke

Thank you to all of the faculty who have supported me in innumerable ways: Stefan Lutz, Michael McCormick, Antonio Brathwaite, Francesco Evangelista, and Catherine Dana.

Thank you to my closest friends for providing a source of solidarity: Jacob Crossno, Markeisha Pollard.

Thank you to my girlfriend for your unabashed support through my trying times: Elena Jordanov.

Thank you to my mentoring opportunities for showing me how to grow as a leader: Emory VMIS, Emory ResLife, STEM Pathways, and Orgo TAs.

Thank you to my PI and advisor for pushing me to exceed expectations in several facets of my life: Stefan Lutz.

Gracias a mis fuentes de inspiración. Aunque vivimos separados y entre dos mundos siempre pienso en ustedes. Lo hicimos: Ma, Pa, y Poncho.

Table of Contents

<u>1. List of Figures</u>	I
<u>2. Abbreviations</u>	II
<u>Chapter 1. Chimeragenesis to Modulate Encapsulin Quaternary Structure</u>	
<u>3. Background</u>	2-5
<u>4. Materials and Methods</u>	6-14
<u>5. Results and Discussion</u>	14-18
<u>Chapter 2. Directed Compartment Self-Assembly</u>	
<u>6. Background</u>	20-22
<u>7. Materials and Methods</u>	23-28
<u>8. Results and Discussion</u>	29-36
<u>9. Conclusions and Future Work</u>	36-38
<u>10. Primer Table</u>	39-41
<u>11. References</u>	42-43

Figures

Figure 1.	4
Figure 2.	5
Figure 3.	5
Figure 4.	6
Figure 5.	7
Figure 6.	9
Figure 7.	10
Figure 8.	14
Figure 9.	15
Figure 10.	17
Figure 11.	20
Figure 12.	21
Figure 13.	22
Figure 14.	23
Figure 15.	24
Figure 16.	29
Figure 17.	30
Figure 18.	33
Figure 19.	34
Figure 20.	35
Figure 21.	35
Figure 22.	37

Abbreviations

1. **BMC** – Bacterial microcompartment
2. **Clp** – Cargo loading peptide
3. **DCSA** – Directed compartment self-assembly
4. **E-loop** – Elongated loop
5. **FLPs** – Ferritin-like proteins
6. **GCaMP** – Fusion protein of GFP, calmodulin, and M13
7. **GFP** – Green fluorescent protein
8. **HEPES** – 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
9. **IPTG** – Isopropyl β -D-1-thiogalactopyranoside
10. **MBP** – Maltose binding protein
11. **MBP-TmE** – Fusion of MBP to N-terminus of TmE subunit
12. **MxE** – *M. xanthus* encapsulin
13. **PCR** – Polymerase chain reaction
14. **PEG** – Polyethylene glycol
15. **PIPE** – Polymerase incomplete primer extension
16. **PMC** – prokaryotic microcompartment
17. **SDS-PAGE** - Sodium dodecyl sulfate polyacrylamide gel electrophoresis
18. **SEC** – Size exclusion chromatography
19. **TEM** – Transmission electron microscopy
20. **TEV** – Tobacco etch virus
21. **TmE** – *T. maritima* encapsulin
22. **TmE-MxE loop** – TmE with MxE E-loop chimera
23. **Tris** – Tris(hydroxymethyl)aminomethane

Chapter 1. Chimeragenesis to Modulate Encapsulin Quaternary Structure

Background

A key feature for biological function in any microorganism is the ability to compartmentalize endogenous chemical processes. In eukaryotes, organelles have long dominated the common perception of cellular compartmentalization. However, prokaryotic organisms also utilize similar compartmentalization strategies to maintain normal cellular function. Prokaryotic organisms rely on proteinaceous compartments, such as bacterial microcompartments (BMCs) or prokaryotic microcompartments (PMCs), to function in the sequestration of toxins, shielding of reactions from the cytosol, or preventing clashes between competing chemical pathways. Recent studies have identified a novel class of prokaryotic compartments named encapsulins that are capable of performing similar biological functions. (Nichols *et al.* 2017)

Encapsulins are small icosahedral proteins with hollow interiors fit for encapsulating cargo. Due to their small size, encapsulins are often referred to as nanocompartments with an exterior diameter from 25 – 35 nm. Some of the first crystal structures of encapsulins arose from the archaea *Pyrococcus furiosus* (Akita *et al.* 2007) and hyperthermophile *Thermotoga maritima* (Sutter *et al.* 2008), revealing structures that resembled viral capsids packed with several ferritin-like proteins (FLPs). Of the characterized native cargos proteins in encapsulins, all have been shown to have some antioxidant activity, suggesting that encapsulins play a role in combating oxidative stress within the cell. (McHugh, *et al.* 2014, Giessen, *et al.* 2017) A pioneering study involving bacterial nutrient starvation in *Myxococcus xanthus* showed significant upregulation of the encapsulin shell protein (MxE) along with its three FLPs; meaning the native *M. xanthus* encapsulin either functions to store iron during nutrient starvation or sequester ferrous iron (Fe^{2+}) to prevent Fenton-driven oxidation, which would result in the generation of reactive oxygen species (ROS). (McHugh, *et al.* 2014).

Other physical properties of encapsulins have alluded to their native function. Once assembled, encapsulins are largely stable proteins. Orthologs from thermophilic organisms such as *T. maritima* are highly thermostable (Sutter, *et al.* 2008) and generally, most encapsulins are resistant to non-specific proteases, tolerate a wide pH range, and can tolerate extracellular environments such as culture media. (Cassidy-Amstutz, *et al.* 2016, Rahmanpour, *et al.* 2013)

The abusive treatment of encapsulins has shown that their protective qualities can be conferred to their cargo proteins, such as in studies with encapsulins from *Rhodococcus erythropolis*, where an exogenous luciferase was encapsulated and well protected from protease treatment. (Tamura *et al.* 2015) The stable nature of these protein shells is thought to protect their native cargo when in placed in extreme conditions.

Analyses of crystallographic data have resulted in advances in the understanding of shell assembly, cargo loading, and native physiological role. Encapsulin shells are composed of 60 or 180 individual subunits that self-assemble into icosahedral structures with 20 equivalent triangular faces. Nanocompartments from *T. maritima* (TmE; T=1 symmetry) assemble from 60 monomers and are 25 nm in external diameter; the encapsulins from *M. xanthus* (MxE; T=3 symmetry) assemble from 180 monomers and are 35 nm in external diameter (Fig. 1). Symmetry of these encapsulins is dictated by the triangulation number, an integer value that describes how many times an equilateral triangle is divided in order to maintain the shape of an icosahedron. Variations in the triangulation number result in proportional changes in size of the icosahedron so long as the molecular mass of each individual subunit is approximately the same. (Prasad *et al.* 2012) Increasing the triangulation number also changes the ratio of pentameric and hexameric faces present on the exterior of the compartment. Encapsulins are organized in either T=1 (no triangulation) or T=3 (3 divisions) symmetries.

The 60-mer T=1 constructs form 12 pentameric vertices, whereas the larger 180-mer T=3 constructs are formed from 12 pentameric vertices and 20 hexameric faces. Each nanocompartment, regardless of triangulation number, contains several pores (5-6 Å large) at three major folds of symmetry. (Akita *et al.* 2007, Sutter *et al.* 2008, McHugh *et al.* 2014) These pores are lined with negatively-charged residues and are thought to sequester metal cations. The chemical

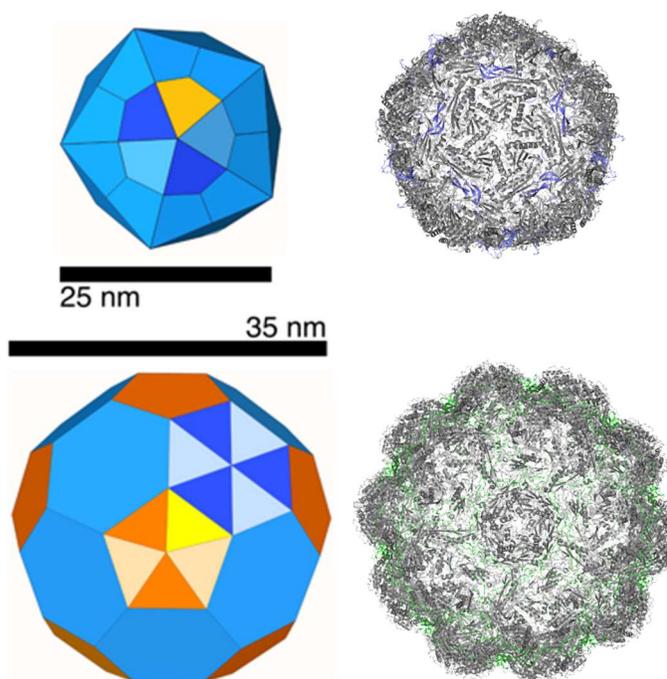


Figure 1. Top; schematic representation of T=1 encapsulin with crystal structure of TmE. (PDB 3dkt; Sutter *et al.* 2008) Bottom; schematic representation of T=3 encapsulin with crystal structure of MxE. (PDB 4pt2; McHugh *et al.* 2014)

composition of these openings regulate the flux of small molecules and ions from cytosol to the interior of the encapsulin while also serving as a permeability barrier for larger molecules, shielding the encapsulin lumen from the cytosolic environment.

The individual subunits that make up the encapsulin contain three distinct domains: The axial domain (A-domain), peripheral domain (P-domain) and elongated loop (E-loop). (Sutter *et al.* 2008) At the monomeric level, these proteins are structurally conserved and are roughly the same size (31-kDa). However, subtle differences in the proteins' sequences result in either T=1 or T=3 constructs. The A-domains and P-domains of the TmE and MxE encapsulins align well with each other (RMSD = 4.8 Å), with the largest variation coming from the E-loop (RMSD = 7.3 Å) (Fig. 2). Differences in size, rigidity, and amino acid composition of the E-loop have been speculated to influence the assemblies into either T=1 or T=3 (Fig. 3). Notably, in the encapsulin

Materials and Methods

Rational Design, Multiple Sequence Analysis

TmE and MxE encapsulins were first analyzed using PDB-accessible crystal structures in PyMOL to generate rational chimeric constructs of TmE-MxE loop (TmE PDB 3dk1; MxE PDB 4pt2). For the first chimera (TmE-MxE loop 1), the optimal stitching points were determined from a crystal structure overlay of the two encapsulin monomers. Proximal residues that displayed no apparent secondary structure interactions with neighboring domains were selected as the stitching points between the two encapsulin E-loops. Second and third-generation TmE-MxE loop chimeras made use of the multiple sequence alignment performed across 200 encapsulin orthologs to determine sequence conservation about the E-loop (Fig. 3).

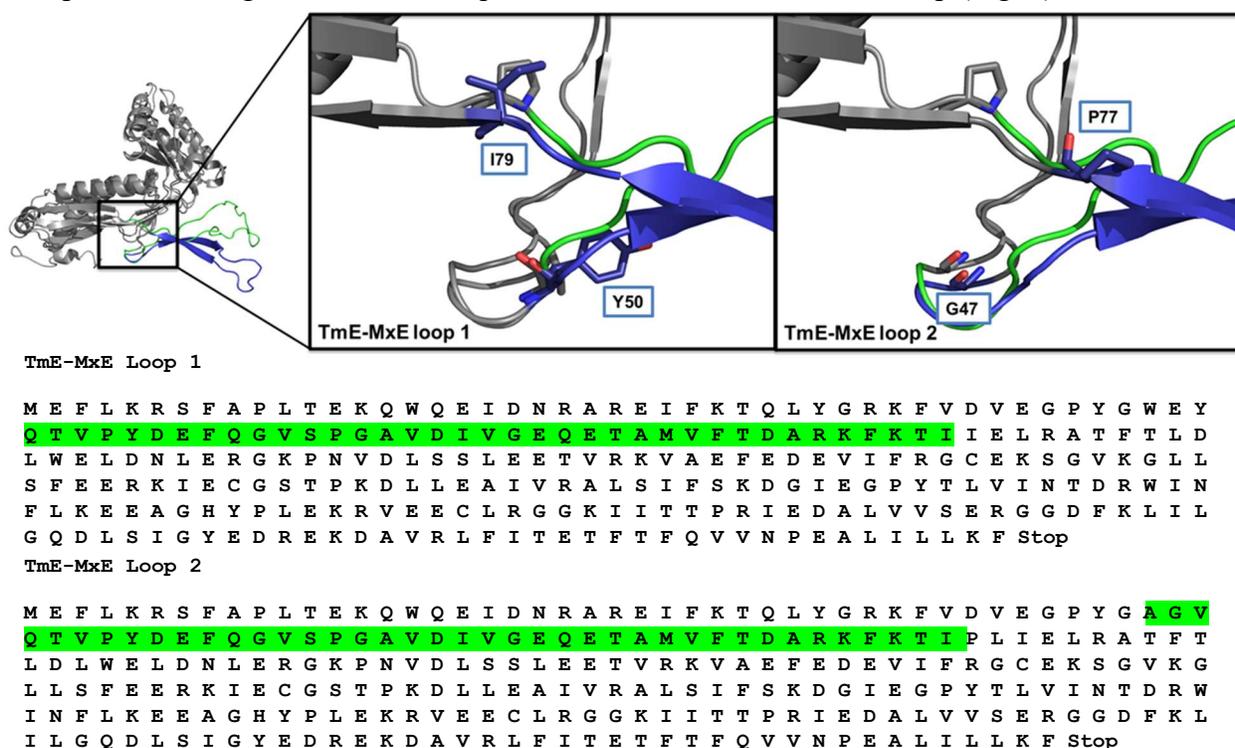
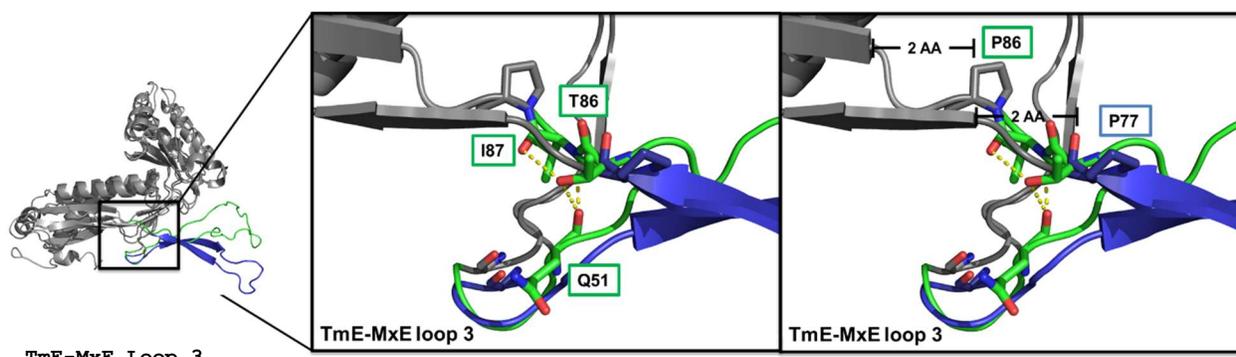


Figure 4. Crystal structure of TmE (blue, PDB 3dk1) and MxE (green, PDB 4pt2) around the E-loop. TmE-MxE loop chimera sequence is shown below with MxE loop sequence shown in green. Flanking residues boxed in blue represent the TmE stitching sites for either sequence.

The multiple sequence alignment was performed using the Lasergene software MegAlign under the ClustalW method. Both TmE and MxE protein sequences were downloaded from the NCBI

RefSeq protein records (TmE, WP_004080898.1; MxE, WP_026114001.1) and individually protein BLASTed to determine 100 non-redundant protein sequences with > 40% sequence identity for each encapsulin. These sequences were then imported into MegAlign, subjected to a ClustalW alignment, and subsequently imported to MegAlign pro for alignment analysis. TmE-MxE loop 2 and TmE-MxE loop 3 constructs were generated using a conserved N-terminal glycine and C-terminal proline around the E-loop as stitching points while still referencing the native encapsulin crystal structures to ensure minimal secondary structure disruption to the neighboring domains.

TmE-MxE loop 3 was constructed upon observing a cross-chain hydrogen bonding interaction occurring in close proximity to the N-terminal G45 and C-terminal P86 in the MxE subunit. Q49 forms a hydrogen bonds across the loop with T84 and I85 residues; likely aiding the stabilization of the much more flexible MxE E-loop by forming a tight knot at the base of the loop (Fig. 5).



TmE-MxE Loop 3

```

M E F L K R S F A P L T E K Q W Q E I D N R A R E I F K T Q L Y G R K F V D V E G P Y G A A G
A V Q T V P Y D E F Q G V S P G A V D I V G E Q E T A M V F T D A R K F K T I P L I E L R A T
F T L D L W E L D N L E R G K P N V D L S S L E E T V R K V A E F E D E V I F R G C E K S G V
K G L L S F E E R K I E C G S T P K D L L E A I V R A L S I F S K D G I E G P Y T L V I N T D
R W I N F L K E E A G H Y P L E K R V E E C L R G G K I I T T P R I E D A L V V S E R G G D F
K L I L G Q D L S I G Y E D R E K D A V R L F I T E T F T F Q V V N P E A L I L L K F Stop

```

Figure 5. Zoomed crystal structure of TmE (blue, PDB 3dkt) and MxE (green, PDB 4pt2) around the conserved glycine and proline of the E-loop. Hydrogen bonding interactions near stitching sites of both encapsulin E-loops are shown with dotted yellow lines. Beta-sheet of the neighboring P-domain is shown to the right, located 2 amino acids downstream from the conserved proline. Added alanine residues are shown in red below.

This interaction is not observed in the TmE E-loop, which is instead stabilized by an antiparallel beta-sheet along the length of the loop. At the C-terminus of the E-loop, both encapsulin subunits

display a short, 2 amino acid long chain that leads into an antiparallel beta-sheet at the N-terminus of the P-domain. Once superimposed with the TmE monomer, it is clear that swapping the E-loop of MxE into the TmE monomer at the conserved G – P stitching sites will result in a kinked MxE loop structure as it will reach the beta-sheet of the neighboring P-domain 2 amino acid residues too early. This in turn will likely disrupt the cross-chain knot observed in the native structure. To circumvent this, TmE-MxE loop 3 was constructed by adding two additional alanine residues to the N-terminus of the loop following the conserved glycine. Alanine additions were spaced apart to avoid formation of any alpha-helices from an uninterrupted stream of linked ala-residues.

Molecular Cloning

E-loop swapping between TmE and MxE was performed via recombination of the corresponding gene fragments followed by product cloning into the pD434-SR (ATUM; Newark, CA) vector. As part of my preliminary studies, the native TmE encapsulin gene was purchased (ATUM; Newark, CA) as a codon-optimized gene for protein expression in *Escherichia coli* BL21 (DE3) and cloned into the pD434-SR plasmid. The native MxE encapsulin gene was isolated from host genomic DNA using synthetic gene-specific DNA primers with restriction sites attached (IDT; Coralville, IA) and PCR amplification. 50 ng of *M. xanthus* genomic DNA was mixed with 1x iProof MasterMix (NEB; Ipswich, MA) and 0.2 μ M of MxE Gene FOR (Table P.1) and MxE Gene REV (Table P.2). The gene product was then double-digested with 1 U/ μ L of restriction enzymes NdeI and HindIII HF (NEB; Ipswich, MA). 3 μ g of pET-23b plasmid was digested in tandem with 1 U/ μ L of NdeI and HindIII HF in 1x Cutsmart buffer (NEB; Ipswich, MA). The digested products were subsequently purified using PCR and Gel extraction kits, respectively (Qiagen; Germantown, MD), mixed together, and ligated with varying ratios of insert:vector.

Ligation products were then transformed into chemically competent *E. coli* DH5 α cells and screened against 0.1 mg/mL ampicillin selection plates. Cell colonies were subsequently screened via Colony PCR using gene-specific (Table P.1 and P.2) and plasmid-specific (Table P.3 and P.4) T7 primers. Hits were sent for sequence verification (GenScript; Piscataway, NJ) for subsequent transformation and heterologous protein expression.

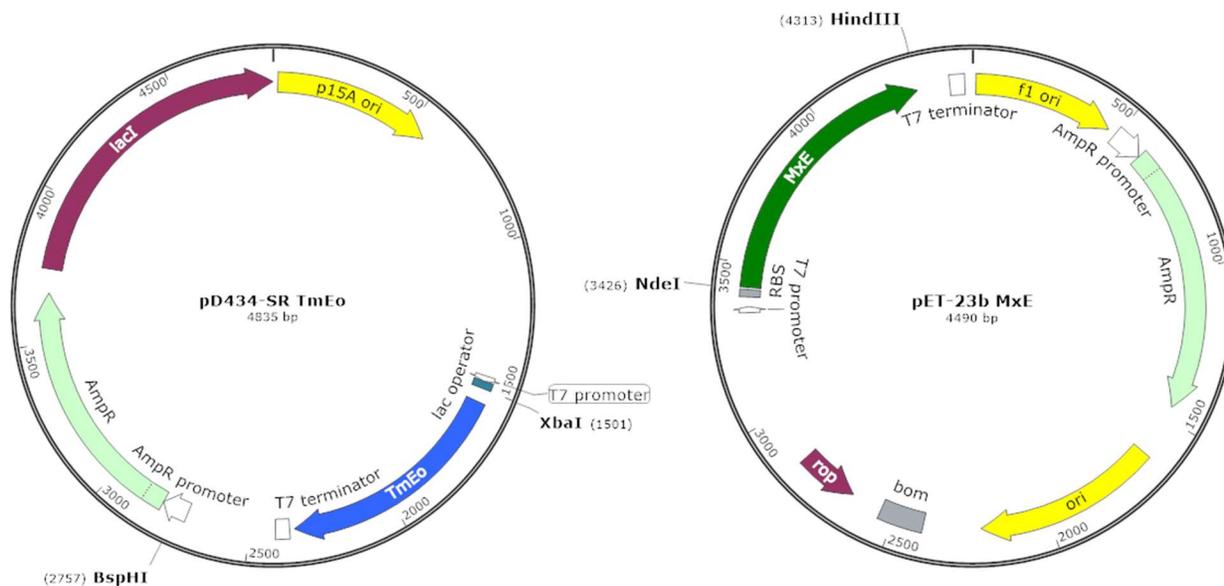


Figure 6. Vector map sequences of pET-23b MxE and pD434-SR TmE. Both plasmids contain the gene for wild-type encapsulins of *T. maritima* and *M. xanthus* with the TmE gene being codon optimized for expression in *E. coli* BL21 (DE3).

TmE-MxE loop chimeras were generated via PCR amplification of the corresponding regions followed by sequential overlap PCR amplification of the fragments. (Heckman, *et al.* 2007)

TmE-MxE loop 1 and TmE-MxE loop 2 were cloned using the following methods: Synthetic DNA primers were designed such that fragment products would contain 15-20 bps of overlap to ensure efficient overlap PCR amplification. TmE upstream and TmE downstream products were produced using 50 ng of pD434-SR TmE along with 0.2 μ M of forward and reverse primer (Upstream; Table P.5 and P.8, Downstream; Table P.7 and P.6). The MxE loop was amplified from 50 ng of pET-23b MxE using 0.2 μ M of forward and reverse primers P.11 and P.12. The first overlap PCR was performed by mixing 50 ng of TmE upstream and 50 ng of MxE loop and

amplifying using the same flanking primers (P.5 and P.12) at 0.2 μ M. The final overlap used 50 ng of this product mixed with 50 ng of the TmE downstream fragment. This was amplified with the flanking pD434-SR primers (P.5 and P.6) containing XbaI and BspHI restriction sites.

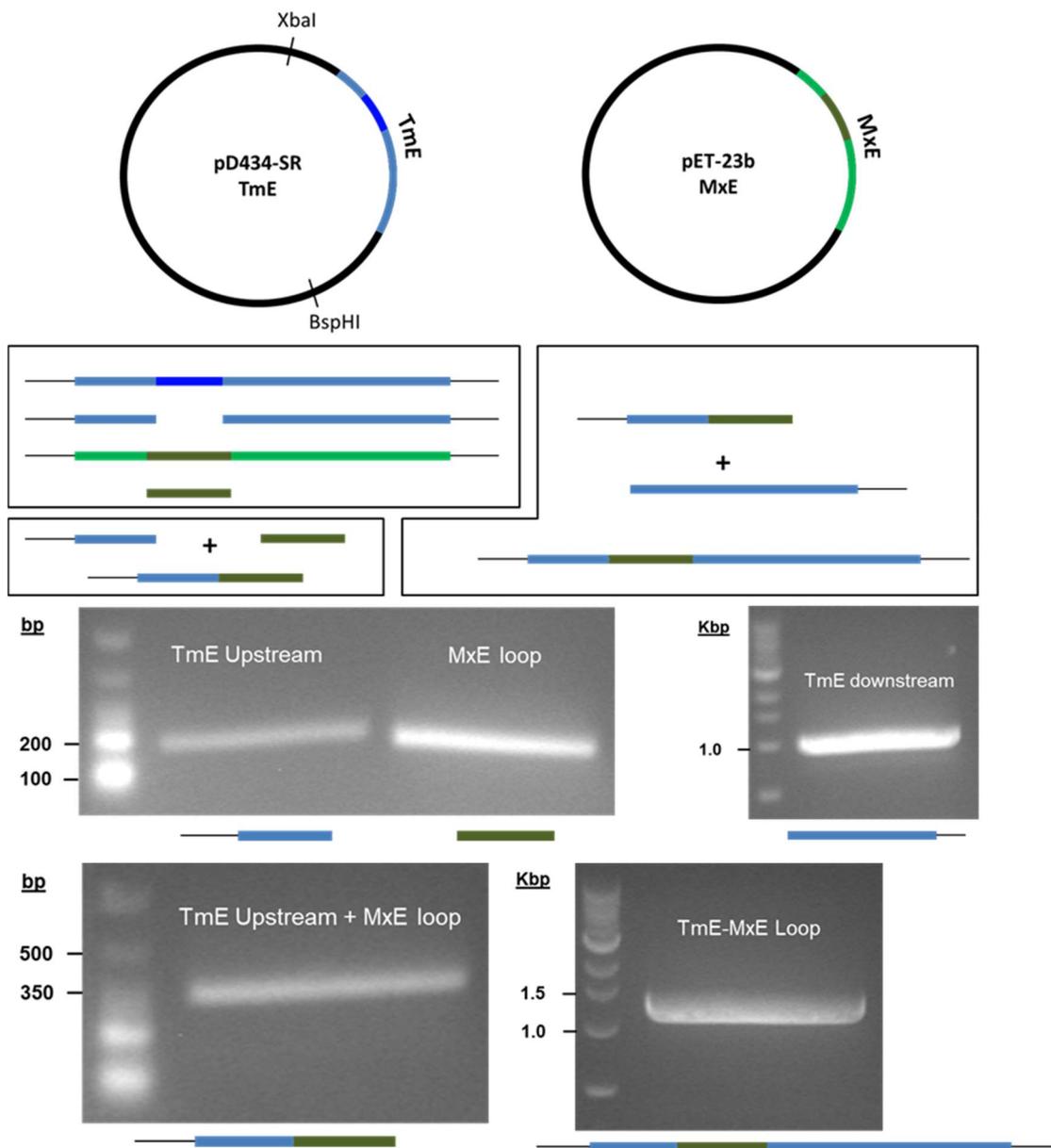


Figure 7. Top; cartoon representation of chimera fragment generation from pD434-SR TmE (blue) and pET-23b MxE (green). E-loop regions are shown as darker shades of the corresponding encapsulin. Bottom; agarose gel PCR product analysis of generated fragments and overlap products.

For products ≥ 1 kbp, a 1x solution of iProof MasterMix was used to amplify; other gene products were mixed with 1mM dNTP mix (NEB; Ipswich, MA), 0.1 U/ μ L Taq polymerase, and

a 1x solution of Thermopol Buffer (NEB; Ipswich, MA). The completed gene product and 3.0 μ g of pD434-SR plasmid were then double restriction digested with 1 U/ μ L BspHI and XbaI enzymes in 1x Cutsmart buffer. Nicked DNA was mixed in varying insert:vector concentrations, ligated, and transformed into chemically competent *E. coli* DH5 α cells and screened against 0.1 mg/mL ampicillin selection plates. Cell colonies were screened using plasmid specific (P.3 and P.4) and gene-specific primers (P.3 and P.12) with hits sent for sequence-verification.

TmE-MxE loop 3 was cloned using these same principles using only two fragments and pD434-SR TmE-MxE loop 2 as template. Synthetic primers were designed to introduce the two amino acid DNA codons near the upstream portion of the MxE loop. An upstream fragment was amplified from 50 ng of pD434-SR TmE-MxE loop 2 using 0.2 μ M of forward P.5 and reverse P.16 using 1mM dNTP mix, 0.1 U/ μ L Taq polymerase in a 1x solution of Thermopol Buffer. The downstream fragment was amplified from 50 ng of pD434-SR TmE-MxE loop 2 using 0.2 μ M of forward P.15 and reverse P.6 in a 1x iProof solution. 50 ng of the upstream and downstream fragments were mixed, overlapped (P.5 and P.6), and double-digested using 1 U/ μ L BspHI and XbaI enzymes in 1x Cutsmart buffer. Digested TmE-MxE loop 3 was then mixed with digested pD434-SR, ligated, transformed, screened, and verified using the same methods outlined for TmE-MxE loop 1 and 2.

Protein Expression and Purification

Stable nanocompartments were expressed and purified using a standard methodology. The cloned encapsulin constructs were transformed into chemically-competent *E. coli* BL21 (DE3) and grown on 0.1 mg/mL ampicillin selection plates. Successful colonies were grown in a 600 mL culture volume of LB media with a final concentration of 0.1 mg/mL ampicillin, then incubated at 37°C until reaching an OD₆₀₀ = 0.6. Protein expression was induced with 0.1 mM

isopropyl β -D-1-thiogalactopyranoside (IPTG) and incubated overnight at 37°C for 18 hrs. 1 mL expression time course samples were pulled following IPTG induction in 2 hr increments.

Overexpressed cell culture was then collected and centrifuged at 4,000 rpms at 4°C. Additional cell pellets were stored in a -20°C freezer.

TmE-MxE loop cell pellets (300 mL) were resuspended in 25 mL of 50 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES)·KOH, 100 mM NaCl buffer pH 7.6. Resuspended cells were then lysed via sonication, treated with 0.01 mg/mL DNase I, and centrifuged at 4,000 rpms at 4°C. The clarified cell lysate was then treated with a final concentration of 0.1 g/mL polyethylene glycol (PEG) 8000, 0.04 g/mL NaCl and incubated on ice for 45 mins followed by another round of centrifugation at 4,000 rpms and 4°C. Precipitant containing fully-assembled nanocompartments was then resuspended in 5 mL 50 mM HEPES·KOH, 100 mM NaCl buffer pH 7.6 and incubated with 0.1 mg/mL trypsin protease at 37°C for 10 mins. These samples were then filtered with 0.22 μ m syringe filters before loading onto a size-exclusion gel filtration system.

Chimera constructs that did not form fully-assembled compartments were purified in a similar manner, with the exception of PEG-mediated precipitation and trypsin incubation, as these methods lead to detrimental effects on protein yield. Instead, the primary purification method utilized anion-exchange chromatography with size-exclusion chromatography as a secondary purification method. Overexpressed protein cell samples were resuspended in 10 mL of 50 mM tris(hydroxymethyl)aminomethane (Tris)·HCl, 20 mM NaCl buffer, pH 7.6. Following cell lysis and centrifugation, clarified lysate samples were loaded on an equilibrated HiTrap Q FF anion exchange chromatography column using an ÄKTA Fast Protein Liquid Chromatography (FPLC) system. An increasing gradient of 50 mM Tris·HCl, 1 M NaCl buffer, pH 7.6 was used to elute

the protein chimera from the anion exchange resin. Protein samples were tracked from a built-in UV lamp 280 nm absorbance readout and verified via sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) analysis.

SDS-PAGE, Size-exclusion chromatography, and TEM Analysis

Throughout the purification process protein samples were examined via SDS-PAGE analysis to ensure adequate transfer of the desired protein. Although encapsulins exist as 60-mer or 180-mer assemblies, the disassembled linearized protein will appear as a ~31-kDa band on an SDS-PAGE gel. 10 μ L protein samples were mixed with 10 μ L 2x SDS dye stock (15 % glycerol, 0.125 M Tris·HCl pH 8.8, 5 mM disodium ethylenediaminetetraacetate dehydrate (EDTA- Na_2), 2 % SDS, 0.1 % bromophenol blue, 1 % β -mercaptoethanol) and boiled for 10 mins at 95°C. The 20 μ L samples were then loaded onto SDS-PAGE gels for gel electrophoresis. Gels were stained in a Coomassie Brilliant Blue stock (0.5 mg/mL Coomassie Brilliant Blue, 10 % glacial acetic acid, 0.25 % isopropanol) and destained in 10 % glacial acetic acid for band visualization.

Filtered protein samples were loaded onto a size-exclusion gel filtration system powered by a Bio-Rad NGC Medium-Pressure Liquid Chromatography system. Protease-treated samples were purified on a HiPrep 16/60 Sephacryl S-500 HR gel filtration column while protease-free samples were purified on a HiPrep 26/60 Sephacryl S-500 HR gel filtration column to avoid protease cross-contamination. Columns were equilibrated with filtered 50 mM HEPES·KOH, 100 mM NaCl buffer pH 7.6 and protein quaternary structures of native and chimeric encapsulins were monitored via analytical chromatograms generated from $\lambda = 280$ nm readouts. Chromatograms are routinely analyzed by instrumentation software allowing for the quantification of peak areas, including modeling of product distributions for overlapping peaks.

Native and chimera encapsulin constructs were then visualized using transmission electron microscopy (TEM). Purified protein samples were applied to carbon-copper TEM grids (EMS; Hatfield, PA), stained with 1 % phosphotungstic acid (PTA) pH 6.5, and desiccated. Grids were visualized on a Hitachi Field Emission Transmission Electron Microscope (HF-7700) and areas of interest were exported for analysis using NIH software (ImageJ).

Results and Discussion

In order to sufficiently examine the structural properties of each protein chimera, the wild-type encapsulins must be characterized. Size-exclusion chromatography (SEC) and transmission electron microscopy (TEM) proved useful tools in analyzing these compartments.

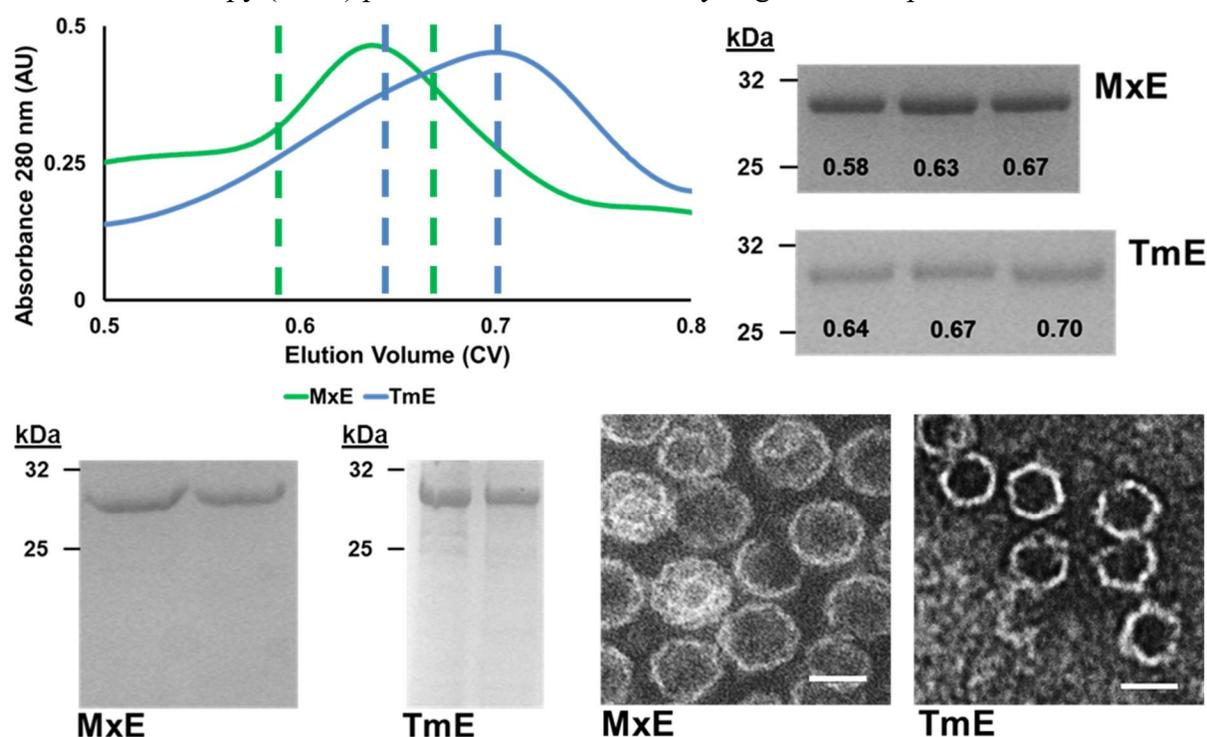


Figure 8. Top; SEC chromatogram of native MxE (green) and TmE (blue) encapsulin with corresponding peak SDS-PAGE analysis in column volumes (CV). Bottom; Trypsin sensitivity tests of native encapsulins with samples incubated at 37°C for 15 mins. TEM images of native MxE and TmE encapsulins (Scale bar = 25 nm).

Indeed, the T=1 (TmE) and T=3 (MxE) compartments have sufficient size differences that can be visualized through SEC. The smaller TmE compartments elute at higher volumes relative to the larger MxE compartments, which is consistent with the expected size:elution trend. When

subjected to trypsin protease treatment, the nanocompartments remained unaffected and structurally intact with their respective SDS-PAGE band intensities remaining unchanged (Fig. 8). Purified nanocompartments were then visualized with TEM. Compartment shape and size were consistent with literature expected values (TmE, 25 nm; MxE, 35 nm), providing an effective method to interpret encapsulin shell assembly. (Sutter *et al.* 2008, McHugh *et al.* 2014)

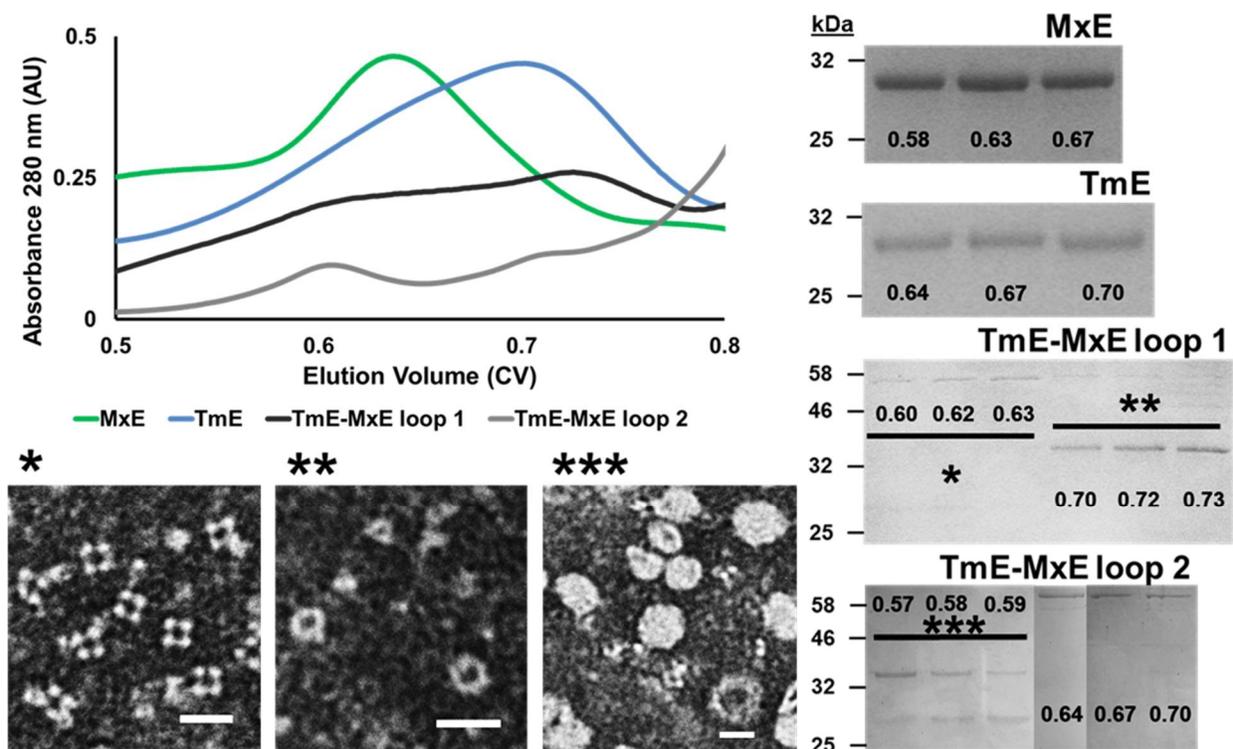


Figure 9. Top: SEC chromatogram of native MxE (green), TmE (blue), TmE-MxE loop 1 (black), and TmE-MxE loop 2 (grey) encapsulin with corresponding peak SDS-PAGE analysis in column volumes (CV). Bottom: TEM images of TmE-MxE loop 1 and TmE-MxE loop 2 chimera products (Scale bar = 25 nm).

After establishing the necessary preliminary data, chimera constructs TmE-MxE loop 1 and 2 were subjected to the same analytical treatment. According to the SEC chromatograms of TmE-MxE loop 1 and 2, two peaks exist for each chimera; one corresponding to TmE, the other to MxE elution volumes. For TmE-MxE loop 1, SDS-PAGE analysis shows that only the second peak corresponds to the encapsulin MW (~31 kDa) meaning that these structures exist as entities of similar size to T=1 capsids. By contrast, TmE-MxE loop 2 also exhibits only one peak corresponding to the MW of encapsulin, overlapping with the elution profile of MxE. These

findings suggest that TmE-MxE loop 2 exists as a larger assembly, mimicking the size of a T=3 capsid.

SEC fractions that contained encapsulin were used to blot carbon-copper TEM grids for TEM analysis of these peaks. The two SEC peaks of TmE-MxE loop 1 were examined, one containing a ~57 kDa protein, the other a ~31 kDa protein. Interestingly, the 57 kDa entity appeared as small squares roughly 15 nm in diameter, much smaller than native MxE or TmE encapsulins. However, this protein elutes roughly at the same volume as MxE encapsulins. It is uncertain if these proteins correlate to encapsulin formation, or are merely recruited throughout the purification process from the *E. coli* lysate. TEM analysis of the 31 kDa protein of TmE-MxE loop 1 showed irregularly shaped nanocompartments that are smaller than the native TmE shell. Furthermore, these compartments exhibit an insignificant presence on the grid, suggesting issues with full shell assembly. The normalized 280 nm absorbance for both TmE-MxE loop 1 and 2 is significantly lower than those for wild-type encapsulin, meaning that these proteins exist in relatively low concentrations within the cellular milieu. The 31 kDa protein of TmE-MxE loop 2 exists as an amorphous globular entity. Unlike native compartments, these entities appear to be packed and irregularly shaped, similar to protein aggregates. Without the defining hollow interior, these compartments are not particularly useful for packaging purposes.

From these data, it appears that these structures are unable to form robust capsids, and may only exist in smaller monomeric or multimeric states. These variants do not suffer from protein expression issues but instead are impacted by PEG-8000 and trypsin treatment (Fig. 10). The fully assembled compartments are capable of withstanding both treatments, suggesting that these protein chimeras are not fully assembled or are incorrectly oriented throughout the assembly process, resulting in protein aggregation.

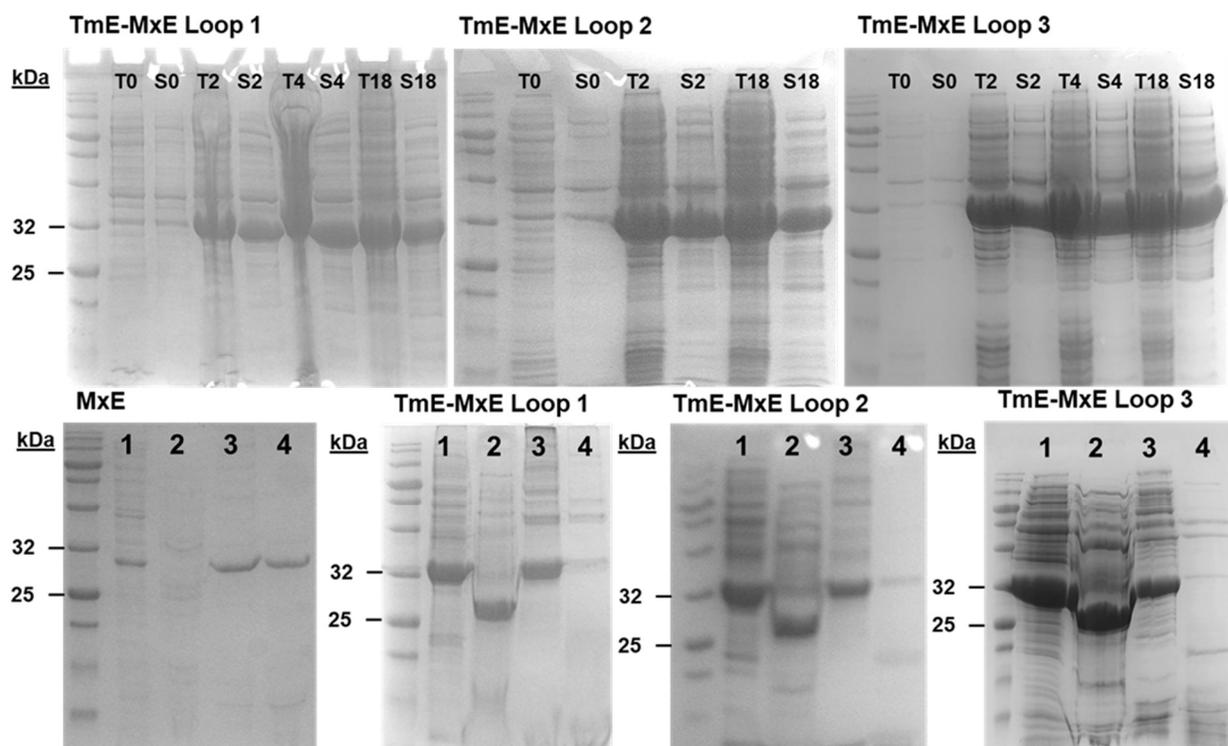


Figure 10. Top; Expression timecourse evaluation of encapsulin chimeras. T0 = Total protein at 0 hrs, S0 = Soluble protein at 0 hrs. Bottom; PEG precipitation and trypsin treatment of different encapsulins. (1) Clarified lysate, (2) PEG pelleted lysate, (3) Resuspended PEG pellet in fresh buffer, (4) Post trypsin treatment for 15 mins at 37 °C

After an 18 hr expression all protein chimeras exhibit a strong presence of soluble protein. Upon PEG precipitation, it appears that only some of the chimera encapsulin is precipitated into the pellet, with a large portion remaining in the aqueous phase. After incubation with trypsin, the vast majority of chimera encapsulin is lost. By contrast, the fully-assembled MxE encapsulin is completely precipitated by PEG treatment and is not susceptible to protease degradation. These findings suggest that all three chimeric constructs formed are insufficiently self-assembling, leading to exceedingly low concentrations of fully assembled compartments. Furthermore, these constructs are largely susceptible to protease degradation unlike the native TmE and MxE structures, suggesting that protease-susceptible residues that may be solvent-inaccessible by the fully formed shell are now exposed.

Although the constructed protein chimeras suffered from stability issues, this work reports significant findings that help resolve compartment structure. Multiple sequence analysis across encapsulin orthologs reveals conserved N-terminal glycine and C-terminal proline residues that flank the ends of the E-loop, elucidating the structural endpoints of the encapsulin E-loop. Swapping the E-loops of TmE and MxE proved to negatively impact shell assembly, leading to smaller quantities of formed entities with amorphous qualities. The more conservative swap of TmE-MxE loop 1 resulted in smaller, irregular shells while TmE-MxE loop 2 produced large aggregates. In any case, the E-loop appears to drastically influence shell assembly, and tampering with this region will dramatically affect the quaternary structure of these proteins.

Chapter 2. Directed Compartment Self-Assembly



Background

The encapsulation of non-native cargo is still well under development. Certain constraints such as encapsulin volume, cargo oligomerization, loading efficiencies, and poor understanding of the native cargo loading mechanisms have challenged the field. In vivo cargo loading is thought to occur in conjunction with shell assembly, which raises questions as to how exactly these proteins are packaged and to what efficiency. Nevertheless, certain methods of encapsulation have proven to be successful in encapsulin and encapsulin-like entities, with cargos representative in all divisions of the central dogma. (Azuma *et al.* 2018, Brach *et al.* 2017, Tamura, *et al.* 2015)

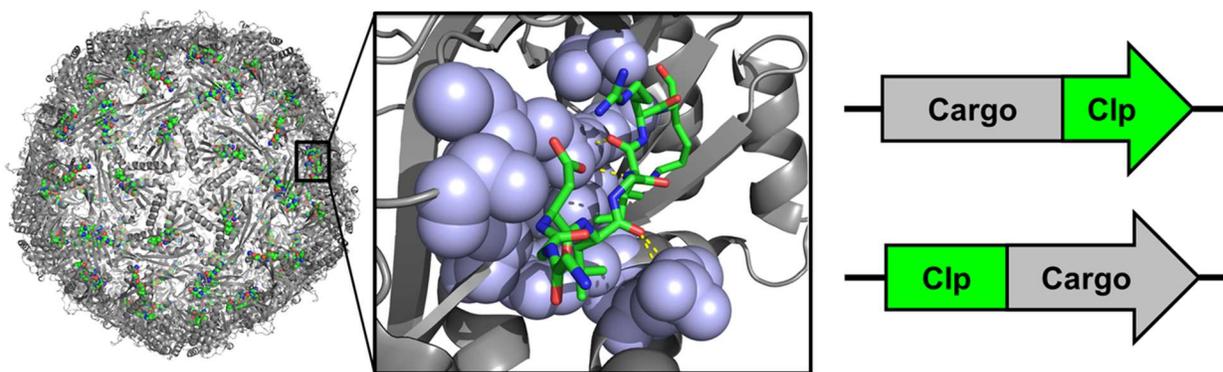


Figure 11. Left; Crystal structure of native TmE encapsulin with lumen-oriented hydrophobic binding pocket. Clp amino acid sequence shown in green. Right; Cartoon gene schematics of C-terminal and N-terminal Clp tags on non-native cargo proteins. (Sutter *et al.* 2008)

Encapsulation of non-native cargo first began by utilizing the mechanisms nature had provided in the wild-type encapsulins. The crystal structure of the encapsulin from *T. maritima* had unresolved electron densities within the capsid lumen save for a small, roughly 10 amino acid motif that lined the interior wall of the shell (Fig. 11). This motif corresponded to the C-terminal region of the encapsulin’s native cargo: a ferritin-like binding protein. (Sutter, *et al.* 2008) This small amino acid tag, called the “cargo loading peptide” (Clp), is conserved across several species within an encapsulin operon in the bacterial genome. Deletion of the Clp tag effectively terminates any encapsulation of cargo. Conversely, fusion of the Clp tag to the C-termini of non-

native cargo promotes their encapsulation. This has been most effectively demonstrated with C-terminal fusions to reporter proteins such as green fluorescent protein (GFP), luciferase, and teal fluorescent protein (TFP). (Cassidy-Amstutz, *et al.* 2016, Tamura, *et al.* 2015, Rurup, *et al.* 2014)

In vitro encapsulation strategies have focused on disassembling the purified encapsulin shell through particularly abusive methods. In the TmE system, potential mechanisms for in vitro loading include compartment disassembly at pH 13 or pH 1, shell unfolding via 7 M guanidinium chloride (GuHCl) treatment, or treatment with 12 M urea to expose the encapsulin lumen. (Cassidy-Amstutz, *et al.* 2016) These treatments are then followed by dilution into a Clp-tagged cargo-rich buffer, where the exposed subunits will reassemble and capture the non-native cargo (Fig. 12). The most efficient of these methods, GuHCl treatment, results in only $60 \pm 3\%$ encapsulin reassembly (the rest lost due to misfolding and subsequent aggregation) with only 7 ± 2 GFPs of the theoretical 60 GFPs encapsulated. (Cassidy-Amstutz, *et al.* 2016)

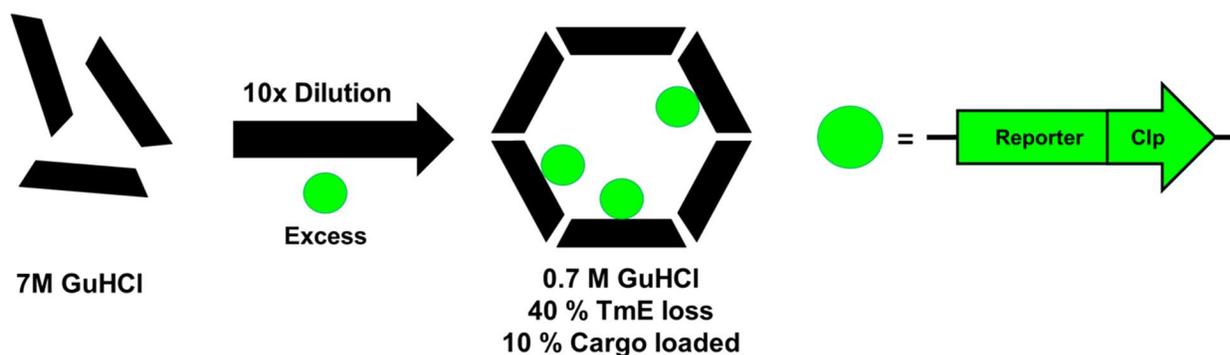


Figure 12. Cartoon representation of current in vitro cargo loading method using guanidinium chloride as a disassembly mechanism. (Cassidy-Amstutz, *et al.* 2016)

These results necessitate the development of novel in vitro loading mechanisms to decrease encapsulin shell protein loss and increase loading efficiency. The N-terminus of each encapsulin monomer is lumen-oriented while the C-termini are oriented outwards in the cytosol. Making use of encapsulin volume constraints, fusion of a large, bulkier entity to the lumen-oriented N-

terminus of the encapsulin monomer could provide a novel sterically-driven mechanism for restricted shell assembly in vivo (Fig. 13).

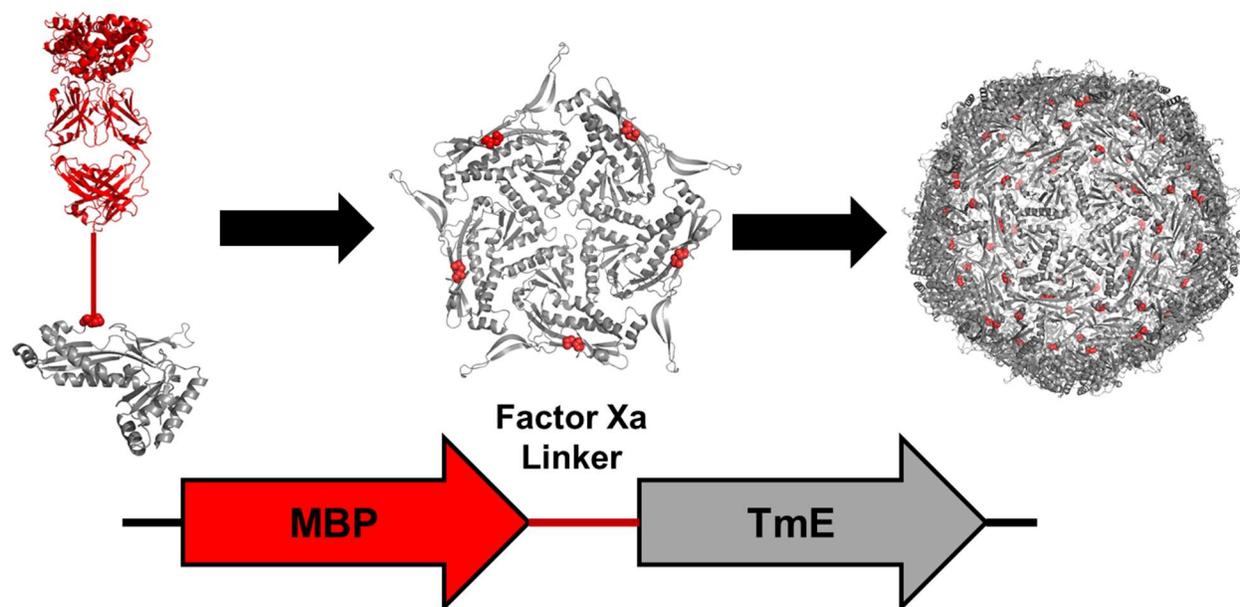


Figure 13. Mock representation of MBP-TmE fusion protein with arrested oligomeric states and full shell assembly. Bottom; Cartoon schematic of MBP-TmE with factor Xa linker portion. (Sutter, *et al.* 2008, Rizk, *et al.* 2011)

The 43-kDa maltose-binding protein (MBP) is a common protein expression tag that has long been used to increase protein solubility and facilitate downstream protein purification. (Kapust, *et al.* 1999) Proteins fused to MBP are separated by a variable linker region usually containing a specific protease cleavage site. Due to the large nature of MBP along with its protease-susceptible linker region, fusion of MBP to the encapsulin monomer N-terminus should provide a controllable mechanism for directed shell reassembly. Furthermore, a novel method for non-native cargo loading that mitigates protein aggregation would be established that may sufficiently kinetically arrest reassembly for more efficient cargo-loading.

Materials and Methods

Molecular Cloning

The MBP-TmE fusion protein was generated from the codon-optimized TmE gene cloned directly downstream from the MBP gene present in the pMAL-c2X vector. Synthetic DNA primers were designed with flanking EcoRI and HindIII restriction sites to clone out the TmE gene. 50 ng of pD434-SR TmE was mixed with 0.2 μ M of TmE Gene FOR (Table P.17) and TmE Gene REV (Table P.18) in a 1x iProof Mastermix solution. The gene product and 3 μ g of pMAL-c2X were then double-digested with 1 U/ μ L of restriction enzymes EcoRI HF and HindIII HF (NEB; Ipswich, MA) and mixed at varying ratios of insert:vector for ligation.

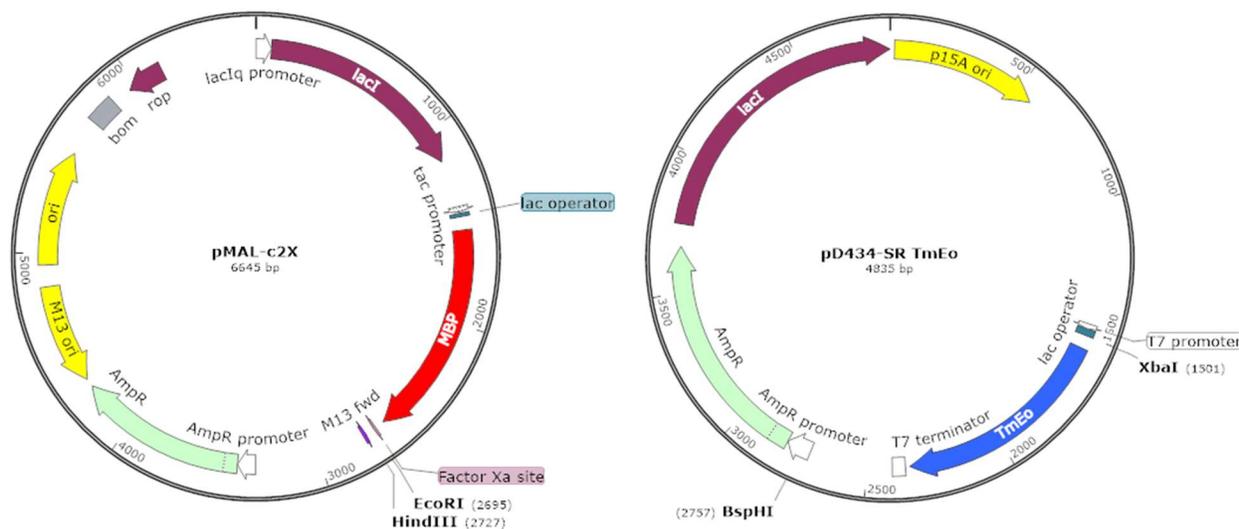


Figure 14. Vector map sequences of pMAL-c2X and pD434-SR TmE. EcoRI and HindIII restriction sites are highlighted in the pMAL-c2X map directly downstream of the MBP gene.

Ligated plasmid were then transformed into chemically competent *E. coli* DH5 α cells and plated on 0.1 mg/mL ampicillin selection plates. Colonies were screened using plasmid specific (P.19 and P.20) primers that flanked both the MBP and TmE gene. Hits were sent for sequence verification.

The factor Xa site is located in the linker region between MBP and TmE. This site was swapped with the most catalytically active TEV protease recognition site ENLYFQG (Kapust, *et al.* 2002)

via polymerase incomplete primer extension (PIPE) cloning (Klock, *et al.* 2009) for facile gene insertion that negates the need for restriction digestion or plasmid backbone ligation. PIPE cloning necessitates the generation of an insert (I-PIPE) product and vector (V-PIPE) product with incomplete, complementary 5' and 3' ends with 20-30 bps overlap. This overlap is found on either end of insert and vector and effectively seals the plasmid for stable transformation into chemically competent cells.

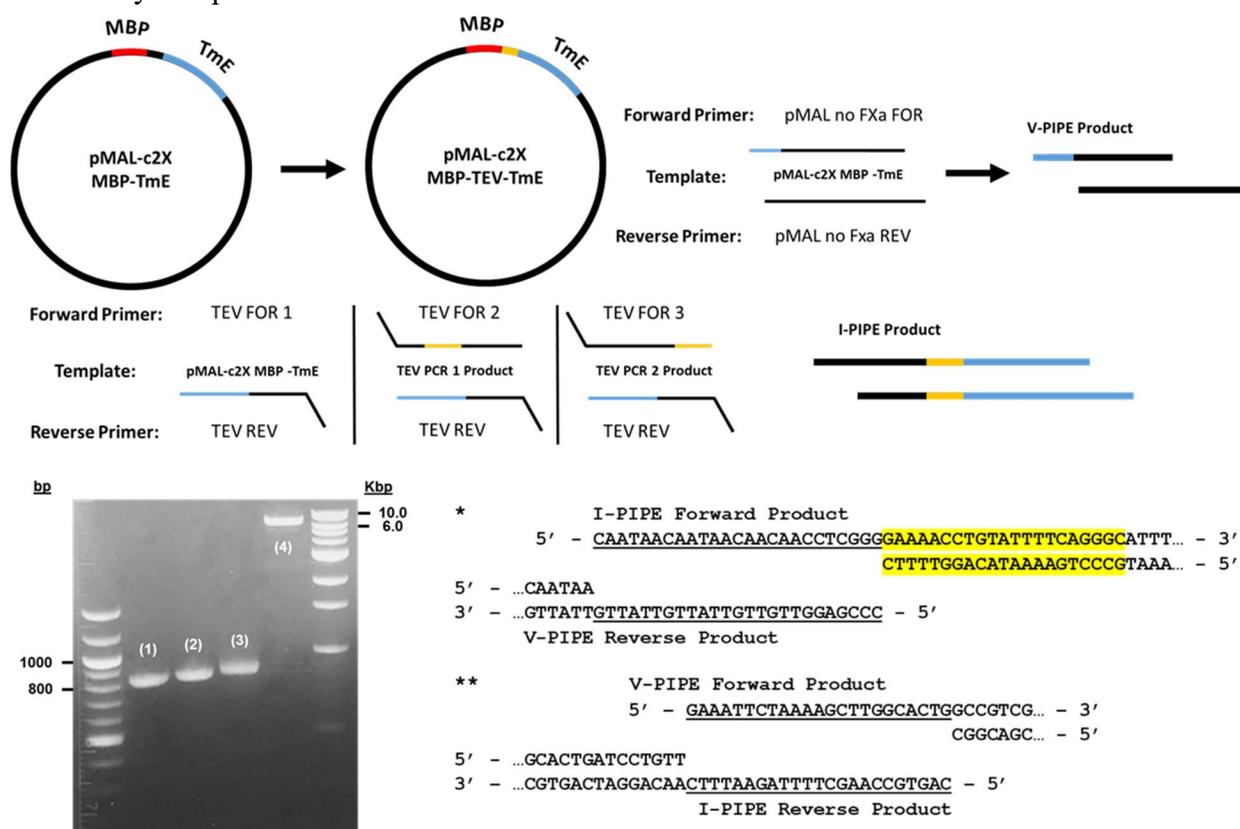


Figure 15. Top; cartoon representation of PIPE cloning using pMAL-c2X MBP-TmE as template and specially-designed synthetic DNA primers with several base pairs of overlap. Bottom; agarose gel PCR product analysis of TEV PCRs (1) and (2), completed I-PIPE product (3), and V-PIPE product (4). * Upstream overlap region of I-PIPE and V-PIPE with TEV site highlighted in yellow. ** Downstream overlap region of I-PIPE and V-PIPE.

Introducing the 7 amino acids specific to TEV protease necessitated a 3-step sequential ~15 bp addition at the 5' end of the TmE gene - due to synthetic oligonucleotide size constraints and high melting temperature (Tm) values. The first addition used 50 ng of pMAL-c2X MBP-TmE as template with 0.2 μ M TEV FOR 1 and TEV REV (P.21 and P.24) in a 1x iProof MasterMix

solution. The second addition 50 ng product (1), primers TEV FOR 2 and TEV REV (P.22 and P.24), and 1x iProof Mastermix under the same reaction conditions. The final addition that generates the I-PIPE product used 50 ng product (2), TEV FOR 3 and TEV REV (P. 23 and P. 24) with 1x iProof Mastermix while removing the final extension step in the PCR reaction cycle to ensure a healthy population of incomplete gene products. V-PIPE product was generated using 50 ng of pMAL-c2X MBP-TmE, 0.2 μ M pMAL no FXa FOR and REV (P.25 and P.26) while also removing the final extension step. V-PIPE product was subjected to DpnI digestion to remove any methylated DNA template. I-PIPE and V-PIPE products were mixed at varying ratios of insert:vector and directly transformed into chemically competent *E. coli* DH5 α cells. Cells were grown on 0.1 mg/mL ampicillin selection plates and plasmids were harvested from grown colonies then sent for sequence verification.

Protein Expression and Purification

MBP-TmE fusion proteins were all expressed and purified under the same conditions. Cloned fusion protein constructs were transformed into chemically-competent *E. coli* BL21 (DE3) and grown on 0.1 mg/mL ampicillin selection plates. Successful colonies were grown in a 300 mL culture volume of LB media with 10 mM glucose and a final concentration of 0.1 mg/mL ampicillin, then incubated at 37°C until reaching an OD₆₀₀ = 0.4. Protein overexpression was induced with 0.1 mM IPTG followed by incubation overnight at 37°C for 18 hrs. 1 mL expression time course samples were collected following IPTG induction in 2 hr increments. Overexpressed cell culture was then collected and centrifuged at 4,000 rpms at 4°C. Additional cell pellets were stored in a -20°C freezer.

The Lutz lab has previously established a robust reporter protein for in vivo co-expression and encapsulation. This construct, pMATT-2 GCaMP, contains the fluorescent reporter protein

GCaMP with a C-terminally fused cargo-loading peptide (Clp) that has been shown to promote encapsulation into TmE encapsulin and encapsulin variants (Williams, *et al.* 2018). The GCaMP-Clp reporter was transformed into chemically-competent *E. coli* BL21 (DE3) and grown on 0.05 mg/mL streptomycin selection plates. Successful colonies were grown in a 300 mL culture volume of LB media with a final concentration of 0.05 mg/mL streptomycin, then incubated at 37°C until reaching an $OD_{600} = 0.5$. Protein overexpression was induced with 0.3 mM IPTG and incubated overnight at 25°C for 20 hrs. Overexpressed cell culture was then collected and centrifuged at 4,000 rpms at 4°C. Additional cell pellets were stored in a -20°C freezer.

GCaMP-Clp cell pellets (100 mL) were resuspended in 25 mL of 50 mM Tris·HCl, 20 mM NaCl buffer pH 7.6. Resuspended cells were then lysed via sonication, treated with 0.01 mg/mL DNase I, and centrifuged at 4,000 rpms at 4°C. Following cell lysis and centrifugation, clarified lysate samples were loaded on an equilibrated HiTrap Q FF anion exchange chromatography column using an ÄKTA FPLC system. An increasing gradient of 50 mM Tris·HCl, 1 M NaCl buffer, pH 7.6 was used to elute GcAMP-Clp from the anion exchange resin. Protein samples were tracked from a built-in UV lamp 280 nm and 495 nm absorbance readout and verified via sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) analysis.

MBP-TmE cell pellets (100 mL) were resuspended in 25 mL of 50 mM HEPES·KOH, 300 mM NaCl buffer pH 7.6. Resuspended cells were then lysed via sonication, treated with 0.01 mg/mL DNase I, and centrifuged at 4,000 rpms at 4°C. The clarified cell lysate was then run through an equilibrated amylose column containing amylose resin (NEB), an affinity matrix for proteins fused to MBP. The resin was allowed to incubate with protein while rocking at 4°C, and flow through was discarded. Once fully bound to the resin, MBP-TmE was eluted in 1 mL fractions using 50 mM HEPES·KOH, 300 mM NaCl, 10 mM maltose buffer pH 7.6 and pooled. Half of

the samples was then filtered with 0.22 μm syringe filters before loading onto a size-exclusion gel filtration system for analytical purposes. The other half was used to perform directed compartment self-assembly assays

Directed Compartment Self-Assembly, Fluorometer Assays

Purified MBP-TmE fusion protein was incubated with 0.1 mg/mL trypsin protease for 15 mins at 37°C. After incubation, the protease was deactivated by spiking in 10 mM CaCl_2 . Freed TmE subunits were allowed to reassemble at 23°C for 1 hr before loading onto a size-exclusion column. MBP-TEV-TmE fusion protein was incubated with 0.15 U/mL TEV protease overnight (20 hrs) at 4 °C. Preliminary free GCaMP-Clp studies were conducted using a Fluoromax-3 spectrofluorometer (Horiba; Edison, NJ) to test protease sensitivity. Purified GCaMP-Clp was dialyzed overnight against 50 mM Tris·HCl, 20 mM NaCl, 0.05 mM ethylene glycol-bis(β -aminoethyl ether)-N,N,N',N'-tetraacetic acid (EGTA) buffer, pH 7.6 to remove residual Ca^{2+} ions in the calmodulin domain of GCaMP, thereby decreasing its fluorescence. A control experiment with 0.25 μM GCaMP-Clp measured the $\lambda = 515$ nm emission spectrum upon spiking in 10 mM CaCl_2 . Trypsin and TEV degradation tests of GCaMP-Clp using 0.1 mg/mL and 0.15 U/mL of protease, respectively, were recorded after incubation at 37°C for 15 mins. The emission spectra at $\lambda = 515$ nm were collected upon spiking in 10 mM CaCl_2 (Chen, *et al.* 2013). All experiments were run in triplicate.

Size-Exclusion and TEM Analysis

Filtered fusion protein samples were loaded onto a size-exclusion gel filtration system powered by a Bio-Rad NGC Medium-Pressure Liquid Chromatography system. Protease-treated and untreated samples were purified on separate HiPrep 16/60 Sephacryl S-500 HR gel filtration column to avoid protease cross-contamination. Columns were equilibrated with filtered 50 mM

HEPES·KOH, 300 mM NaCl buffer pH 7.6 and protein quaternary structures were monitored via analytical chromatograms generated from $\lambda = 280$ nm readouts.

Protein standards were determined using a high-resolution Superdex 200 (10/300 GL) column attached to an ÄKTA FPLC system. The column was equilibrated with filtered 50 mM HEPES·KOH, 300 mM NaCl buffer pH 7.6 and filtered, untreated MBP-TmE was loaded onto the column and run at a slow 0.5 mL/min flow rate. Upon complete bed volume elution, a molecular weight standard kit MW-GF-200 for Gel Filtration Chromatography (Sigma-Aldrich; St. Louis, MO) was loaded using the same buffer conditions and flow rate. SEC chromatograms were generated from $\lambda = 280$ nm readouts. Void volume (V_o) was determined from the Blue Dextran standard (2,000 kDa) at its $\lambda_{\max} = 280$ nm value. Other protein standards (Albumin, Alcohol dehydrogenase, β -Amylase, Carbonic anhydrase, and Cytochrome C) elution volumes were recorded to compute their respective partition coefficients (K_{AV}). From these data, a plot of K_{AV} vs. LogMW was constructed to determine a calibration curve for the size-exclusion column using a linear regression. (Tayyab *et al.* 1991) This formula was then used to compute predicted molecular weights for each SE peak, and divided by the theoretical MW of MBP-TmE to determine a predicted oligomeric state.

Untreated and treated MBP-TmE samples were then visualized using transmission electron microscopy (TEM). Purified protein samples were applied to carbon-copper TEM grids, stained with 1 % phosphotungstic acid (PTA) pH 6.5, and desiccated. Grids were visualized on a Hitachi Field Emission Transmission Electron Microscope (HF-3300) and areas of interest were exported for analysis using NIH software (ImageJ).

Results and Discussion

Upon successfully cloning pMAL-c2X MBP TmE, the fusion protein was determined to express exceedingly well under normal encapsulin expression conditions. It was unsure if purification through the amylose resin would yield any fusion protein as it was possible that MBP-TmE would form fully-assembled or partially-assembled encapsulins. In this case, the lumen-oriented MBP in MBP-TmE may not be exposed to the resin, with only free MBP (~43 kDa) binding to the amylose resin and the rest of the fusion protein eluting through the flow through. This was however not the case, as the fusion protein (~74 kDa) was retained on the resin throughout the

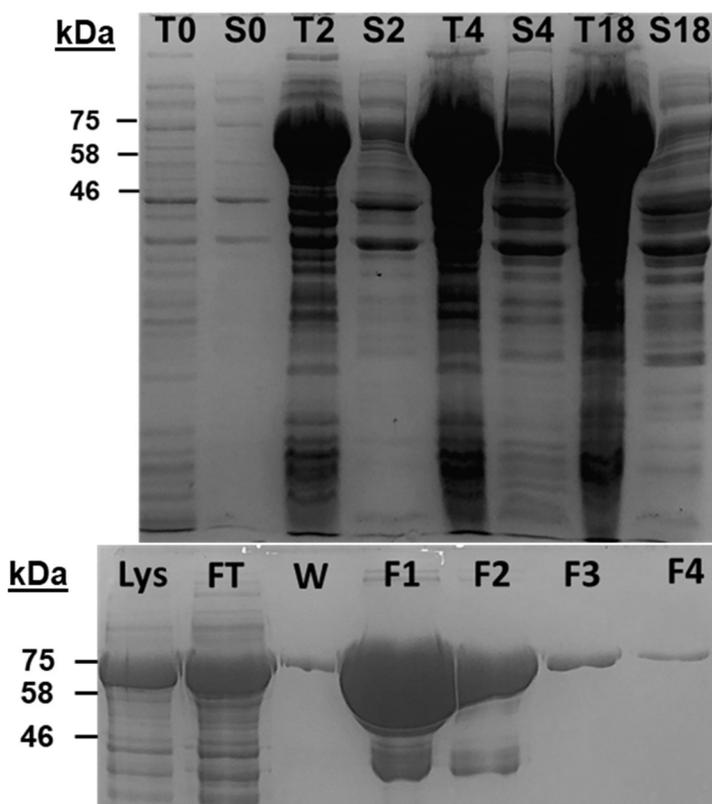


Figure 16. Top; Expression timecourse of MBP-TmE. T0 = Total protein at 0 hrs, S0 = Soluble protein at 0 hrs. Bottom; amylose column purification of MBP-TmE. Lys; cell lysate, FT; flow through, W; buffer wash, F1-4; fractions 1-4 of maltose

purification process (Fig. 16),

supporting the hypothesis that this fusion would result in the formation of unassembled encapsulin subunits *in vivo*.

However, a portion of fusion protein was observed to elute through the flow through of the amylose column.

With the high quantity of fusion protein overexpressed and purified, this can be attributed to amylose resin saturation rather than shell formation.

Indeed, the majority of fusion protein

is eluted within the first two fractions of maltose-containing buffer, giving rise to a quick and facile method of fusion protein purification. The only other band present following this

purification method is self-cleaved MBP that is present within the clarified lysate. This is not unexpected as free MBP readily binds amylose and is the mechanism behind this one-step purification.

The strong affinity of MBP-TmE for the amylose resin suggested that the MBP portion of this fusion construct remained unobstructed allowing for amylose binding. This infers that the shell is disassembled, as a fully assembled nanocompartment would contain lumen-oriented MBP. This then raises the question of how exactly this protein was able to bind. If indeed the MBP prevented shell assembly, then in what oligomeric state do these subunits exist? To test this, the eluted fractions from the amylose resin were examined via analytical SEC and compared against protein standards run under the same size-exclusion conditions.

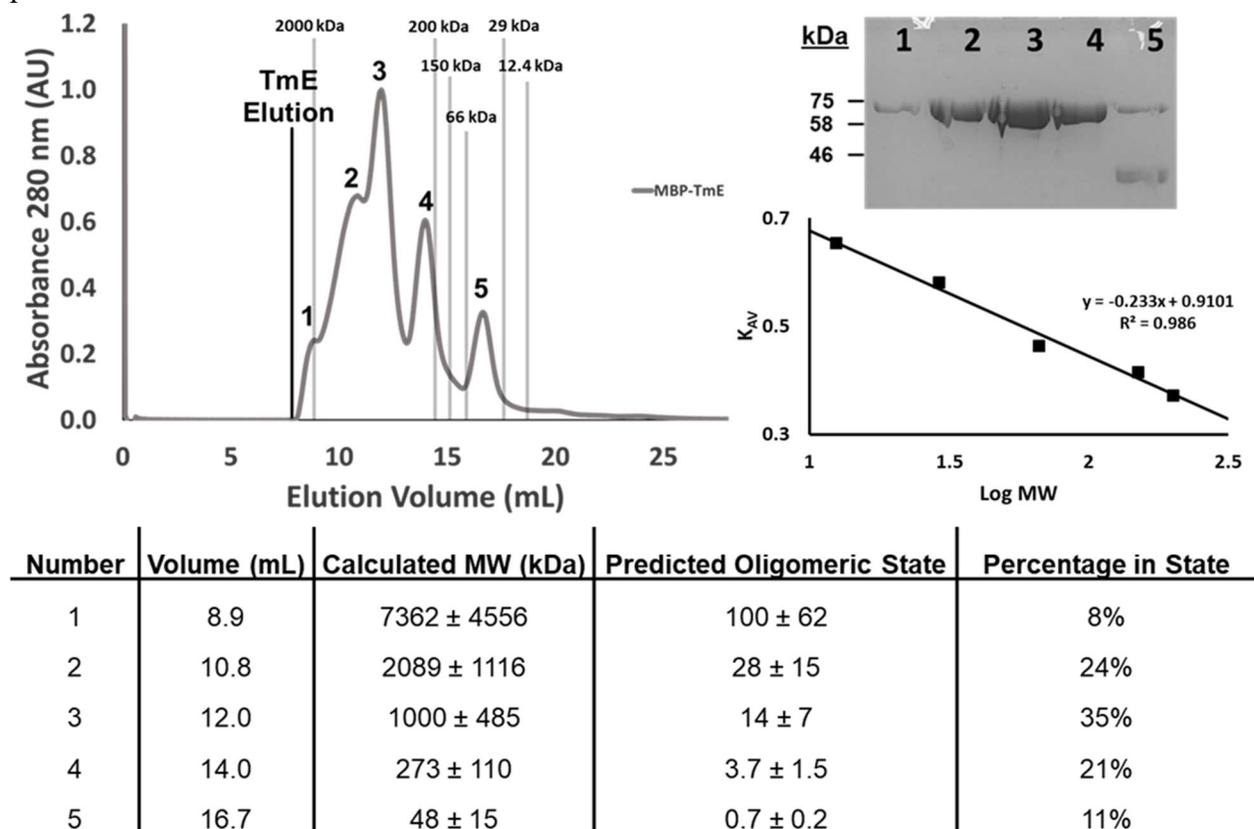


Figure 17. Top; Analytical SEC chromatogram of MBP-TmE with corresponding SDS-PAGE sample data. Protein standards of known molecular weights are superimposed on the MBP-TmE chromatogram and appropriately labeled. Plot of K_{AV} vs. LogMW shown with calibration curve and R^2 value. Bottom; tabulated calculations for each peak of MBP-TmE with associated errors.

From the SEC chromatogram, it appears that MBP-TmE does not exist in a single oligomeric state. Rather, it seems to be in an equilibrium involving multiple oligomeric states. This becomes increasingly apparent upon examination of the calculated values from the protein standard curve. From the SDS-PAGE gel, it is known that all peaks contain fusion protein. For peak 5, this fraction contains both fusion protein and MBP. The calculated MW is within error for both free MBP and a monomer of MBP-TmE (Fig. 17). This elution volume is expected of MBP, which has a molecular weight of ~ 43-kDa and a similar shape to the globular protein standards. (Rizk, *et al.* 2011) However, for MBP-TmE this means that a small percentage of the fusion protein exists as a monomer and has a slightly lower predicted MW. While size-exclusion primarily separates proteins based on size, the shape of the protein, or the Stokes radius of the analyte, can also influence retention time. (Hong, *et al.* 2012) Based on this knowledge, it is known that peak 5 must contain fusion protein, but due to its shape, it likely elutes at a lower predicted molecular weight.

The next most apparent oligomeric assignment is peak 4. From the standard curve, the most likely oligomeric state of peak 4 is a pentamer (Fig. 17). This assignment lies within margin of error and is a known oligomeric state of TmE based on the crystal structure (Sutter *et al.* 2008). Peaks 1-3 are larger assemblies that do not elute at the volume of fully-formed encapsulins, implying that they are not fully-assembled encapsulins but rather a larger oligomer than a pentamer. The MW error associated with each of these peaks is too high for reliable calculated predictions, but it must be outlined that the majority of MBP-TmE does not exist in a canonical oligomeric form. Only 32% of this fusion protein is either in a monomeric or pentameric state (Fig. 17).

As mentioned before, the fully assembled encapsulin consists of pentameric faces that make up the T=1 shell. However, the exact mechanism of shell assembly has yet to be resolved. Herein, I present a system that stalls shell-assembly at multiple oligomeric states, akin to kinetic checkpoints throughout the encapsulin assembly. It is unclear to what degree these states exist in equilibrium, however the relative ratios of protein concentrations determined from the SEC chromatogram reveals that the majority of these fusion proteins do not exist in their monomeric or pentameric state. Rather, they are found as larger assemblies akin to semi-formed encapsulins.

SEC is an imperfect analytical tool with several sources for error. Systematic error sources include changes in column backpressure from several purifications leaving residual protein and fluctuations in flow rate from column over-pressurization. The protein standards utilized are primarily globular proteins, which may not accurately reflect the retention volumes of non-globular proteins. Determining the oligomeric state of a protein has challenged the field, as there are no clearly established methodologies that correctly predict protein oligomerization. (Elcock, *et al.* 2001) One method of predicting MW is utilizing a protein standard kit that contains protein analytes better-suited for our particular SEC column. These standards are typically modeled by second-order, third-order, or logarithmic fits, as certain statistical fits are better suited for certain SEC columns. (Heyden, *et al.* 2002) Another method of interest utilizes analytical ultracentrifugation to identify oligomeric states through their sedimentation velocities. (Cole, *et al.* 2009) These methods would likely give a clearer image of the oligomeric states of MBP-TmE. Following the oligomeric state analysis of MBP-TmE, the next step was to produce a method of triggered, *in vitro* self-assembly that did not rely on encapsulin unfolding from extreme pH conditions or GuHCl treatment.

I call this method of in vitro shell formation directed compartment self-assembly (DCSA). The formation of nanocompartments is user-controlled and triggered through protease introduction

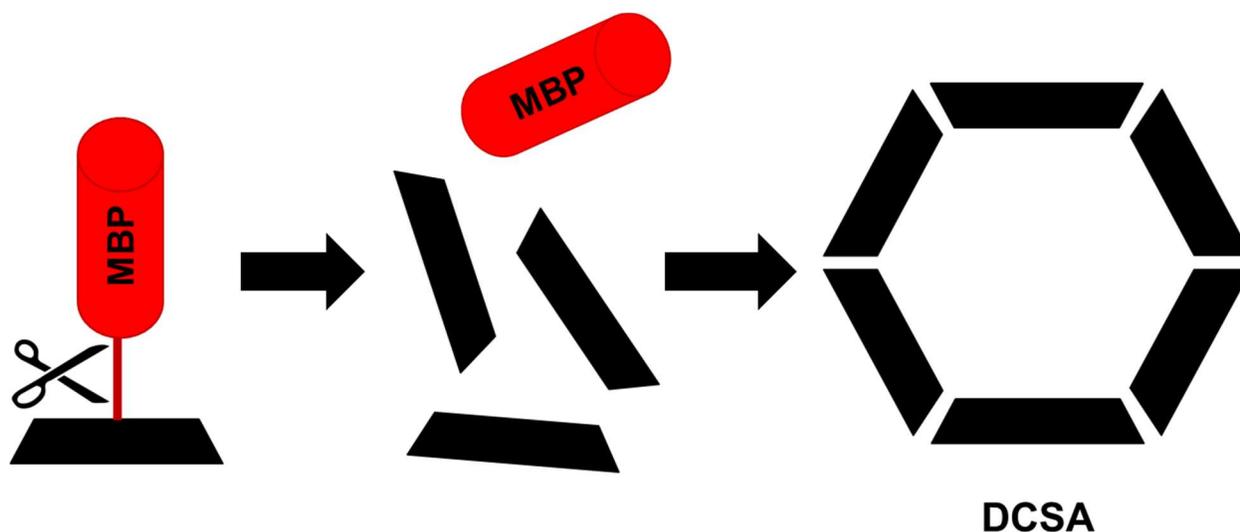


Figure 18. Cartoon representation of encapsulin self-assembly system that operates at physiological conditions: Directed Compartment Self-Assembly (DCSA).

that cleaves at the linker region of MBP-TmE. In the case of first-generation MBP-TmE, this was discovered serendipitously, as trypsin incubation proved to quickly degrade disassembled encapsulin subunits as seen throughout my chimeragenesis studies.

However, trypsin incubation instead cleaves at the linker portion of MBP-TmE, most likely at the exposed charged arginine residue present in the linker site. This, in turn, frees the MBP and allows for the encapsulin subunits to spontaneously self-assemble into icosahedral, regularly-sized T=1 nanocompartments. TEM and SDS-PAGE analysis reveals that these compartments are hollow and do not contain any freed MBP within the encapsulin lumen, meaning that this system is applicable to non-native cargo-loading techniques.

The Pre-Protease SEC chromatogram reveals that there are no fully formed encapsulins present prior to protease cleavage. This was confirmed via TEM analysis, as seen in Fig. 19. However, upon introduction of trypsin protease, the linker portion is severed, allowing the TmE subunits to

spontaneously self-assemble and form encapsulin nanocompartments. These compartments are of the same size and shape as the wild-type encapsulins, and they elute at the same elution volume. However, this method does suffer from protein loss, as the relative ratio of freed MBP is

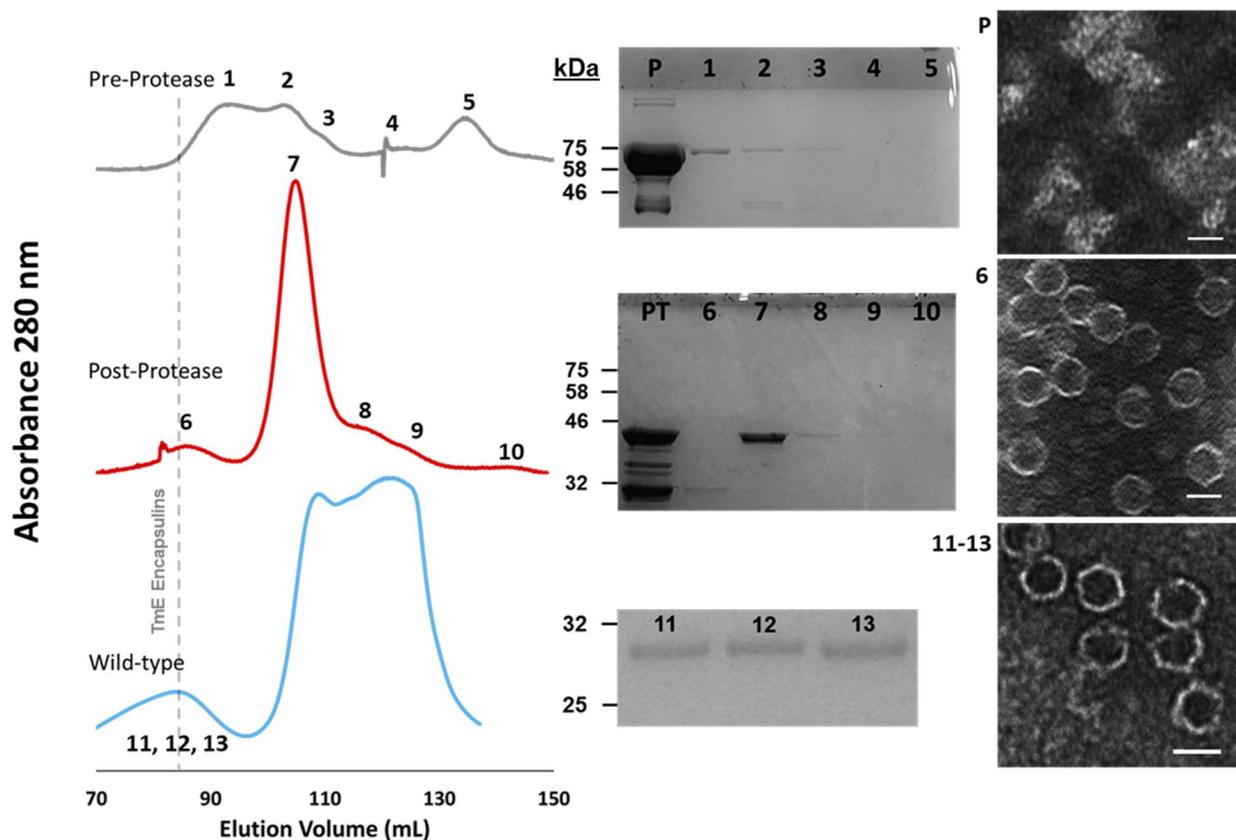


Figure 19. Left; SEC chromatograms of pre-protease, post-protease MBP-TmE and wild-type TmE. Middle; SDS-PAGE analysis of corresponding SEC peaks. P = Pooled amylose column fractions, PT = Pooled amylose post-trypsin incubation at 37 °C for 15 mins. Right; TEM analysis of P = Pooled amylose column fractions (pre-protease), 6 = Post-protease protein peak that overlaps with wild-type TmE encapsulin, and 11-13 = wild-type TmE encapsulins. Scale bar = 25 nm.

much higher the fully formed encapsulins (Fig. 19). This is likely due to the non-specificity of trypsin protease and its rapid turnover rate, as protease-susceptible sidechains of TmE are likely solvent-exposed prior to DCSA. This means that the oligomeric forms of MBP-TmE are likely susceptible to trypsin degradation, necessitating another form of triggered self-assembly.

Since the Lutz lab has established methods of *in vivo* encapsulation using native TmE and GCaMP-Clp, this reporter was selected for *in vitro* encapsulation studies. The fluorescence of

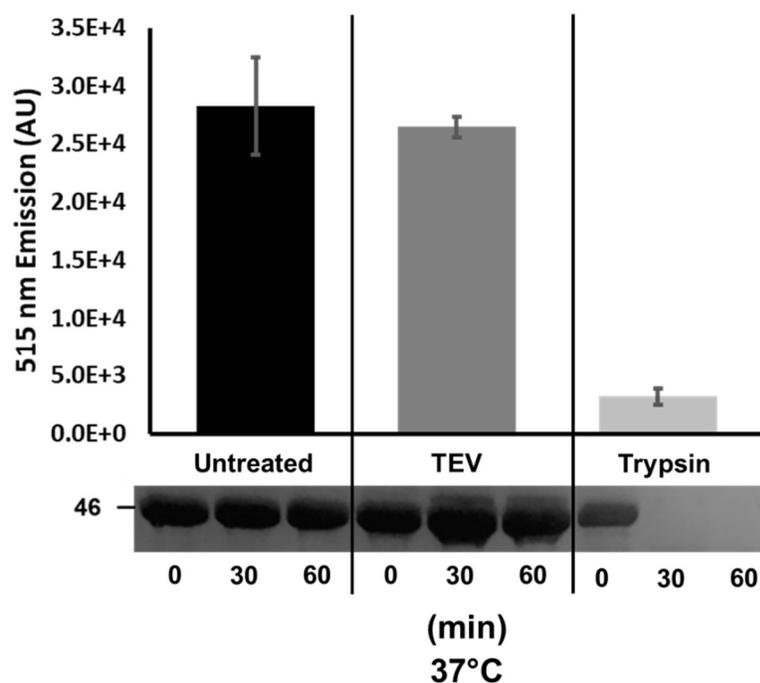


Figure 20. Top; fluorometer assays of untreated, TEV treated, and trypsin treated GCaMP-Clp. Bottom; SDS-PAGE analysis of untreated, TEV, and trypsin treated GCaMP-Clp at 0, 30, and 60 mins of incubation at 37 °C

GCaMP-Clp can be monitored by introducing Ca^{2+} and measuring its 515 nm emission profile.

Preliminary protease sensitivity assays revealed that GCaMP-Clp was susceptible to trypsin treatment but not TEV treatment (Fig. 20). This study was conducted with untreated, TEV treated, and trypsin treated

GCaMP-Clp and verified with SDS-PAGE analysis. This

therefore establishes a method for in vitro encapsulation of GCaMP-Clp that will not affect the reporter protein's integrity, while also providing a mechanism for encapsulation verification.

Since fully-assembled encapsulin is resistant to trypsin protease degradation, it should confer this

resistance to GCaMP-Clp. The

MBP-TmE linker region was

modified to introduce a TEV

cleavage site, a construct called

MBP-TEV-TmE. Upon incubation

with 0.15 U/mL TEV protease

under various reaction conditions, the

construct remained almost entirely as

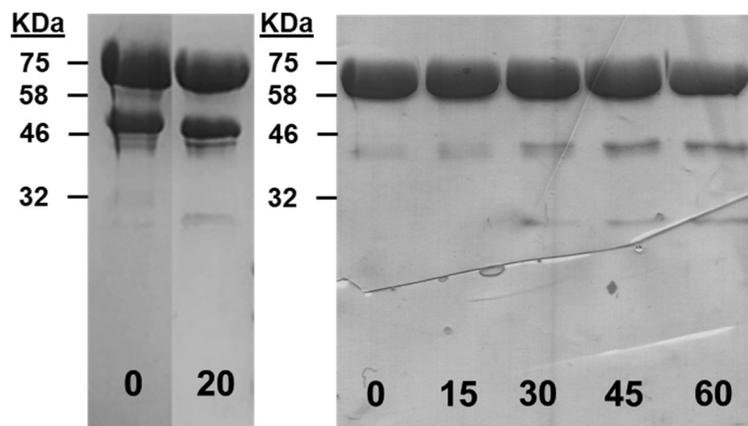


Figure 21. TEV cleavage tests of MBP-TEV-TmE. Left; incubation with TEV protease at 4 °C for 0 to 20 hrs. Right; Incubation with TEV protease with 15 min timepoints for 1 hr at 37 °C.

the fusion protein. It does appear that a higher proportion of fusion protein is freed when incubated at 37 °C for 1 hr as opposed to 4 °C overnight, however both of these treatments yield insignificant amounts of 31-kDa protein. This lack of cleavage could be due to a variety of reasons. Compared to trypsin ($k_{\text{cat}} = 74.2 \text{ s}^{-1}$), TEV protease exhibits significantly lower catalytic activity ($k_{\text{cat}} = 0.16 \pm 0.01 \text{ s}^{-1}$) likely due to its increased specificity. (Evnin, *et al.* 1990; Kapust, *et al.* 2001) This results in poor substrate turnover and is likely accentuated by steric occlusion of the cleavage site presented by the multimeric assemblies of MBP-TmE. Additionally, the cut site is located only 4 amino acids away from the N-terminus of TmE, meaning that this protease can easily be obstructed from its target cut site.

Conclusions and Future Work

The generation of TmE-MxE loop chimeras proved to be difficult and detrimental to compartment integrity. While conserved glycine and proline residues were identified as stitching points between E-loops, this in practice resulted in unstable protein aggregates that were unable to form symmetrical shells. Furthermore, large quantities of protein chimeras existed as subunits that were easily degraded by non-specific protease treatment, or lost as aggregates through PEG precipitation. The most successful variant was TmE-MxE loop 1, which took a more conservative stitching approach. In the future, the most successful protein chimeras will likely result from more conservative E-loop swaps that seek to preserve more of the native amino acid sequence.

The MBP-TmE fusion protein demonstrates promising developments in the areas of in vitro cargo loading and self-assembly mechanistic studies. Through analytical SEC, it was determined that in vivo shell assembly is arrested through N-terminal fusion of MBP. This, however, does not impede the formation of different oligomeric states of MBP-TmE, with predicted

monomeric, and pentameric intermediates forming, along with larger assembly states.

Furthermore, these samples can be easily purified in high quantities for structural studies.

Namely, the high purity of these arrested complexes could be sufficient for high resolution cryo-electron microscopy, providing structural information of shell formation that would otherwise be unattainable.

The MBP-TmE fusion protein also introduces a novel method of in vitro self-assembly that has not been previously reported in any literature. This method is called directed compartment self-assembly (DCSA). DCSA takes advantage of the arrested states of MBP-TmE to selectively trigger in vitro assembly through protease introduction. The protease cleaves at the linker region of the MBP-TmE, freeing the TmE subunits from any steric hindrance, and allowing them to self-assemble.

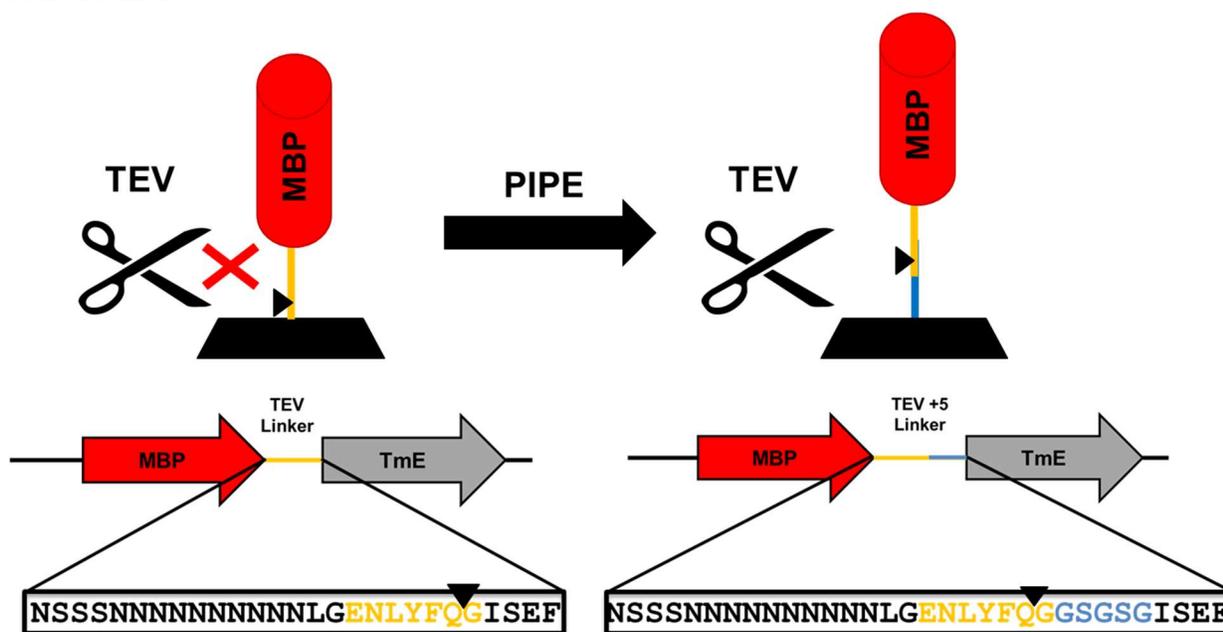


Figure 22. Cartoon schematic of re-engineered TEV site in MBP-TEV-TmE. TEV recognition site shown in yellow, proposed Gly-Ser additions shown in blue.

Due to the inefficiency of TEV protease to trigger DCSA, I will be working on cloning a hydrophilic linker extension between the protease cleavage site and the N-terminus of TmE. By

introducing a 5-amino acid Gly-Ser-Gly-Ser-Gly chain between the cut site and the encapsulin I hope to overcome any steric hindrance that is introduced by the multiple oligomeric states of MBP-TmE (Chen, *et al.* 2013). Upon testing successful cleavage of this construct, I will return to testing the cargo loading capabilities of this system with the established GCaMP-Clp reporter protein. Once this is verified, this system can be introduced as a strong, new method of in vitro encapsulation.

DCSA is particularly exciting as a novel method of in vitro encapsulation for three reasons. (1) This method is conducted at physiological pH 7.6, meaning that cargo protein can be present in the system without fear of disrupting or unraveling either the cargo protein or TmE monomers. These conditions are more favorable over the harsh conditions used previously to disassemble/reassemble fully formed encapsulins, which lead to significant monomer to compartment loss. (2) Non-traditional cargo proteins can be loaded through simple fusion of the Clp and will only be encapsulated following protease treatment. (3) The arrested oligomeric states of MBP-TmE could provide more time for the cargo protein to situate itself within the encapsulin lumen prior to DCSA, overcoming the kinetic barrier found in other methods of in vitro loading. This could lead to a more efficient method of in vitro self-assembly of cargo, and open the door to novel cargo introduction in a variety of biotechnological and biotherapeutic applications.

Primer Table**pET-23b Mx_E Traditional Cloning**

1) Mx_E Gene FOR (19-mer)

NdeI

5' - CA ▼TATGCCTCTGGAGCCAC - 3'

3' - GTAT ▲ACGGAGACCTCGGTG - 5'

2) Mx_E Gene REV (22-mer)

HindIII

5' - CACGGAACGTCGCTAGA A▼CTT - 3'

3' - GTGCCTTGCAGCGATCTTCGA A▲ - 5'

3) T7 FOR (19-mer)

5' - TAATACGACTCACTATAGG - 3'

4) T7 REV (19-mer)

3' - GCTAGTTATTGCTCAGCGG - 3'

pD434-SR Tm_E-Mx_E Loop Traditional Cloning via Overlap PCR

5) pD434-SR XbaI FOR (25-mer)

XbaI

5' - CAATTC~~C~~CCT ▼TAGAAATAATTTTG - 3'

3' - GTTAAGGGGAGATC ▲TTATTAAAC - 5'

6) pD434-SR BspHI REV (22-mer)

BspHI

5' - GTATCCGCT ▼ATGAGACAATAA - 3'

3' - CATAGGCGAGTAC ▲TCTGTTATT - 5'

7) Tm_E (-) Loop 1 FOR (20-mer)

5' - CCATCATGAGCTGCGCGG - 3'

8) Tm_E (-) Loop 1 REV (20-mer)

3' - GGGCATGCCAACCCATAG - 5'

9) Tm_E (-) Loop 2 FOR (20-mer)

5' - TCCCGCTGATTGAGCTGCGC - 3'

10) Tm_E (-) Loop 2 REV (19-mer)

3' - TGCAACTCCCGGCATGCC - 5'

11) Mx_E Loop 1 FOR (37-mer)

5' - CCCGTACGGTTGGGAGTATCAGACGGTTCCTACGAC - 3'

12) Mx_E Loop 1 REV (36-mer)

3' - CGAGCGTTCAAGTTCTGGTAGTAACTCGACGCGGC - 5'

13) MxE Loop 2 FOR (38-mer)
5' - ACGTTGAGGGCCCCTACGGTGCCGGGGTGCAGACGGTT - 3'

14) MxE Loop 2 REV (36-mer)
3' - GCGTTCAAGTCTGGTAGGGCGACTAACTCGACGCG - 5'

15) TmE-MxE Loop A47A49 FOR (44-mer)
5' - ACGTTGAGGGCCCCTACGGTGCCGCCGGGGCAGTGCAGACGGTT - 3'

16) TmE-MxE Loop A47A49 REV (44-mer)
3' - TGCAACTCCCGGGCATGCCACGGCGGCCCGTCACGTCTGCCAA

pMAL-c2X MBP-TmE Traditional Cloning

17) TmE Gene FOR (30-mer)
EcoRI
5' - CGGCG **▼**ATTTCATGGAATTTCTGAAACGCAG - 3'
3' - GCCGCTTAA **▲**CTACCTTAAAGACTTTGCGTC - 5'

18) TmE Gene REV (34-mer)
HindIII
5' - GCACTGATCCTGTTGAAATTCTAAA **▼**AGTTGCCG - 3'
3' - CGTGACTAGGACAACCTTTAAGATTTTCGA **▲**ACGGC - 3'

19) pMAL FOR (19-mer)
5' - CACCAACAAGGACCATAGC - 3'

20) pMAL REV (17-mer)
5' - GTTGGGTAACGCCAGGG - 3'

PIPE Cloning a TEV Recognition Site into pMAL-c2X MBP-TmE

21) TEV FOR 1 (37-mer)
5' - GTATTTTCAGGGCATTTCAGAATTCATGGAATTTCTG - 3'

22) TEV FOR 2 (37-mer)
5' - CCTCGGGGAAAACCTGTATTTTCAGGGCATTTCAGAA - 3'

23) TEV FOR 3 (36-mer)
5' - CAATAACAATAACAACAACCTCGGGGAAAACCTGTA - 3'

24) TEV REV (37-mer)
3' - CGTGACTAGGACAACCTTTAAGATTTTCGAACCGTGAC - 5'

25) pMAL noFXa FOR

(30-mer)

5' - GAAATTCTAAAAGCTTGGCACTGGCCGTCG - 3'

26) pMAL noFXa REV

(25-mer)

3' - GTTATTGTTATTGTTGTTGGAGCCC - 5'

References

1. 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.* **2007**, 368(5), 1469-1483.
2. Azuma, Y. *et al*; Modular Protein Cages for Size-Selective RNA Packaging in Vivo. *J. Am. Chem. Soc.* **2018**, 140(2), 566-569.
3. Brasch, M. *et al*; Assembling Enzymatic Cascade Pathways inside Virus-Based Nanocages Using Dual-Tasking Nucleic Acid Tags. *J. Am. Chem. Soc.* **2017**, 139(4), 1512-1519.
4. Cassidy-Amstutz, C. *et al*; Identification of a Minimal Peptide Tag for in Vivo and in Vitro Loading of Encapsulin. *ACS. Biochem.* **2016**, 55(24), 3461-3468.
5. Chen, T-W. *et al*; Ultra-sensitive fluorescent proteins for imaging neuronal activity. *Nature.* **2013**, 499(7458) : 295-300.
6. Chen, X. *et al*; Fusion Protein Linkers: Property, Design and Functionality. *Adv. Drug. Deliv. Rev.* **2013**, 65(10): 1357-1369.
7. Cole, J. *et al*; Analytical Ultracentrifugation: Sedimentation Velocity and Sedimentation Equilibrium. *Methods. Cell. Biol.* **2008**, 84: 143-179.
8. Elcock, A. *et al*; Identification of protein oligomerization states by analysis of interface conservation. *Proc. Natl. Acad. Sci. USA.* **2001**, 98(6): 2990-4.
9. Evnin, L.B. *et al*; Substrate specificity of trypsin investigated by using a genetic selection. *Proc. Natl. Acad. Sci. USA.* **1990**, (17):6659-6663.
10. Giessen, T. W. Silver, P. A; Widespread distribution of encapsulin nanocompartments reveals functional diversity. *Nat. Microbiol.* **2017**, 2:17029.
11. Heckman, K. *et al*; Gene splicing and mutagenesis by PCR-driven overlap extension. *Nat. Protoc.* **2007**, 2, 924-932.
12. Heyden, Y. *et al*; Evaluation of size-exclusion chromatography and size-exclusion electrochromatography calibration curves. *J. Chromatogr. A.* **2002**, 957, 127-137.
13. Hong, P. *et al*; Size-Exclusion Chromatography for the Analysis of Protein Biotherapeutics and their Aggregates. *J. Liq. Chromatogr. Relat. Technol.* **2012**, 35(20): 2923-2950.
14. Kapust, R.B. *et al*; *Escherichia coli* maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. *Protein. Sci.* **1999**, 8(8), 1668-74.
15. Kapust, R.B. *et al*; The P1' specificity of tobacco etch virus protease. *Biochem. Biophys. Res. Commun.* **2002**, 294, 949-955.

16. Kapust, R.B. *et al*; Tobacco etch virus protease: mechanism of autolysis and rational design of stable mutants with wild-type catalytic proficiency. *Protein. Eng.* **2001**, (12):993-1000.
17. Klock, H.E. *et al*; The Polymerase Incomplete Primer Extension (PIPE) method applied to high-throughput cloning and site-directed mutagenesis. *Methods. Mol. Biol.* **2009**, 498, 91-103.
18. McHugh, C. A. *et al*; A virus capsid-like nanocompartment that stores iron and protects bacteria from oxidative stress. *EMBO. J.* **2014**, 33(17), 1896-1911.
19. Nichols, R. *et al*; Encapsulins: molecular biology of the shell. *Crit. Rev. Biochem. Mol. Bio.* **2017**, 52(5), 583-594.
20. Prasad, B. V. Schmid, M. F; Principles of Virus Structural Organization. *Adv. Exp. Med. Biol.* **2012**, 726, 17-47.
21. Rahmanpour, R. Bugg, T. D. H; Assembly in vitro of *Rhodococcus jostii* RHA1 encapsulin and peroxidase DypB to form a nanocompartment. *FEBS. J.* **2013**, 280(9), 2097-2104.
22. Rizk, S.S. *et al*; Allosteric control of ligand-binding affinity using engineered conformation-specific effector proteins. *Nat. Struct. Mol. Biol.* **2011**, 18: 437-442.
23. Rurup, F.R. *et al*; Self-Sorting of Foreign Proteins in a Bacterial Nanocompartment. *J. Am. Chem. Soc.* **2014**, 136(10), 3828-3832.
24. Sutter, M. *et al*; Structural basis of enzyme encapsulation into a bacterial nanocompartment. *Nat. Struct. Mol. Biol.* **2008**, 15(9), 939-947.
25. Tamura, A. *et al*; Packaging Guest Proteins into the Encapsulin Nanocompartment from *Rhodococcus erythropolis* N771. *Biotechnol. Bioeng.* **2015**, 112(1), 13-20.
26. Tayyab, S. *et al*; Size exclusion chromatography and size exclusion HPLC of proteins. *Biochem. Educ.* **1991**, 19(3), 149-152.
27. Williams, E.M. *et al*; Pore Engineering for Enhanced Mass Transport in Encapsulin Nanocompartments. *ACS. Synth. Biol.* **2018**, 7(11), 2514-2517.